Univerzita Karlova v Praze
2. lékařská fakulta

Studijní program:
Molekulární a buněčná biologie, genetika a virologie

MUDr. Markéta Vlčková

Korelace genotyp-fenotyp u vybraných vzácných onemocnění s využitím molekulární analýzy genomových a genových variant

Genotype-phenotype correlation in selected rare disorders using molecular analysis of genome and gene variants

Disertační práce

Školitel: prof. Ing. Zdeněk Sedláček, DrSc.

Praha, 2016

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem řádně uvedla a citovala všechny použité prameny a literaturu. Současně prohlašuji, že práce nebyla využita k získání jiného nebo stejného titulu.

Souhlasím/~~Nesouhlasím~~* s trvalým uložením elektronické verze mé práce v databázi systému meziuniverzitního projektu Theses.cz za účelem soustavné kontroly podobnosti kvalifikačních prací.

V Praze, 11. 7. 2016 Markéta Vlčková

Identifikační záznam:

Vlčková, Markéta. Korelace genotyp-fenotyp u vybraných vzácných onemocnění s využitím molekulární analýzy genomových a genových variant. *[Genotype-phenotype correlation in selected rare disorders using molecular analysis of genome and gene variants]*. Praha, 2016. rozsah 177 s. Disertační práce (Ph.D.). Univerzita Karlova v Praze, 2. lékařská fakulta, Ústav biologie a lékařské genetiky 2. LF UK. Vedoucí práce/Školitel: Sedláček, Zdeněk.

Poděkování:

Ráda bych touto cestou poděkovala zejména prof. Ing. Zdeňkovi Sedláčkovi, DrSc. za jeho pečlivé, trpělivé a důsledné vedení mojí práce. Nemenší dík pak patří prim. MUDr. Markétě Havlovicové, prof. MUDr. Evě Seemanové, DrSc. a MUDr. Martině Simandlové za jejich rady a vedení v klinické praxi a zejména za předávání zkušeností v dysmorfologii. Děkuji rovněž MUDr. Markovi Turnovcovi a Mgr. Janu Gerykovi za rady a pomoc s bioinformatickou analýzou, IT podporu a pomoc při grafické úpravě práce, MUDr. Pavlovi Tesnerovi, Mgr. Miroslavě Hančárové, Ph.D. a Mgr. Darině Prchalové, Ph.D. za morální podporu. Dále bych chtěla poděkovat všem kolegům z laboratoří ÚBLG a kolegům ze spolupracujících pracovišť, bez jejichž úsilí a často dobré vůle by realizace většiny cílů nebyla možná.

Disertační práce vznikla za podpory:

00064203, CHERISH (EC FP7 223692), CZ.2.16/3.1.00/24022, NT/14200 a NFCZ11-PDP-3-003-2014

Abstrakt

Práce byla zaměřena na podrobnou analýzu pacientů se vzácnými genomovými a genovými variantami. Studován byl vliv těchto variant na fenotyp pacientů. Vzhledem k tomu, že většina našich pacientů, ať už syndromických nebo nesyndromických, byla k podrobnějšímu vyšetření indikována z důvodu mentální retardace a/nebo poruchy autistického spektra, byla práce zaměřena na tyto dvě klinické jednotky.

Nejprve byli k analýze vybráni pacienti s mikroskopicky detekovanými chromozomálními aberacemi, jejichž rozsah, genový obsah a mechanismus vzniku byly upřesněny pomocí molekulárně genetických metod, nejčastěji array CGH s vysokým rozlišením. Později byli analyzováni pacienti se vzácnými nebo jedinečnými submikroskopickými aberacemi, zjištěnými právě pomocí aCGH nebo SNP array. S využitím těchto metod jsme v průběhu práce analyzovali pacienty s delecemi Xp22.1-p22.3, 6q11-q13, 6q14-q16, Xq25, 1q21.1, Xp21.2-p21.3, 2p14-p15, 17q21.31, 9q21.3 a 2p15-p16.1, a pacientku s duplikací Xp21.2-p21.3.

V posledních letech jsme přistoupili k analýze syndromických pacientů metodami sekvenování nové generace. Tak byly odhaleny varianty v genech *HCFC1*, *KAT6B, SOS2* a *KMT2D*, které byly dále studovány.

Práce přispěla k poznatkům o vlivu nalezených genomových a genových variant na fenotyp pacientů, o mechanismech vzniku genomových variant, o roli jednotlivých zasažených genů v utváření fenotypu a rovněž o možnostech využití ale i limitech celogenomových metod.

Klíčová slova: array CGH, SNP array, sekvenování nové generace, vzácná onemocnění, mentální retardace, poruchy autistického spektra, bioinformatická analýza, korelace genotyp-fenotyp

Abstract

The work was focused on detailed analysis of patients with rare genomic and gene variants. We studied the impact of these variants on the phenotype of the patients. As the majority of our patients, both syndromic and non-syndromic, were reffered to the detailed analysis due to intellectual disability and/or autism spectrum disorder, the work was focused on these two clinical diagnoses.

At the beginning we analyzed patients with aberrations detected using cytogenetic analysis, and the extent, gene content and mechanism of origin of the aberrations were refined using molecular genetic methods, most often high-resolution array CGH. Later we analyzed patients with rare or unique submicroscopic aberrations detected using aCGH or SNP array. Using these methodes we analysed in the project patients with deletions of Xp22.1-p22.3, 6q11-q13, 6q14-q16, Xq25, 1q21.1, Xp21.2-p21.3, 2p14-p15, 17q21.31, 9q21.3 a 2p15-p16.1, and a patient with an Xp21.2-p21.3 duplication.

In the last years we proceeded to the analysis of syndromic cases using next generation sequencing. This led to the identification of point mutations in the *HCFC1*, *KAT6B, SOS2* and *KMT2D* genes, which were further studied.

The work contributed to the knowledge about the impact of the genome and gene variants identified on the phenotype of the patients, about the mechanisms of origin of the genome variants, about the role of individual affected genes in the phenotypes and also about the utility and limitations of the genome-wide methods.

Key words: array CGH, SNP array, next generation sequencing, rare diseases, intellectual disability, autism spectrum disorders, bioinformatic analysis, genotype-phenotype correlation

SEZNAM POUŽITÝCH ZKRATEK

1000GP	1000 Genomes Project
aCGH	array CGH
AD	autozomálně dominantní
ADHD	porucha s deficitem pozornosti a hyperaktivitou (attention deficit hyperactivity disorder)
AR	autozomálně recesivní
BAC	umělý bakteriální chromozom (bacterial artificial chromosome)
cDNA	komplementární DNA
CGH	komparativní genomová hybridizace (comparative genome hybridisation)
CNS	centrální nervový systém
CNP	polymorfismus v počtu kopií (copy number polymorphism)
CNV	varianta v počtu kopií (copy number variant)
DECIPHER	Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources
DGV	Database of Genomic Variants
DNA	deoxyribonukleová kyselina
ECARUCA	European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations
EVS	Exome Variant Server
ExAC	Exome Aggregating Consortium
FISH	fluorescenční *in situ* hybridizace
FoSTeS	fork stalling and template switching
GPS	genitopatelární syndrom
HbF	fetální hemoglobin
hUPD	uniparentální heterodizomie
CHARGE	Coloboma, Heart anomaly, choanal Atresia, Retardation, Genital and Ear anomalies
indel	inzerce a delece
IQ	inteligenční kvocient
ISCA	International Standards for Cytogenomic Arrays
iUPD	uniparentální isodizomie
kb	kilobáze
KS	Kabuki syndrom
LCRs	nízkofrekventní repetice (low-copy repeats)
LMD	London Medical Database
LOH	ztráta heterozygozity (loss of heterozygosity)
MAPK	mitogenem aktivované proteinkinázy
Mb	megabáze
mBAND	mnohobarevné pruhování chromozomů
mFISH	vícebarevná FISH (multicolour FISH)
MKN-10	mezinárodní klasifikace nemocí, 10. revize
MLPA	multiplex ligation-dependent probe amplification
MMBIR	microhomology-mediated break-induced replication
MR	mentální retardace
mRNA	messengerová RNA
NAHR	nealelická homologní rekombinace
NCBI	National Center for Biotechnology Information
NHEJ	nehomologní spojování konců (non-homologous end joining)
NHGRI	The National Human Genome Research Institute

NGS	sekvenování nové generace (next generation sequencing)
NMD	nonsense mediated RNA decay
NS	syndrom Noonanové
OMIM	Online Mendelian Inheritance in Man
PAS	porucha autistického spektra
PCR	polymerázová řetězová reakce (polymerase chain reaction)
POSSUM	Pictures Of Standard Syndromes and Undiagnosed Malformations
RNA	ribonukleová kyselina
ROH	oblast homozygozity (region of homozygosity)
SBBYSS	Say-Barber-Biesecker-Young-Simpsonův syndrom
SFARI	Simons Foundation Autism Research Initiative
SNP	jednonukleotidový polymorfismus (single nucleotide polymorphism)
SNPa	SNP array
SNV	jednonukleotidová varianta (single nucleotide variant)
SV	strukturní varianta
UCSC	The University of California Santa Cruz Genome Browser
UV	varianta nejasného významu (variant of unknown significance)
VCC	vrozená srdeční vada
VVV	vrozená vývojová vada
WBDD	The Winter-Baraitser Dysmorphology Database
WES	celoexomové sekvenování (whole exome sequencing)
WGS	celogenomové sekvenování (whole genome sequencing)
XL	X-vázaná dědičnost (X-linked inheritance)
XLMR	X-vázaná mentální retardace (X-linked mental retardation)
XLP-1	X-vázaný lymfoproliferativní syndrom 1

Obsah

1 Literární přehled

1.1 Úvod

Vzácná geneticky podmíněná onemocnění a syndromy mají prevalenci menší než 1 na 2000 osob, často ještě mnohem nižší. Je jich ale obrovský počet (v současné době je popsáno 5000 – 8000 různých jednotek), a proto se významně podílejí na celkové morbiditě a mortalitě v populaci. V Evropě postihují odhadem 27 – 36 milionů jedinců (Richter et al., 2015). Velmi důležitou skupinou vzácných onemocnění jsou choroby spojené s mentální retardací (MR) a poruchami autistického spektra (PAS), které jsou sice v populaci relativně časté, ale díky obrovské genetické heterogenitě se vlastně jedná o velkou skupinu vzácných onemocnění s různou genetickou etiologií, která se projevují podobným a klinicky obvykle nerozlišitelným fenotypem. Určení etiologie vzácných onemocnění je předmětem dlouhodobého zájmu genetiků a lékařů mnoha dalších odborností. Studium pacientů trpících těmito chorobami a jejich genetických variant také může pomoci objasnit funkci genů a jejich rolí v základních buněčných procesech.

Určit genetické příčiny vzácných onemocnění lze několika přístupy. Dříve byly k odhalení nových kauzálních genů využívány vazebné a asociační studie (Botstein a Risch, 2003; Szatmari et al., 2007; Holt et al., 2010). Rozvoj celogenomových molekulárně genetických metod, zejména array CGH (aCGH) nebo SNP array (SNPa) a sekvenování nové generace (NGS), v poslední době umožnil identifikaci genetické příčiny mnoha vzácných mikrodelečních a mikroduplikačních syndromů a monogenních chorob podmíněných nukleotidovými variantami bez nutnosti analyzovat velké soubory pacientů s podobným fenotypem či vyšetřovat rozvětvené rodiny (Ng et al., 2009; Kuhlenbäumer et al., 2011; Vandeweyer a Kooy, 2013).

Ve srovnání s dříve používanými cílenými vyšetřeními však nové celogenomové metody přinášejí řadu nejasných nálezů, které lze interpretovat pouze v souvislosti s přesně definovaným fenotypem pacienta. Řada pacientů navíc má více potenciálně patogenních nálezů. Detailní charakterizace fenotypu je tedy velmi důležitá a součinnost klinického genetika je dnes pro interpretaci nálezů zcela nezastupitelná. Přes všechno úsilí často nelze o roli daného nálezu nebo kombinace nálezů při formování fenotypu pacienta rozhodnout. To se týká jak komplexních syndromických postižení, tak případů nesyndromických, kde je diagnostika stejně důležitá, avšak vzhledem k absenci specifických či klíčových příznaků obvykle ještě mnohem obtížnější.

1.2 Variabilita lidského genomu

Genomy nepostižených lidských jedinců vykazují značnou variabilitu jak na úrovni větších genomových neboli strukturních variant (SV), jejichž zvláštní významnou podskupinu tvoří varianty v počtu kopií (CNV) (Zarrei et al., 2015), tak na úrovni genových či nukleotidových variant, ke kterým patří jednonukleotidové varianty (SNV) a několikanukleotidové delece a inserce (tzv. indel) (Wang et al., 1998). Z hlediska lokalizace lze rozlišit varianty vyskytující se v kódující oblasti a v oblasti nekódující, a z pohledu dědičnosti varianty zděděné a *de novo*.

Genetické varianty často nemají žádný dopad na fenotyp a jsou považovány za benigní změny či polymorfismy. Polymorfismy v počtu kopií nebo jednonukleotidové polymorfismy (CNP nebo SNP) bývají obvykle definovány jako odchylky od referenčního genomu, které mají frekvenci v populaci vyšší než 1%. Tato definice však může být problematická, protože této četnosti dosahují i některé patogenní varianty. Polymorfismy jsou katalogizovány v řadě různých databází (viz kap. 1.3.2), které slouží jako užitečný analytický nástroj při vyhodnocení výsledků molekulárně genetických vyšetření, zejména necílených (de Leeuw et al., 2012; Agarwala et al., 2016).

Pro lékařskou genetiku jsou nejdůležitější varianty charakteru CNV či SNV, které na fenotyp prokazatelně dopad mají. V odborné literatuře jsou pak často uváděny jako patogenní mikrodelece či mikroduplikace a nukleotidové mutace. V popředí zájmu genetiků jsou jednak jednoznačně patogenní, plně penetrantní varianty zodpovědné za konkrétní definované sporadické či hereditární syndromy, ale i varianty predisponující k některým patologickým symptomům (např. poruchy chování, obezita, autoimunitní onemocnění atd.), jejichž penetrance nemusí být úplná a jejichž expresivita je často variabilní. Rovněž informace o těchto variantách jsou dostupné v řadě různých databází (viz kap. 1.3.2).

Závažnou komplikací je skutečnost, že u některých variant nelze o jejich vlivu na fenotyp jednoznačně rozhodnout, např. právě v důsledku jejich neúplné penetrance, variabilní expresivity nebo velmi nízké frekvence v populaci. Takové varianty bývají označovány jako varianty s nejasným klinickým dopadem (variant of unknown significance, UV). Zvláště obtížné na interpretaci mohou být varianty zcela unikátní, u kterých v dostupné literatuře nelze dohledat žádný podobný případ. Pro rozhodnutí, zda varianta má či nemá dopad na fenotyp u konkrétního pacienta, je potřeba využít řady různých analytických nástrojů, podrobnou klinickou analýzu nevyjímaje. Přesto však závěr nemusí být jednoznačný.

1.2.1 Strukturní varianty

SV zahrnují nebalancované a balancované translokace či inverze a intersticiální či subtelomerické CNV. SV zasahují méně genů než početní aberace (aneuploidie a polyploidie, jejichž dopad na fenotyp je dlouhou dobu znám), mnohé z nich jsou unikátní a jejich dopad na fenotyp nemusí být vždy jasný. O kauzalitě mnoha SV nelze v současné době zodpovědně rozhodnout a zůstávají v kategorii UV. Pro jejich interpretaci je důležitá co nejpřesnější charakterizace obtíží a fenotypu pacienta, ale i reevaluace nejasných výsledků s časovým odstupem.

1.2.1.1 Balancované SV

Balancované translokace jsou relativně časté, s frekvencí asi 1/500 - 1/1000 novorozenců (Nielsen a Rasmussen, 1976; Jacobs et al., 1992). Nejčastěji jsou důsledkem chybné reparace chromozomálních zlomů. Většina z nich je zděděná a není asociována s patologickým fenotypem. Nosičství balancované translokace nicméně může způsobovat poruchy reprodukce, zejména vyšší četnost spontánních potratů (Tharapel et al., 1985). Méně časté jsou *de novo* balancované SV (u asi 1/2000 novorozenců), které však rovněž nemusí být asociovány s patologickým fenotypem. Pouze 6,1% - 9,4% nosičů má závažné zdravotní problémy jako například MR, PAS, vrozené vývojové vady (VVV) a další obtíže (Sismani et al., 2008; Feenstra et al., 2011). Někdy nelze rozhodnout, zda je balancovaná translokace příčinou obtíží pacienta, nebo pouze náhodným nálezem, který s obtížemi nijak nesouvisí.

Několik studií zkoumalo kauzalitu zdánlivě balancovaných translokací, zejména u pacientů s MR a PAS (Ciccone et al., 2005; Sismani et al., 2008; Schluth-Bolard et al., 2009; Feenstra et al., 2011). Postižení nosičů bylo vysvětleno několika způsoby. Část pacientů nesla submikroskopické CNV v místech zlomů (Schluth-Bolard et al., 2009) a jejich přestavby vlastně balancované nebyly. U některých pacientů mohl být fenotyp vysvětlen přerušením kauzálního genu zlomem, nebo přesunutím regulačních sekvencí. U některých nosičů nalezená SV s fenotypem nesouvisela a mohli nést jinou patogenní CNV nebo SNV, která jejich fenotyp vysvětlila lépe (Schluth-Bolard et al., 2009; Feenstra et al., 2011).

Analýza unikátních pacientů s balancovanými SV a dobře charakterizovaným fenotypem pomohla identifikovat nové kauzální či kandidátní geny (Kulkarni et al., 2008; Tabet et al., 2015).

1.2.1.2 Nebalancované SV, CNV

Již před nástupem celogenomových metod byla známa řada mikrodelečních a mikroduplikačních syndromů, z nichž nejčastější jsou uvedeny v Tabulce 1.

Tabulka 1: Přehled častých mikrodelečních syndromů

Syndrom	Kritická oblast	Prevalence*
DiGeorge	22q11.2	~ 1/2000-1/4000
Slavotinek	1p36	~ 1/5000-1/10000
Williams - Beuern	7q11.23	~ 1/20000
Prader - Willi	15q11-q13	~ 1/25000
Angelman	15q11-q13	~ 1/10000-1/20000
Smith - Magenis	17p11.2	~ 1/15000-1/25000
Miller - Dieker	17p13.3	~ 1/1000000
Phelan - McDermid	22q13.3	Neznámá*

**Údaje o prevalenci byly získány z www.orpha.net (květen, 2016)*

Rozvoj arrayových metod odhalil, že variabilita lidského genomu na úrovni CNV je značná a zahrnuje 4,8 – 9,5% genomu zdravých jedinců (Zarrei et al., 2015). Postupně jsou odhalovány i vzácnější mikrodeleční a mikroduplikační syndromy, jejichž seznam se neustále rozšiřuje (Slavotinek, 2008; Vissers a Stankiewicz, 2012; Weise et al., 2012; Nevado et al., 2014). Některé varianty, dříve považované za patogenní, byly na základě častých opakovaných nálezů u zdravých kontrol naopak přehodnoceny na benigní. Proto je možné nálezy hodnotit vždy pouze v souvislosti se znalostí klinického obrazu pacientů.

Nejčastějším mechanismem vzniku rekurentních a tedy častých a klinicky dobře popsaných CNV je nealelická homologní rekombinace (NAHR). Ta je často podmíněna lokální architekturou genomu, zejména přítomností repetitivních sekvencí jako jsou low-copy repeats (LCRs) (Lupski a Stankiewicz, 2005). Dalším mechanismem je nehomologní spojování konců (non-homologous end joining, NHEJ), především u nerekurentních CNV. V procesu hrají roli dvouřetězcové zlomy, při jejichž reparaci často dojde k inzerci několika nukleotidů v místě zlomu. CNV také mohou vznikat v důsledku pozastavení či kolapsu replikační vidlice a záměny templátu – vmezeření nového řetězce do vidlice (fork stalling and template switching, FoSTeS) nebo mikrohomologií zprostředkovanou replikací indukovanou zlomem (microhomology-mediated break-induced replication, MMBIR) (Lee et al., 2007).

1.2.2 SNV a několikanukleotidové varianty

K těmto typům variant patří záměna jednoho nukleotidu a delece či inzerce jednoho nebo několika nukleotidů (indel). Kromě exogenních faktorů se na jejich vzniku podílejí především chyby při replikaci či reparaci DNA a spontánní deaminace metylovaného cytosinu.

Tyto drobnější varianty lze opět hodnotit z několika různých hledisek. V rutinní praxi je u předpokládaných AD onemocnění pozornost věnována tomu, zda je SNV *de novo* či zděděná (a pokud je zděděná, tak zda od symptomatického či asymptomatického rodiče, tedy zda segreguje či nesegreguje s fenotypem v rodině), v případě AR chorob je důležité, zda se nalezené varianty nacházejí v pozici *cis* nebo *trans* (k ozřejmení pozice je opět nutno testovat oba rodiče pacienta). U SNV se rovněž hodnotí jejich předpokládaný vliv na funkci proteinu (viz kap. 1.4.1).

Předmětem zájmu klinických genetiků jsou především známé patogenní varianty. Pro hodnocení kauzality variant pro pozorovaný fenotyp je však neméně důležité i studium běžných polymorfismů a jejich katalogizace. Poznatky o variabilitě genomu na úrovni SNP začaly přibývat po dokončení projektu sekvenování lidského genomu po roce 2001 (Collins et al., 2003) a ještě výrazněji po zavedení metod NGS.

1.3 Metody analýzy variability genomu

Laboratorní metody použité pro odhalování variant musí být následovány bioinformatickou analýzou, která je náročná zejména u celogenomových postupů. Větší pozornost je věnována analýze CNV, neboť ta byla pro potřeby práce prioritní. Metody NGS a analýza SNV byly z důvodu jejich dřívější nedostupnosti využity v menší míře.

1.3.1 Laboratorní analýza

Laboratorní metody lze principiálně rozdělit na celogenomové postupy (karyotypizace, aCGH, SNPa a celoexomové a celogenomové sekvenování (WES a WGS)), které se snaží identifikovat varianty v celém genomu, a na cílené postupy, které se soustřeďují na určitý gen, skupinu genů nebo segment genomu.

1.3.1.1 Celogenomové postupy

1.3.1.1.1 Karyotypizace

K vyšetření karyotypu se v současné době využívá metoda G-pruhování. Přestože je metoda známa od 70. let 20. století, má stále své místo jak v diagnostice, tak ve výzkumu. Standardní cytogenetické vyšetření s pomocí G-pruhování umožní (kromě numerických aberací) detekovat delece a duplikace o rozsahu přes 5 - 10 megabází (Mb) DNA a některé balancované přestavby (Speicher a Carter, 2005). Karyotypizace je ve výzkumu etiologie vzácných onemocnění dosud často využívána jako „preselekční" test a pacienti s normálním karyotypem jsou dále vyšetřováni metodami s vyšším rozlišením. Stejně tak mohou být dále vyšetřováni pacienti se zdánlivě balancovanou aberací, u nichž může být přítomna submikroskopická CNV v místě zlomu, nebo CNV či SNV jinde v genomu (viz kap. 1.2.1.1). V neposlední řadě lze k dalšímu vyšetření indikovat pacienty s mikroskopicky detekovanou aberací, u nichž je třeba zjistit přesný rozsah aberace a identifikovat zasažené geny. V některých případech je možno např. metodou NGS identifikovat bodovou mutaci na druhé alele příslušného genu zasaženého delecí a tím vysvětlit fenotyp pacienta.

1.3.1.1.2 Array CGH, SNP array

Metoda aCGH byla představena v roce 1997 a původně sloužila k vyšetřování nádorových tkání (Solinas-Toldo et al., 1997; Pinkel et al., 1998). Záhy se však uplatnila jako efektivní nástroj v diagnostice genetických příčin MR, PAS a dalších onemocnění (South a Brothman, 2011). Metoda je založena na hybridizaci DNA pacienta a DNA normální kontroly značených dvěma různými barvivy k referenčním sondám fixovaným na DNA čipu. S pomocí scanneru se pak vyhodnocuje poměr signálů DNA pacienta vůči kontrole a data jsou zpracována softwarem. Vyšetření slouží k detekci submikroskopických CNV (i když zachytí i aneuploidie) a není jím možno detekovat balancované aberace. Rozlišovací schopnost se liší dle platformy od přibližně 2 Mb (např. BAC aCGH) až po několik kilobází (kb) u chromozom-specifických oligonukleotidových aCGH. Volba platformy záleží na účelu vyšetření – pro rutinní diagnostiku se obvykle využívají platformy s nižším rozlišením, ale s lepším pokrytím oblastí známých mikrodelečních/mikroduplikačních syndromů, pro výzkum a pro přesné určení pozice zlomů je vhodné zvolit rozlišení vyšší.

Metoda SNPa je založená na detekci SNP a zjišťuje genotyp vyšetřované DNA ve statisících až miliónech SNP pomocí hybridizace vyšetřované DNA na čip se zakotvenými oligonukleotidovými sondami (Gijsbers et al., 2009). Stejně jako aCGH i SNPa slouží k detekci CNV. Kromě CNV lze metodou SNPa identifikovat i uniparentální isodizomii (iUPD) nebo oblasti homozygozity (ROH), které mohou poukázat na konsangvinitu rodičů a tudíž možnou autozomálně recesivní (AR) příčinu obtíží u pacienta. Mapováním ROH lze také identifikovat nové kandidátní geny (Vissers et al., 2016). Metodou SNPa lze rovněž určit zygozitu dvojčat a uniparentální heterodizomii (hUPD), pokud jsou vyšetřeni i rodiče.

1.3.1.1.3 Sekvenování nové generace

Vývoj NGS byl iniciován snahou zlevnit a zefektivnit sekvenování lidského genomu. První platforma NGS byla představena v roce 2005 firmou 454 Life Sciences (nyní Roche) (Margulies et al., 2005; van Dijk et al., 2014). V následujících letech bylo představeno několik dalších platforem, využívajících rozličných přístupů (van Dijk et al., 2014). Metodami NGS je možno analyzovat celý genom v jednom experimentu.

V lékařské genetice je možné provést WES či WGS u pacientů vybraných podle dobře definovaného fenotypu resp. s podezřením na konkrétní klinicky definovanou jednotku, u které doposud nebyl objeven kauzální gen či geny. První asociace genů

s konkrétními syndromy s využitím WES byly publikovány v roce 2010 (Hoischen et al., 2010; Lalonde et al., 2010; Ng et al., 2010). Tento přístup je náročný, jak z hlediska výběru pacientů, tak z hlediska následné analýzy získaných dat, ale umožňuje identifikovat nové kauzální geny, pokud je kohorta pacientů dobře vybraná a výsledky jsou konzistentní (Clayton-Smith et al., 2011; Campeau et al., 2012). Další možností je provedení WES či WGS u velkého množství pacientů s méně specifickým fenotypem a bez podezření na konkrétní jednotku (tzv. přístup „genotype first“). Pacienty lze indikovat i neselektivně (indikačním kritériem může být například jen MR či PAS bez další specifikace či dalších sdílených symptomů). Následná analýza dat je však náročnější než v případě předchozím a spoléhá především na nález variant v jednom genu u několika nepříbuzných pacientů a na zpětném posouzení jejich fenotypových shod.

V některých případech lze i NGS považovat za cílenou metodu. Jedná se o situace, kdy jsou k vyšetření indikováni pacienti s podezřením na konkrétní, již známé a dobře definované onemocnění s heterogenní etiologií a několika známými kauzálními geny nebo velmi velkými geny, jejichž analýza klasickým sekvenováním je pracná a nákladná. Tento přístup vyžaduje pečlivou selekci pacientů dle fenotypu a splnění indikačních kritérií, pokud existují. Podobně lze postupovat u klinicky méně precizně definovaných onemocnění se známou vysokou heterogenitou čítající desítky či stovky kauzálních genů. V těchto případech se vyšetřuje širší panel genů asociovaných s chorobou, který může být připraven komerčně nebo dle vlastního návrhu laboratoře. Výhodou cílených postupů u NGS je vyšší pokrytí, absence náhodných nálezů (tzv. incidentalomu) a snazší interpretace výsledků, neboť jsou zahrnuty pouze známé kauzální geny. Nálezy mohou podporovat kauzalitu variant v genech, u kterých je dosud popsáno jen malé množství pacientů. Podrobný klinický popis jednotlivých pacientů přispívá k poznatkům o variabilitě fenotypu, penetranci a prognóze.

S využitím panelů genů lze vyšetřit i pacienty s MR a PAS. Selekce pacientů podle fenotypu nemusí být tak striktní jako v předchozích případech. Interpretace nálezů je však vzhledem ke značné heterogenitě a absenci dalších signálních či klíčových příznaků u nesyndromických MR/PAS náročnější. Vzhledem k přibývajícím poznatkům se design panelů zaměřených na MR a PAS často mění a starší panely nepokrývají všechny známé geny.

1.3.1.2 Cílené postupy

1.3.1.2.1 Fluorescenční in situ hybridizace

Fluorescenční in situ hybridizace (FISH) je metoda založená na hybridizaci fluorochromem značené sondy s vyšetřovanou DNA v interfázních či metafázních jádrech fixovaných na sklíčku. Signál je následně odečten za pomoci fluorescenčního mikroskopu. Rozlišovací schopnost se pohybuje mezi 50 – 200 kb. Do praxe byla metoda zavedena v roce 1986 (Pinkel et al., 1986). Multicolour FISH (mFISH) a mnohobarevné pruhování (mBAND) jsou variantami metody FISH využívajícími směsi kombinatoricky značených malovacích sond. K odečtu se používá mikroskop s více barevnými filtry a data jsou hodnocena softwarem (Pinkel et al., 1986; Speicher a Carter, 2005; Weise et al., 2009).

FISH se využívá zejména k detekci mozaicismu. Jakožto rychlé a spolehlivé vyšetření má své místo i v prenatální a preimplantační diagnostice a v onkocytogenetice. mFISH je využívána nejvíce v onkocytogenetice a k identifikaci marker choromosomů. Pomocí mBAND lze zpřesnit rozsah delecí a duplikací identifikovaných standardním cytogenetickým vyšetřením, nicméně v současnosti jsou pro vyšší rozlišení používány spíše čipové metody. Metody FISH jsou dnes využívány také pro konfirmaci nálezů z čipových metod a k cílenému vyšetření příbuzných, například rodičů k ověření heredity aberace. Pro větší aberace je FISH spolehlivější než metoda multiplex ligation-dependent probe amplification (MLPA, viz níže) (Kim et al., 2015)), a na rozdíl od MLPA existují komerčně dostupné sondy FISH pokrývající prakticky celý genom.

1.3.1.2.2 Multiplex ligation-dependent probe amplification

MLPA byla poprvé představena v roce 2002 (Schouten et al., 2002). Jedná se o semikvanititativní metodu pro detekci drobných delecí či duplikací na úrovni exonů či genů. Metoda je založena na amplifikaci specifických sond, které hybridizují k vyšetřované DNA, přičemž se amplifikují pouze sondy, které úspěšně hybridizovaly. Kvantita amplifikací získaného produktu odráží počet kopií přítomných ve vyšetřované DNA. Hodnotí se relativní poměr ploch peaků proti kontrolnímu vzorku (Schouten et al., 2002; Sellner a Taylor, 2004). Pro svou finanční a technickou nenáročnost MLPA postupně nahradila dříve využívané metody (Schouten et al., 2002; Sellner a Taylor, 2004), a slouží jako efektivní konfirmační test. Má však i své limity. V důsledku

přítomnosti vzácných SNV v místě nasedání primerů mohou být výsledky vyšetření falešně pozitivní či falešně negativní (Sellner a Taylor, 2004; Kim et al., 2015).

1.3.1.2.3 Polymerázová řetězová reakce a sekvenování dle Sangera

Polymerázová řetězová reakce (PCR) je známá od 80. let minulého století a je masivně využívána v rutinní diagnostice. Amplifikovaný cílový úsek DNA lze dále analyzovat pomocí elektroforézy (detekce přítomnosti/nepřítomnosti produktu či změny jeho délky) nebo sekvenování dle Sangera (identifikace nukleotidových variant). Sekvenování bylo představeno v 70. letech minulého století (Sanger et al., 1977). Metoda slouží k detekci SNV a variant typu indel. Jedná se o metodu cílenou a nákladnou, což značně omezuje její efektivní využití u heterogenních jednotek s klinicky obtížně odlišitelným fenotypem. Možné využití ve výzkumu je například jako druhý krok po NGS analýze souboru pacientů s konzistentním fenotypem či přesvědčivou klinickou diagnózou (v prvním NGS poukáže na možný kandidátní gen, ve druhém kroku je sekvenováním dle Sangera vyšetřen vytipovaný kandidátní gen u větší kohorty pacientů) (Ng et al., 2009; Hoischen et al., 2010; Clayton-Smith et al., 2011). Sangerovo sekvenování slouží i jako konfirmační vyšetření nálezů z NGS.

1.3.2 Bioinformatická analýza

Moderní laboratorní celogenomové vyšetřovací metody poskytnou velké množství dat. Bioinformatická analýza je nutná pro identifikaci variant v hrubých datech, jejich anotaci a především pro interpretaci jejich kauzality pro fenotyp pacienta. Nástrojů a databází využitelných pro bioinformatickou analýzu existuje celá řada. Možnosti využití jednotlivých databází, jakož i jejich limity, jsou shrnuty např. v těchto pracích (South a Brothman, 2011; de Leeuw et al., 2012).

1.3.2.1 Databáze SV a CNV

Database of Genomic Variants (DGV, http://dgv.tcag.ca/dgv/app/home) obsahuje přehled benigních CNV a inverzí. Data jsou získávána z populačních studií, které splňují kvalitativní požadavky správců databáze. Databáze je pravidelně aktualizována, některé (zejména starší) záznamy jsou přehodnocovány. Vzhledem k tomu, že CNV uvedené v této databázi jsou konsenzuálně považovány za benigní varianty, má tato databáze nezastupitelné místo při vyhodnocení výsledků získaných metodami aCGH, SNPa a NGS (MacDonald et al., 2014).

Databáze dbVar (http://www.ncbi.nlm.nih.gov/dbvar) shromažďuje informace o strukturních variantách genomu včetně inzercí mobilních elementů, translokací

a komplexních chromozomálních přestaveb. Do této databáze je sice možné posílat varianty o libovolné velikosti, ale doporučuje se varianty menší než 50 nukleotidů posílat do databáze dbSNP (Church et al., 2010).

European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA, http://umcecaruca01.extern.umcn.nl:8080/ecaruca/ecaruca.jsp) je databáze vytvořená v roce 2003. Obsahuje údaje o pacientech s aberacemi detekovanými jak karyotypováním, tak i čipovými metodami. Detailní klinická data pacientů jsou upravována dle Winter-Baraitser London Dysmorphology Database (WBDD). Nálezy zde uváděné jsou považovány za kauzální pro fenotyp pacientů. Vzhledem k tomu, že databáze propojuje cytogenetické nálezy s fenotypy pacientů, ji lze využít při hodnocení kauzality nálezů u dosud nediagnostikovaných pacientů a pro porovnání fenotypů (Vulto-van Silfhout et al., 2013).

Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER, https://decipher.sanger.ac.uk/) byla vytvořena v roce 2004 a je zdrojem informací zejména o potenciálně patogenních submikroskopických CNV. Kromě popisu CNV obsahuje i klinické údaje o pacientech, někdy i fotodokumentaci. Data do databáze poskytuje síť vybraných spolupracujících genetických pracovišť. Stejně jako v případě ECARUCA je využití databáze DECIPHER při rozhodování o kauzalitě jednotlivých CNV velmi přínosné (Firth et al., 2009).

1.3.2.2 Databáze SNV a indel

Databáze dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) je databáze provozovaná National Center for Biotechnology Information (NCBI) a National Human Genome Research Institute (NHGRI). Databáze neobsahuje pouze SNP, ale je možné v ní nalézt i informace o dalších typech variant - indel, mikrosatelitových markerech (short tandem repeats, STR), vícenukleotidových polymorfismech a dalších. Obsahuje i informace o variantách u dalších organismů. Do databáze jsou přijímány jak neutrální varianty, tak varianty související se změnou fenotypu (Sherry et al., 2001).

ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) je databází shromažďující informace o konkrétních genetických variantách a jejich vztahu k fenotypu. Databáze je úzce provázána s databázemi dbSNP, dbVar a některými dalšími databázemi a zdroji provozovanými NCBI (Landrum et al., 2014). Obě databáze, dbSNP i ClinVar jsou volně přístupné, nové informace do nich může poslat jakákoliv laboratoř na světě.

K dalším v současné době využívaným databázím patří databáze 1000 Genomes Project (1000GP, http://www.1000genomes.org/). 1000GP si dal za cíl

vytvořit databázi variant s frekvencí vyšší než 1%. Ve skutečnosti data pocházejí z 2504 vzorků, které reprezentují 26 různých populací z Evropy, Asie, Ameriky i Afriky. Projekt probíhal mezi lety 2008 až 2015, dnes další udržování a rozvoj finančně podporuje nadace Wellcome Trust. V rámci bioinformatických analýz se tato data používají k filtraci variant na základě populačních frekvencí (frekvenční filtr) (Auton et al., 2015; Birney a Soranzo, 2015; Sudmant et al., 2015).

Exome Aggregation Consortium (ExAC, http://exac.broadinstitute.org/) se snaží agregovat data, harmonizovat postupy exomového sekvenování a souhrnná data dávat dále k dispozici vědecké komunitě. Nyní projekt nabízí data pocházející z více než 60 tisíc nepříbuzných pacientů, která byla získána z různých projektů orientovaných na konkrétní typy onemocnění, nebo i populačních studií (Exome Aggregation Consortium et al., 2015).

Exome Variant Server (EVS, http://evs.gs.washington.edu/EVS/) poskytuje data získaná v rámci Grand Opportunity Exome Sequencing Project, na kterém pod vedením National Heart, Lung and Blood Institute pracovalo mnoho dalších univerzitních pracovišť v USA. Hlavím cílem bylo objevit pomocí WES nové geny a mechanismy přispívající k onemocněním v kardiologii, pneumologii a hematologii u dobře fenotypicky popsaných populací evropského a afrického původu. Databáze neobsahuje informace o individuálních genotypech či fenotypech (Tennessen et al., 2012).

1.3.2.3 Databáze chorob, fenotypů a kauzálních genů

Online Mendelian Inheritance in Man (OMIM, http://www.omim.org/) obsahuje katalog genů a asociovaných genetických chorob a syndromů. V databázi lze najít podrobné klinické informace o jednotlivých monogenních onemocněních, ale i o mikrodelečních syndromech a predisponujících alelách multifaktoriálních chorob. Databáze obsahuje informace o více než 12000 genech a je zaměřena na korelaci mezi genotypem a fenotypem. Získat lze i informace o typu dědičnosti onemocnění. Data vycházejí z recenzovaných publikací a jsou pravidelně aktualizována. Jedná se tedy o cenný zdroj informací zejména při hodnocení nálezu variant u pacientů se suspekcí na monogenní onemocnění (Amberger et al., 2009).

London Medical Database (LMD, http://www.lmdatabases.com/) je klinicky orientovaná databáze zaměřená na monogenní syndromy. Databáze je rozdělena do 3 sekcí. Pro účely této práce byla využita WBDD, která obsahuje informace o více než 5500 syndromických jednotkách. Kromě monogenních onemocnění obsahuje databáze rovněž informace o jednotkách způsobených vlivy prostředí, mikrodelečních

a mikroduplikačních syndromech, syndromech způsobených UPD a o onemocněních, která doposud nemají známou příčinu. Databáze obsahuje bohatou fotodokumentaci, jež umožňuje porovnání fenotypů. U každého syndromu je uveden stručný popis, výčet klinických příznaků, typ dědičnosti, genetická příčina (je-li známa) a odkazy na příslušné publikace. Databáze umožňuje vyhledávání syndromů podle klíčových příznaků a je užitečným nástrojem v diferenciální diagnostice (Winter et al., 1984).

Pictures of Standard Syndromes and Undiagnosed Malformations (POSSUM, http://www.possum.net.au/) je rovněž klinická databáze. V současné době obsahuje informace o přibližně 4000 syndromech nejrůznější etiologie. Databáze rovněž obsahuje bohatou fotodokumentaci, souhrnné informace o jednotlivých syndromech a umožňuje vyhledávání podle klíčových příznaků (Pelz et al., 1996).

Orphanet (http://www.orpha.net) je portál obsahující informace o vzácných onemocněních. Jedná se o mezinárodní projekt, na kterém se podílí 35 zemí a jehož cílem je zlepšit diagnostiku a péči o pacienty. Kromě souhrnných klinických informací o jednotlivých onemocněních lze s pomocí portálu vyhledat a oslovit konkrétní pracoviště, které se zabývá molekulární diagnostikou konkrétního syndromu či onemocnění (ať už rutinně či výzkumně). Portál také obsahuje odkazy na pacientská sdružení a informace o léčivých přípravcích, které jsou používány v léčbě vzácných onemocnění (Aymé a Schmidtke, 2007).

Simons Foundation Autism Research Initiative (SFARI, https://sfari.org/) je iniciativa zaměřená na zlepšení diagnostiky a léčby PAS. Portál obsahuje mimo jiné i databázi genů asociovaných s PAS (SFARIGene), která v současné době obsahuje přes 800 genů (Abrahams et al., 2013).

Developmental Brain Disorder Genes Database (http://www.geisingeradmi.org/care-innovation/studies/dbd-genes/) je databáze obsahující informace o mutacích v potvrzených i kandidátních genech pro neurovývojová onemocnění (MR, PAS, porucha s deficitem pozornosti a hyperaktivitou (ADHD), schizofrenie, bipolární porucha a epilepsie). U jednotlivých genů je uvedena asociace s konkrétními neurovývojovými poruchami a počet dosud popsaných ztrátových mutací. Geny jsou klasifikovány do 4 kategorií podle počtu již popsaných mutací a jejich charakteru (*de novo* či zděděné). U genů v kategoriích 1 a 2 se předpokládá existence silných důkazů pro asociaci s neurovývojovými poruchami, geny v kategorii 3 a 4 jsou považovány za kandidátní (Gonzalez-Mantilla et al., 2016).

1.3.2.4 Genomové browsery

The University of California Santa Cruz Genome Browser (UCSC, https://genome.ucsc.edu/) je komplexní databáze s mnoha různými funkcemi, jejíž využití pro bioinformatickou analýzu je značné. Lze ji využít například k získání informací o počtu a poloze již dříve popsaných CNV (ať už patogenních či benigních), k ověření genového obsahu konkrétních CNV, k identifikaci SNP či ověření funkčnosti PCR primerů *in silico*. Databáze je propojena s dalšími databázemi (včetně DGV, DECIPHER, ECARUCA, OMIM). Systém umožňuje uživatelům nahrát vlastní data získaná např. vyšetřením aCGH či SNPa a ta pak díky propojení s ostatními databázemi snadněji analyzovat (Speir et al., 2016). Obdobné spektrum funkcí a využití nabízí browser Ensembl (Yates et al., 2016).

1.3.2.5 Programy pro predikci patogenity variant

Predikční programy jsou využívány k hodnocení patogenity jednonukleotidových variant. Jedním z často využívaných programů je PolyPhen-2 (Polymorphism Phenotyping v2, http://genetics.bwh.harvard.edu/pph2/), který slouží k hodnocení patogenity missense variant. Funguje na bázi pravděpodobnostního klasifikátoru a pro každou variantu určí pravděpodobnost, s jakou je tato varianta patogenní. Pro danou variantu software určí skóre od 0 do 1, čím více se toto skóre blíží 1, tím je pravděpodobnější, že daná varianta je patogenní. Tento klasifikátor je trénován na množině známých patogenních variant a množině známých nepatogenních variant a posuzuje 11 vlastností včetně evoluční konzervovanosti aminokyseliny na dané pozici apod. Pro klinicko-diagnostické účely je doporučován klasifikátor trénovaný na množině HumVar, která obsahuje 13032 mutací způsobujících lidské nemoci v databázi Uniprot a 8946 lidských nesynonymních polymorfismů bez asociace k nějaké nemoci (Adzhubei et al., 2010).

Dalším z často využívaných softwarů je SIFT (http://sift.jcvi.org/). Program SIFT je založen pouze na odhadu evoluční konzervovanosti na dané pozici. Program nejprve vyhledá sekvence podobné sekvenci, ve které posuzujeme variantu, a odhadne evoluční konzervovanost dané pozice. Výsledkem je opět pravděpodobnost, která je měřítkem neškodnosti dané varianty. Varianty s hodnotou této pravděpodobnosti menší než 0,05 jsou považovány za patogenní (Kumar et al., 2009).

1.4 Vztah genotypu a fenotypu se zaměřením na MR a PAS

1.4.1 Důsledky genetických variant pro expresi genů

Přítomnost genetických variant může ovlivňovat expresi a funkci genů několika mechanismy. U CNV zasahujících na dávku citlivé geny vedou delece či duplikace těchto genů k nedostatečné či nadměrné produkci proteinu. V případě intragenové lokalizace chromozomálního zlomu může dojít k inaktivaci alely či produkci zkráceného nefunkčního proteinu. Pokud jsou oba zlomy lokalizovány intragenově v různých genech, může dojít ke vzniku fúzního genu. V některých případech může jinak benigní delece genu demaskovat bodovou mutaci na druhé alele, což může vést k rozvoji AR onemocnění (Lupski a Stankiewicz, 2005; Kumar, 2008).

SNV mohou funkci genů ovlivňovat také několika způsoby. Ztrátové mutace vedoucí k předčasnému zařazení terminačního kodonu mohou vést k inaktivaci alely či produkci nefunkčního proteinu. Tento typ mutací se ve fenotypu může projevit v případě na dávku citlivých genů, nebo v případě ztráty funkce genu na obou alelách (AR onemocnění). Podobným způsobem mohou fenotyp ovlivnit i mutace sestřihové, které postihují sestřihová místa či vytvářejí alternativní sestřihová místa a vedou k produkci aberantní messengerové RNA (mRNA), nebo mutace typu missense, které vedou k zařazení nesprávného aminokyselinového zbytku a v důsledku opět k produkci nefunkčního proteinu. Dalším možným důsledkem přítomnosti bodových mutací či drobných indel variant je vznik proteinu s toxickou funkcí a/nebo nadměrná produkce takového proteinu, což opět může mít negativní důsledky na fenotyp pacientů.

1.4.2 Definice MR a PAS

MR je dle mezinárodní klasifikace nemocí, 10. revize (MKN-10; http://www.uzis.cz/cz/mkn/index.html) definována jako *„stav zastaveného nebo neúplného duševního vývoje, který je charakterizován zvláště porušením dovedností, projevujícím se během vývojového období, postihujícím všechny složky inteligence, to je poznávací, řečové, motorické a sociální schopnosti. Retardace se může vyskytnout bez, nebo současně s jinými somatickými nebo duševními poruchami.“* MR je diagnostikována klinicky pomocí psychologického vyšetření. Stupeň postižení se obvykle určí pomocí standardizovaných testů inteligence. Dle zjištěného inteligenčního kvocientu (IQ) lze tíži MR klasifikovat jako lehkou (IQ 50 - 69), střední (IQ 35 - 49), těžkou (IQ 20 - 34) a hlubokou (IQ < 20). V anglosaské literatuře je někdy používána klasifikace na lehkou MR (IQ 70 - 50) a těžkou MR (IQ < 50). Toto dělení je s ohledem na genetické příčiny MR účelnější, neboť u pacientů s těžkou MR se jednoznačná

genetická příčina nalezne častěji než u pacientů s lehkou MR, kde se předpokládá spíše multifaktoriální etiologie (Leonard a Wen, 2002; Vissers et al., 2016).

PAS spadá do skupiny pervazivních vývojových poruch. Dle MKN-10 je tato skupina charakterizována *„kvalitativním porušením recipročních sociální interakce na úrovni komunikace a omezeným, stereotypním a opakujícím se souborem zájmů a činností. Tyto kvalitativní abnormality jsou pervazivním rysem chování jedince v každé situaci“*. Stejně jako MR je i PAS diagnózou klinickou, diagnostikovatelnou s použitím standardizovaných testů. Dle posledního doporučení American Psychiatric Association jsou všechny dříve rozlišované subtypy (Autistická porucha, Aspergerova porucha, Atypický autismus, Pervazivní vývojová porucha jinak nespecifikovaná a Desintegrační porucha v dětství) sloučeny do jedné jednotky – PAS (American Psychiatric Association, 2013).

Odhadovaná prevalence MR činí přibližně 2 - 3% (Ropers, 2008; de Vries et al., 2005) a prevalence PAS 1 - 1,5% (Baron-Cohen et al., 2009). Obě jednotky jsou však často komorbidní (přibližně 70% pacientů s PAS má diagnostikován i různý stupeň MR (Fombonne, 2002) a přibližně 40% pacientů s MR má zároveň diagnostikovanou i PAS (La Malfa et al., 2004)). Vhledem k výše uvedené prevalenci představují MR a PAS závažný socioekonomický problém.

MR i PAS jsou často rozdělovány na syndromické (asociované s VVV či dalšími patologiemi a faciální stigmatizací) a nesyndromické (izolovaná MR a/nebo PAS). Kritéria pro rozhodnutí zda se jedná o jednotku syndromickou či nesyndromickou však nejsou zcela striktní a v případě přítomnosti nespecifické faciální stigmatizace jako jediného dalšího příznaku do značné míry subjektivní. Obě skupiny tedy mohou vykazovat určitý překryv. Pro účely této práce je za syndromickou jednotku považována jednotka asociovaná s VVV a výraznou faciální dysmorfií.

Příčiny MR a PAS jsou heterogenní a u řady pacientů se jedná o kombinaci genetických faktorů či predispozic a faktorů prostředí. Mezi získané příčiny MR a PAS patří prenatální poškození centrálního nervového systému (CNS) např. teratogenními agens či léky, perinatální poškození např. v důsledku hypoxie v průběhu porodu či intrakraniálního krvácení a postnatální poškození CNS např. v důsledku neuroinfekce, hypoxie či traumatu. Genetické příčiny MR a PAS jsou obvykle rozdělovány na mikroskopicky viditelné chromozomální aberace, CNV a monogenní onemocnění (se zvláštní podskupinou dědičných poruch metabolismu). Toto rozdělení je ale do značné míry umělé a nedokonalé, protože stejná afekce může být způsobena různými typy variant a často můžeme pozorovat i jejich překryv. Podrobněji jsou genetické příčiny

MR a PAS diskutovány níže (viz kap. 1.4.3). U 30% - 50% pacientů zůstane příčina jejich obtíží i přes veškeré diagnostické úsilí neobjasněna. U pacientů s těžším či syndromickým postižením je jednoznačná genetická příčina identifikována častěji než u pacientů s postižením lehčím (Vissers et al., 2016).

Efektivní terapie neurovývojových poruch zatím neexistuje (s výjimkou některých vrozených poruch metabolismu, k jejichž klinickému obrazu MR a/nebo PAS mohou patřit (Das, 2016)). Identifikace příčin má tak význam zejména v genetickém poradenství a určení reprodukční prognózy resp. rizika opakování v rodině.

1.4.3 Dnes známé genetické příčiny MR a PAS

MR a PAS jsou mezi častými projevy velké řady vzácných genetických onemocnění a i jejich etiologie je do značné míry podobná. Nejčastější a nejdéle známou genetickou příčinou MR je trizomie 21 - Downův syndrom. I ten může být asociován s PAS (Rauch et al., 2006; Warner et al., 2014). MR či PAS mohou podmiňovat i SV, zejména nebalancované. Balancované chromozomální aberace jsou označeny jako příčina MR a/nebo PAS u méně než 1% pacientů (Ropers, 2010). Mikroskopicky viditelné nebalancované SV jsou identifikovány u přibližně 3% pacientů s MR a/nebo PAS (Miller et al., 2010), submikroskopické SV (resp. CNV) se na etiologii podílejí u 10 – 15% pacientů (Zahir a Friedman, 2007; Kaufman et al., 2010). Častější a dlouho známé rekurentní mikrodeleční syndromy asociované s MR a PAS jsou shrnuty v Tabulce 1 v kapitole 1.2.1.2. Díky zavedení aCGH a SNPa do rutinní praxe byly v posledních deseti letech popsány nové mikrodeleční a mikroduplikační syndromy, jejichž fenotyp však často není příliš specifický a je tedy klinicky obtížně rozpoznatelný. Některé asociované SV navíc vykazují neúplnou penetranci a velmi variabilní expresivitu. Díky vysoké genetické heterogenitě jsou jednotlivé SV jako příčiny MR a PAS vzácné (s výskytem obvykle menším než 0,1%). Mezi nejčastější CNV asociované s MR a/nebo PAS patří mikroduplikační syndrom 15q11.2-q13 s frekvencí 1% a mikrodeleční syndrom 16p11.2 s frekvencí 0,5%. Další asociované SV jsou shrnuty v této práci (Nevado et al., 2014).

V některých případech je kauzální gen lokalizován v oblasti již dříve popsaných mikrodelečních či mikroduplikačních syndromů a novější práce ukázaly, že bodové mutace v daném genu jsou postačující pro rozvoj patologického fenotypu (např. *UBE3A* (Kishino et al., 1997), *RAI1* (Slager et al., 2003), *SHANK3* (Durand et al., 2007) nebo *KANSL1* (Koolen et al., 2012)).

Z monogenních jednotek je s MR a PAS nejčastěji asociován syndrom fragilního X chromozomu (FRAX) který vysvětlí přibližně 0,5% případů (Vissers et al., 2016). MR a PAS jsou asociovány s celou řadou dalších monogenních syndromických jednotek, přehled nejčastějších je uveden v Tabulce 2. Některé z těchto jednotek mohou být způsobeny jak nebalancovanou SV (mikrodelecí či mikroduplikací zasahující kauzální gen) tak i bodovou mutací v konkrétním genu. Monogenních syndromických jednotek se známou etiologií je v databázích OMIM, LMD či Orphanet popsáno mnoho stovek.

Tabulka 2: Přehled nejčastějších monogenních syndromů a onemocnění asociovaných s MR a/nebo PAS (Moss a Howlin, 2009).

Syndrom	Lokus/lokusy	Gen/geny
syndrom fragilního X	Xq27.3, Xq28	*FMR1, FMR2*
tuberózní skleróza	9q34.13, 16p13.3	*TSC1, TSC2*
Rettův syndrom	Xq28, Xp22.13	*MECP2, CDKL5*
fenylketonurie	12q23.2	*PAH*
CHARGE* asociace	8q12.2	*CHD7*
Angelmanův syndrom	15q11.2	*UBE3A*

**CHARGE - Coloboma, Heart anomaly, choanal Atresia, Retardation, Genital and Ear anomalies*

V posledních letech díky rozvoji molekulárně genetických metod přibyly poznatky i o genech zapříčiňujících nesyndromickou MR a/nebo PAS (Vissers et al., 2016). Zdrojem informací o těchto genech jsou databáze uvedené v kapitole 1.3.2. Asociace některých genů je již dostatečně prokázána, u dalších jsou z důvodu malého počtu pozorování nutné další studie. Příklady genů, jejichž asociace s MR a PAS je již dobře podložena, jsou uvedeny v Tabulce 3.

Tabulka 3: Příklady genů asociovaných s MR/PAS (Amberger et al., 2009; Abrahams et al., 2013; Gonzalez-Mantilla et al., 2016; Vissers et al., 2016)

Gen*	Lokus	Dědičnost
CC2D1A	19p13.12	AR
CDKL5	Xp22.13	XL
DLG3	Xq13.1	XL
DYRK1A	21q22.13	AD
FOXP1	3p13	AD
CHD2	15q26.1	AD
CHD8	14q11.2	AD
IL1RAPL1	Xp21.3-21.2	XL
KPTN	19q13.32	AR
MAN1B1	9q34.3	AR
METTL23	17q25.1	AR
NDST1	5q33.1	AR
PAK3	Xq23	XL
SYNGAP1	6p21.32	AD
SYP	Xp11.23	XL
TAF2	8q24.12	AR
TRAPPC9	8q24.3	AR
TSPAN7	Xp11.4	XL

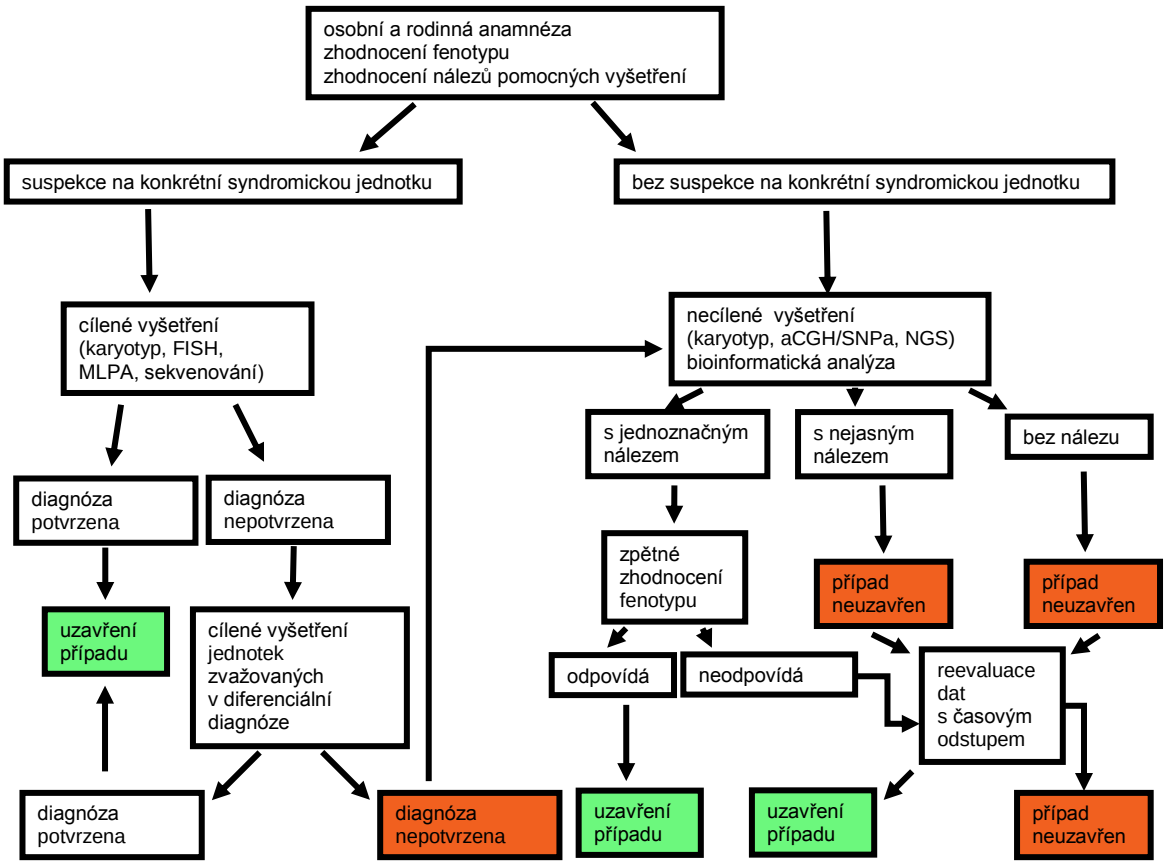
*XL- X vázaná dědičnost; * Do přehledu byly zahrnuty geny, jejichž asociace s MR/PAS je uvedena alespoň ve dvou výše uvedených zdrojích. Varianty v těchto genech jsou asociovány převážně s nesyndromickou MR.*

1.4.4 Korelace genotyp-fenotyp

Klinický genetik indikuje specifická vyšetření na základě zhodnocení fenotypu pacienta. V případě, že se jedná o jednoznačně klinicky identifikovatelnou diagnózu, je indikováno cílené vyšetření k jejímu potvrzení. Pro řadu diagnóz jsou vypracována indikační kritéria a klinické dotazníky, které velmi usnadňují rozhodnutí o indikaci k vyšetření. Vzhledem k variabilní expresivitě většiny vzácných onemocnění je však třeba indikaci k vyšetření vždy posuzovat individuálně, zejména u hraničních případů. Precizní zhodnocení fenotypu je v těchto případech velmi důležité i pro následnou interpretaci nálezu (zejména u SNV), neboť může být nalezena dosud nepopsaná varianta, jejíž vliv mohou predikční programy hodnotit diskrepantně. Zhodnocení, zda nalezená varianta je skutečně kauzální, je pak výsledkem spolupráce mezi klinickým genetikem (posouzení, zda je fenotyp skutečně v souladu s nálezem) a molekulárním genetikem, který zhodnotí možný vliv nalezené varianty na funkci proteinu (v případě SNV) či rozsah a genový obsah aberace (v případě CNV).

V řadě případů, zejména u pacientů s nesyndromickou MR/PAS, nebo u syndromických pacientů, jejichž fenotyp však neodpovídá žádné dosud popsané jednotce, nelze diagnózu klinicky jednoznačně určit. Pro tyto případy jsou vypracována

vyšetřovací schémata a indikační kritéria pro celogenomové metody (např. pro aCGH (de Vries et al., 2005)). Celogenomové metody mohou u těchto pacientů odhalit varianty v již známých kauzálních genech podmiňujících MR a PAS, nebo mohou identifikovat nové kandidáty, které je nutno potvrdit dalším studiem. Možný vyšetřovací postup je shrnut na Obrázku 1.



Obrázek 1:Možný postup při vyšetření pacientů se suspekcí na vzácné onemocnění.

Korelace genotyp-fenotyp je důležitý krok v interpretaci výsledků laboratorních vyšetření. V principu jde o zhodnocení, zda fenotyp pacienta odpovídá zjištěným nálezům. V rutinní praxi u dobře popsaných onemocnění, kdy je fenotyp jasně definován již před vyšetřením a výsledek laboratorních vyšetření klinické podezření potvrdí, je uzavření případu snadné. Pokud se najde známá patogenní mutace, lze konstatovat, že klinická diagnóza byla potvrzena na molekulární úrovni.

Složitější situace nastává v případech, kdy fenotyp není dostatečně specifický a/nebo se jedná o nález vzácné varianty, o jejíchž dopadech na fenotyp nejsou dosud k dispozici dostatečné údaje. Při hodnocení je třeba mít na paměti variabilní expresivitu zjištěných variant a také fakt, že ne všechny pacientovy obtíže musí být nutně

důsledkem přítomnosti zjištěné varianty. V těchto případech je nezbytné využití bioinformatických nástrojů a databází. Porovnávána je přítomnost klíčových či signálních příznaků, tíže postižení, ale i podobnost fenotypu publikovaných pacientů s fenotypem vyšetřovaného pacienta.

Byla vypracována řada doporučení, jakým způsobem hodnotit patogenitu nalezených variant. Nalezené varianty jsou porovnány se záznamy o benigních polymorfismech a databázemi variant nalezených ve velkých souborech. Pokud varianta není v žádné z databází uvedena, pak je třeba pátrat, zda již byl popsán pacient se stejným či podobným nálezem a pokud ano, tak porovnat, zda jeho fenotyp byl podobný vyšetřovanému pacientovi. V případě CNV se hodnotí překryv se známými mikrodelečními či mikroduplikačními syndromy, genový obsah, zejména přítomnost na dávku citlivých genů, charakter aberace (heterozygotní delece, homozygotní delece, duplikace, amplifikace) (Miller et al., 2010). V případě dosud nepopsaných SNV se hodnotí charakter varianty (missense, nonsense, sestřihová, posunová atd.) a její pravděpodobný vliv na funkci proteinu, zda varianta postihuje konzervované místo apod. Nápomocné jsou i predikční programy (viz kap. 1.3.2.5)

Jak u CNV tak u SNV se hodnotí, zda byla varianta zděděna od zdravého rodiče či zda (v případě familiárního výskytu onemocnění) nález segreguje s fenotypem v rodině. Pro potvrzení kauzální role kandidátních genů je nutné nalezení dalších pacientů s mutacemi v příslušných genech a s podobným fenotypem. Cennou podporu poskytnou také funkční studie na úrovni RNA, proteinu, buňky nebo experimentálního organismu.

I přes využití veškerých dostupných analytických nástrojů nemusí být případ uzavřen a nález může zůstat v kategorii UV. Rodiny pacientů takový závěr obvykle nevnímají příznivě, zvláště pokud nejsou dopředu náležitě poučeny o možnostech a limitech celogenomových metod. Obtížné až nemožné je rovněž zodpovědět dotazy na dlouhodobou prognózu u pacientů s unikátními či dosud nedostatečně popsanými nálezy (studie zaměřené na genetické příčiny vzácných onemocnění se obvykle nezabývají dlouhodobým sledováním pacientů a klinické údaje o dospělých pacientech s konkrétními genomovými a genovými variantami často zcela chybí, i když lze oprávněně předpokládat, že takoví pacienti existují). Dlouhodobá prognóza je však pochopitelně hlavním předmětem zájmu rodičů postiženého dítěte. I z těchto důvodů je důležité pacienty s unikátními nálezy dlouhodobě sledovat a pravidelně hodnotit jejich klinický stav.

2 Cíle práce

Práce byla zaměřena na detailní analýzu genotypu a fenotypu pacientů se vzácnými onemocněními vykazujícími MR a PAS ve spektru jejich příznaků. Cílem práce bylo:

- nalézt kauzální genomové a genové varianty u pacientů se vzácnými onemocněními asociovanými s MR a/nebo PAS s využitím různých molekulárně cytogenetických a molekulárně genetických vyšetření
- podrobně analyzovat genotyp i fenotyp těchto pacientů a přispět tak k poznatkům o zasažených genech a konkrétních nalezených variantách, stejně tak jako k poznatkům o spektrech symptomů, projevech a prognóze příslušných vzácných onemocnění
- ve vybraných případech pomocí přesného určení povahy aberací s využitím metod molekulární genetiky včetně například klonování zlomových míst zjistit pravděpodobný mechanismus vzniku aberací, jejich genový obsah a následně určit i podíl jednotlivých genů na projevech pozorovaných u pacienta
- potvrdit kauzalitu pro konkrétní onemocnění u již publikovaných, avšak zatím nedostatečným počtem pacientů podložených vzácných nálezů
- optimalizovat vyšetřovací schéma zejména u pacientů s MR a PAS, u kterých nebylo možno uzavřít diagnózu klinicky resp. s využitím cílené metody
- přispět k poznatkům o možnostech a limitech využití celogenomových metod v diagnostice pacientů se vzácnými onemocněními a k optimalizaci praktického postupu těchto vyšetření (genetické poradenství před testem, proces informovaného souhlasu, sdělování výsledků, postoj k náhodným nálezům a nejistotě spojené s některými nálezy atd.).

3 Publikace a komentáře

Publikace 1
***FMR1* gene expansion, large deletion of Xp, and skewed X-inactivation in a girl with mental retardation and autism**.
Vazna A, Musova Z, **Vlckova M**, Novotna D, Dvorakova L, Hrdlicka M, Havlovicova M, Sedlacek Z.
Am J Med Genet A. 2010 May;152A(5):1273-7.

Publikace 2
Mechanism and genotype-phenotype correlation of two proximal 6q deletions characterized using mBAND, FISH, array CGH, and DNA sequencing.
Vlckova M, Trkova M, Zemanova Z, Hancarova M, Novotna D, Raskova D, Puchmajerova A, Drabova J, Zmitkova Z, Tan Y, Sedlacek Z.
Cytogenet Genome Res. 2012;136(1):15-20.

Publikace 3
Skin lesions in a boy with X-linked lymphoproliferative disorder: comparison of 5 *SH2D1A* deletion cases.
Mejstrikova E, Janda A, Hrusak O, Buckova H, **Vlckova M**, Hancarova M, Freiberger T, Ravcukova B, Vesely K, Fajkusova L, Kopeckova L, Sumerauer D, Kabickova E, Sediva A, Stary J, Sedlacek Z.
Pediatrics. 2012 Feb;129(2):e523-8.

Publikace 4
Array comparative genome hybridization in patients with developmental delay: two example cases.
Hancarova M, Drabova J, Zmitkova Z, **Vlckova M**, Hedvicakova P, Novotna D, Vlckova Z, Vejvalkova S, Marikova T, Sedlacek Z.
N Biotechnol. 2012 Feb 15;29(3):321-4.

Publikace 5
Identification of a patient with intellectual disability and *de novo* 3.7 Mb deletion supports the existence of a novel microdeletion syndrome in 2p14-p15.
Hancarova M, Vejvalkova S, Trkova M, Drabova J, Dleskova A, **Vlckova M**, Sedlacek Z.
Gene. 2013 Mar 1;516(1):158-61.

Publikace 6
Monozygotic twins with 17q21.31 microdeletion syndrome.
Vlckova M, Hancarova M, Drabova J, Slamova Z, Koudova M, Alanova R, Mannik K, Kurg A, Sedlacek Z.
Twin Res Hum Genet. 2014 Oct;17(5):405-10.

Publikace 7
Deletions of 9q21.3 including *NTRK2* are associated with severe phenotype.
Hancarova M, Puchmajerova A, Drabova J, Karaskova E, **Vlckova M**, Sedlacek Z.
Am J Med Genet A. 2015 Jan;167A(1):264-7.

Publikace 8

***BCL11A* deletions result in fetal hemoglobin persistence and neurodevelopmental alterations.**

Basak A, Hancarova M, Ulirsch JC, Balci TB, Trkova M, Pelisek M, **Vlckova M**, Muzikova K, Cermak J, Trka J, Dyment DA, Orkin SH, Daly MJ, Sedlacek Z, Sankaran VG.

J Clin Invest. 2015 Jun;125(6):2363-8.

Publikace 9

***HCFC1* loss-of-function mutations disrupt neuronal and neural progenitor cells of the developing brain.**

Jolly LA, Nguyen LS, Domingo D, Sun Y, Barry S, Hancarova M, Plevova P, **Vlckova M**, Havlovicova M, Kalscheuer VM, Graziano C, Pippucci T, Bonora E, Sedlacek Z, Gecz J.

Hum Mol Genet. 2015 Jun 15;24(12):3335-47.

Publikace 10

A patient showing features of both SBBYSS and GPS supports the concept of a *KAT6B* related disease spectrum, with mutations in mid-exon 18 possibly leading to combined phenotypes

Vlckova M, Simandlova M, Zimmermann P, Stranecky V, Hartmannova H, Hodanova K, Havlovicova M, Hancarova M, Kmoch S, Sedlacek Z

Eur J Med Genet. 2015 Oct;58(10):550-5.

Publikace 11

Activating mutations affecting the Dbl homology domain of *SOS2* cause Noonan syndrome.

Cordeddu V, Yin JC, Gunnarsson C, Virtanen C, Drunat S, Lepri F, De Luca A, Rossi C, Ciolfi A, Pugh TJ, Bruselles A, Priest JR, Pennacchio LA, Lu Z, Danesh A, Quevedo R, Hamid A, Martinelli S, Pantaleoni F, Gnazzo M, Daniele P, Lissewski C, Bocchinfuso G, Stella L, Odent S, Philip N, Faivre L, **Vlckova M,** Seemanova E, Digilio C, Zenker M, Zampino G, Verloes A, Dallapiccola B, Roberts AE, Cavé H, Gelb BD, Neel BG, Tartaglia M.

Hum Mutat. 2015 Nov;36(11):1080-7.

Publikace 12

Molecular genetic analysis in 14 Czech Kabuki syndrome patients is confirming the utility of phenotypic scoring.

Paderova J, Holubova A, Simandlova M, Puchmajerova A, **Vlckova M**, Malikova M, Pourova R, Vejvalkova S, Havlovicova M, Senkerikova M, Ptakova N, Drabova J, Krepelova A, Macek M Jr.

Clin Genet. 2016 Feb 4. doi: 10.1111/cge.12754. [Epub ahead of print]

Publikace 1

***FMR1* gene expansion, large deletion of Xp, and skewed X-inactivation in a girl with mental retardation and autism**.

Vazna A, Musova Z, Vlckova M, Novotna D, Dvorakova L, Hrdlicka M, Havlovicova M, Sedlacek Z.

Am J Med Genet A. 2010 May;152A(5):1273-7. IF: 2,159

V práci popisujeme unikátní pacientku s plnou mutací genu *FMR1* a rozsáhlou delecí v oblasti Xp22.1-p22.3. Práce je zaměřena na detailní molekulárně genetickou analýzu delece a studium mechanismu jejího vzniku. Dále je diskutován také podíl obou genetických defektů na fenotypu pacientky a jejich vliv na inaktivaci chromozomu X.

Fenotyp pacientky v době vyšetření zahrnoval mírnou nespecifickou faciální stigmatizaci, lehkou MR, atypický autismus a poruchy chování s dominujícími záchvaty vzteku a agresivity a s dysforií. Výskyt tremoru a předčasného ovariálního selhání v maternální linii ukazoval na možnost syndromu fragilního X. Následným vyšetřením byla u pacientky skutečně prokázána plná mutace genu *FMR1*. Překvapivá však byla metylace nejen expandované, ale i normální alely *FMR1*. Současně byla cytogenetickým vyšetřením u pacientky zjištěna rozsáhlá delece Xp22. Vyšetřením rodičů bylo zjištěno, že delece vznikla *de novo* na paternálním chromozomu X, zatímco matka předala pacientce plnou mutaci genu *FMR1*. Analýza delece s následným klonováním zlomů ukázala, že rozsah delece je přibližně 17,4 Mb a zahrnuje více než 90 protein kódujících genů. Přítomnost tak velké delece může vést k nerovnoměrné inaktivaci chromozomů X u pacientky, konkrétně k převažující inaktivaci deletovaného chromozomu X, což bylo prokázáno analýzou metylace na lokusu *AR*. Tento stav má za následek pozorovanou metylaci obou alel genu *FMR1* (jedna alela je metylována kvůli plné mutaci a druhá, normální alela, kvůli deleci na chromozomu, na kterém se nachází).

Tato neobvyklá kombinace dvou genetických defektů vede k relativně mírnému fenotypu u pacientky způsobenému spíše „demaskováním“ plné mutace v genu *FMR1* než přítomností rozsáhlé delece, přestože ta zahrnuje velké množství genů. Zatímco delece se fenotypově díky nerovnoměrné X-inaktivaci neprojevuje, efekt plné mutace *FMR1* může být díky témuž fenoménu výraznější, než by se dalo u dívky - heterozygotky pro plnou mutaci očekávat. Určitý vliv genů zasažených delecí a unikajících X-inaktivaci nelze zcela vyloučit.

CLINICAL REPORT

AMERICAN JOURNAL OF medical genetics PART A

FMR1 Gene Expansion, Large Deletion of Xp, and Skewed X-Inactivation in a Girl With Mental Retardation and Autism

Alzbeta Vazna,[1] Zuzana Musova,[1] Marketa Vlckova,[1] Drahuse Novotna,[1] Lenka Dvorakova,[2] Michal Hrdlicka,[3] Marketa Havlovicova,[1] and Zdenek Sedlacek[1]*

[1]Department of Biology and Medical Genetics, Charles University, 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic

[2]Institute of Inherited Metabolic Disorders, Charles University, 1st Faculty of Medicine and General University Hospital, Prague, Czech Republic

[3]Department of Child Psychiatry, Charles University, 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic

Received 28 April 2009; Accepted 21 December 2009

We describe a girl with mild facial anomalies, mild mental retardation, and atypical autism with a remarkable behavioral phenotype of persistent anger, aggression, and dysphoria. The occurrence of late-onset tremor and premature ovarian failure in the maternal branch of the family pointed to a possible defect in the *FMR1* gene. Indeed, the patient carried a full *FMR1* mutation. Unexpectedly, both alleles of the gene were almost completely methylated. Cytogenetic examination of the patient revealed in addition a large de novo deletion in band Xp22 on one of her X chromosomes. The deletion was fine mapped using oligonucleotide array CGH, and its breakpoints were localized using sequencing. The size of the deletion was about 17.4 Mb, and it contained more than 90 protein-coding genes. Microsatellite analysis indicated paternal origin of the aberrant chromosome. The large rearrangement was the most probable cause of the X-inactivation skewing, thus explaining the methylation of not only the expanded (maternal) but also the normal (paternal) *FMR1* alleles. This pattern of skewed X-inactivation was confirmed using the analysis of methylation at the *AR* locus. The relatively mild phenotype of the patient resulted most likely from unmasking of the *FMR1* defect. Although the deleted region contained many important genes, the phenotypic contribution of the rearranged X chromosome was probably limited by its almost complete inactivation. However, reduced dose of several genes escaping X-inactivation might also play a role in the phenotype of the patient. © 2010 Wiley-Liss, Inc.

Key words: fragile X syndrome; Xp deletion; array CGH; skewed X-inactivation; mental retardation; autism

How to Cite this Article:
Vazna A, Musova Z, Vlckova M, Novotna D, Dvorakova L, Hrdlicka M, Havlovicova M, Sedlacek Z. 2010. *FMR1* gene expansion, large deletion of Xp, and skewed X-inactivation in a girl with mental retardation and autism.
Am J Med Genet Part A 152A:1273–1277.

INTRODUCTION

The fragile X syndrome (FXS, OMIM 300624) is the most common single-gene cause of mental retardation. It is associated with the expansion of a polymorphic CGG repeat in the 5′ untranslated region of the *FMR1* gene in Xq27.3. Premutations (59–200 CGG) can cause fragile X-associated tremor-ataxia syndrome (FXTAS) in older men or premature ovarian failure (POF) in women, while full mutations (more than 200 CGG) lead to methylation silencing of the *FMR1* gene and FXS. Affected males have mental impairment, developmental, cognitive, language, and motor delay, a unique behavioral profile often with autistic traits, mild dysmorphic features (elongated face, prominent jaw, large protruding ears), hyperextensibility of joints, and large testes after puberty. Females with full *FMR1* mutation have variable milder phenotype due to the presence of a normal allele. About 50% of them suffer from some degree of mental retardation and behavioral problems [reviewed in Jacquemont et al., 2007].

In this report we describe a girl with mental retardation and atypical autism. She was found to carry a full mutation in the *FMR1* gene on one of her X chromosomes and a large deletion on the other

Grant sponsor: Ministries of Education and Health of the Czech Republic; Grant numbers: NR/9457-3, MZO00064203, MSM0021620806; Grant sponsor: European Commission; Grant numbers: 043318 (INCORE), 223692 (CHERISH).
*Correspondence to:
Zdenek Sedlacek, Department of Biology and Medical Genetics, Charles University, 2nd Faculty of Medicine and University Hospital Motol, V Uvalu 84, 15006 Prague 5, Czech Republic.
E-mail: zdenek.sedlacek@lfmotol.cuni.cz
Published online 13 April 2010 in Wiley InterScience (www.interscience.wiley.com)
DOI 10.1002/ajmg.a.33352

© 2010 Wiley-Liss, Inc.

X chromosome, leading to an almost complete X-inactivation skewing and silencing of the normal *FMR1* allele.

MATERIALS AND METHODS

Clinical Report

The patient first presented to our institution with developmental delay, mild mental retardation, and suspected pervasive developmental disorder at the age of 6.5 years. The girl came from the first uneventful pregnancy of healthy unrelated parents (the mother aged 26 years, the father 31 years). The labor was induced at the 42nd week of gestation. The birth weight was 3,750 g (75th centile), the length was 51.5 cm (75th centile). The Apgar scores were 9, 10, and 10 at 1, 5, and 10 min, respectively. She exhibited poor suck, weight loss, and neonatal icterus. Her milestones were significantly delayed (crawling at the age of 10 months, standing at the age of 15 months, walking at the age of 20 months). Speech development was not delayed, but the speech was odd and dyslalic; problems with grammatical functioning were also detected.

At the age of 6.5 years the height of the patient was 120 cm (50th centile), her weight was 28 kg (90–97th centile), and her head circumference was 52 cm (75th centile). She had mild dysmorphic features with frontal bossing, strabismus, a narrow nasal bridge, prominent dysplastic ears, and a thin upper lip. She also had thick hands, short thick fingers, and broad thumbs. Radiography showed signs of Thiemann disease.

The patient was diagnosed with atypical autism at the age of 7.5 years. Using Childhood Autism Rating Scale (CARS) [Schopler et al., 1980] she had a score of 32 points (cut-off score is 30 points). In Autism Diagnostic Interview—Revised [ADI-R, Lord et al., 1994], Proposed Algorithm for ICD-10, she scored positively in the "Repetitive Behaviors and Stereotyped Patterns" domain (5 points, cut-off: 3). In the "Qualitative Impairments in Reciprocal Social Interaction" domain she gained 9 points (cut-off: 10), and in the "Verbal Communication" domain she had 7 points (cut-off: 8). Her behavior was characterized by hyperkinetic syndrome, whimsicality, mulishness, anger attacks, aggression, self-injury, and persistent dysphoria. Inappropriate communication and vulgarisms in contact with both familiar and strange people were repeatedly observed.

The genealogy of the patient showed no history of mental retardation or autism, but tremor was apparent in her maternal grandfather since the age of 60 years. A maternal aunt and the mother of the maternal grandfather suffered from POF. Depression in a paternal granduncle led to his suicide at the age of 49 years.

Laboratory Investigations

Genomic DNA was isolated from blood lymphocytes of the patient, her parents, and maternal grandfather using the Gentra Puregene Blood Kit (Qiagen, Hilden, Germany). The analysis of the *FMR1* gene was performed using the Fragile X PCR Kit (Abbott, Abbott Park, IL) and Southern blotting with *Hin*dIII and *Sac*II double digest and the StB12.3 probe [Rousseau et al., 1992].

Cytogenetic examination of blood lymphocytes was performed using standard protocols. Custom array CGH analysis of the patient's DNA was performed by Nimblegen on the array HG18_CHRX_FT, and the results were analyzed using SignalMap (Nimblegen, Madison, WI). Primers X6653 (CTGTAAACGCTT CCTCTTGGGT), X24069 (GACCGACTTTCTGTTTTTGTCCT), and X24069a (GACCCAGACATCCAATGAAGCA) were used to clone and sequence the deletion breakpoints. The UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) was used for the identification of deleted genes. Hereditary diseases mapping to the deleted interval were assessed using OMIM (http://www.ncbi.nlm.nih.gov/omim).

To determine the parental origin of the deletion, three polymorphic markers, DXS9902, DXS9895, and a nameless $(TAGA)_n$ repeat, located in the deleted region of Xp, were amplified in the family using primers DXS9902R (GTCAAAAAACATAAAATTGAT GATGTC), DXS9902F (GGGTGAAGAGAAGCAGGAATTT), DXS9895F (GTCTGACAAATATTGAATGGCTCT), DXS9895R (ACCAGGTCCCTCTCACAACAC), VNTRX14063F (GTCCGG AGAGTCAAATGTTTTC), and VNTRX14063R (TGAAATGACT GTTATAGGCCAAA) and analyzed by sequencing.

The X-inactivation ratio was also determined using the analysis of the methylation status of the *AR* locus located in Xq12 as described [Racchi et al., 1998]. The patient's father and maternal grandfather were used as male controls.

RESULTS

CGG Expansion and Methylation in the *FMR1* Gene

The occurrence of mental retardation, late-onset tremor, and POF in the family of the patient prompted us to analyze her *FMR1* gene. The identification of a full mutation confirmed the diagnosis of FXS in the patient. The maternal grandfather and the asymptomatic mother of the patient carried premutation alleles (90 and 96 CGG repeats, respectively; data not shown). DNA samples of the maternal aunt and great grandmother were not available. Southern blot confirmed a full expansion with about 460 CGG repeats in the patient. Unexpectedly, this analysis also revealed almost complete methylation of the normal *FMR1* allele contrasting to the expected 50% ratio (Fig. 1). The saturation of signals from the methylated fragments did not allow exact quantification of the ratio, but it exceeded 90% in favor of methylation.

Karyotype of the Patient and Detailed Analysis of the Xp Deletion

Karyotyping revealed a large deletion of Xp22.1-p22.3 in the patient (data not shown). Karyotypes of both parents were normal. Array CGH showed that the deletion breakpoints were located about 6.6 and 24.0 Mb from the Xp telomere, and that the total size of the deletion was about 17.4 Mb (Fig. 2). No other larger DNA gains or losses were detected in the vicinity of the deletion. *ZFX* and *VCX3A* genes flanked the deletion (Fig. 2), but neither of the breakpoints interrupted any gene. The deletion contained more than 90 protein-coding genes, including several known disease genes (Fig. 2). The deletion was spanned by a 1.6-kb long PCR product. Its sequencing showed that the breakpoints were located at bases 6654160 and 24069365. Three nucleotides (TTC) added at the junction showed homology neither with the proximal nor with the distal deletion

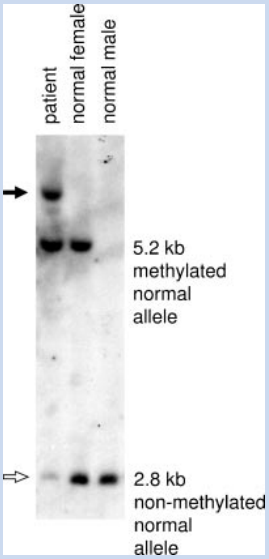


FIG. 1. Southern blot analysis of *FMR1* gene expansion and methylation. In normal individuals the probe detects a 2.8-kb *Hin*dIII/*Sac*II fragment from the active *FMR1* allele (in both sexes) and a 5.2-kb *Hin*dIII fragment from the inactive (methylated) allele (in females). The patient shows a 6.6-kb fragment originating from the methylated expanded allele (black arrow). The signal from the non-methylated fraction of the normal allele is very weak (open arrow).

breakpoint (Fig. 2). Both breakpoints were located in different dispersed repeats (Fig. 2). There was no remarkable sequence homology between the breakpoints.

Parental Origin of the Deletion and X-Inactivation Pattern at the *AR* Locus

The patient carried only the maternal alleles of the informative markers tested from the deleted region (data not shown). The defect thus affected the paternal chromosome, and the patient carried only the maternal copy of Xp22.11-p22.31. As the *FMR1* expansion was located on the maternal chromosome, both defects were in trans on the X chromosomes of the patient. The analysis of X-inactivation at the *AR* locus indicated approximately 94% skewing of her X-inactivation toward the paternal X chromosome in blood lymphocytes (data not shown). This was in accord with the methylation pattern detected at the *FMR1* locus (see above).

DISCUSSION

The diagnosis of FXS was confirmed in the patient by the identification of the full *FMR1* mutation. The involvement of a *FMR1* defect in the family was in accord with the occurrence of tremor or POF in several maternal relatives. The unexpected finding of methylation of both *FMR1* alleles was explained by the identification of a large deletion and preferential inactivation of the paternal X chromosome. Large rearrangements are causal for X-inactivation skewing [Wolff et al., 2000], probably due to selection against cells not expressing the deleted genes. In contrast to recurrent deletions of Xp22.31 caused by non-allelic homologous recombination and leading to ichthyosis and mental retardation [Van Esch et al., 2005], this deletion was unique and its features pointed to non-homologous end joining [Lupski and Stankiewicz, 2005] as the likely mechanism.

The *FMR1* gene status of our female patient resembled that of males with full mutation. One *FMR1* allele was silenced due to the repeat expansion, and the normal allele was silenced by non-random X-inactivation. The manifestation of FXS in females is less severe than in males, but 71% have IQ below 85, and 50% below 70, depending on the X-inactivation ratio [de Vries et al., 1996]. A female with full *FMR1* mutation and inactivation of the normal allele due to X-inactivation skewing of unknown cause showed at the age of 16 years symptoms of FXS, such as long narrow face, prominent ears and jaw, hyperextensibility of joints, shyness, and severe mental retardation [Heine-Suner et al., 2003]. X-inactivation and mental retardation (but not other features of FXS) correlated also in female monozygotic twins with full mutations [Willemsen et al., 2000]. Our patient had mild mental retardation and no physical features characteristic of boys with FXS; however, the typical FXS facies is known to develop at the onset of puberty.

Several other genetic mechanisms could also influence the phenotype of the patient. First, the skewed X-inactivation could unmask possible additional recessive mutations. However, the patient did not show any symptoms of other known X-linked disorders. Second, the large deletion of many genes could itself contribute to the phenotype, although the inactivation of this chromosome is likely to limit the effect.

Many patients with various Xp deletions have been described, but the great majority carried terminal deletions (including the *SHOX* gene, and resulting in small stature), and only a minority had interstitial deletions [James et al., 1998; Ogata et al., 2001; Lachlan et al., 2006]. Four related females with a deletion partly overlapping with the deletion in our patient (but extending into the pseudoautosomal region) exhibited phenotypic variability not explainable by varying X-inactivation [Chocholska et al., 2006]. Some features like strabismus and broad hands could also be seen in our patient. Another female suffering from microphthalmia with linear skin defects syndrome (MLS) carried a 3-Mb deletion fully contained within the deletion of our patient [Morleo et al., 2005].

The deletion in our patient contained several genes undergoing X-inactivation and causing X-linked dominant conditions (MLS, Nance–Horan syndrome, early infantile epileptic encephalopathy 2, pyruvate decarboxylase deficiency, Coffin–Lowry syndrome, hypophosphatemic rickets). Interestingly, it also contained about 20 genes confirmed or suspected to escape X-inactivation [Carrel and Willard, 2005; Carrel et al., 2006; Wang et al., 2006], and the patient might suffer from their decreased dosage. These genes included *OFD1*, responsible for X-linked dominant oral–facial–digital syndrome type 1. Our patient did not show symptoms of any of these disorders, similar to other females with large Xp deletions [James et al., 1998;

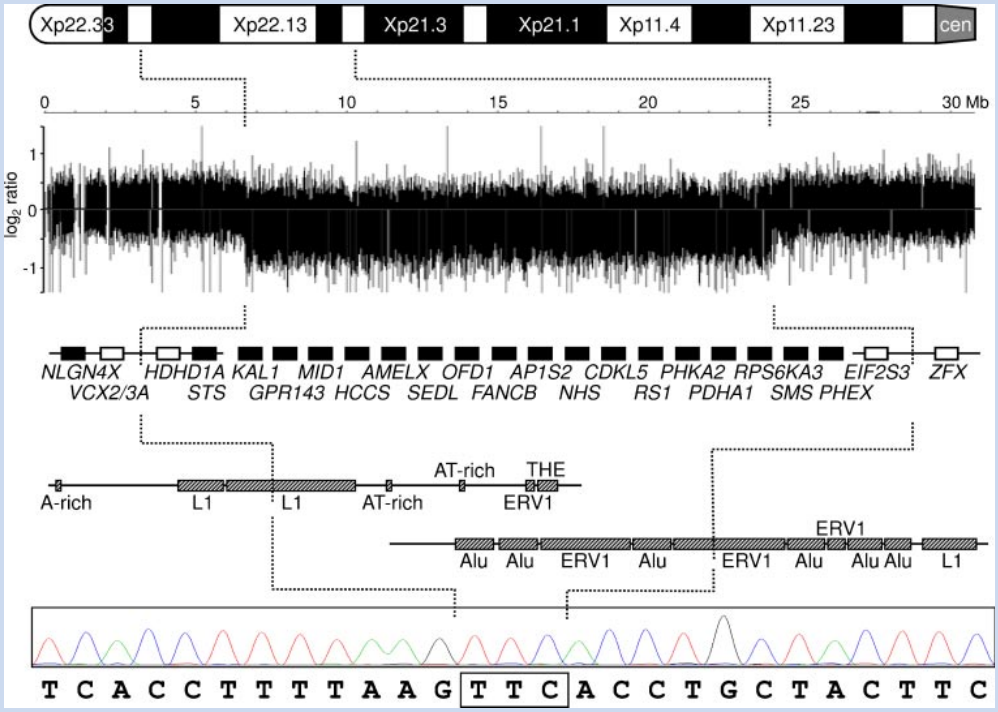


FIG. 2. Schematic overview of the Xp deletion, its gene content, and molecular features of its breakpoints. The array CGH analysis clearly indicated a deletion of Xp22.11-p22.31 between 6.6 and 24.0 Mb of one of the X chromosomes (top). Of more than 90 genes deleted, only genes involved in hereditary disorders (black rectangles) and genes flanking the deletion (open rectangles) are shown (middle). Both breakpoints were located within dispersed repeats but these belonged to different classes (L1MA5 and LOR1b-int to the L1 (LINE) and ERV1 (LTR) families, respectively). Sequencing of the deletion breakpoints revealed three added nucleotides (TTC, boxed) at the junction (bottom).

Ogata et al., 2001; Lachlan et al., 2006]. Although at least some these disorders are caused by loss-of-function mutations, their symptoms in females can be influenced by an interplay between residual expression of truncated protein products, different X-inactivation patterns induced by different genetic defects, variable X-inactivation during development and in different cell types, and the presence of a cell line where the mutated allele is active in tissues critical for the condition [James et al., 1998; Ferrante et al., 2001; Wimplinger et al., 2006].

Chromosome rearrangements can also lead to positional effects. One breakpoint in our patient was located in the proximity of the *NLGN4* and *VCX3A* genes implicated in autism and mental retardation, but this might be of no relevance due to the inactivation of the deleted chromosome. Finally, genomic imprinting could also play a role in the phenotype of the patient as her paternal X chromosome was completely inactivated.

Although a complex composite phenotype due to the co-occurrence of two defects on her X chromosomes could be expected in the patient, skewed X-inactivation likely led to her rather mild clinical picture, mainly limited to the behavioral phenotype. Autism criteria are met in 26% of boys and 6% of girls with FXS [Hatton et al., 2006], or in 18% of males and 10% of females with FXS [Clifford et al., 2007]. Most of the symptoms in our patient are thus likely to result from her *FMR1* defect.

ACKNOWLEDGMENTS

This work was supported by grants NR/9457-3, MZO00064203, and MSM0021620806 from the Ministries of Education and Health of the Czech Republic, and INCORE (043318) and CHERISH (223692) from the European Commission.

REFERENCES

Carrel L, Willard HF. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434:400–404.

Carrel L, Park C, Tyekucheva S, Dunn J, Chiaromonte F, Makova KD. 2006. Genomic environment predicts expression patterns on the human inactive X chromosome. PLoS Genet 2:e151.

Chocholska S, Rossier E, Barbi G, Kehrer-Sawatzki H. 2006. Molecular cytogenetic analysis of a familial interstitial deletion Xp22.2-22.3 with a highly variable phenotype in female carriers. Am J Med Genet Part A 140A:604–610.

Clifford S, Dissanayake C, Bui QM, Huggins R, Taylor AK, Loesch DZ. 2007. Autism spectrum phenotype in males and females with fragile X full mutation and premutation. J Autism Dev Disord 37:738–747.

de Vries BB, Wiegers AM, Smits AP, Mohkamsing S, Duivenvoorden HJ, Fryns JP, Curfs LM, Halley DJ, Oostra BA, van den Ouweland AM, Niermeijer MF. 1996. Mental status of females with an FMR1 gene full mutation. Am J Hum Genet 58:1025–1032.

Ferrante MI, Giorgio G, Feather SA, Bulfone A, Wright V, Ghiani M, Selicorni A, Gammaro L, Scolari F, Woolf AS, Sylvie O, Bernard L, Malcolm S, Winter R, Ballabio A, Franco B. 2001. Identification of the gene for oral–facial–digital type I syndrome. Am J Hum Genet 68:569–576.

Hatton DD, Sideris J, Skinner M, Mankowski J, Bailey DB Jr, Roberts J, Mirrett P. 2006. Autistic behavior in children with fragile X syndrome: Prevalence, stability, and the impact of FMRP. Am J Med Genet Part A 140A:1804–1813.

Heine-Suner D, Torres-Juan L, Morla M, Busquets X, Barcelo F, Pico G, Bonilla L, Govea N, Bernues M, Rosell J. 2003. Fragile-X syndrome and skewed X-chromosome inactivation within a family: A female member with complete inactivation of the functional X chromosome. Am J Med Genet Part A 122A:108–114.

Jacquemont S, Hagerman RJ, Hagerman PJ, Leehey MA. 2007. Fragile-X syndrome and fragile X-associated tremor/ataxia syndrome: Two faces of FMR1. Lancet Neurol 6:45–55.

James RS, Coppin B, Dalton P, Dennis NR, Mitchell C, Sharp AJ, Skuse DH, Thomas NS, Jacobs PA. 1998. A study of females with deletions of the short arm of the X chromosome. Hum Genet 102:507–516.

Lachlan KL, Youings S, Costa T, Jacobs PA, Thomas NS. 2006. A clinical and molecular study of 26 females with Xp deletions with special emphasis on inherited deletions. Hum Genet 118:640–651.

Lord C, Rutter M, Le Couteur A. 1994. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 24:659–685.

Lupski JR, Stankiewicz P. 2005. Genomic disorders: Molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet 1:e49.

Morleo M, Pramparo T, Perone L, Gregato G, Le Caignec C, Mueller RF, Ogata T, Raas-Rothschild A, de Blois MC, Wilson LC, Zaidman G, Zuffardi O, Ballabio A, Franco B. 2005. Microphthalmia with linear skin defects (MLS) syndrome: Clinical, cytogenetic, and molecular characterization of 11 cases. Am J Med Genet Part A 137A:190–198.

Ogata T, Muroya K, Matsuo N, Shinohara O, Yorifuji T, Nishi Y, Hasegawa Y, Horikawa R, Tachibana K. 2001. Turner syndrome and Xp deletions: Clinical and molecular studies in 47 patients. J Clin Endocrinol Metab 86:5498–5508.

Racchi O, Mangerini R, Rapezzi D, Rolfo M, Gaetani GF, Ferraris AM. 1998. X chromosome inactivation patterns in normal females. Blood Cells Mol Dis 24:439–447.

Rousseau F, Heitz D, Biancalana V, Oberle I, Mandel JL. 1992. On some technical aspects of direct DNA diagnosis of the fragile X syndrome. Am J Med Genet 43:197–207.

Schopler E, Reichler RJ, DeVellis RF, Daly K. 1980. Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). J Autism Dev Disord 10:91–103.

Van Esch H, Hollanders K, Badisco L, Melotte C, Van Hummelen P, Vermeesch JR, Devriendt K, Fryns JP, Marynen P, Froyen G. 2005. Deletion of VCX-A due to NAHR plays a major role in the occurrence of mental retardation in patients with X-linked ichthyosis. Hum Mol Genet 14:1795–1803.

Wang Z, Willard HF, Mukherjee S, Furey TS. 2006. Evidence of influence of genomic DNA sequence on human X chromosome inactivation. PLoS Comput Biol 2:e113.

Willemsen R, Olmer R, De Diego Otero Y, Oostra BA. 2000. Twin sisters, monozygotic with the fragile X mutation, but with a different phenotype. J Med Genet 37:603–604.

Wimplinger I, Morleo M, Rosenberger G, Iaconis D, Orth U, Meinecke P, Lerer I, Ballabio A, Gal A, Franco B, Kutsche K. 2006. Mutations of the mitochondrial holocytochrome c-type synthase in X-linked dominant microphthalmia with linear skin defects syndrome. Am J Hum Genet 79:878–889.

Wolff DJ, Schwartz S, Carrel L. 2000. Molecular determination of X inactivation pattern correlates with phenotype in women with a structurally abnormal X chromosome. Genet Med 2:136–141.

Publikace 2

Mechanism and genotype-phenotype correlation of two proximal 6q deletions characterized using mBAND, FISH, array CGH, and DNA sequencing.

Vlckova M, Trkova M, Zemanova Z, Hancarova M, Novotna D, Raskova D, Puchmajerova A, Drabova J, Zmitkova Z, Tan Y, Sedlacek Z.

Cytogenet Genome Res. 2012;136(1):15-20. IF: 1,839

V práci se zabýváme detailní analýzou dvou unikátních pacientů s významně odlišným fenotypem a s nepřekrývajícími se delecemi v proximální oblasti 6q. Určením přesného rozsahu aberací se podařilo identifikovat všechny deletované geny a tím umožnit co nejpřesnější korelaci genotypu s fenotypem. Následné klonování zlomů pak umožnilo analýzu mechanismu vzniku aberací.

Intersticiální delece 6q jsou vzácné. V literatuře se obvykle rozdělují do 3 skupin podle lokalizace a asociovaného fenotypu. Typickými znaky proximálních delecí (6q11-q16) jsou antimongoloidní sklon očních štěrbin, úzké rty a umbilikální či tříselné kýly. Střední delece (6q15-q25) jsou charakterizované mikrocefalií, hypertelorismem, intrauterinní růstovou retardací, respiračními problémy a malformacemi končetin. Distální delece (6q25-qter) způsobují rozštěpy patra, anomálie sítnice, hypoplázii genitálu a epilepsii. Hypotonie, anomálie ušních boltců, faciální dysmorfie a MR jsou společné pro všechny tři skupiny.

Prvním naším pacientem byla sedmiletá dívka s MR, hypotonií, faciální dysmorfií, pectus excavatum, opožděnou myelinizací a bilaterální atrofií čelních laloků. Její karyotyp 46,XX,del(6)(q11q14.1) byl následně potvrzen technikou mBAND. Analýza aCGH ukázala, že delece zaujímá asi 15 Mb a obsahuje 34 protein kódujících genů. Distální zlom leží v intronu 1 genu *MYO6*. Proximální zlom se nachází v centromerické oblasti, která není pokryta referenční genomovou sekvencí. Matka dívky nese tutéž deleci, ale v 77% lymfocytů má marker chromozom odpovídající deletovanému segmentu. Signál centromerické sody na jejím deletovaném chromozomu byl slabší ve srovnání s normálním homologem a další slabší signál byl přítomen na markeru. To svědčí pro vznik delece mechanismem rozštěpení centromery. U dcery již měly signály na obou chromozomech stejnou intensitu, což by mohlo svědčit pro reparaci alfa satelitních sekvencí.

Druhým pacientem byl pětiměsíční chlapec s faciální dysmorfií, rozštěpem patra, srdeční vadou (VCC), bráničí hernií, atypickými dermatoglyfy a mikropenisem. Jeho karyotyp, opět potvrzený technikou mBAND, byl 46,XX,del(6)(q14.1q16.1).

Následným vyšetřením metodou aCGH bylo zjištěno, že delece je přibližně 19,5 Mb dlouhá a obsahuje 58 protein kódujících genů. Proximální zlom leží v repetitivní sekvenci L1MA6, distální zlom leží v intronu 1 genu *FUT9*. Absence homologních sekvencí v okolí zlomů a přítomnost 7 přidaných bází v sekvenci zlomu nás vedlo k závěru, že aberace vznikala mechanismem NHEJ.

V době uveřejnění práce bylo v literatuře popsáno celkem 20 pacientů s proximální delecí 6q. Žádný z nich nebyl analyzován pomocí aCGH s vysokým rozlišením. Delece našich dvou pacientů se nepřekrývaly, ale rozdělovaly publikovaný interval proximálních delecí na dvě téměř stejné části. Také fenotyp našich pacientů byl významně odlišný – dívka s proximální delecí měla pouze mírné příznaky i v porovnání s doposud publikovanými případy. Chlapec byl naopak velmi těžce postižen a měl i četné VVV, které nejsou typickým příznakem u pacientů s proximální delecí 6q. Ani jeden z pacientů nevykazoval symptomy známých AR či autozomálně dominantních (AD) onemocnění, které jsou způsobeny geny lokalizovanými v inkriminované oblasti. Absenci příznaků AR onemocnění lze vysvětlit nepřítomností patogenní varianty na druhé alele, naproti tomu absenci příznaků AD onemocnění můžeme vysvětlit částečně jejich pozdnějším nástupem, částečně tím, že mohou být způsobeny spíše mutacemi přinášejícími novou škodlivou funkci (gain of function).

Naše práce představuje důležitý příspěvek k podrobnější subklasifikaci intersticiálních delecí 6q, k přesnějšímu stanovení prognózy jeho nosičů a ke zpřesnění role jednotlivých deletovaných genů ve spektru příznaků postižení.

Case Report

Cytogenetic and Genome Research

Cytogenet Genome Res 2012;136:15–20
DOI: 10.1159/000334709

Accepted: August 29, 2011
by M. Schmid
Published online: December 8, 2011

Mechanism and Genotype-Phenotype Correlation of Two Proximal 6q Deletions Characterized Using mBAND, FISH, Array CGH, and DNA Sequencing

M. Vlckova[a] M. Trkova[b] Z. Zemanova[c] M. Hancarova[a] D. Novotna[a] D. Raskova[b] A. Puchmajerova[a] J. Drabova[a] Z. Zmitkova[a] Y. Tan[a] Z. Sedlacek[a]

[a]Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, [b]Gennet, and [c]Center of Oncocytogenetics, Institute of Clinical Biochemistry and Laboratory Diagnostics, Charles University 1st Faculty of Medicine and General University Hospital, Prague, Czech Republic

Key Words
Array CGH · Centromere fission · Deletion junction · Marker chromosome · Proximal 6q deletion

Abstract
Proximal 6q deletions have a milder phenotype than middle and distal 6q deletions. We describe 2 patients with non-overlapping deletions of about 15 and 19 Mb, respectively, which subdivide the proximal 6q region into 2 parts. The aberrations were identified using karyotyping and analysed using mBAND and array CGH. The unaffected mother of the first patient carried a mosaic karyotype with the deletion in all metaphases analysed and a small supernumerary marker formed by the deleted material in about 77% of cells. Her chromosome 6 centromeric signal was split between the deleted chromosome and the marker, suggesting that this deletion arose through the centromere fission mechanism. In this family the location of the proximal breakpoint in the centromere prevented cloning of the deletion junction, but the junction of the more distal deletion in the second patient was cloned and sequenced. This analysis showed that the latter aberration was most likely caused by non-homologous end joining. The second patient also had a remarkably more severe phenotype which could indicate a partial overlap of his deletion with the middle 6q interval. The phenotypes of both patients could be partly correlated with the gene content of their deletions and with phenotypes of other published patients.
Copyright © 2011 S. Karger AG, Basel

Three different phenotypic groups have been suggested in carriers of interstitial 6q deletions according to the location of the defect: proximal deletions (6q11q16) with upslanted fissures, thin lips, and hernias; middle deletions (6q15q25) with microcephaly, hypertelorism, intrauterine growth retardation, respiratory problems, and limb malformations; and distal deletions (6q25qter) with cleft palate, retinal abnormalities, genital hypoplasia, and seizures [Hopkin et al., 1997]. Hypotonia, ear and facial dysmorphism, and mental retardation are common to all 3 groups [Hopkin et al., 1997]. However, most of the older studies were based solely on karyotyping, and the size and location of the deletions were determined only at low resolution. The recent boom of microarray methods allows fine mapping of the deletion breakpoints and a much more precise delineation of the extent and gene content

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2011 S. Karger AG, Basel
1424–8581/12/1361–0015$38.00/0

Accessible online at:
www.karger.com/cgr

Zdenek Sedlacek
Department of Biology and Medical Genetics
Charles University 2nd Faculty of Medicine and University Hospital Motol
Plzenska 130/221, CZ–15000 Prague 5 (Czech Republic)
Tel. +420 257 296 153, E-Mail zdenek.sedlacek@lfmotol.cuni.cz

Downloaded by:
Univerzita Karlova v Praze
195.113.40.7 - 7/9/2014 11:23:27 AM

of the aberrations which in turn allows better genotype-phenotype correlations. In addition, molecular analysis of the deletion breakpoints can shed light on the aberration mechanisms.

We present 2 patients with proximal 6q deletions identified using karyotyping. Detailed analysis of the deletions showed that they subdivided the proximal 6q region into 2 parts of similar size. One of the deletions involved the centromere and arose most likely through the centromere fission mechanism, while the second deletion was probably caused by non-homologous end joining. The patients had remarkably different phenotypes which could be partly correlated with the gene content of their deletions and with phenotypes of other patients with proximal 6q deletions.

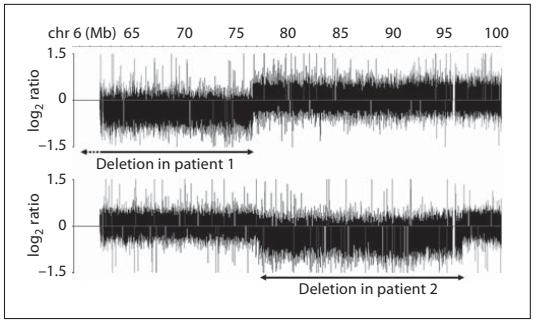


Fig. 1. Fine mapping of deletions in patients 1 and 2 using array CGH. The deletions are marked by double arrows. In patient 1 the proximal breakpoint mapped to the centromere gap and the exact size of the deletion was not known (dashed arrow). The deletions did not overlap and were separated by about 0.7 Mb of DNA.

Materials and Methods

Case Report

Patient 1 was the second child of healthy unrelated parents. The age of the mother and father was 41 and 45 years, respectively. The pregnancy was uneventful. Cytogenetic analysis of amniotic fluid cells performed due to advanced maternal age showed an apparently normal female karyotype. The delivery was at the 31st week of gestation by Caesarean section due to breech and fetal distress. The birth weight of the girl was 1,740 g (>75th centile) and length was 44 cm (>75th centile). The neonatal period was unremarkable. At the age of 2 months she developed hypotonia and affective paroxysms. EEG showed an abnormal pattern, and subsequent MRI proved delayed myelinization and bilateral frontal lobe atrophy, but no other major anomalies. The girl was referred to a geneticist at the age of 3 years because of developmental delay and absent speech. She had mild facial dysmorphism (high forehead, hypertelorism, epicanthal folds, dysplastic ears), single palmar crease on the right hand, pectus excavatum, hypotonia, and mild mental retardation. The audiologic exam was normal.

Patient 2 was the first child of healthy unrelated 30-year-old parents. Fetal ultrasound showed a heart defect and cleft palate. Prenatal karyotype was 46,XY,?del(6)(q?). During the third trimester, polyhydramnion occurred. The boy was born at the 40th week of gestation by spontaneous delivery. His weight was 3,390 g (>50th centile) and length was 53 cm (>95th centile). He suffered from a congenital heart defect (common atrium), cleft palate, and diaphragmatic hernia. Facial dysmorphism (low forehead, epicanthal folds, prominent eyelids, broad nasal tip, anteverted flared nostrils, long philtrum, micrognathia, dysplastic ears), atypical dermatoglyphs, micropenis, and cryptorchidism were also present. On examination at the age of 8 months his length was 74 cm (>50th centile), weight was 8,720 g (25th centile), and head circumference was 48 cm (95th centile). The boy suffered from developmental delay, seizures, laryngomalacia, bilateral iris coloboma and optic disc hypoplasia, and showed hypermobility of joints. EEG proved focal epilepsy, and CT scan showed global atrophy of the brain, partial agenesis of corpus callosum, and dilatation of the ventricles.

Karyotyping, Array CGH, mBAND, FISH, and Molecular Analysis of the Breakpoints

Cytogenetic analysis of blood lymphocytes was performed using standard protocols. mBAND analysis with the XCyte6 probe (MetaSystems, Altlußheim, Germany) was used for fine assessment of the deletions in both patients. Chromosome 6 painting probe WCP6 (Cambio, Cambridge, UK) was used to exclude a balanced insertional translocation in the parents of patient 2. FISH analysis of patient 1 and her mother was performed with chromosome 6 centromeric probe D6Z1 (Cytocell, Cambridge, UK). Genomic DNA was isolated from blood using Gentra Puregene Blood Kit (Qiagen, Hilden, Germany). Custom array CGH analysis was performed by Nimblegen on the catalogue array HG18_CHR6_FT with median probe spacing of 404 bp, and the results were analysed using SignalMap (Nimblegen, Madison, Wisc., USA). Long-range PCR (LR-PCR) used the Expand Long Template PCR System (Roche, Basel, Switzerland). The primer sequences are available on request. PCR fragments were purified using QIAamp PCR Purification Kit (Qiagen) and sequenced on an ABI PRISM 3100 automatic sequencer (Applied Biosystems, Foster City, Calif., USA). Bioinformatic analysis used the hg18 human genome assembly available in the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

Results

Karyotyping of patient 1 showed a small deletion of 6q11q13. Array CGH mapped the proximal deletion breakpoint to the centromere genome assembly gap (chr 6 Mb 58.89–61.94). The distal breakpoint mapped around Mb 76.55. The total size of the deleted region was ~15 Mb (fig. 1). It contained 34 protein-coding RefSeq genes. The unaffected mother of patient 1 carried a mosaic karyo-

Downloaded by:
Univerzita Karlova v Praze
195.113.40.7 - 7/9/2014 11:23:27 AM

type with the same deletion in all metaphases analysed and, in addition, a small supernumerary marker in 77% of cells, probably a ring chromosome. mBAND analysis confirmed that the marker chromosome constituted of proximal 6q material deleted from the derived chromosome 6 (fig. 2A).

Postnatal cytogenetic analysis of patient 2 revealed a deletion of 6q14q16, and mBAND analysis confirmed the result of the karyotyping (fig. 2B). Chromosome painting excluded a balanced insertional translocation in both parents who had normal karyotypes. Array CGH analysis of patient 2 indicated that the deletion affected chromosome 6 between Mb 77.23–96.63. The length of the deletion was ~19.5 Mb (fig. 1). The deleted region contained 58 protein-coding RefSeq genes.

FISH analysis of the mother of patient 1 with the chromosome 6 centromeric probe showed in all cells analysed a significantly weaker signal from the derivative chromosome 6 compared to the normal homologue, and a weak signal was also present on the marker chromosome (fig. 2C). However, the signals from the normal and deleted chromosome 6 homologues of patient 1 were of similar intensity (fig. 2D).

Fine mapping of the distal deletion breakpoint in patient 1 was a prerequisite for the intended cloning of the proximal breakpoint using inverse PCR. Array CGH suggested that the distal breakpoint mapped to intron 1 of the *MYO6* gene, within a 3.5-kb DNA stretch proximal to nucleotide 76,555,196. Sequence analysis of a series of PCR products of genomic DNA spanning known SNPs and/or possible STRs showed that patient 1 was heterozygous for SNP rs2748963 (G/T) at nucleotide 76,559,033 while all markers in the 10.5-kb interval proximal to this SNP were non-informative (homo- or hemizygous). Heterozygosity for SNP rs2748963 was retained in LR-PCR products extending to nucleotide 76,554,600. This nucleotide was located at the distal end of a 3-kb long contiguous cluster of various Alu and L1 repeats (fig. 3). This information together with the array data indicated that the distal deletion breakpoint in patient 1 was most likely located within this repetitive element cluster, precluding the use of inverse PCR for cloning of the proximal breakpoint located in the centromeric heterochromatin.

In patient 2 the proximal breakpoint mapped between the *IMPG1* and *HTR1B* genes, and the distal breakpoint was located in intron 1 of *FUT9*. The deletion junction was bridged with a LR-PCR product. Sequencing of this DNA fragment showed that the deletion joined nucleotides 77,231,427 and 96,635,141, with 7 nucleotides not belonging to any of the breakpoints added at the junction (fig. 3). No low copy repeats or dispersed repetitive elements were involved in the breakpoints, and their sequences had no homology (fig. 3).

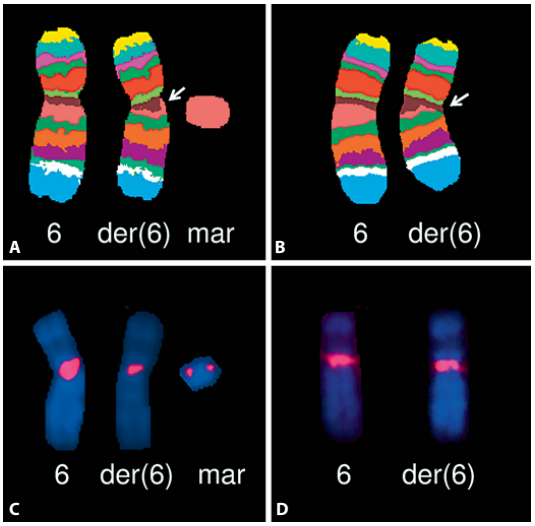


Fig. 2. mBAND analysis of chromosome 6 in the mother of patient 1 (**A**) and in patient 2 (**B**), and FISH analysis of chromosome 6 centromere in the mother of patient 1 (**C**) and in patient 1 (**D**). The mBAND analyses showed small deletions of proximal 6q (arrows) in both cases (**A**, **B**). The marker chromosome in the mother of patient 1 consisted of proximal 6q material deleted from the derived chromosome 6 (**A**). The signals of the chromosome 6 centromeric probe were present both on the deleted homologue and the marker (2 signals may suggest a double ring chromosome) in the mother of patient 1, and the signal on the deleted homologue was weaker than that on the normal chromosome 6 (**C**). In patient 1 no difference in signal intensity could be observed (**D**).

Discussion

To our knowledge, 20 patients with deletions affecting the proximal 6q region and possibly extending to the adjacent part of middle 6q have been analyzed using array CGH and described in the literature [Le Caignec et al., 2005; Klein et al., 2007; Bonaglia et al., 2008; Derwinska et al., 2009; Lespinasse et al., 2009; Traylor et al., 2009; Wang et al., 2009; Spreiz et al., 2010; Woo et al., 2010] or the Decipher database [Firth et al., 2009] (fig. 4). Most of these patients were analysed using low-resolution arrays, and in none of them an attempt was made to clone the

Downloaded by:
Uniwersytet Kalifa v Prze
195.116.40.7 - 7/9/2014 11:23:27 AM

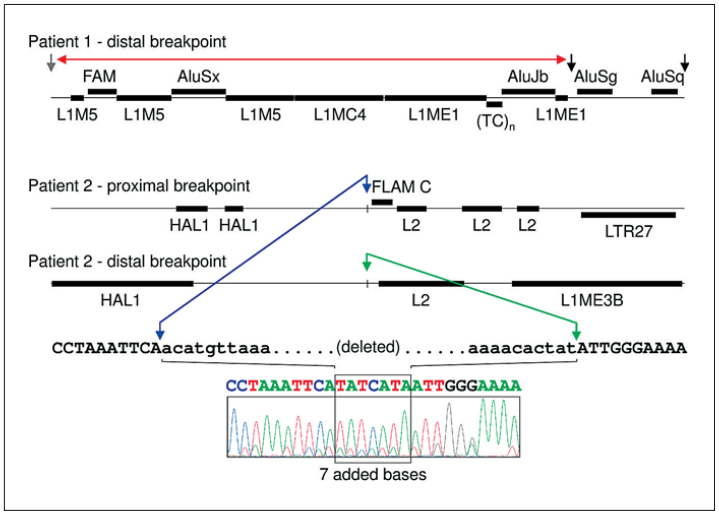


Fig. 3. Repeats at the distal deletion breakpoint in patient 1 and repeat content and sequence of both breakpoints in patient 2. LINE repeats in the 3.5-kb regions around the breakpoints are drawn on the line, SINE repeats above and other repeats below the line. In patient 1 (with the proximal breakpoint in the centromere gap), the distal breakpoint could be mapped only approximately to a cluster of repeats (red horizontal double arrow). The gray arrow marks the position very likely to be single-copy based on the array CGH data. The right black arrow points to heterozygous SNP rs2748963, and heterozygosity was retained on LR-PCR products extending to the position marked by the left black arrow. In patient 2 both the proximal (blue arrows) and distal breakpoints (green arrows) were located in repeat-poor regions. DNA sequencing showed that 7 bases were added at the junction. The deleted sequence is in lowercase.

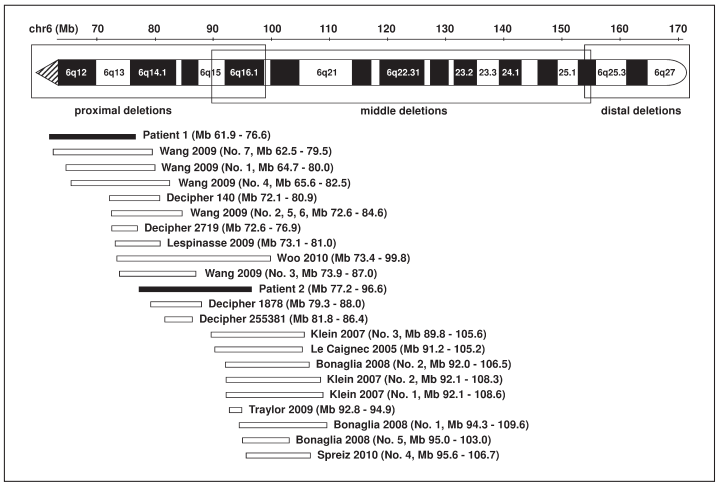


Fig. 4. Alignment of deletions in patients 1 and 2 with published deletions. Only patients with deletions not associated with other rearrangements, fine mapped using array methods and overlapping with but not extending significantly the region covered by deletions in our patients, were included. The ideogram on top shows the classification of 6q deletions.

deletion breakpoints and to elucidate the mechanisms of the aberrations.

The aberration in the mother of our patient 1 was likely caused by centromere fission. This mechanism of parallel formation of a deleted chromosome and a marker, most often a ring chromosome derived from the deleted material, was proposed by Barbara McClintock in 1938 and recently suggested to be referred to by her name [Baldwin et al., 2008]. The marker probably compensated for the deletion and caused the normal phenotype of the mother. However, due to its likely ring structure and the inherent instability of rings [Kosztolanyi, 1987], this marker was present only in a mosaic state, and was not transmitted or was lost in the early development of patient 1. About a dozen of patients who inherited an unbalanced karyotype from a parent with a balanced constitu-

Downloaded by:
Univerzita Karlova v Praze
195.113.40.7 - 7/9/2014 11:23:27 AM

tion involving centromere misdivision have been described [reviewed in Burnside et al., 2011]. Interestingly, while in the mother the centromeric signal from the deleted chromosome 6 was significantly weaker than that from the normal homologue, possibly reflecting the centromere fission, both chromosomes 6 in patient 1 showed signals of equal intensity, which could indicate some kind of healing or regaining of the centromeric alpha satellite arrays on the deleted chromosome. This and the absence of any new constriction on the deleted chromosome suggests that in this case the amount of the alpha satellite material retained on the deleted chromosome was sufficient to support the centromere function. This contrasted with a recently published similar event on chromosome 8 which, however, resulted in neocentromere formation [Burnside et al., 2011].

The distal deletion breakpoint in patient 1 likely involved a dense cluster of dispersed repeats rich in transposable sequences which precluded the cloning of the deletion junction. Interestingly, LINEs and SINEs are known to be present at increased frequency in the arrays of the centromeric alpha satellite repeats [Ugarkovic, 2009], and thus non-allelic homologous recombination may have played a role in the formation of this aberration [Lupski and Stankiewicz, 2005]. On the contrary, the absence of homology and different repeat content in the vicinity of the deletion breakpoints in patient 2 argued against non-allelic homologous recombination as the mechanism causing the aberration. The presence of 7 newly added bases at the deletion junction could be a signature of non-template directed repair associated with non-homologous end joining [Lupski and Stankiewicz, 2005].

The deletions in our 2 patients did not overlap and subdivided the proximal 6q region into 2 parts of almost equal length (fig. 4). The phenotypic abnormalities in our patients were of remarkably different severity. While the phenotype of patient 1 was relatively mild compared to the previously published cases and patient 2, the phenotype of patient 2 was more severe than that of other patients with proximal 6q deletions. Major anomalies were more often described in patients with middle or distal 6q deletions [Hopkin et al., 1997]. As the boundary between the proximal and middle 6q deletions was not exactly defined [Hopkin et al., 1997], the deletion in patient 2 might overlap with the region of the middle deletions.

The deletion of the distal part of proximal 6q in patient 2 may had more severe consequences due to the higher gene density in this region. Although the deletions in patients 1 and 2 were of similar size, the number of genes deleted in patient 2 was almost a double of that in patient 1. Both deletions were too large and with too many genes to allow assigning of individual symptoms to specific genes. Neither of our patients showed symptoms of recessive disorders caused by genes located in the region (*EYS*, *LMBRD1,* and *SLC17A5* in patient 1, and *LCA5*, *BCKDHB*, *SLC35A1,* and *RARS2* in patient 2) indicating lack of mutations on the remaining alleles of these genes. Interestingly, at the age when they were last examined, the patients also did not show any recognizable symptoms of dominant disorders mapping to the deletions: multiple epiphyseal dysplasia type 6 *(COL9A1)* [Czarny-Ratajczak et al., 2001], cone-rod dystrophy type 7 *(RIMS1)* [Johnson et al., 2003], and autosomal dominant nonsyndromic sensorineural hearing loss *(MYO6)* [Melchionda et al., 2001] in patient 1, and autosomal dominant atrophic macular degeneration *(ELOVL4)* [Bernstein et al., 2001; Zhang et al., 2001] in patient 2. The disorders can have a later onset, but other published deletion cases also showed no symptoms, indicating that these disorders may result from gain-of-deleterious-function mutations, although truncating mutations have also been described in at least some of them [Bernstein et al., 2001; Melchionda et al., 2001; Zhang et al., 2001]. Multiple genes in both deletions were good candidates for the developmental delay and mental retardation observed in both patients (e.g. *KHDRBS2*, *BAI3*, *B3GAT2*, *KCNQ5,* and *FILIP1* in patient 1, and *HTR1B*, *ORC3L*, *GABRR1*, *GABRR2,* and *EPHA7* in patient 2).

Some of the genotype-phenotype correlations were contradictory. A study of a small deletion involving only 2 genes suggested that the defect of *EPHA7* could negatively affect brain size and shape and lead to microcephaly. However, microcephaly was not observable in all patients with deletions overlapping the *EPHA7* locus [Traylor et al., 2009], and patient 2, who also has the *EPHA7* gene deleted, rather showed macrocephaly. This implies that some symptoms may be masked by the general population variability, and/or that the effect of some genes can be overridden by the effects of other loci deleted together with the specific gene.

Alignment of individual symptoms observed in our patients and other deletion carriers described in the literature indicated that hydronephrosis and other urinary tract defects might be enriched in patients with deletions overlapping middle 6q13, and similarly other symptoms could be assigned to deletions of other subregions (joint laxity and umbilical hernia to 6q14.1, cleft palate to 6q15, and various eye problems to 6q16.1). Similarly, while macrostomia was prevailing in patients with deletions of

Downloaded by:
Univerzita Karlova v Praze
195.113.40.7 - 7/9/2014 11:23:27 AM

the proximal part of the region studied, microstomia was characteristic for the distal part; and the same applied to upslanting and downslanting palpebral fissures. No obvious candidate genes for these phenotypes could be identified in these regions. Obesity could be characteristic for defects in the *SIM1* gene in the Prader-Willi-like candidate region of 6q16 [Bonaglia et al., 2008], which is distal to the region deleted in our patient 2. However, careful alignment of the published phenotypes also showed that the patients often clustered according to the author of the report rather than the chromosome region deleted, further stressing the importance of standardized phenotyping.

It can be expected that the widespread use of microarray and next-generation sequencing technologies will lead to the identification of small aberrations in patients with less complex phenotypes, which will be more informative for the correlation of individual symptoms with specific genes in the proximal 6q region.

Acknowledgements

Supported by grants MZOFNM2005, MZOVFN2005, and CHERISH (EC FP7 223692). The study makes use of data generated by the DECIPHER Consortium (http://decipher.sanger.ac.uk) funded by the Wellcome Trust.

References

Baldwin EL, May LF, Justice AN, Martin CL, Ledbetter DH: Mechanisms and consequences of small supernumerary marker chromosomes: from Barbara McClintock to modern genetic-counseling issues. Am J Hum Genet 82:398–410 (2008).

Bernstein PS, Tammur J, Singh N, Hutchinson A, Dixon M, et al: Diverse macular dystrophy phenotype caused by a novel complex mutation in the *ELOVL4* gene. Invest Ophthalmol Vis Sci 42:3331–3336 (2001).

Bonaglia MC, Ciccone R, Gimelli G, Gimelli S, Marelli S, et al: Detailed phenotype-genotype study in five patients with chromosome 6q16 deletion: narrowing the critical region for Prader-Willi-like phenotype. Eur J Hum Genet 16:1443–1449 (2008).

Burnside RD, Ibrahim J, Flora C, Schwartz S, Tepperberg JH, et al: Interstitial deletion of proximal 8q including part of the centromere from unbalanced segregation of a paternal deletion/marker karyotype with neocentromere formation at 8p22. Cytogenet Genome Res 132:227–232 (2011).

Czarny-Ratajczak M, Lohiniva J, Rogala P, Kozlowski K, Perala M, et al: A mutation in *COL9A1* causes multiple epiphyseal dysplasia: further evidence for locus heterogeneity. Am J Hum Genet 69:969–980 (2001).

Derwinska K, Bernaciak J, Wisniowiecka-Kowalnik B, Obersztyn E, Bocian E, et al: Autistic features with speech delay in a girl with an approximately 1.5-Mb deletion in 6q16.1, including *GPR63* and *FUT9*. Clin Genet 75:199–202 (2009).

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, et al: DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 84:524–533 (2009).

Hopkin RJ, Schorry E, Bofinger M, Milatovich A, Stern HJ, et al: New insights into the phenotypes of 6q deletions. Am J Med Genet 70:377–386 (1997).

Johnson S, Halford S, Morris AG, Patel RJ, Wilkie SE, et al: Genomic organisation and alternative splicing of human *RIM1*, a gene implicated in autosomal dominant cone-rod dystrophy (CORD7). Genomics 81:304–314 (2003).

Klein OD, Cotter PD, Moore MW, Zanko A, Gilats M, et al: Interstitial deletions of chromosome 6q: genotype-phenotype correlation utilizing array CGH. Clin Genet 71:260–266 (2007).

Kosztolanyi G: Does 'ring syndrome' exist? An analysis of 207 case reports on patients with a ring autosome. Hum Genet 75:174–179 (1987).

Le Caignec C, Swillen A, Van Asche E, Fryns JP, Vermeesch JR: Interstitial 6q deletion: clinical and array CGH characterisation of a new patient. Eur J Med Genet 48:339–345 (2005).

Lespinasse J, Gimelli S, Bena F, Antonarakis SE, Ansermet F, et al: Characterization of an interstitial deletion 6q13-q14.1 in a female with mild mental retardation, language delay and minor dysmorphisms. Eur J Med Genet 52:49–52 (2009).

Lupski JR, Stankiewicz P: Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet 1:e49 (2005).

Melchionda S, Ahituv N, Bisceglia L, Sobe T, Glaser F, et al: *MYO6*, the human homologue of the gene responsible for deafness in Snell's waltzer mice, is mutated in autosomal dominant nonsyndromic hearing loss. Am J Hum Genet 69:635–640 (2001).

Spreiz A, Muller D, Zotter S, Albrecht U, Baumann M, et al: Phenotypic variability of a deletion and duplication 6q16.1–q21 due to a paternal balanced ins(7;6)(p15;q16.1q21). Am J Med Genet A 152A:2762–2767 (2010).

Traylor RN, Fan Z, Hudson B, Rosenfeld JA, Shaffer LG, et al: Microdeletion of 6q16.1 encompassing *EPHA7* in a child with mild neurological abnormalities and dysmorphic features: case report. Mol Cytogenet 2:17 (2009).

Ugarkovic DI: Centromere-competent DNA: structure and evolution. Prog Mol Subcell Biol 48:53–76 (2009).

Wang JC, Dang L, Lomax B, Turner L, Shago M, et al: Molecular breakpoint mapping of 6q11-q14 interstitial deletions in seven patients. Am J Med Genet A 149A:372–379 (2009).

Woo KS, Kim JE, Kim KE, Kim MJ, Yoo JH, et al: A de novo proximal 6q deletion confirmed by array comparative genomic hybridization. Korean J Lab Med 30:84–88 (2010).

Zhang K, Kniazeva M, Han M, Li W, Yu Z, et al: A 5-bp deletion in *ELOVL4* is associated with two related forms of autosomal dominant macular dystrophy. Nat Genet 27:89–93 (2001).

Downloaded by:
Univerzita Karlova v Praze
195.113.40.7 - 7/9/2014 11:23:27 AM

Publikace 3

Skin lesions in a boy with X-linked lymphoproliferative disorder: comparison of 5 *SH2D1A* deletion cases.

Mejstrikova E, Janda A, Hrusak O, Buckova H, Vlckova M, Hancarova M, Freiberger T, Ravcukova B, Vesely K, Fajkusova L, Kopeckova L, Sumerauer D, Kabickova E, Sediva A, Stary J, Sedlacek Z.

Pediatrics. 2012 Feb;129(2):e523-8. IF: 5,119

Studie je výsledkem mezioborové spolupráce a je zaměřena na velmi vzácný X-vázaný lymfoproliferativní syndrom 1 (XLP-1) způsobený mutacemi v genu *SH2D1A*. V publikaci je popsáno celkem 5 pacientů ze dvou rodin. U tří z nich byla provedena podrobná molekulárně genetická analýza. V práci je diskutován zejména atypický fenotyp pacienta A1. Detailní molekulárně genetická analýza měla napomoci k vysvětlení jeho neobvyklého fenotypu.

XLP-1 je onemocnění projevující se těžkou dysregulací imunity s řadou dalších konsekvencí. V práci popisujeme pacienta A1, který netrpěl typickými symptomy XLP-1 a který byl opakovaně vyšetřován pro velmi těžké kožní onemocnění svými projevy imitující epidermolysis bullosa. Pacient měl navíc opožděný vývoj řeči, ADHD, poruchy učení a epileptický záchvat ve věku 3 let. U maternálního bratrance se ve věku 16 let manifestoval Burkittův lymfom s následnou těžkou aplastickou anemií, což vedlo k podezření na XLP-1. Ten byl potvrzen molekulárně genetickým průkazem delece celého genu *SH2D1A*. Stejná delece byla následně zjištěna i u pacienta A1.

Detailní analýzou jsme prokázali, že delece je u obou bratranců identická a zasahuje i část genu *ODZ1* (dle novější nomenklatury *TENM1),* jehož podíl na fenotypu je nejasný. Významný rozdíl ve fenotypových projevech u obou bratranců a dosud nepopsaná asociace XLP-1 s kožními projevy, kterými trpěl pacient A1, nás vedla k vyšetření pacienta B1 z nepříbuzné rodiny, který měl rovněž projevy XLP-1 a prokázanou deleci celého genu *SH2D1A*. Porovnání delecí u všech tří pacientů, stejně tak jako údaje v literatuře, svědčily spíše proti vlivu genu *ODZ1* na fenotyp, neboť pacient B1 nesl deleci zasahující větší část tohoto genu a nevykazoval žádné symptomy netypické pro XLP-1, stejně tak jako několik dříve publikovaných pacientů s delecemi v této oblasti nemělo žádné netypické projevy. Atypické kožní projevy u pacienta A1 jsou tak zřejmě důsledkem samotného XLP-1 a mohly vzniknout na podkladě autoimunitní reakce při komplexním narušení regulace imunitní odpovědi,

která je pro XLP-1 typická. Naše práce tedy rozšiřuje poznatky o možných projevech XLP-1 a zároveň přispívá k poznatkům o funkci genu *ODZ1*.

Pozdější práce uvádějí souvislost mezi bodovými mutacemi v genu *ODZ1* (resp. *TENM1)* s anosmií (Alkelai et al., 2016) a s mozkovou obrnou (McMichael et al., 2015). Ani jeden z našich pacientů však v době vyšetření tyto příznaky nevykazoval. Zdravotní stav pacienta B1 byl při naší poslední klinické kontrole v roce 2015 velmi dobrý.

Skin Lesions in a Boy With X-linked Lymphoproliferative Disorder: Comparison of 5 *SH2D1A* Deletion Cases

AUTHORS: Ester Mejstříková, MD, PhD,[a,b] Aleš Janda, MD, PhD,[a,c] Ondřej Hrušák, MD, PhD,[a,b] Hana Bučková, MD, PhD,[d] Markéta Vlčková, MD,[e] Miroslava Hančárová, MSc,[e] Tomáš Freiberger, MD, PhD,[f] Barbora Ravčuková, MSc,[f] Karel Veselý, MD, PhD,[g] Lenka Fajkusová, PhD,[h] Lenka Kopečková, PhD,[h] David Sumerauer, MD, PhD,[a] Edita Kabíčková, MD,[a] Anna Šedivá, MD, PhD,[i] Jan Starý, MD, PhD,[a] and Zdeněk Sedláček, PhD[e]

Departments of [a]Pediatric Hematology and Oncology, [e]Biology and Medical Genetics, and [i]Immunology, and [b]Childhood Leukemia Investigation Prague (CLIP), Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic; [c]Centre of Chronic Immunodeficiency (CCI), University Medical Centre Freiburg and University of Freiburg, Freiburg, Germany; [d]Dermatological Division, Department of Pediatrics, University Hospital Brno, Medical Faculty of Masaryk University, Brno, Czech Republic; [f]Molecular Genetics Laboratory, Centre for Cardiovascular Surgery and Transplantation, and Department of Clinical Immunology and Allergology, Masaryk University, Brno, Czech Republic; [g]1st Institute of Pathological Anatomy, St. Anne's Hospital and Medical Faculty of Masaryk University, Brno, Czech Republic; [h]Centre of Molecular Biology and Gene Therapy, University Hospital Brno and Masaryk University, Brno, Czech Republic

KEY WORDS

SH2D1A, *ODZ1*, SAP, teneurin, X-linked lymphoproliferative syndrome, aplastic anemia, lymphoma, dermatitis, psoriasis, dermatosclerosis, vesiculobullous skin disease

ABBREVIATIONS

EB—epidermolysis bullosa
EBV—Epstein-Barr virus
FIM—fulminant infectious mononucleosis
HLH—hemophagocytic lymphohistiocytosis
NKT—natural killer T lymphocytes
XLP—X-linked lymphoproliferative disorder

Drs Mejstříková and Janda contributed equally to this work.

www.pediatrics.org/cgi/doi/10.1542/peds.2011-0870

doi:10.1542/peds.2011-0870

Address correspondence to Aleš Janda, MD, PhD, Centre of Chronic Immunodeficiency (CCI), University Medical Center Freiburg and University of Freiburg, Breisacher Str. 117, 79095 Freiburg, Germany. E-mail: ales.janda@uniklinik-freiburg.de

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2012 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: *Dr Janda received an unrestricted fellowship grant from the European Society for Immunodeficiencies provided by Baxter; the other authors have indicated they have no financial relationships relevant to this article to disclose.*

abstract

SH2D1A gene defects are the cause of X-linked lymphoproliferative disorder (XLP-1), a rare condition characterized by severe immune dysregulation. We present a patient lacking the typical symptoms of XLP-1, but experiencing a severe unusual skin condition encompassing features of dermatosclerosis and vesiculobullous skin disease. A maternal cousin of the patient was diagnosed with XLP-1 and found to carry a deletion of the *SH2D1A* gene. *SH2D1A* deletion was also identified in our patient, which offered a possible explanation for his skin symptoms. Subsequent analysis showed that the deletion in both cousins was identical and involved the whole *SH2D1A* gene and a part of the adjacent *ODZ1* gene. High phenotypic variability of XLP-1 observed in this family prompted us to analyze the genotype-phenotype correlation of 2 different-sized deletions involving *SH2D1A* and *ODZ1* in 5 patients from 2 families, and we report the clinical and laboratory data on these individuals. Our findings illustrate the wide clinical variability of XLP-1, both inter- and intrafamilial, which may complicate the diagnosis of this condition. The comparison of phenotypes of our patients argues against a strong involvement of the *ODZ1* gene in the skin disorder and other symptoms observed in our index patient. His hitherto not described severe skin condition extends the phenotypic range of XLP-1. *Pediatrics* 2012;129:e523–e528

Downloaded from pediatrics.aappublications.org at 2. Lekarska Fakulta UK on September 10, 2015

The X-linked *SH2D1A* gene encodes SAP (Signaling Lymphocyte Activation Molecule [SLAM]–Associated Protein), a small adaptor protein. SAP interacts with SLAM family members and assists their tyrosine-phosphorylation signaling important in most lymphocyte subsets.[1,2] SAP also exerts proapoptotic function in the restimulation-induced death of activated lymphocytes.[3] Impairment of SAP affects T, natural killer, natural killer T lymphocyte (NKT), and indirectly also B cells, resulting in varied immune defects.[4] SAP deficiency typically manifests in childhood or early adolescence as a rare primary immunodeficiency, X-linked lymphoproliferative disease (XLP-1, Online Mendelian Inheritance in Man [OMIM] #308240), affecting 1 to 3 per million boys.[5,6] The typical phenotype includes hemophagocytic lymphohistiocytosis (HLH) and/or fulminant infectious mononucleosis (FIM) and dysgammaglobulinemia (frequently progressing into hypogammaglobulinemia). Approximately 25% of XLP-1 patients have non-Hodgkin lymphoma.[7] Aplastic anemia, pulmonary lymphoid granulomatosis, colitis, necrotizing vasculitis, and chronic active gastritis are less common.[7–9] Recently, *BIRC4* was identified as another causal X-linked gene in a minority of patients (XLP-2, OMIM #300635). It encodes the inhibitor-of-apoptosis protein (XIAP), a caspase inhibitor.[10] Although SAP and XIAP have antagonistic functions in apoptosis, the phenotypes of XLP-1 and XLP-2 are strikingly similar. The minor differences include the lack of NKT cells and higher incidence of hemorrhagic colitis in XLP-1, and the absence of lymphoma in XLP-2.[11,12] Although Epstein-Barr virus (EBV) has always been assumed to play a decisive role in XLP, it is now clear that a significant proportion of patients are EBV negative at presentation, and some do not develop FIM despite the EBV infection.[7,8,13] XLP therefore represents a general immune dysregulation rather than a defect of the anti-EBV response.[7] Recently, lymphoproliferation and FIM have also been ascribed to the deficiency of interleukin-2-inducible T-cell kinase[14] (autosomal EBV-associated lymphoproliferative syndrome 1, OMIM #613011).

We present a male with an atypical skin condition who was found to carry a deletion of the whole *SH2D1A* and a part of the adjacent *ODZ1* gene, and report clinical and laboratory data of 2 families with 5 affected males carrying similar deletions. Our findings emphasize the wide clinical variability of XLP-1, both inter- and intrafamilial. The hitherto not described severe skin condition in our index patient extends the phenotypic range of XLP-1.

PATIENT PRESENTATIONS

Patient A1

Patient A1, a boy from family A, had an episode of transient unconsciousness at the age of 3 years. Focal epileptiform discharges were found on electroencephalogram and antiepileptic therapy was initiated. Apart from the initial attack, no clinical symptoms recurred, and the cranial computed tomography scan was normal. The antiepileptic therapy was discontinued at 13 years of age, and the patient remained free of central neurologic symptoms. He also had attention deficit hyperactivity disorder, dyslexia, dysgraphia, and retarded speech development in childhood. However, his cognitive and intellectual abilities, memory, and emotional responses were normal when tested at 26 years of age.

At the age of 8 years he developed atrophic skin and scaling on the dorsal parts of his hands and feet, accompanied by dyspigmentation on the extremities and trunk. The condition was classified as psoriasis. After histologic examination, it was redefined as sclerodermia-like syndrome. The idiosyncratic effects of antiepileptic drugs (carbamazepine, phenobarbital, and valproic acid) were suspected, but changing the medication had no effect on the skin eruptions. Two years after manifestation, the infliction progressed into tiny droplet-like depigmentation around joints with bullae at the acral parts of the extremities. The severity of the condition culminated at the age of 13 years. A form of dystrophic epidermolysis bullosa (EB) was considered, but laboratory investigations did not support this diagnosis (Supplemental Information). The patient was treated with topical medication and intermittently with systemic steroids. The skin infliction oscillated with gradual improvement and stabilization; however, with remarkable scarring of the subcutaneous tissue resulting in joint movement limitation (elbows, knees, hands) and severe mutilation of hands and feet (Fig 1). The bullous eruptions have not reappeared in adulthood. Repeated extensive investigation of the residual lesions at the age of 26 years (histology, immunofluorescence, electron microscopy, molecular genetics; Supplemental Information) yielded additional evidence against EB. Only nonspecific inflammation in the areas of previously present bullae was seen; no structural abnormities and pathologic antibodies in situ or in serum were detected. Mucous membranes have never been involved. Muscular hypotrophy, nail atrophy, and carious dentition were also present, and electromyogram revealed a mild neurogenic lesion. The patient was EBV negative.

The diagnosis of XLP-1 and identification of a *SH2D1A* deletion (Supplemental Information) in patient A2, a maternal cousin of patient A1 (see below), prompted us to also test patient A1. His *SH2D1A* gene was deleted, which yielded a possible new explanation for his

Downloaded from pediatrics.aappublications.org at 2. Lekarska Fakulta UK on September 10, 2015

FIGURE 1
Current residual skin lesions of patient A1. Atrophic skin with hyperkeratotic scales on elbows and knees (A, B), as well as on acral parts of his extremities (C, D) (status at 26 years of age). Atypical pseudosyndactyly and finger contracture with nail atrophy on 1 of the hands (C).

symptoms. To assist the genotype-phenotype correlation, patient A1 was compared with four other patients from 2 families. Clinical and laboratory features of all patients are summarized in Table 1.

Patient A2

Patient A2 presented at the age of 16 years with generalized cervical, axillary, and abdominal lymphadenopathy, tiredness, wasting, ascites, and rapidly progressing pleural effusions, with no night sweats or fever. The diagnosis of Burkitt-like lymphoma with t(8;14)(q24;q32) translocation was established. The patient was treated according to the Berlin-Frankfurt-Munich B-cell Non-Hodgkin Lymphoma (BFM B-NHL) 2004 protocol,[15] achieving full remission in 5 months. Three months later, after a viral infection, marked pancytopenia occurred. fluorodeoxyglucose positron-emission tomography/computed tomography excluded lymphoma relapse. Bone marrow trephine biopsy confirmed the diagnosis of severe aplastic anemia. The patient underwent an uncomplicated transplantation of granulocyte-colony stimulating factor–mobilized peripheral blood progenitor cells from his healthy 9/10 HLA-matched brother. Currently, 24 months after transplantation, he is alive and well, requiring no supportive therapy. Prompted by the uncommon association of lymphoma and severe aplastic anemia, *SH2D1A* testing revealed a deletion of all exons. Additional analysis showed that both patients A1 and A2 carried identical deletions of *SH2D1A* and a part of *ODZ1* (Fig 2).

Patient A3

Retrospectively, another affected boy, patient A3, was identified in family A. This maternal uncle of patients A1 and A2 was born in 1961 and died of acute liver failure due to probable FIM at 5 years of age. No other details or materials were available.

Patient B1

Patient B1, from family B unrelated to family A, developed bronchopneumonia, fluidothorax, and hepatosplenomegaly at the age of 4 years, resulting in a multiorgan failure. Identification of activated histiocytes in ascitic fluid led to the diagnosis of malignant histiocytosis, and the patient was promptly treated with a combined chemotherapy. Bone marrow transplantation was recommended but not approved by his parents. Within a few months after chemotherapy, panhypogammaglobulinemia manifested. Since then he has required immunoglobulin replacement. Genetic analysis revealed a deletion of *SH2D1A* and a part of *ODZ1*, larger than that in family A (Fig 2).

Patient B2

A younger brother of patient B1 succumbed to FIM and malignant histiocytosis at the age of 2 years. No other details or materials were available.

DISCUSSION

Only a few reports exist on dermatological symptoms in XLP-1 patients. Skin lichenification, depigmentation, or psoriasis have been documented.[7,16] There are no reports on a severe bullous skin condition with scaling associated with XLP. We suppose that the lesions in patient A1 could be of autoimmune origin and associated with immune dysregulation related to his SAP deficiency.

No autoantibodies expected of a bullous condition could be detected, possibly because we could only thoroughly examine the patient years after the disease flare when the skin changes were residual. The bioptic material from the enhanced disease-activity phase was not available for reanalysis. The hypothetical temporary presence of autoantibodies could be explained by the disruption of B-lymphocyte homeostasis. Hypergammaglobulinemia gradually progressing to hypogammaglobulinemia is a known feature of XLP-1, and autoreactive B-cell clones may disappear in parallel with normal B cells.

Downloaded from pediatrics.aappublications.org at 2. Lekarska Fakulta UK on September 10, 2015

TABLE 1 Clinical and Laboratory Characteristics of the Patients

Characteristics	Patient				
	A1	A2	A3	B1	B2
Year of birth	1984	1991	1961	1988	1993
Age at presentation, y	8	16	5	4	2
HLH/FIM	No	no	Yes[a]	Yes[a]	Yes[a]
Lymphoma	No	Yes[a]	No	No	No
Aplastic anemia	No	Yes	No	No	No
Dysgammaglobulinemia	Yes	No	NA	Yes[b]	Yes
Other symptoms	Severe skin lesions,[a] extensive dental caries, 1 seizure, abnormal EEG, ADHD	Eufunctional goiter	NA	Extensive dental caries	Coombs positive anemia
B cells	↓ class-switched memory cells, ↑ transitional cells, normal marginal zone B cells	NA	NA	↓ class-switched memory cells, ↓ marginal zone B cells	NA
NKT	↓↓	NA	NA	↓↓	NA
EBV status	Negative[c]	Negative[c]	NA	Positive[d]	Positive[e]
Treatment	Antiepileptics, systemic steroids, skin treatment	Chemotherapy, HSCT	NA	Chemotherapy, IVIg	Antibiotics, supportive treatment
Age at last follow-up, y	26	19	5	22	2
Outcome	Alive, experiencing sequelae of severe skin lesions	Alive and well	Deceased	Alive	Deceased

↑, increased; ↓, decreased; ↓↓, strongly decreased; NA, not available or not done; ADHD, attention-deficit/hyperactivity disorder; HSCT, hematopoietic stem cell transplantation; Ig, immunoglobulin; PCR, polymerase chain reaction.

[a] Presenting symptoms.

[b] Progressing into hypogammaglobulinemia.

[c] Serology and PCR.

[d] Viral capsid antibody IgA, IgG, IgM positive, Epstein-Barr nuclear antigen IgG negative, EBV PCR not performed.

[e] Heterophilic antibodies positive.

Grosieux et al[16] described a patient with signs of XLP presenting with pityriasis lichenoides and polyneuropathy explained by progressive systemic vasculitis. Several other cases of systemic vasculitis in XLP have been reported (reviewed in ref 17). Repeated skin biopsies yielded no evidence of vasculitis in patient A1. In addition, the course of his disease differed from the reported cases where it was acute, with fast progression and fatal outcome despite immunosuppression therapy.

The picture of the skin condition, the absence of multiorgan involvement, and the long-term usage of antiepileptics before the occurrence of skin symptoms argue against anticonvulsant hypersensitivity syndrome.[18] However, we cannot exclude the impact of this therapy on the skin in a SAP-deficient environment. We also cannot fully exclude other concurrent genetic defects responsible for the skin defect, but our extensive analysis (Supplemental Information) did not support this possibility, and we consider this scenario unlikely.

We also speculated that the additional symptoms in patient A1 could have been due to an expansion of the *SH2D1A* deletion during its transmission in family A, a phenomenon rarely described in other microdeletions.[19] This hypothesis was not confirmed; the deletions in both cousins were identical, involving only *SH2D1A* and *ODZ1*. The phenotypic impact of the *ODZ1* defect is unclear. *ODZ1* codes 1 of 4 teneurins, a family of transmembrane proteins with predominantly neural expression. Their exact function is unknown, but it may include cell adhesion, signaling, and gene regulation in neuronal development.[20] It is tempting to attribute the neurologic, learning, and behavioral deficits of patient A1 to the truncation of his *ODZ1* gene, but the absence of these symptoms in patient A2, who has the same deletion, and in patient B1, whose *ODZ1* deletion is even larger, is at odds with this explanation. Our doubts on the effects of *ODZ1* deficiency are also supported by previous observations of no additional symptoms in several XLP-1 families with 3.5-Mb deletions that include *ODZ1*.[5]

The only curative option for patients with XLP-1 is an allogeneic transplantation of hematopoietic cells. However, the course of the disease is highly variable even in families with the same genetic defects, and a substantial number of long-term untransplanted survivors have been reported.[7] The patients who profit most from the transplantation are those presenting with HLH, FIM, or lymphoma.[7] This treatment option was not offered to our index patient, because his clinical situation was stable, and we were not convinced that the benefit of the transplantation would outweigh its risks. He is carefully monitored, and transplantation would be undertaken

Downloaded from pediatrics.aappublications.org at 2. Lekarska Fakulta UK on September 10, 2015

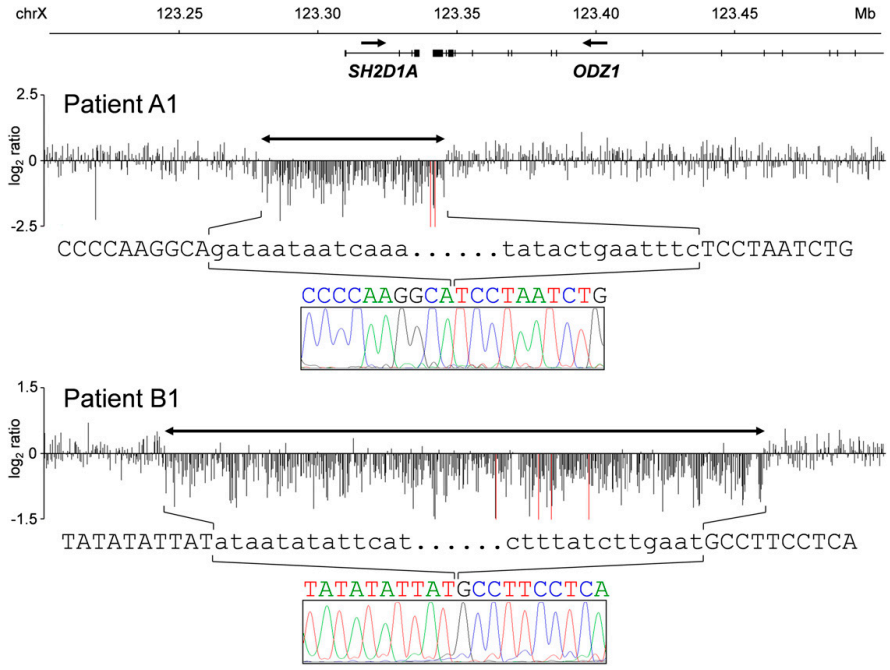


FIGURE 2

Array CGH and sequence analysis of the SH2D1A and ODZ1 deletions. Regions of decreased array CGH signals mark the deletions (double arrows; red signals exceed the graph range). Deleted sequences are in lower case; sequences of the junctions are in color. The deletion in patient A1 (and A2) is smaller and fully contained in the deletion of patient B1. CGH, comparative genomic hybridization.

in case of disease progression and the onset of other manifestations.

We conclude that XLP-1 should be considered in males with classic presenting symptoms, but also in males with non-Hodgkin lymphoma associated with severe aplastic anemia or skin lesions. A positive family history should prompt genetic testing, and once XLP-1 diagnosis is established, all males at risk should be monitored even when asymptomatic or presenting with atypical symptoms.

ACKNOWLEDGMENTS

This work was supported by grants IGA NS/9997-4, NS/10480-3, NS/10398-3, MSM0021620813, MZOFNM2005, and BMBF 01 EO 0803.

REFERENCES

1. Sayos J, Wu C, Morra M, et al. The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. *Nature*. 1998;395(6701):462–469
2. Bolino A, Yin L, Seri M, et al. A new candidate region for the positional cloning of the XLP gene. *Eur J Hum Genet*. 1998;6(5):509–517
3. Snow AL, Marsh RA, Krummey SM, et al. Restimulation-induced apoptosis of T cells is impaired in patients with X-linked lymphoproliferative disease caused by SAP deficiency. *J Clin Invest*. 2009;119(10):2976–2989
4. Nichols KE, Ma CS, Cannons JL, Schwartzberg PL, Tangye SG. Molecular and cellular pathogenesis of X-linked lymphoproliferative disease. *Immunol Rev*. 2005;203:180–199
5. Sumegi J, Huang DL, Lanyi A, et al. Correlation of mutations of the SH2D1A gene and Epstein-Barr virus infection with clinical phenotype and outcome in X-linked lymphoproliferative disease. *Blood*. 2000;96(9):3118–3125
6. Purtilo DT. X-linked lymphoproliferative disease (XLP) as a model of Epstein-Barr virus-induced immunopathology. *Springer Semin Immunopathol*. 1991;13(2):181–197
7. Booth C, Gilmour KC, Veys P, et al. X-linked lymphoproliferative disease due to SAP/SH2D1A deficiency: a multicenter study on the manifestations, management and outcome of the disease. *Blood*. 2011;117(1):53–62
8. Brandau O, Schuster V, Weiss M, et al. Epstein-Barr virus-negative boys with non-Hodgkin lymphoma are mutated in the SH2D1A gene, as are patients with X-linked lymphoproliferative disease (XLP). *Hum Mol Genet*. 1999;8(13):2407–2413
9. Rougemont AL, Fournet JC, Martin SR, et al. Chronic active gastritis in X-linked

Downloaded from pediatrics.aappublications.org at 2. Lekarska Fakulta UK on September 10, 2015

lymphoproliferative disease. *Am J Surg Pathol*. 2008;32(2):323–328

10. Rigaud S, Fondanèche MC, Lambert N, et al. XIAP deficiency in humans causes an X-linked lymphoproliferative syndrome. *Nature*. 2006;444(7115):110–114
11. Filipovich AH, Zhang K, Snow AL, Marsh RA. X-linked lymphoproliferative syndromes: brothers or distant cousins? *Blood*. 2010; 116(18):3398–3408
12. Pachlopnik Schmid J, Canioni D, Moshous D, et al. Clinical similarities and differences of patients with X-linked lymphoproliferative syndrome type 1 (XLP-1/SAP deficiency) versus type 2 (XLP-2/XIAP deficiency). *Blood*. 2011;117(5):1522–1529
13. Strahm B, Rittweiler K, Duffner U, et al. Recurrent B-cell non-Hodgkin's lymphoma in two brothers with X-linked lymphoproliferative disease without evidence for Epstein-Barr virus infection. *Br J Haematol*. 2000;108(2):377–382
14. Huck K, Feyen O, Niehues T, et al. Girls homozygous for an IL-2-inducible T cell kinase mutation that leads to protein deficiency develop fatal EBV-associated lymphoproliferation. *J Clin Invest*. 2009;119(5): 1350–1358
15. Reiter A. Diagnosis and treatment of childhood non-Hodgkin lymphoma. *Hematology(Am Soc Hematol Educ Program*. 2007;285–296
16. Grosieux C, Amoric JC, Mechinaud F, et al. Cutaneous and neurologic vasculitis disclosing EBV-selective immunodeficiency [in French]. *Ann Dermatol Venereol*. 1996;123 (6–7):387–392
17. Talaat KR, Rothman JA, Cohen JI, et al. Lymphocytic vasculitis involving the central nervous system occurs in patients with X-linked lymphoproliferative disease in the absence of Epstein-Barr virus infection. *Pediatr Blood Cancer*. 2009;53(6):1120–1123
18. Mansur AT, Pekcan Yasar S, Göktay F. Anticonvulsant hypersensitivity syndrome: clinical and laboratory features. *Int J Dermatol*. 2008;47(11):1184–1189
19. South ST, Rope AF, Lamb AN, et al. Expansion in size of a terminal deletion: a paradigm shift for parental follow-up studies. *J Med Genet*. 2008;45(6):391–395
20. Tucker RP, Kenzelmann D, Trzebiatowska A, Chiquet-Ehrismann R. Teneurins: transmembrane proteins with fundamental roles in development. *Int J Biochem Cell Biol*. 2007;39(2):292–297

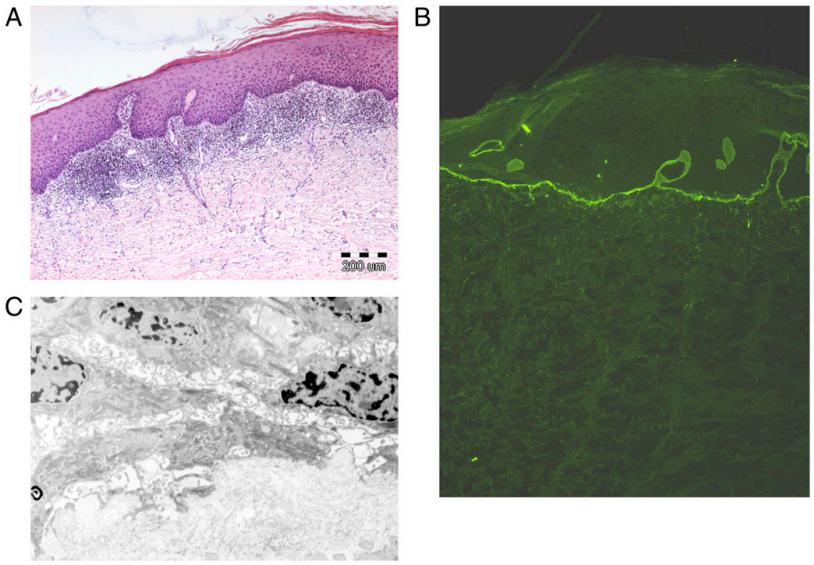
Downloaded from pediatrics.aappublications.org at 2. Lekarska Fakulta UK on September 10, 2015

Supplemental Information

SKIN CONDITION IN PATIENT A1

Epidermolysis bullosa (EB) was considered in differential diagnosis at the age of 13 years because of blister formation, although other symptoms including hyperkeratous deposits above the joints and in the acrae, finger adhesions, or atrophic changes above the shoulder blades did not support this diagnosis. Serum autoantibodies (ANA, ENA, ds-DNA, Scl-70, ANCA, ACLA, and anti-BP180/BP230) were tested negative by enzyme-linked immunosorbent assay. The course of the disorder with its late onset was also at odds with the diagnosis of EB, as was the histologic analysis which rather revealed atypical hyper- and parakeratosis with strong neutrophilic infiltration.

To definitely exclude the diagnosis of EB at the age of 26 years, several excisional skin biopsies involving peri-lesional regions were taken and analyzed using light microscopy (Supplemental Fig 3A), direct immunofluorescence (Supplemental Fig 3B), transmission electron microscopy (Supplemental Fig 3C), and immunofluorescent antigen mapping of the dermoepidermal basement membrane zone by the use of a standard panel of antibodies for EB diagnostics that included antibodies against cytokeratins 5 and 14, dystonin, laminin 5, and collagens IV, VII, and XVII. These analyses failed to reveal any defect related to EB. Direct immunofluorescence analysis by using a standard panel of antibodies for the diagnostics of vesiculobullous diseases, lupus erythematodes, and vasculitis that included antibodies against IgG, IgM, IgA, C3b, and fibrinogen was also performed. In the area of skin erosion only nonspecific deposits of fibrin were present at the dermoepidermal junction; other immunodeposits were not revealed. Observed acanthosis could be ascribed to regenerative changes in the epidermis rather than to EB, because bullous eruptions were not noted any more during the patient's adulthood. The above-mentioned serum autoantibodies were tested negative again.



SUPPLEMENTAL FIGURE 3

Histologic examination of one of the lesions of patient A1 arguing against EB. A, Epidermis with acanthosis, hyperparakeratosis, and mild focal spongiosis. Dense lymphocytic inflammatory infiltrate in the upper dermis and dermal fibrosis can be observed (magnification 40×). B, Immunofluorescence antigen mapping showing strong signal of collagen VII in the epidermal basement membrane zone (magnification 100×). C, Electron microscopic investigation shows normal components of dermoepidermal basement membrane zone.

GENETICS

EB-Related Genes Sequencing in Patient A1

Genomic DNA was isolated from blood lymphocytes by using the Gentra Puregene Blood kit (Qiagen, Hilden, Germany). Because of the diagnostic uncertainty sequence analysis of three genes involved in EB—*COL7A1* (dystrophic form of EB), and *KRT5* and *KRT14* (EB simplex)—was performed even when defects of the respective proteins were not detected by using immunohistochemistry. Sequencing of all exons and adjacent intron regions[22,23] did not identify any mutation.

SH2D1A Gene Analysis in Patients A1, A2, and B1

Exons of the *SH2D1A* gene were amplified in separate reactions by using the following primers: 1F: AAGCATCCCC GTCTGAGTG, 1R: CCTCAGCAGGCCCATGTC, 2F: CTTCCTATGAATGCAATGACACCA, 2R: CC TCCTTGACACCCCCAGAA, 3F: AATGTTGTTTT TCCTTCTTCCTTATGTT, 3R: AACTCTGAGCTT CCAAACCCTGTC, 4F: TGTGTCATTGTGAGTTT TATGCAGTTGGA, and 4R: CTTTAGAAATGCA TTTGTAGCTCACCGAAC. Control primers C5: TGCCAAGTGGAGCACCCAA and C3: GCA TCTTGCTCTGTGCAGAT amplifying intron 3 of the *HLA-DRB1* gene were used in each reaction as a positive control. No *SH2D1A*-specific products were detected in these three patients, indicating large *SH2D1A* defects.

Array CGH Analysis and Sequencing of Deletion Breakpoints in Patients A1, A2, and B1

Custom array comparative genomic hybridization (CGH) analysis was performed by Nimblegen on chromosome X array HG18_CHRX_FT, and the results were analyzed by using SignalMap (Nimblegen, Madison, WI). The deletion was located between Mb 123.28 and 123.35 in patient A1, and between Mb 123.25 and 123.46 in patient B1. Array CGH did not point to any other larger DNA gains or losses of chromosome X. Mapping and sequencing of the breakpoints were performed by using PCR in patients A1 and A2 with primers XAF (GAAAGTCTTAGACCACTAAGCAGAT) and XAR (CATCAACCTCTTAAGCCATGG), and in patient B1 with primers XBF (GCACT ATCAATACCCAGGAATTTGCATC) and XBR (CAAGGAAACTACGAGAAACTCAAAGCC). The deletions removed segments between nucleotides 123 275 647 and 123 345 822 in patients A1 and A2, and between 123 243 812 and 123 460 144 in patient B1 (HG18). The deletion in family A was 70 176 bp long and was fully contained in the 216 333 bp long deletion of Patient B1. Both deletion junctions contained no added or rearranged nucleotides and no sequence homology. Both deletions removed the whole *SH2D1A* and the 3′ end of *ODZ1*. Patients A1 and A2 lacked the terminal exon 32, exon 31 and a part of exon 30 of *ODZ1*. Patient B1 lacked *ODZ1* exons 20 to 32.

SUPPLEMENTAL REFERENCES

22. Jerabkova B, Kopeckova L, Buckova H, Vesely K, Valickova J, Fajkusova L. Analysis of the COL7A1 gene in Czech patients with dystrophic epidermolysis bullosa reveals novel and recurrent mutations. *J Dermatol Sci.* 2010;59(2):136–140
23. Jerabkova B, Marek J, Buckova H, et al. Keratin mutations in patients with epidermolysis bullosa simplex: correlations between phenotype severity and disturbance of intermediate filament molecular structure. *Br J Dermatol.* 2010;162(5):1004–1013

Publikace 4

Array comparative genome hybridization in patients with developmental delay: two example cases.

Hancarova M, Drabova J, Zmitkova Z, Vlckova M, Hedvicakova P, Novotna D, Vlckova Z, Vejvalkova S, Marikova T, Sedlacek Z.

N Biotechnol. 2012 Feb 15;29(3):321-4. IF: 1,706

Prezentované dva vybrané modelové případy pacientů s komplexním postižením podmíněným chromozomální aberací demonstrují možnosti využití aCGH v diagnostice genetických příčin MR a PAS a určení klinické prognózy pro pacienty.

MR a PAS jsou často asociovány s CNV a metody aCGH jsou vhodným diagnostickým nástrojem jak pro detailní analýzu rozsahu a genového obsahu aberací identifikovaných klasickými cytogenetickými metodami (karyotypováním), tak i pro identifikaci karyotypováním neodhalitelných submikroskopických aberací. Pro ilustraci těchto dvou možností využití metod aCGH jsme vybrali dva případy. Pacient 1 měl cytogeneticky detekovatelnou deleci Xp21.2-p21.3 a metoda aCGH byla v jeho případě použita k určení přesného rozsahu delece a identifikaci všech deletovaných genů, což umožnilo určit jeho klinickou prognózu. Karyotyp pacienta 2 byl vyhodnocen jako normální, ale metodou aCGH u něho byla detekována malá delece v oblasti 1q21.1. Jednalo se o rekurentní deleci charakteristickou pro stejnojmenný mikrodeleční syndrom, která tudíž byla pravděpodobně kauzální, přestože byla zděděna od podstatně méně postižené matky.

Tato práce demonstruje užitečnost metody aCGH jako velmi důležité součásti rutinní diagnostiky genetických příčin syndromické i nesyndromické MR. Zároveň poukazuje na problémy, jež jsou spojeny s variabilní expresivitou a neúplnou penetrancí některých nově popsaných mikrodelečních syndromů a které si vynucují komplexní přístup a detailní posouzení fenotypu při hodnocení potenciální kauzality laboratorních nálezů.

New Biotechnology • Volume 29, Number 3 • February 2012 RESEARCH PAPER

Array comparative genome hybridization in patients with developmental delay: two example cases

Research Paper

Miroslava Hancarova, Jana Drabova, Zuzana Zmitkova, Marketa Vlckova, Petra Hedvicakova, Drahuse Novotna, Zdenka Vlckova, Sarka Vejvalkova, Tatana Marikova and Zdenek Sedlacek

Department of Biology and Medical Genetics, Charles University, 2nd Faculty of Medicine and University Hospital Motol, V Uvalu 84, Prague, Czech Republic

Developmental delay is often a predictor of mental retardation (MR) or autism, two relatively frequent developmental disorders severely affecting intellectual and social functioning. The causes of these conditions remain unknown in most patients. They have a strong genetic component, but the specific genetic defects can only be identified in a fraction of patients. Recent developments in genomics supported the establishment of the causal link between copy number variants in the genomes of some patients and their affection. One of the techniques suitable for this analysis is array comparative genome hybridization, which can be used both for detailed mapping of chromosome rearrangements identified by classical cytogenetics and for the identification of novel submicroscopic gains or losses of genetic material. We illustrate the power of this approach in two patients. Patient 1 had a cytogenetically visible deletion of chromosome X and the molecular analysis was used to specify the gene content of the deletion and the prognosis of the child. Patient 2 had a seemingly normal karyotype and the analysis revealed a small recurrent deletion of chromosome 1 likely to be responsible for his phenotype. However, the genetic dissection of MR and autism is complicated by high heterogeneity of the genetic aberrations among patients and by broad variability of phenotypic effects of individual genetic defects.

Introduction

Child development is a process of acquiring skills in several domains (cognitive, social, emotional, speech/language, and motor) during specific time periods. Developmental delay is defined as significant delay in two or more domains, and is thought to predict a future diagnosis of mental retardation (MR) and/or autism [1]. MR is a disability with limitations in intellectual functioning and adaptive behavior diagnosed before the age of 18 years [2]. It can be classified into four categories: mild (IQ 50–70), moderate (IQ 35–50), severe (IQ 20–35), and profound (IQ < 20). Autism is a disorder characterized by impairments of social interaction and communication, and unusual, often stereotyped behavior and interests [2]. Around 0.5–1% and 2–3% of population suffer from autism and MR, respectively [3–5]. Both conditions can be syndromic (if associated with other symptoms like facial dysmorphy) or non-syndromic (if no other consistent clinical or metabolic features are present).

Autism and MR affect the brain, the most complex human organ, and their causes are therefore also complex. Environmental factors include maternal infections or drug use in pregnancy and perinatal complications [4,5]. However, the genetic component is much more important in both conditions, as shown in twin and family studies [6,7]. Defined genetic syndromes, mutations in several specific genes, and cytogenetically visible or submicroscopic chromosome aberrations currently account for about 20% of the cases, but these causes are very heterogeneous with no one specific genetic defect being responsible for more than 1–2% of the cases [3,8,9].

Corresponding author: Sedlacek, Z. (zdenek.sedlacek@lfmotol.cuni.cz)

1871-6784/$ - see front matter © 2010 Elsevier B.V. All rights reserved. doi:10.1016/j.nbt.2010.10.006

If autism and MR were polygenic or multifactorial, the causal alleles could be detectable by whole-genome association studies. However, the results of these studies were rather disappointing because of low odds ratio and low level of replication [3]. Nevertheless, these analyses pointed to a more frequent occurrence of copy number variants (CNVs) in patients with autism and MR. CNVs are variably deleted or amplified regions of the genome, often involving genes. This variability is likely to contribute to phenotypic variability including the predisposition to diseases or the occurrence of cognitive and behavioral disorders or congenital defects [10,11].

The association of autism and MR with rare CNVs led to a shift in the understanding these two conditions from the multifactorial model (based on an interplay of common genetic variants with low effect) to the multiple rare variant model. Most affected individuals may carry unique genetic defects with strong effect, often arising *de novo* [8,9]. This model also explains the difficulties of whole-genome association studies to identify the predisposing loci, possibly just with regions containing clustering of different rare variants giving positive scores [12].

CNVs are detected using genomic approaches including array comparative genome hybridization (aCGH), which has underwent a remarkable shift in genome coverage and resolution since its first description [13,14]. In patients with defects identified using karyotyping (which has a resolution limit of several Mb), aCGH can specify the extent of the rearrangement and more precisely map the breakpoints. In cytogenetically normal patients, aCGH can reveal submicroscopic gains or losses of chromosome material. In both cases aCGH can point to genes potentially causing the phenotypes, and help to identify the mechanisms of the rearrangements.

In this report we illustrate the power of aCGH in two patients with developmental delay. In the first patient a large deletion was identified using karyotyping and aCGH allowed the determination of its exact size and gene content which had immediate consequences for the prognosis in this patient. In the second patient with a seemingly normal karyotype, the aCGH analysis revealed a small cytogenetically invisible deletion likely to be responsible for his phenotype.

Materials and methods

Patients

Patient 1 was admitted to the hospital at the age of six weeks with an electrolyte imbalance, renal failure and elevated levels of creatine kinase, adrenocorticotropic hormone and liver markers. He also suffered from hyponatremia, dehydration, proteinuria, high level of glycerol in the urine and weak spontaneous mobility. At a follow-up check at the age of two years he showed mild psychomotor retardation. His parents were healthy and the family history was not remarkable.

Patient 2 had a remarkable family history of MR and congenital defects (Fig. 1). At the age of 12 years, the boy (III.1) suffered from mild MR, attention deficit hyperactivity disorder (ADHD) with autistic features and microcephaly. After birth he presented with hypotonia, and his psychomotor development was disharmonic. His mother (II.2) also suffered from delayed development in childhood and showed behavioral problems and mental deficit. Both she and her maternal half-sister (II.4) suffered from congenital defects

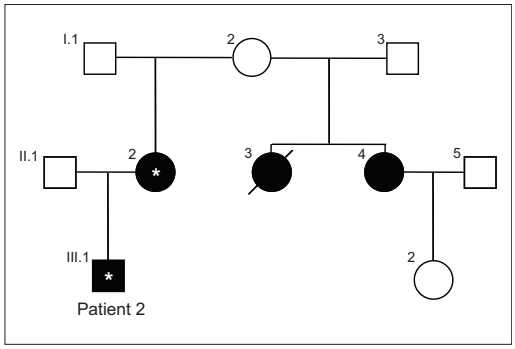


FIGURE 1

Pedigree of the family of Patient 2. Individuals with MR and/or congenital defects (see text for details) are represented by black symbols, and those in whom the deletion was identified are marked with asterisks. DNA from other family members was not available for study.

(syndactyly and cleft palate, respectively). Another maternal half-sister (II.3) with multiple malformations (caudal regression, gastroschisis, anal atresia and colon agenesis, persistent cloaca, and meningocele) died soon after birth. Their mother, the grandmother of the patient (I.2), had no remarkable phenotype. Only the DNA of Patient 2 and his mother were available for testing.

Laboratory methods

Chromosomal aberrations in peripheral blood lymphocytes of the patients and their parents were analyzed using standard cytogenetic techniques. Further delineation of the aberrations was performed using fluorescence in situ hybridization (FISH) with a chromosome X painting probe (WCP X, Cytocell) in Patient 1, and a locus-specific probe (RP11-533N14, BlueGnome) in Patient 2, according to the recommendations of the manufacturer.

Genomic DNA was isolated from blood lymphocytes of the patients and their parents using the Gentra Puregene Blood Kit (Qiagen). Multiplex ligation-dependent probe amplification (MLPA) was performed using the SALSA MLPA KIT P034/P035 DMD/Becker (MRC-Holland) according to the instructions of the manufacturer. Whole-genome aCGH analysis of Patient 2 was performed using a bacterial artificial chromosome (BAC) array with median backbone probe spacing of 565 kb (CytoChip V3, BlueGnome) according to the instructions of the manufacturer. Custom high resolution aCGH analysis of Patient 1 and Patient 2 was performed by Nimblegen on catalogue oligonucleotide CGH arrays specific for chromosome X (HG18_CHRX_FT, median probe spacing of 340 bp) and chromosome 1 (HG18_CHR1_FT, median probe spacing of 541 bp), respectively. The results were analyzed using the SignalMap software (Nimblegen). Long-range PCR used the Expand Long Template PCR System (Roche). The gene content of the deletions was analyzed using the UCSC and Ensembl genome browsers (http://genome.ucsc.edu/cgi-bin/hgGateway, http://www.ensembl.org/index.html) and the hg18 assembly.

Results

The phenotype of Patient 1 and his biochemical findings were suggestive of glycerol kinase deficiency (GKD) and Duchenne

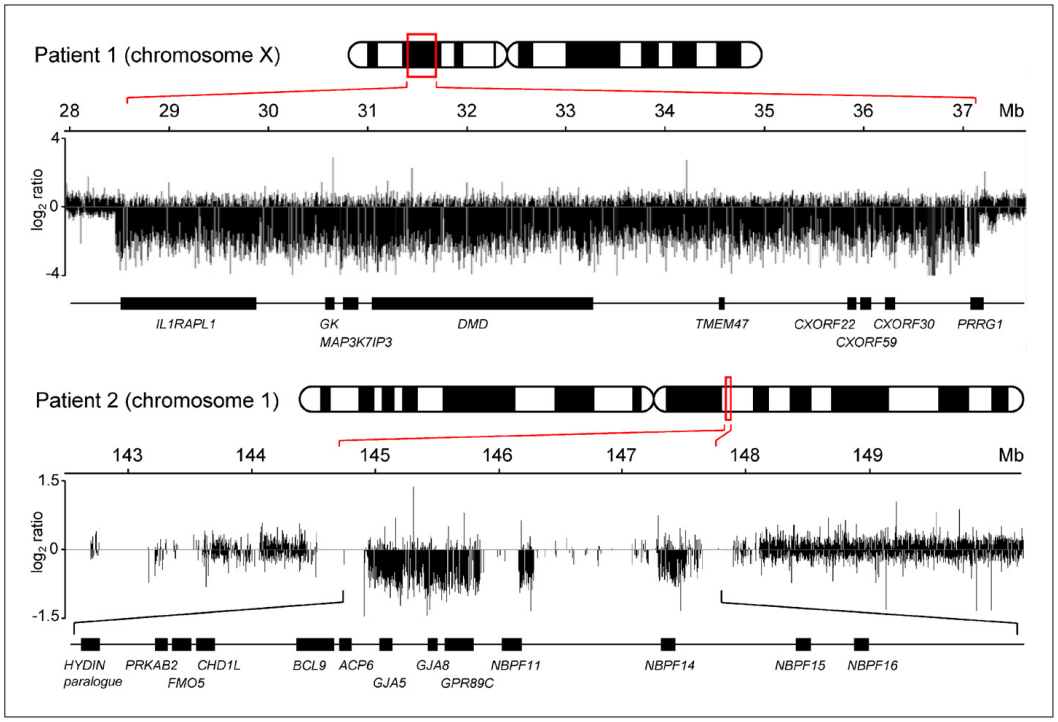


FIGURE 2

The results of aCGH analysis of the deletions in Patient 1 and Patient 2 using chromosome-specific high resolution arrays. The regions of decreased signal ratios correspond to the deletions. This analysis allowed precise mapping of the breakpoints (which was not possible using karyotyping and low resolution arrays used initially to identify the deletions in Patient 1 and Patient 2, respectively), and the assessment of gene content of the deletions. Regions devoid of signals correspond to segmental duplications and low-copy repeats not represented on the chromosome 1 specific array.

muscular dystrophy (DMD). This prompted us to perform cytogenetic analysis aimed primarily at deletions of the X chromosome. Indeed, his karyotype was 46,XY,del(X)(p21.2p21.3), and the deletion was likely to encompass the above-mentioned two loci. The same deletion was also identified in his asymptomatic mother. FISH analysis showed that the deleted segment was not translocated to any other chromosome. MLPA analysis of the *DMD* gene of Patient 1 revealed a deletion of all exons, and classical PCR analysis indicated the deletion of several additional genes (*GK*, *NR0B1* and *IL1RAPL1*). aCGH analysis confirmed an 8.7 Mb long deletion between Mb 28.5 and 37.2 of the X chromosome containing a total of nine protein coding genes (Fig. 2). The distal breakpoint of the deletion was located between the *DCAF8L1* and *IL1RAPL1* genes, and the proximal breakpoint was located in the *PRRG1* gene. Both breakpoints mapped to regions of unique DNA or dispersed repeats and there was no obvious homology between the breakpoint regions. Repeated attempts to clone the deletion junction using long-range PCR were not successful.

The karyotype of Patient 2 was normal, but the low resolution aCGH analysis showed a deletion of chromosome region 1q21.1. While the signals from BACs RP11-373C9 and RP11-763B22 were normal, BAC RP11-94I2 yielded a borderline signal, and low signals from BACs RP11-47D6, RP11-563P13, RP11-441L11, RP11-533N14, RP11-314N2 and RP11-301M17 clearly indicated a deletion. The remaining chromosomes did not show any copy number aberrations. The deletion was fine mapped using high resolution aCGH to Mb 144.8–147.8 of chromosome 1 (Fig. 2). The size of the deletion was about 3 Mb and it contained 13 protein coding genes. Both breakpoints were located in extended complex regions of segmental duplication and low-copy repeats. The deletion was confirmed in Patient 2 using FISH with BAC probe RP11-533N14, and the same technique was used to identify the same deletion in his mother.

Discussion

The results obtained in our patients clearly illustrate the power of aCGH for the detailed analysis of cytogenetically visible defects and for the detection of rare submicroscopic CNVs associated with developmental delay, autism and MR.

The molecular analyses in Patient 1 confirmed the suspected diagnosis of a rare X-linked recessive disorder, complex GKD. In contrast to isolated GKD, which is caused by point mutations in the glycerol kinase gene (*GK*), complex GKD is a contiguous gene

syndrome caused by a deletion of the *GK* locus together with other Xp21 sequences including the adrenal hypoplasia congenita locus (*NR0B1* gene) and/or *DMD* [15]. About 100 patients with complex GKD have been reported, and their phenotypes usually reflected the variable size of their deletions. The analysis of gene content of the deletions is therefore an important prognostic factor. Patient 1 carried a rather large deletion including the *IL1RAPL1* gene, which has been associated with MR and autism [16,17], and his prognosis with respect to these two conditions was therefore rather unfavorable. In accord with this prognosis, Patient 1 showed at the age of 2 years mild psychomotor retardation. The breakpoints of the deletion were unique and contained no segmental duplications or low-copy repeats.

A genetic basis of the phenotype and the remarkable family history of Patient 2 was suspected but completely unknown. The molecular analysis using whole-genome aCGH identified a submicroscopic deletion of chromosome 1 in the distal 1q21.1 region, which was subsequently fine mapped using high resolution aCGH. This recently described microdeletion is recurrent, and its formation is mediated by recombination between segmental duplications flanking the deleted region [18,19]. In large collections of affected individuals this deletion has been associated with a complex and variable phenotype which often included MR and autism or ADHD, microcephaly and a wide range of other congenital anomalies. This deletion removes several protein coding genes including a paralog of the *HYDIN* gene which is suspected to cause microcephaly and other neuropsychiatric features in the deletion carriers [18]. The deletion of 1q21.1 is often inherited from a less affected or even seemingly normal parent [18,19]. The mildly affected mother of Patient 2 was shown to carry the deletion, and the family history suggested that also her two half-sisters could be carriers. Their unaffected mother, the grandmother of Patient 2, could also carry the deletion, or at least a germline mosaic of this defect.

Patient 2 illustrates very clearly several key problems of the recent concept of genetics of autism and MR. First, there is a very substantial heterogeneity at the phenotype level, and the clinical picture of carriers of a particular genetic defect can be very variable, even within one family. The phenotypic spectrum in 1q21.1 deletion carriers ranges from completely normal individuals through various mild to fully expressed psychiatric phenotypes to very severe congenital defects [18,19], and this is similar for other CNVs. The reason for this variability is generally unknown but it is speculated that the genetic background (the genotype of the individual at other genome loci), epigenetic status and interaction with environmental triggers are playing the key role [20]. It has been proposed recently that full phenotypic expression of one CNV can be dependent on the co-occurrence of other CNVs in the same individual [21]. Second, the spectrum of genetic defects underlying autism and MR in individual patients can be extremely broad, and individual genetic defects can be extremely rare. Although the microdeletion of 1q21.1 has clearly been associated with autism and MR, its frequency among affected individuals reaches only 0.2–0.5% [18,19], and thus the diagnostic yield of this rearrangement is very low. The same applies to most other genetic defects currently known to be responsible for these conditions [3,8,9]. This precludes the possibility to design a simple targeted diagnostic genetic test. It is probable that the continuing improvements and increasing affordability of high resolution genome analysis methods including whole-genome sequencing will soon lead to the identification of the 'missing heritability' in an increasing fraction of sufferers from complex diseases, and will at the same time offer efficient tools for their routine diagnostics.

Acknowledgements

Supported by grants MZOFNM2005 from the Ministry of Health of the Czech Republic, and INCORE (FP6 043318) and CHERISH (FP7 223692) from the European Commission.

References

1 Accardo, P. and Capute, A. (1995) Intellectual disability. In *Principles and Practice of Pediatrics* (Oski, F. and Diangelis, C.D., eds), pp. 673–679, Lippincott, Philadelphia
2 American Psychiatric Association, (1994) *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* (4th ed.), American Psychiatric Association, Washington DC
3 Abrahams, B.S. and Geschwind, D.H. (2008) Advances in autism genetics: on the threshold of a new neurobiology. *Nat. Rev. Genet.* 9, 341–355
4 Levy, S.E. *et al.* (2009) Autism. *Lancet* 374, 1627–1638
5 Roeleveld, N. *et al.* (1997) The prevalence of mental retardation: a critical review of recent literature. *Dev. Med. Child Neurol.* 39, 125–132
6 Jorde, L.B. *et al.* (1991) Complex segregation analysis of autism. *Am. J. Hum. Genet.* 49, 932–938
7 Bailey, A. *et al.* (1995) Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol. Med.* 25, 63–77
8 Buxbaum, J.D. (2009) Multiple rare variants in the etiology of autism spectrum disorders. *Dialog. Clin. Neurosci.* 11, 35–43
9 Vissers, L.E. *et al.* (2010) Genomic microarrays in mental retardation: from CNV to gene, from research to diagnosis. *J. Med. Genet.* 47, 285–297
10 Iafrate, A.J. *et al.* (2004) Detection of large-scale variation in the human genome. *Nat. Genet.* 36, 949–951
11 Conrad, D.F. *et al.* (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704–712
12 Dickson, S.P. *et al.* (2010) Rare variants create synthetic genome-wide associations. *PLoS Biol.* 8, e1000294
13 Chen, X. *et al.* (2009) New tools for functional genomic analysis. *Drug Discov. Today* 14, 754–760
14 Solinas-Toldo, S. *et al.* (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromos. Cancer* 20, 399–407
15 Sjarif, D.R. *et al.* (2000) Isolated and contiguous glycerol kinase gene disorders: a review. *J. Inherit. Metab. Dis.* 23, 529–547
16 Piton, A. *et al.* (2008) Mutations in the calcium-related gene IL1RAPL1 are associated with autism. *Hum. Mol. Genet.* 17, 3965–3974
17 Carrie, A. *et al.* (1999) A new member of the IL-1 receptor family highly expressed in hippocampus and involved in X-linked mental retardation. *Nat. Genet.* 23, 25–31
18 Brunetti-Pierri, N. *et al.* (2008) Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat. Genet.* 40, 1466–1471
19 Mefford, H.C. *et al.* (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.* 359, 1685–1699
20 Manolio, T.A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature* 461, 747–753
21 Girirajan, S. *et al.* (2010) A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat. Genet.* 42, 203–209

Publikce 5

Identification of a patient with intellectual disability and *de novo* 3.7 Mb deletion supports the existence of a novel microdeletion syndrome in 2p14-p15.

Hancarova M, Vejvalkova S, Trkova M, Drabova J, Dleskova A, Vlckova M, Sedlacek Z.

Gene. 2013 Mar 1;516(1):158-61. IF: 2,196

Mikrodelece 2p14-p15 byla nedávno popsána u dvou pacientů s opožděním psychomotorického vývoje a MR bez přidružených VVV či nápadné faciální stigmatizace. V této kazuistice prezentujeme čtyřletého chlapce s *de novo* mikrodelecí v oblasti 2p14-p15, která se překrývá se dvěma dříve publikovanými mikrodelecemi. Stejně jako v předchozích dvou případech byla i u našeho pacienta zjištěna MR, absence řeči, mikrocefalie, podlouhlý obličej, bulbózní špička nosu a tenký horní ret. Celkově však náš pacient vykazoval těžší postižení než doposud publikované případy.

Mikrodelece 2p14-p15 byla, kromě dvou publikovaných případů uvedených výše, také několikrát identifikována v rámci vyšetření rozsáhlých kohort pacientů s MR, avšak u těchto případů není k dispozici detailní klinický popis. Naše identifikace dalšího pacienta s mikrodelecí 2p14-p15 a jeho fenotypová podobnost s dosud známými případy tak podporuje domněnku, že se jedná o nový vzácný mikrodeleční syndrom s kritickým regionem v blízkosti regionu dobře popsaného a častějšího mikrodelečního syndromu 2p15-p16. V kritické oblasti se nachází několik genů, které mohou být zodpovědné za MR u pacientů, u kterých jsou deletovány. Zdá se, že pacienti s mikrodelecemi zasahujícími kritické regiony obou syndromů mají těžší a komplexnější postižení.

Tato práce představuje důležitý příspěvek k poznatkům o vzácných mikrodelečních syndromech, k hodnocení kauzality jednotlivých mikrodelecí a k dalšímu upřesnění očekávatelného fenotypu a prognózy pacientů s mikrodelecí 2p14-p15. Z praktického pohledu je náš případ zajímavý také tím, že jsme byli jako autoři práce osloveni jednou zahraniční rodinou s dítětem s podobným nálezem a postižením s žádostí o zprostředkování kontaktu na rodinu našeho pacienta. Výměna zkušeností a vzájemná podpora je u rodin se vzácnými onemocněními nesmírně důležitá a potěšilo nás, že kontakt úspěšně proběhl, a že naše publikace měla i tento nečekaný pozitivní efekt.

Gene 516 (2013) 158–161

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Short Communication

Identification of a patient with intellectual disability and de novo 3.7 Mb deletion supports the existence of a novel microdeletion syndrome in 2p14–p15

Miroslava Hancarova [a], Sarka Vejvalkova [a], Marie Trkova [b], Jana Drabova [a], Alzbeta Dleskova [a], Marketa Vlckova [a], Zdenek Sedlacek [a,*]

[a] Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic
[b] Gennet, Prague, Czech Republic

ARTICLE INFO

Article history:
Accepted 2 December 2012
Available online 22 December 2012

Keywords:
2p14–p15 deletion
2p15–p16.1 microdeletion syndrome
Copy number variation
Developmental delay
Intellectual disability
SNP array

ABSTRACT

Microdeletions spanning 2p14–p15 have recently been described in two patients with developmental and speech delay and intellectual disability but no congenital malformations or severe facial dysmorphism. We report a 4-year-old boy with a de novo 3.7 Mb long deletion encompassing the region deleted in the previous cases. The patient had clinical features partly consistent with the published cases including intellectual disability, absent speech, microcephaly, long face, bulbous nasal tip and thin upper lip, but his overall clinical picture was more severe compared to the published patients. The identification of this additional patient and a detailed analysis of deletions identified in various patient cohorts and in normal individuals support the existence of a new rare microdeletion syndrome in 2p14–p15. Its critical region is in the vicinity of but clearly separate from the minimal region deleted in the well established 2p15–p16.1 microdeletion syndrome. A thorough comparison of the deletions and phenotypes indicates that multiple genes located in this region may be involved in intellectual functioning, and that some patients may show composite and more complex phenotypes due to deletions spanning both critical regions.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

During the last years many novel microdeletion and microduplication syndromes with intellectual disability (ID) have been identified using microarray methods (Cooper et al., 2011; Ropers, 2010). However, individually these syndromes account usually only for small fractions of percent of cases with ID, and many of the syndromes have been described only in a handful of patients. Reporting of new cases with known syndromes is important for the definition of the phenotype spectra of the syndrome and for the delineation of the minimal deleted region responsible for the symptoms. A report of a new patient can also support the causality of a defect which has only been described in a very limited number of individuals. In this paper we describe a patient with ID and a deletion of chromosome 2p14–p15. Our report supports the existence of a novel microdeletion syndrome with critical region adjacent to but clearly separate from that of the well established 2p15–p16.1 microdeletion syndrome (Piccione et al., 2012).

2. Material and methods

2.1. Patient

The patient was referred to a geneticist at 19 months of age because of microcephaly and developmental and speech delay. The boy was born as a second child to healthy non-consanguineous Czech parents aged 29 years (mother) and 30 years (father). The delivery was severely premature in the 25th week of gestation with birth weight of 680 g. The normal elder brother of the patient was also born prematurely in the 29th week of gestation. The patient required ventilatory support during the first month of life, a surgical ligation of persistent ductus arteriosus and two repairs of inguinal hernias. The first measurement of head circumference (HC) is available from 3 months after birth (32 cm, 25th centile, after correction for the gestational age). The boy had serious feeding problems. At the age of 6 months he was noted to have a very slow increase of HC, and severe microcephaly was clearly evident after 1 year of age. At 19 months his height, weight and HC were at −1.5 SD, −3.1 SD and −5 SD, respectively (Fig. 1). At the last examination at 4 years and 3 months the boy showed marked muscle and fat wasting, microcephaly and brachycephaly, and his height, weight

Abbreviations: Mb, megabase; SNP, single nucleotide polymorphism; ID, intellectual disability; HC, head circumference; SD, standard deviation; MRI, magnetic resonance imaging; *FMR1*, fragile X mental retardation 1; PCR, polymerase chain reaction; FISH, fluorescence in situ hybridization; *COMMD1*, COMM domain-containing protein 1; *SPRED2*, sprouty-related, EVH1 domain-containing protein 2; *MEIS1*, myeloid ecotropic viral integration site 1; *AFTPH*, aftiphilin; *SLC1A4*, solute carrier family 1, member 4; *RAB1A*, member of RAS oncogene family, isoform 1; DD, developmental delay; DECIPHER, Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources; SRO, shortest region of overlap; DGV, Database of Genomic Variants.

* Corresponding author at: Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Plzenska 130/221, 15000 Prague 5, Czech Republic. Tel.: +420 257296153; fax: +420 224433520.
E-mail address: zdenek.sedlacek@lfmotol.cuni.cz (Z. Sedlacek).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.12.027

Fig. 1. Front and lateral views of the patient at the age of 19 months (A), 2 years and 3 months (B), and 4 years and 3 months (C, D).

and HC were at −1.9 SD, −2 SD and −5.4 SD, respectively. His facial features included long face with slight bitemporal narrowing, a visual impression of hypotelorism (not measured), bulbous nasal tip, smooth philtrum, thin upper lip, prominent chin and large, protruding, low set and dysplastic ears (Fig. 1). The eruption of teeth was delayed (at 18 months of age). The patient had undescended testes and several small hemangiomas on legs and body. MRI of the brain showed at the age of 18 months a hemosiderin deposition in the left lateral ventricle and rims of hyperintense signal around the temporal and occipital horns of both lateral ventricles. These changes could be a consequence of intracranial hemorrhage in the newborn period. The psychomotor development of the boy was delayed. He started to walk shortly before the age of 3 years and at 4 years his developmental quotient corresponded to a less than 2-year-old child. The most obvious was his severe speech impairment. At the age of 4 years his speech corresponded to a 1-year-old child. He could follow simple instructions but his speech was nearly absent. His ID was classified as moderate to severe. Phoniatric examination showed no hearing impairment. Ophthalmological exam revealed bilateral myopia (−2D). The patient was severely hyperactive. He was under neurological surveillance because of cerebral palsy possibly caused by his severe perinatal history. He was diagnosed with common type of spastic cerebral palsy with only discrete signs of diparesis and cerebellar ataxia but without any seizures in his personal history. The patient was referred to rehabilitation therapy and a special education center, where he made slow progress in his cognitive and motor development.

2.2. Laboratory methods

Peripheral blood samples were obtained from the patient and his parents based on their informed consent. Conventional cytogenetic analysis was performed using standard G-banding, and *FMR1* gene testing

Table 1
The comparison of phenotypes of the three patients.

Wohlleber et al. Patient 1	Wohlleber et al. Patient 2	Present case
Male 8 y 3 m	Female 12 y	Male 4 y 3 m
Head and face		
Microcephaly	x	Microcephaly, brachycephaly
Long face	x	Long face
Bulbous nasal tip	Broad nasal tip	Bulbous nasal tip
Thin upper lip	Thin upper lip	Thin upper lip
Slightly hooded eyelids	Hooded eyelids	x
x	x	Smooth philtrum, prominent chin
x	x	Large dysplastic low set ears
Other features		
Mild funnel chest	x	x
Smooth fetal finger pads	x	x
Slight syndactyly of the 2nd and 3rd toes of the right foot	x	x
x	Joint hypermobility	x
x	x	Muscle and fat wasting
x	n. a.	Undescended testes
Milestones and cognitive and sensory development		
Feeding problems	x	Feeding problems
x	x	Delayed teeth eruption
Walked at 15 m	Walked at 18 m	Walked just before the age of 3 y
Speech delay (2–3-word sentences at 5 y)	Speech delay (single words at 3 y, simple sentences at 4 y)	Speech delay (single words at 4 y)
Mild ID	Mild ID	Moderate to severe ID
Brain MRI normal	Brain MRI normal	Small changes in the lateral ventricles
Generalized seizures	x	x
Sensorineural hearing loss	x	x
Mild hyperopia	x	Bilateral myopia
x	Clumsy	x
x	x	Severe hyperactivity

y, years; m, months; x, character absent; n. a., not applicable.

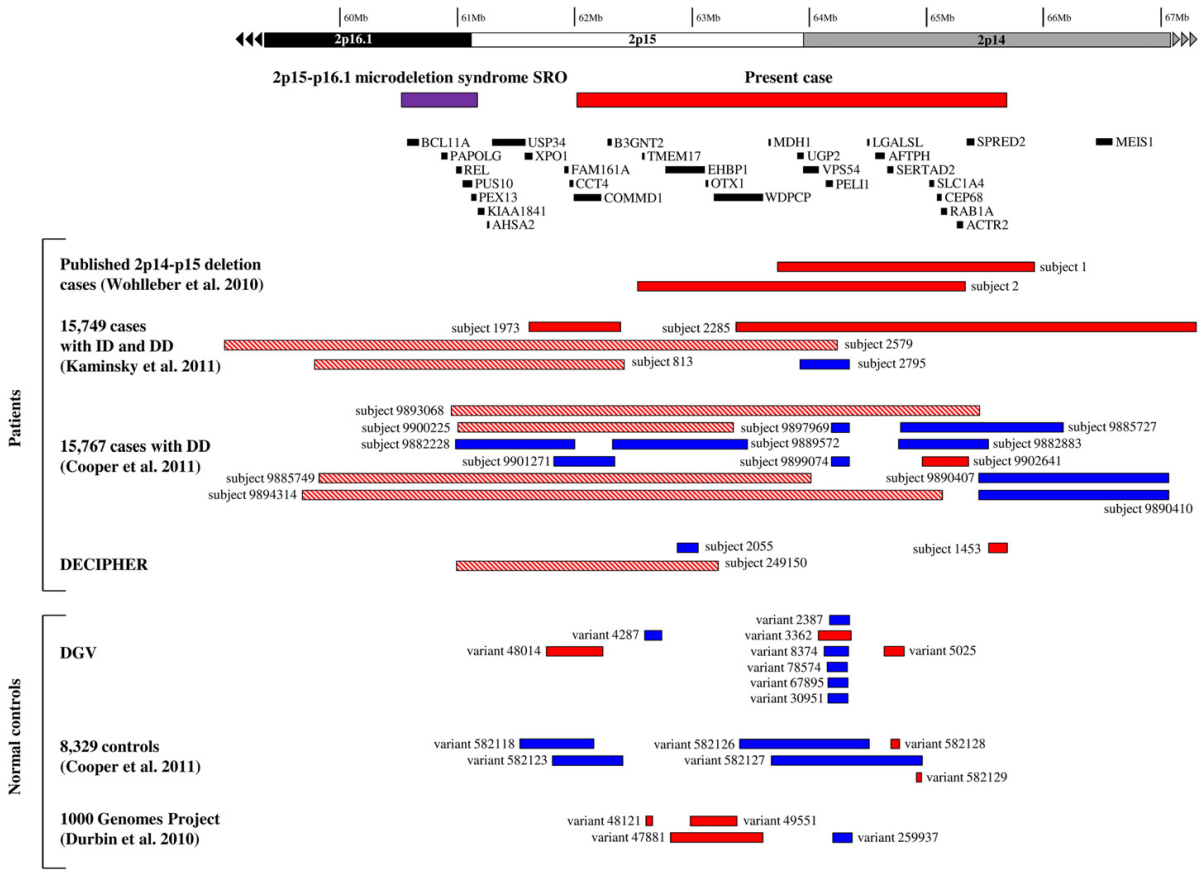


Fig. 2. Schematics of the region affected by the deletion in our patient including the protein-coding gene content and the shortest region of overlap (SRO) of the 2p15–p16.1 microdeletion syndrome. Deletions are depicted in red, duplications in blue. Deletions in the present case and in the previously published patients with the proposed 2p14–p15 microdeletion syndrome are shown together with aberrations identified in large published ID/DD cohorts and aberrations listed in DECIPHER, DGV and the 1000 Genomes Project (Durbin et al., 2010). Only aberrations overlapping with the deletion in our patient are shown. Deletions overlapping the 2p15–p16.1 microdeletion syndrome SRO are hatched.

used the Fragile X PCR Kit (Abbott, Abbott Park, IL, USA). Genomic DNA of the patient was analyzed on a HumanCytoSNP-12 BeadChip (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Data were analyzed using Genome Studio (Illumina) and QuantiSNP (Colella et al., 2007). The analysis used genome build hg18. The deletion identified was independently confirmed using FISH probe RP11-44A12 (BlueGnome, Cambridge, UK).

3. Results

The cytogenetic analysis revealed a normal male karyotype, and the *FMR1* testing did not show any CGG trinucleotide expansion in the patient. The SNP array analysis revealed a 3.7 Mb long microdeletion of chromosome 2p14–p15 with the distal breakpoint between rs6714793 and rs6722228, and the proximal breakpoint between rs7579084 and rs840776 (minimum span: chr2: 62,013,233–65,731,466; maximum span: chr2: 61,966,678–65,743,837). The distal deletion breakpoint was located in the *COMMD1* gene and the proximal breakpoint between the *SPRED2* and *MEIS1* genes. The deletion involved 18 protein-coding RefSeq genes (Fig. 2) and 6 non-coding RNA genes. The deletion was the sole aberration observed in the patient. It was confirmed using FISH in the patient but in none of his parents, indicating its de novo nature.

4. Discussion

The deletion in our patient overlapped with the 2p14–p15 deletions described recently in two patients proposed to have a novel microdeletion syndrome (Wohlleber et al., 2011). The published patients showed developmental delay, mild ID and speech delay but no congenital malformations or severe facial dysmorphism (Table 1). The first case, an 8-year-old boy, had mild microcephaly. The other patient, a 12-year-old girl, had normal HC, and also her other phenotypic features and facial dysmorphy were much milder. The developmental and speech delay and the overall clinical picture of our patient were more severe compared to both of these published patients. This could be partly attributed to his prematurity and more severe perinatal history, as indicated also by the MRI findings. Growth and developmental retardation, microcephaly and cerebral palsy are also among known consequences of preterm delivery (Kytnarova et al., 2011). Concurrently, preterm births are frequent in carriers of genetic defects (Honein et al., 2009). Nevertheless, some aspects of the phenotype of our patient including ID, speech delay, long face, bulbous nasal

tip and thin upper lip were reminiscent of the features seen in the reported patients. The similarity was higher with the more severely affected boy (Table 1), with whom our patient also shared feeding difficulties, microcephaly and very severe speech impairment, although the microcephaly in our patient was even more pronounced.

The phenotypes of the three patients could also differ due to the differences in the size and gene content of the deletions. The three deletions shared no common breakpoints, and the 3.7 Mb long deletion in our patient was the longest with the highest number of deleted genes. The shortest region of overlap of all three deletions contained 10 protein-coding genes (Fig. 2). Some of them could be candidates for ID based on their expression in the brain or their role in neuronal development or functioning. The *AFTPH* gene product aftiphilin is expressed exclusively in the brain where it is enriched on clathrin-coated vesicles and is likely involved in trafficking in neurons. Aftiphilin is also found at synapses (Burman et al., 2005). The *SLC1A4* gene encodes an amino acid transporter and is expressed mostly in the brain. Its product is depleted in neurons in the cingulate cortex and in astrocytes of the white matter in schizophrenia, and in neurons in bipolar disorder and major depression, reflecting the changes in glutamatergic neurotransmission in these disorders (Weis et al., 2007). The *RAB1A* gene encodes a member of the Ras superfamily of GTPases. Its product is a key regulator of the intracellular vesicle transport from the endoplasmic reticulum to the Golgi apparatus that it is crucial for proper neuron growth and neuronal differentiation (Maier et al., 2009).

Two recently published large studies of copy number variation in ID and developmental delay (DD) (Cooper et al., 2011; Kaminsky et al., 2011) and the DECIPHER database (Firth et al., 2009) list several overlapping deletions involving 2p14–p15. These deletions have variable breakpoints, length and gene contents. This together with the phenotypic heterogeneity usually seen in the novel microdeletion syndromes complicates the genotype–phenotype correlation. The symptoms of the 2p14–p15 syndrome are not specific and some are similar to the phenotypic features of the well established 2p15–p16.1 microdeletion syndrome, although some symptoms of the latter disorder such as telecanthus, broad nasal root, long philtrum, camptodactyly, structural brain anomalies and autism (Piccione et al., 2012) are missing in patients with 2p14–p15 deletions. Some specific recognizable phenotypic signs such as long face, bulbous nasal tip and thin upper lip seem to be characteristic of the 2p14–p15 syndrome. Several deletions overlap both the 2p14–p15 and the 2p15–p16.1 regions (Fig. 2), but the current narrowing down of the shortest region of overlap of the 2p15–p16.1 microdeletion syndrome (Piccione et al., 2012) and the clear separation of the cluster of deletions including those in our patient and the previously published boy and girl (Fig. 2) support the existence of a novel microdeletion syndrome in the 2p14–p15 region. The definition of this condition will require the analysis of additional cases, because multiple genes from this broader region are likely to be involved in ID, and some patients may show composite and more complex phenotypes.

Acknowledgments

We thank the family of the patient for cooperation and the reviewers for helpful comments. Supported by CHERISH 223692, CZ.2.16/3.1.00/24022 and a Czech grant for conceptual development of research organization 00064203.

References

Burman, J.L., Wasiak, S., Ritter, B., de Heuvel, E., McPherson, P.S., 2005. Aftiphilin is a component of the clathrin machinery in neurons. FEBS Lett. 579, 2177–2184.
Colella, S., et al., 2007. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 35, 2013–2025.
Cooper, G.M., et al., 2011. A copy number variation morbidity map of developmental delay. Nat. Genet. 43, 838–846.
Durbin, R.M., et al., 2010. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073.
Firth, H.V., et al., 2009. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources. Am. J. Hum. Genet. 84, 524–533.
Honein, M.A., et al., 2009. The association between major birth defects and preterm birth. Matern. Child Health J. 13, 164–175.
Kaminsky, E.B., et al., 2011. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet. Med. 13, 777–784.
Kytnarova, J., et al., 2011. Post-natal growth of 157 children born as extremely premature neonates. J. Paediatr. Child Health 47, 111–116.
Maier, S., et al., 2009. GTRAP3-18 serves as a negative regulator of Rab1 in protein transport and neuronal differentiation. J. Cell. Mol. Med. 13, 114–124.
Piccione, M., et al., 2012. Interstitial deletion of chromosome 2p15–16.1: report of two patients and critical review of current genotype–phenotype correlation. Eur. J. Med. Genet. 55, 238–244.
Ropers, H.H., 2010. Genetics of early onset cognitive impairment. Annu. Rev. Genomics Hum. Genet. 11, 161–187.
Weis, S., Llenos, I.C., Dulay, J.R., Verma, N., Sabunciyan, S., Yolken, R.H., 2007. Changes in region- and cell type-specific expression patterns of neutral amino acid transporter 1 (ASCT-1) in the anterior cingulate cortex and hippocampus in schizophrenia, bipolar disorder and major depression. J. Neural Transm. 114, 261–271.
Wohlleber, E., et al., 2011. Clinical and molecular characterization of two patients with overlapping de novo microdeletions in 2p14–p15 and mild mental retardation. Eur. J. Med. Genet. 54, 67–72.

Publikace 6

Monozygotic twins with 17q21.31 microdeletion syndrome.

Vlckova M, Hancarova M, Drabova J, Slamova Z, Koudova M, Alanova R, Mannik K, Kurg A, Sedlacek Z.

Twin Res Hum Genet. 2014 Oct;17(5):405-10. IF: 1,635

Práce popisuje případ dvou sester, monozygotních dvojčat s mikrodelečním syndromem 17q21.31 (Koolen-de Vriesův syndrom, MIM 610443). Je zaměřena na detailní popis fenotypu pacientek a na porovnání genotypu a fenotypu obou dvojčat navzájem se zaměřením na případné rozdíly.

Mikrodeleční syndrom 17q21.13 je jednou z nedávno objevených příčin syndromické MR. Je způsoben rekurentní delecí v oblasti 900 kb dlouhé inverze vyskytující se u přibližně 20% evropské populace. Invertovaná varianta nese haplotyp H2 a takovou orientaci nízkofrekventních repetic, která predisponuje ke vzniku delecí mechanismem NAHR. V době uveřejnění publikace byla prevalence odhadována na 1/16000, což by tento syndrom řadilo mezi časté mikrodeleční syndromy. V současné době je prevalence odhadována na 1/55000 (Koolen et al., 2016). K charakteristickým fenotypovým projevům patří MR, hypotonie, přátelská povaha a typický faciální fenotyp zahrnující podlouhlý obličej, nápadně hrubé rysy a hruškovitý nos s bulbózní špičkou. Fenotyp pacientů se zvýrazňuje s věkem.

Obě sledované sestry vykazovaly typický fenotyp syndromu, nicméně u dvojčete B jsme zaznamenali o něco těžší postižení. Bylo zjištěno, že mikrodelece je *de novo* a že matka dvojčat je homozygotkou pro predisponující haplotyp H2. Genomová DNA obou dvojčat byla dále analyzována metodami vysokorozlišovací SNPa a aCGH k ozřejmění případných CNV, které by mohly být zodpovědné za fenotypové odlišnosti. Přestože genotyp dvojčat nemusel být zcela identický (pozorovali jsme potenciální rozdíly v počtu kopií na několika vysoce repetitivních lokusech), žádné další potenciálně patogenní CNV, jež by mohly být asociovány s odlišným fenotypem, nebyly detekovány. Těžší postižení u dvojčete B je tedy zřejmě důsledkem komplikovanější perinatální anamnézy a variabilní expresivity onemocnění. Podrobným klinickým popisem a sledováním pacientek v relativně dlouhém čase (od narození do 25 let) tato práce přispívá k dosavadním znalostem o fenotypu, a zejména prognóze mikrodelečního syndromu 17q21.31. Zároveň se zabývá potenciálními genotypovými odlišnostmi u monozygotních dvojčat, jejichž studium má v genetice stále velmi významnou roli.

Recentní publikace se (kromě rozšíření fenotypového spektra) zaměřují na porovnání fenotypů pacientů s mikrodelecí v oblasti 17q21.31 a pacientů s bodovými mutacemi v genu *KANSL* (který je vždy zasažen delecí a je označován za kritický) ve snaze rozhodnout, zda se jedná o tzv „contiguous gene syndrome“ nebo zda je přítomnost mutace v genu *KANSL* postačující pro rozvoj fenotypu. Dle výsledků se zdá, že mezi fenotypy pacientů s mikrodelecí a bodovou mutací v genu *KANSL* není žádný signifikantní rozdíl (Koolen et al., 2016; Zollino et al., 2015).

Twin Research and Human Genetics
page 1 of 6 ▌© The Authors 2014 ▌doi:10.1017/thg.2014.29

Monozygotic Twins with 17q21.31 Microdeletion Syndrome

Marketa Vlckova,[1,#] Miroslava Hancarova,[1,#] Jana Drabova,[1] Zuzana Slamova,[1] Monika Koudova,[1,*] Renata Alanova,[1,] Katrin Mannik,[2] Ants Kurg,[2] and Zdenek Sedlacek[1]**
[1]Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic
[2]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

Chromosome 17q21.31 microdeletion syndrome is a genomic disorder caused by a recurrent 600 kb long deletion. The deletion affects the region of a common inversion present in about 20% of Europeans. The inversion is associated with the H2 haplotype carrying additional low-copy repeats susceptible to non-allelic homologous recombination, and this haplotype is prone to deletion. No instances of 17q21.31 deletions inherited from an affected parent have been reported, and the deletions always affected a parental chromosome with the H2 haplotype. The syndrome is characterized clinically by intellectual disability, hypotonia, friendly behavior and specific facial dysmorphism with long face, large tubular or pear-shaped nose and bulbous nasal tip. We present monozygotic twin sisters showing the typical clinical picture of the syndrome. The phenotype of the sisters was very similar, with a slightly more severe presentation in Twin B. The 17q21.31 microdeletion was confirmed in both patients but in neither of their parents. Potential copy number differences between the genomes of the twins were subsequently searched using high-resolution single nucleotide polymorphism (SNP) and comparative genome hybridisation (CGH) arrays. However, these analyses identified no additional aberrations or genomic differences that could potentially be responsible for the subtle phenotypic differences. These could possibly be related to the more severe perinatal history of Twin B, or to the variable expressivity of the disorder. In accord with the expectations, one of the parents (the mother) was shown to carry the H2 haplotype, and the maternal allele of chromosome 17q21.31 was missing in the twins.

■ **Keywords:** 17q21.31 microdeletion syndrome, monozygotic twins, CNV, epigenetics

The chromosome 17q21.31 microdeletion syndrome (Koolen-de Vries syndrome, MIM 610443) is a genomic disorder characterized by intellectual disability (ID), friendly behavior, hypotonia and distinct facial features with thin long face, large pear-shaped nose and prominent chin (Koolen et al., 2006; Sharp et al., 2006; Shaw-Smith et al., 2006). The typical facial phenotype is usually less apparent in the infancy and becomes remarkable during adolescence (Koolen et al., 2008; Slavotinek, 2008). Major anomalies, seizures, joint hyperlaxity and eye anomalies can be also present, but are less common. The prevalence of the syndrome is estimated to 1 in 16,000 (Koolen et al., 2008).

The syndrome is caused by a recurrent 600 kb deletion of 17q21.31. The region is predisposed to rearrangement by its specific genome architecture. The deletion breakpoints map to large clusters of low copy repeats (LCRs) predisposing to non-allelic homologous recombination (NAHR). The 17q21.31 region is known for its inversion polymorphism of about 900 kb and the presence of two highly divergent SNP haplotypes designated H1 and H2. H2 is associated with the inversion and is found at a frequency of 20% in the European population (Stefansson et al., 2005). H2 differs from the non-inverted H1 allele by the arrangement of LCRs, which makes H2 prone to NAHR events (Koolen et al., 2008; Steinberg et al., 2012). At least one of the parents of deletion patients always carried at least one H2 allele, which seems to be necessary for the deletion formation.

RECEIVED 26 June 2013; ACCEPTED 5 September 2013.

ADDRESS FOR CORRESPONDENCE: Miroslava Hancarova, Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Plzenska 130/221, 15000 Prague 5, Czech Republic. E-mail: miroslava.hancarova@lfmotol.cuni.cz

# These authors contributed equally to the work.
* Present address: GHC Genetics, Prague, Czech Republic
** Present address: Institute for the Care of Mother and Child, Prague, Czech Republic

The deletion encompasses several genes, among which haploinsufficiency of *KANSL1* has recently been shown to be responsible for the syndrome (Koolen et al., 2012b; Zollino et al., 2012).

Herein we present the first report of monozygotic twins carrying the 17q21.31 microdeletion and showing only slightly different phenotypes. Analysis on high-resolution arrays did not reveal any genetic differences between the twins. The subtle clinical differences can probably be explained by different perinatal history of the twins or by the variable expressivity of the disorder.

Materials and Methods

Patients

The girls were born from a twin pregnancy to healthy, nonconsanguineous parents of Czech origin. The age of the mother and father were 22 and 25 years, respectively. The delivery was in the 38th week of gestation by cesarean section due to hypoxia in Twin B.

Twin A was born with a weight of 1980 g and length of 43 cm (both below the 3rd centile). The Apgar score was 3-7-7 (Apgar, 1953). Partial exchange transfusion had to be administered due to polyglobulia and hyperviscosity syndrome. The newborn suffered from left-side hypotonic hydronephrosis with reflux. Twin B was born with a weight of 1910 g and length of 43 cm (both below the 3rd centile). The Apgar score was 3-7-7. Perinatal hypoxia followed by intracranial hemorrhage occurred during the delivery. Right-side hydronephrosis, strabismus and horizontal nystagmus were noted in the newborn.

Postaxial polydactyly of toes and fingers, congenital hip dysplasia, delay in motor milestones and speech delay were observed in both twins. Psychological examination at the age of 10 years showed moderate ID in both twins, but Twin A performed slightly better than Twin B (Twin A was assessed as functioning in the upper range of moderate ID and being slightly more diligent and adaptable, and less anxious). At the examination at 19 years of age both twins had disproportionately short stature (Twin A 153.3 cm, the 1st centile; Twin B 157.7 cm, the 6th centile) with shortening of upper and lower limbs, thoracic hyperkyphosis, low-pitched voice and similar facial expression (Figure 1), and with very long, thin and coarse face, coarse hair, thick eyebrows, large nose, bulbous nasal tip, smooth broad philtrum, thick lips, mandibular prognathism, and hirsutism. Twin A had a high palate. Twin B had wide-spaced teeth and diastema, and slightly more coarse facial features compared to Twin A. However, especially with respect to their age, the overall clinical picture of both twins was remarkably similar. None of them showed other symptoms often described in the 17q21.31 microdeletion syndrome, such as seizures, joint hypermobility, cleft lip/palate, heart defects, or pectus excavatum (Koolen et al., 2012b).

Laboratory Analyses

Informed consent for genetic analyses was obtained from the parents of the patients. Genomic DNA of both twins and the parents was extracted from blood lymphocytes using the Gentra Puregen Blood Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Conventional cytogenetic analysis was performed using standard G-banding. The *FMR1* gene testing used the Fragile X PCR Kit (Abbot, Abbot Park, IL, USA). The BAC array comparative genome hybridisation (CGH; BlueGnome, Cambridge, UK) analysis of Twin A was performed according to the manufacturer's instructions. The FISH analysis with the BAC clone RP11-111L23 (BlueGnome) was used to independently confirm the deletion in the twins and to test for its presence in the parents. Diagnostic alleles of single nucleotide polymorphisms (SNPs) rs1800547 (G) and rs9468 (C) and the presence of the 238 bp deletion in intron 9 of the *MAPT* gene characteristic for the H2 allele (Koolen et al., 2008) were analysed in the family using DNA sequencing and gel electrophoresis, respectively (PCR primer sequences are available upon request). The high-resolution SNP array analysis of both twins using the HumanCytoSNP-12 BeadChip (~300 K; Illumina, San Diego, CA, USA) and direct array CGH comparison of their genomes using the Nimblegen 2.1M Whole-Genome CGH Array (Roche NimbleGen, Madison, WI, USA) were used for confirmation of monozygosity and for a more detailed analysis of potential differences in copy number variants (CNVs) in the genomes of both twins. Data were analysed using GenomeStudio (Illumina), QuantiSNP (Colella et al., 2007) and SignalMap (Roche NimbleGen). Multiplex ligation-dependent probe amplification (MLPA) analysis was performed using custom synthetic probes and the P200 Human DNA Reference Probemix (MRC Holland, Amsterdam, The Netherlands; probe sequences are available upon request). All analyses used genome build hg18/NCBI36.

Results

The cytogenetic analysis revealed normal female karyotypes, and the *FMR1* gene testing excluded the fragile X syndrome in both twins. The BAC array CGH analysis of Twin A identified a deletion characteristic for the 17q21.31 microdeletion syndrome with breakpoints between bases 40,740,861-41,074,265 and 41,679,148-42,178,065. The FISH analysis confirmed the deletion in both twins but in neither of their parents. The haplotype analysis revealed homozygosity for the inverted H2 allele in the mother, homozygosity for the non-inverted H1 allele in the father, and hemizygosity for H1 in both twins. Thus the deletion was *de novo* in the twins and it affected one of the maternal chromosomes 17.

The SNP array analysis confirmed the monozygosity of the twins. This high-resolution analysis found no differences in the extent of the 17q21.31 microdeletion between

FIGURE 1

(Colour online) Facial photographs of the patients at the age of 19 (top) and 23 years (bottom). Twin A is on the left, Twin B on the right. Features typical for the 17q21.31 microdeletion syndrome (long, narrow and coarse face, coarse hair, large nose with bulbous nasal tip, broad philtrum, thick lips, mandibular prognathism) and subtle differences between the twins (slightly more coarse facial features in Twin B) can be observed.

the patients (chr17:41,041,709-41,560,151; Figure 2). Both twins shared two additional CNVs, a 0.1 Mb long duplication in 10q26.3 (chr10:135,102,337-135,215,135) encompassing *CYP2E1*, and a 1.7 Mb long deletion in 16p11.2 (chr16:31,977,497-33,704,396) involving *TP53TG3*. Both these CNVs were located in highly polymorphic copy-number variable regions. The analysis with the highest resolution used (2.1M array CGH) did not detect any obvious CNV differences between the genomes of the twins. In several small regions copy number differences between the patients could not be excluded (chr18:14,184,640-15,370,613 and chr21:13,302,864-14,139,384 being most suspicious), but most of these segments coincided with complex segmental duplications, where the validity of the findings was questionable, impossible to confirm using standard methods and of uncertain clinical impact even if they were confirmed. The analysis of three of these regions where unique sequences could be targeted with custom MLPA probes (chr14:18,127,587-19,272,166; chr16:32,082,491-34,128,024 and chr22:49,414,658-49,584,579) failed to confirm any copy number differences between the twins in these regions.

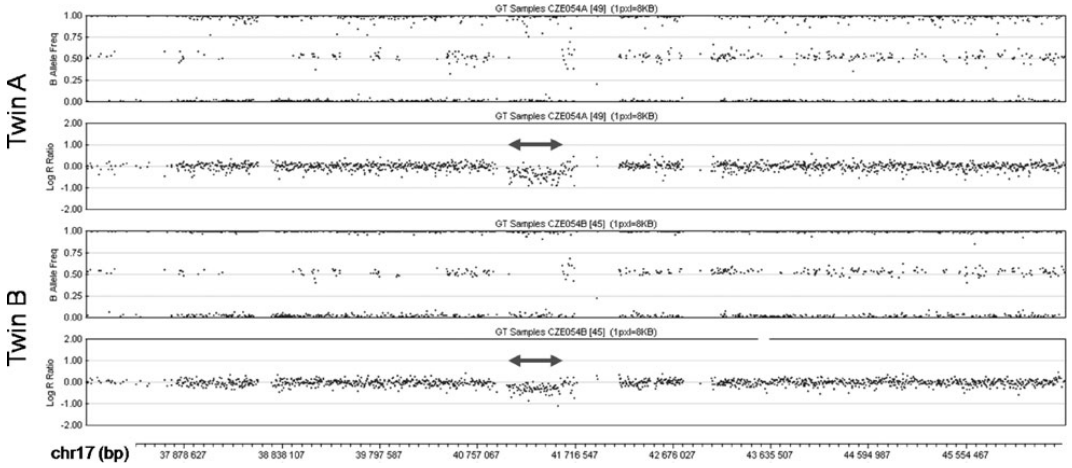


FIGURE 2

SNP array analysis of the middle part of 17q in the patients. The deletions are marked by double arrows. In the diagrams of the B Allele Frequency (top panel in each twin) deletions are indicated by the absence of dots around the value of 0.5 (absence of heterozygous genotype AB). Concurrently, in the diagrams of Log R Ratio (bottom panel in each twin) the deletions are indicated by dots clustering below the value of 0.0 (decreased intensity of the signal).

Discussion

To our knowledge this is the first description of monozygotic twins with the 17q21.31 microdeletion syndrome. The deletion was *de novo* on a maternal chromosome 17, although a low-level somatic and gonadal mosaicism could not be excluded (Koolen et al., 2012a). The twin sisters showed only a subtle phenotypic discordance. Generally, discordant monozygotic twins are a valuable resource for the analysis of genetic, epigenetic or environmental variation contributing to the disease. The 17q21.31 microdeletion syndrome is one of a few clinically recognizable new syndromes with well-defined clinical outcome, and rare instances of affected twins could contribute to understanding the variability of this disorder.

The phenotypes of our patients were very similar and fully corresponded to the typical picture of the syndrome (Koolen et al., 2008). The subtle phenotypic differences between the twins included a slightly more severe cognitive impairment and more coarse facial features with strabismus and horizontal nystagmus in Twin B. These differences prompted us to search for possible genomic differences. The 17q21.31 microdeletion was of the same size in both twins, and also the two other CNVs detectable at the 300K level were present in common and were unlikely to contribute to the phenotype. The 10q26.3 duplication encompassing *CYP2E1* is a common polymorphism possibly associated with alcohol addiction (Deng & Deitrich, 2008). The 16p11.2 deletion around *TP53TG3* affected a very variable gene-poor pericentric region. Also, the direct comparison of both genomes using an even higher resolution (2.1M) did not yield any findings. Several suspicious CNV differences were located in highly polymorphic regions of segmental duplications, the structure of which made the confirmation of these aberrations difficult or impossible, and analysis of three of these regions failed to confirm any differences between the two genomes. However, it should be noted that these regions are susceptible to *de novo* events, and that any genomic differences between the twins could be expected to be in a mosaic state, further complicating their detection. In any case, due to the paucity of genes, these potentially differential CNVs were unlikely to affect the phenotype.

Several other studies focused on monozygotic twins with microdeletion syndromes and a different degree of phenotypic discordance. Ghebranious et al. (2007) presented monozygotic twins with a 16p11.2 microdeletion and no other CNV differences, who showed similar phenotypes but severe aortic stenosis developed only in one twin. Most monozygotic twin pairs reported with 22q11 deletions were also phenotypically discordant. Singh et al. (2002) reviewed five such pairs in whom no high-resolution whole genome analyses were performed to uncover potential genomic differences. The discordance in a recently identified monozygotic twin pair with a 22q11 microdeletion was explained by size differences of the deletions (Halder et al., 2012), which however have not been confirmed using an independent method and are thus questionable. Rio et al. (2013) reported a phenotypically and genetically discordant monozygotic twin pair carrying a 2p25.3 deletion in one twin and mosaicism with one third of cells with the 2p25.3 deletion, one third with a 2p25.3 duplication, and one third of normal cells in the other one. Other recent studies of monozygotic twin pairs discordant for breast cancer (Lasa et al., 2010),

schizophrenia (Ono et al., 2010) or congenital heart defect (Breckpot et al., 2012) also identified no CNV differences explaining the discordance.

In the absence of genetic differences, the twin discordance can be explained by epigenetics or environment (Czyz et al., 2012). The study of *DRD2* methylation in two pairs of monozygotic twins, one discordant and one concordant for schizophrenia, showed that the affected twin from the discordant pair was epigenetically 'closer' to the affected concordant twins than to his unaffected co-twin (Petronis et al., 2003). Similarly, the affected twin from a monozygotic pair discordant for caudal duplication anomaly showed higher methylation of the *AXIN1* promoter than the unaffected twin, whose *AXIN1* methylation was higher than that of normal controls (Oates et al., 2006). An epigenome-wide approach found that approximately one third of monozygotic twins had epigenetic differences in DNA methylation and histone modification (Fraga et al., 2005). Epigenetic marks were more distinct in twins who were older, had different lifestyles, and had spent less of their lives together, underlining the significant role of environmental factors in the process (Fraga et al., 2005; Kaminsky et al., 2009). Environmental factors could include the differences in the intrauterine environment and in perinatal and postnatal history, and the twinning process itself could play a role as well as stochastic factors can do (Czyz et al., 2012). Mosaicism resulting from later postzygotic genomic rearrangements or epigenetic changes can be difficult to detect, and it can differentially affect specific tissues (e.g., the brain) that are not accessible to testing. Another limitation of twin studies, including ours, which are using blood as the source of DNA, is blood chimerism, which can mask genetic or epigenetic discordance (Erlich, 2011).

In the case of our patients who show no CNV differences, all other factors mentioned above could contribute to their subtle phenotypic discordance. The currently emerging whole exome and whole genome sequencing approaches could identify possible genetic variation on the nucleotide level not addressed in our study, and epigenetic differences could also play a role. However, the simplest and likely sufficient explanation of the slightly discordant phenotype of the twins is in their perinatal history, which was clearly more severe in Twin B (perinatal hypoxia followed by intracranial hemorrhage). The differences in the clinical picture of our patients can also be the consequence of stochastic factors acting in the common inter-individual variability, and the variable expressivity of the 17q21.31 microdeletion syndrome.

Acknowledgments

We thank the family of the patients for cooperation. Supported by grants CHERISH 223692 and CZ.2.16/3.1.00/24022 from the European Commission, DRO UH Motol 00064203 and NT/14200 from the Czech Ministry of Health, and SF0180027s10 from the Estonian Ministry of Education and Research.

References

Apgar, V. (1953). A proposal for a new method of evaluation of the newborn infant. *Current Researches in Anesthesia & Analgesia*, *32*, 260–267.

Breckpot, J., Thienpont, B., Gewillig, M., Allegaert, K., Vermeesch, J. R., & Devriendt, K. (2012). Differences in copy number variation between discordant monozygotic twins as a model for exploring chromosomal mosaicism in congenital heart defects. *Molecular Syndromology*, *2*, 81–87.

Colella, S., Yau, C., Taylor, J. M., Mirza, G., Butler, H., Clouston, P., ... Ragoussis, J. (2007). QuantiSNP: An objective Bayes Hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Research*, *35*, 2013–2025.

Czyz, W., Morahan, J. M., Ebers, G. C., & Ramagopalan, S. V. (2012). Genetic, environmental and stochastic factors in monozygotic twin discordance with a focus on epigenetic differences. *BMC Medicine*, *10*, 93.

Deng, X. S., & Deitrich, R. A. (2008). Putative role of brain acetaldehyde in ethanol addiction. *Current Drug Abuse Reviews*, *1*, 3–8.

Erlich, Y. (2011). Blood ties: Chimerism can mask twin discordance in high-throughput sequencing. *Twin Research and Human Genetics*, *14*, 137–143.

Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., ... Esteller, M. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Science of the U S A*, *102*, 10604–10609.

Ghebranious, N., Giampietro, P. F., Wesbrook, F. P., & Rezkalla, S. H. (2007). A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. *American Journal of Medical Genetics Part A*, *143A*, 1462–1471.

Halder, A., Jain, M., Chaudhary, I., & Varma, B. (2012). Chromosome 22q11.2 microdeletion in monozygotic twins with discordant phenotype and deletion size. *Molecular Cytogenetics*, *5*, 13.

Kaminsky, Z. A., Tang, T., Wang, S. C., Ptak, C., Oh, G. H., Wong, A. H., ... Petronis, A. (2009). DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genetics*, *41*, 240–245.

Koolen, D. A., Dupont, J., de Leeuw, N., Vissers, L. E., van den Heuvel, S. P., Bradbury, A., ... Parker, M. J. (2012a). Two families with sibling recurrence of the 17q21.31 microdeletion syndrome due to low-grade mosaicism. *European Journal of Human Genetics*, *20*, 729–733.

Koolen, D. A., Kramer, J. M., Neveling, K., Nillesen, W. M., Moore-Barton, H. L., Elmslie, F. V., ... de Vries, B. B. (2012b). Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nature Genetics*, *44*, 639–641.

Koolen, D. A., Sharp, A. J., Hurst, J. A., Firth, H. V., Knight, S. J., Goldenberg, A., ... de Vries, B. B. (2008). Clinical

and molecular delineation of the 17q21.31 microdeletion syndrome. *Journal of Medical Genetics*, *45*, 710–720.

Koolen, D. A., Vissers, L. E., Pfundt, R., de Leeuw, N., Knight, S. J., Regan, R., ... de Vries, B. B. (2006). A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nature Genetics*, *38*, 999–1001.

Lasa, A., Ramon y Cajal, T., Llort, G., Suela, J., Cigudosa, J. C., Cornet, M., ... Baiget, M. (2010). Copy number variations are not modifiers of phenotypic expression in a pair of identical twins carrying a BRCA1 mutation. *Breast Cancer Research and Treatment*, *123*, 901–905.

Oates, N. A., van Vliet, J., Duffy, D. L., Kroes, H. Y., Martin, N. G., Boomsma, D. I., ... Chong, S. (2006). Increased DNA methylation at the AXIN1 gene in a monozygotic twin from a pair discordant for a caudal duplication anomaly. *American Journal of Human Genetics*, *79*, 155–162.

Ono, S., Imamura, A., Tasaki, S., Kurotaki, N., Ozawa, H., Yoshiura, K., & Okazaki, Y. (2010). Failure to confirm CNVs as of aetiological significance in twin pairs discordant for schizophrenia. *Twin Research and Human Genetics*, *13*, 455–460.

Petronis, A., Gottesman, II, Kan, P., Kennedy, J. L., Basile, V. S., Paterson, A. D., & Popendikyte, V. (2003). Monozygotic twins exhibit numerous epigenetic differences: Clues to twin discordance? *Schizophrenia Bulletin*, *29*, 169–178.

Rio, M., Royer, G., Gobin, S., de Blois, M., Ozilou, C., Bernheim, A., ... Malan, V. (2013). Monozygotic twins discordant for submicroscopic chromosomal anomalies in 2p25.3 region detected by array CGH. *Clinical Genetics*, *84*, 31–36.

Sharp, A. J., Hansen, S., Selzer, R. R., Cheng, Z., Regan, R., Hurst, J. A., ... Eichler, E. E. (2006). Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nature Genetics*, *38*, 1038–1042.

Shaw-Smith, C., Pittman, A. M., Willatt, L., Martin, H., Rickman, L., Gribble, S., ... Carter, N. P. (2006). Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nature Genetics*, *38*, 1032–1037.

Singh, S. M., Murphy, B., & O'Reilly, R. (2002). Monozygotic twins with chromosome 22q11 deletion and discordant phenotypes: Updates with an epigenetic hypothesis. *Journal of Medical Genetics*, *39*, e71.

Slavotinek, A. M. (2008). Novel microdeletion syndromes detected by chromosome microarrays. *Human Genetics*, *124*, 1–17.

Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., ... Stefansson, K. (2005). A common inversion under selection in Europeans. *Nature Genetics*, *37*, 129–137.

Steinberg, K. M., Antonacci, F., Sudmant, P. H., Kidd, J. M., Campbell, C. D., Vives, L., ... Eichler, E. E. (2012). Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nature Genetics*, *44*, 872–880.

Zollino, M., Orteschi, D., Murdolo, M., Lattante, S., Battaglia, D., Stefanini, C., ... Marangi, G. (2012). Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nature Genetics*, *44*, 636–638.

Publikace 7

Deletions of 9q21.3 including *NTRK2* are associated with severe phenotype.

Hancarova M, Puchmajerova A, Drabova J, Karaskova E, Vlckova M, Sedlacek Z.

Am J Med Genet A. 2015 Jan;167A(1):264-7. IF: 2,159

Kazuistika popisuje těžce postiženou pacientku se vzácnou mikrodelecí 9q21.3. V literatuře byly v době publikace popsány pouze dva případy postižených s podobnou delecí (jeden jako kazuistika a jeden v rámci rozsáhlejší studie). V databázích ISCA a DECIPHER byly uveřejněny další případy, ale bez popisu fenotypu. V práci se zabýváme zejména funkcí genů zasažených delecí a jejich možným příspěvkem k fenotypu pacientů, krátce je rovněž diskutován i mechanismus vzniku aberace.

Pacientka měla mnohočetné VVV (defekt septa síní, vrozenou dysplázii kyčlí, vezikoureterální reflux, equionovary), poruchu růstu, psychomotorické opoždění, těžkou MR a faciální stigmatizaci. Její mikrodelece 9q21.3 byla menší než u dříve publikovaného případu a zahrnovala 7 protein kódujících genů. Fenotyp obou pacientů byl podobný: psychomotorické opoždění, některé faciální rysy, mikrocefalie, hypotonie, VCC, vrozená dysplázie kyčlí a deformity končetin. Hlavním rozdílem bylo, že dříve popsaný pacient zemřel ve věku 14 měsíců na respirační selhání, ale naše pacientka žádné respirační obtíže neměla. Ze zasažených genů jsme se jako na kandidátní pro MR a další neurologické obtíže zaměřili na *AGTBP1* (exprimovaný v mozku plodu; homozygotní mutace homologu *AGTPBP1* způsobují nemoc dolního motoneuronu u ovcí), *NTRK2* (exprimovaný v mozku, zodpovědný za vývoj centrálního i periferního nervového systému, součást dráhy MAPK) a *HNRNPK* (jehož down-regulace negativně ovlivňuje formování synapsí). K respiračnímu selhání u dříve publikovaného pacienta mohla přispět haploinsuficience genu *KIF27,* který ovlivňuje ciliogenesi.

Práce podporuje domněnku, že mikrodelece v oblasti 9q21.3 je patogenní a může podmiňovat vzácný mikrodeleční syndrom s nerekurentními zlomovými místy. Tento koncept byl podpořen popisem fenotypově velmi podobných pacientů s intragenovými defekty *HNRNPK* (natolik podobných, že autoři přetiskli i fotografii námi popsané pacientky) (Au et al., 2015), takže i u naší pacientky je za nejzávažnější příznaky nejspíše odpovědná delece právě tohoto genu. Další práce popisuje pacienta s *de novo* frameshift mutací *HNRNPK*, jehož faciální fenotyp jakož i spektrum VVV jsou podobné dříve publikovaným případům, včetně naší pacientky (Lange et al., 2016).

CORRESPONDENCE

AMERICAN JOURNAL OF medical genetics PART A

Deletions of 9q21.3 Including *NTRK2* Are Associated With Severe Phenotype

Miroslava Hancarova,* Alena Puchmajerova, Jana Drabova, Eliska Karaskova, Marketa Vlckova and Zdenek Sedlacek

Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic

Manuscript Received: 28 April 2014; Manuscript Accepted: 4 September 2014

TO THE EDITOR:

We have read with interest the report "Novel Interstitial 2.6 Mb Deletion on 9q21 Associated With Multiple Congenital Anomalies" by Pua et al. [2014], describing a patient with craniofacial abnormalities, heart defect, bicornuate uterus, extremity deformities, hip dislocation, delayed myelination, and recurrent pneumonia, who died after respiratory decompensation at 14 months of age. The region deleted in the patient showed limited copy number variation (CNV) in both normal and affected individuals, and the article prompted reporting of additional variants of this region to better understand their pathogenicity and the role of individual genes.

In the course of CNV analysis in children with developmental delay (DD) and/or autistic features using SNP arrays (Human CytoSNP-12 BeadChips, Illumina, San Diego, CA) we identified a 9q21.3 deletion of about 2 Mb, arr[hg19] 9q21.32–q21.33 (86,369,356 × 2, 86,595,071–88,357,495 × 1, 88,477,869 × 2) dn, as a sole aberration in a severely affected Czech patient (Fig. 1). The deletion breakpoints were most likely located in segmental duplications represented by the *KIF27* gene on the proximal side and a rearranged non-processed *KIF27* pseudogene (*LOC389765*) on the distal side (Fig. 2). The deletion removed 7 protein-coding RefSeq genes, *KIF27*, *C9orf64*, *HNRNPK*, *RMI1*, *SLC28A3*, *NTRK2*, and *AGTPBP1* (or 8 genes including *GKAP1* if the proximal breakpoint were centromeric from the segmental duplication), and also the *MIR7–1* gene (Fig. 2). The deletion was confirmed using FISH with BAC clone RP11–59M22 (BlueGnome, Cambridge, UK) in the patient but in none of her parents. Subsequent SNP array analysis of the parents confirmed the de novo nature of the deletion and showed its paternal origin.

The currently 13-year-old girl was the first child of healthy nonconsanguineous Czech parents aged 29 and 28, both university graduates. The second trimester biochemical screening in pregnancy showed values pointing to an increased risk of chromosome aberrations (AFP 0.89 MoM; uE3 0.68 MoM; HCG 3.16 MoM) but the fetal karyotype from amniotic fluid cells was normal. The girl was delivered at 42 weeks of gestation by Caesarean due to abnormal cardiotocography. The birth parameters were nearly normal: weight of 3,650 g (25th–50th centile), length of 50 cm (25th centile) and head circumference (HC) of 34 cm (<10th centile). Systolic murmur pointed to a heart defect soon after birth. The patient underwent cardiosurgical closure of an atrioventricular septal defect and extirpation of a right atrial hemangioma at 3 weeks of age. Starting from the age of 3 months she had repeated orthopedic surgery for congenital hip dysplasia (bilateral acetabulum reconstruction) and equinovarus deformity of the right foot. Bilateral vesicoureteral reflux was also noted. The growth parameters dropped under the 3rd centile at 5 months of age and have remained low ever since. At this age a severe psychomotor delay and quadriparesis, mostly affecting the lower extremities, were also noted. The hypotonic syndrome developed into spasticity. At 13 years of age the patient had a height of 140 cm, weight of 27 kg, and HC of 51 cm (all <3rd centile). She was able to sit only with support and for a short period of time, was not able to walk and was permanently bound to a wheelchair. This condition led to generalized muscular atrophy. She had severe intellectual disability (ID) and severe speech delay (she used 12 words). Due to the severity of ID, formal testing for autism could not be performed. The psychomotor development markedly regressed after each administering of anesthesia. The craniofacial features of the patient included low posterior hairline, narrow forehead, long downslanting palpebral

How to Cite this Article:
Hancarova M, Puchmajerova A, Drabova J, Karaskova E, Vlckova M, Sedlacek Z. 2015. Deletions of 9q21.3 including *NTRK2* are associated with severe phenotype.
Am J Med Genet Part A 167A:264–267.

Conflict of interests: none.
Grant sponsor: Czech Ministry of Health; Grant numbers: NT/14200, DRO UH Motol 00064203; Grant sponsor: European Commission; Grant numbers: CHERISH 223692, CZ.2.16/3.1.00/24022.
*Correspondence to:
Miroslava Hancarova, Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Plzenska 130/221, 15000 Prague 5, Czech Republic.
E-mail: miroslava.hancarova@lfmotol.cuni.cz
Article first published online in Wiley Online Library (wileyonlinelibrary.com): 27 October 2014
DOI 10.1002/ajmg.a.36797

© 2014 Wiley Periodicals, Inc.

FIG. 1. Photograph of the patient at the age of 3.5 years (A), 6 years (B), and 10.5 years (C), and details of the irregular finger width, the low-set thumb (D), and overlapping toes on the lower limbs (E). The facial features include narrow forehead, long downslanting palpebral fissures, abnormal nasal bridge, downturned corners of the mouth, permanently open mouth, and abnormal large low-set ears.

fissures, palpebral edema, nasal bridge changing with age from depressed into prominent, broad philtrum, full red cheeks, larger mouth with downturned corners, thick lips, permanently open mouth, high palate, abnormal large low-set ears, and hypomimia (Fig. 1). The patient had a single transverse palmar crease, fingers of irregular width, low-set thumbs and overlapping toes on the lower limbs. Electromyography showed peripheral demyelinating neuropathy together with a central lesion. The conduction velocity of peripheral motor nerve fibers was decreased. Brain ultrasound did not identify any significant anomalies. Ophthalmological examination proved hypermetropia.

The deletion in our patient was smaller but overlapped significantly with the deletion of the patient of Pua et al. [2014] which removed 4–5 additional protein-coding RefSeq genes (*RASEF*, *FRMD3*, *IDNK*, *UBQLN1* (and *GKAP1*)) but left the *AGTPBP1* gene intact (Fig. 2). Both patients manifested with psychomotor retardation, craniofacial abnormalities (e.g., low-set ears, downturned corners of the mouth, full cheeks, and low hairline), microcephaly, hypotonia, heart defect, extremity deformities, and hip dislocation. The severity of psychomotor delay could not be compared due to the low age and lacking information on the patient of Pua et al. [2014]. Also the examinations performed in both patients differed markedly making a detailed comparison difficult. The main differences between both patients were respiratory distress and chronic lung disease in the patient of Pua et al. [2014] completely absent in our patient.

Literature and database search revealed just one deletion similar to the deletions of the two patients. In a large study of children with DD, Cooper et al. [2011] identified a patient (9896441) with unspecified craniofacial and neurological abnormalities and multiple congenital anomalies, and a deletion of the interval *HNRNPK–AGTPBP1* (Fig. 2). Three other deletions (259303, 269941, and nssv1602706 from the DECIPHER and ISCA databases, respectively) had an overlap but lacked any phenotype information, as did three other deletions from these databases which also were too large to allow relevant comparison (252750, 263384, and nssv1494932, Fig. 2). Three intragenic *NTRK2* deletions were described in autism [Sanders et al., 2011]. A cluster of four duplications spanning the interval *NTRK2–AGTPBP1* found in DECIPHER and ISCA patients with ID or DD (Fig. 2) but not in normal subjects from DGV could also be of interest (264794 and nssv583231 with no phenotype information; nssv1603921 with delayed speech development and muscular hypotonia; and nssv1609376 with ADHD, autism, pica, delayed speech and fine motor development, and sleep disturbances), as well as the de novo duplication mirroring the deletion described by Pua et al. [2014] in patient 260708 with cryptorchidism, delayed speech, and short attention span. The influence of duplications on the phenotype could be different from that of deletions, and the same could apply to the heterozygous missense *NTRK2* mutation described in an obese child with ID [Yeo et al., 2004]. Concerning the normal variation of the region in controls, DGV lists just two clusters of small deletions overlapping the *SLC28A3* and *AGTPBP1* genes (Fig. 2), and EVS lists multiple truncating variants in *KIF27*, and 1, 1, 8, and 9 such variants in *C9orf64*, *AGTPBP1*, *SLC28A3*, and *RMI1*, respectively, but none in *HNRNPK* and *NTRK2* (in a total of about 13,000 alleles sequenced).

Several genes from 9q21.3 are good candidates for the symptoms observed in the patients. Supplementary Table S1 (in supporting information online) summarizes the detailed information available on the genes deleted in our patient. Below, we focus on prime candidates and genes which could possibly explain the differences in patient phenotypes. The ATP/GTP binding protein 1 (*AGTPBP1*) gene is highly expressed in human fetal brain. The protein coded by the mouse homologue, Nna1, has a role in neuronal bioenergetics and its mutations or abnormal expression cause the Purkinje cell degeneration (pcd) phenotype which includes ataxia and degeneration of mitral cells of the olfactory bulb, thalamic neurons and retinal photoreceptor cells [Fernandez-Gonzalez et al., 2002]. Moreover, a homozygous missense mutation in the sheep *AGTPBP1* homologue causes lower motor neuron disease with poor muscle tone, muscle atrophy, and severe tetraparesis [Zhao et al., 2012]. Therefore, *AGTPBP1* is a candidate for some of the neurological symptoms in our patient. The *NTRK2* gene coding type 2 neurotrophic tyrosine kinase receptor is a good candidate for the neurodevelopmental delay in both patients. This kinase, highly expressed in human brain, is responsible for the development of the central and peripheral nervous system; it is involved in the MAPK

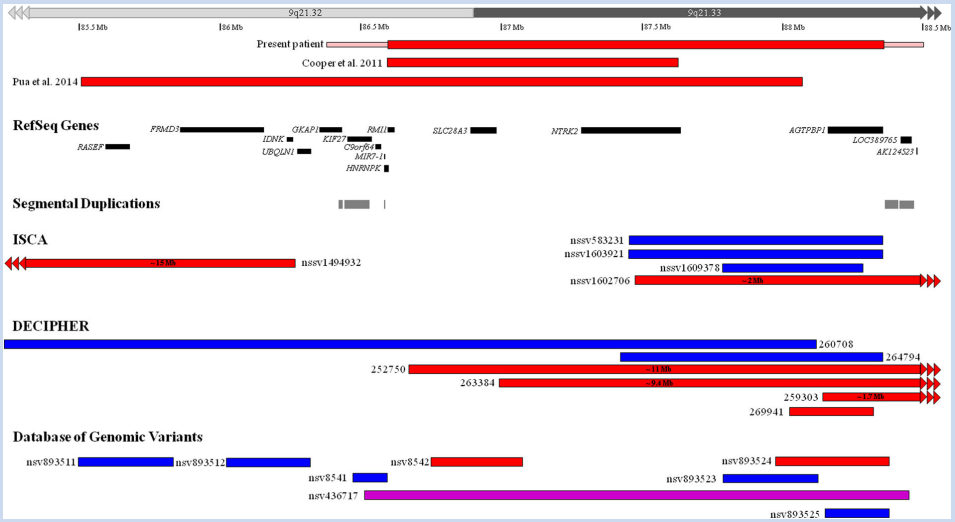


FIG. 2. Schematics of the region affected by the deletion in our patient including the protein-coding genes, segmental duplications and previously published structural variants. Deletions are depicted in red, duplications in blue, and inversions in purple. The uncertainty in the position of deletion breakpoints in our patient is in pink. Only aberrations overlapping with the deletion in our patient are shown. The total size in Mb is indicated in variants significantly extending outside from the region studied in the proximal or distal direction (triple arrowheads). Structural variant regions from DGV are indicated by bars representing variable numbers of supporting variants (one for nsv893511 from one study with a frequency of 1/6533; one for nsv893512 (1 study, 1/6533); one for nsv8541 (1 study, 1/31); two for nsv436717 [Korbel et al., 2007]; four for nsv8542 (1 study, 4/31); one for nsv893523 (1 study, 1/6533); 11 for nsv893524 [3 studies, 9/6533, 1/2026 and 1/95); one for nsv893525 (1 study, 1/6533)].

pathway and belongs to the postsynaptic density proteome as well as to FMRP targets [Huang and Reichardt 2001]. *NTRK2* variants have been associated with a range of neuropsychiatric and neurodegenerative disorders [Gupta et al., 2013]. Similarly also the *RMI1* gene (RecQ mediated genome instability 1) is highly expressed in human fetal brain and is associated with embryo development [Guiraldelli et al., 2013]. Finally, another member of the postsynaptic density proteome, heterogeneous nuclear ribonucleoprotein K, the product of the *HNRNPK* gene, is also a candidate due to its role in gene expression, chromatin remodeling or actin dynamics, and because its downregulation leads to reduced synapse formation [Proepper et al., 2011]. The most significant phenotypic differences between the two patients, namely the respiratory distress and chronic lung disease, could be possibly related to the function of the *KIF27* gene in motile ciliogenesis [Katoh and Katoh 2004]. While this gene is clearly deleted in the patient by Pua et al. [2014] the consequences of the rearrangement in our patient (possibly involving a recombination between *KIF27* and its pseudogene) for the *KIF27* function are unclear.

The limited rate of CNV around the *NTRK2* gene contrasts with the remarkable genome architecture of the region. Segmental duplications represented by *KIF27* and its pseudogene have a total length of about 100 kb, nucleotide similarity exceeding 97% and direct orientation in the reference genome [Bailey et al., 2001], and as such they may predispose to deletions and duplications [Liu et al., 2012]. However, our patient is the only individual reported with a deletion of the region flanked by the segmental duplications, which themselves show structural variation. Remarkably, three unrelated Yoruba individuals carry an inversion of the region deleted in our patient [Korbel et al., 2007; Kidd et al., 2008]. The notion that inverted alleles predispose to several recurrent deletions [Girirajan and Eichler 2010] prompted us to test the inversion status in the parents of our patient, especially the father, whose chromosome 9 was affected by the deletion. FISH with BAC clones RP11–158D2 (BlueGnome), RP11–522I20, RP11–202I11, and RP11–213G2 (Source BioScience, Nottingham, UK) showed that none of the parental chromosomes 9 were inverted (data not shown). The paucity of deletions of this region could indicate that its hemizygous state may indeed be associated with a severe phenotype, or even may be a subject of prenatal selection.

ACKNOWLEDGMENTS

We thank the family of the patient for cooperation, and Tereza Jancuskova and Sona Pekova for the preparation of FISH probes.

WEB RESOURCES

DECIPHER, http://decipher.sanger.ac.uk; DGV, http://dgv.tcag.ca; EVS, http://evs.gs.washington.edu/EVS/; ISCA, http://www.iscaconsortium.org.

REFERENCES

Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. 2001. Segmental duplications: Organization and impact within the current human genome project assembly. Genome Res 11:1005–1017.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, Abdel-Hamid H, Bader P, McCracken E, Niyazov D, Leppig K, Thiese H, Hummel M, Alexander N, Gorski J, Kussmann J, Shashi V, Johnson K, Rehder C, Ballif BC, Shaffer LG, Eichler EE. 2011. A copy number variation morbidity map of developmental delay. Nat Genet 43:838–846.

Fernandez-Gonzalez A, La Spada AR, Treadaway J, Higdon JC, Harris BS, Sidman RL, Morgan JI, Zuo J. 2002. Purkinje cell degeneration (pcd) phenotypes caused by mutations in the axotomy-induced gene, Nna1. Science 295:1904–1906.

Girirajan S, Eichler EE. 2010. Phenotypic variability and genetic susceptibility to genomic disorders. Hum Mol Genet 19:R176–R187.

Guiraldelli MF, Eyster C, Pezza RJ. 2013. Genome instability and embryonic developmental defects in RMI1 deficient mice. DNA Repair (Amst) 12:835–843.

Gupta VK, You Y, Gupta VB, Klistorner A, Graham SL. 2013. TrkB Receptor Signalling: Implications in neurodegenerative, psychiatric and proliferative disorders. Int J Mol Sci 14:10122–10142.

Huang EJ, Reichardt LF. 2001. Neurotrophins: Roles in neuronal development and function. Annu Rev Neurosci 24:677–736.

Katoh Y, Katoh M. 2004. KIF27 is one of orthologs for Drosophila Costal-2. Int J Oncol 25:1875–1880.

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tuzun E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M, Saranga D, Brand A, Tao W, Gustafson E, McKernan K, Chen L, Malig M, Smith JD, Korn JM, McCarroll SA, Altshuler DA, Peiffer DA, Dorschner M, Stamatoyannopoulos J, Schwartz D, Nickerson DA, Mullikin JC, Wilson RK, Bruhn L, Olson MV, Kaul R, Smith DR, Eichler EE. 2008. Mapping and sequencing of structural variation from eight human genomes. Nature 453:56–64.

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M. 2007. Paired-end mapping reveals extensive structural variation in the human genome. Science 318:420–426.

Liu P, Carvalho CM, Hastings PJ, Lupski JR. 2012. Mechanisms for recurrent and complex human genomic rearrangements. Curr Opin Genet Dev 22:211–220.

Proepper C, Steinestel K, Schmeisser MJ, Heinrich J, Steinestel J, Bockmann J, Liebau S, Boeckers TM. 2011. Heterogeneous nuclear ribonucleoprotein k interacts with Abi-1 at postsynaptic sites and modulates dendritic spine morphology. PLoS ONE 6:e27045.

Pua HH, Krishnamurthi S, Farrell J, Margeta M, Ursell PC, Powers M, Slavotinek AM, Jeng LJ. 2014. Novel interstitial 2.6 Mb deletion on 9q21 associated with multiple congenital anomalies. Am J Med Genet Part A 164A:237–242.

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, Mason CE, Bilguvar K, Celestino-Soper PB, Choi M, Crawford EL, Davis L, Wright NR, Dhodapkar RM, DiCola M, DiLullo NM, Fernandez TV, Fielding-Singh V, Fishman DO, Frahm S, Garagaloyan R, Goh GS, Kammela S, Klei L, Lowe JK, Lund SC, McGrew AD, Meyer KA, Moffat WJ, Murdoch JD, O'Roak BJ, Ober GT, Pottenger RS, Raubeson MJ, Song Y, Wang Q, Yaspan BL, Yu TW, Yurkiewicz IR, Beaudet AL, Cantor RM, Curland M, Grice DE, Gunel M, Lifton RP, Mane SM, Martin DM, Shaw CA, Sheldon M, Tischfield JA, Walsh CA, Morrow EM, Ledbetter DH, Fombonne E, Lord C, Martin CL, Brooks AI, Sutcliffe JS, Cook EH, Jr., Geschwind D, Roeder K, Devlin B, State MW. 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70:863–885.

Yeo GS, Connie Hung, Rochford CC, Keogh J, Gray J, Sivaramakrishnan J, O'Rahilly S, Farooqi S. 2004. A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. Nat Neurosci 7:1187–1189.

Zhao X, Onteru SK, Dittmer KE, Parton K, Blair HT, Rothschild MF, Garrick DJ. 2012. A missense mutation in AGTPBP1 was identified in sheep with a lower motor neuron disease. Heredity (Edinb) 109:156–162.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Supplementary Table 1: Detailed overview of function and phenotype associations of genes deleted in our patient.

Gene symbol	Gene name	Function	Phenotype association	References
GKAP1	G kinase anchoring protein 1	cGMP-dependent protein kinase; germ cell development	associated with acute myeloid leukemia and myeloid leukemia	[Sweetser et al., 2005; Yuasa et al., 2000]
KIF27	kinesin family member 27	microtubule-associated kinesin-like motor; Hedgehog signaling pathway; motile ciliogenesis, ATP binding	knockout mice are small and die by 8 weeks, defective function of ependymal and respiratory cilia	[Katoh and Katoh 2004; Klejnot and Kozielski 2012; Vogel et al., 2012]
C9orf64	chromosome 9 open reading frame 64	somatic cell memory gene; required for efficient human induced pluripotent stem (iPS) cell generation; possible tumor suppressor in ovarian cancer	deleted in some cases of acute myeloid leukemia; its promoter region is methylated in some breast cancer cell lines	[Cai et al., 2007; Ohi et al., 2011; Sweetser et al., 2005]
HNRNPK	heterogeneous nuclear ribonucleoprotein K	a member of subfamily of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNPs); conserved RNA binding protein; involved in multiple processes of gene expression, including chromatin remodeling, transcription, mRNA splicing, mRNA stability and mRNA translation; important for pre-mRNA processing and transport; a target of several intracellular signaling cascades (posttranscriptional regulation processes); modulation of expression of neurofilament mRNAs during development of the cerebral cortex; central role in actin dynamics; transcriptional co-activator of p53; member of postsynaptic density proteome; predicted to be likely haploinsufficient	the downregulation of hnRNPK results in extensive filopodia formation and enlargement of the dendritic tree, in reducing of mature synapse formation, and in abolished activation of p53 transcription	[Bayes et al., 2011; Fukuda et al., 2009; Moumen et al., 2013; Proepper et al., 2011; Thyagarajan and Szaro 2008]

Gene symbol	Gene name	Function	Phenotype association	References
RMI1	RecQ mediated genome instability 1	a component of evolutionarily conserved BLM/TOP3α/RMI1 protein complex; limits DNA crossover formation via dissolution of double Holliday junctions; promotes genome stability; a dual-role in embryonic development and tumor suppression; predicted to be likely haploinsufficient	depletion leads to increased sister chromatids exchange level similar to BLM knock out (Bloom syndrome, MIM 210900); RMI1 knock out in mice showed a severe developmental delay resulting in embryonic lethality	[Guiraldelli et al., 2013; Chen et al., 2011; Raynard et al., 2006; Yin et al., 2005]
SLC28A3	solute carrier family 28 (concentrative nucleoside transporter), member 3	sodium-dependent nucleoside transporter; regulation of multiple cellular processes, including neurotransmission, vascular tone or adenosine concentration in the vicinity of cell surface receptors	therapeutic response in patients with HCV infection; prognosis of pancreatic cancer; prediction of anthracycline-induced cardiotoxicity in children	[Mohelnikova-Duchonova et al., 2013; Rau et al., 2013]
NTRK2 / TRKB	neurotrophic tyrosine kinase, receptor, type 2	receptor of brain-derived neurotrophic factor BDNF; regulation of both short-term synaptic functions and long-term potentiation of brain synapses; indispensable for the survival, development and synaptic plasticity of several subtypes of neurons; activation of intracellular signaling cascades (MAP kinase pathway); related to cell growth, cell survival, and cytoskeletal dynamics; member of postsynaptic density proteome; FMRP (fragile X mental retardation protein) target	molecular variants predispose to depressions; suicide behavior; progression of anxiety disorders; schizophrenia; bipolar disorder; autism; drug abuse like nicotine dependence; loss of TrkB signaling associated with Alzheimer's, Huntington's and other neurodegenerative disorders; vulnerability to eating disorders; a heterozygous missense mutation was associated with early-onset obesity, hyperphagia and severe developmental delay	[Correia et al., 2010; Ernst et al., 2011; Huang and Reichardt 2001; Chen et al., 2008; Li et al., 2008; Murphy et al., 2012; Murphy et al., 2011; Ribases et al., 2005; Soontornniyomkij et al., 2011; Yeo et al., 2004]
AGTPBP1 / NNA1	ATP/GTP binding protein 1	a member of the cytosolic carboxypeptidase subfamily; a role in protein turnover, neuronal bioenergetics, axonal regeneration; FMRP target	mutations or abnormal expression cause a Purkinje cell degeneration phenotype in mice; homozygous missense mutation associated with lower motor neuron disease in sheep	[Fernandez-Gonzalez et al., 2002; Kalinina et al., 2007; Zhao et al., 2012]

Additional References for Supplementary Table 1

Bayes A, van de Lagemaat LN, Collins MO, Croning MD, Whittle IR, Choudhary JS, Grant SG. 2011. Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nat Neurosci 14:19-21.

Cai LY, Abe M, Izumi S, Imura M, Yasugi T, Ushijima T. 2007. Identification of PRTFDC1 silencing and aberrant promoter methylation of GPR150, ITGA8 and HOXD11 in ovarian cancers. Life Sci 80:1458-1465.

Correia CT, Coutinho AM, Sequeira AF, Sousa IG, Lourenco Venda L, Almeida JP, Abreu RL, Lobo C, Miguel TS, Conroy J, Cochrane L, Gallagher L, Gill M, Ennis S, Oliveira GG, Vicente AM. 2010. Increased BDNF levels and NTRK2 gene association suggest a disruption of BDNF/TrkB signaling in autism. Genes Brain Behav 9:841-848.

Ernst C, Wanner B, Brezo J, Vitaro F, Tremblay R, Turecki G. 2011. A deletion in tropomyosin-related kinase B and the development of human anxiety. Biol Psychiatry 69:604-607.

Fernandez-Gonzalez A, La Spada AR, Treadaway J, Higdon JC, Harris BS, Sidman RL, Morgan JI, Zuo J. 2002. Purkinje cell degeneration (pcd) phenotypes caused by mutations in the axotomy-induced gene, Nna1. Science 295:1904-1906.

Fukuda T, Naiki T, Saito M, Irie K. 2009. hnRNP K interacts with RNA binding motif protein 42 and functions in the maintenance of cellular ATP level during stress conditions. Genes Cells 14:113-128.

Guiraldelli MF, Eyster C, Pezza RJ. 2013. Genome instability and embryonic developmental defects in RMI1 deficient mice. DNA Repair (Amst) 12:835-843.

Huang EJ, Reichardt LF. 2001. Neurotrophins: roles in neuronal development and function. Annu Rev Neurosci 24:677-736.

Chen H, You MJ, Jiang Y, Wang W, Li L. 2011. RMI1 attenuates tumor development and is essential for early embryonic survival. Mol Carcinog 50:80-88.

Chen Z, Simmons MS, Perry RT, Wiener HW, Harrell LE, Go RC. 2008. Genetic association of neurotrophic tyrosine kinase receptor type 2 (NTRK2) With Alzheimer's disease. Am J Med Genet B Neuropsychiatr Genet 147:363-369.

Kalinina E, Biswas R, Berezniuk I, Hermoso A, Aviles FX, Fricker LD. 2007. A novel subfamily of mouse cytosolic carboxypeptidases. FASEB J 21:836-850.

Katoh Y, Katoh M. 2004. KIF27 is one of orthologs for Drosophila Costal-2. Int J Oncol 25:1875-1880.

Klejnot M, Kozielski F. 2012. Structural insights into human Kif7, a kinesin involved in Hedgehog signalling. Acta Crystallogr D Biol Crystallogr 68:154-159.

Li MD, Lou XY, Chen G, Ma JZ, Elston RC. 2008. Gene-gene interactions among CHRNA4, CHRNB2, BDNF, and NTRK2 in nicotine dependence. Biol Psychiatry 64:951-957.

Mohelnikova-Duchonova B, Brynychova V, Hlavac V, Kocik M, Oliverius M, Hlavsa J, Honsova E, Mazanec J, Kala Z, Melichar B, Soucek P. 2013. The association between the expression of solute carrier transporters and the prognosis of pancreatic cancer. Cancer Chemother Pharmacol 72:669-682.

Moumen A, Magill C, Dry KL, Jackson SP. 2013. ATM-dependent phosphorylation of heterogeneous nuclear ribonucleoprotein K promotes p53 transcriptional activation in response to DNA damage. Cell Cycle 12:698-704.

Murphy ML, Carballedo A, Fagan AJ, Morris D, Fahey C, Meaney J, Frodl T. 2012. Neurotrophic tyrosine kinase polymorphism impacts white matter connections in patients with major depressive disorder. Biol Psychiatry 72:663-670.

Murphy TM, Ryan M, Foster T, Kelly C, McClelland R, O'Grady J, Corcoran E, Brady J, Reilly M, Jeffers A, Brown K, Maher A, Bannan N, Casement A, Lynch D, Bolger S, Tewari P, Buckley A, Quinlivan L, Daly L, Kelleher C, Malone KM. 2011. Risk and protective genetic variants in suicidal behaviour: association with SLC1A2, SLC1A3, 5-HTR1B &NTRK2 polymorphisms. Behav Brain Funct 7:22.

Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ, Yu J, Hebrok M, Hochedlinger K, Costello JF, Song JS, Ramalho-Santos M. 2011. Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat Cell Biol 13:541-549.

Proepper C, Steinestel K, Schmeisser MJ, Heinrich J, Steinestel J, Bockmann J, Liebau S, Boeckers TM. 2011. Heterogeneous nuclear ribonucleoprotein k interacts with Abi-1 at postsynaptic sites and modulates dendritic spine morphology. PLoS One 6:e27045.

Rau M, Stickel F, Russmann S, Manser CN, Becker PP, Weisskopf M, Schmitt J, Dill MT, Dufour JF, Moradpour D, Semela D, Mullhaupt B, Geier A. 2013. Impact of genetic SLC28 transporter and ITPA variants on ribavirin serum level, hemoglobin drop and therapeutic response in patients with HCV infection. J Hepatol 58:669-675.

Raynard S, Bussen W, Sung P. 2006. A double Holliday junction dissolvasome comprising BLM, topoisomerase IIIalpha, and BLAP75. J Biol Chem 281:13861-13864.

Ribases M, Gratacos M, Badia A, Jimenez L, Solano R, Vallejo J, Fernandez-Aranda F, Estivill X. 2005. Contribution of NTRK2 to the genetic susceptibility to anorexia nervosa, harm avoidance and minimum body mass index. Mol Psychiatry 10:851-860.

Soontornniyomkij B, Everall IP, Chana G, Tsuang MT, Achim CL, Soontornniyomkij V. 2011. Tyrosine kinase B protein expression is reduced in the cerebellum of patients with bipolar disorder. J Affect Disord 133:646-654.

Sweetser DA, Peniket AJ, Haaland C, Blomberg AA, Zhang Y, Zaidi ST, Dayyani F, Zhao Z, Heerema NA, Boultwood J, Dewald GW, Paietta E, Slovak ML, Willman CL, Wainscoat JS, Bernstein ID, Daly SB. 2005. Delineation of the minimal commonly deleted segment and identification of candidate tumor-suppressor genes in del(9q) acute myeloid leukemia. Genes Chromosomes Cancer 44:279-291.

Thyagarajan A, Szaro BG. 2008. Dynamic endogenous association of neurofilament mRNAs with K-homology domain ribonucleoproteins in developing cerebral cortex. Brain Res 1189:33-42.

Vogel P, Read RW, Hansen GM, Payne BJ, Small D, Sands AT, Zambrowicz BP. 2012. Congenital hydrocephalus in genetically engineered mice. Vet Pathol 49:166-181.

Yeo GS, Connie Hung CC, Rochford J, Keogh J, Gray J, Sivaramakrishnan S, O'Rahilly S, Farooqi IS. 2004. A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. Nat Neurosci 7:1187-1189.

Yin J, Sobeck A, Xu C, Meetei AR, Hoatlin M, Li L, Wang W. 2005. BLAP75, an essential component of Bloom's syndrome protein complexes that maintain genome integrity. EMBO J 24:1465-1476.

Yuasa K, Omori K, Yanaka N. 2000. Binding and phosphorylation of a novel male germ cell-specific cGMP-dependent protein kinase-anchoring protein by cGMP-dependent protein kinase Ialpha. J Biol Chem 275:4897-4905.

Zhao X, Onteru SK, Dittmer KE, Parton K, Blair HT, Rothschild MF, Garrick DJ. 2012. A missense mutation in AGTPBP1 was identified in sheep with a lower motor neuron disease. Heredity (Edinb) 109:156-162.

Publikace 8

***BCL11A* deletions result in fetal hemoglobin persistence and neurodevelopmental alterations.**

Basak A, Hancarova M, Ulirsch JC, Balci TB, Trkova M, Pelisek M, Vlckova M, Muzikova K, Cermak J, Trka J, Dyment DA, Orkin SH, Daly MJ, Sedlacek Z, Sankaran VG.

J Clin Invest. 2015 Jun;125(6):2363-8. IF: 13,215

Z našeho pracoviště vycházející předchozí publikace pacientky s mikrodelečním syndromem 2p15-p16.1 (Hancarova et al., 2013) a poukazem na gen *BCL11A* jako pravděpodobně klíčový pro tento syndrom iniciovala mezinárodní spolupráci, ve které jsme u dvou českých a jednoho kanadského pacienta s delecemi této oblasti prokázali perzistenci fetálního hemoglobinu (HbF) u jedinců s poloviční dávkou genu *BCL11A*. Práce je zaměřena i na možnost využít gen *BCL11A* jako potenciální cíl v léčbě srpkovité anemie a beta-thalasemií, a podrobně popisuje klinický obraz pacientů s mikrodelecí *BCL11A*.

Funkční studie na buněčných kulturách a zvířecích modelech ukazují, že *BCL11A* hraje roli v regulaci exprese HbF, avšak skutečný regulační potenciál *BCL11A* v hematopoeze a jeho další fenotypové vlivy *in vivo* u lidí nejsou stále ozřejměny. V rámci studie byli vyšetření tři pacienti se vzácnou mikrodelecí 2p15-p16 zahrnující i gen *BCL11A* a prokázali jsme u nich významně zvýšenou hladinu HbF při jinak normálních parametrech v krevním obraze a imunologických testech. Pacienti však vykazovali komplexní postižení zahrnující vývojové opoždění, MR a autistické rysy. U pacientky, kterou jsme měli možnost sledovat v dlouhodobém časovém horizontu, navíc tyto obtíže vykazovaly určitou progresi.

Práce přináší další důkaz o roli genu *BCL11A* v regulaci exprese HbF *in vivo*. Zároveň je však diskutován vliv *BCL11A* na funkci CNS a neurologické poruchy spojené s jeho haploinsuficiencí, což limituje jeho případné využití v cílené léčbě srpkovité anemie a beta-thalasemie. Pozorování také ilustruje obrovskou pleiotropii účinku variant v genech podmiňujících PAS/MR, která sahá u *BCL11A* od neurovývojových poruch přes hematologické znaky až k nádorům prsu, se všemi důsledky které tato pleiotropie má pro úvahy o cíleném ovlivnění těchto genů při budoucí léčbě PAS/MR a dalších fenotypů.

Downloaded from http://www.jci.org on September 10, 2015. http://dx.doi.org/10.1172/JCI81163

The Journal of Clinical Investigation BRIEF REPORT

BCL11A deletions result in fetal hemoglobin persistence and neurodevelopmental alterations

Anindita Basak,[1,2] Miroslava Hancarova,[3] Jacob C. Ulirsch,[1,2] Tugce B. Balci,[4] Marie Trkova,[5] Michal Pelisek,[6] Marketa Vlckova,[3] Katerina Muzikova,[3] Jaroslav Cermak,[7] Jan Trka,[3] David A. Dyment,[4] Stuart H. Orkin,[1] Mark J. Daly,[2,8] Zdenek Sedlacek,[3] and Vijay G. Sankaran[1,2]

[1]Divison of Hematology/Oncology, Manton Center for Orphan Disease Research, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA. [2]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [3]Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic. [4]Department of Genetics, Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, Ontario, Canada. [5]Gennet, Prague, Czech Republic. [6]Regional Hospital Strakonice, Strakonice, Czech Republic. [7]Institute of Hematology and Blood Transfusion, Prague, Czech Republic. [8]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

A transition from fetal hemoglobin (HbF) to adult hemoglobin (HbA) normally occurs within a few months after birth. Increased production of HbF after this period of infancy ameliorates clinical symptoms of the major disorders of adult β-hemoglobin: β-thalassemia and sickle cell disease. The transcription factor BCL11A silences HbF and has been an attractive therapeutic target for increasing HbF levels; however, it is not clear to what extent BCL11A inhibits HbF production or mediates other developmental functions in humans. Here, we identified and characterized 3 patients with rare microdeletions of 2p15-p16.1 who presented with an autism spectrum disorder and developmental delay. Moreover, these patients all exhibited substantial persistence of HbF but otherwise retained apparently normal hematologic and immunologic function. Of the genes within 2p15-p16.1, only *BCL11A* was commonly deleted in all of the patients. Evaluation of gene expression data sets from developing and adult human brains revealed that *BCL11A* expression patterns are similar to other genes associated with neurodevelopmental disorders. Additionally, common SNPs within the second intron of *BCL11A* are strongly associated with schizophrenia. Together, the study of these rare patients and orthogonal genetic data demonstrates that BCL11A plays a central role in silencing HbF in humans and implicates BCL11A as an important factor for neurodevelopment.

Introduction

The switch from fetal hemoglobin (HbF) to adult hemoglobin (HbA) expression that occurs during the months following birth is of considerable therapeutic interest, since elevated HbF ameliorates the clinical symptoms in β-thalassemia and sickle cell disease (SCD) (1, 2). Genome-wide association and functional follow-up studies in cell and animal models have shown that BCL11A, a multiple zinc-finger–containing transcription factor, is an important silencer of HbF expression (3, 4). This has resulted in a concerted effort to develop targeted approaches to induce HbF by inhibiting BCL11A (1, 2). However, the extent to which BCL11A silences HbF and its other functions in vivo in humans is unknown. BCL11A plays a key dosage-dependent role in the immune system in mouse models (5, 6), and recent studies implicate it as an autism spectrum disorder (ASD) and developmental delay (DD) candidate gene (7, 8).

To address the in vivo role of *BCL11A* in humans, we sought to study patients with small deletions involving this gene. A microdeletion syndrome of the 2p15-p16.1 region has been described in rare patients and consists of a number of features, including an ASD, DD, hypotonia, fine motor dysfunction, and facial dysmorphism (OMIM 612513) (9, 10). Most such deletions are large and involve a number of genes. We identified 3 patients with small de novo deletions that only removed *BCL11A* and 1–2 adjacent genes. Analysis of these patients, along with orthogonal genetic data, allowed us to assess the in vivo role of *BCL11A*. We demonstrated that BCL11A plays a key role in both silencing HbF and in human neurodevelopment.

Results and Discussion

We identified 3 patients with small de novo deletions of the 2p15-p16.1 region that only removed *BCL11A* and 1–2 adjacent genes (Figure 1A). Patient 1 had an approximately 440 kb deletion (chr2: 60,689,727–61,128,229 in hg19 coordinates) (10), Patient 2 had an approximately 1 Mb deletion (chr2: 60,029,857–61,059,383), and Patient 3 had an approximately 875 kb deletion (chr2: 59,958,420–60,834,298). *BCL11A* was the only deleted gene shared in all 3 patients, while *PAPOLG* and *MIR4432* were each deleted in 2 of the 3 patients. *PAPOLG* has been suggested to encode a protein that mediates posttranscriptional 3′ adenylation of specific RNAs, although it does not have a known physiologic role (11). *MIR4432* encodes a microRNA that has been identified from deep RNA sequencing of B lymphocytes (12). We noted that both *BCL11A* and *PAPOLG* were expressed in

Authorship note: Anindita Basak, Miroslava Hancarova, Jacob C. Ulirsch, and Tugce B. Balci contributed equally to this work.
Conflict of interest: The authors have declared that no conflict of interest exists.
Submitted: January 25, 2015; **Accepted:** April 6, 2015.
Reference information: *J Clin Invest.* 2015;125(6):2363–2368. doi:10.1172/JCI81163.

Downloaded from http://www.jci.org on September 10, 2015. http://dx.doi.org/10.1172/JCI81163

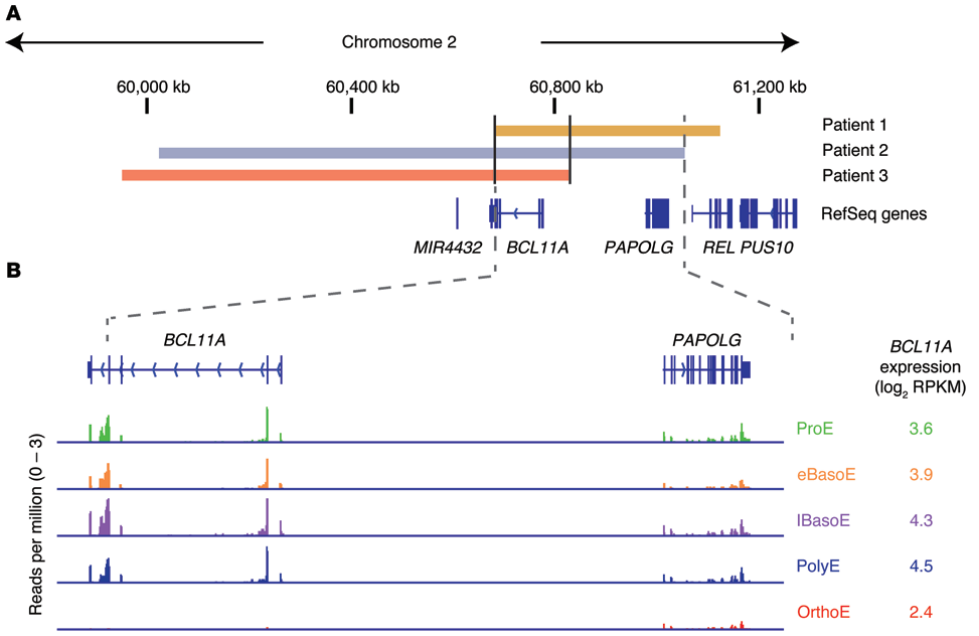


Figure 1. Involvement of *BCL11A* in the 2p microdeletion syndrome. (**A**) A depiction of the 2p15-p16.1 region with coordinates shown (hg19). The position of the patient deletions are shown in orange (Patient 1), blue (Patient 2), and red (Patient 3), and RefSeq genes are shown below. The commonly deleted region of patients 1 and 2 is shown between dotted lines. (**B**) The common deleted region of patients 1 and 2 involving *BCL11A* and *PAPOLG*. RNA expression is shown below at various stages of human erythroid differentiation. This includes proerythroblasts (ProE), early basophilic erythroblasts (eBasoE), late basophilic erythroblasts (lBasoE), polychromatic erythroblasts (PolyE), and orthochromatic erythroblasts (OrthoE). The height of RNA peaks in each region demonstrates the number of reads per million at that site and the reads per kb per million (RPKM) mapped reads for *BCL11A* is shown in $\log_2$ scale.

developing red blood cell (erythroid) precursors from humans, although *BCL11A* was expressed at higher levels than *PAPOLG* (Figure 1B and Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI81163DS1). *MIR4432* was not detectable, and we did not identify any other RNAs that would be removed by these deletions in erythroid cells (Supplemental Figure 1), although we cannot entirely rule out effects on regulatory elements.

Analysis of mononuclear cell RNA from patients 1 and 2 and age-matched controls revealed that *BCL11A* was haploinsufficient in the patients, while *PAPOLG* was not significantly reduced compared with controls (Figure 2A). Concomitantly, we noted that there were higher mRNA levels of the HbF-encoding genes, *HBG1* and *HBG2* (Figure 2A). Consistent with this (though, at lower levels, consistent with known maturational and posttranscriptional regulation) (1, 13), we found that HbF was substantially elevated at 23.8%, 16.1%, and 29.7% in blood from patients 1–3, respectively (Figure 2B), whereas in normal age-matched controls, it would be < 1% (patients 1, 2, and 3 were 14, 6, and 3.5 years, respectively, when this test was done). Since *PAPOLG* was expressed in erythroid cells, we suppressed this gene using shRNAs in primary human erythroid cells and observed no change in the expression of the HbF-encoding genes (Supplemental Figure 2).

Given the variation in HbF levels observed in the patients, we genotyped HbF-associated common variants in the remaining intact *BCL11A* gene locus in all of the patients. Patient 1 had 2 of 3 minor alleles associated with higher HbF levels (alleles A, T, and G at rs4671393, rs1427407, and rs7606173, respectively), while patients 2 and 3 had the reference alleles that are associated with lower HbF levels (alleles G, G, and G at rs4671393, rs1427407, and rs7606173, respectively) (14). This suggests that the variation in HbF levels between patients cannot be fully explained by common genetic variation at the remaining intact *BCL11A* locus. We assessed the loci containing the *HBG1*, *HBG2*, and *HBB* genes and found no deletions or mutations in the patients that would result in elevated HbF. The levels of HbF observed in the microdeletion patients would be sufficient to ameliorate symptoms in patients with β-thalassemia or SCD (1, 15). Importantly, there were no changes in blood counts or other hematologic parameters in the patients (Table 1). Lymphocyte subset levels in patients 1 and 2 and immunoglobulin levels in Patient 2 were tested and were normal (Table 1), suggesting that *BCL11A* haploinsufficiency does not impair immune function in humans, in contrast to its effects in mice (5, 6). None of the patients had a history of severe or unusual infections, supporting the observation that they all appeared to have normal immunologic function.

Downloaded from http://www.jci.org on September 10, 2015. http://dx.doi.org/10.1172/JCI81163

Table 1. Hematologic and immunologic parameters for 2p15-p16.1 deletion patients

Parameter	Patient 1 at 9 years	Patient 1 at 14 years	Patient 2 at 1 year	Patient 2 at 6 years	Patient 3 at 3.5 years	Normal range[A]
wbc count (10^9/l)	6.3	5.1	7.4	7.7	7.5	4.0–12.0
rbc count (10^{12}/l)	4.27	4.13	4.58	4.63	4.73	4.20–5.40
Hemoglobin (g/dl)	13.6	13.5	13.0	14.5	13.1	12.0–15.5
Hematocrit (%)	39.1	38.2	37.5	39.9	36.3	35.0–45.0
MCV (fl)	91.7	92.5	81.9	86.2	76.7	75.0–90.0
MCH (pg)	31.9	32.7	28.4	31.3	27.7	25.0–31.0
MCHC (g/dl)	34.8	35.3	34.7	36.3	36.1	32.0–36.0
RDW (%)	13.1	12.5	14.3	13.2	14.4	11.5–14.5
Platelet Count (10^9/l)	253	166	423	339	427	150–450
Mean Platelet Volume (fl)	7.2	9.1	8.7	9.6	9.1	7.8–11.0
Absolute Lymphocyte Count (10^9/l)	2.898	2.866	5.62	3.966	4.1	1.200–4.200
Absolute Monocyte Count (10^9/l)	0.605	0.515	0.230	0.239	0.500	0.120–1.200
Absolute Neutrophil Count (10^9/L)	2.797	1.719	1.550	3.495	2.800	2.120–6.960
Reticulocyte (%)	–	1.16	–	1.16	1.31	0.50–2.50
Absolute reticulocyte count (10^9/L)	–	47.9	–	53.7	62	–
Reticulocyte cell hemoglobin (pg)	–	35.5	–	35.4	32.7	28.0–36.0
$CD3^+$ lymphocytes (%)	–	74	–	81	–	56–84
$CD4^+$ lymphocytes (%)	–	32	–	37	–	31–52
$CD8^+$ lymphocytes (%)	–	41	–	40	–	18–35
$CD19^+$ lymphocytes (%)	–	6.4	–	13	–	6–23
$CD3^+HLA\text{-}DR^+$ lymphocytes (%)	–	1.1	–	0.8	–	0–4
$CD3\text{-}CD16/56^+$ lymphocytes (%)	–	19	–	5.4	–	3–22
IgG (g/l)	–	–	–	9.43	–	6.37–11.05
IgA (g/l)	–	–	–	2.90	–	0.58–1.16
IgM (g/l)	–	–	–	1.61	–	0.47–1.67

[A]Note that normal ranges do vary at different ages and in different labs. General normal ranges are shown here. –, data not obtained or range unavailable.

All 3 patients exhibited common features, including an ASD, moderate to severe DD, hypotonia, and facial dysmorphism (with common features including an asymmetric face, telecanthus, strabismus, mild ptosis, and long eyelashes). We also noted that there were progressive neurological features in the older patients (1 and 2), including worsening of fine motor activity and coordination, as well as hyperactivity and aggression. Patients 1 and 2 had MRI scans of the brain performed without signs of structural abnormalities, with the exception of microcephaly (both had head circumferences < 3rd percentile for age). Patient 3 had a normal head size (25th percentile) but was noted to have a posterior fossa malformation on MRI. EEGs were also performed on patients 1 and 2 and showed no focal abnormalities in electrical activity.

Recent studies have implicated *BCL11A* as a potential DD and ASD candidate gene (7, 8). We aimed to evaluate whether the DD, ASD, and other features seen in the patients may be attributable to *BCL11A* or *PAPOLG* haploinsufficiency. We found that 4.4% of individuals in a healthy population of 6,503 harbored loss-of-function (LOF) mutations in *PAPOLG*, including 2 individuals with homozygous LOF mutations (Figure 3A and Supplemental Table 1). Given the observed high frequency of LOF variants in a sample of the general population, all of the observed phenotypes, which are rarely observed in the general population, are extremely unlikely to be due to LOF for *PAPOLG*. In contrast, no LOF alleles were found in *BCL11A* in this population (Figure 3A). Furthermore, orthogonal data revealed that it was among the most constrained in the human genome (ranked 106/15877, *P* value LOF = 2.26×10^{-6}) (16). These findings are consistent with the neurologic phenotypes seen in Patient 3 and in a previously described patient (17), who had deletions involving *BCL11A* without disrupting the protein-coding region of *PAPOLG*. We do note that there have been distinct neurologic phenotypes seen in rare patients with 2p15-p16.1 microdeletions that do not disrupt the protein-coding region of *BCL11A* (18), suggesting that these deletions may either disrupt regulatory elements of *BCL11A* or that other genes in the region may also have neurologic functions.

To better delineate a neurodevelopmental role for *BCL11A*, we examined its expression in 524 RNA sequencing data sets from numerous regions of the developing and adult human brain (Figure 3B). *BCL11A* was expressed at high levels and sim-

Downloaded from http://www.jci.org on September 10, 2015. http://dx.doi.org/10.1172/JCI81163

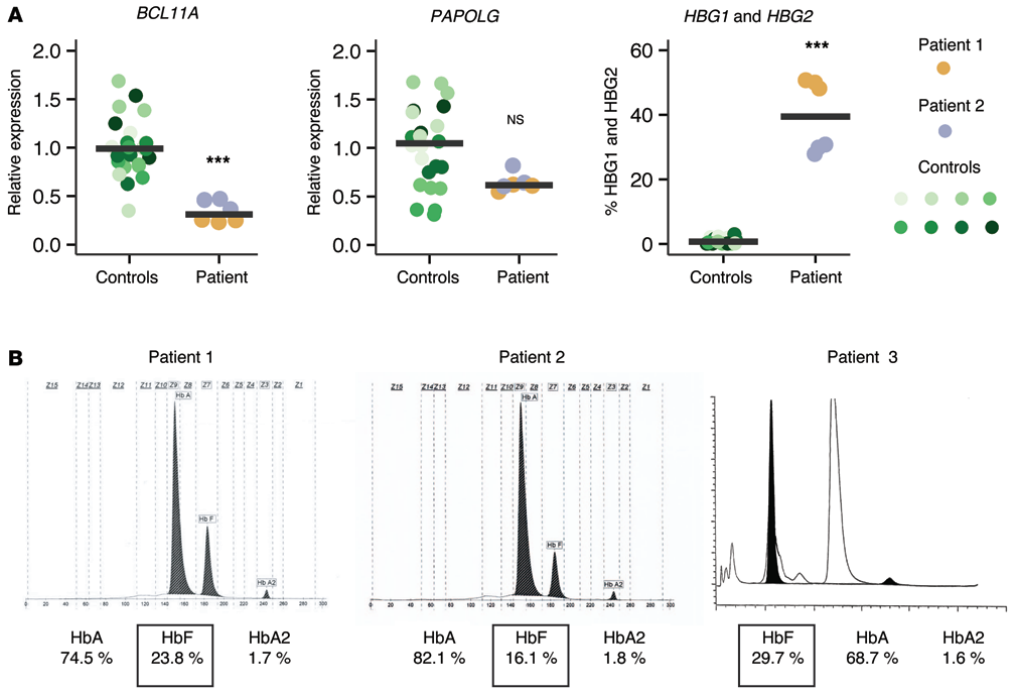


Figure 2. Persistence of HbF with *BCL11A* haploinsufficiency. (**A**) Relative gene expression from qPCR analysis done for *BCL11A* and *PAPOLG*. In addition, the percentage of *HBG1* and *HBG2* are shown. The color-coding of various samples in all the panels is shown on the right, and independent replicates (n = 3 per individual) are plotted individually. ***$P < 0.001$. (**B**) Hemoglobin electrophoresis (patients 1 and 2) and high-performance liquid chromatography (HPLC; Patient 3) chromatograms with the level of different hemoglobin subtypes quantified from peripheral blood samples. The level of hemoglobin A (HbA), HbF, and HbA2 are shown below the chromatograms. In the HPLC chromatogram, the peaks for HbF and HbA2 are filled in, while HbA remains without any filling. The ordering of the labels below the chromatograms is in the order of peak positions. All comparisons were performed using the 2-tailed nonparametric Mann-Whitney U test.

ilar to other ASD/DD candidate genes, including *CHD8* (19) and *DYRK1A* (20), during brain development and in adult brain tissue (Figure 3B and Supplemental Figures 3 and 4). We noted that *PAPOLG* was expressed at lower levels and *KLF1*, which is mutated in cases of persistent HbF without neurologic phenotypes (1), was not expressed in the human brain, illustrating the specificity of this expression data (Figure 3B). *BCL11A* has been suggested to have a role in neurogenesis in model systems, although the consequences of this have not been fully characterized (21). Our results from the rare patients with 2p15-p16.1 microdeletions and orthogonal data strongly implicate *BCL11A* as a high-confidence candidate gene underlying disorders of altered human neurodevelopment.

Finally, since ASD and DD are known to have connections with other neurodevelopmental disorders, we examined data from a recent study of common genetic variation underlying schizophrenia (22). While not initially identified, upon reanalysis, we noted that there were intronic SNPs in *BCL11A* that were significantly associated with schizophrenia and that were located close to or overlapping the common SNPs associated with HbF levels in humans (Figure 3C and ref. 14). In addition, SNPs in this region have also been implicated in attention deficit/hyperactivity disorder (ADHD), another condition thought to be due to underlying alterations in neurodevelopment (23). Therefore, by studying rare patients with *BCL11A* haploinsufficiency in concert with orthogonal genetic data of human neurodevelopmental disorders, we were able to strongly implicate BCL11A as a key gene whose function is necessary for normal human neurologic function and where alterations underlie a number of neuropsychiatric disorders.

Recent functional studies have raised hope that targeting BCL11A may be highly effective to induce HbF in patients with hemoglobin disorders (2). Our findings from rare patients with 2p15-p16.1 microdeletions demonstrate that haploinsufficiency of *BCL11A* is sufficient to allow persistence of HbF at a high enough level to ameliorate β-thalassemia or SCD. Indeed, the levels of HbF observed in these patients are similar to cases of elevated HbF due to mutations in the β-globin locus itself, which have been shown to result in a benign clinical course when acquired with β-thalassemia or SCD (1, 24). Moreover, we observe no immune

Downloaded from http://www.jci.org on September 10, 2015. http://dx.doi.org/10.1172/JCI81163

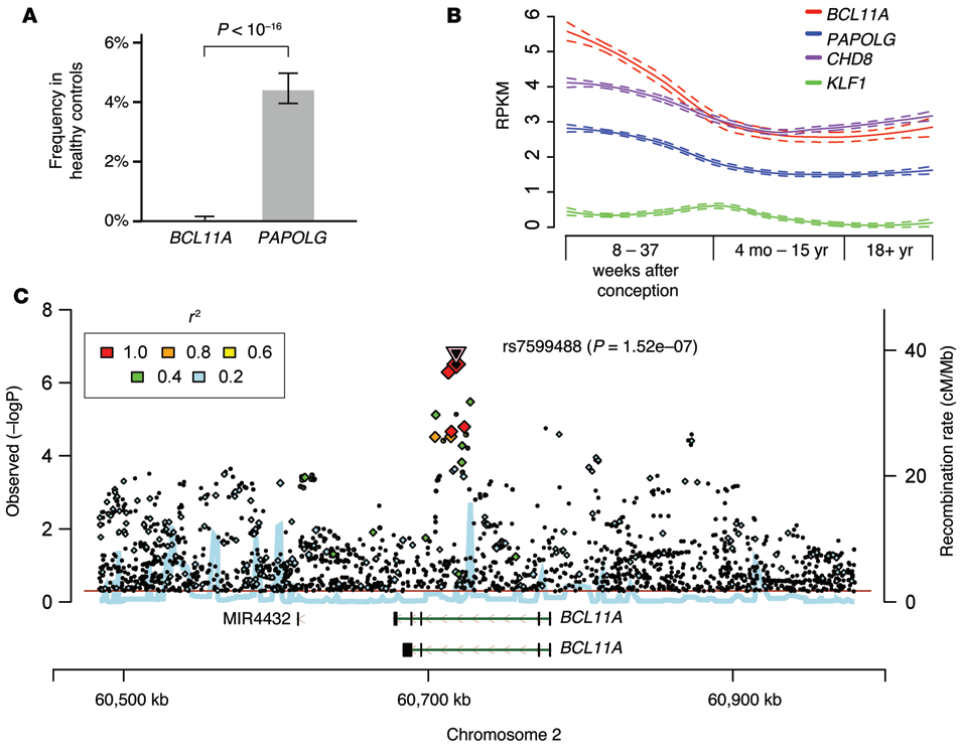


Figure 3. Role of *BCL11A* in human neurodevelopment. (**A**) The percentage of LOF alleles in *BCL11A* and *PAPOLG* from the 6,503 individuals in the Exome Sequencing Project. Error bars represent 95% confidence intervals around the percentage of individuals with one or more LOF alleles. Comparison of these frequencies is performed by Fisher's exact test. (**B**) The expression of select genes in brain tissue at different developmental stages (from various brain regions that are aggregated here for simplicity). Data are plotted as the number of reads per kb per million (RPKM) for the genes shown. A locally weighted scatterplot smoothing regression was applied to expression of each gene. Results from this regression are plotted with 95% confidence intervals. (**C**) Regional association plot depicting data analyzed from a recent schizophrenia GWAS (22).

dysfunction in these patients, in contrast to haploinsufficient mouse models (5, 6). By studying these rare patients in concert with orthogonal genetic data from neurodevelopmental disorders — including ASD, DD, schizophrenia, and ADHD — we are able to strongly implicate *BCL11A* as a key neurodevelopmental gene. This finding emphasizes the importance of using hematopoietic-specific or CNS-nonpenetrating approaches when attempting to target BCL11A therapeutically (2).

Methods

Further information can be found in Supplemental Methods.

Statistics. All pairwise comparisons were performed using the 2-tailed nonparametric Mann-Whitney *U* test unless otherwise stated in the text. Differences were considered significant if the *P* value was less than 0.05.

Study approval. All family members had provided written informed consent to participate in this study. The institutional review boards at Boston Children's Hospital, Charles University in Prague, and University of Ottawa approved the study protocols.

Acknowledgments

We are grateful to the patients and their families for their willingness to participate in this study and to D. Nathan, A. Chakravarti, E. Benz, and C. Walsh for their advice and input. This article is dedicated to the memory of the late Professor Bill Wood.

This work was supported by grants from the NIH (U01 HL117720, R21 HL120791, and R01 DK103794) (to V.G. Sankaran) and from the Czech Ministry of Health (NT/14200 and 00064203) (to Z. Sedlacek). A. Basak is a Translational Research Development Scholar at Boston Children's Hospital.

Address correspondence to: Vijay G. Sankaran, Boston Children's Hospital, 3 Blackfan Circle, CLS 03001, Boston, Massachusetts 02115, USA. Phone: 617.919.6270; E-mail: sankaran@broadinstitute.org. Or to: Zdenek Sedlacek, Charles University 2nd Faculty of Medicine and University Hospital Motol, Plzenska 130/221, 15000 Prague 5, Czech Republic. Phone: 420.257296153; E-mail: zdenek.sedlacek@lfmotol.cuni.cz.

Downloaded from http://www.jci.org on September 10, 2015. http://dx.doi.org/10.1172/JCI81163

1. Sankaran VG, Orkin SH. The switch from fetal to adult hemoglobin. *Cold Spring Harb Perspect Med*. 2013;3(1):a011643.
2. Sankaran VG, Weiss MJ. Anemia: progress in molecular mechanisms and therapies. *Nat Med*. 2015;21(3):221–230.
3. Sankaran VG, et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science*. 2008;322(5909):1839–1842.
4. Sankaran VG, et al. Developmental and species-divergent globin switching are driven by BCL11A. *Nature*. 2009;460(7259):1093–1097.
5. Yu Y, et al. Bcl11a is essential for lymphoid development and negatively regulates p53. *J Exp Med*. 2012;209(13):2467–2483.
6. Ippolito GC, et al. Dendritic cell fate is determined by BCL11A. *Proc Natl Acad Sci U S A*. 2014;111(11):E998–E1006.
7. Coe BP, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet*. 2014;46(10):1063–1071.
8. De Rubeis S, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*. 2014;515(7526):209–215.
9. Rajcan-Separovic E, et al. Clinical and molecular cytogenetic characterisation of a newly recognised microdeletion syndrome involving 2p15-16.1. *J Med Genet*. 2007;44(4):269–276.
10. Hancarova M, Simandlova M, Drabova J, Mannik K, Kurg A, Sedlacek Z. A patient with de novo 0.45 Mb deletion of 2p16.1: the role of BCL11A, PAPOLG, REL, and FLJ16341 in the 2p15-p16.1 microdeletion syndrome. *Am J Med Genet A*. 2013;161A(4):865–870.
11. Kyriakopoulou CB, Nordvarg H, Virtanen A. A novel nuclear human poly(A) polymerase (PAP), PAP γ. *J Biol Chem*. 2001;276(36):33504–33511.
12. Jima DD, et al. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. *Blood*. 2010;116(23):e118–127.
13. Stamatoyannopoulos G. Control of globin gene expression during development and erythroid differentiation. *Exp Hematol*. 2005;33(3):259–271.
14. Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nat Genet*. 2010;42(12):1049–1051.
15. Musallam KM, Sankaran VG, Cappellini MD, Duca L, Nathan DG, Taher AT. Fetal hemoglobin levels and morbidity in untransfused patients with beta-thalassemia intermedia. *Blood*. 2012;119(2):364–367.
16. Samocha KE, et al. A framework for the interpretation of de novo mutation in human disease. *Nat Genet*. 2014;46(9):944–950.
17. Peter B, Matsushita M, Oda K, Raskind W. De novo microdeletion of BCL11A is associated with severe speech sound disorder. *Am J Med Genet A*. 2014;164(8):2091–2096.
18. Chabchoub E, Vermeesch JR, de Ravel T, de Cock P, Fryns JP. The facial dysmorphy in the newly recognised microdeletion 2p15-p16.1 refined to a 570 kb region in 2p15. *J Med Genet*. 2008;45(3):189–192.
19. Bernier R, et al. Disruptive CHD8 mutations define a subtype of autism early in development. *Cell*. 2014;158(2):263–276.
20. Willsey AJ, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. 2013;155(5):997–1007.
21. John A, et al. Bcl11a is required for neuronal morphogenesis and sensory circuit formation in dorsal spinal cord development. *Development*. 2012;139(10):1831–1841.
22. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421–427.
23. Hinney A, et al. Genome-wide association study in German patients with attention deficit/hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2011;156B(8):888–897.
24. Sankaran VG, et al. A functional element necessary for fetal hemoglobin silencing. *N Engl J Med*. 2011;365(9):807–814.

Supplementary Information for:

Fetal hemoglobin persistence and neurodevelopmental alterations due to *BCL11A* deletions

Anindita Basak[§], Miroslava Hancarova[§], Jacob C. Ulirsch[§], Tugce B. Balci[§], Marie Trkova, Michal Pelisek, Marketa Vlckova, Katerina Muzikova, Jaroslav Cermak, Jan Trka, David A. Dyment, Stuart H. Orkin, Mark J. Daly, Zdenek Sedlacek*, Vijay G. Sankaran*

[§]Contributed equally

*Address correspondence to:
Vijay G. Sankaran, Boston Children's Hospital, 3 Blackfan Circle, CLS 03001, Boston, MA 02115; Email: sankaran@broadinstitute.org
Or to: Zdenek Sedlacek, Charles University 2nd Faculty of Medicine and University Hospital Motol, Plzenska 130/221, 15000 Prague 5, Czech Republic; Email: zdenek.sedlacek@lfmotol.cuni.cz

SUPPLEMENTARY METHODS

Deletion mapping and genetic analysis.

The DNA of the patients was analyzed using microarray comparative genomic hybridization (array CGH) using Human CytoSNP-12 BeadChips (Illumina) according to the manufacturer's protocol. Deletions were confirmed using fluorescence in situ hybridization. The *de novo* inheritance of these deletions was demonstrated by performing similar analysis as above on parental DNA. For analysis of loss-of-function variants in *PAPOLG* and *BCL11A* from the general population, data from the Exome Sequencing Project (http://evs.gs.washington.edu/EVS/) was used and statistics were calculated as discussed for a population of 6,503 total individuals.

Globin and *BCL11A* locus sequencing.

Sanger sequencing of the human *HBG1*/*HBG2*/*HBB* gene locus was carried out to assess for mutations in this locus that could result in elevated HbF levels in the patients (1, 2). The regions where full coverage sequencing was performed included (hg19 coordinates): chr11:5,246,635-5,247,070 and 5,247,668-5,248,372 (*HBB*), chr11: 5,270,550-5,271,727 (*HBG1*), and chr11:5,275,474-5,277,265 (*HBG2*). The following HbF associated SNPs in *BCL11A* were genotyped using Sanger sequencing: rs4671393, rs1427407, and rs7606173.

Hematologic and immunologic phenotyping.

Blood samples were obtained from the patients and were subjected to standard clinical hematologic and immunologic assays. This included a complete blood count with a white blood cell differential analysis, reticulocyte count, lymphocyte subset analysis, and immunoglobulin subtype analysis.

Mononuclear cell isolation and RNA analysis.

Mononuclear cells were obtained from peripheral blood by isolation of the buffy coat using Ficoll-Paque (GE Life Sciences) from both patients and normal healthy control children. Briefly, approximately 5 ml blood, diluted in RPMI was layered on Ficoll-Paque and centrifuged at 1300 rpm for 30 min. The buffy coat containing the peripheral blood mononuclear cells (PBMC) was collected and washed in PBS. The PBMC pellet obtained was subject to total RNA isolation using RNeasy Plus Mini Kit (Qiagen). Genomic DNA was eliminated by RNase-free DNase I digestion (Qiagen) during the RNA isolation procedure. Isolated total RNA was quantified on a Nanodrop 2000C instrument (Thermo Scientific). First strand cDNA synthesis and reverse transcription was carried out with the iScript cDNA synthesis Kit (BioRad) in a total volume of 20 µl according to manufacturer's instructions. Gene expression was quantified by quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) using a 96-well plate on a CFX96 Real Time System (BioRad) with iQ SYBR Green Supermix (BioRad) as previously described (3). Primers for qRT-PCR were: *BCL11A* (exon 3 - 4) - 5'-GCCTGGGATGAGTGCAGAAT-3' and 5'- ATGCACTGGTGAATGGCTGT-3'; *PAPOLG* - 5'-CACCACTACCTTCCTGCAGA-3' and 5'-GGATTGAAGTCCGCCCGAG-3'; *GAPDH* - 5'-TGCACCACCAACTGCTTAGC-3' and 5'-GGCATGGACTGTGGTCATGAG-3'; *HBB* - 5'-CTGAGGAGAAGTCTGCCGTTA-3' and 5'-AGCATCAGGAGTGGACAGAT-3'; *HBG1*/*HBG2* - 5'-TGGATGATCTCAAGGGCAC-3' and 5'-TCAGTGGTATCTGGAGGACA-3'.

Cell culture and lentiviral transduction

293T cells were maintained in 2 mM L-glutamine containing DMEM supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (P/S). For lentivirus

production, 293T cells were transfected with the shRNAs in the pLKO.1 construct with the pVSVG and pDelta8.9 vectors using FuGene 6 reagent (Promega), as we have described previously (3, 4). G-CSF mobilized CD34+ cells of peripheral blood from donors were obtained by magnetic sorting and frozen after isolation.

Subsequent to thawing, CD34+ cells were resuspended in primary cell culture medium and differentiated to the erythroid lineage, using a previously described culture protocol (4, 5). From day 0 – 7, cells were cultured at a density of 10^5 - 10^6 cells per milliliter in IMDM supplemented with 2% human AB plasma, 3% human AB serum, 1% P/S, 3 IU/mL heparin, 10 ug/mL insulin, 200ug/mL holo-transferrin, 1 IU erythropoietin (EPO), 10 ng/mL stem cell factor (SCF), and 1 ng/mL IL-3. From day 7 on, IL-3 was omitted from the medium. Lentiviral infection occurred on day 2 with subsequent puromycin selection and further differentiation, as previously described (3, 4). For RNA analysis, cells were harvested on day 9 of culture when they were at the basophilic to polychromatophilic erythroblast stages of differentiation.

Transcriptome analysis.

Analysis of RNA sequencing (RNA-seq) data from differentiating human erythroid cells was obtained from the Gene Expression Omnibus accession GSE53983 and analyzed using the Tuxedo suite of tools (6, 7). Single-end reads were aligned to the human genome (build hg19) and transcriptome, both obtained from UCSC genome browser, using Tophat version 2.0.10 and allowing for novel junctions. For representation in a genome browser, aligned replicates were combined and normalized to reads per million. RNA-seq data from developing and adult post-mortem human brain regions was obtained and analyzed, as described (8, 9). This included a total of 524 RNA-seq data

samples that were combined according to age of specimens, and all brain regions were averaged for each age or analyzed separately.

Schizophrenia genome-wide association study (GWAS) reanalysis

We obtained data from a large meta-analysis of 36,989 schizophrenia cases and 113,075 controls (10). In the original analysis, 128 linkage-disequilibrium-independent SNPs exceeded genome-wide significance ($P < 5X10^{-8}$). To examine whether there may be SNPs that reach a threshold of significance that is highly significant on a genome-wide scale, but not meeting the conservative threshold set in the initial report, we examined SNPs that reached a threshold of $P < 5X10^{-7}$. In this subsequent analysis, we noted that there was a significant association peak in the intron of *BCL11A* that is reported here. The schizophrenia GWAS data was analyzed and depicted with the assistance of the Ricopili tool (http://www.broadinstitute.org/mpg/ricopili/).

SUPPLEMENTARY REFERENCES

1. Sankaran, V.G., Ghazvinian, R., Do, R., Thiru, P., Vergilio, J.A., Beggs, A.H., Sieff, C.A., Orkin, S.H., Nathan, D.G., Lander, E.S., et al. 2012. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *The Journal of clinical investigation* 122:2439-2443.

2. Sankaran, V.G., Xu, J., Byron, R., Greisman, H.A., Fisher, C., Weatherall, D.J., Sabath, D.E., Groudine, M., Orkin, S.H., Premawardhena, A., et al. 2011. A functional element necessary for fetal hemoglobin silencing. *The New England journal of medicine* 365:807-814.

3. Sankaran, V.G., Menne, T.F., Scepanovic, D., Vergilio, J.A., Ji, P., Kim, J., Thiru, P., Orkin, S.H., Lander, E.S., and Lodish, H.F. 2011. MicroRNA-15a and -16-1 act via MYB to elevate fetal hemoglobin expression in human trisomy 13. *Proceedings of the National Academy of Sciences of the United States of America* 108:1519-1524.

4. Ludwig, L.S., Gazda, H.T., Eng, J.C., Eichhorn, S.W., Thiru, P., Ghazvinian, R., George, T.I., Gotlib, J.R., Beggs, A.H., Sieff, C.A., et al. 2014. Altered translation of GATA1 in Diamond-Blackfan anemia. *Nature medicine* 20:748-753.

5. Hu, J., Liu, J., Xue, F., Halverson, G., Reid, M., Guo, A., Chen, L., Raza, A., Galili, N., Jaffray, J., et al. 2013. Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood* 121:3246-3253.

6. An, X., Schulz, V.P., Li, J., Wu, K., Liu, J., Xue, F., Hu, J., Mohandas, N., and Gallagher, P.G. 2014. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood* 123:3466-3477.

7. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* 7:562-578.

8. Shen, E.H., Overly, C.C., and Jones, A.R. 2012. The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain. *Trends in neurosciences* 35:711-714.

9. Bernier, R., Golzio, C., Xiong, B., Stessman, H.A., Coe, B.P., Penn, O., Witherspoon, K., Gerdts, J., Baker, C., Vulto-van Silfhout, A.T., et al. 2014. Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* 158:263-276.

10. 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421-427.

SUPPLEMENTARY TABLE

Supplementary Table 1. Loss of Function Variants in *PAPOLG*

hg19 Position	Prevalence[¶]	Transcript	Function	cDNA alteration	Protein Alteration
chr2:60995628	A1A1=2, A1R=285, RR=5973	NM_022894.3	frameshift	c.271_283del13	p.(T91Kfs*24)
chr2:60995951	A1A1=0, A1R=1, RR=6254	NM_022894.3	frameshift	c.365_366del2	p.(H122Rfs*5)
chr2:61021851	TT=0, TC=1, CC=6494	NM_022894.3	stop gained	c.1996C>T	p.(R666*)

[¶] R corresponds to the wild type and A1 corresponds to the loss of function allele

SUPPLEMENTARY FIGURE LEGENDS

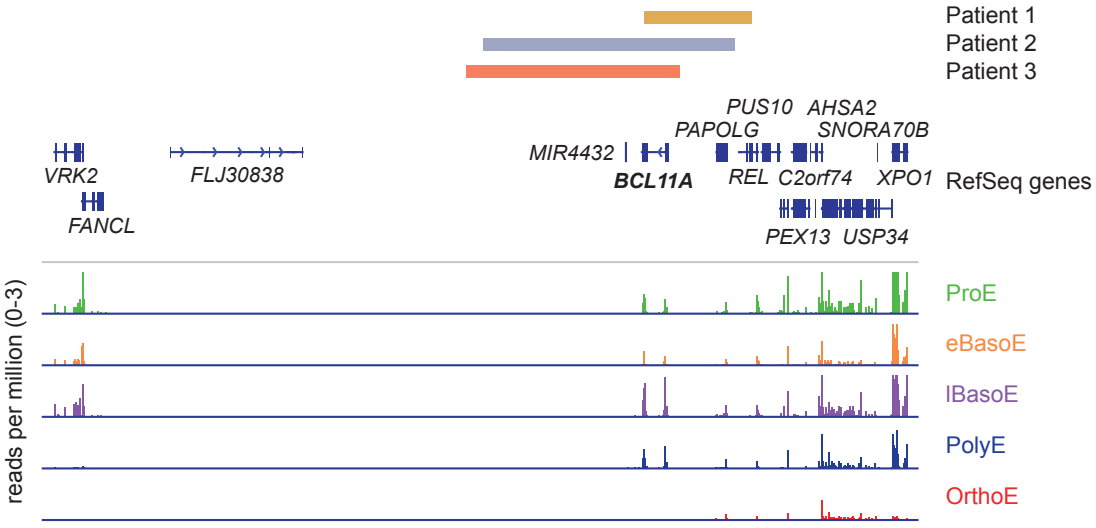
Supplementary Figure 1. Erythroid RNA Expression in the Region of the 2p Microdeletion Syndrome. A depiction of the 2p15-p16.1 region. The position of the patient deletions are shown in orange (Patient 1), blue (Patient 2), and vermillion (Patient 3), and RefSeq genes are shown below. RNA expression is shown below at various stages of human erythroid differentiation. This includes proerythroblasts (ProE), early basophilic erythroblasts (eBasoE), late basophilic erythroblasts (lBasoE), polychromatic erythroblasts (PolyE), and orthochromatic erythroblasts (OrthoE). The height of RNA peaks in each region demonstrates the number of reads per million at that site. This is a zoomed out view of what is shown in Figure 1.

Supplementary Figure 2. *PAPOLG* Does Not Affect Fetal Hemoglobin Levels. (**A**) The relative expression of *PAPOLG* mRNA in primary human erythroid cells treated with control or *PAPOLG* shRNA containing lentiviruses is shown at day 9 of differentiation. The shRNAs targeting *PAPOLG* (sh1 – 4) are compared with the shRNA targeting luciferase (shLuc) using a Mann-Whitney U test with the p-value as shown. (**B**) The percentage of *HBG1* and *HBG2* in cells with shRNAs targeting controls or *PAPOLG* as in the prior panel shown from day 12 of differentiation (percentages were calculated by measuring *HBG1*/*HBG2* and *HBB* expression and then dividing *HBG1*/*HBG2* by the sum of *HBG1*/*HBG2* and *HBB*). Comparisons are performed using the Mann-Whitney U test as in the prior panel.

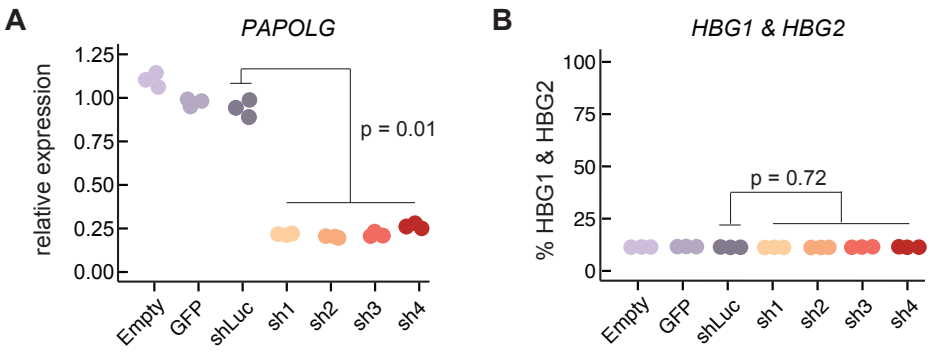
Supplementary Figure 3. Brain RNA Expression of Known Autism Genes. The expression in brain tissue at different developmental stages (from various brain regions that are aggregated here for simplicity) is presented and demonstrates the number of reads per kilobase per million for the genes shown. A locally weighted scatterplot smoothing regression was applied to expression of each gene where each age is an independent and equal time point. Results from this regression are plotted with 95% confidence intervals. All high confidence genes implicated in autism from the Simons Foundation Autism Research Institute are plotted.

Supplementary Figure 4. Spatial Brain Gene Expression During Development. Fine spatial gene expression for *CHD8*, *DYRK1A*, and *BCL11A*, are shown at 15 pcw and 16 pcw. *CHD8* has been previously shown to localize to the intermediate zone, but enrichment was not observed in this region at either time point for *DYRK1A* or *BCL11A*. For simplicity, regions are organized first by lobe and grouped similar to previous investigations of *CHD8*. Abbreviations are as follows: frontal, f; parietal, p; temporal, t; occipital, o; supragenicular nucleus of the thalamus, SG; marginal zone, MZ; outer cortical plate, CPo; inner cortical plate, CPi; SP, subplate zone; intermediate zone, IZ; outer subventricular zone, SZo; inner subventricular zone, SZi, ventricular zone, VZ; frontal polar cortex, fp; dorsolateral prefrontal cortex, dl; dorsomedial frontal cortex, dm-f; ventrolateral prefrontal cortex, vl; orbital frontal cortex, or; posterior frontal cortex (motor cortex), m1; primary somatosensory cortex, s1; dorsomedial parietal cortex, dm-p; posterosuperior (dorsal) parietal cortex, pd; posteroinferior (ventral) parietal cortex, pv; medial temporal-occipital cortex, mt; lateral temporal-occipital cortex, lt; superolateral temporal cortex, sl; inferolateral temporal cortex, il; posterior parahippocampal cortex, ph; midinferior temporal cortex (area 36), t36; caudal midinferior temporal cortex (area TF), tf; midlateral temporal cortex, mlt; primary visual cortex, v1; dorsomedial extrastriate cortex, dm-o; ventromedial extrastriate cortex, vm; midlateral extrastriate cortex, mle.

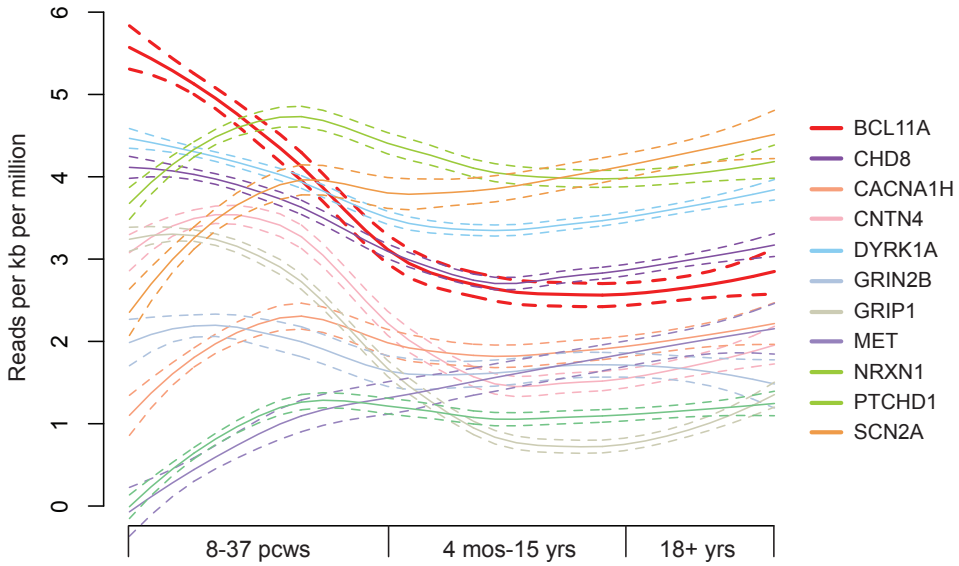
Supplementary Figure 1



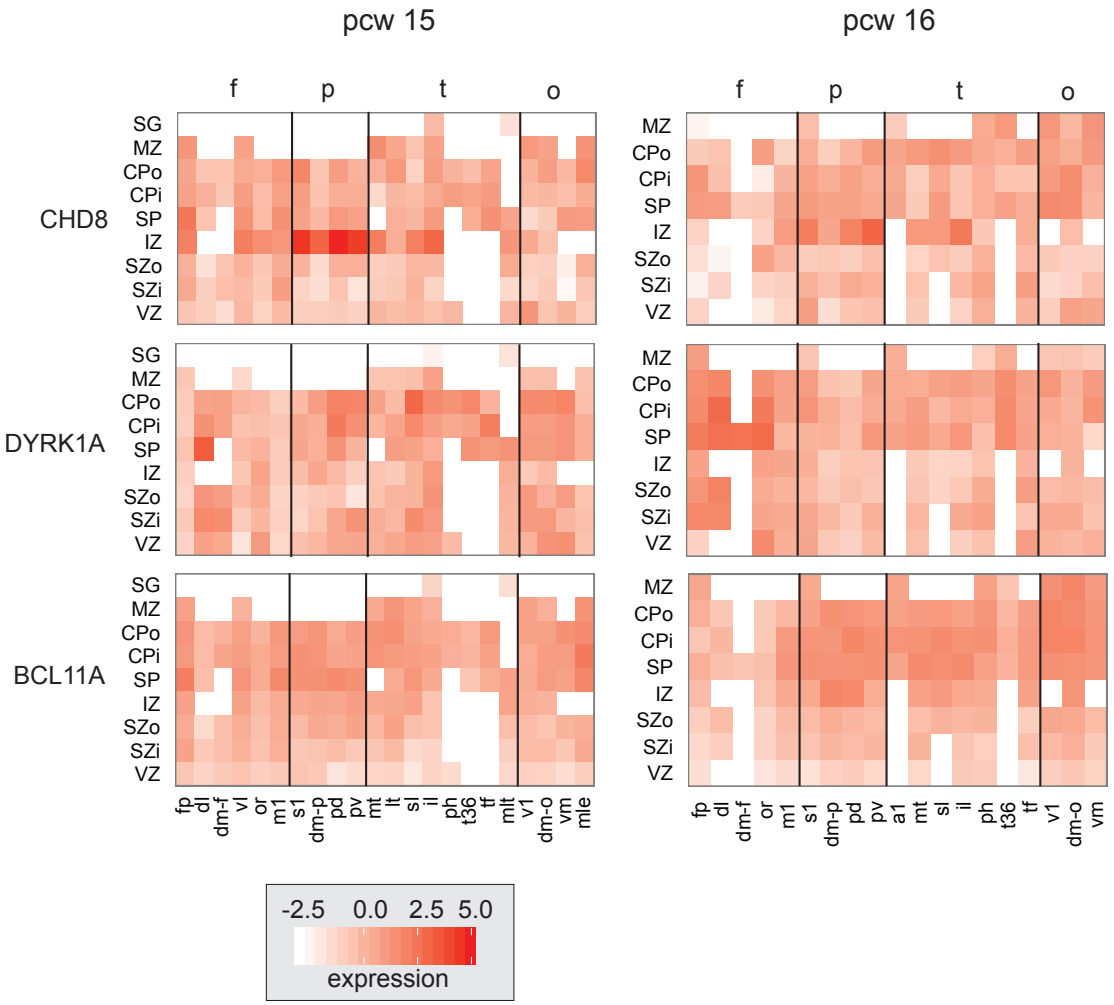
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Publikace 9

***HCFC1* loss-of-function mutations disrupt neuronal and neural progenitor cells of the developing brain.**

Jolly LA, Nguyen LS, Domingo D, Sun Y, Barry S, Hancarova M, Plevova P, Vlckova M, Havlovicova M, Kalscheuer VM, Graziano C, Pippucci T, Bonora E, Sedlacek Z, Gecz J.

Hum Mol Genet. 2015 Jun 15;24(12):3335-47. IF: 6,393

Další mezinárodní spolupráce se zaměřila na vliv genu *HCFC1* na rozvoj X-vázané MR. Práce navazuje na studii zkoumající vliv regulačních variant a over-exprese *HCFC1* na vývoj CNS. Na rozdíl od předchozí práce je tato publikace zaměřena zejména na vliv částečné či úplné ztráty funkce genu *HCFC1* způsobené mutacemi v kódující oblasti. Publikace vychází jednak z dat získaných *in vitro* a jednak z analýzy několika rodin s X-vázanou MR (XLMR) s prokázanými variantami v genu *HCFC1*. V pokusu byla s pomocí shRNA snížena exprese *HCFC1* v embryonálních nervových buňkách. Následná analýza ukázala, že na rozdíl od nadměrné exprese, ztráta funkce *HCFC1* upřednostňuje proliferaci neurálních progenitorových buněk na úkor diferenciace a podporuje axonální růst post-mitotických neuronů. Jako další důkaz podílu genu *HCFC1* na rozvoji neurologických onemocnění popisujeme dvě nové missense varianty v *HCFC1* nalezené u pacientů s MR. Funkční dopady jedné z nově popsaných variant a tří dříve popsaných variant byly dále zkoumány dalšími funkčními testy. Z výsledků vyplývá, že tři ze čtyř testovaných variant vedou k částečné ztrátě funkce *HCFC1*. Over-exprese *HCFC1* vede k redukci proliferace HEK293T buněk a axonálnímu růstu neuronů. Tento efekt byl částečně zmírněn over-expresí 3 ze 4 testovaných variantních alel. Varianta identifikovaná u českých sourozenců zasahuje jaderný lokalizační signál a vede k retenci mutovaného proteinu v cytoplasmě. Další dvě varianty negativně ovlivňují expresi genu *MMACHC*, který hraje roli v metabolismu kobalaminu.

Tato práce ukazuje pravděpodobné důsledky, které mají missense varianty v genu *HCFC1* na buněčné úrovni, identifikuje pravděpodobné mechanismy postižení způsobené poruchou *HCFC1* ovlivňující správný vývoj mozku v embryonálním období a dále významně podporuje roli *HCFC1* v XLMR.

Human Molecular Genetics, 2015, Vol. 24, No. 12 3335–3347

doi: 10.1093/hmg/ddv083
Advance Access Publication Date: 3 March 2015
Original Article

ORIGINAL ARTICLE

HCFC1 loss-of-function mutations disrupt neuronal and neural progenitor cells of the developing brain

Lachlan A. Jolly[1,2], Lam Son Nguyen[4], Deepti Domingo[3], Ying Sun[1,2], Simon Barry[1], Miroslava Hancarova[5], Pavlina Plevova[6], Marketa Vlckova[5], Marketa Havlovicova[5], Vera M. Kalscheuer[7], Claudio Graziano[8], Tommaso Pippucci[8], Elena Bonora[8], Zdenek Sedlacek[5] and Jozef Gecz[1,2,3,*]

[1]School of Paediatrics and Reproductive Health, [2]Robinson Research Institute and [3]School of Molecular and Biomedical Sciences, University of Adelaide, Adelaide 5000, Australia, [4]INSERM UMR 1163, Laboratory of Molecular and Pathophysiological Bases of Cognitive Disorders, Paris Descartes—Sorbonne Paris Cité University, Imagine Institute, Necker-Enfants Malades Hospital, 75015 Paris, France, [5]Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic, [6]Department of Medical Genetics, University Hospital Ostrava, tr. 17. listopadu 1790, 708 52 Ostrava, Czech Republic, [7]Department of Human Molecular Genetics, Max Planck Institute for Molecular Genetics, Ihnestrasse 73, D-14195 Berlin, Germany and [8]Unit of Medical Genetics, Department of Medical and Surgical Sciences, DIMEC, St.Orsola-Malpighi Hospital, University of Bologna, Bologna 40138, Italy

*To whom correspondence should be addressed. Tel: +61 881616339; Fax: +61 881617342; Email: jozef.gecz@adelaide.edu.au

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

Abstract

Both gain- and loss-of-function mutations have recently implicated *HCFC1* in neurodevelopmental disorders. Here, we extend our previous *HCFC1* over-expression studies by employing short hairpin RNA to reduce the expression of *Hcfc1* in embryonic neural cells. We show that in contrast to over-expression, loss of Hcfc1 favoured proliferation of neural progenitor cells at the expense of differentiation and promoted axonal growth of post-mitotic neurons. To further support the involvement of *HCFC1* in neurological disorders, we report two novel *HCFC1* missense variants found in individuals with intellectual disability (ID). One of these variants, together with three previously reported *HCFC1* missense variants of unknown pathogenicity, were functionally assessed using multiple cell-based assays. We show that three out of the four variants tested result in a partial loss of HCFC1 function. While over-expression of the wild-type *HCFC1* caused reduction in HEK293T cell proliferation and axonal growth of neurons, these effects were alleviated upon over-expression of three of the four *HCFC1* variants tested. One of these partial loss-of-function variants disrupted a nuclear localization sequence and the resulting protein displayed reduced ability to localize to the cell nucleus. The other two variants displayed negative effects on the expression of the HCFC1 target gene *MMACHC*, which is responsible for the metabolism of cobalamin, suggesting that these individuals may also be susceptible to cobalamin deficiency. Together, our work identifies plausible cellular consequences of missense *HCFC1* variants and identifies likely and relevant disease mechanisms that converge on embryonic stages of brain development.

Received: November 10, 2014. **Revised:** February 5, 2015. **Accepted:** March 2, 2015

© The Author 2015. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Introduction

Intellectual disability (ID) affects ~2–3% of the population and is considered one of the most genetically heterogeneous human disorders (1). Contribution of genes on the X chromosome (i.e. X-linked ID; XLID) has been the best characterized with over 100 genes implicated thus far. We have recently reported a non-coding regulatory mutation in an X-linked gene, *HCFC1* (OMIM 3 00 019), as the likely cause of mild non-syndromic ID in the large X-linked family MRX3 (2). HCFC1 is a transcriptional co-regulator with many important functions in cell proliferation and mitochondrial biogenesis (3–6). Recent data suggest that HCFC1 containing transcriptional complexes may regulate more than a quarter of all human promoters (7). The non-coding regulatory mutation in the 5′ untranslated region (UTR) of *HCFC1* abolished a DNA-binding site for the transcriptional regulator YY1, which subsequently caused de-repression of *HCFC1* gene expression in patient cell lines (2). To model the consequences of this change, over-expression of *HCFC1* was shown to alter the behaviour of embryonic neural cells, namely promoting differentiation of neural progenitor cells (NPCs) and reducing the neurite growth of *ex vivo* cultured hippocampal neurons. Furthermore, consistent with these newly identified roles for *HCFC1* and those previously established, transcriptome analysis of patient cell lines highlighted deregulation of genes that were important for embryonic brain growth, transcriptional regulation, and mitochondrial function (2,8). Additionally, three unrelated families containing male individuals with ID and different *HCFC1* missense changes were identified (2). Whether these additional variants were pathogenic, or simply low-frequency benign single-nucleotide polymorphisms (SNPs), was not established. Subsequently, multiple *HCFC1* missense mutations, affecting the N-terminal Kelch domain of the HCFC1 protein, have been identified in a cohort of patients with X-linked Cobalamin type C (CblC)-like disorder, which was named CblX (9,10) (OMIM: 3 09 541). CblC (OMIM: 2 77 400) is a metabolic disorder most frequently caused by mutations in *MMACHC* (OMIM: 6 09 831), which encodes an enzyme required for metabolism of cobalamin (11). HCFC1 was shown to bind to the promoter of *MMACHC* and knockdown of *HCFC1* in cultured cells led to a loss of *MMACHC* expression. Similarly and most importantly, *MMACHC* expression was significantly reduced also in skin fibroblasts of two CblX patients tested (9). Thus, loss of HCFC1 function was demonstrated as the cause of CblX. Both CblC and CblX patients display neurological impairment; however, CblX-related neurological phenotype is much more severe and includes intractable epilepsy, brain malformations and severe cognitive impairment (9). Together, these findings suggest that in addition to over-expression of *HCFC1*, also loss-of-function *HCFC1* mutations cause disruptions to normal brain development and likely involve multiple mechanisms.

In this study, we extend our original investigations of *HCFC1* over-expression by investigating the effect that loss-of-function *HCFC1* mutations have on embryonic brain development. We also report two novel *HCFC1* variants identified in two unrelated patients with ID and collate available clinical findings. Finally, we use multiple cell-based assays to address the functional consequence of these and also previously reported *HCFC1* variants. Our findings shed further light into the cellular mechanisms as well as the clinical consequences of *HCFC1*-related neurological disorders.

Results

Reduction of Hcfc1 stimulates neuronal growth

To model the underlying neural cell pathology of *Hcfc1* loss-of-function mutations, and to compare it with established mechanisms behind gain-of-function mutations, we developed short hairpin RNA (shRNA) lentiviral reagents to knock down *Hcfc1* expression. We identified one shRNA trigger sequence (T2) that targeted the untranslated region (UTR) of *Hcfc1*, that modestly knockdown *Hcfc1* expression by about half when tested in NIH3T3 cells (Supplementary Material, Fig. S1). As *HCFC1* gain-of-function is known to reduce the axonal growth and neurite arbourization of hippocampal neurons, we asked what would be the effect of loss of function in these assays. We isolated post-mitotic hippocampal neurons from embryonic day 18.5 (E18.5) mice embryos and plated them *in vitro*. Following 12 h growth, we transduced the cells with lentiviral particles encoding enhanced green fluorescent protein (EGFP) and either a control shRNA sequence (sequence against luciferase that does not target mammalian genes) or a *Hcfc1* shRNA (T2; $Hcfc1^{shRNA}$). Following transduction, neurons were allowed to grow until 4 days *in vitro*, subsequently fixed and immunofluorescently labelled against the marker proteins MAP2 and TAU1 to identify dendrites and axons, respectively (Fig. 1A). Next we conducted morphometric analysis on *EGFP* expressing (i.e. transduced) neurons. Compared with neurons transduced with control lentiviral particles, neurons transduced with $Hcfc1^{shRNA}$ viral particles displayed a 33% increase in axonal length and similar increases in axonal and dendritic (and hence total) termini number, which reports on the degree of neuronal arborization (Fig. 1B and C). Re-expression of *HCFC1* could rescue this axonal defect that confirms the specificity of the $Hcfc1^{shRNA}$ (see below). Thus, reduction of *Hcfc1* expression also affected hippocampal neuronal growth and displayed the opposite effect to that previously described for *HCFC1* over-expression (2).

Loss of Hcfc1 results in expansion of neural progenitor cells

Promotion of NPC differentiation was also observed when *HCFC1* was over-expressed to model the effect of the gain-of-function *HCFC1* mutation (2). Thus, we investigated whether these cells would also be sensitive to loss of *Hcfc1* function. We isolated NPCs from the E18.5 embryonic dorsal cortex and grew them as non-adherent neurospheres. We transduced the cells with control or $Hcfc1^{shRNA}$ lentiviral particles, and achieved transduction rates of ~70% (data not shown). In the first instance, we created purified transduced cultures using fluorescence-activated cell sorting (FACS) and subjected purified cells to a cell proliferation assay. Cells were plated at an equivalent initial density and allowed to grow for 6 days during which cell growth in the cultures was assayed at Days 0, 3 and 6 (Cell Titre 96AQueous Assay; Promega). When seeded at low (clonal) density, cell growth was linear in both control and $Hcfc1^{shRNA}$ cultures (linear regression of line of best fit: $R^2 = 0.984$ and 0.997, respectively); however, the rate of growth in $Hcfc1^{shRNA}$ cultures was higher (slope of linear fit; control = 0.56, $Hcfc1^{shRNA}$ = 1.12), resulting in a 36 and 65% increase in cell numbers at Days 3 and 6 of culture, respectively (Fig. 2A). Similar results were obtained when cells were cultured at higher initial densities; however, cell growth plateaued at later stages, perhaps due to nutrient growth restriction (Supplementary Material, Fig. S2). We reasoned this increase in cell numbers may result from (i) an inhibition of differentiation even under proliferative culture conditions (i.e. in the presence of the epidermal growth factor; EGF), (ii) reduction in apoptosis or (iii) increased progenitor proliferation. To address these alternatives, we first collected RNA from the purified cultures and applied qRT-PCR to discover the expression levels of several cell type-specific marker genes (Fig. 2B). We saw no statistically significant

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

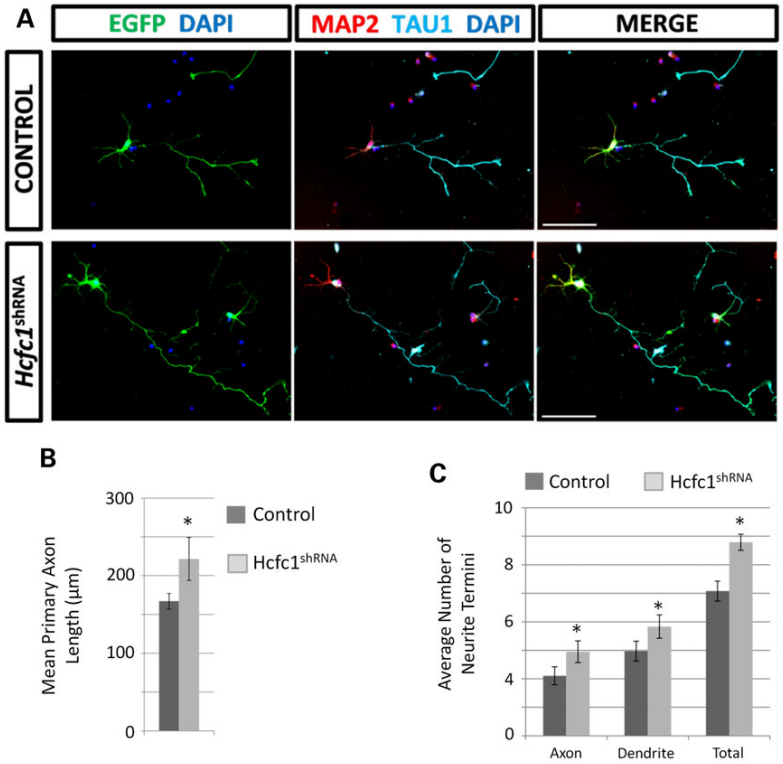


Figure 1. Loss of *Hcfc1* causes enhanced hippocampal neuronal growth *in vitro*. Isolated primary hippocampal neurons were cultured following transduction with lentiviral particles encoding either an inert shRNA sequence (against luciferase; control) or *Hcfc1* shRNA (Hcfc1^{shRNA}). (**A**) Representative immunofluorescent images of control and Hcfc1^{shRNA} neurons at Day 4 of *in vitro* growth. Bars = 100 μm. Transduced cells express *EGFP* (green). Cultures were co-stained using antibodies that label dendritic (MAP2; red) and axonal (TAU1; cyan) neuronal structures. (**B**) Quantification of primary axonal length. (**C**) Quantification of neurite termini. *$P < 0.05$ by Student's *t*-test.

difference in the expression of the early neuroepithelia marker *Sox1*, the cortical NPC marker *Pax6*, the radial glia cell marker *BLBP* nor the pan NPC marker *Nestin*. Likewise, we saw no significant differences in the terminally differentiated cell-type marker genes, namely the early neuronal marker *βIII-tubulin*, the astrocyte marker *GFAP* and the oligodendrocyte marker *CNPase*. While not reaching statistical significance, we did note a trend for elevated expression of marker genes for progenitor cell types and a reciprocal decrease in the expression of terminally differentiated cell-type marker genes in *Hcfc1*^{shRNA} cultures (Fig. 2B). To more definitively identify and quantify the numbers of progenitor cells and monitor their behaviours at single-cell resolution, we employed immunofluorescence. We dissociated neurosphere cells (from parallel non-purified cultures) and plated them onto a poly-l-lysine substrate in the presence of EGF. Only transduced cells (i.e. those expressing EGFP) were analysed. The percentage of transduced cells of NPC identity was reported by the co-expression of *EGFP* and *Pax6*. We identified a significant 20% increase in Pax6-positive NPCs in *Hcfc1*^{shRNA} cultures (Fig. 2C and Supplementary Material, Fig. S3A). We next asked whether apoptotic rates within the cultures were affected by reduction of *Hcfc1*; however, we were unable to identify any differences in the number of transduced cells antigenic to an activated caspase3 antibody, a hallmark of cells undergoing cell death (Fig. 2D and Supplementary Material, Fig. S3B). Next, we reasoned that if more NPCs were present in the *Hcfc1*^{shRNA} cultures, then a higher percentage of transduced cells will be undergoing mitosis at any given time, so we used an antibody against phospho-histone3, a specific nuclear mark of cells at mitosis. Transduced cells in *Hcfc1*^{shRNA} cultures were found to have a 34% increase in cells undergoing mitosis compared with controls (Fig. 2E and Supplementary Material, Fig. S3C). Together the results suggest that under proliferative conditions, transduced NPCs in *Hcfc1*^{shRNA} cultures were more likely to remain in the proliferative progenitor state. We further explored the possibility that transduced cells in *Hcfc1*^{shRNA} cultures might be conducive to NPC maintenance under conditions that promote differentiation. We thus plated control and *Hcfc1*^{shRNA}-transduced cells onto poly-l-lysine substrate as described above but removed growth factor so as to promote differentiation. We allowed the cells to differentiate for 3 days before staining them immunofluorescently using antibodies against cell type-specific marker proteins Pax6, βIII-tubulin, GFAP and CNPase, and counted their abundance within the cultures (Fig. 2F and Supplementary Material, Fig. S4A–C). Under these conditions, transduced cells in the *Hcfc1*^{shRNA} cultures had 25% more Pax6-expressing NPCs, and 34% less differentiated cell types compared with transduced cells in control cultures. This reduction in differentiation was found to be significant across all three neural cell lineages. In aggregate, our data suggest that reduction of *Hcfc1* in NPCs promotes the cell cycling

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

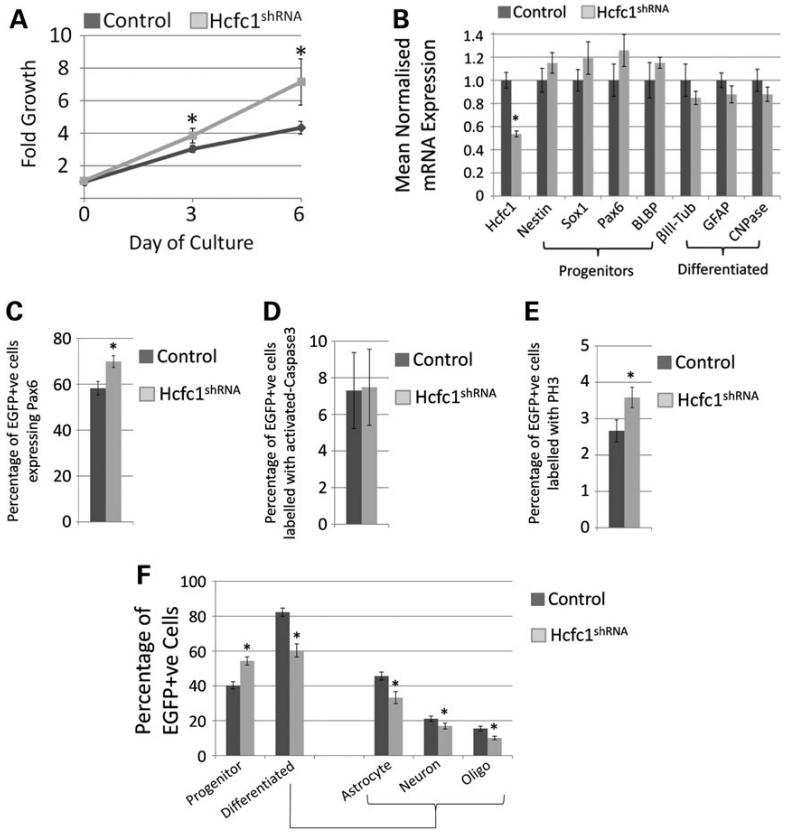


Figure 2. Loss of *Hcfc1* causes increased growth kinetics and reduced differentiation of NPCs. Isolated primary cortical NPCs grown *in vitro* as neurospheres were transduced with lentiviral particles encoding either inert shRNA sequences (Control) or *Hcfc1* targeted shRNA ($Hcfc1^{shRNA}$). (**A** and **B**) Transduced cells were purified using FACS prior to assays. (A) Cell proliferation assay conducted in the presence of EGF at Days 0, 3 and 6 following passage. Cells were plated at an initial low clonal density of 1×10^4 cells. (B) qRT–PCR analysis of marker genes for NPC populations (*Nestin*, *Sox1*, *Pax6* and *BLBP*) and for differentiated cell types [*βIII-tubulin* (βIII-tub), *GFAP* and *CNPase*]. (**C–F**) Non-purified transduced cells were assayed, and all values given as percentage of transduced cells (i.e. *EGFP* expressing cells; EGFP+ve). (C–E) Dissociated cells were plated onto poly-l-lysine substrate and cultured in the presence of EGF. (C) Quantification of percentage of transduced cells expressing *Pax6* by immunofluorescence. (D) Quantification of percentage of transduced cells expressing activated caspase3 by immunofluorescence. (E) Quantification of transduced cells expressing phospho-histone3 (PH3) by immunofluorescence. (F) Cells were plated onto poly-l-lysine substrate and cultured in the absence of EGF. Percentage of cell types present in differentiated cultures identified by immunofluorescence: progenitor cells (Pax6) and differentiated cells (astrocytes: GFAP; neurons: βIII-tubulin and oligodendrocytes: CNPase). *$P < 0.05$ by Student's *t*-test.

and drives their fate choices towards producing daughter NPCs at the expense of differentiated cell types. As in our previous neuronal cell assays, this result is in general opposite to previous results observed when *HCFC1* is over-expressed in these same cell types and suggests that NPCs are sensitive to *HCFC1* dosage.

Identification of novel HCFC1 variants in individuals with ID

We previously described three missense variants in the *HCFC1* coding region (NM_005334.2) in patients with probable or likely XLID using X-exome sequencing (2). A missense change (c.674G>A, p.Ser225Asn) segregated with ID in family D144, associating with ID in four affected males across two generations (Fig. 3A). A second missense variant (c.2626G>A, p.Gly876Ser) was found in a simplex patient (family D147) whose mother had 100% skewing of X chromosome inactivation. However, this variant was also found in 1/1466 chromosomes (or 7/10 271; National Heart, Lung, and Blood Institute [NHLBI] Exome Variant Server), and the patient also had a potentially deleterious variant in the XLID gene *MED12* (NM_ 005120.2; c.3101T>G, p.Phe1034Cys; OMIM: 3 00 188), and a likely deleterious truncating variant in the ID gene *ARID1B* (NM_017519.2; c.2723delC, p.Pro908fs*6; OMIM: 6 14 556) (12) (Fig. 3A). A third *HCFC1* variant (c.5267C>T, p.Ala1756-Val) was found in the index patient of family D82 with four affected males across two generations; however, co-segregation could not be performed as DNA was unavailable on all relevant family members (Fig. 3A). The index patient also had a variant in *ZMYM3* (NM_005096.3; c.356A>G, p.Gln119Arg) of unknown significance. We now report the discovery of two additional novel *HCFC1* variants in patients with probable XLID. The first novel variant was found in a family ascertained in the Czech Republic (family CZE168) that contained two affected brothers. The clinical descriptions of these individuals are detailed in

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

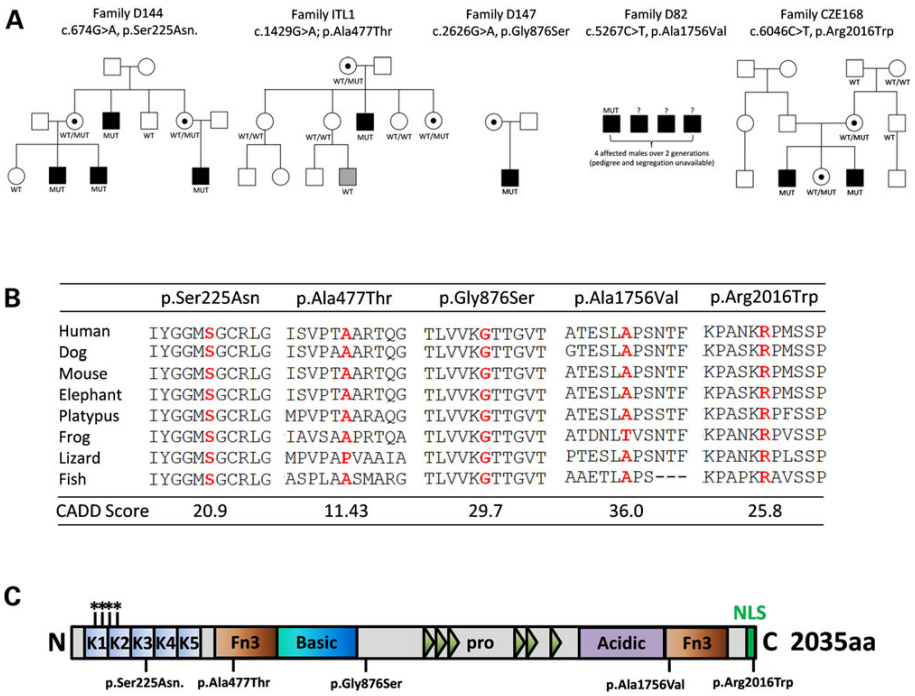


Figure 3. *HCFC1* variants segregate with ID and result in changes to conserved regions of the protein. (**A**) Pedigrees of families identified in (2) (Families D144, D147 and D87) and that of new families (ITL1 and CZE168), with *HCFC1* variants. cDNA (NM_005334.2) and protein (XP_005274721.1) annotation for individual variants are shown. Black fill indicates ID, black dots indicate heterozygote carriers, and grey fill indicates autism (not ID). (**B**) Sequence alignment of variant amino acids across different species (using EBI CLUSTALW server), and *in silico* prediction (CADD) of the effect that the variant amino acid changes have on protein function and integrity. Note that CADD scores >20 are considered likely pathogenic, while scores >10 are considered potentially pathogenic. (**C**) Linear scale cartoon of HCFC1 protein showing known structural features (K1-5: Kelch domains; Fn3: Fibronectin 3 domain; Pro: HCFC1 proteolytic repeats; NLS: nuclear localization sequence) and the locations of variant amino acids. Note that * indicates locations of variants previously reported (9,10).

Supplementary Material, Table S1. Briefly, both affected individuals shared clinical features of mild–moderate ID, delayed speech and delayed psychomotor development. They were both of short stature (3rd percentile) and shared dysmorphic features of thick lips, long nose and a long philtrum. Both displayed mild liver steatosis and diffuse hypocontractibility of the left ventricle. Metabolic profiling also revealed transiently increased plasma homocysteine levels and normal levels of methylmalonic acid in urine. Whole-exome sequencing was performed to find causative variants. No variants were found under an autosomal recessive model, whereas five variants were found under a model of X-linkage (Supplementary Material, Table S2). Of these, the *HCFC1* variant (chrX 153 215 026G>A (HG19); c.6046C>T; pArg2016Trp) was the most likely candidate based on extended segregation analysis and predicted tolerances to variation (Supplementary Material, Table S2). X-inactivation studies on the affected individual's mother showed significant skewing (87%). The *HCFC1* variant abolishes a predicted bi-partite nuclear localization sequence (NLS; amino acids 2011–2034; cNLS Mapper (13) prediction score = 9.6). A second novel *HCFC1* variant was found in a family ascertained in Italy (family ITL1; chrX 153 225 268 C>T (HG19); c.1429G>A; p.Ala477Thr; Fig. 3A and Supplementary Material, Fig. S5). The affected individual has severe ID, absent speech and displayed frequent febrile seizures during infancy. He had an elongated face, large ears, arachnodactyly and a lean body habitus. Metabolic profiling revealed normal methylmalonioc levels in urine. The affected individual's great uncle also suffered from ID (deceased; Supplementary Material, Fig. S5), while the individual's nephew suffered from autism spectrum disorder and did not carry the variant. This *HCFC1* variant disrupts the fibronectin type 3 domain of HCFC1 which is involved in heterodimerization of HCFC1 N- and C-terminal proteolytic cleavage products and DNA binding.

We were able to access material (whole blood) from individuals of the D144, ITL1 and CZE168 families (but not the others) to test for the expression levels of *HCFC1*. Intriguingly, mRNA expression of the variant forms of *HCFC1* in affected individuals was elevated in families ITL1 and CZE168, and modestly elevated in two of three individuals from D144, compared with control samples (Supplementary Material, Fig. S6A–C), which may suggest an autoregulatory mechanism controlling *HCFC1* transcription (see discussion). Western blot analysis of an affected individual in the CZE168 family, however, did not reveal significant increase in HCFC1 protein levels (Supplementary Material, Fig. S6D).

Overall, this new genetic evidence together with previous reports suggests the involvement of the *HCFC1* in the pathology of ID; however, while the genetic evidence for the p.Ser225Asn variant (located within the N-terminal Kelch domain) and p.Arg2016Trp (located in the NLS) is more persuasive, whether the p.Gly876Ser, p.Ala1756Val and p.Ala477Thr variants are likely causative is less clear. All of the variants alter conserved amino acids of HCFC1, and *in silico* predictions (CADD score (14))

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

suggested four of the variants to be likely deleterious to protein function, with the p.Ala477Thr variant predicted to be possibly deleterious (Fig. 3B and C). We thus embarked on functional characterization of these variants and their effect on HCFC1 function.

The p.Arg2016Trp variant of HCFC1 disrupts its nuclear localization

We cloned four of the five *HCFC1* variants into a cDNA construct that permits expression in mammalian cells under a constitutive promoter. Unfortunately, we were unable to obtain clones with the p.Ala477Thr variant. The expression constructs contained in frame Myc and HA tags at the N and C terminus, respectively, to facilitate discrimination between endogenous and exogenous expression (Fig. 4A). When expressed in HEK293T cells, we observed no difference in *HCFC1* variant mRNA or protein expression levels compared with wild-type *HCFC1*, suggesting the variants did not overtly affect exogenously expressed *HCFC1* mRNA or protein stability (Supplementary Material, Fig. S7A and B). We next studied the subcellular localization of wild-type and variant HCFC1 proteins. First, we compared the expression of endogenous HCFC1 with that of exogenously expressed wild-type HCFC1. HCFC1 is known to undergo proteolytic processing in cells and can be found as either a full-length protein, or N- and C-terminal fragments (15,16). These fragments are known to form a heterodimer but can also display independent functions (17–19). We used our anti-Myc and HA antibodies to detect exogenous HCFC1, and our anti-HCFC1 antibody, which binds the C-terminal of HCFC1, to localize exogenous and endogenous HCFC1. In untransfected cells, endogenous HCFC1 was localized predominantly in cell nuclei, with highest expression found in cells undergoing mitosis (Fig. 4B). Exogenous HCFC1 detected with the C-terminal Myc tag displayed similar expression, while the HCFC1 detected with the N-terminal HA tag was enriched in the nucleus but also localized to the cytoplasm (Fig. 4B). All variants except the p.Arg2016Trp variant localized akin to wild type. The p.Arg2016Trp variant detected through both its N-terminal HA tag and C-terminal Myc tag failed

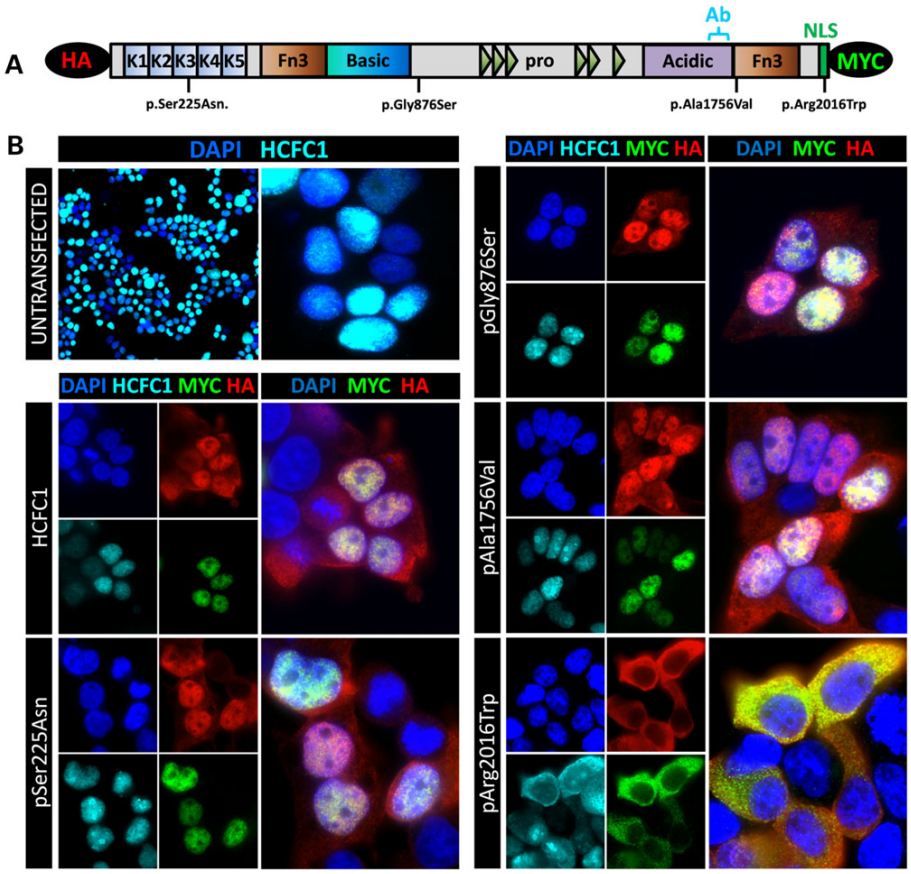


Figure 4. *HCFC1* p.Arg2016Trp variant fails to localize to the nucleus in HEK293T cells. (**A**) Diagram of HCFC1 protein with the locations of the variants being tested. Note also that exogenously expressed HCFC1 proteins are tagged by HA at the N-terminus and Myc at the C-terminus, and also the C-terminal location of the epitope used to generate the HCFC1 antibody (Ab). (**B**) HEK293T cells were transfected with empty control vectors or vectors encoding wild-type or variant forms of *HCFC1*. Immunofluorescent detection of endogenous (and exogenous) HCFC1 (using the anti-HCFC1 antibody; cyan), and exogenous wild-type and variant forms using both anti-HA (red) and anti-Myc (green) antibodies. Cell nuclei are counterstained with DAPI (blue).

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

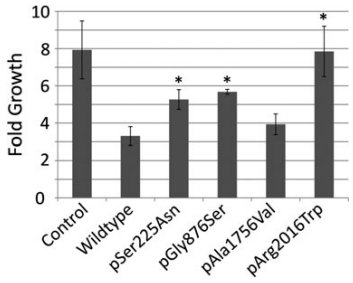


Figure 5. *HCFC1* variants disrupt its ability to suppress cell growth of HEK293T cells. HEK293T cells were transfected with an empty expression vector (control) or with expression vectors encoding either wild-type or variant HCFC1 forms. Transfected cells were plated at equivalent densities and allowed to grow for 2 days at which point cell numbers were assayed. $n = 6$, * is significantly different to wild type, $P < 0.05$ Student's *t*-test.

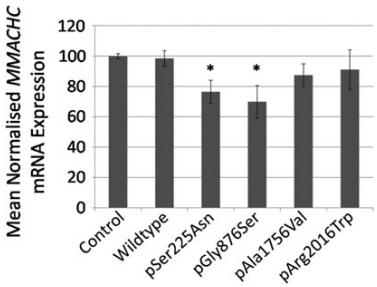


Figure 6. Over-expression of *HCFC1* variants disrupts *MMACHC* mRNA expression. HEK293T cells were transfected with an empty expression vector (control) or with expression vectors encoding either wild-type or variant HCFC1 forms and RNA was isolated. *MMACHC* mRNA expression was analysed by qRT–PCR.

to localize to the nucleus and was found in the cytoplasm. Similar results were obtained using HeLa cells (data not shown) and in hippocampal neurons (Supplementary Material, Fig. S8A and B). These data indicate that the *HCFC1* variant affecting the nuclear localization sequence (p.Arg2016Trp) results in a disruption to its nuclear localization.

HCFC1 variants disrupt proliferation of HEK293T cells

Given previous studies showing the requirement of HCFC1 at multiple stages of the cell cycle (4,6,17–19), we also tested the ability of *HCFC1* and its variant forms to affect the proliferation of HEK293T cells. We plated transfected cells out at an equivalent density and monitored cell growth using the aforementioned cell growth assay. Following 2 days of growth, over-expression of wild-type *HCFC1* caused a 58% decrease in growth compared with control cultures (Fig. 5 and Supplementary Material, Fig. S9). This growth reduction was not significantly different from that resulting from over-expression of p.Ala1756Val (50% decrease). Over-expression of the p.Ser225Asn and p.Gly876Ser mutants both resulted in decreases in growth of 33 and 28%, respectively, which was a modest, but significant loss of growth suppression compared with over-expression of the wild-type *HCFC1*. The over-expression of the p.Arg2016Trp had no effect on growth and was akin to control cells (Fig. 5 and Supplementary Material, Fig. S9). These data suggest that three of four of the tested *HCFC1* variants may cause loss of function to various degrees, with the exception being the p.Ala1756Val variant.

HCFC1 variants disrupt expression of MMACHC

The investigations on loss-of-function *HCFC1* mutations in CblX revealed that *HCFC1* is required for the expression *MMACHC*, which encodes an enzyme involved in the metabolism of cobalamin, and which is also the most common genetic factor found mutated in CblC (9). We thus looked at the ability of the *HCFC1* variants to affect the expression of *MMACHC*. We again utilized our HEK293T cells and over-expressed wild type or variant *HCFC1* and analysed *MMACHC* mRNA expression by qRT–PCR. Over-expression of WT *HCFC1* had no effect on *MMACHC* expression, suggesting that endogenously expressed wild-type HCFC1 binding (and function) is already saturated at the *MMACHC* promoter under normal conditions (Fig. 6). Likewise, expression of the p.Ala1756Val and p.Arg2016Trp had no effect. The other two variants tested resulted in the reduction of *MMACHC* expression. As we were able to achieve large over-expression of wild-type and variant *HCFC1* in these cells (~160-fold as tested by qRT–PCR, and ~10-fold as tested by western blot; Supplementary Material, Fig. S7), the stoichiometry suggests that the exogenously expressed proteins will outcompete endogenous wild-type HCFC1 for binding at the *MMACHC* promoter. Given that HCFC1 loss-of-function mutations are described to result in reduced *MMACHC* expression in patient cells (9), the loss of *MMACHC* expression in this assay can be interpreted likewise, suggesting that the p.Ser225Asn and p.Gly876Ser variants are also likely loss of function.

HCFC1 variants affect neuronal outgrowth

The above data suggest that with the exception of the p.Ala1756-Val variant, the other HCFC1 variants may cause a loss or partial loss of function. To further investigate this in a model system more relevant to patient phenotypes, we assessed the ability of the four variants to effect the axonal growth of hippocampal neurons. As we have previously demonstrated and as shown above, this assay is both sensitive to *Hcfc1* gain and loss of function, which results in either reduction or extension of axonal growth, respectively (2). We assayed the effect that over-expression of the *HCFC1* variants has on axonal growth and their ability to rescue defects associated with loss of *Hcfc1* expression. We isolated primary hippocampal neurons from E18.5 embryos (as described above) and immediately following isolation, we nucleofected them with an expression vectors encoding Red Fluorescent Protein (RFP) together with either an empty expression vector control or expression vectors encoding either wild-type *HCFC1*, or one of the four *HCFC1* variants. Cells were plated out and allowed to attach overnight. The following day, these transfected cultures were transduced with lentiviral particles that encode both *EGFP* and either control or $Hcfc1^{shRNA}$. Cells were grown for an additional 3 days *in vitro*, fixed and immunofluorescently stained for the axonal marker protein TAU1. Axon lengths of dual-transfected and transduced cells were measured (i.e. those expressing both *RFP* and *EGFP*) (Fig. 7A). Over-expression of *HCFC1* in control-transduced cells resulted in a 25% reduction of axonal length, in agreement with our original findings (2). A similar 23% reduction was observed when the p.Ala1756Val variant was over-expressed, while the effect of over-expression of the other three variants resulted in less severe axonal length reductions (ranging only between 4 and 8%) that were significantly different from

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

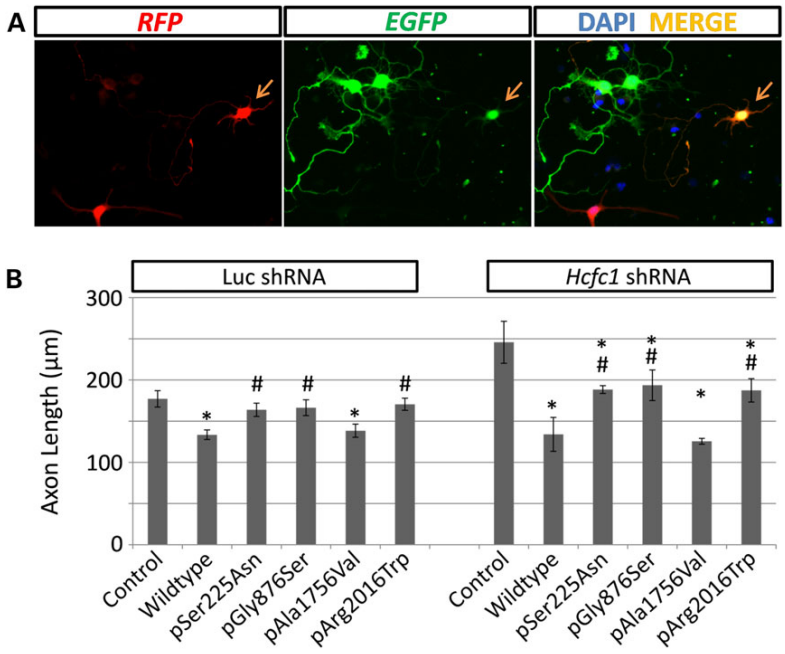


Figure 7. *HCFC1* variants disrupt axon growth. Hippocampal neurons were isolated and transfected with *RFP* expression vector together with either an empty expression vector (control) or vectors encoding either wild-type or variant forms of *HCFC1*. Following cell attachment *in vitro*, neurons were transduced with lentiviral particles that encode *EGFP* and either control or *Hcfc1*^{shRNA}. (**A**) Representative immunofluorescent image from control condition showing identification of a dual transfected (i.e. RFP expressing; red) and transduced (i.e. EGFP expressing; green) neuron (yellow arrow). Nuclei were stained using DAPI (blue). (**B**) Quantification of mean axonal length. $P < 0.05$ by Student's *t*-test comparing values within a shRNA condition (Luc or *Hcfc1*^{shRNA}); *significantly different to control condition; #significantly different to wild-type condition.

HCFC1 wild-type over-expression. Next, we analysed samples transduced with *Hcfc1*^{shRNA} and revealed similar outcomes (Fig. 7B). Transduction with this virus caused a 39% increase in axonal length compared with transduction with a control virus. This elongated axon length was reduced by 45% when *HCFC1* wild-type was re-expressed and a similar 49% reduction observed when the p.Ala1756Trp variant was expressed. Re-expression of the other three *HCFC1* variants caused significant reductions (i.e. compared with control *Hcfc1*^{shRNA} condition) that ranged between 21 and 24%; however, these reductions were significantly lower, compared with re-expression of the wild-type *HCFC1* (Fig. 7B). Together, these results confirm that axonal length is regulated by *HCFC1* dosage and that three of the four *HCFC1* variants tested here (i.e. p.Ser225Asn, p.Gly876Ser and p.Arg2016Trp) lead to partial HCFC1 loss of function.

Discussion

Previous findings suggest that both HCFC1 gain and loss of function mutations can cause neurodevelopmental disorders. Here, we complement our previous studies on *HCFC1* gain of function by reporting the discovery of altered neural cell behaviours resulting from *HCFC1* loss of function. We find, in general opposite to the effects of over-expression, that modest knockdown of *Hcfc1* in NPCs caused over-proliferation and an associated reduction of differentiation, while knockdown in hippocampal neurons promoted axon growth. In support of these findings, brain malformations including microcephaly, macrocephaly and cortical gyration malformation have been reported in CblX patients (9,10) and in individuals from families CZE168 and D147 that we describe here. Thus, embryonic brain growth and neuronal connectivity are processes sensitive to HCFC1 function (and dosage), and alterations to these processes may underlie aspects of the patient neurological phenotypes.

Having established neural-specific effects of loss of *Hcfc1* function, we sought to provide evidence in support of the pathogenicity of additional *HCFC1* variants described previously (2), and newly discovered, including a novel variant affecting the NLS of HCFC1. Three of the four variants we tested resulted in altered HCFC1 function consistent with a partial, but not complete, loss-of-function effect with the exception being the p.Ala1756Val variant. The other three variants (p.Ser225Asn, p.Gly876Ser and p.Arg2016Trp) all showed reduced (but not abolished) potency compared with wild type when assayed for effects on HEK293T proliferation and neuronal axonal growth. A similar effect was observed for the p.Ser225Asn and p.Gly876Ser variants when assayed for an ability to regulate *MMACHC* transcription levels, but not the p.Arg2016Trp variant, which behaved akin to wild type (and p.Ala1756Val variant) in this instance. The inability of the HCFC1 p.Arg2016Trp variant to effect *MMACHC* expression in a negative manner is however consistent with its exclusion from the nucleus, which would mask its potential ability to disrupt transcription carried out by endogenous wild-type HCFC1 in these cells. In fact, by measure of the effect on HEK293T proliferation and localization, the p.Arg2016Trp variant may be considered the most detrimental to protein function of the four

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

variants tested. Furthermore, the two affected brothers with this variant both displayed elevated homocysteine levels, a signature metabolic feature of individuals with CblX (and CblC), and thus suggestive of loss of HCFC1 transcriptional activation of *MMACHC*. Interestingly, *HCFC1* mRNA levels were found to be elevated in whole blood of the affected individuals and also in the individual with the p.Ala477Thr mutation, family ITL1, and in lymphoblastoid cell lines of two of the three individuals with the p.Ser225Asn mutation, family D144. This finding may suggest that HCFC1 normally participates in the repression of its own transcription. Such an autoregulatory mechanism would be in line with the need to tightly regulate dosage of HCFC1 during development and homeostasis. That HCFC1 binds YY1, a known repressor of *HCFC1* transcription, provides a candidate mechanism warranting further investigation, particularly in light of our previous finding wherein mutation of a YY1 binding site in the *HCFC1* 5′UTR was associated with *HCFC1* over-expression and ID (2,20). Although transcription was elevated, initial investigations did not show an increase in the overall protein level of the p.Arg2016Trp variant.

Despite the deleterious effect on HCFC1 function in our cell assays, the contribution of the p.Gly876Ser variant to the pathology of the patient remains complicated by the likely pathogenic contributions of a *MED12* variant (c.3101T>G, p.Phe1034Cys) and a protein truncating *ARID1B* variant (c.2723delC, p.Pro908fs*6) also found in the individual. Mutations in *MED12* cause XLID (21) (OMIM: 3 00 895 and 3 09 520). This particular variant has not been previously reported and is predicted to be potentially pathogenic (CADD score = 14.53); as such its role in this individual's phenotype is not conclusive. In contrast, haploinsufficiency of *ARID1B* invariably causes a broad range of disorders from mild ID to more severe phenotypes of Coffin-Siris syndrome (OMIM: 1 35 900) and can be associated with a spectrum of clinical presentations (12,22). Because of this, it is difficult to disentangle the contributions of *HCFC1*, *MED12* or *ARID1B* variants. Unfortunately, we were unable to obtain information on the affected individual's homocysteine or methylmalonic acid levels, which could potentially identify CblX. The aggregate clinical data available to us are still consistent with CblX, but also severe *ARID1B*-associated phenotypes, and have some features not yet described for *MED12* mutations (e.g. epilepsy), and missing some features normally associated with MED12 mutations (e.g. hypertelorism and constipation). It is credible that *HCFC1*, ARID1B and *MED12* variants may all be contributing to the ID; the possibility of multiple hits with additive or independent effect has been previously documented and might be under-ascertained (23,24). Interestingly, the EVS server reports the p.Gly876Ser variant in 7/10 271 chromosomes (including two males with the change, a male frequency of 1/1193). Furthermore, following completion of our functional studies, the ExAC database became available (http://exac.broadinstitute.org, last accessed February 2015), which also reported the p.Gly876Ser variant at a frequency of 52/1 21 298 chromosomes (of which 17 are likely to be males, giving an approximate male frequency of ~1/2300). These data reveal that the p.Gly876Ser occurs at a very low frequency in the population and suggest that the variant might be benign; however, it is also possible that the males with this variant in these databases may have mild to borderline ID, which aligned with the high incidence (affecting 2–3%) and wide spectrum of ID is plausible, albeit undeterminable at this point in time. Our conservative conclusion based on the functional and genetic data presented here is to regard the p.Gly876Ser variant as a low-frequency variant that provides an increased susceptibility to intellectual disability. This is consistent with recent evidence which reveals that complex disease traits are associated with collections of common variants found in monogenic disease genes (25), and that milder neurodevelopmental phenotypes (for example autism spectrum disorder, and probably mild ID) are more likely to involve combinations of rare and common variants (26,27). That the p.Gly876Ser variant has been identified in a patient with autism may provide additional support of its involvement (28).

Our data indicate that of the four variants we tested, three affect HCFC1 protein function, namely the p.Ser225Asn, p.Gly876Ser and p.Arg2016Trp variants found in families D144, D147 and CZE168, respectively. The p.Arg2016Trp variant disrupted a nuclear localization signal sequence in the C-terminus of the protein. Our localization studies show that this indeed did disrupt the ability of HCFC1 to become enriched in the nucleus; however, that the protein retained some function suggested that nuclear import was reduced, but not abolished. The p.Ser225Asn variant lies in the Kelch repeat domain of HCFC1 similar to previously reported loss-of-function HCFC1 variants (9). This domain is characterized as a protein–protein interaction motif and has been shown to mediate binding to transcription factors (for example E2F1 and E2F4 and THAP11) and chromatin regulators (such as MLL H3K4 Methyltransferase) (4,29). HCFC1 has been described as an obligate co-factor of THAP11, which recruits HCFC1 to target promoters, including that of the *MMACHC* gene (9,29,30). Thus, the p.Ser225Asn variant may disrupt these interactions, as has been postulated for other variants of HCFC1 affecting the Kelch domain (9). Likewise, the p.Gly876Ser variant may affect the protein–protein interactions mediated by the basic domain, which includes DNA-binding proteins (e.g. ZBTB17) and chromatin modifiers such as SIN3A (18). It will be interesting to identify which promoters and genes are affected on a global scale using patient-derived cell lines, to better understand the biology of HCFC1 and understand the genetic pathways that underlie ID. In all of our assays, the p.Ala1756Val variant behaved as wild-type HCFC1. While this does not support its pathogenicity, it does not exclude its involvement in some finer way and may reflect limitations of the assay conditions to detect subtle disruptions to HCFC1 function (see below). Such a small disruption may still be solely responsible for pathogenesis, or alternatively its effect may be compounded by contributions of other rare variants (e.g. the c.356A>G, p.Gln119Arg variant in ZMYM3 of unknown significance discovered in the patient) or common variants (24,25).

As aforementioned, loss-of-function mutations that alter the Kelch domain of HCFC1 have been identified in patients with CblC-like disorder called CblX (9,10). CblC is most commonly caused by mutations in *MMACHC* and results in methylmalonic aciduria and homocystinuria. The defect causes decreased levels of the coenzymes adenosylcobalamin (AdoCbl) and methylcobalamin (MeCbl), which results in decreased activity of the respective enzymes methylmalonyl-CoA mutase and methyltetrahydrofolate:homocysteine methyltransferase, also known as methionine synthesis. The main biochemical findings in CblC patients with *MMACHC* mutations are increased plasma homocysteine level, low or normal plasma metionin level, homocysteinuria and methylmalonic aciduria (31). Evidence from CblX patient cell lines and knockdown experiments in HEK293T cells revealed that these *HCFC1* mutations resulted in an almost complete loss of *MMACHC* expression, suggesting almost complete loss of HCFC1 function (9). The patients from whom the variants we tested here have in general much milder neurological presentations compared with CblX (and CblC) individuals and consistently, in our assays, it appears that the HCFC1 retained at least some biological activity, i.e. our assays suggest only partial loss of function

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

(Supplementary Material, Table S3). Furthermore, both the individuals with p.Arg2016Trp mutation had transiently elevated homocysteine levels, but repeatedly normal levels of plasma metionin and methylmalonic acid in urine (the measurement of AdoCbl and MeCbl was not available). Normal urine methylmalonic acid levels were also reported for the individual with the p.Ala477Thr variant. These nearly normal biochemical findings correspond to the much milder neurological defects compared with CblC patients, again suggesting partial MMACHC activity, and only a partial defect in HCFC1 function. It remains to be tested to what extent *MMACHC* expression may be affected in these individuals, and the other individuals, and we studied here (materials unavailable at this time), and more so, to what extent the metabolism of cobalamin is affected in all. Such information may have some clinical value, as expression levels of mutant *MMACHC* alleles have been associated with age of onset and severity in CblC patients (11). However, as CblC causing *MMACHC* mutations are all reported as homozygous (or compound heterozygous), it suggests that unless MMACHC function is severely compromised, the metabolic symptoms may not exist (11). That the metabolic features of CblX disorder are relatively mild compared with CblC patients, but the neurological phenotype is much more severe (includes brain malformation and intractable epilepsies) suggests that *MMACHC* deficiency alone does not explain the neurological phenotype, and that *HCFC1* is important for the expression of additional genes important for normal brain development (9). Thus, mutations resulting in partial loss of HCFC1 function may result in neurological phenotypes in the absence of the clinical presentation of cobalamin deficiency. Furthermore, while treatment of early-onset CblC patients with a combinatorial approach consisting of supplementation of hydroxycobalamin, betain and folic acid provides some improvements in the visceral and haematological symptoms, the efficacy on neurological outcomes is not established, and patients invariably continue to show neurological and cognitive impairment regardless of time of diagnosis and/or treatment initiation (32).

The mounting challenge of assigning pathogenicity to large numbers of variants is typically approached by employing genetic and statistical evidence often without in-depth functional studies or even detailed clinical assessment. By comprehensive investigations of four *HCFC1* variants using all above criteria, we found conforming evidence for pathogenicity of p.Ser225Asn and p.Arg2016Trp; however, some discordance was observed with the remaining two variants studied. The variant with the highest CADD score, p.Ala1756Val, which is also not present in any publicly available control dataset we analysed, behaved as wild type in our cell-based studies. Conversely, the p.Gly876Ser variant which is found at very low frequency in northern European descent population displayed loss of function in our cell assays. This raises important issues pertaining to the relevance of such functional tests for reaching correct clinical diagnosis and actions thereafter. Resolution of these fundamental questions, also relevant to many other genes and disorders, will involve deeper understanding of the function of these genes, better genotype–phenotype databases, more sophisticated models (e.g. patient-derived stem cells) or simply more patients and variants investigated. Our work underlines the importance of combined approach to the resolution of the relevance of genetic variation in neurological disease, involving clinical, genetic, bioinformatics and cell and molecular functional studies.

In conclusion, our data provide additional support for the involvement of *HCFC1* in ID. Together, with previous studies, it is now apparent that both gain and loss of HCFC1 function can result in changes to the behaviour of embryonic neural cells, which likely underpins aspects of patient phenotypes (2,33). Regarding loss-of-function mutations, the severity of the phenotypes likely reflects the extent to which the mutations affect HCFC1 function, with almost complete loss of function causing CblX featuring severe neurological phenotypes (e.g. retractable epilepsy, brain malformations, severe neurocognitive impairment) and clinically relevant deficiencies in cobalamin, whereas partial loss of function resulting in less severe neurological phenotypes (e.g. mild ID) (Supplementary Material, Table S3) (9). The continued emergence of large-scale sequencing practices in research and diagnostics will likely provide further insight into the spectrum of phenotypic outcomes stemming from both gain- and loss-of-function mutations in *HCFC1*.

Materials and Methods

Animal use

This study was performed under regulations of the South Australian Animal Welfare Act 1986, and in strict accordance with the Australian Code of Practice for the Care of Animals for Scientific Purposes, 2004. The protocol was approved by the Women's and Children's Health Network (WCHN) Animal Ethics Committee (Approval Number: 944/03/15). All euthanasia was performed using cervical dislocation, and every effort was made to minimize suffering. Time-mated pregnant female Swiss mice were obtained from the Women's and Children's Health Network Animal Care Facility (Women's and Children's Hospital, Adelaide, Australia).

Generation of lentivirus

Constitutive 3rd generation lentiviral vectors were employed as previously described (34–36). To generate *Hcfc1*-specific shRNA transfer vectors, three shRNA sequences predicted to specifically target *Hcfc1* UTRs were designed using BLOCK-it RNAi designer software [Invitrogen, Carlsbad, CA, USA)]. A shRNA sequence targeted to luciferase was used as a control (36). The shRNA sequences are T1: GCAGAAGGCAGATTGGAAAGA; T2: GGAGATAACACCCA TACTTAA and T3: GCACTTGTT TGTAGT TCATCG. These sequences were cloned into the lentiviral transfer vector plv-C-shRNA, and lentiviral particle stocks were generated and titrated as previously described (35,36). Transduction of neurons was performed at either multiplicity of infection (MOI) = 1 overnight (O/N) at Day 1 of culture. Transduction of NPCs was achieved at MOI = 20 using O/N incubation with dissociated cells derived from secondary neurospheres. For purification of transduced cells, neurospheres were dissociated and EGFP expressing cells purified using FACS (BD FACSAriaII flow cytometer; BD Bioscience, San Jose, CA, USA), for further culture as neurospheres.

Cloning of HCFC1 expression constructs

pCGN-HCFC1-FL vector was a kind gift from W. Herr, Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland (15). To generate mutant constructs, fragments containing the desired nucleotide(s) for mutagenesis were first amplified and then subcloned into pGEMT by pGEM®-T Easy Vector Systems (Promega, Madison, WI, USA) using the following primer pairs: for p.Ser225Asn, F-TATGACGTGCCTGACTATGC and R-TAGG TACCAGAGGTCCTTGC; for p.GlySer876, F-AGTAGCCCACAGAT GAGTGG and R-CTTGAGAGGGTGGCAATGG; for p.Ala1756Val and p.Ala2016Trp, F-TTATCACTTCCCCAAGAGC and R-CTCACCCT GAAGTTCTCAGG. Mutagenesis reaction was performed directly on pGEMT vector using QuikChange Multi Site-Directed

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

Mutagenesis Kit (Stratagene, La Jolla, CA, USA) according to the manufacturer's protocol using the following primers: for p.Ser225Asn, F-TACGGCGGGATGAATGGCTGCAGGCTGG; for p.GlySer876, F-CGTTGGTTGTGAAAAGCACCACAGGTGTCACG; for p.Ala1756Val, F-CACTGAGAGCCTGGTTCCATCCAACACATTTGTGG; and for p.Ala2016Trp, F-CCAAGCCAGCCAACAAGTGGCCCATGTCCTCTC. Subsequently, mutated fragments were subcloned back into pCGN-HCFC1-FL construct by restriction digest followed by ligation with T4 ligase (NEB, Ipswich, MA, USA): for p.Ser225Asn, SexAI and XbaI ; for p.GlySer876, *NotI* and *SpeI*; for p.Ala1756Val or p.Ala2016Trp, *BamHI* and *SgrA1*. All cloning steps were performed using DH5α bacterial strain. Vector extraction was performed using Miniprep Kit (QIAGEN, Germantown, MD, USA) according to the manufacturer's protocol. For transfection and virus production, vectors were purified using Endofree Plasmid Maxi Kit (QIAGEN).

Cell culture

Isolation of NPCs from the E18 cortex was as previously described (36–38). Single cells for NPC assays were generated by passing dissociated cells through a 0.75 µm cell filter (BD Biosciences). All neural progenitor cell assays were conducted in biological triplicate, with each replicate representing a separate transduction experiment. Averages of these triplicate results are reported. The cell proliferation assay was conducted using the Cell Titre 96AQueous Kit (Promega) as per manufacturer's instruction; each biological replicate was analysed using technical quadruplicates and normalized against starting cell numbers (±4.2% cells per well). Adherent NPC assays were performed by plating dissociated neurosphere cells onto poly-l-lysine (Sigma-Aldridge, St Louis, MO, USA) coated coverslips (Menzel-glasser, Thermo Fisher Scientific, Waltham, MA, USA) at $1 \times 10^4/cm^2$, respectively, and culturing cells in neurosphere media with or without EGF. Immunofluorescent staining and fluorescent microscopy (see below) were used for cell count analysis. For each cell count analysis, at least 200 cells were scored for each replicate experiment (+EGF condition: pax6 total $n = 1230$ control, and $n = 1486$ *Hcfc1*^{shRNA}; phospho-Histone3: total $n = 1168$ control, and $n = 1208$ *Hcfc1*^{shRNA}; activated caspase3: total $n = 602$ control, and $n = 613$ *Hcfc1*^{shRNA}; and –EGF condition: CNPase: total $n = 780$ control, and $n = 602$ *Hcfc1*^{shRNA}; GFAP: total $n = 1160$ control; $n = 1130$ *Hcfc1*^{shRNA}; βIII-tubulin: total $n = 640$ control, and $n = 603$ *Hcfc1*^{shRNA}; Pax6: total $n = 680$ control; $n = 1190$ *Hcfc1*^{shRNA}). Isolation of primary hippocampal neurons was as described previously (2). Nucleofection was conducted as previously described (2) using 5 µg of pCGN-HCFC1 expression vectors (wild type or variant) and 1 µg of pDsRed-Monomer-C1 (RFP expression vector; Clontech, Mountain View, CA, USA). Morphometric analysis was conducted using immunofluorescent staining and fluorescent microscopy (see below). All data represent average of three biological replicates. For each replicate, at least 30 neurons were scored using the ImageJ software package (NIH). HEK293T and NIH3T3 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal calf serum (Gibco, Life Technologies, Mulgrave, Victoria, Australia). Transfections were conducted using Lipofectamine 3000 as per manufacturer's instruction (Invitrogen).

Immunofluorescence

Cultured cells were fixed with 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS) for 15 min at room temperature (R/T). Cells were block permeabilized using PBS containing 0.1% Tween20 (PBST) with either 10% normal horse serum (NHS) or 5% bovine serum albumin (BSA) for 1 h at R/T. Primary and secondary antibodies were incubated in PBST with either 3% NHS or 1% BSA overnight (O/N) at 4°C and 1 h at RT, respectively, at the following dilutions; chicken anti-MAP2, mouse anti-CNPAse, mouse anti-Tau1 (all at 1:2000; Chemicon, Millipore Bioscience Research Reagents, Temecula, CA, USA), rabbit anti-GFAP, mouse anti-βIII-tubulin, (both at 1:300; Sigma-Aldridge), rabbit anti-Pax6 (1:200; Chemicon), rabbit anti-phospho-histone3 (1:200, Abcam, Cambridge, UK) Rabbit anti-activated caspase3 (1:200, BD Biosciences), mouse anti-Myc (1:500, Santa Cruz, Santa Cruz Biotechnology, Santa Cruz, CA, USA), chicken anti-HA (Abcam, 1:500), rabbit anti-HCFC1 (1:200, Bethyl Laboratories, Montgomery, TX, USA), donkey anti-sheep Alexafluor555, donkey anti-rabbit Alexafluor488/555/647 and donkey anti-mouse Alexafluor488/555/647 (all 1:1000; Invitrogen), donkey anti-chickenCy3 (Jackson Laboratories, Bar Harbour, ME, USA). Cells were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) and mounted with Slow-fade mounting media (both from Invitrogen). Non-specific staining was controlled by using secondary-only controls (data not shown). Fluorescence was viewed using the Axioplan2 microscope (Carl Zeiss, Jena, Germany) fitted with an HBO 100 lamp (Carl Zeiss). Images were captured using an Axiocam Mrm camera and Axio Vs40 v4.5.0.0 software (Axiovision, Carl Zeiss).

Biochemical analysis

Protein was isolated from cells using a lysis buffer (120 mM NaCl, 50 mM Tris–HCl (pH 8.0), 0.5% NP-40 (v/v), 1 × protease inhibitor cocktail (Sigma), 1 mM Na_3VO_4, 1 mM NaF, 1 mM PMSF) and quantitated using Bradford assay (Biorad, Hercules, CA, USA). Samples were separated and transferred to nitrocellulose (Biotrace NT, Pall Corporation, New York, NY, USA) using the NuPage precast gel system as per manufacturer's instructions (Invitrogen). Blots were blocked using PBST containing 5% NHS and 5% skim milk. Antibodies were incubated in this same solution at the following dilutions; rabbit anti-HCFC1 (1:2000), mouse anti-β-actin (1:5000; Sigma-Aldridge), goat anti-rabbit-HRP and goat anti-mouse HRP (both 1:1000; Dako, Glostrup, Denmark). RNA was isolated from cells/tissues using TRIzol reagent as per manufacturer's instructions (Invitrogen) and further processed using the RNAeasy kit including DNAse treatment (Qiagen). cDNA was generated using SuperscriptIII reverse transcriptase as per manufacturer's instructions (Invitrogen) using random hexamer priming (Geneworks, Adelaide, Australia). The StepOne platform and software (Applied Biosciences, Invitrogen) was employed for qRT-PCR analysis. PCR reactions were generated using the sybr green master mix as per manufacturer's instructions (Biorad) and run using the following parameters: 95°C—5 min; 35 cycles of: 95°C—10 s, 60°C—30 s; followed a melt curve increment step 60°C–100°C. All primers used met strict quality control parameters amplifying control cDNA at 100 ± 10% efficiencies and producing only single PCR products. Primers for *β-Actin*, *HCFC1* and *MMACHC* are as previously described (2,9).

Supplementary material

Supplementary material is available at *HMG* online.

Conflict of Interest statement. None declared.

Funding

L.A.J. acknowledges support from Women's and Children's Hospital Foundation and J.G. of NHMRC Program Grant 628952 and

Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

NHMRC senior research fellowship 1041920. Z.S. was supported by grants NT/14200 and DRO UH Motol 00064203 from the Czech Ministry of Health, and CHERISH 223692 from the European Commission. V.M.K. was supported by the EU FP7 project GENCODYS, grant number 241995.

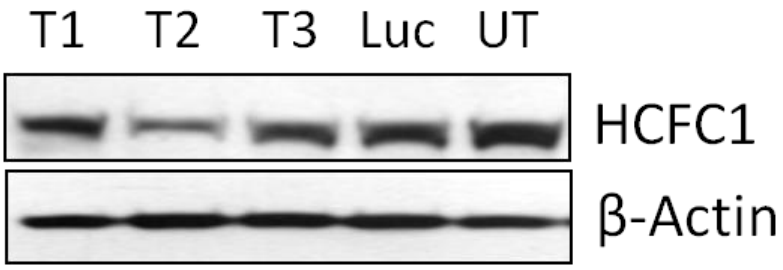
References

1. Ropers, H.H. (2008) Genetics of intellectual disability. *Curr. Opin. Genet. Dev.*, **18**, 241–250.
2. Huang, L., Jolly, L.A., Willis-Owen, S., Gardner, A., Kumar, R., Douglas, E., Shoubridge, C., Wieczorek, D., Tzschach, A., Cohen, M. *et al.* (2012) A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability. *Am. J. Hum. Genet.*, **91**, 694–702.
3. Vercauteren, K., Gleyzer, N. and Scarpulla, R.C. (2008) PGC-1-related coactivator complexes with HCF-1 and NRF-2beta in mediating NRF-2(GABP)-dependent respiratory gene expression. *J. Biol. Chem.*, **283**, 12102–12111.
4. Tyagi, S., Chabes, A.L., Wysocka, J. and Herr, W. (2007) E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol. Cell*, **27**, 107–119.
5. Julien, E. and Herr, W. (2004) A switch in mitotic histone H4 lysine 20 methylation status is linked to M phase defects upon loss of HCF-1. *Mol. Cell.*, **14**, 713–725.
6. Goto, H., Motomura, S., Wilson, A.C., Freiman, R.N., Nakabeppu, Y., Fukushima, K., Fujishima, M., Herr, W. and Nishimoto, T. (1997) A single-point mutation in HCF causes temperature-sensitive cell-cycle arrest and disrupts VP16 function. *Genes Dev.*, **11**, 726–737.
7. Michaud, J., Praz, V., James Faresse, N., Jnbaptiste, C.K., Tyagi, S., Schutz, F. and Herr, W. (2013) HCFC1 is a common component of active human CpG-island promoters and coincides with ZNF143, THAP11, YY1, and GABP transcription factor occupancy. *Genome Res.*, **23**, 907–916.
8. Hu, H., Haas, S.A., Chelly, J., Van Esch, H., Raynaud, M., de Brouwer, A.P., Weinert, S., Froyen, G., Frints, S.G., Laumonnier, F. *et al.* (2015) X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Mol. Psychiatry*, doi:10.1038/mp.2014.193.
9. Yu, H.C., Sloan, J.L., Scharer, G., Brebner, A., Quintana, A.M., Achilly, N.P., Manoli, I., Coughlin, C.R. 2nd, Geiger, E.A., Schneck, U. *et al.* (2013) An X-linked cobalamin disorder caused by mutations in transcriptional coregulator HCFC1. *Am. J. Hum. Genet.*, **93**, 506–514.
10. Gerard, M., Morin, G., Bourillon, A., Colson, C., Mathieu, S., Rabier, D., Billette de Villemeur, T., Ogier de Baulny, H. and Benoist, J.F. (2015) Multiple congenital anomalies in two boys with mutation in HCFC1 and cobalamin disorder. *Eur. J. Med. Genet.*, **58**,148–153.
11. Lerner-Ellis, J.P., Anastasio, N., Liu, J., Coelho, D., Suormala, T., Stucki, M., Loewy, A.D., Gurd, S., Grundberg, E., Morel, C.F. *et al.* (2009) Spectrum of mutations in MMACHC, allelic expression, and evidence for genotype-phenotype correlations. *Hum. Mutat.*, **30**, 1072–1081.
12. Wieczorek, D., Bogershausen, N., Beleggia, F., Steiner-Haldenstatt, S., Pohl, E., Li, Y., Milz, E., Martin, M., Thiele, H., Altmuller, J. *et al.* (2013) A comprehensive molecular study on Coffin-Siris and Nicolaides-Baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. *Hum. Mol. Genet.*, **22**, 5121–5135.
13. Kosugi, S., Hasebe, M., Tomita, M. and Yanagawa, H. (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc. Natl Acad. Sci. USA*, **106**, 10171–10176.
14. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M. and Shendure, J. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, **46**, 310–315.
15. Wilson, A.C., LaMarco, K., Peterson, M.G. and Herr, W. (1993) The VP16 accessory protein HCF is a family of polypeptides processed from a large precursor protein. *Cell*, **74**, 115–125.
16. Capotosti, F., Guernier, S., Lammers, F., Waridel, P., Cai, Y., Jin, J., Conaway, J.W., Conaway, R.C. and Herr, W. (2011) O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell*, **144**, 376–388.
17. Julien, E. and Herr, W. (2003) Proteolytic processing is necessary to separate and ensure proper cell growth and cytokinesis functions of HCF-1. *EMBO J.*, **22**, 2360–2369.
18. Mangone, M., Myers, M.P. and Herr, W. (2010) Role of the HCF-1 basic region in sustaining cell proliferation. *PLoS ONE*, **5**, e9020.
19. Wilson, A.C., Freiman, R.N., Goto, H., Nishimoto, T. and Herr, W. (1997) VP16 targets an amino-terminal domain of HCF involved in cell cycle progression. *Mol. Cell. Biol.*, **17**, 6139–6146.
20. Yu, H., Mashtalir, N., Daou, S., Hammond-Martel, I., Ross, J., Sui, G., Hart, G.W., Rauscher, F.J. 3rd, Drobetsky, E., Milot, E. *et al.* (2010) The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol. Cell. Biol.*, **30**, 5071–5085.
21. Risheg, H., Graham, J.M. Jr., Clark, R.D., Rogers, R.C., Opitz, J. M., Moeschler, J.B., Peiffer, A.P., May, M., Joseph, S.M., Jones, J.R. *et al.* (2007) A recurrent mutation in MED12 leading to R961W causes Opitz-Kaveggia syndrome. *Nat. Genet.*, **39**, 451–453.
22. Hoyer, J., Ekici, A.B., Endele, S., Popp, B., Zweier, C., Wiesener, A., Wohlleber, E., Dufke, A., Rossier, E., Petsch, C. *et al.* (2012) Haploinsufficiency of ARID1B, a member of the SWI/SNF-a chromatin-remodeling complex, is a frequent cause of intellectual disability. *Am. J. Hum. Genet.*, **90**, 565–572.
23. Wright, C.F., Fitzgerald, T.W., Jones, W.D., Clayton, S., McRae, J.F., van Kogelenberg, M., King, D.A., Ambridge, K., Barrett, D.M., Bayzetinova, T. *et al.* (2014) Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet*, doi: 10.1016/S0140-6736 (14)61705-0.
24. Girirajan, S. and Eichler, E.E. (2010) Phenotypic variability and genetic susceptibility to genomic disorders. *Hum. Mol. Genet.*, **19**, R176–R187.
25. Blair, D.R., Lyttle, C.S., Mortensen, J.M., Bearden, C.F., Jensen, A.B., Khiabanian, H., Melamed, R., Rabadan, R., Bernstam, E.V., Brunak, S. *et al.* (2013) A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. *Cell*, **155**, 70–80.
26. Gaugler, T., Klei, L., Sanders, S.J., Bodea, C.A., Goldberg, A.P., Lee, A.B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J. *et al.* (2014) Most genetic risk for autism resides with common variation. *Nat. Genet.*, **46**, 881–885.
27. Schaaf, C.P., Sabo, A., Sakai, Y., Crosby, J., Muzny, D., Hawes, A., Lewis, L., Akbar, H., Varghese, R., Boerwinkle, E. *et al.* (2011) Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Hum. Mol. Genet.*, **20**, 3366–3375.
28. Piton, A., Gauthier, J., Hamdan, F.F., Lafreniere, R.G., Yang, Y., Henrion, E., Laurent, S., Noreau, A., Thibodeau, P., Karemera, L. *et al.* (2011) Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol. Psychiatry*, **16**, 867–880.

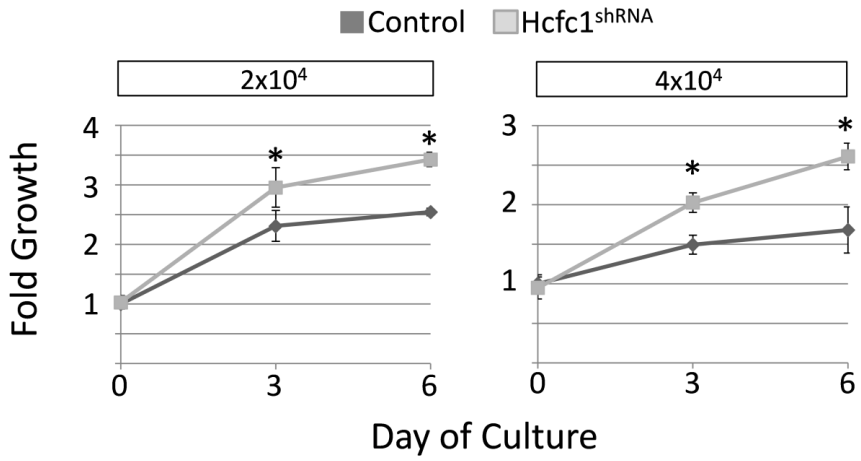
Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015

29. Parker, J.B., Palchaudhuri, S., Yin, H., Wei, J. and Chakravarti, D. (2012) A transcriptional regulatory role of the THAP11-HCF-1 complex in colon cancer cell function. *Mol. Cell. Biol.*, **32**, 1654–1670.
30. Dejosez, M., Levine, S.S., Frampton, G.M., Whyte, W.A., Stratton, S.A., Barton, M.C., Gunaratne, P.H., Young, R.A. and Zwaka, T.P. (2010) Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. *Genes Dev.*, **24**, 1479–1484.
31. Rosenblatt, D.S. and Fowler, B. (2006) In Fenandes, J., Saudubray, J.M., van den Berghe, G. and Walter, J.H. (eds), *Inborn Metabolic Diseases, Diagnosis and Treatment*. Springer, Heidelberg, Germany, pp. 379–396.
32. Martinelli, D., Deodato, F. and Dionisi-Vici, C. (2011) Cobalamin C defect: natural history, pathophysiology, and treatment. *J. Inherit. Metab. Dis.*, **34**, 127–135.
33. Quintana, A.M., Geiger, E.A., Achilly, N., Rosenblatt, D.S., Maclean, K.N., Stabler, S.P., Artinger, K.B., Appel, B. and Shaikh, T.H. (2014) Hcfc1b, a zebrafish ortholog of HCFC1, regulates craniofacial development by modulating MMACHC expression. *Dev. Biol.*, **396**, 94–106.
34. Brown, C.Y., Sadlon, T., Gargett, T., Melville, E., Zhang, R., Drabsch, Y., Ling, M., Strathdee, C.A., Gonda, T.J. and Barry, S.C. (2010) Robust, reversible gene knockdown using a single lentiviral short hairpin RNA vector. *Hum. Gene Ther.*, **21**, 1005–1017.
35. Barry, S.C., Harder, B., Brzezinski, M., Flint, L.Y., Seppen, J. and Osborne, W.R. (2001) Lentivirus vectors encoding both central polypurine tract and posttranscriptional regulatory element provide enhanced transduction and transgene expression. *Hum. Gene Ther.*, **12**, 1103–1108.
36. Jolly, L.A., Homan, C.C., Jacob, R., Barry, S. and Gecz, J. (2013) The UPF3B gene, implicated in intellectual disability, autism, ADHD and childhood onset schizophrenia regulates neural progenitor cell behaviour and neuronal outgrowth. *Hum. Mol. Genet.*, **22**, 4673–4687.
37. Basak, O. and Taylor, V. (2007) Identification of self-replicating multipotent progenitors in the embryonic nervous system by high Notch activity and Hes5 expression. *Eur. J. Neurosci.*, **25**, 1006–1022.
38. Giachino, C., Basak, O. and Taylor, V. (2009) Isolation and manipulation of mammalian neural stem cells in vitro. *Methods Mol. Biol.*, **482**, 143–158.

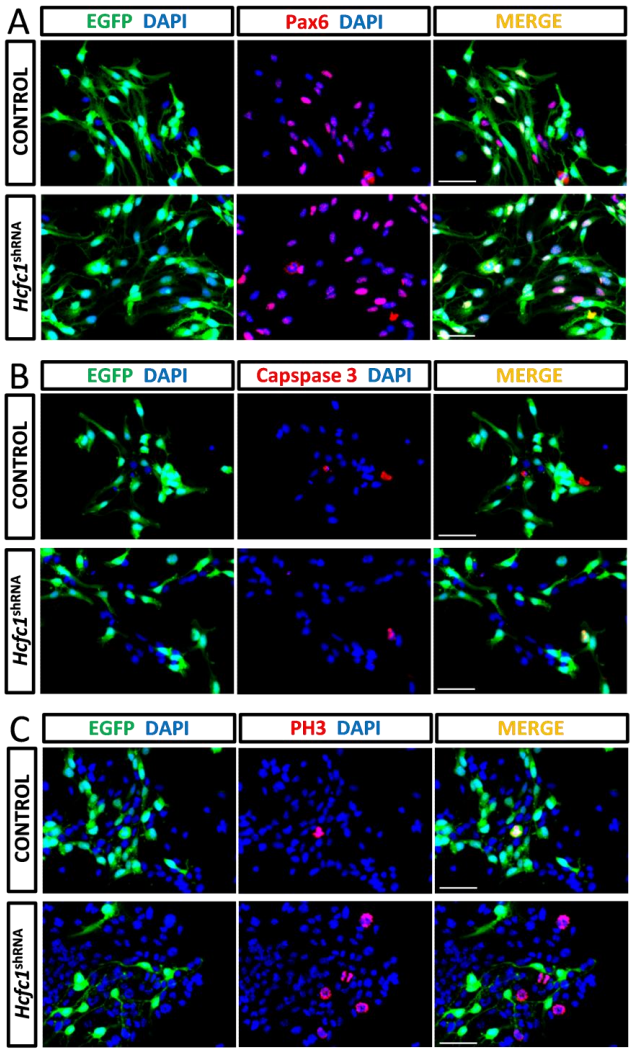
Downloaded from http://hmg.oxfordjournals.org/ at Univerzita Karlova v Praze on September 10, 2015



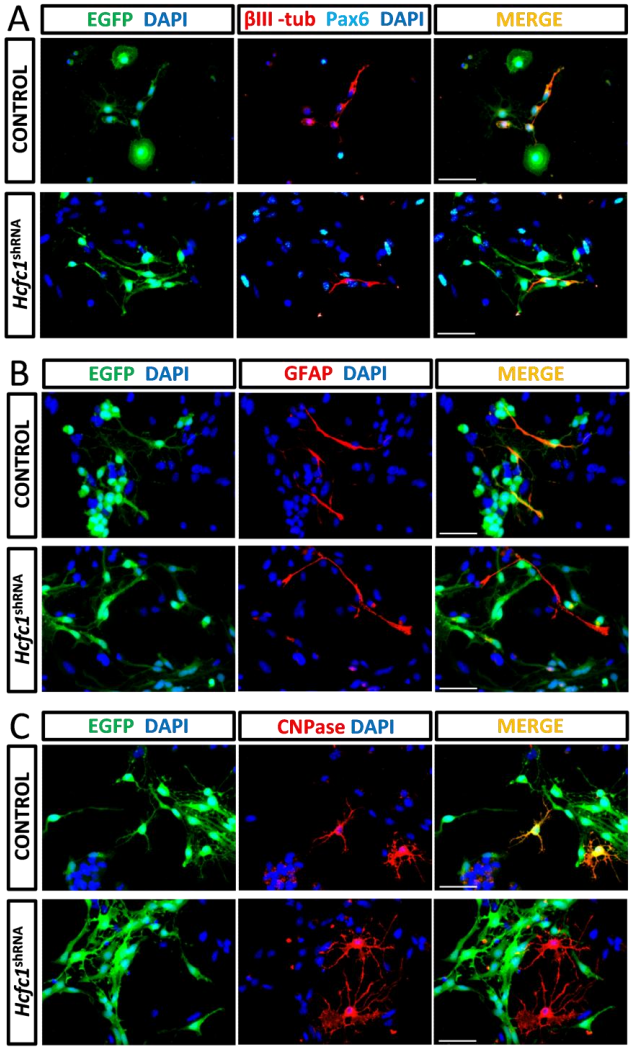
Supplementary Figure 1. Identification of an shRNA trigger sequence that knocks down Hcfc1. Western blot analysis of Hcfc1 and β-actin expression in lysates collected from NIH3T3 cells that were either untransduced (UT) or transduced with lentiviral particles encoding one of three *Hcfc1* shRNA triggers (T1-3; all targeted to untranslated regions) or control shRNA trigger (against luciferase; Luc) . The T2 sequence was chosen for subsequent *Hcfc1* knockdown experiments.



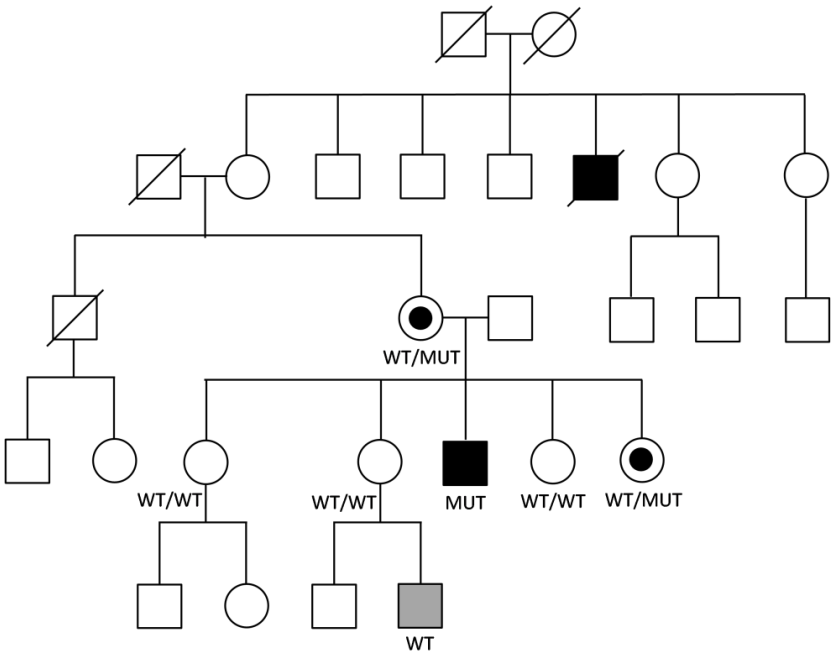
Supplementary Figure 2. Loss of Hcfc1 in NPCs causes increased growth kinetics. Isolated primary cortical NPCs were grown as neurospheres and transduced with lentiviral particles encoding *EGFP* and either inert shRNA sequences (Control) or *Hcfc1* targeted shRNA. Transduced cells were purified using FACS to produce pure populations prior to assay. Cells were then plated at the higher initial densities of either $2x10^4$ or $4x10^4$ cells per well and a cell proliferation assay was used to quantitate cell growth at days 0, 3 and 6 days. *$p<0.05$ by Students t-test.



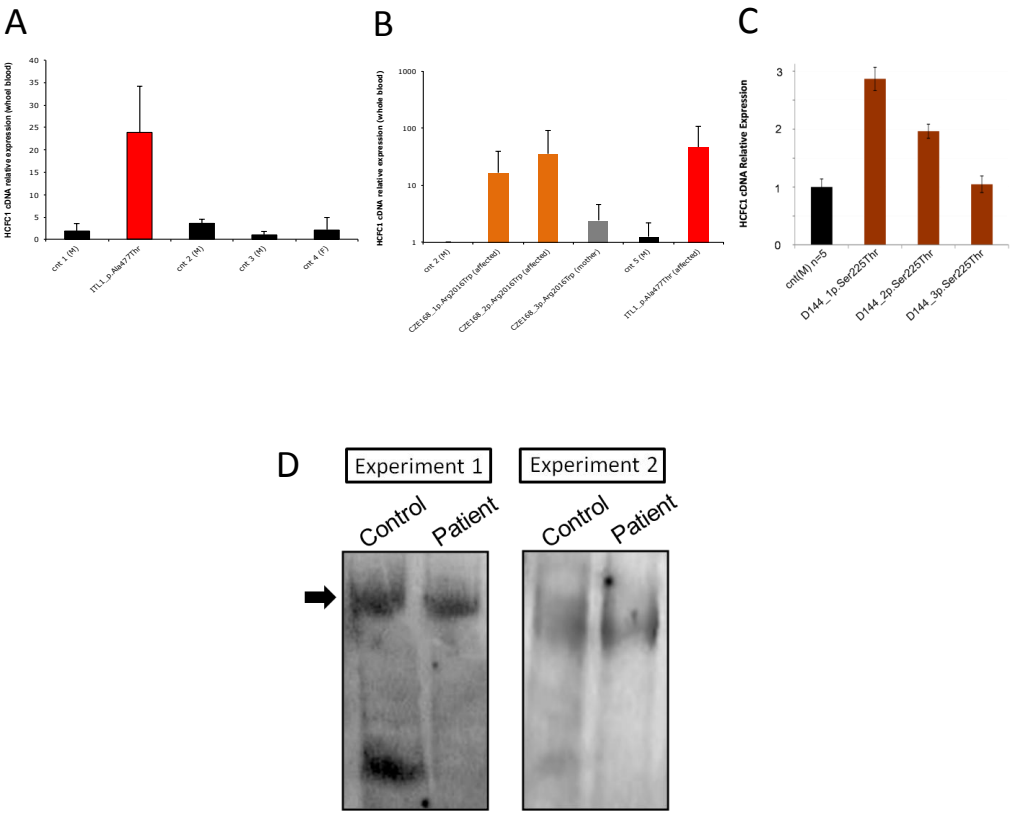
Supplementary Figure 3. Neurosphere cultures contain more proliferative NPCs following loss of *Hcfc1* expression. Isolated primary cortical NPCs grown as neurospheres were transduced with lentiviral particles encoding *EGFP* and either inert shRNA sequences (Control) or *Hcfc1* targeted shRNA. Neurospheres were dissociated and cells attached to poly-l-lysine substrate and allowed to grow for three days in the presence of EGF to promote further proliferation. Transduced cells express *EGFP* (green). Cultures were fixed and stained using immunofluoresence. Representative images of cultures stained for (A) progenitor cells (Pax6: red), (B) apoptotic cells (activated-Caspase3: red) and (C) mitotic cells (phospho-Histone3 (PH3): red). All cultures were counterstained with DAPI (Blue) to identify cell nuclei.



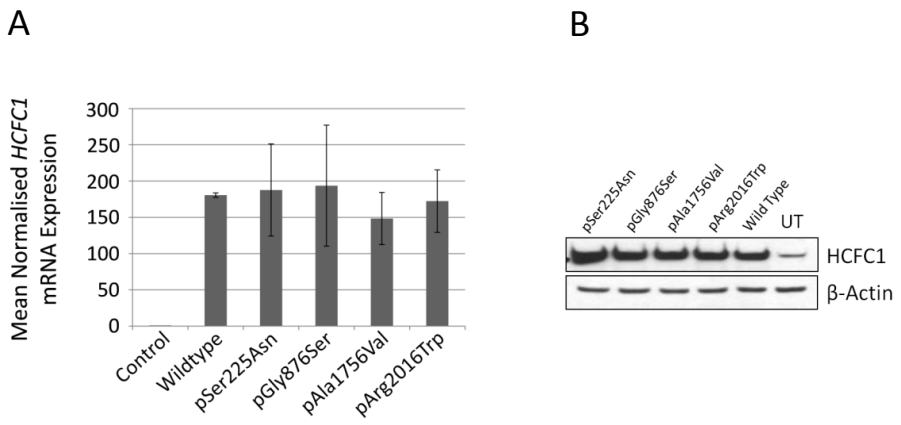
Supplementary Figure 4. **Loss of Hcfc1 in neurospheres results in reduced differentiation of NPCs**. Isolated primary cortical NPCs grown as neurospheres were transduced with lentiviral particles encoding *EGFP* and either inert shRNA sequences (Control) or *Hcfc1* targeted shRNA. Neurospheres were dissociated and cells attached to poly-l-lysine substrate and allowed to grow for three days in the absence of growth factors to promote differentiation. Transduced cells express *EGFP* (green). Cultures were fixed and stained using immunofluorescence . Representative images of cultures stained for (A) progenitor cells (Pax6: cyan), and neurons (βIII-tubulin), and (B) astrocytes (GFAP: red) and (C) oligodendrocytes (CNPase: red). All cultures were counterstained with DAPI (Blue) to identify cell nuclei.



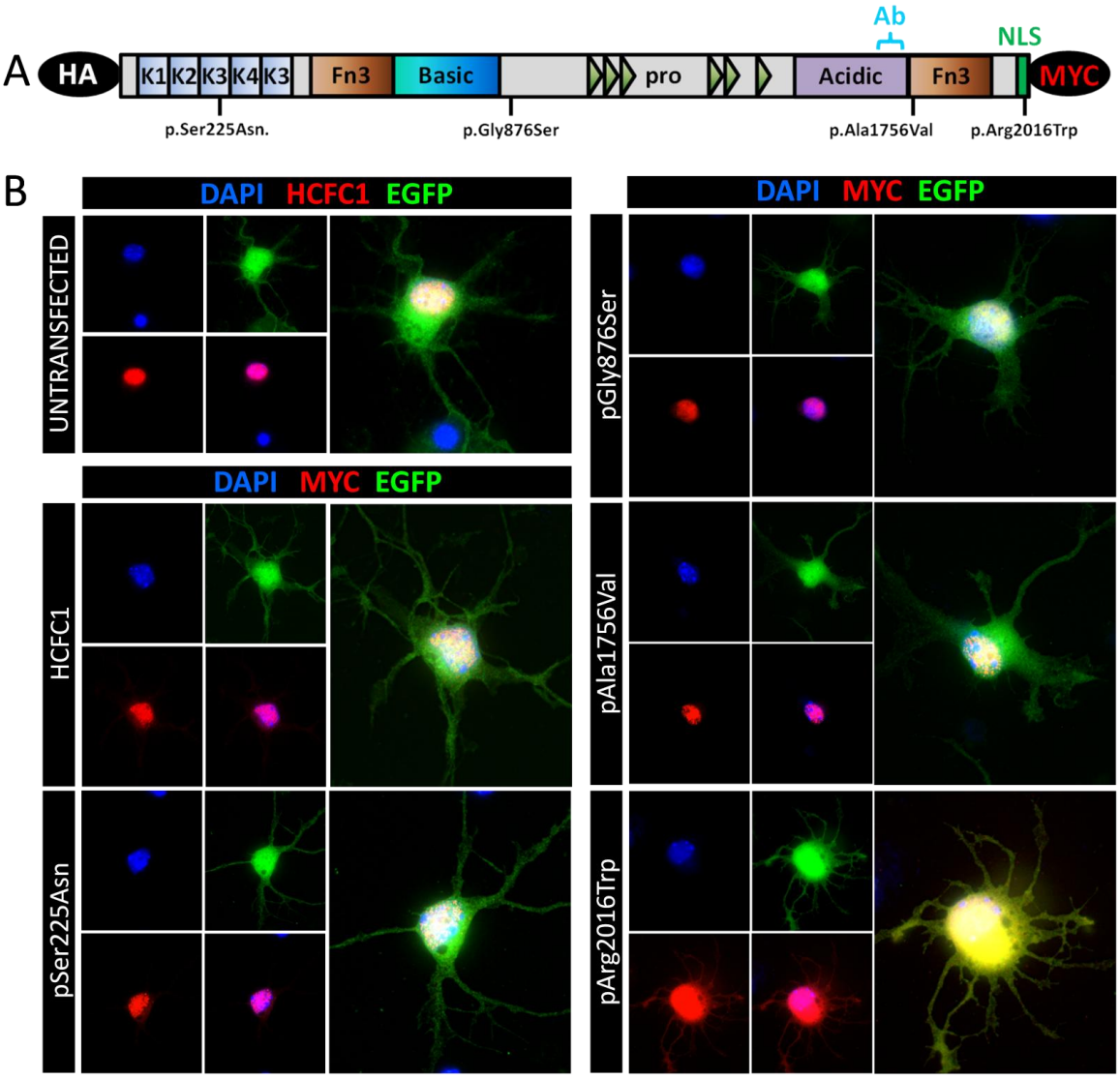
Supplementary Figure 5. Extended pedigree of family ITL1. Black fill indicates ID, whilst grey fill indicates autism spectrum disorder. Segregation of HCFC1 variants in the family; 153,225,268 C>T (HG19); c.1429G>A; p.Ala477Thr. WT: wildtype; MUT: mutant.



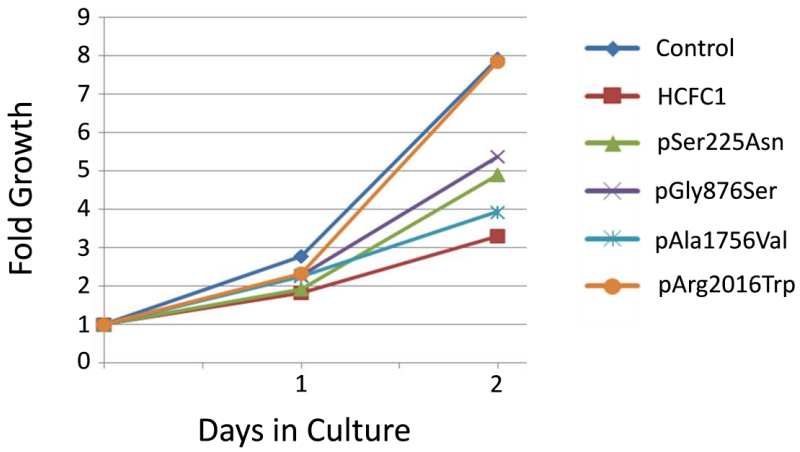
Supplementary Figure 6. Expression of HCFC1 variants in patient blood. A-C. qRT-PCR analysis of *HCFC1* mRNA expression. Black bars: control samples (M: male and F: female); Red bars: individual with p.Ala477Thr variant; Orange bar: brothers with p.Arg2016Trp variant; Brown bar: brothers with p.Ser225Thr variant. D. Duplicate western blot analysis of HCFC1 protein expression from single individual with p.Arg2016Trp variant and control sample. Arrow indicates full length HCFC1.



Supplementary Figure 7. *HCFC1* variants do not alter the over-expression levels compared to each other and to wildtype. HEK293T cells were transfected with an empty expression vector (control) or with expression vectors encoding either wildtype or variant *HCFC1* forms and RNA and protein were isolated. A. qRT-PCR analysis of *HCFC1* mRNA expression. B. Western blot analysis HCFC1 and β-actin protein expression.



Supplementary Figure 8. ***HCFC1*** **p.Arg2016Trp variant fails to localise to the nucleus in neuronal cells**. A. Diagram of HCFC1 protein with the locations of the variants being tested. Note also that exogenously expressed HCFC1 proteins are tagged by HA at the N-terminus and Myc at the C-terminus, and also the C-terminal location of the epitope used to generate the HCFC1 antibody (Ab). B. Primary hippocampal neurons were transfected with vectors encoding *EGFP* and either empty vector control or wild-type or variant forms of HCFC1. Immunofluorescent detection of endogenous HCFC1 (using the anti-HCFC1 antibody; red, upper right panel), and exogenous wildtype and variant forms (all other panels) using anti-HA (red) antibodies. Transfected cells express *EGFP* (green). Cell nuclei are counterstained with Dapi (blue). Note that p.Arg2016Trp fails to exclusively localise to the nucleus and is found localised throughout the cell soma and dendrites in a pattern equivalent to *EGFP.*



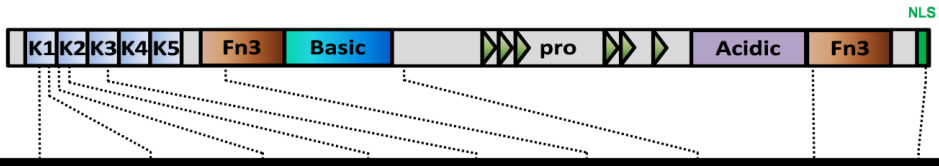
Supplementary Figure 9. **HCFC1 variants disrupt its normal function in the control of cell growth of HEK293T cells.** HEK293T cells were transfected with an empty expression vector (control) or with expression vectors encoding either wildtype or variant HCFC1 forms. Transfected cells were plated at equivalent densities and allowed to grow for 1 or 2 days at which point cell numbers were assayed. n=6, * is significantly different to wildtype, $p<0.05$ student's t-test.

Family CZE168; c.6046C>T, p.Arg2016Trp		
Brother A	**Medical History**	First uneventful pregnancy leading to a spontaneous birth at the 40th week of gestation (2550g / 46 cm). Individual displayed hypotrophy and failed to thrive during the newborn period. He suffered from pyelonephritis and then bronchopneumonia at seven months of age. Psychomotor development delay was noted at one year of age, independent walking occurred a two and a half years of age. He had delayed speech development and was later diagnosed with dyslalia. He also had cryptorchidism, congenital cleft lip and partial cleft palate (surgery at 1 year of age). During adolescence he had two episodes of generalized epileptic seizures and was diagnosed with mild intellectual disability. He is short stature (165 cm, 3rd centile), has a long philtrum, thick lips and long nose, with head circumference within the normal range. Opthalmic exam identified hypermetropia and strabismus divergens. Audiology exam was normal. Examinations of skin, appendages, skeleton and dermaoglyphs were normal. Echocardiography revealed diffuse hypocontractility of the left ventricle (ejection fraction 43-45%). He is overweight and suffers from diabetes type 2 from 32 years of age (on a diet). An abdominal ultrasound revealed mild liver steatosis, but otherwise normal.
	Lab Tests	Molecular Genetics: Has returned normal results for screening of karyotype, FMR1 expansion mutation, subtelomeric rearangements (FISH), 22q11.2 deletion (FISH) and SNP array (300K). Metabolic Screening: increased homocysteine levels (18.6 and 15.0µmol/l (sampled in June and October of 2011 respectively; normal range 3.5-15.0µmol/l)); normal folic acid (502 µg/l (normal range 176-589µg/l)); holotranscobalamin 70pmol/l (normal range 19-119), the level of plasma and urine methionin and urine methylmalonate was repeatedly normal; lower urine pterin excretion (both total biopterin (0.31 or 0.29mmol/mol of creatine) and neopterin (0.14 or 0.21 mmol/mol of creatine)); lysosomal storage disorder screening was normal, and electron microscopic analysis of deep skin biopsy was normal, however vacuolisation of peripheral blood lymphocytes was found. Neurotransmitter screening: normal levels of vanilalanin, vanilglykol, 5-hydroxyindolacetic acid (HIAA) and vanilacetic acid (HVA) were found in cerebrospinal fluid.
Brother B	**Medical History**	Third uneventful pregnancy leading to birth at 39th week by Caesarean section due to maternal hypertension (2850g / 50cm). Psychomotor development delay noted at 8 months along with hypotonia. Independent walking occurred at 18 months of age and he displayed delayed speech development and was later diagnosed with dyslalia. He was diagnosed with moderate intellectual disability (IQ 53 at age of 8 years), and was emotionally unstable and transiently aggressive during adolescence. From the age of 15 he had severe generalised dystonia. He is short stature (160cm, 3rd percentile), normal weight. He has a long philtrum, thick lips, long nose with broad tip and has excessive sweating. He has relative macrocephaly (head circumference in 75th centile), but MRI at age 19 years was normal. Opthalmic and audiology exams were normal. Examinations of skin, appendages, genetalia and dermatoglyphs were normal. Echocardiography revealed diffuse hypocontractility of the left ventricle (ejection fraction 30-35%). An abdominal ultrasound revealed mild liver steatosis, but otherwise normal.
	Lab Tests	Molecular Genetics: Has returned normal results for screening of karyotype, *ARX* sequencing, subtelomeric rearangements (FISH), 22q11.2 deletion (FISH), SNP array (300K), mtDNA mutation analysis and *TOR1A* mutation analysis. Metabolic Screening: increased homocysteine levels (21.0 and 14.5 µmol/l (sampled in June and October of 2011 respectively; normal range 3.5-15.0µmol/l)); slightly elevated folic acid (612.7µg/l (normal range 176-589µg/l)); holotranscobalamin 55pmol/l (normal range 19-119), the level of plasma and urine methionin and urine methylmalonate was repeatedly normal; lower urine pterin excretion (both total biopterin (0.40 or 0.3.2mmol/mol of creatine) and neopterin (0.21 or 0.12 mmol/mol of creatine)); lysosomal storage disorder screening was normal, and electron microscopic analysis of deep skin biopsy was normal, however vacuolisation of peripheral blood lymphocytes was found. Neurotransmitter screening: normal levels of vanilalanin, vanilglykol, 5-hydroxyindolacetic acid (HIAA) and vanilacetic acid (HVA) were found in cerebrospinal fluid.

Supplementary Table 1. Clinical descriptions of affected brothers in family CZE168.

Gene	Accession	Variant	Comment
AFF2	NM_001170628	c.G1090A; p.D364N	Present in healthy maternal grandfather
AMOT	NM_001113490	c.C1400T; p.T4671I	Present in healthy maternal uncle
DPR2	NM_001171184	c.C784T; p.R262W	Present in healthy maternal grandmother;
MAGEC3	NM_177456	c.G347A; p.S116N	Present in healthy maternal grandfather
HCFC1	NM_005334	c.C6046T; p.R2016W	De-novo in mother

Supplementary Table 2. Results of whole exome sequencing of family CZE168. No candidate genes were identified under an autosomal recessive model. Five candidate genes were identified using an X-linked model. Note that truncation mutations in *DPR2* have been identified in healthy male subjects (Tarpey *et al.,* 2009, *Nat Genet*. 41(5):535-43).



	p.Gln68	p.Ala73	p.Tyr103	p.Ala115	p.Ser225	p.Ala477	p.Gly876	p.Ala1756	p.Arg2016
Number of Cases	1	3	2	10	4	1	1	1	2
Affect on HCFC1 function	ND	LOF (+++)	ND	LOF (+++)	LOF (+)	ND	LOF (+)	Normal	LOF (++)
Additional Genes with Potential Contribution	-	-	-	-	-	-	ARID1B MEF2	ZMYM3	-
Developmental Delay	Yes (?)	Severe (3/3)	Severe (1/1)	Severe (8/8)	Yes (2/2)	Severe	Yes	Yes	Yes (2/2)
ID	ND	ND	ND	ND	Mild (2/2)	Severe	Yes	Mild	Mild (1/2) Moderate (1/2)
Seizures	yes	Epilepsy (3/3)	Epilepsy (1/1)	Epilepsy (6/9) IS (3/9)	No (2/2)	Epilepsy	Yes	No	Yes
Movement Disorder	No	Choreoathetosis (1/3)	No (1/1)	Choreoathetosis (4/10)	No	Dystonia	No	No	No
Muscle Tone	Normal	Hypotonic (1/3)	Hypotonic (1/1)	Hypotonic (2/10)	Normal	Normal	Hypotonic	Normal	Normal
Head Size	Microcephaly	Normal	Microcephaly (2/2)	Microcephaly (6/10)	Normal	Normal	Macrocephaly	Normal	Macrocephaly (2/2)
Plasma Homocysteine	Elevated	Elevated (1/3)	Elevated (1/1)	Elevated (4/8)	ND	ND	ND	ND	Transiently Elevated (2/2)
Urine Methymalonic Acid	Elevated	Elevated 3/3)	Elevated (1/1)	Elevated (7/7)	ND	Normal	ND	ND	Normal
Reference	Yu *et al.* 2013	Yu *et al.* 2013	Gerard *et al.* 2015	Yu *et al.* 2013	Huang *et al.* 2012	Current study	Huang *et al.* 2012	Huang *et al.* 2012	Current study

Supplementary Table 3. Overview of clinical spectrum resulting from coding variants in HCFC1 associated with neurodevelopmental disorders. ND: not described; LOF: loss of function; +: partial; ++: moderate; +++: high degree of affect; (?): severity not described.

Publikace 10

A patient showing features of both SBBYSS and GPS supports the concept of a *KAT6B* related disease spectrum, with mutations in mid-exon 18 possibly leading to combined phenotypes

Vlckova M., Simandlova M., Zimmermann P., Stranecky V., Hartmannova H., Hodanova K., Havlovicova M., Hancarova M., Sedlacek Z., Kmoch S.

Eur J Med Genet. 2015 Oct;58(10):550-5. IF: 1,446

V kazuistice představujeme pacientku s potvrzenou *de novo* mutací v genu *KAT6B* s komplexním fenotypem, který vykazuje známky dvou vzácných syndromů – genitopatelárního syndromu (GPS) a Say-Barber-Biesecker-Young-Simpsonova syndromu (SBBYSS). V práci jsme shrnuli podrobné klinické údaje o všech dosud publikovaných pacientech s mutacemi v genu *KAT6B*. Tato data jsme následně podrobili clusterové analýze. Cílem práce bylo jednak rozhodnout, zda se v případě GPS a SBBYSS skutečně jedná o dvě klinicky odlišné jednotky, či zda se jedná o jednu jednotku s variabilním fenotypem, a také ověřit, zda a jak je fenotyp pacientů závislý na lokalizaci mutace v rámci genu *KAT6B*.

GPS a SBBYSS jsou klinicky dobře definované vzácné syndromy, které vykazují určitý fenotypový překryv. Teprve nedávno odhalenou genetickou podstatou těchto onemocnění jsou mutace v genu *KAT6B*, který kóduje lysin acetyltransferázu 6B, která je součástí komplexu histon-H3 acetyltransferáz. *KAT6B* je vysoce konzervovaný gen, který hraje důležitou roli v neurogenezi (Merson et al., 2006). Většina dosud identifikovaných mutací byla ztrátových (zejména nonsense), vznikla *de novo* a byla lokalizována v exonu 18. Exon 18 kóduje acidickou (A) doménu a doménu aktivující transkripci (TA).

Již z předchozích prací vyplývalo, že fenotyp pacientů je do značné míry závislý na lokalizaci mutace. Doposud publikované případy se nám podařilo nově rozdělit do 4 skupin v závislosti na lokalizaci mutace a na výsledném fenotypu. Pacienti s mutacemi lokalizovanými na 3´ konci exonu 18 vedoucími ke ztrátě TA domény a mutacemi mimo exon 18 (s výjimkou pacienta 11) mají klinickou symptomatologii typickou pro SBBYSS, pacienti s mutacemi lokalizovanými na 5´ konci exonu 18 (zhruba v první šestině), které vedou ke ztrátě obou domén, vykazují fenotyp GPS. Skupina pacientů s mutacemi lokalizovanými ve zhruba druhé šestině exonu 18 je fenotypicky nejméně konzistentní. V této skupině jsou jak pacienti s klinickou diagnózou GPS, tak SBBYSS a dále několik pacientů s intermediárním fenotypem,

u kterých jsou příznaky obou syndromů zastoupeny přibližně ve stejné míře. Do této skupiny spadá i naše pacientka, která jako jedna z mála doposud publikovaných případů vykazuje rovněž intermediární fenotyp.

Naše výsledky podporují domněnku, že v případě GPS a SBBYSS se skutečně jedná o dvě různé klinické jednotky a že toto rozdělení není dáno pouze historicky. Naše práce také podporuje a dále zpřesňuje předchozí pozorování, že fenotyp pacientů do značné míry závisí na lokalizaci mutace. Rozdělení genu na segmenty, které se liší fenotypovým projevem v nich lokalizovaných mutací, pravděpodobně koreluje s účinností mechanismu nonsense-mediated mRNA decay (NMD) u jednotlivých mutovaných transkriptů a také se zastoupením funkčních domén ve zkráceném proteinu, pokud je produkován.

Naše hypotéza získala další podporu identifikací slovenské pacientky s intermediárním fenotypem, která je v současné době šestým celosvětově popsaným případem intermediárního fenotypu, a jejíž mutace je také lokalizována v druhé šestině exonu 18 (J. Radvanszky, abstrakt 26. Izakovičov memoriál 2015, rukopis v přípravě).

Publikace z prosince 2015 popisuje další 3 pacienty s již dříve popsanou synonymní záměnou 3147G>A, p.(Pro1049Pro) v exonu 16. Popsaní pacienti mají fenotyp odpovídající SBBYSS. Patogenita této varianty nebyla v době naší publikcace zcela jednoznačná, dle autorů výše uvedené práce tato varianta ovlivňuje sestřih a vzhledem k relativně častému výskytu je považována za hotspot (Yilmaz et al., 2015).

European Journal of Medical Genetics 58 (2015) 550–555

Contents lists available at ScienceDirect

European Journal of Medical Genetics

journal homepage: http://www.elsevier.com/locate/ejmg

Clinical report

A patient showing features of both SBBYSS and GPS supports the concept of a *KAT6B*-related disease spectrum, with mutations in mid-exon 18 possibly leading to combined phenotypes

Marketa Vlckova [a, *], Martina Simandlova [a], Pavel Zimmermann [b], Viktor Stranecky [c], Hana Hartmannova [c], Katerina Hodanova [c], Marketa Havlovicova [a], Miroslava Hancarova [a], Stanislav Kmoch [c], Zdenek Sedlacek [a]

[a] Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic
[b] Department of Statistics and Probability, Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic
[c] Institute of Inherited Metabolic Disorders, Charles University 1st Faculty of Medicine and General University Hospital, Prague, Czech Republic

ARTICLE INFO

Article history:
Received 28 May 2015
Received in revised form
31 August 2015
Accepted 10 September 2015
Available online 11 September 2015

Keywords:
KAT6B
Genitopatellar syndrome
Say-Barber-Biesecker-Young-Simpson syndrome
Genotype-phenotype correlation

ABSTRACT

Genitopatellar syndrome (GPS) and Say-Barber-Biesecker-Young-Simpson syndrome (SBBYSS) are two distinct clinically overlapping syndromes caused by *de novo* heterozygous truncating mutations in the *KAT6B* gene encoding lysine acetyltransferase 6B, a part of the histone H3 acetyltransferase complex. We describe an 8-year-old girl with a *KAT6B* mutation and a combined GPS/SBBYSS phenotype. The comparison of this patient with 61 previously published cases with *KAT6B* mutations and GPS, SBBYSS or combined GPS/SBBYSS phenotypes allowed us to separate the *KAT6B* mutations into four groups according to their position in the gene (reflecting nonsense mediated RNA decay and protein domains) and their clinical outcome. We suggest that mutations in mid-exon 18 corresponding to the C-terminal end of the acidic (Asp/Glu-rich) domain of KAT6B may have more variable expressivity leading to GPS, SBBYSS or combined phenotypes, in contrast to defects in other regions of the gene which contribute more specifically to either GPS or SBBYSS. Notwithstanding the clinical overlap, our cluster analysis of phenotypes of all known patients with *KAT6B* mutations supports the existence of two clinical entities, GPS and SBBYSS, as poles within the *KAT6B*-related disease spectrum. The awareness of these phenomena is important for qualified genetic counselling of patients with *KAT6B* mutations.

© 2015 Elsevier Masson SAS. All rights reserved.

1. Introduction

Genitopatellar syndrome (GPS, OMIM #606170) and Say-Barber-Biesecker-Young-Simpson syndrome (SBBYSS, OMIM #603736, Ohdo syndrome variant) are two syndromes with distinct phenotypes; nevertheless, a clinical overlap between GPS and SBBYSS has repeatedly been noted. For differential diagnosis the presence of major and minor features can be used (Campeau and Lee, 2013). GPS is a rare skeletal dysplasia characterised mainly by corpus callosum agenesis, microcephaly, specific anomalies of external genital and kidney, hypoplastic or absent patellae and flexion deformities of the limbs. Minor features including congenital heart defects, hearing impairment, hypotonia and thyroid, dental and anal anomalies can also be present. SBBYSS is a subtype of blepharophimosis - intellectual disability syndromes (Verloes et al., 2006). Characteristic major features include long thumbs, long great toes, lacrimal duct abnormalities, patellar hypoplasia or agenesis and distinct facial dysmorphism (blepharophimosis, ptosis, expressionless/mask-like face, bulbous nasal tip and small mouth with thin upper lip). Minor features include hypotonia, thyroid abnormalities, hearing impairment, dental anomalies, cryptorchidism and cleft palate. Global developmental delay, intellectual disability (usually severe) and feeding problems are common to both syndromes. Previous reports showed that the clinical overlap is usually more apparent in the minor features and less in the major ones (reviewed in Gannon et al., 2015).

KAT6B variants, predominantly heterozygous truncating *de novo* mutations, have been identified in SBBYSS (Clayton-Smith et al.,

* Corresponding author. Charles University 2nd Faculty of Medicine and University Hospital Motol, V Uvalu 84, 15006 Prague 5, Czech Republic.
E-mail address: marketa.vlckova@lfmotol.cuni.cz (M. Vlckova).

http://dx.doi.org/10.1016/j.ejmg.2015.09.004
1769-7212/© 2015 Elsevier Masson SAS. All rights reserved.

2011) and later also in GPS (Simpson et al., 2012; Campeau et al., 2012a). The *KAT6B* gene is expressed in adult neural stem cells and encodes lysine acetyltransferase 6B, a part of the histone H3 acetyltransferase complex and a member of the MYST protein family. Up to now, 42 SBBYSS and 15 GPS patients with *KAT6B* mutations have been reported, and 4 additional patients with *KAT6B* mutations and a significant clinical overlap between the two syndromes (combined phenotypes) have been identified recently (reviewed in Gannon et al., 2015). In addition, a group of GPS and SBBYSS patients exists in whom no *KAT6B* mutations could be identified (Gannon et al., 2015). The majority (52/61) of *KAT6B* mutations were in exon 18 encoding the acidic (A, Asp/Glu-rich) and transcriptional activation (TA, Ser/Met-rich) domains of KAT6B (Clayton-Smith et al., 2011; Simpson et al., 2012; Campeau et al., 2012a; Szakszon et al., 2013; Yu et al., 2014; Gannon et al., 2015). Only 3 variants were not truncating: p.Trp987Arg, p.Pro1049Pro and p.Glu1367_Glu1368del (Clayton-Smith et al., 2011; Gannon et al., 2015); however, the recurrent p.Pro1049Pro variant was predicted to alter splicing. Truncating mutations in exon 18 leading to loss of both domains have been suggested to cause GPS. Truncating mutations preserving the A domain may cause SBBYSS, similarly to mutations proximal to exon 18 (Campeau et al., 2012a; Szakszon et al., 2013; Gannon et al., 2015). Campeau et al. (2012b) suggested that gain-of-function *KAT6B* mutations lead to symptoms unique for GPS while haploinsufficiency or loss-of function may cause features common to both conditions, with the most distal mutations leading to symptoms specific for SBBYSS.

In this report we describe a patient with a *de novo* truncating *KAT6B* mutation who showed phenotypic features of both GPS and SBBYSS. We suggest that mutations in mid-exon 18 corresponding to the C-terminal end of the A domain of KAT6B may have more variable expressivity leading to GPS, SBBYSS or combined phenotypes, in contrast to mutations in three other regions of the gene which contribute more specifically to either GPS or SBBYSS. We also show that clustering of patients with *KAT6B* mutations based on their phenotypic features but blinded towards their diagnoses identifies two clusters which correlate with the classification of the patients as GPS or SBBYSS. These two observations support the existence of two different clinical entities associated with the *KAT6B* gene.

2. Clinical report

The girl was born as the first child of healthy unrelated Czech parents. The family history was unremarkable. At birth, the mother and father were 33 and 34 years old, respectively. The pregnancy was uneventful and the girl was delivered spontaneously in the 42nd week of gestation. Her weight was 3430 g (50–75th centile) and length was 51 cm (50–75th centile). Due to tachypnea, hypotonia, seizures and dysmorphic features she underwent several examinations including abdominal ultrasound and brain MRI scan. An atrial septal defect, renal cysts, nephrolithiasis and corpus callosum agenesis were detected. The ophthalmological examination revealed lacrimal ducts stenosis and optic atrophy. Developmental delay, feeding difficulties, recurrent urinary tract infections, laryngeal stridor, sleep apnea and contractures occurred during infancy. At the age of 8 years the height of the patient was 138 cm (90th centile), weight was 24 kg (below the 3rd centile) and head circumference was 52 cm (25th centile). Remarkable dysmorphic features (Fig. 1) included expressionless face, blepharophimosis, ptosis, bulbous nose with flat nasal bridge, thin upper lip, and low-set and posteriorly rotated dysplastic ears. Long overlapping toes and contractures of knees were also present. Psychological examination showed moderate intellectual disability and autistic features. The features of the patient are summarized in Table 1 where her phenotype is put in context with other published cases.

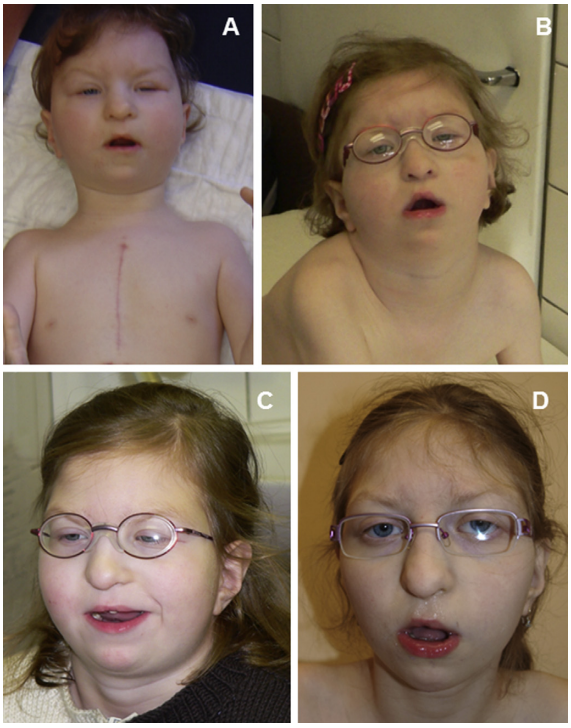


Fig. 1. Phenotype of the patient. Facial photographs taken at the age of 7 months (A), 3 years (B), 4 years (C) and 10 years (D) show the open mouth appearance, hypomimia, low set ears and ptosis; blepharophimosis and bulbous nose are most apparent in pictures A and D, respectively.

Karyotyping of the patient yielded normal results. High-resolution array CGH analysis of the patient using the Nimblegen Human Whole-Genome 385K Array yielded normal results. Whole exome sequencing performed as previously described (Park et al., 2014) on the captured barcoded DNA library (SeqCap EZ Human Exome Library v3.0, Roche Nimblegen) using SOLiD 4 System (Applied Biosystems) and followed by Sanger sequencing of the family trio showed a *de novo* heterozygous truncating *KAT6B* mutation in the patient (*KAT6B*: NM_012330.3:exon18: c.4171G > T:p(E1391X)) (Fig. 2).

3. Discussion

GPS and SBBYSS are very rare but clinically well-described conditions associated with *KAT6B* mutations. Although clinical overlap between these two syndromes has repeatedly been noted, only four patients with a significantly combined GPS/SBBYSS phenotype have been described (Gannon et al., 2015). We describe another such patient showing symptoms frequent in GPS but rare in SBBYSS, symptoms common to both syndromes as well as symptoms typical for SBBYSS (Table 1), who carried a truncating mutation in the proximal part of *KAT6B* exon 18 close to the boundary between the regions encoding the A and TA domains. The phenotype of our patient further supports the broad phenotypic overlap between GPS and SBBYSS, justifying the recently formulated concept of *KAT6B* spectrum (Gannon et al., 2015).

A detailed comparison of this atypical case with previous

Table 1
Symptoms observed in the current patient in the context of their frequency in published cases with *KAT6B* mutations.

	Our case (GPS/SBBYSS)	Published GPS cases		Published SBBYSS cases		Published combined GPS/SBBYSS cases	
	Feature present	Fraction	%	Fraction	%	Fraction	%
Features commonly observed in both syndromes							
DD/ID	Yes	14(15)/15	93(100)%	41(42)/42	97(100)%	4(4)/4	100(100)%
Congenital heart defect	Yes	11(11)/15	73(73)%	21(23)/42	50(55)%	4(4)/4	100(100)%
Feeding difficulties	Yes	7(13)/15	47(87)%	38(41)/42	90(98)%	2(3)/4	50(75)%
Genital anomalies	No	13(13)/15	87(87)%	20(20)/42	48(48)%	4(4)/4	100(100)%
Average % score*	75.0%	75.0(86.6)%		71.4(75.0)%		87.5(93.7)%	
Features more frequently observed in GPS							
Corpus callosum agenesis	Yes	14(15)/15	93(100)%	3(28)/42	7(67)%	2(4)/4	50(100)%
Microcephaly	No	11(13)/15	73(87)%	8/(23)42	19(55)%	2(2)/4	50(50)%
Micrognathia	No	4(14)/15	27(93)%	1(42)/42	2(100)%	0(4)/4	0(100)%
Absent or hypoplastic patellae	Yes	15(15)/15	100(100)%	10(14)/42	24(33)%	1(1)/4	25(25)%
Flexion contractures	Yes	15(15)/15	100(100)%	19(36)/42	45(86)%	2(2)/4	50(50)%
Club foot	No	10(14)/15	67(93)%	0/(42)42	0(100)%	0(4)/4	0(100)%
Skeletal anomalies (axial skeleton)**	No***	7(12)/15	44(80)%	0(42)/42	0(100)%	0(4)/4	0(100)%
Renal anomalies	Yes	13(13)/15	87(93)%	4(21)/42	10(50)%	1(3)/4	25(75)%
Laryngo/tracheomalacia	Yes	2/(11)15	13(73)%	1(42)/42	2(100)%	0(4)/4	0(100)%
Small bowel malrotation	No	2(10)/15	13(67)%	1(42)/42	2(100)%	0(4)/4	0(100)%
Anteriorly positioned anus	No	2(11)/15	13(73)%	0/(42)42	0(100)%	0(4)/4	0(100)%
Anal atresia or stenosis	No	2(11)/15	13(73)%	0(42)/42	0(100)%	0(4)/4	0(100)%
Average % score*	41.7%	53.6(86.1)%		9.3(82.5)%		16.7(83.3)%	
Features more frequently observed in SBBYSS							
Hypotonia	Yes	3(15)/15	20(100)%	39(40)/42	93(95)%	2(3)/4	50(75)%
Blepharophimosis	Yes	0(15)/15	0(100)%	3(42)/42	7(100)%	0(4)/4	0(100)%
Ptosis	Yes	0(15)/15	0(100)%	2(42)/42	5(100)%	0(4)/4	0(100)%
Expressionless facies	Yes	0(15)/15	0(100)%	1(42)/42	2(100)%	0(4)/4	0(100)%
Typical SBBYSS face	Yes	0(13)/15	0(87)%	38(42)/42	90(100)%	4(4)/4	100(100)%
Long thumbs	Yes	4(15)/15	27(100)%	31(32)/42	74(76)%	1(2)/4	25(50)%
Long toes	Yes	2(13)/15	13(87)%	27(31)/42	64(74)%	2(2)/4	50(50)%
Lacrimal ducts stenosis	Yes	0(15)/15	0(100)%	8(37)/42	19(88)%	0(4)/4	0(100)%
Dental anomalies	No	1(14)/15	7(93)%	23(30)/42	55(71)%	3(4)/4	75(100)%
Cleft palate	No	1(12)/15	7(80)%	12(15)/42	31(36)%	1(1)/4	25(25)%
Hearing impairment	No	2(11)/15	13(73)%	10(36)/42	24(86)%	0(4)/4	0(100)%
Tyhroid anomalies	No	4(10)/15	27(67)%	19(20)/42	45(48)%	3(3)/4	75(75)%
Average % score*	66.7%	9.4(90.5)%		42.2(81.2)%		33.3(81.2)%	

* Average % score is calculated as arithmetic average of % occurrence of symptoms listed in the group. ** Skeletal anomalies include pelvic and costovertebral anomalies and thoracolumbal kyphosis/scoliosis. *** No remarkable skeletal anomalies were observed; X-ray examination was not performed. The numbers in parentheses indicate patient numbers and percentages with inclusion of patients in whom the symptoms were not reported explicitly to be present or absent (missing information, "NK" in Supplementary Table).

patients allowed us to further elaborate the genotype–phenotype correlation in individuals with *KAT6B* defects located at different positions within the gene. A careful analysis of our patient together with 61 previously published patients (42 SBBYSS, 15 GPS and 4 combined SBBYSS/GPS cases, Supplementary Table) suggests that the currently known *KAT6B* mutations can be separated into four groups (Fig. 3, Table 2). Group 1 mutations are located between codons 1 and 1205, with the majority clustering in the region corresponding to the proximal 1/3 of the A domain (in exons 15, 16 and most of exon 17 except its most distal part). As only premature termination codons located proximal to the last 50–55 bases of the penultimate exon cause nonsense-mediated RNA decay (NMD) (Nagy and Maquat, 1998), truncating Group 1 mutations are the only variants expected to cause NMD and KAT6B haploinsufficiency. Group 1 mutations have been associated only with SBBYSS and average % scores indicate strongly prevailing SBBYSS features (Table 2). Group 2 mutations are located approximately between codons 1205 and 1350, and are expected to produce truncated KAT6B lacking a significant part (more than 1/3) of the A domain and the whole TA domain. These mutations have been associated exclusively with GPS and this is correlated with high average % scores for GPS features. Group 3 mutations located between codons 1350 and 1520 can result in truncated proteins retaining a larger part of the A domain but lacking the TA domain. They have been reported to cause a mixture of phenotypes ranging from typical GPS across combined GPS/SBBYSS to typical SBBYSS. Our patient carries one of the Group 3 mutations. Average % scores indicate that these patients often show symptoms typical for both conditions. Finally, Group 4 mutations are located between codons 1520 and 2073 in the region corresponding to the TA domain and can lead to proteins retaining the whole A domain and lacking different, often significant parts of the TA domain. These mutations cause almost exclusively SBBYSS with the exception of one case of GPS/SBBYSS. The latter patient (individual 48 from Fig. 3 and Supplementary Table) was classified as combined (Gannon et al., 2015) but according to our scoring system he is closer to SBBYSS (average % scores for GPS and SBBYSS features of 8.3% and 41.7%, respectively; compare with Table 1). The average % scores for SBBYSS features in this group are the highest of all groups, in accord with the notion that the association with symptoms specific for SBBYSS is most pronounced in the most distal *KAT6B* mutations (Campeau et al., 2012b). Our analysis seems to indicate that long fingers and heart defects might be more common among patients with Group 4 mutations.

The boundary between Group 1 and 2 mutations sharply separates variants causing SBBYSS and GPS, which may reflect a relatively well-defined site determined by exon structure and rules governing NMD. On the contrary, the interval of Group 3 mutations could represent a transition region where the *KAT6B* variants can lead to the retention of significant parts of the A domain and can have a more variable expressivity, causing combined phenotypes which can be diagnosed as such or promiscuously as either GPS or

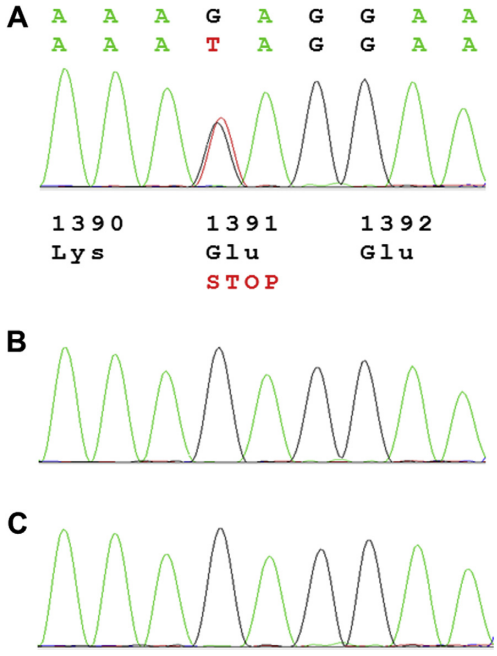


Fig. 2. Sanger sequencing of the region around the *KAT6B* mutation in the patient (A) and her mother (B) and father (C) clearly shows the *de novo* nature of the variant.

SBBYSS. The ambiguity in this region is likely enhanced by phenotype heterogeneity which can be expected in *KAT6B* disorders similarly to other genetic conditions, and also by limited consistency among phenotype descriptions published by different teams.

The observation of correlation between the position of mutation in the gene and diagnosis given to the patient supports the notion of GPS and SBBYSS as two distinct entities and not just a relict of the historical developments. We also performed cluster analysis of all 61 patients with *KAT6B* mutations based on their phenotypic features but blinded towards their diagnoses (Supplementary Text, Fig. 4). Two clusters could be observed which correlated with the original diagnoses of the patients. Cluster 1 contained 15 patients, of whom 13 were diagnosed as GPS, one as a combined GPS/SBBYSS phenotype and one as SBBYSS. Cluster 2 contained 46 patients, of whom 40 were diagnosed as SBBYSS, four as combined phenotypes and two as GPS. The analysis shows that patients with combined phenotypes do not form a specific group but rather differ in different features from the typical GPS or SBBYSS phenotype. Interestingly, one of the two GPS patients in Cluster 2, Patient 11, carried a Group 2 mutation located very close to the boundary from where NMD is expected to be inactive, pointing to a possible transition zone similar to the region of Group 3 mutations where the phenotype can be atypical or combined. This patient was also the only patient with a Group 2 mutation in Cluster 2. The other GPS patient in Cluster 2, Patient 26, carried a Group 3 mutation which can be associated with combined phenotypes. The significance of the fact that the only SBBYSS patient in Cluster 1, Patient 47, had a mutation located very close to the mutation of Patient 48, the only patient with a combined phenotype and a Group 4 mutation, is unclear.

Our analysis extends and refines previous observations that phenotype correlates with the position of *KAT6B* mutations, and supports the notion that although a significant clinical overlap exists between GPS and SBBYSS which can even lead to combined GPS/SBBYSS phenotypes, two distinct poles exist within the *KAT6B* spectrum which correspond to the historical diagnostic entities. This heterogeneity can make the clinical diagnosis of GPS and SBBYSS more difficult, and efforts to reach a conclusive diagnosis can contribute to the current low rate of reporting of combined GPS/SBBYSS cases. The awareness of these phenomena is therefore important for qualified genetic counselling and management of patients with *KAT6B* mutations.

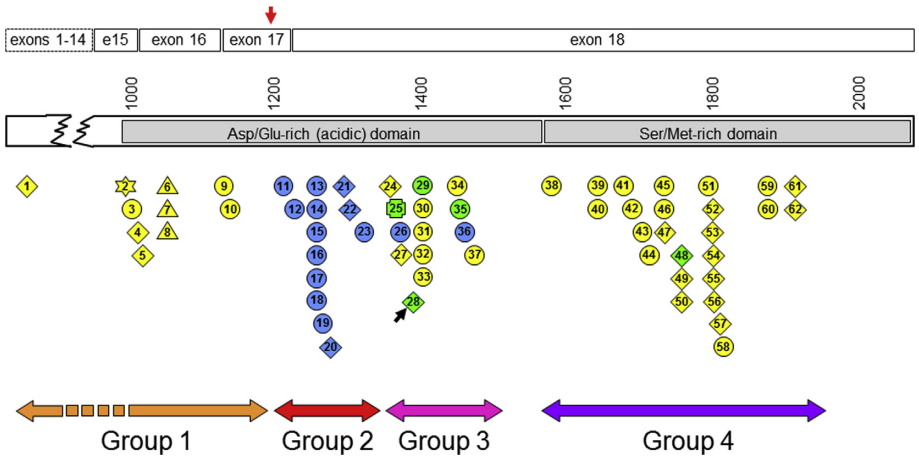


Fig. 3. Schematics showing the distribution and type of *KAT6B* mutations in patients with SBBYSS (yellow), GPS (blue) and combined GPS/SBBYSS phenotypes (green). Diamonds, nonsense mutations; circles, frameshift mutations; triangles, missense mutations affecting splicing; cross, in-frame deletion; star, missense mutation. The numbering of cases corresponds to that in the Supplementary Table. The mutation in our patient is marked by a black arrow. The bars on the top show exon boundaries in *KAT6B* transcript NM_012330.3, codon numbering and protein domains. The position in exon 17 downstream of which nonsense mediated RNA decay is likely to be inactive on transcripts with premature termination codons is labelled by a red vertical arrow. The grouping of mutations discussed in the text is indicated by horizontal double arrows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Clinical diagnosis and symptom frequency in patients with Group 1–4 *KAT6B* mutations (including the current patient).

	Group 1		Group 2		Group 3		Group 4	
	Fraction	%	Fraction	%	Fraction	%	Fraction	%
Clinical diagnosis								
GPS	0/10	0%	13/13	100%	2/14	14%	24/25	96%
SBBYSS	10/10	100%	0/13	0%	8/14	57%	0/25	0%
GPS/SBBYSS	0/10	0%	0/13	0%	4/14	29%	1/25	4%
Features commonly observed in both syndromes								
DD/ID	10(10)/10	100(100)%	13(13)/13	100(100)%	13(14)/14	93(100)%	24(25)/25	96(100)%
Congenital heart defect	2(2)/10	20(20)%	10(10)/13	77(77)%	10(11)/14	71(78)%	15(16)/25	60(64)%
Feeding difficulties	9(10)/10	90(100)%	6(11)/13	46(85)%	11(13)/14	78(93)%	22(24)/25	88(96)%
Genital anomalies	5(5)/10	50(50)%	12(12)/13	92(92)%	9(9)/14	64(64)%	11(11)/25	44(44)%
Average % score*	65.0(67.5)%		78.8(88.5)%		76.8(83.9)%		72.0(76.0)%	
Features more frequently observed in GPS								
Corpus callosum agenesis	0(5)/10	0(50)%	12(13)/13	92(100)%	4(10)/14	28(71)%	4(18)/25	16(72)%
Microcephaly	2(4)/10	20(40)%	10(11)/13	77(85)%	8(9)/14	57(64)%	1(14)/25	4(96)%
Micrognathia	0(10)/10	0(100)%	4(12)/13	31(92)%	0(13)/14	0(93)%	1(25)/25	4(100)%
Absent or hypoplastic patellae	1(1)/10	10(10)%	13(13)/13	100(100)%	7(7)/14	50(50)%	6(10)/25	24(40)%
Flexion contractures	4(7)/10	40(70)%	13(13)/13	100(100)%	9(13)/14	64(93)%	11(21)/25	44(84)%
Club foot	0(10)/10	0(100)%	9(12)/13	69(92)%	1(13)/14	7(93)%	0(25)/25	0(100)%
Skeletal anomalies (axial skeleton)**	0(10)/10	0(100)%	7(11)/13	54(85)%	0(12)/14	0(86)%	0/(25)25	0(100)%
Renal anomalies	2(4)/10	20(40)%	11(12)/13	85(92)%	4(8)/14	28(57)%	2(15)/25	8(60)%
Laryngo/tracheomalacia	0(10)/10	0(100)%	1(9)/13	8(69)%	2(14)/14	14(100)%	1(25)/25	4(100)%
Small bowel malrotation	1(10)/10	10(100)%	2(9)/13	15(69)%	0(12)/14	0(86)%	0(25)/25	0(100)%
Anteriorly positioned anus	0(10)/10	0(100)%	2(10)/13	15(77)%	0(12)/14	0(86)%	0(25)/25	0(100)%
Anal atresia or stenosis	0(10)/10	0(100)%	2(10)/13	15(77)%	0(12)/14	0(86)%	0(25)/25	0(100)%
Average % score*	8.3(75.8)%		55.1(86.5)%		20.7(80.3)%		8.7(87.7)%	
Features more frequently observed in SBBYSS								
Hypotonia	9(9)/10	90(90)%	2(13)/13	15(100)%	9(12)/14	64(86)%	24(25)/25	96(100)%
Typical SBBYSS face	9(10)/10	90(100)%	0(11)/13	0(85)%	12(14)/14	86(100)%	22(25)/25	88(100)%
Long thumbs	4(4)/10	40(40)%	3(13)/13	23(100)%	11(13)/14	78(93)%	19(20)/25	76(80)%
Long toes	5(5)/10	50(50)%	1(11)/13	8(85)%	11(13)/14	78(93)%	15(19)/25	60(76)%
Lacrimal ducts stenosis	1(9)/10	10(90)%	0(13)/13	0(100)%	3(12)/14	21(86)%	5(23)/25	20(92)%
Dental anomalies	3(3)/10	30(30)%	1(12)/13	8(92)%	7(13)/14	50(93)%	16(20)/25	64(80)%
Cleft palate	1(1)/10	10(10)%	1(11)/13	8(85)%	3(4)/14	21(28)%	9(12)/25	40(48)%
Hearing impairment	1(9)/10	10(90)%	1(9)/13	8(69)%	3(11)/14	21(78)%	7/(22)25	28(88)%
Tyhroid anomalies	3(4)/10	30(40)%	4(9)/13	31(69)%	6(7)/14	43(50)%	13(13)/25	52(52)%
Average % score*	40.0(60.0)%		11.9(87.2)%		51.6(78.6)%		57.8(79.6)%	

See Table 1 for legend.

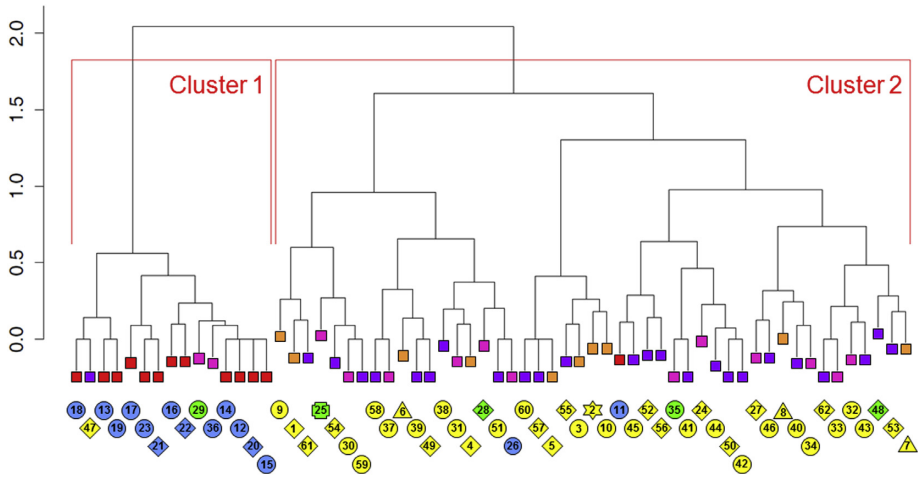


Fig. 4. Cluster dendrogram constructed based on phenotypic features of the patients. See Supplementary Text for details. The inclusion of patient's mutations into Groups 1–4 is indicated by coloured squares (orange, Group 1; red, Group 2; pink, Group 3; violet, Group 4; as in Fig. 3). Individual patients, their diagnoses and mutation types are marked by the same symbols as in Fig. 3. The vertical axis displays the distance at which clusters are formed. The two main clusters observed (Clusters 1 and 2) are indicated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Consent

Written consent was obtained from the parents of our patient for publication of this report.

Conflicts of interest

None.

Acknowledgements

We thank the family of the patient for cooperation, and two anonymous reviewers for helpful comments. The work was supported by grants 00064203, NT/14200, CZ.2.16/3.100/24022, NF-CZ11-PDP-3-003-2014, UNCE 20401, PRVOUK-P24/LF1/3 and SVV-UK 260148/2015 programs of the Charles University in Prague; by BIOCEV – Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University (CZ.1.05/1.1.00/02.0109) from the European Regional Development Fund and by grant NT13116-4/2012 from the Ministry of Health of the Czech Republic.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ejmg.2015.09.004.

References

Campeau, P.M., Kim, J.C., Lu, J.T., Schwartzentruber, J.A., Abdul-Rahman, O.A., Schlaubitz, S., et al., 2012a. Mutations in KAT6B, encoding a histone acetyltransferase, cause genitopatellar syndrome. Am. J. Hum. Genet. 90, 282–289.

Campeau P.M., Lee B.H.. KAT6B-Related Disorders. GeneReviews [Internet]. http://www.ncbi.nlm.nih.gov/books/NBK114806/ Last Revision: January 10, 2013; (accessed 26.08.15).

Campeau, P.M., Lu, J.T., Dawson, B.C., Fokkema, I.F., Robertson, S.P., Gibbs, R.A., et al., 2012b. The KAT6B-related disorders genitopatellar syndrome and Ohdo/SBBYS syndrome have distinct clinical features reflecting distinct molecular mechanisms. Hum. Mutat. 33, 1520–1525.

Clayton-Smith, J., O'Sullivan, J., Daly, S., Bhaskar, S., Day, R., Anderson, B., et al., 2011. Whole-exome-sequencing identifies mutations in histone acetyltransferase gene KAT6B in individuals with the Say-Barber-Biesecker variant of Ohdo syndrome. Am. J. Hum. Genet. 89, 675–681.

Gannon, T., Perveen, R., Schlecht, H., Ramsden, S., Anderson, B., Kerr, B., Day, R., et al., 2015. Further delineation of the KAT6B molecular and phenotypic spectrum. Eur. J. Hum. Genet. 23, 1165–1170.

Nagy, E., Maquat, L.E., 1998. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem. Sci. 23, 198–199.

Park, E.J., Grabińska, K.A., Guan, Z., Stranecky, V., Hartmannova, H., Hodanova, K., et al., 2014. Mutation of Nogo-B receptor, a subunit of cis-prenyltransferase, causes a congenital disorder of glycosylation. Cell Metab. 20, 448–457.

Simpson, M.A., Deshpande, C., Dafou, D., Vissers, L.E., Woollard, W.J., Holder, S.E., et al., 2012. De novo mutations of the gene encoding the histone acetyltransferase KAT6B cause genitopatellar syndrome. Am. J. Hum. Genet. 90, 290–294.

Szakszon, K., Salpietro, C., Kakar, N., Knegt, A.C., Olah, E., Dallapiccola, B., et al., 2013. De novo mutations of the gene encoding the histone acetyltransferase KAT6B in two patients with Say-Barber/Biesecker/Young-Simpson syndrome. Am. J. Med. Genet. A 161A, 884–888.

Verloes, A., Bremond-Gignac, D., Isidor, B., David, A., Baumann, C., Leroy, M.A., et al., 2006. Blepharophimosis-mental retardation (BMR) syndromes: a proposed clinical classification of the so-called Ohdo syndrome, and delineation of two new BMR syndromes, one X-linked and one autosomal recessive. Am. J. Med. Genet. A 140, 1285–1296.

Yu, H.-C., Geiger, E.A., Medne, L., Zackai, E.H., Shaikh, T.H., 2014. An individual with blepharophimosis–ptosis epicanthus inversus syndrome (BPES) and additional features expands the phenotype associated with mutations in KAT6B. Am. J. Med. Genet. A 164A, 950–957.

A patient showing features of both SBBYSS and GPS supports the concept of a KAT6B-related disease spectrum, with mutations in mid-exon 18 possibly leading to combined phenotypes

Marketa Vlckova, Martina Simandlova, Pavel Zimmermann, Viktor Stranecky, Hana Hartmannova, Katerina Hodanova, Marketa Havlovicova, Miroslava Hancarova, Stanislav Kmoch, Zdenek Sedlacek

Supplementary text

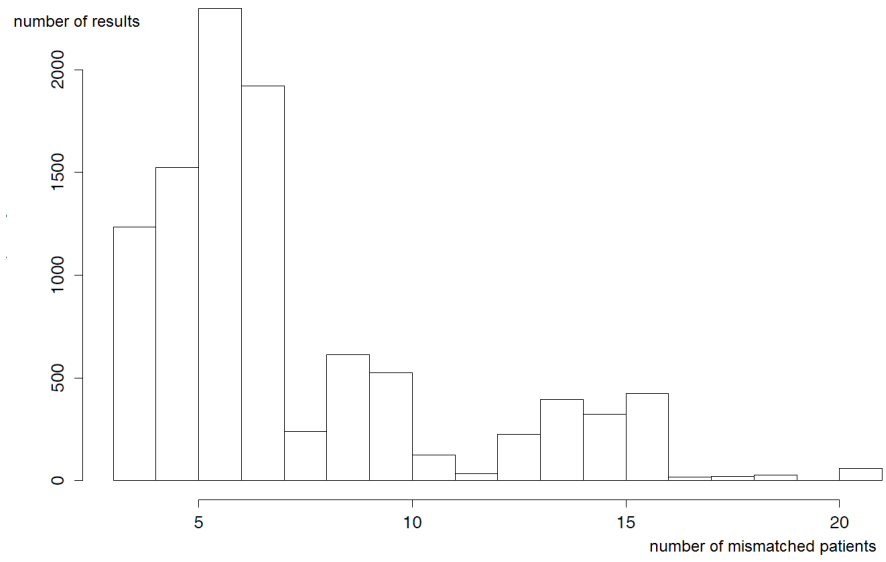
1. **Input data for cluster analyses**

Data available for the analysis were generally dichotomous with a high amount of missing observations. Prior to analysis, variables (symptoms) were pre-selected in order to exclude variables missing too often (more than 75 % missing data in the GPS or SBBYSS diagnosis groups; "NK (not known/not reported)", see Supplementary table), or which had only minimum variance observed (over 95 % of “y” or “n” observed; see Supplementary table). This resulted in selection of 13 variables for the quantification of dissimilarities among the patients:

- congenital heart defect
- feeding difficulties
- genital anomalies
- corpus callosum agenesis
- microcephaly
- absent or hypoplastic patellae
- flexion contractures
- renal anomalies
- long thumbs
- long toes
- cleft palate
- hearing impairment
- tyhroid anomalies

2. Cluster analyses

Values of “y” and “n” were treated symmetrically as all patients were already diagnosed with one of the two diagnoses, and both “y” and “n” were assumed informative to distinguish between the diagnoses in all variables (symptoms). For this reason, simple matching coefficient was selected as the similarity measure. The Ward’s method [Ward, 1963] applied in hierarchical clustering split the patients into two clusters which correlated very well with their diagnoses as GPS or SBBYSS. The limited amount of patients and the fact that clustering variables were dichotomous resulted in many ties in the distances measured between the patients. In such circumstances the results of clustering depend on the order of processing of the objects (patients). To avoid this influence, cluster analysis was performed 10 000 times and the match of the resulting clusters with the diagnoses (GPS or SBBYSS) was analysed. In all cases the amount of mismatches between the clusters and the diagnoses was rather limited. The best results achieved had only three mismatches and the median was six mismatches (Supplementary Fig. 1).



Supplementary Fig. 1. A histogram showing the distribution of results with different numbers of mismatched patients among the 10 000 repeated analyses

The worst results were achieved in cases where split up into two clusters was insufficient; however, split up into three clusters that could again be clearly matched with the diagnoses resulted in approximately five mismatches. Therefore we could conclude that no matter which order of patients was chosen, the clusters corresponding to the two diagnoses could be identified, and the match was very good (median of six mismatches among 57 patients with a clear-cut diagnosis of GPS or SBBYSS). The analysis of mismatches could also be used to identify patients who were most frequently mismatched and hence could be considered inconsistent with their original diagnoses. There were three problematic patients mismatched in most results (Patients 11, 26 and 47; see Supplementary table and Figs. 3 and 4). In results with more than three mismatches, these patients were accompanied by additional mismatched patients, who were however very variable among the different results. This observation further supported the existence of two diagnostic entities in patients with *KAT6B* mutations.

In order to characterise the resulting clusters we analysed the result displayed in Fig. 4. Intentionally we selected one of the dendrograms with the best match with the diagnosis as the three patients mismatched in this example are mismatched systematically in the majority of the 10 000 runs. If a more probable dendrogram with a higher amount of mismatched patients (e.g. six) was chosen, the additional more or less random mismatches would reduce the transparency of the output. For both clusters the ratio of "y" to all observed values ("y" plus "n") was calculated for all variables used in the analysis. This ratio in fact represented the centroid for each cluster. The two clusters differed if the difference in the ratio was close to 1 (much more "y" in Cluster 1 than in Cluster 2) or -1 (much more "n" in Cluster 1 than in Cluster 2). The highest absolute value of the difference was observed for the following variables (symptoms):

Cluster (~diagnosis)	corpus callosum agenesis	absent or hypoplastic patellae	renal anomalies
Cluster 1 (~GPS)	1.00	1.00	0.92
Cluster 2 (~SBBYSS)	0.30	0.28	0.27
Abs. difference	**0.70**	**0.72**	**0.65**

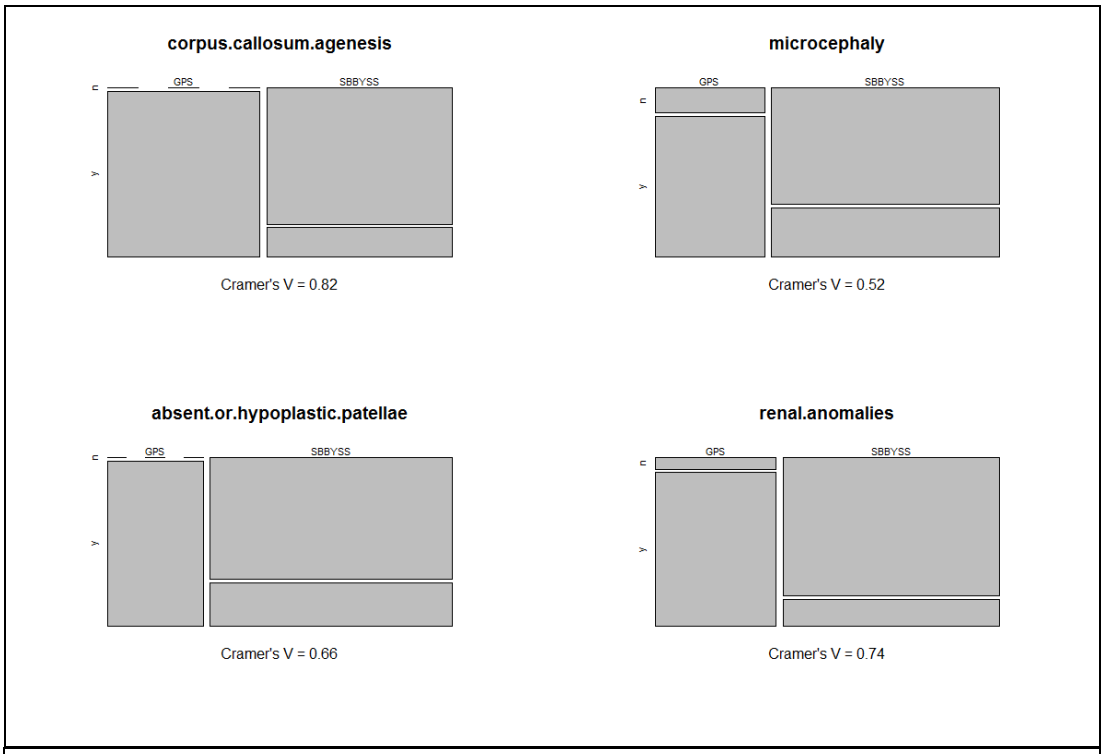
3. Further analyses

A. Group 3 *KAT6B* mutations

We observed that Group 3 *KAT6B* mutations located between codons 1350 and 1520 tend to cause a mixture of phenotypes ranging from typical GPS across combined GPS/SBBYSS clinical picture to typical SBBYSS (see the main article). Only 14 patients have been reported with Group 3 mutations, of whom two were diagnosed with GPS and eight with SBBYSS (and four with combined GPS/SBBYSS phenotypes). Yet the observations suggest a likely systematic difference between the patients from this group diagnosed with GPS and SBYSS. Specifically, corpus callosum agenesis and renal anomalies were observed in both patients diagnosed with GPS while these two symptoms were never observed in patients with SBBYSS. As the mutation Groups 1 and 4 correspond with the SBBYSS diagnosis, Group 2 corresponds with the GPS diagnosis and Group 3 can be split by one of the above mentioned symptoms, we can conclude that a combination of mutation group with one of these two symptoms may be considered a very good diagnostic tool.

B. Contingency of the symptoms with the diagnoses

This analysis was performed in order to reveal which symptoms could be the most important to distinguish between the two diagnoses. In other words contingency between the diagnoses and symptoms was measured. Only 13 variables selected for the cluster analyses were considered. The highest contingency of the diagnosis was observed with corpus callosum agenesis, microcephaly, absent or hypoplastic patellae and renal anomalies. The association measured with Cramer's V coefficient [Cramer, 1946] and the difference in ratios of "y" and "n" for the two diagnoses is shown in Supplementary Fig. 2.



Supplementary Fig. 2. Mosaic plots showing different ratios of the symptoms in different diagnoses. Cramer's V measure of association is displayed under the plots.

Supplementary references

Ward Jr, JH. "Hierarchical grouping to optimize an objective function." *Journal of the American Statistical Association* 58 (1963): 236-244.

Cramer, H. "Mathematical Methods of Statistics." (Princeton: Princeton UniversityPress, 1946).

Summary of mutations and clinical features of 15 previously reported patients with GPS, 42 previously reported patients with SBBYSS and 4 previously reported patients plus the current patient with GPS/SBBYSS.

Patient identification in our paper	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Author	Smith et al. 2011	Gannon et al. 2015	Smith et al. 2011	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015
Patient identification in the original report	11	26	3	22	20	12	13	14	17	1	48
Mutation	c.527dup	c.2959T>C	c.3018del	c.3040C>T	c.3064G>T	c.3147G>A	c.3147G>A	c.3147G>A	c.3366_3369del	c.3401_3402del	c.3598_3622dup25
Mutation	p.(Tyr176*)	p.(Trp987Arg)	p.(Glu1007Argfs*5)	p.(Gln1014*)	p.(Glu1022*)	p.(Pro1049Pro)	p.(Pro1049Pro)	p.(Pro1049Pro)	p.(Lys1124Glyfs*5)	p.(Gly1134Ala fs*2)	p.(Gly1208Glufs*19)
Exon	3	15	15	16	16	16	16	16	16	17	17
Diagnosed as	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	GPS
Mutation/patient group (this report)	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**2**
Gender	**M**	**F**	**F**	**F**	**F**	**M**	**M**	**F**	**M**	**F**	**F**
developmental delay/intellectual disability	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**
congenital heart defect	**n**	**n**	**n**	**n**	**y**	**n**	**n**	**y**	**n**	**n**	**y**
feeding difficulties	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**NK**	**y**	**y**
genital anomalies *	**y**	**n**	**n**	**n**	**n**	**y**	**y**	**y**	**y**	**n**	**n**
corpus callosum agenesis	**NK**	**n**	**NK**	**NK**	**n**	**NK**	**n**	**n**	**NK**	**n**	**y**
microcephaly	**n**	**y**	**n**	**n**	**NK**	**NK**	**n**	**y**	**n**	**n**	**y**
micrognathia	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
absent or hypoplastic patellae	**n**	**n**	**n**	**n**	**n**	**n**	**n**	**n**	**y**	**n**	**y**
flexion contractures	**NK**	**NK**	**NK**	**n**	**y**	**n**	**n**	**y**	**y**	**y**	**y**
club foot	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
skeletal anomalies (axial skeleton) **	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
renal anomalies	**n**	**n**	**NK**	**y**	**n**	**NK**	**n**	**n**	**y**	**n**	**y**
laryngo/tracheomalacia	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
small bowel malrotation	**NK**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**y**
anteriorly positioned anus	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
anal atresia or stenosis	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
hypotonia	**y**	**y**	**y**	**y**	**n**	**y**	**y**	**y**	**y**	**y**	**y**
blepharophimosis	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
ptosis	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
expressionless facies	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
typical SBBYSS face	**y**	**y**	**NK**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**n**
long thumbs	**n**	**n**	**n**	**y**	**n**	**y**	**y**	**y**	**n**	**n**	**y**
long toes	**n**	**y**	**n**	**y**	**n**	**y**	**y**	**n**	**n**	**y**	**n**
lacrimal ducts stenosis	**n**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
dental anomalies	**n**	**n**	**y**	**n**	**n**	**y**	**n**	**n**	**y**	**n**	**n**
cleft palate	**n**	**n**	**n**	**n**	**n**	**y**	**n**	**n**	**n**	**n**	**y**
hearing impairment	**n**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
tyhroid anomalies	**n**	**y**	**y**	**n**	**NK**	**n**	**y**	**n**	**n**	**n**	**y**
Common signs (4)	**3(3)**	**2(2)**	**2(2)**	**2(2)**	**3(3)**	**3(3)**	**3(3)**	**4(4)**	**2(3)**	**2(2)**	**3(3)**
GPS signs (12)	**0(9)**	**1(8)**	**0(10)**	**1(9)**	**2(9)**	**0(10)**	**0(7)**	**2(9)**	**3(11)**	**1(8)**	**6(12)**
SBBYSS sings (12)	**2(5)**	**4(9)**	**5(9)**	**4(9)**	**1(7)**	**6(11)**	**5(10)**	**3(8)**	**3(8)**	**3(8)**	**4(9)**
Total (28)	**5(17)**	**7(19)**	**7(21)**	**7(20)**	**6(19)**	**9(24)**	**8(20)**	**9(21)**	**8(22)**	**6(18)**	**13(24)**

y, yes (feature present); n, no (feature absent); NK, not known / not reported

* Some genital anomalies (hypoplastic labia minora, clitoral hypertrophy and scrotal hypoplasia) are typical for GPS and some (hypospadias) are typical for SBBYSS or are common to both syndromes (cryptorchidism). However, in the majority of previously published cases the type of genital anomaly was not specified, and therefore we have decided to include "genital anomalies" among the common symptoms.

** Skeletal anomalies include pelvic and costovertebral anomalies and thoracolumbal kyphosis/scoliosis.

12	13	14	15	16	17	18	19	20	21	22	23
Simpson et al. [2012]	Simpson et al. [2012]	Simpson et al. [2012]	Campeau et al. [2012]	Campeau et al. [2012]	Gannon et al. 2014	Simpson et al. [2012]	Campeau et al. [2012]	Campeau et al. [2012]	Simpson et al. [2012]	Campeau et al. [2012]	Gannon et al. 2015
GPS-01[a]	GPS-02	GPS-03	4	6	50	GPS-05	5	3	GPS-04	1	49
c.3680_3695del	c.3768_3771del	c.3768_3771del	c.3769_3772delTCTA	c.3769_3772delTCTA	c.3769_3772del	c.3773dup	c.3788_3789delAA	c.3802G>T	c.3877A>T	c.3892G>T	c.3962_3963del
p.(Asp1227Glufs*11)	p.(Lys1258Glyfs*13)	p.(Lys1258Glyfs*13)	p.(Lys1258Glyfs*13)	p.(Lys1258Glyfs*13)	p.(Lys1258Glyfs*13)	p.(Trp1259Valfs*12)	p.(Lys1263Argfs*7)	p.(Gly1268*)	p.(Lys1293*)	p.(Gly1298*)	p.(Gln1321Argfs*20)
18	18	18	18	18	18	18	18	18	18	18	18
GPS	GPS	GPS	GPS	GPS	GPS	GPS	GPS	GPS	GPS	GPS	GPS
2	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
F	**F**	**M**	**M**	**M**	**M**	**F**	**F**	**F**	**M**	**F**	**M**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**
y	**n**	**y**	**y**	**y**	**y**	**n**	**n**	**y**	**y**	**y**	**y**
NK	**NK**	**NK**	**y**	**n**	**y**	**NK**	**y**	**y**	**NK**	**n**	**y**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**NK**
y	**y**	**y**	**y**	**y**	**NK**	**y**	**y**	**y**	**n**	**y**	**n**
y	**y**	**y**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**n**	**NK**	**NK**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**
y	**y**	**y**	**y**	**y**	**NK**	**n**	**y**	**y**	**y**	**y**	**NK**
n	**y**	**NK**	**y**	**n**	**NK**	**y**	**y**	**y**	**y**	**y**	**NK**
y	**y**	**y**	**y**	**y**	**y**	**n**	**y**	**y**	**y**	**y**	**NK**
NK	**NK**	**NK**	**n**	**n**	**NK**	**NK**	**n**	**y**	**NK**	**n**	**NK**
NK	**NK**	**NK**	**n**	**n**	**NK**	**NK**	**n**	**y**	**NK**	**n**	**NK**
NK	**NK**	**NK**	**n**	**n**	**NK**	**NK**	**n**	**y**	**NK**	**y**	**NK**
NK	**NK**	**NK**	**y**	**y**	**NK**	**NK**	**n**	**n**	**NK**	**n**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**y**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**n**
NK	**NK**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**y**
NK	**NK**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**n**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**y**
NK	**NK**	**NK**	**NK**	**NK**	**n**	**NK**	**NK**	**NK**	**NK**	**NK**	**n**
NK	**NK**	**NK**	**n**	**n**	**NK**	**NK**	**n**	**n**	**NK**	**y**	**NK**
NK	**NK**	**NK**	**n**	**n**	**y**	**NK**	**y**	**n**	**NK**	**n**	**y**
3(4)	**2(3)**	**3(4)**	**4(4)**	**3(3)**	**4(4)**	**2(3)**	**3(3)**	**4(4)**	**3(4)**	**3(3)**	**4(4)**
7(11)	**8(12)**	**7(12)**	**8(9)**	**7(8)**	**4(12)**	**6(10)**	**7(8)**	**10(11)**	**6(10)**	**8(9)**	**2(11)**
0(12)	**0(12)**	**0(12)**	**0(10)**	**0(10)**	**4(11)**	**0(12)**	**1(11)**	**0(10)**	**0(12)**	**1(11)**	**4(9)**
10(27)	**10(27)**	**10(28)**	**12(23)**	**10(21)**	**12(27)**	**8(25)**	**11(22)**	**14(25)**	**9(26)**	**12(23)**	**10(24)**

24	25	26	27	28	29	30	31	32	33	34	35	36
Smith et al. 2011	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015	**Present case**	Gannon et al. 2015	Smith et al. 2011	Smith et al. 2011	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015	Campeau et al. [2012
4	55	51	23		54	5	10	7	21	16	53	2
c.4069G>T	c.4074_4079del	c.4110dupA	c.4114G>T	c.3295G>T	c.4205_4206del	c.4205_4206del	c.4205_4206del	c.4205_4206del	c.4205_4206del	c.4312del	c.4320_4321del	4368delinsAAAAAC
p.(Glu1357*)	p.(Glu1367_Glu1368)	p.(Glu1371Argfs*16)	p.(Glu1372*)	p.(Glu1099*)	p.(Ser1402Cysfs*5)	p.(Ser1402Cysfs*5	p.(Ser1402Cysfs*5)	p.(Ser1402Cysfs*5)	p.(Ser1402Cysfs*5)	p.(Glu1438Lysfs*23)	p.(Lys1441Glyfs*23)	p.(Glu1454Leyfs*8)
18	18	18	18	18	18	18	18	18	18	18	18	18
SBBYSS	GPS/SBBYSS	GPS	SBBYSS	GPS/SBBYSS	GPS/SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	GPS/SBBYSS	GPS
3	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**
M	**M**	**F**	**M**	**F**	**M**	**M**	**F**	**M**	**M**	**F**	**M**	**M**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**
y	**y**	**n**	**y**	**y**	**y**	**y**	**n**	**y**	**NK**	**y**	**y**	**y**
y	**n**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**NK**	**y**
y	**y**	**n**	**y**	**n**	**y**	**y**	**n**	**y**	**n**	**n**	**y**	**y**
NK	**NK**	**y**	**NK**	**y**	**y**	**NK**	**NK**	**n**	**n**	**n**	**NK**	**y**
y	**n**	**NK**	**y**	**n**	**y**	**y**	**n**	**n**	**n**	**y**	**y**	**y**
NK	**NK**	**NK**	**NK**	**n**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
y	**n**	**y**	**n**	**y**	**y**	**y**	**n**	**y**	**n**	**n**	**n**	**y**
NK	**y**	**y**	**y**	**y**	**y**	**NK**	**NK**	**y**	**y**	**y**	**n**	**y**
NK	**NK**	**NK**	**NK**	**n**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**y**
NK	**NK**	**NK**	**NK**	**n**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**n**
n	**NK**	**y**	**NK**	**y**	**NK**	**n**	**n**	**n**	**n**	**n**	**y**	**y**
NK	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**y**
NK	**NK**	**NK**	**NK**	**n**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**n**
NK	**NK**	**NK**	**NK**	**n**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**n**
NK	**NK**	**NK**	**NK**	**n**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**n**
y	**n**	**NK**	**y**	**y**	**y**	**y**	**y**	**y**	**n**	**y**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
y	**y**	**NK**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**NK**
y	**y**	**y**	**y**	**y**	**n**	**y**	**y**	**y**	**y**	**y**	**NK**	**NK**
y	**y**	**y**	**y**	**y**	**n**	**y**	**y**	**y**	**y**	**y**	**NK**	**NK**
y	**NK**	**NK**	**NK**	**y**	**NK**	**y**	**n**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**y**	**NK**	**NK**	**n**	**y**	**y**	**y**	**y**	**NK**	**y**	**NK**	**NK**
y	**n**	**n**	**y**	**n**	**n**	**n**	**n**	**n**	**n**	**n**	**y**	**NK**
y	**NK**	**NK**	**NK**	**n**	**NK**	**y**	**n**	**NK**	**NK**	**NK**	**NK**	**y**
y	**n**	**NK**	**n**	**n**	**y**	**n**	**n**	**y**	**y**	**y**	**y**	**n**
4(4)	**3(3)**	**2(2)**	**4(4)**	**3(3)**	**4(4)**	**4(4)**	**2(2)**	**4(4)**	**2(3)**	**3(3)**	**3(4)**	**4(4)**
2(11)	**1**	**4(12)**	**2(11)**	**5(5)**	**4(12)**	**2(11)**	**0(9)**	**2(9)**	**1(8)**	**2(9)**	**2(10)**	**7(8)**
8(12)	**4**	**2(11)**	**5(11)**	**8(8)**	**4(9)**	**7(10)**	**5(8)**	**6(11)**	**4(10)**	**6(11)**	**3(12)**	**1(11)**
14(27)	**8**	**8(25)**	**11(26)**	**16(16)**	**12(25)**	**13(25)**	**7(19)**	**12(24)**	**7(21)**	**11(23)**	**8(26)**	**12(23)**

37	38	39	40	41	42	43	44	45	46	47	48	49
Smith et al. 2011	Gannon et al. 2015	Smith et al. 2011	Gannon et al. 2015	Smith et al. 2011	Szakszon et al. 2013	Gannon et al. 2015	Gannon et al. 2015	Smith et al. 2011	Smith et al. 2011	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015
1	15	12	2	7	1	8	24	8	9	6	56	10
c.4405dup	c.4728_4729del	c.4911_4921del	c.4911_4921del	c.5030del	5071delTACTATGGin	c.5115_5116del	c.5131del	c.5201_5210dup	c.5201_5210dup	c.5209 C>T	c.5302 C>T	c.5302 C>T
p.(Ser1469Phefs*18)	p.(Arg1577Cysfs*21)	p.(Val1638Alafs*27	p.(Val1638Alafs*27	p.Thr1677Metfs*38	p.(Met1690Glufs*24)	p.(Tyr170611efs*9)	p.(Ser1711Profs*4)	p.(Gln1737Hisfs*41)	p.(Gln1737Hisfs*41)	p.(Gln1737*)	p.(Gln1768*)	p.(Gln1768*)
18	18	18	18	18	18	18	18	18	18	18	18	18
SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	GPS/SBBYSS	SBBYSS
3	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**
M	**F**	**F**	**F**	**M**	**M**	**F**	**M**	**F**	**F**	**M**	**M**	**F**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**
n	**n**	**n**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**n**	**y**	**n**
y	**n**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**NK**	**y**	**y**
y	**n**	**n**	**n**	**y**	**y**	**n**	**y**	**n**	**n**	**y**	**y**	**n**
NK	**NK**	**NK**	**n**	**NK**	**NK**	**NK**	**n**	**NK**	**NK**	**y**	**y**	**y**
y	**NK**	**n**	**y**	**NK**	**n**	**NK**	**n**	**NK**	**NK**	**NK**	**n**	**n**
NK	**NK**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
n	**n**	**n**	**n**	**n**	**NK**	**y**	**NK**	**y**	**n**	**y**	**n**	**n**
NK	**y**	**NK**	**y**	**NK**	**NK**	**y**	**y**	**NK**	**NK**	**y**	**n**	**n**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**n**	**NK**	**NK**	**NK**	**NK**	**n**	**y**	**n**	**NK**	**n**	**n**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**
NK	**NK**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
y	**y**	**y**	**y**	**y**	**NK**	**y**	**y**	**y**	**y**	**y**	**y**	**y**
y	**y**	**y**	**y**	**n**	**y**	**y**	**y**	**y**	**y**	**y**	**n**	**y**
y	**y**	**y**	**y**	**n**	**NK**	**y**	**n**	**y**	**y**	**y**	**y**	**y**
n	**NK**	**n**	**NK**	**y**	**NK**	**NK**	**NK**	**y**	**y**	**NK**	**NK**	**NK**
y	**n**	**y**	**y**	**y**	**NK**	**n**	**y**	**y**	**y**	**y**	**y**	**y**
n	**n**	**y**	**n**	**y**	**NK**	**n**	**y**	**y**	**y**	**n**	**n**	**y**
n	**NK**	**n**	**NK**	**y**	**NK**	**NK**	**y**	**y**	**y**	**NK**	**NK**	**NK**
n	**n**	**n**	**n**	**y**	**y**	**y**	**y**	**y**	**n**	**y**	**y**	**n**
3(3)	**1(1)**	**2(2)**	**3(3)**	**4(4)**	**4(4)**	**3(3)**	**4(4)**	**3(3)**	**3(3)**	**2(3)**	**4(4)**	**2(2)**
1(11)	**1(11)**	**0(9)**	**2(10)**	**0(11)**	**1(11)**	**2(12)**	**1(9)**	**2(12)**	**0(10)**	**3(12)**	**1(8)**	**1(8)**
5(8)	**4(9)**	**6(9)**	**5(10)**	**7(10)**	**5(12)**	**5(10)**	**7(11)**	**9(12)**	**8(11)**	**6(11)**	**5(10)**	**6(11)**
9(22)	**6(21)**	**8(20)**	**10(23)**	**11(25)**	**10(27)**	**10(25)**	**12(24)**	**14(27)**	**11(24)**	**11(26)**	**10(22)**	**9(21)**

50	51	52	53	54	55	56	57	58	59	60	61	62
Gannon et al. 2015	Smith et al. 2011	Gannon et al. 2015	Gannon et al. 2015	Gannon et al. 2015	Smith et al. 2011	Szakszon et al. 2013	Gannon et al. 2015	Gannon et al. 2015	Hung-Chun 2014	Gannon et al. 2015	Gannon et al. 2015	Smith et al. 2011
25	2	11	18	19	13	2	3	5	1	4	9	6
c.5302 C>T	c.5370_5373dup	c.5389 C>T	c.5389 C>T	c.5389 C>T	c.5389C>T	c.5389C>T	c.5424C>A	c.5442_5445del	c.5623_5624dupCA	c.5624_5625del	c.5731C>T	c.5734A>T
p.(Gln1768*)	p.(Ile1792Glnfs*12)	p.(Arg1797*)	p.(Arg1797*)	p.(Arg1797*)	p.(Arg1797*)	p.(Arg1797*)	p. (Tyr1808*)	p.(Phe1815Alafs*8)	p.(Gln1875Hisfs*5)	p.(Ala1876Leufs*3)	p.(Gln1911*)	p.(Arg1912*)
18	18	18	18	18	18	18	18	18	18	18	18	18
SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS	SBBYSS
4	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**
M	**F**	**F**	**M**	**M**	**F**	**F**	**F**	**M**	**M**	**F**	**M**	**F**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**NK**	**y**	**y**	**y**
y	**n**	**y**	**y**	**y**	**n**	**n**	**y**	**n**	**NK**	**y**	**n**	**y**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**NK**	**y**	**y**	**y**
y	**n**	**n**	**y**	**y**	**n**	**n**	**n**	**y**	**y**	**n**	**y**	**n**
n	**NK**	**n**	**n**	**NK**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**n**	**NK**
n	**NK**	**n**	**NK**	**n**	**n**	**NK**	**NK**	**NK**	**NK**	**NK**	**n**	**n**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
n	**y**	**y**	**n**	**y**	**n**	**NK**	**n**	**n**	**NK**	**n**	**n**	**n**
y	**NK**	**n**	**n**	**y**	**NK**	**NK**	**y**	**y**	**NK**	**y**	**y**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
n	**NK**	**n**	**y**	**NK**	**n**	**NK**	**NK**	**n**	**NK**	**NK**	**NK**	**n**
NK	**NK**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
y	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**y**	**NK**	**y**	**y**	**y**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**y**	**NK**	**NK**	**y**	**NK**	**NK**	**NK**
NK	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**
y	**y**	**y**	**y**	**y**	**y**	**NK**	**y**	**y**	**NK**	**y**	**y**	**y**
y	**y**	**y**	**y**	**y**	**n**	**y**	**n**	**y**	**NK**	**n**	**y**	**y**
y	**y**	**n**	**y**	**y**	**n**	**NK**	**n**	**y**	**NK**	**n**	**NK**	**y**
NK	**y**	**NK**	**NK**	**NK**	**n**	**NK**	**NK**	**NK**	**NK**	**NK**	**NK**	**y**
NK	**y**	**y**	**NK**	**y**	**y**	**y**	**y**	**n**	**NK**	**n**	**n**	**y**
y	**n**	**y**	**n**	**n**	**n**	**y**	**n**	**NK**	**NK**	**n**	**n**	**n**
NK	**y**	**y**	**NK**	**NK**	**n**	**y**	**NK**	**NK**	**NK**	**NK**	**NK**	**n**
y	**n**	**y**	**y**	**n**	**y**	**y**	**n**	**n**	**n**	**n**	**n**	**y**
4(4)	**2(2)**	**3(3)**	**4(4)**	**4(4)**	**2(2)**	**2(2)**	**3(3)**	**3(3)**	**1(4)**	**3(3)**	**3(3)**	**3(3)**
1(8)	**1(12)**	**2(8)**	**1(9)**	**2(11)**	**0(9)**	**0(12)**	**1(11)**	**2(10)**	**0(12)**	**1(11)**	**1(9)**	**0(9)**
6(12)	**7(10)**	**7(11)**	**5(11)**	**5(10)**	**4(7)**	**8(12)**	**3(8)**	**4(10)**	**2(11)**	**2(7)**	**3(9)**	**7(10)**
11(24)	**10(24)**	**12(22)**	**10(24)**	**11(25)**	**6(18)**	**10(26)**	**7(22)**	**9(23)**	**3(27)**	**6(21)**	**7(21)**	**10(22)**

Publikace 11

Activating mutations affecting the Dbl homology domain of *SOS2* cause Noonan syndrome.

Cordeddu V, Yin JC, Gunnarsson C, Virtanen C, Drunat S, Lepri F, De Luca A, Rossi C, Ciolfi A, Pugh TJ, Bruselles A, Priest JR, Pennacchio LA, Lu Z, Danesh A, Quevedo R, Hamid A, Martinelli S, Pantaleoni F, Gnazzo M, Daniele P, Lissewski C, Bocchinfuso G, Stella L, Odent S, Philip N, Faivre L, Vlckova M, Seemanova E, Digilio C, Zenker M, Zampino G, Verloes A, Dallapiccola B, Roberts AE, Cavé H, Gelb BD, Neel BG, Tartaglia M.

Hum Mutat. 2015 Nov;36(11):1080-7. IF: 5,144

Tato multicentrická studie je výsledkem spolupráce na poli genetických příčin RASopatií (Rauen, 2013), skupiny AD onemocnění s typickou faciální dysmorfií, VCC, poruchou růstu, anomáliemi skeletu, zvýšeným rizikem nádorů a MR. Syndrom Noonanové (NS), nejčastější RASopatie, je geneticky heterogenní. Je způsoben dysregulací proteinů dráhy RAS/ERK (extracellular signal-regulated kinase). Mutace v dosud známých genech jsou nalézány asi u 80% pacientů s odpovídajícím fenotypem.

Tato studie probíhala 10 let a měla za cíl nalézt asociaci mezi RASopathiemi a dosud nepopsanými geny hrajícími roli v RAS signální dráze u dosud neobjasněných případů NS. U všech pacientů byl podrobně popsán fenotyp pomocí standardizovaného dotazníku. Klinicky velmi suspektní pacienti byli dále vyšetřováni NGS a Sangerovým sekvenováním. Po objevení genu *SOS1* v roce 2007, jehož mutace vysvětlí asi 10 % případů NS, byl jako další vhodný kandidát vytipován gen *SOS2*, který vykazuje 70% homologii se *SOS1*.

Publikace popisuje několik pacientů s typickým NS a s missense variantami *SOS2*. Práce podporuje domněnku, že varianty v *SOS2* způsobují NS u malé části pacientů. Celkem byly identifikovány 4 missense mutace v 5 rodinách se sporadickým i familiálním výskytem NS. Podobně jako u *SOS1* vykazují pacienti s mutacemi *SOS2* normální růst a vývoj a výrazné postižení ektodermu. Na rozdíl od *SOS1*, který má více hot spotů, jsou mutace v *SOS2* lokalizovány výhradně v Db1 doméně. Práce rozšiřuje poznatky o genetické podstatě jednoho z nejčastějších klinicky dobře definovaných geneticky heterogenních syndromů, a přispívá k pochopení rozdílů mezi pacienty s NS. Přesnější korelace genotyp-fenotyp může vést k lepší individuální dispenzární péči o pacienty s NS a k optimalizaci budoucích obecných doporučení.

RESEARCH ARTICLE

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

Activating Mutations Affecting the Dbl Homology Domain of SOS2 Cause Noonan Syndrome

Viviana Cordeddu,[1,2†] Jiani C. Yin,[3†] Cecilia Gunnarsson,[4†] Carl Virtanen,[3†] Séverine Drunat,[5] Francesca Lepri,[6] Alessandro De Luca,[7] Cesare Rossi,[8] Andrea Ciolfi,[1] Trevor J. Pugh,[3] Alessandro Bruselles,[1] James R. Priest,[9,10] Len A. Pennacchio,[11,12] Zhibin Lu,[3] Arnavaz Danesh,[3] Rene Quevedo,[3] Alaa Hamid,[3] Simone Martinelli,[1] Francesca Pantaleoni,[1] Maria Gnazzo,[6] Paola Daniele,[7] Christina Lissewski,[13] Gianfranco Bocchinfuso,[14] Lorenzo Stella,[14] Sylvie Odent,[15] Nicole Philip,[16] Laurence Faivre,[17] Marketa Vlckova,[18] Eva Seemanova,[18] Cristina Digilio,[6] Martin Zenker,[13] Giuseppe Zampino,[19] Alain Verloes,[5] Bruno Dallapiccola,[6] Amy E. Roberts,[20‡] Hélène Cavé,[5,21‡] Bruce D. Gelb,[22*‡] Benjamin G. Neel,[3,23‡§] and Marco Tartaglia[1,6‡‖]

[1]Dipartimento di Ematologia, Oncologia e Medicina Molecolare, Istituto Superiore di Sanità, Rome 00161, Italy; [2]Dipartimento di Scienze Psicologiche, della Salute e del Territorio, Università degli Studi "G. d'Annunzio", Chieti-Pescara 66100, Italy; [3]Princess Margaret Cancer Centre, University Health Network and Department of Medical Biophysics, University of Toronto, Toronto, Ontario, ON M5S, Canada; [4]Department of Clinical and Experimental Medicine, Division of Clinical Genetics, Faculty of Health Sciences, Linköping University, Linköping 581 83, Sweden; [5]Département de Génétique, Hôpital Robert Debré, Paris 75019, France; [6]Bambino Gesù Children's Hospital, Istituto di Ricovero e Cura a Carattere Scientifico, Rome 00165, Italy; [7]IRCCS-Casa Sollievo della Sofferenza Hospital, Mendel Institute, Rome 00161, Italy; [8]UO Genetica Medica, Policlinico S.Orsola-Malpighi, Bologna 40138, Italy; [9]Division of Pediatric Cardiology, Stanford University School of Medicine, Stanford University, Stanford, California 94305; [10]Child Health Research Institute, Stanford Cardiovascular Institute, Stanford University School of Medicine, Stanford, California 94305; [11]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720; [12]US Department of Energy Joint Genome Institute, Walnut Creek, California 94598; [13]Institute of Human Genetics, University Hospital of Magdeburg, Otto-von-Guericke-University, Magdeburg 39106, Germany; [14]Dipartimento di Scienze e Tecnologie Chimiche, Università di Roma 'Tor Vergata', Rome 00133, Italy; [15]Service de Génétique Clinique, Hôpital SUD, Rennes 35200, France; [16]Département de Génétique Médicale, Hôpital d'Enfants de la Timone, Marseille 13385, France; [17]Centre de Génétique, Hôpital d'Enfants, Dijon 21000, France; [18]Department of Biology and Medical Genetics, Charles University 2nd Faculty of Medicine and University Hospital Motol, Prague 150 06, Czech Republic; [19]Istituto di Pediatria, Università Cattolica del Sacro Cuore, Rome 00168, Italy; [20]Department of Cardiology and Division of Genetics, Boston Children's Hospital, Boston, Massachusetts 02115; [21]INSERM UMR_S1131, Institut Universitaire d'Hématologie, Université Paris Diderot, Paris-Sorbonne-Cité, Paris 75205, France; [22]The Mindich Child Health and Development Institute, and the Departments of Pediatrics and Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029; [23]The Laura and Isaac Perlmutter Cancer Center, New York University School of Medicine, New York, New York 10016

Communicated by Nancy Spinner

Received 11 May 2015; accepted revised manuscript 30 June 2015.
Published online 14 July 2015 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22834

ABSTRACT: The RASopathies constitute a family of autosomal-dominant disorders whose major features include facial dysmorphism, cardiac defects, reduced postnatal growth, variable cognitive deficits, ectodermal and skeletal anomalies, and susceptibility to certain malignancies. Noonan syndrome (NS), the commonest RASopathy, is genetically heterogeneous and caused by functional dysregulation of signal transducers and regulatory proteins with roles in the RAS/extracellular signal-regulated kinase (ERK) signal transduction pathway. Mutations in known disease genes account for approximately 80% of affected individuals. Here, we report that missense mutations altering Son of Sevenless, Drosophila, homolog 2 (*SOS2*), which encodes a RAS guanine nucleotide exchange factor, occur in a small percentage of subjects with NS. Four missense mutations were identified in five unrelated sporadic cases and families transmitting NS. Disease-causing mutations affected three conserved residues located in the Dbl homology (DH) domain, of which two are directly involved in the intramolecular binding network maintaining SOS2 in its autoinhibited conformation. All mutations were found to promote enhanced signaling from RAS to ERK. Similar to NS-causing *SOS1* mutations, the phenotype associated with *SOS2* defects is characterized by normal development and growth, as well as marked ectodermal involvement. Unlike *SOS1* mutations, however, those in *SOS2* are restricted to the DH domain.

Additional Supporting Information may be found in the online version of this article.

†These authors contributed equally to this study.

‡These authors contributed equally as senior authors to this study.

Additional Supporting Information may be found in the online version of this article.

§Correspondence to: Benjamin G. Neel, The Laura and Isaac Perlmutter Cancer Center, New York University School of Medicine, New York, NY 10016. E-mail: Benjamin.Neel@nyumc.org

‖Correspondence to: Marco Tartaglia, Department of Hematology, Oncology and Molecular Medicine, Istituto Superiore di Sanità, Viale Regina Elena, 299, Rome 00161, Italy. E-mail: marco.tartaglia@iss.it

*Correspondence to: Bruce D. Gelb, The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, One Gustave Levy Place, Box 1040, New York, NY 10029. E-mail: bruce.gelb@mssm.edu

Contract grant sponsors: National Institutes of Health (R01 HL071207; R01 HL0832732; U01 DE020060, R01 HG003988, and U54 HG006997; U54 HG006504); Telethon-Italy (GGP13107); AIRC (IG 13360); Ministry of Health (RF-2011-02349938); Princess Margaret Cancer Foundation and the Ontario Ministry of Health and Long Term Care; CIHR CGS-D.

© 2015 WILEY PERIODICALS, INC.

Hum Mutat 0:1–8, 2015. © 2015 Wiley Periodicals, Inc.

KEY WORDS: genotype–phenotype correlations; Noonan syndrome; RAS signaling; SOS2

Introduction

Aberrant signaling from RAS has been causally linked to a family of clinically related developmental disorders, collectively termed the RASopathies, which are characterized by facial dysmorphism, a wide spectrum of cardiac defects, reduced growth postnatally, variable cognitive deficits, ectodermal and musculoskeletal anomalies, and increased risk for certain malignancies. The shared mechanism of disease for this group of disorders is dysregulation of the RAS/extracellular signal-regulated kinase (ERK) and/or phosphoinositide 3-kinase/AKT signaling pathways [Tartaglia and Gelb, 2010]. Among the RASopathies, Noonan syndrome (NS; MIM #163950) is the most common and clinical variable [Roberts et al., 2013]. So far, missense germline mutations in genes encoding proteins of the RAS family of GTPases (*KRAS, NRAS, RIT1*, and *RRAS*), modulators of RAS function (*PTPN11, SOS1, CBL, RASA2,* and *SHOC2*), or downstream signal transducers (*RAF1, BRAF,* and *MEK1*), resulting in gain-of-function (GOF) effects on RAS/ERK signaling, have been reported to underlie approximately 80% of NS or related conditions [Tartaglia and Gelb, 2010; Aoki et al., 2013; Chen et al., 2014; Flex et al., 2014]. Genotype–phenotype associations have been established for several of these genes. For example, individuals with mutations in Son of Sevenless, Drosophila, homolog 1 (*SOS1*; MIM #182530), which encodes a member of the son-of-sevenless family of RAS guanine nucleotide exchange factors (GEFs), typically have near-normal stature and neurocognitive development but show striking ectodermal involvement [Pandit et al., 2007; Roberts et al., 2007], whereas certain mutations affecting *RAF1*, a serine/threonine kinase functioning as a RAS effector and initiator of the RAF/MAK/ERK cascade, and possibly those affecting *RIT1,* a RAS subfamily GTPase, are generally associated with early-onset hypertrophic cardiomyopathy [Pandit et al., 2007; Razzaque et al., 2007; Aoki et al., 2013].

During the last 10 years, we and others have been using a candidate gene approach, focused on genes encoding signal transducers and regulators with a role in the RAS signaling cascade to resolve the unexplained proportion of cases of NS and the other RASopathies. Since our discovery of *SOS1* as a NS disease gene [Roberts et al., 2007; Tartaglia et al., 2007], Son of Sevenless, Drosophila, homolog 2 (*SOS2*; MIM #601247), which encodes a protein with similar function in RAS signaling, was considered an excellent candidate. Our first screening efforts, however, did not identify putative disease-causing variants, suggesting that activating mutations in this gene do not cause RASopathies or that they account for a relatively small proportion of affected subjects. To address the latter hypothesis, we initiated a multicenter collaborative screening effort using larger numbers of mutation-negative RASopathy cohorts. While this work was in progress, one paper reported the identification of heterozygous *SOS2* variants in a cohort of 50 subjects with NS without an identifiable mutation in previously known disease genes [Yamamoto et al., 2015]. Among these, one was found to occur de novo in a sporadic case, and two cosegregated with the trait in two small-sized families. Here, we report that *SOS2* mutations underlie a small proportion of NS cases and are associated with a clinical phenotype overlapping that resulting from *SOS1* mutations. Furthermore, we demonstrate that *SOS2* mutations promote enhanced activation of RAS and ERK, similar to what was observed for NS-causing *SOS1* mutations [Roberts et al., 2007; Tartaglia et al., 2007]. Unlike NS-associated *SOS1* mutations, however, *SOS2* defects in NS are not found in multiple domains of the GEF, but specifically cluster within the Dbl homology (DH) domain.

Materials and Methods

Clinical Data and Biological Material Collection

Clinical data were obtained and biological materials were collected and stored in accordance with the ethical standards of the institutional review boards (Ospedale Pediatrico Bambino Gesù, Rome, Italy; Università Cattolica del Sacro Cuore, Rome, Italy; Istituto Mendel, Rome, Italy; Policlinico S. Orsola-Malpighi, Bologna, Italy; Hôpital Robert Debré, Paris, France; Hôpital SUD, Rennes, France; Hôpital d'Enfants, Dijon, France; Hôpital d'Enfants de la Timone, Marseille, France; University Hospital of Magdeburg, Magdeburg, Germany; Boston Children's Hospital, Boston, MA; Icahn School of Medicine at Mount Sinai, New York, NY) and after written informed consent. All subjects exhibited features fitting NS (cohort 3) or within the RASopathy phenotypic spectrum (cohorts 1 and 2), and had tested negative for mutations in previously identified RASopathy genes. The clinical diagnosis was made on the basis of standardized clinical criteria as assessed by experienced clinical geneticists. Genomic DNA was isolated from peripheral blood leukocytes and buccal mucosal epithelial cells using standard protocols. Permission was obtained to publish the photographs of subjects shown in Figure 1.

DNA Sequencing and Mutation Analysis

Mutation scanning of the entire *SOS2* coding sequence (NM_006939.2; NC_000014.9 [50117128..50231878, complement]) was performed by Sanger sequencing at the US Department of Energy's Joint Genome Institute (Walnut Creek, CA), Istituto Superiore di Sanità and Ospedale Pediatrico Bambino Gesù, as previously described [Tartaglia et al., 2007].

Whole-exome sequencing (WES) was performed on 54 probands as part of the NHGRI-sponsored Centers for Mendelian Genomics program at the Yale Center for Genome Analysis (New Haven, CT) [Bamshad et al., 2012]. The WES analysis pipeline was based on the 1000 Genomes Project data analysis data pipeline, was composed from the widely used open source software projects bwa 0.7.5a [Li and Durbin, 2009], Picard 1.96, GATK 2.7 [McKenna et al., 2010; DePristo et al., 2011], snpEff 3.0 [Cingolani et al., 2012], BEDTools 2.16.2 [Quinlan and Hall, 2010], and custom-developed software, and implemented the "GATK Best Practices," including indel realignment, deduplication, and base-quality score recalibration. Short reads were aligned to a gender- and pseudo-autosomal region-masked build of the hg19 human reference genome using bwa mem. The exome capture targets were expanded with 100-bp flanks for variant calling. Single-nucleotide variants (SNVs) and indels were called jointly with the GATK HaplotypeCaller. Variant quality score recalibration (VQSR) was used to estimate the probability that an SNV is a true variant instead of an artifact and to set the corresponding variant filter thresholds. The PASS threshold for VQSR was set to capture 99.5% of known true positives. We observed that this threshold offered a good compromise between precision and recall. Mean coverage and fractions of bases at different coverage levels were calculated with the unflanked intervals;

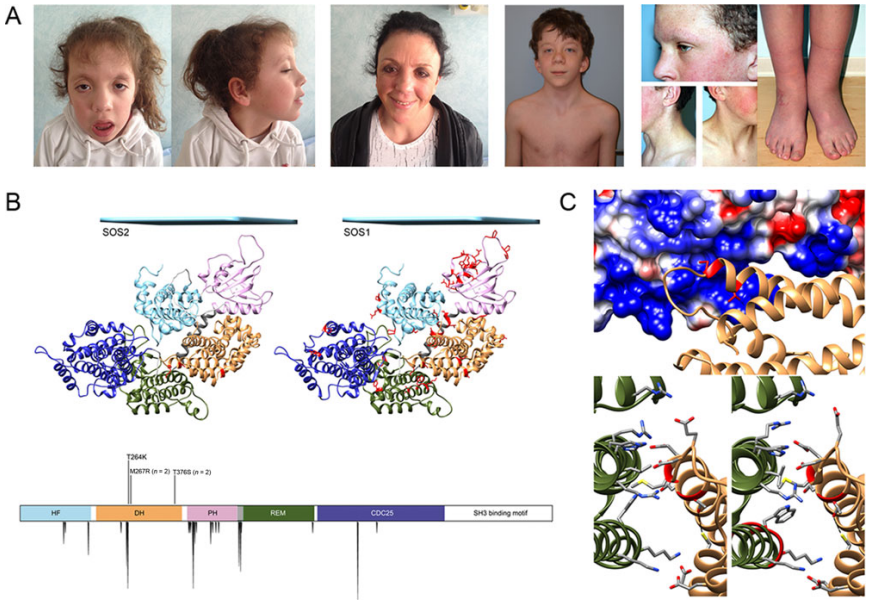


Figure 1. Mutations affecting residues located in the Dbl homology domain of SOS2 cause Noonan syndrome (NS). **A**: Clinical features of subjects carrying a mutated *SOS2* allele. **B**: Three-dimensional structure (above) and domain organization (below) of SOS proteins and positions of NS-causing mutations. Proteins are shown in the inactive conformation of SOS1 (residues 6–1,045) (PDB ID: 3KSY) [Gureasko et al., 2010] and the homology model derived for SOS2 (residues 6–1,043). Protein domains are color-coded: histone-like folds (cyan), Dbl homology (DH) (brown), pleckstrin homology (pink), helical linker (gray), RAS exchange motif (REM) (green), and CDC25 (blue). The cyan plane indicates the membrane position. The side chains of mutated residues (this study and [Lepri et al., 2011]) are shown in the SOS structures, whereas the NS-causing amino acid substitutions identified in SOS2 and the distribution of *SOS1* mutations reported in 83 NS subjects [Lepri et al., 2011] are shown above and below the scheme, respectively. **C**: The DH–REM interface is shown with the surface of the REM domain colored by its electrostatic potential and the Cα trace of the DH domain as a ribbon with the mutated residues colored in red (upper panel). Interactions at the DH/REM interface occurring in the SOS2 model (lower panel, left) and SOS1 X-ray structure (right). The residues participating in the interdomain interactions are shown as color-coded atoms (O: red; N: blue; C: gray; S: yellow). The ribbon is colored red at mutated residues. The interface is stabilized by several salt bridges, and is anionic at the DH domain (Asp263, Glu266, Asp269, Glu270, Asp283, Glu366), and cationic at the REM domain (Arg623, Arg686, Arg692, His693, His697, Lys726, Lys733). Met267 interacts hydrophobically with Leu685 and Trp727. Thr264 stabilizes the structure of the DH domain at the interface with REM by forming an H-bond with Cys280 (black dashed line).

the callable coverage of RefSeq coding exons was calculated with the flanked intervals.

The third cohort of subjects was screened by targeted resequencing. Primer pairs were designed with the IntegraGen (Evry, France) internal pipeline. Samples were amplified on an Access Array system (Fluidigm, San Francisco, CA), and equimolar pools of amplified products were sequenced on a MiSeq instrument (Illumina, San Diego, CA), using MiSeq Reagent Kit V2 cycles and paired end 2 × 150 bp. Image analysis and base calling were performed using the Real Time Analysis pipeline v. 1.14 (Illumina). Alignment of paired-end reads to the reference human genome and variant calling were carried out using the CASAVA v.1.8 pipeline (Illumina). Variant annotation was achieved using an in-house pipeline by IntegraGen. Sequencing at 25× depth covered at least 96.8% of *SOS2* in all patients.

All variants identified by WES and targeted sequencing were confirmed using Sanger sequencing. When parental genomic DNAs were available, their status vis-à-vis the relevant variant was similarly assessed. When possible, paternity was confirmed by simple tandem repeat (STR) genotyping, using the AmpF/STR Identifier PCR Amplification Kit (Applied Biosystems, Waltham, MA) or the PowerPlex 16 System (Promega, Fitchburg, WI).

The frequency of the identified *SOS2* variants in the general population was assessed using the Exome Aggregation Consortium's database (ExAC; http://exac.broadinstitute.org). The likelihood that variants were deleterious was evaluated using the combined annotation-dependent depletion method (CADD; http://cadd.gs.washington.edu/info) [Kircher et al., 2014], reported as scaled C-scores. A C-score threshold of 15.0 was used for declaring likely pathogenicity of missense variants. The *SOS1* and *SOS2* variants observed in human cancers are reported in the Catalogue of Somatic Mutations in Cancer (COSMIC; http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/).

Structural Modeling

A homology model of SOS2 (residues 6–1,043) was obtained based on the crystallographic structure of SOS1 (residues 6–1,045, pdb code 3KSY; http://www.rcsb.org/pdb/home/home.do) [Gureasko et al., 2010]. The sequence alignment was performed by using Clustal W [Larkin et al., 2007] (identity 78%) and checked manually. The structure model was built by using the Swiss Model Workspace [Arnold et al., 2006]. The software Chimera [Pettersen et al., 2004] was used for electrostatic calculations and for the structural representations reported in the figures.

Biochemical Studies

A *SOS2* cDNA (Transomics, Huntsville, AL) was amplified by PCR, with the addition of a FLAG tag at the N-terminus. The

amplified PCR product was subcloned into the expression vector pcDNA5/FRT/TO (Invitrogen, Carlsbad, CA). NS-associated mutations were introduced by site-directed mutagenesis.

HEK-293T cells (ATCC, Manassas, VA) were maintained in DMEM containing 10% (v/v) FBS and 100 units/ml penicillin-streptomycin (Invitrogen). Cells were seeded for 1 day before transfection with wild-type or mutant *SOS2* plasmids, along with a hemagglutinin (HA)-tagged ERK expression construct at a 4:1 ratio, using FuGENE HD transfection reagent (Promega), according to the manufacturer's protocol. This design eliminates transfection efficiency as a variable, by ensuring that all cells analyzed for HA-ERK phosphorylation also express the relevant *SOS2* allele. For assessing RAS activation, we generated stable transductants. Briefly, Flp-In T-REx 293 cells (Invitrogen) were cotransfected with the appropriate pcDNA5/FRT/TO expression plasmid, and the Flp recombinase-expressing plasmid pOG44 (Invitrogen) using FuGENE HD (Promega). Stable pools of transfectants were selected in hygromycin (250 μg/ml). Expression of exogenous *SOS2* was induced with 1 μg/ml tetracycline 24 hr before analysis. Cells were maintained in 10% serum or serum-starved overnight, before stimulation with EGF (20 ng/ml) for various times, as indicated.

Total cell extracts were prepared in radioimmunoprecipitation assay buffer (50 mM Tris–HCl [pH 7.5], 150 mM NaCl, 2 mM EDTA, 1% NP-40, 0.5% Na deoxycholate, 0.1% SDS), supplemented with a protease and phosphatase inhibitor cocktail (40 μg/ml phenylmethylsulfonyl fluoride, 20 mM NaF, 1 mM Na_3VO_4, 10 mM β-glycerophosphate, 10 mM sodium pyrophosphate, 2 μg/ml antipapain, 2 μg/ml pepstatin A, 20 μg/ml leupeptin, and 20 μg/ml aprotinin). Lysates (20–25 μg protein) were resolved by SDS-PAGE and analyzed by immunoblotting using the following antibodies: anti-FLAG, clone M2 (Sigma–Aldrich, St. Louis, MO); anti-phospho-p44/42 MAPK and anti-phospho-MEK1/2 (Cell Signaling Technology, Beverly, MA); anti-ERK2, clone D-2 (Santa Cruz, Biotechnology, Dallas, TX); anti-ERK1/2 (Millipore, Billerica, MA). Binding of primary antibodies was detected by using IRDye infrared secondary antibodies and the Odyssey Infrared imaging system (Li-Cor Biosciences, Lincoln, NE). Quantification was performed using Odyssey V3.0 software.

RAS loading was assessed using a RAS activation assay kit (Millipore), according to manufacturer's instructions. RAS-GTP was recovered from lysates using the GST-tagged RAF-Ras binding domain of RAF1 as bait in pull-down experiments followed by SDS-PAGE and immunoblotting for RAS.

Statistical Analysis

The biochemical data for the SOS2 mutants and wild-type controls were compared using an ANOVA or a Fisher exact test, as appropriate; the threshold for declaring significance was set at $P < 0.05$. A Bonferroni post-hoc correction was applied to all of the ANOVA *P* values.

Results

SOS2 Mutation Scanning in the RASopathies

We scanned *SOS2* for mutations in 150 individuals with clinically diagnosed RASopathy who had screened negative for previously discovered disease genes. We identified 19 individuals harboring nine missense variants, of which five were known polymorphisms and four were novel (Supp. Table S1). Among the latter, each observed once, three were deemed benign, as two were inherited from unaffected fathers and one was likely tolerated. The remaining variant, c.1127C>G, predicting the p.[Thr376Ser] amino acid change (Supp. Fig. S1A) had not been observed in greater than 120,000 alleles from the ExAC's database and was deemed likely deleterious using the CADD method [Kircher et al., 2014]. The variant was inherited from his mother, who exhibited features suggestive of NS. Genotyping of the mother's parents documented its de novo origin and confirmed paternity (Supp. Fig. S1A). Of note, mutation of the corresponding *SOS1* residue, Thr378, had been previously reported to cause NS [Denayer et al., 2010].

Next, we scanned WES data with excellent *SOS2* coverage (Supp. Fig. S1B) from another 54 mutation-negative RASopathy cases. We identified five *SOS2* missense variants, including four likely polymorphisms, two recurrent and two inherited from unaffected parents. The remaining individual had a different missense nucleotide substitution affecting codon 376 (c.1126A>T, p.[Thr376Ser]), inherited from her affected mother, in whom it arose de novo (Supp. Fig. S1A). Consistent with the causative role of the *SOS2* lesion in the family, no other variant affecting previously identified RASopathy genes was annotated in WES data from Subject 2.

Because the clinical features of the three affected individuals were consistent with NS (Fig. 1A; Table 1) (see below), we screened *SOS2* in a third cohort, which included 61 mutation-negative subjects with clinical diagnosis of NS, using targeted resequencing. Sequencing at 25× depth covered at least 96.8% of the *SOS2* coding sequence in all patients. We found a c.791C>A (p.[Thr264Lys]) and two independent c.800T>G (p.[Met267Arg]) variants (Supp. Fig. S1A) that were deemed likely pathologic due to high CADD scores, correspondence to an NS-related *SOS1* mutational hotspot (p.[Thr266Lys], p.[Met269Arg], and p.[Met269Thr]), and probable GOF effects on GEF activity, based on SOS1 structural studies and expression of p.[Met269Arg] [Pandit et al., 2007; Roberts et al., 2007]. The c.791C>A allele arose de novo in that sporadic case, for whom paternity was confirmed.

All *SOS2* variants described in this report have been submitted to the NSEuronet mutation database (https://nseuronet.com/php/index.php).

Molecular Modeling of NS-Associated *SOS2* Mutations

Similar to the structurally and functionally related SOS1, SOS2 stimulates the release of GDP from RAS, promoting the conversion of the GTPase from the inactive, GDP-bound to the active, GTP-bound form [Nimnual and Bar-Sagi, 2002]. SOS2 is a large multidomain protein characterized by an *N*-terminal regulatory portion including two tandemly arranged histone-like folds, which are followed by DH domain and a pleckstrin-homology (PH) domain, and a *C*-terminal catalytic region comprising the RAS exchanger motif (REM) and CDC25 domains, followed by a tail containing docking sites for adaptor proteins required for receptor anchoring (Fig. 1B). In SOS1, GEF activity is controlled principally by two binding sites for RAS: the catalytic site, which is located within the CDC25 domain, and a distal site involving two adjacent regions of the CDC25 and REM domains. The latter domain positively modulates GEF activity by promoting a conformational change at the active site that allows GDP-RAS to bind [Margarit et al., 2003]. The majority of *SOS1* mutations causing NS affect residues that are implicated in the maintenance of SOS1 in its autoinhibited conformation [Roberts et al., 2007; Tartaglia et al., 2007; Lepri et al., 2011]. Among these, a class of mutations involves residues that participate in the autoinhibitory interaction of the DH and REM domains that blocks RAS access to the allosteric site. These mutations directly

Table 1. List of Clinical Features of *SOS2* Mutation-Positive Subjects

Case	1	2 (mother of case 1)	3	4 (daughter of case 3)	5	6	7
Mutation (cDNA)[a]	1127C>G	1127C>G	1126A>T	1126A>T	800T>G	791C>A	800T>G
Mutation (protein)	p.[Thr376Ser]	p.[Thr376Ser]	p.[Thr376Ser]	p.[Thr376Ser]	p.[Met267Arg]	p.[Thr264Lys]	p.[Met267Arg]
Sex	M	F	F	F	F	M	M
Age (years)	13	34	30	6	6	24	42
Perinatal findings							
Polyhydramnios	+	na	na	+	–	na	na
Fetal macrosomia	–	–	+	+	+	na	na
Other prenatal	–	–	–	–	INT	–	–
Neonatal/infantile growth failure	+	na	–	–	–	–	–
Birth weight (kg)/birth weight SD	2.9/–1.13	na	na	na	4.29/+1.89	3.5/–0.06	na
Birth length (cm)/birth length SD	na	na	na	na	51/+0.61	51/+0.31	na
Birth HC (cm)/birth HC SD	na	na	na	na	35/+0.12	36/+0.12	na
Poor sucking	+	+	–	–	+	na	na
Height (cm)/height SD	114.6/–2.53	161.5/–0.50	158/–0.82	113/–0.99	112/–0.97	180/+0.44	165/–1.65
Weight (kg)/weight SD	na	n.a.	na	na	17.9/+1.48	na	53/–2.03
GH deficiency	+	n.a.	–	–	na	na	na
Craniofacial features							
Tall forehead	+	+	+	+	–	+	+
Sparse eyebrows	+	+	+	+	+	+	+
Downslanting palpebral fissures	+	–	+	+	+	–	–
Hypertelorism/telecanthus	+	+	+	+	–	–	–
Epicanthal folds	+	+	+	+	–	–	–
Palpebral ptosis	+	–	+	+	+	–	+
Flat nasal bridge	+	–	+	+	–	–	–
Prominent philtrum	–	–	+	+	+	–	–
Thick lips	–	+	+	+	–	–	+
Low-set/posteriorly rotated ears	+	+	+	+	+	–	+
Thickened helix	+	+	+	+	+	–	–
Large, thick ear lobe	–	+	+	+	–	–	–
Cardiac anomalies							
Pulmonary valve stenosis	–	–	–	–	–	–	+
Pulmonary valve dysplasia	–	–	–	–	–	–	–
Hypertrophic cardiomyopathy	–	–	–	–	ISH	–	+
Septal defects	VSD	–	ASD	–	ASD	–	–
Other defects	–	–	–	–	RM	–	–
Dermatologic findings							
Hyperpigmented skin	–	–	–	–	–	–	–
Keratosis pilaris/dry skin	–	–	+	+	+	+	–
Sparse/absent scalp hair	+	–	+	+	–	+	na
Sparse/absent eyebrows	+	–	+	+	na	+[b]	+
Ulerythema ophryogenes	–	–	+	+	+	+	na
Curly hair	–	+	+	+	–	+	+
Deep palmo/plantar creases	–	–	–	–	–	–	+
Other features	–	–	–	–	HS	Lentigines[c]	–
Musculoskeletal features							
Short webbed neck	+	+	+	+	+	na	–
Cubitus valgus	–	–	+	+	na	na	na
Hyperextensible joints	–	–	+	+	–	na	+
Pectus deformity	+	–	+	+	+[d]	+[d]	+[d]
Central nervous system features							
Intellectual disability	–[e]	–	–	–[f]	–	–	–
Brain MRI abnormalities	–	–	–	na	na	–	na
Seizures/EEG abnormalities	–	–	–	–	–	–	–
Ophthalmological anomalies	+[g]	–	–	–	+[h]	–	na
Gastrointestinal anomalies	–	+[i]	–	–	na	–	–
Cryptorchidism	+					+	+
Urogenital anomalies		–	–	–	na	+[j]	
Hematologic anomalies	–	–	–	–	na	+[k]	–
Lymphatic anomalies	–	+	–	–	+[l]	+[m]	–
Miscellaneous	–	–	–	–	+[n]	BIH	+[o]

[a]Position referred to the A of the ATG translation initiation codon in the reference cDNA sequence (NM_006939.2).
[b]Normalized by 24 years.
[c]Multiple lentigines spread on the face and neck.
[d]Excavatum.
[e]Mild learning difficulties.
[f]IQ 120.
[g]Hyperopia.
[h]Right: posterior embryontoxin; left: enlarged corneal nerves.
[i]Gastric malrotation.
[j]Bilateral uretero-pelvic junction stenosis.
[k]Easy bruising.
[l]Marked lymphedema of the right leg (around 6 years).
[m]Lymphedema of hands and feet (onset at 15 years).
[n]Superinfection of right buttock cyst.
[o]Charcot-Marie-Tooth disease type 2, onset in infancy (unidentified gene), hypereosinphilia.
BIH, bilateral inguinal hernia; HC, head circumference; HS, hyperelastic skin; INT, increased nuchal translucency; ISH, isolated septal hypertrophy; NA, data not available; RM, rhabdomyoma of the right ventricle (spontaneously resolved); SD, standard deviation (based on CDC growth charts 2000).

affect the stability of the inactive conformation of the GEF directly by disrupting the inhibitory interdomain bonding network at the distal site.

To analyze the functional impact of the three disease-associated missense variants, all altering conserved residues in the DH domain (Fig. 1B), a model of autoinhibited SOS2 (residues 6–1,043) was obtained using the crystallographic structure of SOS1 (residues 6–1,045, pdb code 3KSY) as the template [Gureasko et al., 2010]. Based on this model, the p.[Thr264Lys] and p.[Met267Arg] substitutions insert positively charged residues in the anionic DH region that interfaces with a cationic region of the REM domain, thereby reducing electrostatic attraction (Fig. 1C). Similar to what was observed for *SOS1* mutations affecting certain residues (p.[Thr266Lys], p.[Met269Arg/Thr], p.[Lys728Ile], p.[Trp729Leu], and p.[Ile733Phe]), these substitutions were predicted to activate SOS2 by destabilizing the DH–REM interaction, allowing RAS to access the allosteric site [Sondermann et al., 2004]. Thr376 is solvent-exposed, so the effect of the p.[Thr376Ser] substitution was unclear.

Functional Effects of NS-Associated *SOS2* Mutations

To test the GOF role of NS-associated *SOS2* mutations predicted by our molecular modeling analysis, the effects exerted by these mutations on RAS-ERK signaling were assessed by comparing the levels of ERK and MEK phosphorylation as well as that of GTP-bound RAS in transient transfection assays. Consistent with the predicted activating role of the three amino acid substitutions, expression of the $SOS2^{Thr264Lys}$, $SOS2^{Met267Arg}$, and $SOS2^{Thr376Ser}$ mutants promoted enhanced phosphorylation of both endogenous and exogenous ERK compared with that observed in cells expressing the wild-type protein (Fig. 2A). Similarly, enhanced activation of endogenous MEK was observed in cells expressing these SOS2 mutants. Of note, the SOS2 mutants engendered different degrees of hyperactivation. ERK and MEK were constitutively active in $SOS2^{Thr264Lys}$- and $SOS2^{Met267Arg}$-expressing cells, whether randomly growing or serum-starved. By contrast, cells expressing the $SOS2^{Thr376Ser}$ mutant exhibited enhanced and protracted activation of these kinases, but activation remained dependent on growth factor stimulation. The increased GEF activity of the NS-associated SOS2 mutants is consistent with the structural analyses. Moreover, our biochemical analyses indicate a differential impact of NS-associated mutations at the level of SOS2 activation.

We also assessed RAS activation by generating stable transductants in Flp-In T-Rex 293 cells, which allow tetracycline-inducible expression of SOS2 constructs at near endogenous levels. Compared with the effects of wild-type SOS2, expression of all three SOS2 mutants led to increased RAS activation, shown as higher levels of GTP-bound RAS. These findings are consistent with a direct GOF effect on the RAS-GEF activity of NS-associated *SOS2* mutations.

Overall, these data provide evidence that, similar to what has been reported previously for NS-causative *SOS1* mutations, NS-associated *SOS2* mutations promote enhanced activation of RAF-MEK-ERK signaling cascade, an effect that is directly mediated by their enhanced RAS-GEF activity.

Clinical Features Observed in NS due to *SOS2* Mutations

Detailed clinical data were collected for the six affected individuals with the four *SOS2* missense mutations to explore possible genotype–phenotype correlations. All subjects displayed typical NS facial features (Fig. 1A). Notably, height was normal in all but one individual, and all individuals exhibited normal or nearly normal neurocognitive status at 6–42 years (Table 1). By contrast, ectodermal involvement was striking and similar to that seen in *SOS1* mutation-associated NS patients, including facial keratosis pilaris, sparse scalp hair, and ulerythema ophryogenes (Fig. 1A; Table 1). Cardiac involvement was typical for NS. Overall, the genotype–phenotype associations observed with the *SOS2* mutations resembled those established for *SOS1* mutations.

Discussion

Here, we report that *SOS2* is mutated in NS, confirming the recent report of a Brazilian cohort from Yamamoto et al. (2015). We provide the first data indicating that disease-causing *SOS2* mutations promote enhanced GEF function in SOS2, resulting in enhanced signaling through RAS and the RAF-MEK-ERK cascade. Finally, the available clinical records support the idea that *SOS2* mutations are associated with a phenotype resembling that previously associated with *SOS1* mutations.

Mutations in *SOS1* constitute approximately 10% of NS cases. In contrast, *SOS2* mutations are estimated to account for a far smaller percentage of cases in the three cohorts we analyzed. Remarkably, all of the eight independent *SOS2* mutations underlying NS were found to alter only one of three amino acids (i.e., Thr264, Met267, and Thr376), all located in the DH domain. This is strikingly different from NS-associated *SOS1* mutations that arise at multiple hotspots (Fig. 1B) [Pandit et al., 2007; Roberts et al., 2007; Lepri et al., 2011]. The two SOS proteins are nearly 70% homologous, expressed ubiquitously, and activate RAS proteins. SOS2, however, has weaker biological effects. For example, *Sos1* deficiency in mice is embryonic lethal, whereas mice without *Sos2* are normal [Esteban et al., 2000; Qian et al., 2000]. SOS2 has been shown to be much less stable than SOS1 because of its accelerated degradation via a ubiquitin-dependent process [Nielsen et al., 1997]. Furthermore, expression of myristoylated SOS1 promotes cell transformation, whereas similarly tagged SOS2 cannot. Consistent with this idea, several germline NS-causing *SOS1* GOF mutations at three major hotspots within the DH and PH domains as well as the helical linker occur as sporadic mutations in cancers (as reported in COSMIC), whereas cancer-associated *SOS2* mutations only affect the DH domain. The driving force behind the *SOS2* DH domain "hotspot" shared with *SOS1* is likely biochemical, not genetic. Both mutants are constitutively active; neither mutation arose at a CpG dinucleotide and both are transversions. Although the precise mechanism is unclear, the same is likely true of the two nucleotide changes resulting in the p.[Thr376Ser] substitution, which also did not arise at CpG dinucleotides, are transversions and for which a comparable *SOS1* substitution has been observed once in cancer.

Our data, which represent the first biochemical characterization of NS-causing *SOS2* mutations, suggest that different mechanisms likely drive the functional dysregulation of SOS2 and, in turn, aberrant RAS activation. Mutations affecting Thr264 and Met267, which are located in the surface of the DH domain that mediates the autoinhibitory interaction with the REM domain, were shown to constitutively upregulate MEK and ERK activity. The corresponding residues in SOS1, Thr266 and Met269, are among the most common site of mutations in NS. In particular, Met269, which is mutated in approximately 10% of cases bearing *SOS1* mutations, interacts directly with residues of the REM domain implicated in RAS binding [Sondermann et al., 2004]. As previously documented for SOS1, these mutations are predicted to affect the stability of the inactive conformation of the protein directly by disrupting the inhibitory interdomain bonding network at the distal site [Lepri

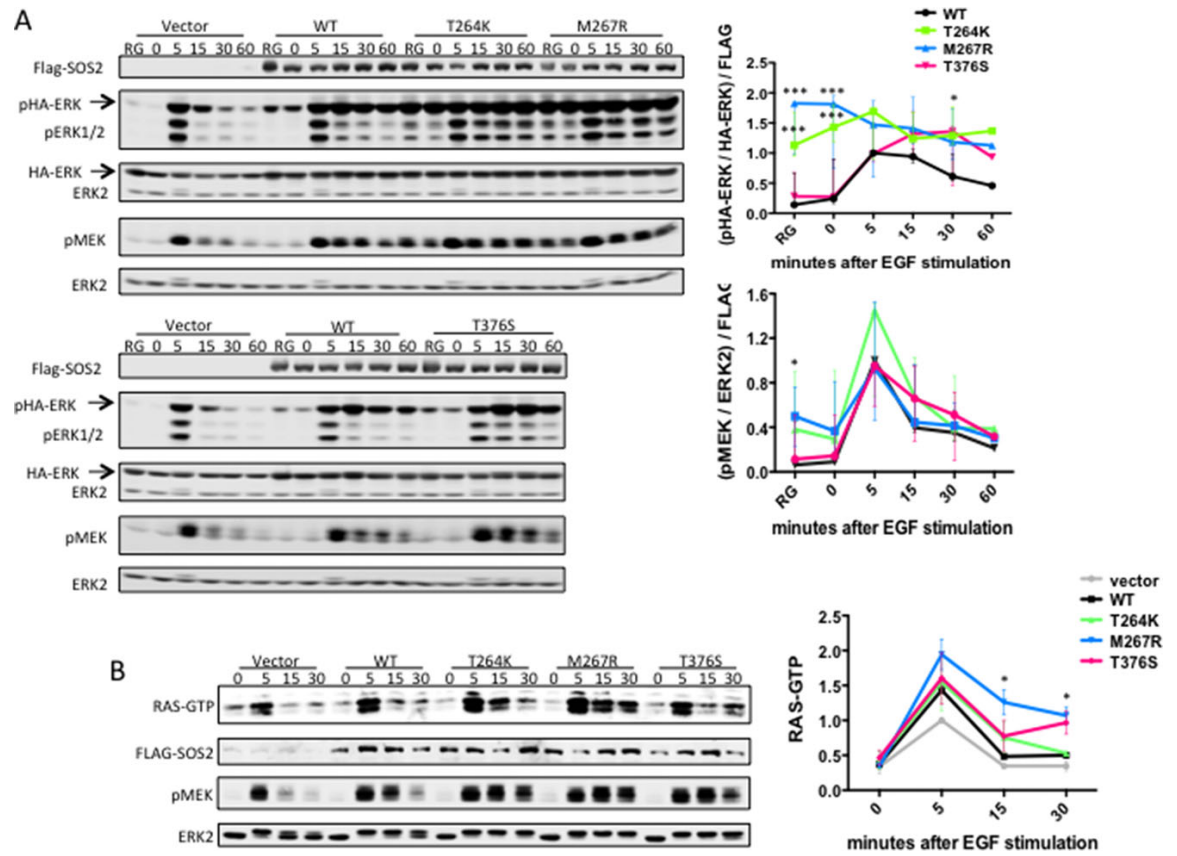


Figure 2. Noonan syndrome-associated SOS2 mutants cause enhanced MEK/ERK activation. **A**: SOS2 mutants enhance MEK/ERK activation. *FLAG-SOS2* and *HA-ERK1* expression constructs were cotransfected into 293T cells. Cells were randomly growing (RG), or starved and then stimulated with EGF (20 ng/ml), as indicated, prior to lysis. Left: a representative immunoblot. Right: quantification of MEK and ERK phosphorylation from three biological replicates (mean ± s.d; $^{**}P < 0.01$, $^{*}P < 0.05$, Bonferroni post-test when ANOVA was significant). **B**: SOS2 mutants increase RAS activation. Flp-In T-REx 293 cells stably expressing *FLAG-SOS2* constructs were starved, stimulated with 20 ng/ml EGF, and RAS loading was assessed. Left: a representative immunoblot. Right: quantification of RAS loading pooled from two biological replicates (mean ± s.d; $^{*}P < 0.05$, one-tailed Student's *t*-test).

et al., 2011]. On the other hand, a different pattern of signaling dysregulation was observed in cells overexpressing $SOS2^{Thr376Ser}$. In these cells, ERK activation was enhanced and protracted but retained dependence on EGF stimulation. While the solvent-exposed position of Thr376 suggests a possible role in the interaction of the GEF with signaling partners and/or modulators, the available structural data do not allow us to recognize any functional clue for this recurrent NS-causing SOS2 substitution.

Analysis of the clinical data documented that mutations in *SOS2* are associated with a consistent phenotype that unambiguously falls within the NS clinical spectrum, but is apparently characterized by a high prevalence of ectodermal features and low occurrence of short stature and cognitive impairment compared with what is observed in the NS general population. Overall, the *SOS2*-associated NS phenotype closely resembles that associated with *SOS1* mutations. Notably, both groups of patients enjoy excellent neurodevelopment, which is likely attributable to the developmental stage-dependent expression of RAS GEFs. While SOS proteins activate RAS downstream from *N*-methyl-D-aspartate glutamate receptors in cortical neurons of the developing and neonatal central nervous systems (CNS), there is a switch to RAS guanine nucleotide-releasing factors 1 and 2 later in life [Tian et al., 2004]. Consistent with evidence that NS-associated *PTPN11* mutations perturb CNS homeostasis, not brain development [Pagani et al., 2009; Lee et al., 2014], restoration of normal CNS RAS signaling in children with *SOS* mutations enables normal development. Hence, our findings provide further hope that therapies reducing RAS signaling in children with NS-causing mutations in non-*SOS* genes would improve their developmental trajectories.

Overall, the present work provides evidence that a narrow spectrum of activating missense mutations in *SOS2* account for a small proportion of NS, and that subjects heterozygous for a germline *SOS2* mutation exhibit a distinctive phenotype resembling that of *SOS1* mutations, providing new clinically valuable information for diagnosis and more effective patient management.

Acknowledgments

We thank the families who participated in the study, Serenella Venanzi (Istituto Superiore di Sanità, Rome, Italy), Natacha Fillot and Nathalie Pouvreau (Robert Debré Hospital, Paris, France) for technical assistance, and Guillaume Gauchotte (Nancy University Hospital, Nancy, France) for the tissue biopsy specimen (father of patient 6).

Several authors (B.D.G., B.G.N., A.E.R., and M.T.) are inventors for patents that have been licensed by their institutions for gene testing for NS and for which they receive royalties. By virtue of expanding the utility of testing, the work described in this manuscript could theoretically increase those royalties.

B.G.N. is a Canada Research Chair, Tier 1. L.A.P.'s research was conducted at the E.O. Lawrence Berkeley National Laboratory, performed under Department of Energy Contract DE-AC02-05CH11231.

References

Aoki Y, Niihori T, Banjo T, Okamoto N, Mizuno S, Kurosawa K, Ogata T, Takada F, Yano M, Ando T, Hoshika T, Barnett C, et al. 2013. Gain-of-function mutations in RIT1 cause Noonan syndrome, a RAS/MAPK pathway syndrome. Am J Hum Genet 93:173–180.

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201.

Bamshad MJ, Shendure JA, Valle D, Hamosh A, Lupski JR, Gibbs RA, Boerwinkle E, Lifton RP, Gerstein M, Gunel M, Mane S, Nickerson DA, et al. 2012. The Centers for Mendelian Genomics: a new large-scale initiative to identify the genes underlying rare Mendelian conditions. Am J Med Genet A 158A:1523–1525.

Chen PC, Yin J, Yu HW, Yuan T, Fernandez M, Yung CK, Trinh QM, Peltekova VD, Reid JG, Tworog-Dube E, Morgan MB, Muzny DM, et al. 2014. Next-generation sequencing identifies rare variants associated with Noonan syndrome. Proc Natl Acad Sci USA 111:11473–11478.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6:80–92.

Denayer E, Devriendt K, de Ravel T, Van Buggenhout G, Smeets E, Francois I, Sznajer Y, Craen M, Leventopoulos G, Mutesa L, Vandecasseye W, Massa G, et al. 2010. Tumor spectrum in children with Noonan syndrome and SOS1 or RAF1 mutations. Genes Chromosomes Cancer 49:242–252.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498.

Esteban LM, Fernandez-Medarde A, Lopez E, Yienger K, Guerrero C, Ward JM, Tessarollo L, Santos E. 2000. Ras-guanine nucleotide exchange factor sos2 is dispensable for mouse growth and development. Mol Cell Biol 20:6410–6413.

Flex E, Jaiswal M, Pantaleoni F, Martinelli S, Strullu M, Fansa EK, Caye A, De Luca A, Lepri F, Dvorsky R, Pannone L, Paolacci S, et al. 2014. Activating mutations in RRAS underlie a phenotype within the RASopathy spectrum and contribute to leukaemogenesis. Hum Mol Genet 23:4315–4327.

Gureasko J, Kuchment O, Makino DL, Sondermann H, Bar-Sagi D, Kuriyan J. 2010. Role of the histone domain in the autoinhibition and activation of the Ras activator Son of Sevenless. Proc Natl Acad Sci USA 107:3430–3435.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46:310–315.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

Lee YS, Ehninger D, Zhou M, Oh JY, Kang M, Kwak C, Ryu HH, Butz D, Araki T, Cai Y, Balaji J, Sano Y, et al. 2014. Mechanism and treatment for learning and memory deficits in mouse models of Noonan syndrome. Nat Neurosci 17:1736–1743.

Lepri F, De Luca A, Stella L, Rossi C, Baldassarre G, Pantaleoni F, Cordeddu V, Williams BJ, Dentici ML, Caputo V, Venanzi S, Bonaguro M, et al. 2011. SOS1 mutations in Noonan syndrome: molecular spectrum, structural insights on pathogenic effects, and genotype-phenotype correlations. Hum Mutat 32:760–772.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Margarit SM, Sondermann H, Hall BE, Nagar B, Hoelz A, Pirruccello M, Bar-Sagi D, Kuriyan J. 2003. Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS. Cell 112:685–695.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303.

Nielsen KH, Papageorge AG, Vass WC, Willumsen BM, Lowy DR. 1997. The Ras-specific exchange factors mouse Sos1 (mSos1) and mSos2 are regulated differently: mSos2 contains ubiquitination signals absent in mSos1. Mol Cell Biol 17: 7132–7138.

Nimnual A, Bar-Sagi D. 2002. The two hats of SOS. Sci STKE 2002 :PE36.

Pagani MR, Oishi K, Gelb BD, Zhong Y. 2009. The phosphatase SHP2 regulates the spacing effect for long-term memory induction. Cell 139: 186–198.

Pandit B, Sarkozy A, Pennacchio LA, Carta C, Oishi K, Martinelli S, Pogna EA, Schackwitz W, Ustaszewska A, Landstrom A, Bos JM, Ommen SR, et al. 2007. Gain-of-function RAF1 mutations cause Noonan and LEOPARD syndromes with hypertrophic cardiomyopathy. Nat Genet 39:1007–1012.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.

Qian X, Esteban L, Vass WC, Upadhyaya C, Papageorge AG, Yienger K, Ward JM, Lowy DR, Santos E. 2000. The Sos1 and Sos2 Ras-specific exchange factors: differences in placental expression and signaling properties. EMBO J 19:642–654.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Razzaque MA, Nishizawa T, Komoike Y, Yagi H, Furutani M, Amo R, Kamisago M, Momma K, Katayama H, Nakagawa M, Fujiwara Y, Matsushima M, et al. 2007. Germline gain-of-function mutations in RAF1 cause Noonan syndrome. Nat Genet 39:1013–1017.

Roberts AE, Allanson JE, Tartaglia M, Gelb BD. 2013. Noonan syndrome. Lancet 381:333–342.

Roberts AE, Araki T, Swanson KD, Montgomery KT, Schiripo TA, Joshi VA, Li L, Yassin Y, Tamburino AM, Neel BG, Kucherlapati RS. 2007. Germline gain-of-function mutations in SOS1 cause Noonan syndrome. Nat Genet 39:70–74.

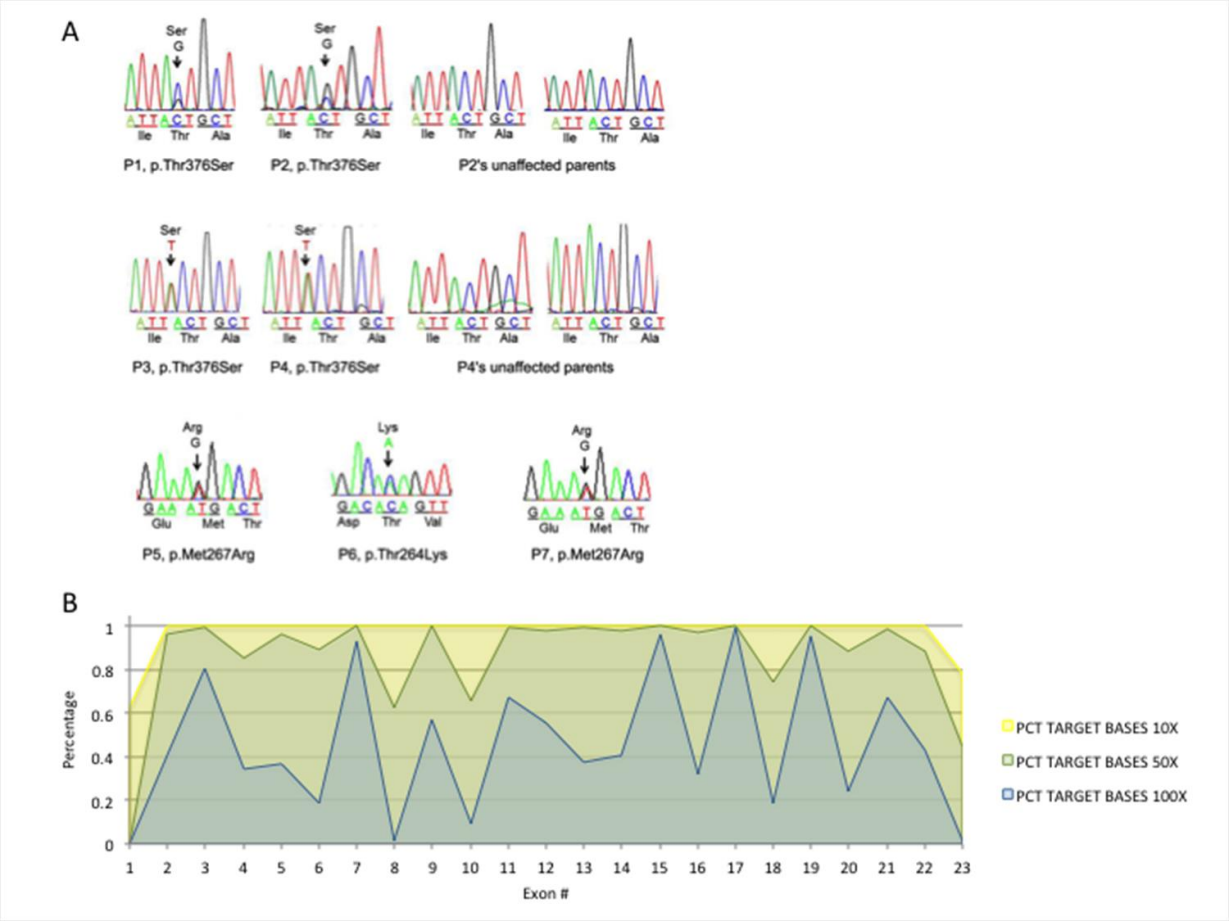
Sondermann H, Soisson SM, Boykevisch S, Yang SS, Bar-Sagi D, Kuriyan J. 2004. Structural analysis of autoinhibition in the Ras activator Son of sevenless. Cell 119:393–405.

Tartaglia M, Gelb BD. 2010. Disorders of dysregulated signal traffic through the RAS-MAPK pathway: phenotypic spectrum and molecular mechanisms. Ann N Y Acad Sci 1214:99–121.

Tartaglia M, Pennacchio LA, Zhao C, Yadav KK, Fodale V, Sarkozy A, Pandit B, Oishi K, Martinelli S, Schackwitz W, Ustaszewska A, Martin J, et al. 2007. Gain-of-function SOS1 mutations cause a distinctive form of Noonan syndrome. Nat Genet 39:75–79.

Tian X, Gotoh T, Tsuji K, Lo EH, Huang S, Feig LA. 2004. Developmentally regulated role for Ras-GRFs in coupling NMDA glutamate receptors to Ras, Erk and CREB. EMBO J 23:1567–1575.

Yamamoto GL, Aguena M, Gos M, Hung C, Pilch J, Fahiminiya S, Abramowicz A, Cristian I, Buscarilli M, Naslavsky MS, Malaquias AC, Zatz M, et al. 2015. Rare variants in SOS2 and LZTR1 are associated with Noonan syndrome. J Med Genet 52:413–421.



Supp. Figure S1. Germline *SOS2* mutations causing Noonan syndrome. **(A)** Electropherograms showing missense mutations in *SOS2* deemed pathologic (indicated with black arrows) in the various probands and relevant family members. **(B)** Coverage of exons of *SOS2* from whole exome sequencing.

Supp. Table S1. *SOS2* Mutations and Polymorphisms

FIRST COHORT: RASOPATHIES (N= 150); APPROACH: SANGER SEQUENCING; HIGH THROUGHPUT RESEQUENCING								
Variant (AA)	**Domain**	**NT Change**	**cDNA**	**HG19**	**Annotation (dpSNP124)**	**CADD Phred**	**# Cases**	**Notes**
D189E	HF/DH linker	A>T	c.567T>A	chr14:50655362	Novel	4.4	1	Uncharacterized variant
A208T	DH	C>T	c.622G>A	chr14:50655307	rs61755579	20.6	5	Polymorphism
T376S	**DH**	**G>C**	**c.1127C>G**	**chr14:50628269**	**Novel**	**11.0**	**1 (familial)[a]**	**Pathogenic**
D400N	DH	C>T	c.1198G>A	chr14:50626803	rs200368064	17.8	1	Uncharacterized variant
T476S	PH	G>C	c.1427C>G	chr14:50626574	Novel	11.8	1	Likely disease-unrelated
S483N	PH	C>T	c.1448G>A	chr14:50626553	rs17122201	1.5	4	Polymorphism
E707G	REM	T>C	c.2120A>G	chr14:50619829	rs369462490	15.4	4	Polymorphism
Q1235H	*C*-terminal region	C>G	c.3705G>C	chr14:50585356	Novel	14.2	1	Uncharacterized variant
W1248C	*C*-terminal region	C>A	c.3744G>T	chr14:50585317	rs138133010	6.5	1	Uncharacterized variant
SECOND COHORT: RASOPATHIES (N=54); APPROACH: WES								
Variant (AA)	**Domain**	**NT Change**	**cDNA**	**HG19**	**Annotation (dpSNP124)**	**CADD Phred**	**# Cases**	**Notes**
P191R	HF/DH linker	G>C	c.572C>G	chr14:50655357	rs72681869	16.0	1	Likely disease-unrelated
A208T	DH	C>T	c.622G>A	chr14:50655307	rs61755579	20.6	2	Likely disease-unrelated
T376S	**DH**	**T>A**	**c.1126A>T**	**chr14:50628270**	**Novel**	**11.1**	**1 (familial)[b]**	**Pathogenic**
D952N	PH	C>T	c.2854G>A	chr14:50605434	rs200387871	21.4	1	Likely disease-unrelated
N1161D	*C*-terminal region	T>C	c.3481A>G	chr14:50587001	rs371703424	12.2	1	Likely disease-unrelated
THIRD COHORT: NOONAN SYNDROME (N=61); APPROACH: TARGETED RESEQUENCING								
Variant (AA)	**Domain**	**NT Change**	**cDNA**	**HG19**	**Annotation (dpSNP124)**	**CADD Phred**	**# Cases**	**Notes**
P191R	HF/DH linker	G>C	c.572C>G	chr14:50655357	rs72681869	16.0	3	Uncharacterized variant
L197L	DH	T>C	c.591A>G	chr14:50655338	rs113460230	9.6	7	Uncharacterized variant
A208T	DH	C>T	c.622G>A	chr14:50655307	rs61755579	20.6	12	Polymorphism
T264K	**DH**	**G>T**	**c.791C>A**	**chr14:50649248**	**Novel**	**31.0**	**1**	**Pathogenic**
M267R	**DH**	**A>C**	**c.800T>G**	**chr14:50649239**	**Novel**	**23.8**	**2**	**Pathogenic**
S483N	PH	C>T	c.1448G>A	chr14:50626553	rs17122201	1.5	2	Polymorphism
N1257D	*C*-terminal region	T>C	c.3769A>G	chr14:50585292	rs150393358	2.0	1	Uncharacterized variant
P1318S	*C*-terminal region	G>A	c.3952C>T	chr14:50585109	rs140995728	4.0	1	Uncharacterized variant

[a] Mother-son pair. The variant was documented to occur *de novo* in the affected mother.

[b] Mother-daughter pair. The variant was documented to occur *de novo* in the affected mother.

Publikace 12

Molecular genetic analysis in 14 Czech Kabuki syndrome patients is confirming the utility of phenotypic scoring.

Paderova J, Holubova A, Simandlova M, Puchmajerova A, Vlckova M, Malikova M, Pourova R, Vejvalkova S, Havlovicova M, Senkerikova M, Ptakova N, Drabova J, Geryk J, Maver A, Krepelova A, Macek M Jr.

Clin Genet. 2016 Feb 4. doi: 10.1111/cge.12754. [Epub ahead of print]

V publikaci jsou shrnuty naše dosavadní zkušenosti s klinickou a molekulární diagnostikou syndromu Kabuki (KS), jehož genetická příčina byla objevena teprve nedávno. KS je AD onemocnění nejčastěji způsobené *de novo* mutací v genech *KMT2D* (lysine/K/specific methyltransferase 2D) a *KDM6A* (lysine/K/specific demethylase 6A). Na pracovišti jsme vyšetřili 14 pacientů s fenotypem poukazujícím na KS. U všech 14 pacientů byla provedena sekvenace celé kódující oblasti genů *KMT2D* a *KDM6A* a MLPA analýza k ozřejmění intragenových delecí či duplikací. Patogenní mutace v genu *KMT2D* byla nalezena u 6/14 (43%) pacientů. Všechny mutace byly ztrátové (zejména nonsense). Tři z námi detekovaných mutací byly již v literatuře u KS popsány a tři mutace byly nové. U žádného pacienta nebyla detekována mutace v genu *KDM6A* a nebyla nalezena ani žádná intragenová delece či duplikace v žádném z genů. Pacienti bez prokázané mutace (8/14) pak byli vyšetřeni metodou aCGH s cílem identifikovat případné další lokusy, které by mohly být asociovány s KS. U jedné pacientky byla detekována potenciálně patogenní varianta – 6,6 Mb dlouhá duplikace Xp21.2-p21.3.

Pro analýzu fenotypu jsme použili podrobný klinický dotazník vycházející ze skórovacího systému „MLL2-Kabuki-skóre“ (Makrythanasis et al., 2013). Pacienti s potvrzenou mutací dosáhli skóre 7 a více (v průměru 8), pacienti bez mutace měli v průměru skóre 4,9. Statistická analýza ukázala, že rozdíl je statisticky významný.

Práce potvrzuje, že sekvenace kódující oblasti genu *KMT2D* a aCGH u negativních pacientů má vysokou diagnostickou výtěžnost, zároveň však poukazuje na nutnost precizní analýzy fenotypu, která by měla předcházet indikaci k molekulárně genetickému vyšetření. Z naší práce dále vyplývá, že použití „MLL2-Kabuki skóre“ má vysokou prediktivní hodnotu a je efektivním vodítkem k indikaci cíleného molekulárně genetického vyšetření. Práce zároveň ukazuje účelnost podrobných klinických dotazníků v případech, kdy se jedná o dobře klinicky definované syndromické jednotky.

CLINICAL GENETICS An International Journal of Genetics, Molecular and Personalized Medicine

Clin Genet 2016
Printed in Singapore. All rights reserved

© 2016 John Wiley & Sons A/S.
Published by John Wiley & Sons Ltd

CLINICAL GENETICS
doi: 10.1111/cge.12754

Original Article

Molecular genetic analysis in 14 Czech Kabuki syndrome patients is confirming the utility of phenotypic scoring

Paděrová J., Holubová A., Simandlová M., Puchmajerová A., Vlčková M., Malíková M., Pourová R., Vejvalková S., Havlovicová M., Šenkeříková M., Ptáková N., Drábová J., Geryk J., Maver A., Křepelová A., Macek Jr. M. Molecular genetic analysis in 14 Czech Kabuki syndrome patients is confirming the utility of phenotypic scoring.
Clin Genet 2016. © John Wiley & Sons A/S. Published by John Wiley & Sons Ltd, 2016

Kabuki syndrome (KS) is a dominantly inherited disorder mainly due to *de novo* pathogenic variation in *KMT2D* or *KDM6A* genes. Initially, a representative cohort of 14 Czech cases with clinical features suggestive of KS was analyzed by experienced clinical geneticists in collaboration with other specialties, and observed disease features were evaluated according to the 'MLL2-Kabuki score' defined by Makrythanasis et al. Subsequently, the aforementioned genes were Sanger sequenced and copy number variation analysis was performed by MLPA, followed by genome-wide array CGH testing. Pathogenic variants in *KMT2D* resulting in protein truncation in 43% (6/14; of which 3 are novel) of all cases were detected, while analysis of *KDM6A* was negative. MLPA analysis was negative in all instances. One female patient bears a 6.6 Mb duplication of the Xp21.2–Xp21.3 region that is probably disease causing. Subjective KS phenotyping identified predictive clinical features associated with the presence of a pathogenic variant in *KMT2D*. We provide additional evidence that this scoring approach fosters prioritization of patients prior to *KMT2D* sequencing. We conclude that *KMT2D* sequencing followed by array CGH is a diagnostic strategy with the highest diagnostic yield.

Conflict of interest

The authors have no conflict of interest that could influence the content of this manuscript.

J. Paděrová[a], A. Holubová[a], M. Simandlová[a], A. Puchmajerová[a], M. Vlčková[a], M. Malíková[a], R. Pourová[a], S. Vejvalková[a], M. Havlovicová[a], M. Šenkeříková[b], N. Ptáková[a], J. Drábová[a], J. Geryk[a], A. Maver[c], A. Křepelová[a] and M. Macek Jr. [a]

[a]Department of Biology and Medical Genetics, Charles University Prague-2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic, [b]Department of Medical Genetics, Charles University Prague-Faculty of Medicine and University Hospital Hradec Králové, Hradec Králové, Czech Republic, and [c]Department of Obstetrics & Gynecology, Centre for Mendelian Genomics, Clinical Institute of Medical Genetics, University Medical Centre Ljubljana, Ljubljana, Slovenia

Key words: array CGH – Kabuki syndrome – *KDM6A* – *KMT2D* – *MLL2* – MLL2-Kabuki score – MLPA – phenotyping – Sanger DNA sequencing

Corresponding author: Prof Milan Macek Jr., MD, DSc., Department of Biology and Medical Genetics, Charles University Prague-2nd Faculty of Medicine and University Hospital Motol, V Úvalu 84, 150 06 Prague 5, Czech Republic.
Tel.: +420 2 2443 3501;
fax: +420 2 2443 3520;
e-mail: milan.macek.jr@lfmotol.cuni.cz;
www.fmotol.cz/ublg

Received 7 October 2015, revised and accepted for publication 30 January 2016

Kabuki (make-up) syndrome (KS) is a multiple malformation syndrome, originally described in Japan more than three decades ago (1, 2). Clinically, KS is characterized by intellectual disability, developmental delay, postnatal growth retardation, hypotonia, congenital heart and renal malformations, persisting

fetal finger pads and, most uniquely, by typical facial features.

KS is mainly due to *de novo* pathogenic variants (henceforward termed in legacy nomenclature as 'mutations'). However, familial cases have also been reported (3–6), suggesting that variable expressivity may play a role in some instances.

Recently, two molecular subtypes of KS have been described. Kabuki syndrome 1 (KABUK1; MIM# 147920) is caused by mutations or large deletions in the *KMT2D* gene (formerly termed *MLL2* gene; lysine/K/specific MethylTransferase 2D) (MIM# 602113) located in the 12q13.12 chromosomal region and spanning 54 exons. *KMT2D* belongs to the Trithorax group of genes (methyltransferases and transcriptional activators) that are particularly important in *HOX* gene activation during embryonic development. Methyltransferase activity of the KMT2D protein is conferred by its C-end SET domain coded by exons 51–53.

The involvement of *KMT2D* in the development of KS was previously documented by Sanger and/or exome sequencing (7), whereby mutations were found in 44–76% of KABUK1 patients (3, 7–16). KABUK1-associated mutations can arise anywhere within the *KMT2D* gene. Thus far, missense, nonsense, and frameshift mutations in various exons, including splice site alterations, have been linked to this syndrome. Truncating mutations are most commonly observed. Missense mutations occur more frequently in exon 48, which codes several functional protein domains (7, 9, 13). The majority of mutations are unique and confined to individual cases, while few alleles are recurrent at very low rates (3, 7–16). Mosaic mutations have also been published (16). Finally, *KMT2D* can be affected by large deletions leading to a loss of the entire gene or its parts (16).

Recently, Makrythanasis et al. established a phenotypic scoring system (henceforth referred to as the 'MLL2-Kabuki score') as a means of distinguishing between those patients that have a *KMT2D* mutation, and those who are unlikely to bear a mutation in the gene (3). This system is based on a discrete subset of KS phenotypic features associated with *KMT2D* mutations, each of which assigns a different relative weight to the overall score and thus the likelihood of *KMT2D* being involved in the pathogenesis of KS.

Kabuki syndrome 2 (KABUK2; MIM# 300867) is caused by pathogenic variants or large deletions in the *KDM6A* gene (UTX; MIM# 300128) located at Xp11.3. The gene escapes X-inactivation and comprises 29 exons. The KDM6A protein acts as a demethylase and promotes transcription of downstream genes. Various point mutations scattered all over the gene have been found in several patients with KS (6, 11, 17). In addition, the *KDM6A* gene was found to be deleted (in part or whole) in some KS patients (17, 18). Mutations in *KDM6A* explain the development of KS in approximately 1–6% of all cases (6, 11, 17).

Currently, 24–56% of clinically diagnosed KS patients do not have genomic alterations in the aforementioned genes, suggesting that this syndrome is genetically heterogeneous. Several studies have reported results from molecular cytogenetic (array comparative genomic hybridization; henceforward array CGH) analyses in *KMT2D* mutation-negative (KMT2D[–]) cases of KS (4, 8, 19). Various forms of structural variation elsewhere in the genome (including mosaics) were found indicating that other clinical entities, such as the Sotos syndrome (MIM# 117550) (4), could in some instances, clinically present under a 'KS-like phenotype'.

We studied a representative cohort of 14 Czech KS patients who underwent interdisciplinary clinical evaluation. Our objectives were to: (i) replicate and complement previous studies using Sanger DNA sequencing of *KMT2D* and *KDM6A* coding regions and adjacent splice sites, together with MLPA-based (multiplex ligation-dependent probe amplification) copy number variation (CNV) analysis; (ii) perform genome-wide array CGH in cases where the *KMT2D* and *KDM6A* loci remained negative following sequencing and MLPA analysis; (iii) confirm the clinical diagnosis of KS in our patients, and ultimately facilitate informed reproductive choices in affected families; (iv) assess genotype–phenotype correlations; (v) establish the most efficient clinical and molecular diagnostics algorithm, and (vi) evaluate the use of the 'MLL2-Kabuki score' as a prioritization tool for selecting patients for *KMT2D* molecular analysis.

Materials and methods

Subjects

Fourteen patients (KS1–KS14) aged 1–15 years (10 males/4 females) had been referred by collaborating pediatric primary care or other pediatric specialties to two tertiary Czech academic medical genetic centers (located in Prague and Hradec Králové Charles University teaching hospitals) for differential diagnosis of their dysmorphic features, intellectual disability and/or their unexplained postnatal growth delay. These centers cover most of the western parts of the country (approximately 4 million inhabitants) in terms of clinical dysmorphology. In addition, three adult apparently healthy controls were analyzed in order to validate 'in house' designed sequencing primers and assess the presence of observed variants in analyzed genes in the general population. Following institutional ethical approval, an informed consent was obtained from all parents of affected minors enrolled in the study. In families, where the *KMT2D* mutation was found in the index case, parental testing (when possible) was carried out.

Methods

Studied patients underwent thorough clinical phenotyping by experienced clinical geneticists in order to establish the 'MLL2-Kabuki score' (3) prior to *KMT2D* molecular genetic analysis in all instances (Table 1). Subsequently, additional clinical, laboratory and medical imaging features known to be associated with KS

Table 1. Clinical features observed in Czech KS patients[a]

Patient	KS1	KS3	KS4	KS5	KS12	KS13	KMT2D[+]	KS2	KS6	KS7	KS8	KS9	KS10	KS11	KS14	KMT2D[–]
Gender	M	M	M	M	F	M	–	M	M	F	F	F	M	M	M	–
General clinical features																
Postnatal growth retardation	√	√	√	√	√	√	6/6	√			√		√	√		4/8
Microcephaly		√	√	√	√	√	5/6	√					√	√		3/8
Feeding difficulties		√	√	√	√	√	5/6	√			√		√		√	4/8
Facial features																
Arched eyebrow	√	√	√	√		√	5/6	√		√	√	√	√	√	√	7/8
Everted lower eyelid	√	√	√	√		√	5/6	√	√	√	√	√		√		6/8
Long palpebral fissures	√	√	√	√	√	√	6/6		√	√	√	√	√	√	√	7/8
Ptosis	√	√		√		√	4/6			√		√			√	3/8
Strabismus	√	√	√	√	√		5/6			√					√	2/8
Blue sclerae				√	√	√	3/6		√		√					2/8
Coloboma			√				1/6									0/8
Large dysplastic ears	√	√	√	√	√	√	6/6	√		√	√	√	√	√	√	7/8
Broad nasal root	√	√	√	√	√	√	6/6	√	√	√	√	√	√	√	√	8/8
Flat nasal tip	√	√	√	√	√		4/6		√		√				√	3/8
Short columella	√	√	√	√	√	√	6/6	√	√	√	√			√	√	6/8
Flat philtrum		√	√	√	√	√	5/6	√	√	√	√	√	√	√	√	8/8
Abnormal dentition	√		√				2/6		√					√	√	3/8
Oligodontia	√						1/6									0/8
High or cleft palate	√	√	√	√	√		5/6		√	√		√			√	4/8
Small mandible			√	√		√	3/6				√	√	√		√	4/8
Thin upper lip, full lower lip	√	√	√	√	√		5/6		√	√	√	√	√	√	√	7/8
Skeletal and limb anomalies																
Brachytactyly or Clinodactyly	√	√	√	√	√	√	6/6			√	√	√		√		4/8
Hypoplastic fingers	√	√		√	√	√	5/6									0/8
Hypoplastic nails		√	√	√	√	√	5/6									0/8
Persistent fetal finger pads	√	√	√	√		√	5/6	√	√	√			√	√		5/8
Dysplastic hips				√	√		2/6									0/8
Joint hypermobility	√	√		√	√		4/6		√	√		√	√		√	5/8
Abnormal dermatoglyphics					√		1/6		√					√	√	3/8
Neurological features																
Intellectual disability	√	√	√	√	√	√	6/6	√	√	√	√	√	√	√	√	8/8
Developmental delay	√	√	√	√	√	√	6/6	√	√		√	√	√	√	√	7/8
Hypotonia	√	√	√	√	√	√	6/6		√	√	√	√	√	√	√	7/8
Seizures or abnormal electroencephalogram	√						1/6							√	√	2/8
Organ anomalies																
Congenital heart defect	√	√	√	√	√	√	6/6	√	√				√			3/8
Congenital renal anomalies			√		√	√	3/6						√			1/8
Other features																
Immune disorders	√	√	√	√			4/6		√		√					2/8
Hearing loss/frequent otitis media	√					√	2/6									0/8
Abnormalities of testes	√			√	N/A	√	3/5		√	N/A	N/A	N/A				1/5
Premature telarche in females	N/A	N/A	N/A	N/A		N/A	0/1	N/A	N/A				N/A	N/A	N/A	0/3
Cataract	√						1/6									0/8
Autism	√	√				√	3/6		√	√						2/8
MLL2-Kabuki score	7	8	9	8	8	8	8	5	5	4	4	4	7	6	4	4.875

F, female; M, male; N/A, non-applicable.

[a]KMT2D[+]: frequency of phenotypic features in *KMT2D*-mutation-positive patients; KMT2D[–]: frequency of phenotypic features in *KMT2D*-mutation-negative patients; common features are highlighted in gray. Note: this score can reach a maximum of 10 points comprising up to 5 points for 1–15 defined KS facial features; maximum of 1 point for 2–4 'positive' limb and extremities features and 1 point for each for the presence of microcephaly, short stature, heart and kidney anomalies (i.e. listed under 'General clinical features' and 'Organ anomalies' in the table).

were substantiated by other medical specialties within a standard multidisciplinary collaboration (Table 1.)

A 'cascade' molecular genetic testing strategy had been performed using template DNA isolated from peripheral blood white cells. Initially, Sanger DNA sequencing of the entire *KMT2D* coding sequence, including adjacent splice sites, was performed. Previously published (4, 7) and newly designed ('Primer BLAST' tool; NCBI, USA) primer sequences were used. Subsequently, in patients who remained negative, the *KDM6A* gene was analyzed in a similar manner (primer sequences for the analysis of both genes are available upon request).

Sequencing reactions were carried out using a Big Dye Terminator v3.1 Cycle Sequencing Kit, capillary electrophoresis on a 3130xl Genetic Analyzer and Sequencing Analysis software version v5.3 (Applied Biosystems, Forster City, CA, USA). The National Centre for Biotechnology Information (NCBI) cDNA sequences (GenBank:

Table 2. Overview of pathogenic *KMT2D* mutations in Czech KS patients

Case/gender	Mutation cDNA level	Mutation protein level	Exon	Mutation type	Parents tested (father/mother)	Mode of inheritance
KS1/M	c.16371_16374delTGAA[a]	p.Glu5458Metfs*2	52	FS	NA	NA
KS3/M	c.8743C>T[a,b]	p.Arg2915*	34	NS	NA	NA
KS4/M	c.2488G>T	p.Glu830*	10	NS	F-/M-	*de novo*
KS5/M	c.4549_4549delG	p.Glu1517Argfs*4	16	FS	F-/M-	*de novo*
KS12/F	c.6349_6350delinsA	p.Pro2117Thrfs*27	31	FS	F-/M-	*de novo*
KS13/M	c.5625_5628delAGAC[a,c]	p.Asp1876Glyfs*38	25	FS	M-	NA

cDNA position assignment according to Ref Seq NM_003482.3; F, female; F-, father tested negative for *KMT2D* mutation; FS, frameshift mutation; M, male; M-, mother tested negative for *KMT2D* mutation; NA, not available (i.e. parents not tested); NS, nonsense mutation.

[a]Makrythanasis et al. (3)

[b]Li et al. (15)

[c]Banka et al. (9)

RefSeq NM_003482.3 for *KMT2D* gene and RefSeq NM_021140.2 for *KDM6A* gene) were used as a reference. All observed variants were described using Human Genome Variation Society nomenclature (www.hgvs.org/mutnomen). Synonymous mutations were considered as benign providing that the splice site was not newly created or disrupted. Nonsense and frameshift mutations were considered pathogenic because they are most probably associated with haploinsufficiency. Missense mutations were initially assessed using the Short Genetic Variation database – dbSNP (www.ncbi.nlm.nih.gov/projects/SNP/), and mutation pathogenicity was evaluated using MutationTaster (www.mutationtaster.org/), PolyPhen-2 (genetics.bwh.harvard.edu/pph2/), SIFT/Provean (provean.jcvi.org/index.php), Align GVGD (agvgd.iarc.fr/) and MutPred (mutpred.mutdb.org/) prediction algorithms. Potential splice site mutations were evaluated using Alamut software version v2.3 (Interactive Biosoftware, Rouen, France).

Simultaneously, the MLPA method (MRC, the Netherlands; probe mixes P389-A1 and P445-A1) was used to detect CNV in *KMT2D* and *KDM6A*. Outcomes of these assays were analyzed using a freely available in-house developed software 'e-MLPA' for data normalization and their analysis (emlpa.lf2.cuni.cz).

Furthermore, in eight *KMT2D*/*KDM6A* mutation-negative patients, we made use of array CGH using the CytoChip ISCA 8x60K assay. Results of these tests were analyzed using BlueFuse Multi version v4.0 software (Illumina, San Diego, USA). Finally, in patient KS9, the chromosome localization of the molecular cytogenetic finding was confirmed by fluorescent *in situ* hybridization (FISH) with the use of Aquarius Whole Chromosome Painting probe for chromosome X (WCP X; Cytocell, Cambridge, UK).

Hypothesis assessment was carried out by the R-Project for Statistical Computing software suite (www.r-project.org). In order to test the hypothesis that patients who bear a mutation in *KMT2D* (KMT2D[+]) have higher values of the 'MLL2-Kabuki score' than KMT2D[−] cases, we performed Wilcoxon rank sum test. The effect of gender and age on the 'MLL2-Kabuki score' had been modeled by cumulative link models implemented in package ordinal. The 'MLL2-Kabuki score' was statistically assessed as function of linear combination of gender, age and binary variable determining to which group, i.e. KMT2D[+] or KMT2D[−], a KS patient belongs. Finally, we utilized the hypergeometric test in order to characterize the difference between selected groups of specific clinical features.

Results

Detailed clinical features of all studied KS cases are outlined in Table 1. and the 'MLL2-Kabuki score' ranged between 4 and 9 points. Six out of 14 patients (43%) were found to have a disease-causing mutation in *KMT2D* (Table 2, Fig. 1.). Three mutations were previously published (3, 9), while the remaining mutations are novel. All mutations, located in exons 10, 16, 25, 31, 34 and 52, result in the truncation of the KMT2D protein product (4× frameshift and 2× nonsense mutations; five of them were detected in male patients and one in a female patient). In three KMT2D[+], parental DNA was available and thus tested, which allowed us to determine the *de novo* status of the detected mutations.

We found a missense mutation in *KMT2D* in four cases (c.3790C>G, c.6629C>T, c.7716G>T, c.15053C>T); all of them originally absent in the dbSNP database (accessed January 31 2014). Initially, these missense mutations had been considered as being potentially pathogenic and further evaluated using various *in silico* prediction tools. However, contradictory outcomes in terms of their pathogenicity were found in all instances (data now shown). Subsequently, variants c.3790C>G and c.15053C>T were detected by DNA sequencing in apparently healthy controls and the remaining later became listed in dbSNP as benign.

Intragenic *KMT2D* or *KDM6A* structural variation was not observed by MLPA in any of the patients examined. Likewise, mutations within the *KDM6A* were not detected by sequencing.

Array CGH examination of *KMT2D*/*KDM6A* mutation-negative patients revealed a 6.6 Mb *de novo* duplication on chromosome X (hg19, chrX:24,942,199-31,315,935) in a female patient (KS9). Subsequent analysis by FISH utilizing the WCP

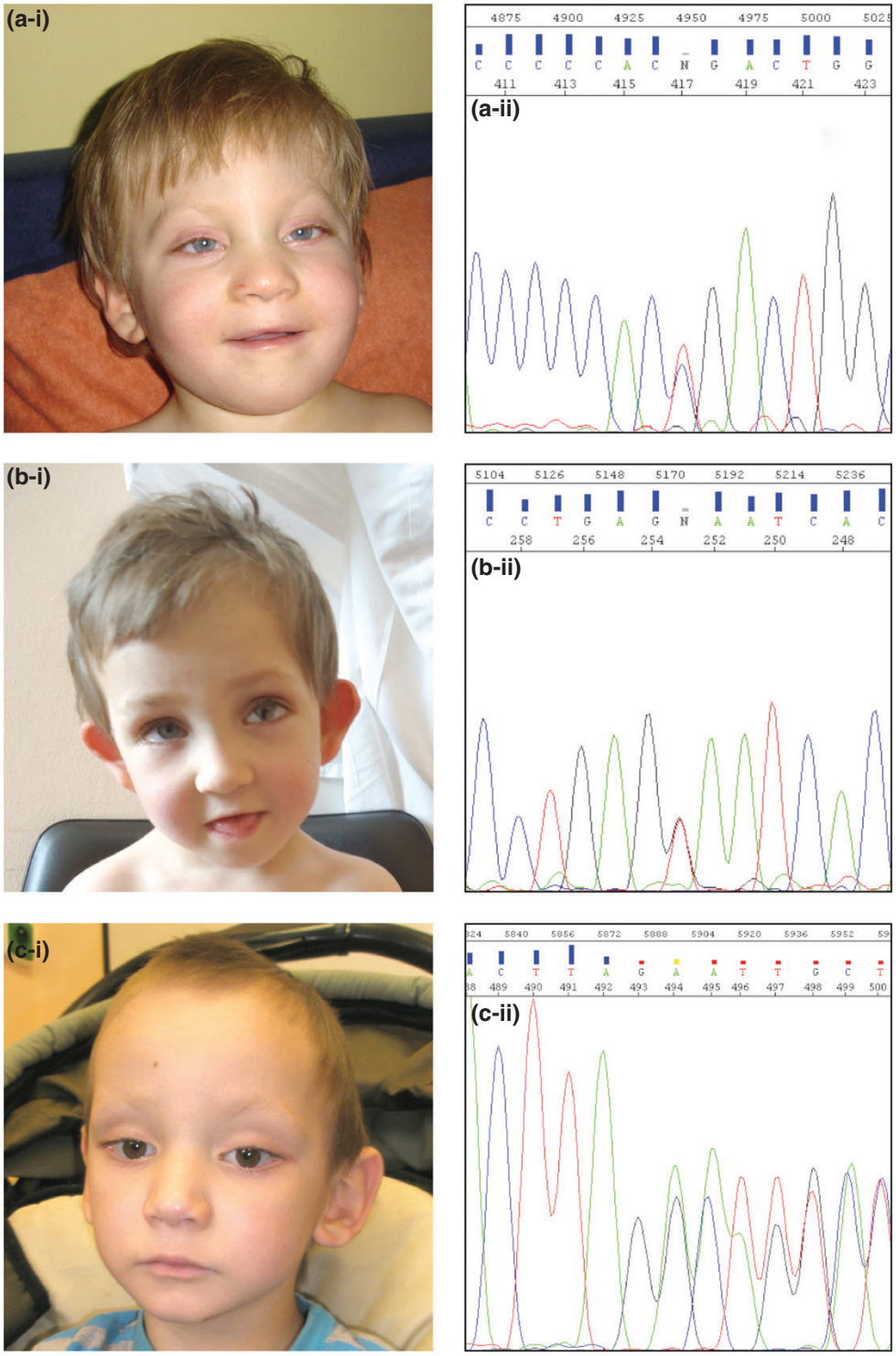


Fig. 1. Photographs of representative *KMT2D* mutation-positive patients and sequencing electrophoretograms. (**a-i**, **b-i**, **c-i**) Photographs of KS patients with a *KMT2D* mutation showing typical KS facial features such as arched eyebrows, long palpebral fissures, broad nasal root and large dysplastic ears. (**a-i**) Patient KS3; (**a-ii**) nonsense mutation in KS3 c.8743C>T in exon 34; (**b-i**) patient KS4; (**b-ii**) nonsense mutation in KS4 c.2488G>T in exon 10; (**c-i**) patient KS13; (**c-ii**) frameshift mutation in KS13 c.5625_5628delAGAC in exon 25.

X probe showed that the duplicated region is not located elsewhere in the genome, thus being highly suggestive of a tandem duplication present on this chromosome (data available upon request).

All KMT2D[+] patients shared characteristic facial features (as highlighted in gray in Table 1) comprising long palpebral fissures, abnormal ears, broad nasal root and short columella, brachydactyly or clinodactyly, short stature, hypotonia, developmental delay, intellectual disability, and congenital heart defects such as coarctation of the aorta, ventricular septal or valvular defects. Hypoplastic nails and fingers were present only in KMT2D[+] patients, each in five of six patients. Strabismus was significantly more prevalent in KMT2D[+] patients ($p = 0.0023$) (Fig. 2). The presence of other disease features varied among KMT2D[+] as well as among KMT2D[−] patients (Table 1).

The 'MLL2-Kabuki score', yielded an average score of 8 (range: 7–9) for the KMT2D[+] group and an average score of 4.875 (range: 4–7) for the KMT2D[−] group. On the basis of the Wilcoxon rank sum test, this score is significantly greater in the KMT2D[+] group than in KMT2D[−] group ($p = 0.0012$). The cumulative link modeling provided additional evidence that the score is significantly higher in the KMT2D[+] group ($p = 0.0022$) and that male patients have a higher score ($p = 0.0183$). Lip nodules, a feature originally published by Makrythanasis et al. (3), were not observed in any of our patients and thus are not included in Table 1.

Discussion

This is the first comprehensive molecular genetic study in a representative cohort of Czech KS patients that had been ascertained at two tertiary medical genetics departments with specialized clinical dysmorphology expertise. Moreover, to our knowledge, this is the largest Central and Eastern European cohort to date analyzed.

The observed overall mutation detection rate in the *KMT2D* (43%) was comparable with previous studies – 44% (14), 48% (13) and 52% (3). All mutations detected in our study led to the truncation of the *KMT2D* gene product. The position of the respective mutation within this gene is unlikely to have any phenotypic significance, because all 'deactivate' the functional SET domain which is located at the 3′ end of the gene.

The gender bias in terms of more KMT2D[+] males being observed, is most probably due to the small number of analyzed KS cases in our study. Overall, the phenotypic scoring was significantly higher in all KMT2D[+] patients, making the 'MLL2-Kabuki score' confidently predictive for the presence of a *KMT2D* mutation. Based on this outcome, we have included the assessment of this score as a part of routine pre-testing clinical evaluation in patients suspected of having KS. In this regard, scores ranging between 6 and 9 are considered as an indication for targeted molecular genetic analysis of *KMT2D*, generally in line with Makrythanasis et al. (3) but bearing in mind the smaller size of the Czech cohort with regards to the score range. Hence, scores of 5 and lower are warranting additional clinical examinations, including medical imaging in the search for other organ anomalies (Table 1). In addition, such cases rather ought to be subjected to genome-wide analyses comprising array CGH, followed by, e.g. exome sequencing.

As our study was carried out prior to introduction of next generation sequencing (NGS) in our laboratories, and given the fact that NGS is currently becoming more cost effective than Sanger DNA sequencing, this strategy is currently preferable not only in 'lower' 'MLL2-Kabuki scores', but also for cases within its 'indicative range'. Nonetheless, accurate phenotyping prior to genetic testing will still play an important role as a vast amount of variants produced by novel technologies need to be interpreted accordingly.

The utility of 'MLL2-Kabuki score' may further be improved by complementing subjective expert interpretation of dysmorphic features with two dimensional (2D) and/or three dimensional (3D) facial morphometrics, as half of the scoring points are based on the assessment of facial gestalt. It may be as well useful to discern major and minor facial features for KS in larger cohorts and attribute these features a different value in the scoring system, as some of them do not appear to be as common (e.g. lip nodules) as others (e.g. long palpebral fissures). Furthermore, 3D morphometrics analysis may foster recognition of facial features in younger KS patients that may be subjectively less noticeable. The Prague center has recently installed the 3dMD system for 3D facial analyses (www.3dmd.com) which we intend to use in clinical practice and research in KS, and beyond.

All six KMT2D[+] patients from our group shared particular clinical features such as short stature, hypotonia, psychomotor retardation, brachydactyly or clinodactyly, including typical KS facial features and various heart disorders (as highlighted in gray in Table 1), suggesting that these are linked to the presence of a *KMT2D* mutation. On the other hand, some features considered that typical and highly sensitive for KS were observed in patients from both groups, i.e. in KMT2D[+] and KMT2D[−] cases, making them less specific 'predictors' for the presence of a *KMT2D* mutation (e.g. long palpebral fissures, everted lower eyelids, arched eyebrows, large ears or persistent fetal finger pads; Table 1). Previously reported renal anomalies in KMT2D[+] patients (12) were present in 50% (3/6) of our KMT2D[+] and in 12.5% (1/8) of KMT2D[−] patients similarly, as reported elsewhere (4).

Comparable with some other studies (4, 12), mutations within the *KDM6A* were not found by sequencing, because the overall detection rate in this gene is generally lower than 6% (11, 14, 17).

MLPA analysis of the *KMT2D* also did not lead to the detection of CNVs, comparable with the study of a sizeable Italian cohort (20). Nonetheless, rare CNVs detected by MLPA in *KMT2D* in 3 out of 64 patients were observed by others (16). In addition, we did not detect CNVs within *KDM6A* using MLPA, while others found partial or entire deletions of this gene (6, 17).

A 6.6-Mb duplication on chromosome X was found by array CGH in a female patient (KS9). This structural variation encompasses 37 genes involving *ARX* (MIM#

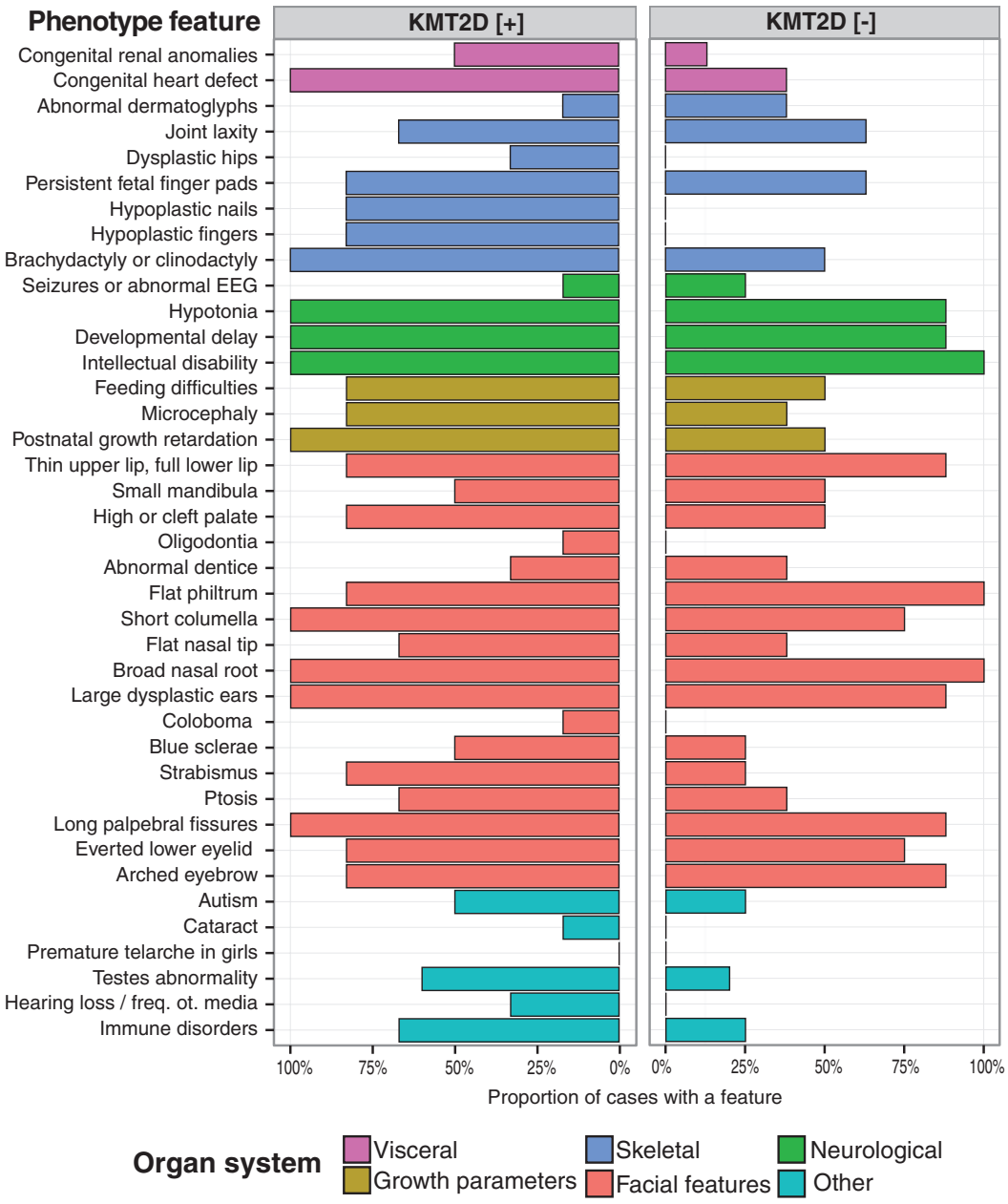


Fig. 2. Graphical presentation of phenotypic features in patients with and without mutation in the *KMT2D* gene. KMT2D [+]: percentage of patients with given phenotypic features within the group of six patients with a mutation in *KMT2D*; KMT2D [−]: percentage of patients with a given phenotypic features within the group of eight patients without mutation in *KMT2D*.

300382) and *IL1RAPL1* (MIM# 300206) and produces a breakpoint in *DMD* (MIM# 300377). *ARX* is a homeobox transcription factor responsible for embryonic brain development and maintenance of subpopulations of cortex neurons. *IL1RAPL1* regulates synapse establishment and synaptic transfer. Various types of mutation in *ARX* and *IL1RAPL1* are associated with intellectual disability (21–24). Some female carriers of a non-functional DMD gene, with skewed X-inactivation, show signs of muscle weakness (25–27). If the X chromosomes, affected by this molecular alteration, were not fully subjected to random X inactivation in all tissues, we could speculate that intellectual disability and hypotonia in this patient are the consequence.

In the DECIPHER database (decipher.sanger.ac.uk), we found a duplication of a similar extent as in KS9 present in a female suffering from global development delay (ID: 292724). However, in this instance, the entire

DMD gene was involved in this duplication and thus did not contain a breakpoint. Another female patient (ID: 262074) listed in this resource had a duplication of a region distal to duplication seen in KS9, with only one overlapping gene (*ARX*). This patient presented with intellectual disability, low-set ears, short stature and down slanted palpebral fissures. Nonetheless, the relationship of this variation to the KS-like phenotype in our case still may be spurious.

In 7 out of 14 KS patients, no pathogenic variation was observed. We speculate that this could be due to: (i) pathogenic variants present in the non-coding parts of the analyzed genes; (ii) low frequency mosaics that went undetected by our methodological approach; (iii) mutations in exons not covered by the current version of MLPA assays; (iv) intra- or intergenic rearrangements with breakpoints within *KMT2D* or *KDM6A*; (v) mutations in other genes involved in KS development; or (vi) spurious KS-like clinical features resulting from another clinical syndrome unrelated to KS.

In summary, our results indicate that targeted *KMT2D* sequencing has the highest diagnostic yield in KS patients. In addition, our results corroborate that array CGH in KMT2D[−] cases is indicated as suggested by others (4, 9, 13), followed by exome sequencing which is currently becoming more cost effective. We have provided additional evidence that distinct phenotypic features assessed subjectively by an experienced clinical geneticist and weighed by the 'MLL2-Kabuki score' are predictive for the presence of a *KMT2D* mutation and allow prioritization of patients undergoing *KMT2D* molecular diagnostics.

Acknowledgements

We thank all families for their participation in the study. We are grateful to Dr Jeremy A. Squire (Sao Paulo, Brasil) for helpful comments on this article, acknowledge Dr Jan Vejvalka (Prague, Czech Republic) for the development and management of the 'e-MLPA' software and Ms Ivana Funková for administrative support. We thank our colleagues in primary care, pediatrics, pediatric-cardiology, neurology, psychiatry, orthopedics and medical radiology for establishing other clinical and imaging features of the disease. This study was supported by CZ.2.16/3.1.00/24022OPPK; Conceptual Development Project of Research Organization #00064203 for University Hospital Motol; Ministry of Health projects no. NT/13770-4/2012; Norway Grants – NF-CZ11-PDP-3-003-2014, LM2015091 and COST – LD14073 to M. M. Jr., and NT/14200 to M. H.

References

1. Kuroki Y, Suzuki Y, Chyo H, Hata A. A new malformation syndrome of long palpebral fissures, large ears, depressed nasal tip, and skeletal anomalies associated with postnatal dwarfism and mental retardation. J Pediatr 1981: 99: 570–573.
2. Niikawa N, Matsuura N, Fukushima Y, Ohsawa T, Kajii T. Kabuki make-up syndrome: a syndrome of mental retardation, unusual facies, large and protruding ears, and postnatal growth deficiency. J Pediatr 1981: 99: 565–569.
3. Makrythanasis P, van Bon B, Steehouwer M et al. MLL2 mutation detection in 86 patients with Kabuki syndrome: a genotype-phenotype study. Clin Genet 2013: 84: 539–545.
4. Hannibal MC, Buckingham KJ, Ng SB et al. Spectrum of MLL2 (ALR) mutations in 110 cases of Kabuki syndrome. Am J Med Genet A 2011: 155A: 1511–1516.
5. Kokitsu-Nakata NM, Petrin AL, Heard JP et al. Analysis of MLL2 gene in the first Brazilian family with Kabuki syndrome. Am J Med Genet A 2012: 158A: 2003–2008.
6. Lederer D, Shears D, Benoit V, Verellen-Dumoulin C, Maystadt I. A three generation X-linked family with Kabuki syndrome phenotype and a frameshift mutation in KDM6A. Am J Med Genet A 2014: 164A: 1289–1292.
7. Ng SB, Bigham AW, Buckingham KJ et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 2010: 42: 790–793.
8. Paulussen ADC, Stegmann APA, Blok MJ et al. MLL2 mutation spectrum in 45 patients with Kabuki syndrome. Hum Mutat 2011: 32: E2018–E2025.
9. Banka S, Veeramachaneni R, Reardon W et al. How genetically heterogeneous is Kabuki syndrome?: MLL2 testing in 116 patients, review and analyses of mutation and phenotypic spectrum. Eur J Hum Genet 2012: 20: 381–388.
10. Micale L, Augello B, Fusco C et al. Mutation spectrum of MLL2 in a cohort of Kabuki syndrome patients. Orphanet J Rare Dis 2011: 6: 38.
11. Miyake N, Koshimizu E, Okamoto N et al. MLL2 and KDM6A mutations in patients with Kabuki syndrome. Am J Med Genet A 2013: 161: 2234–2243.
12. Courcet J-B, Faivre L, Michot C et al. Clinical and molecular spectrum of renal malformations in Kabuki syndrome. J Pediatr 2013: 163: 742–746.
13. Bögershausen N, Wollnik B. Unmasking Kabuki syndrome. Clin Genet 2013: 83: 201–211.
14. Micale L, Augello B, Maffeo C et al. Molecular analysis, pathogenic mechanisms, and read through therapy on a large cohort of Kabuki syndrome patients. Hum Mutat 2014: 35 (7): 841–850.
15. Li Y, Bögershausen N, Alanay Y et al. A mutation screen in patients with Kabuki syndrome. Hum Genet 2011: 130: 715–724.
16. Banka S, Howard E, Bunstone S et al. MLL2 mosaic mutations and intragenic deletion-duplications in patients with Kabuki syndrome. Clin Genet 2013: 83: 467–471.
17. Banka S, Lederer D, Benoit V et al. Novel KDM6A (UTX) mutations and a clinical and molecular review of the X-linked Kabuki syndrome (KS2). Clin Genet 2015: 87 (3): 252–258. DOI: 10.1111/cge.12363.
18. Lederer D, Grisart B, Digilio MC et al. Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. Am J Hum Genet 2012: 90: 119–124.
19. Kuniba H, Yoshiura K, Kondoh T et al. Molecular karyotyping in 17 patients and mutation screening in 41 patients with Kabuki syndrome. J Hum Genet 2009: 54: 304–309.
20. Priolo M, Micale L, Augello B et al. Absence of deletion and duplication of MLL2 and KDM6A genes in a large cohort of patients with Kabuki syndrome. Mol Genet Metab 2012: 107: 627–629.
21. Bienvenu T, Poirier K, Friocourt G et al. ARX, a novel Prd-class-homeobox gene highly expressed in the telencephalon, is mutated in X-linked mental retardation. Hum Mol Genet 2002: 11: 981–991.
22. Shen E, Shulha H, Weng Z, Akbarian S. Regulation of histone H3K4 methylation in brain development and disease. Philos Trans R Soc Lond B Biol Sci 2014: 369 (1652).
23. Carrie A, Jun L, Bienvenu T et al. A new member of the IL-1 receptor family highly expressed in hippocampus and involved in X-linked mental retardation. Nat Genet 1999: 23: 25–31.
24. Behnecke A, Hinderhofer K, Bartsch O et al. Intragenic deletions of IL1RAPL1: report of two cases and review of the literature. Am J Med Genet A 2011: 155A (2): 372–379. DOI: 10.1002/ajmg.a.33656.
25. Hoffman EP, Arahata K, Minetti C, Bonilla E, Rowland LP. Dystrophinopathy in isolated cases of myopathy in females. Neurology 1992: 42 (5): 967–975.
26. Heide S, Afenjar A, Edery P et al. Xp21 deletion in female patients with intellectual disability: two new cases and a review of the literature. Eur J Med Genet 2015: 58 (6–7): 341–345. DOI: 10.1016/j.ejmg.2015.04.003.
27. Hu X, Burghes AHM, Ra PN, Thompson MW, Murphy EG, Worton RG. Partial gene duplication in Duchenne and Becker muscular dystrophies. J Med Genet 1988: 25: 369–376.

4 Závěry

Práce byla zaměřena na podrobnou analýzu pacientů se vzácnými, resp. unikátními nálezy na úrovni CNV nebo SNV. U většiny analyzovaných pacientů se jednalo o komplexní syndromické postižení asociované s vývojovým opožděním, MR a/nebo PAS. K analýze byly využity celogenomové i cílené postupy a u všech pacientů byla provedena co nejpodrobnější analýza fenotypu, v případech dobře definovaných syndromů s využitím standardizovaného klinického dotazníku.

V průběhu práce se podařilo úspěšně dokončit analýzu u pacientů s delecemi Xp22.1-p22.3, 6q11-q13, 6q14-q16, Xq25, 1q21.1, Xp21.2-p21.3, 2p14-p15, 17q21.31, 9q21.3 a 2p15-p16.1 a s duplikací Xp21.2-p21.3, a dále pacientů s nukleotidovými variantami v genech *HCFC1*, *KAT6B*, *SOS2* a *KMT2D*.

U pacientů s mikroskopicky viditelnými SV (delece Xp22.1-p22.3, 6q11-q13, 6q14-q16 a Xp21.2-p21.3) byl metodami s vyšším rozlišením (mBAND a aCGH) upřesněn rozsah těchto SV, což umožnilo lepší zhodnocení genového obsahu i mechanismu vzniku aberací (viz níže). Precizní korelace genotyp-fenotyp je u rozsáhlých aberací zahrnujících několik desítek genů obtížná až nemožná, přesto má smysl. Příkladem je delece 6q11-q13, která zahrnovala několik genů způsobujících dobře definovaná AD onemocnění, jejichž symptomy naše pacientka v době vyšetření nevykazovala. Tento fakt může být vysvětlen jednak tím, že za symptomy těchto AD onemocnění jsou zodpovědné spíše aktivační než ztrátové mutace, a jednak tím, že některá z těchto onemocnění mají až pozdější nástup symptomů.

U mikrodelecí 2p14-p15 a 9q21.3, jejichž vliv na fenotyp nebyl v době publikace dostatečně podložen, podpořila naše práce jejich kauzalitu a rovněž přispěla k poznatkům o fenotypu pacientů nesoucích tyto SV. Stejně tak práce podpořila asociaci SNV v genu *HCFC1* s MR/PAS a asociaci variant v genu *SOS2* s NS.

U častějších nebo již dříve dobře definovaných SV jako jsou aberace spojené s mikrodelečními syndromy 1q21.1, 17q21.31, 2p15-p16.1 a u SNV v etablovaných kauzálních genech *KAT6B a KMT2D* přispěla naše práce k rozšíření fenotypového spektra asociovaného s těmito genetickými defekty.

U pacientky se dvěma zásahy na obou X chromozomech (delecí Xp22.1-p22.3 a plnou mutací genu *FMR1*) jsme přispěli k poznatkům o možných dopadech nevyvážené X inaktivace na fenotyp dívek s X vázanými recesivními onemocněními. V případě monozygotních dvojčat s mikrodelecí 17q21.31 jsme díky detailní analýze genomových variant přispěli k poznatkům o možných genotypových odlišnostech

u monozygotních dvojčat a díky preciznímu porovnání fenotypu obou dvojčat také k poznatkům o variabilní expresivitě mikrodelečního syndromu 17q21.31.

V případě pacientů s delecí Xq25, pacientek s delecemi 17q21.31 a 2p15-p16.1, pacientky se SNV v genu *KAT6B* a pacienta s variantou v genu *SOS2* (kterou nesla i jeho symptomatická, ale velmi mírně postižená matka) jsme vzhledem k jejich sledování v delším časovém horizontu přispěli i k poznatkům o vývoji fenotypu a dlouhodobé prognóze pacientů nesoucích tyto genomové a genové varianty.

U většiny pacientů s CNV byl diskutován i mechanismus vzniku aberace, neboť metody aCGH a SNPa umožnily relativně přesnou lokalizaci zlomových míst. V případě pacientů s delecemi v oblasti Xp22.1-p22.3, 6q14-q16 a Xq25 se podařilo klonovat zlomová místa s pomocí PCR a určit tak i jejich přesnou nukleotidovou sekvenci. U pacientů s rekurentními delecemi 17q21.31 a 1q21.1 vznikla aberace na podkladě NAHR. Matka pacientky s delecí 6q11-q13 nesla marker chromozom v mozaice a podrobná analýza ukázala, že delece je zřejmě důsledkem rozštěpení centromery. U většiny ostatních případů předpokládáme, že aberace vznikly nejspíše v důsledku NHEJ.

U všech pacientů s CNV byl detailně analyzován genový obsah aberací a byla diskutována role vybraných genů při utváření patologického fenotypu. Z genů zasažených delecemi byl například diskutován možný dopad parciální delece genu *ODZ1* (resp. *TENM1*) u pacienta s delecí Xq25 a atypickým fenotypem, genu *BCL11A* jako možného cíle v genové terapii thalasemie a sprkovité anemie a zároveň jeho role v patogenesi neurovývojových onemocnění a delece genů *NTRK2* a *HNRNPK* jakožto možných kandidátů zodpovědných za fenotyp pacientů s mikrodelecí 9q21.3 (pozdější práce naznačují, že zodpovědným genem je právě *HNRNPK).*

V případě pacientů s genovými variantami v *HCFC1*, *SOS2* a *KMT2D* jsme přispěli k souhrnným studiím (ve dvou případech mezinárodním), které měly za cíl ozřejmit funkci těchto genů či upřesnit korelaci genotyp-fenotyp u již známých kauzálních genů. Případ pacientky s variantou v genu *KAT6B* a detailní analýza všech dalších dosud publikovaných případů demonstruje významný vliv lokalizace varianty v genu *KAT6B* na výsledný fenotyp.

V případě pacientů s KS práce demonstruje naše zkušenosti s diagnostickou posloupností u pacientů se suspekcí na tento syndrom a zdůrazňuje důležitost podrobného fenotypového popisu pacientů a užitečnost klinických dotazníků včetně výpočtu skóre jak při rozhodování o indikaci k vyšetření tak při následném hodnocení nalezených variant.

Práce rovněž demonstruje užitečnost ale i limity celogenomových metod v diagnostice vzácných onemocnění a zároveň i nutnost precizního zhodnocení fenotypu pacienta jak před vyšetřením, tak i po vyšetření ale také nutnosti důsledného informování rodin o způsobu vyšetření a možnosti náhodných či neinterpretovatelných nálezů. Důležitost genetického poradenství demonstruje případ pacienta A1 s delecí Xq25, který nevykazoval typické symptomy XLP-1 a testován byl pouze na základě pozitivní rodinné anamnézy. Průkaz delece genu *SH2D1A* vedl k nutnosti zařazení tohoto pacienta do dlouhodobé dispenzární péče hematoonkologů.

5 Literatura

Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, et al. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism 4, 36.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. (2010). A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249.

Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, et al. (2016). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 44, D7-D19.

Alkelai A, Olender T, Haffner-Krausz R, Tsoory MM, Boyko V, et al. (2016). A role for TENM1 mutations in Congenital General Anosmia. Clin. Genet. Apr 4. doi: 10.1111/cge.12782. [Epub ahead of print].

Amberger J, Bocchini CA, Scott AF, Hamosh A. (2009). McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res. 37, D793-D796.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Au PYB, You J, Caluseriu O, Schwartzentruber J, Majewski J, et al. (2015). GeneMatcher aids in the identification of a new malformation syndrome with intellectual disability, unique facial dysmorphisms, and skeletal and connective tissue abnormalities caused by de novo variants in HNRNPK. Hum. Mutat. 36, 1009–1014.

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. (2015). A global reference for human genetic variation. Nature 526, 68–74.

Aymé S, Schmidtke J. (2007). Networking for rare diseases: a necessity for Europe. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 50, 1477–1483.

Baron-Cohen S, Scott FJ, Allison C, Williams J, Bolton P, et al. (2009). Prevalence of autism-spectrum conditions: UK school-based population study. Br. J. Psychiatry 194, 500–509.

Birney E, Soranzo N. (2015). Human genomics: The end of the start for population sequencing. Nature 526, 52–53.

Botstein D, Risch N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33 Suppl, 228–237.

Campeau PM, Kim, JC, Lu JT, Schwartzentruber JA, Abdul-Rahman OA, et al. (2012). Mutations in KAT6B, encoding a histone acetyltransferase, cause Genitopatellar syndrome. Am. J. Hum. Genet. 90, 282–289.

Ciccone R, Giorda R, Gregato G, Guerrini R, Giglio S, et al. (2005). Reciprocal translocations: a trap for cytogenetists? Hum. Genet. 117, 571–582.

Clayton-Smith J, O'Sullivan J, Daly S, Bhaskar S, Day R, et al. (2011). Whole-exome-sequencing identifies mutations in histone acetyltransferase gene KAT6B in individuals with the Say-Barber-Biesecker variant of Ohdo syndrome. Am. J. Hum. Genet. 89, 675–681.

Collins FS, Morgan M, Patrinos A. (2003). The Human Genome Project: lessons from large-scale biology. Science 300, 286–290.

Das AM. (2016). Pharmacotherapy of inborn errors of metabolism illustrating challenges in orphan diseases. J. Pharmacol. Toxicol. Methods. Feb 26. doi: 10.1016/j.vascn.2016.02.182. [Epub ahead of print].

de Leeuw N, Dijkhuizen T, Hehir-Kwa JY, Carter NP, Feuk L, et al. (2012). Diagnostic interpretation of array data using public databases and internet sources. Hum. Mutat. 33, 930–940.

de Vries BBA, Pfundt R, Leisink M, Koolen DA, Vissers LELM, et al. (2005). Diagnostic genome profiling in mental retardation. Am. J. Hum. Genet. 77, 606–616.

Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, et al. (2007). Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat. Genet. 39, 25–27.

Exome Aggregation Consortium, Lek M, Karczewski K, Minikel E, Samocha K, et al. (2015). Analysis of protein-coding genetic variation in 60,706 humans. bioRxiv 30338. doi: http://dx.doi.org/10.1101/030338.

Feenstra I, Hanemaaijer N, Sikkema-Raddatz B, Yntema H, Dijkhuizen T, et al. (2011). Balanced into array: genome-wide array analysis in 54 patients with an apparently balanced de novo chromosome rearrangement and a meta-analysis. Eur. J. Hum. Genet. 19, 1152–1160.

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, et al. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am. J. Hum. Genet. 84, 524–533.

Fombonne E. (2002). Epidemiological trends in rates of autism. Mol. Psychiatry 7 Suppl 2, S4-S6.

Gijsbers ACJ, Lew JYK, Bosch CAJ, Schuurs-Hoeijmakers JHM, van Haeringen A, et al. (2009). A new diagnostic workflow for patients with mental retardation and/or multiple congenital abnormalities: test arrays first. Eur. J. Hum. Genet. 17, 1394–1402.

Gonzalez-Mantilla, AJ, Moreno-De-Luca A, Ledbetter DH, Martin CL. (2016). A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. JAMA Psychiatry 73, 275–283.

Hancarova M, Simandlova M, Drabova J, Mannik K, Kurg A, et al. (2013). A patient with de novo 0.45 Mb deletion of 2p16.1: the role of BCL11A, PAPOLG, REL, and FLJ16341 in the 2p15-p16.1 microdeletion syndrome. Am. J. Med. Genet. A. 161A, 865–870.

Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, et al. (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat. Genet. 42, 483–485.

Holt R, Barnby G, Maestrini E, Bacchelli E, Brocklebank D, et al. (2010). Linkage and candidate gene studies of autism spectrum disorders in European populations. Eur. J. Hum. Genet. 18, 1013–1019.

Church DM, Lappalainen I, Sneddon TP, Hinton J, Maguire M, et al. (2010). Public data archives for genomic structural variation. Nat. Genet. 42, 813–814.

Jacobs PA, Browne C, Gregson N, Joyce C, White H. (1992). Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. J. Med. Genet. 29, 103–108.

Kaufman L, Ayub M, Vincent JB. (2010). The genetic basis of non-syndromic intellectual disability: a review. J. Neurodev. Disord. 2, 182–209.

Kim MJ, Cho SI, Chae, JH, Lim BC, Lee JS, et al. (2016). Pitfalls of Multiple Ligation-Dependent Probe Amplifications in Detecting DMD Exon Deletions or Duplications. J. Mol. Diagn. 18, 253–259.

Kishino T, Lalande M, Wagstaff J. (1997). UBE3A/E6-AP mutations cause Angelman syndrome. Nat. Genet. 15, 70–73.

Koolen DA, Kramer JM, Neveling K, Nillesen WM, Moore-Barton HL, et al. (2012). Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. Nat. Genet. 44, 639–641.

Koolen DA, Pfundt R, Linda K, Beunders G, Veenstra-Knol HE, et al. (2016). The Koolen-de Vries syndrome: a phenotypic comparison of patients with a 17q21.31 microdeletion versus a KANSL1 sequence variant. Eur. J. Hum. Genet. 24, 652–659.

Kuhlenbäumer G, Hullmann J, Appenzeller S. (2011). Novel genomic techniques open new avenues in the analysis of monogenic disorders. Hum. Mutat. 32, 144–151.

Kulkarni S, Nagarajan P, Wall J, Donovan DJ, Donell RL, et al. (2008). Disruption of chromodomain helicase DNA binding protein 2 (CHD2) causes scoliosis. Am. J. Med. Genet. A. 146A, 1117–1127.

Kumar D. (2008). Disorders of the genome architecture: a review. Genomic Med. 2, 69–76.

Kumar P, Henikoff S, Ng PC. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081.

La Malfa G, Lassi S, Bertelli M, Salvini R, Placidi GF. (2004). Autism and intellectual disability: a study of prevalence on a sample of the Italian population. J. Intellect. Disabil. Res. 48, 262–267.

Lalonde E, Albrecht S, Ha KCH, Jacob K, Bolduc N, et al. (2010). Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing. Hum. Mutat. 31, 918–923.

Landrum MJ, Lee, JM, Riley GR, Jang W, Rubinstein WS, et al. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 42, D980-D985.

Lange L, Pagnamenta AT, Lise S, Clasper S, Stewart H, et al. (2016). A de novo frameshift in HNRNPK causing a Kabuki-like syndrome with nodular heterotopia. Clin. Genet. Mar 8. doi: 10.1111/cge.12773. [Epub ahead of print].

Lee JA, Carvalho CMB, Lupski JR. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131, 1235–1247.

Leonard H, Wen X. (2002). The epidemiology of mental retardation: challenges and opportunities in the new millennium. Ment. Retard. Dev. Disabil. Res. Rev. 8, 117–134.

Lupski JR, Stankiewicz P. (2005). Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet. 1, e49.

MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. 42, D986-D992.

Makrythanasis P, van Bon BW, Steehouwer M, Rodríguez-Santiago B, Simpson M, et al. (2013). MLL2 mutation detection in 86 patients with Kabuki syndrome: a genotype-phenotype study. Clin. Genet. 84, 539–545.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

McMichael G, Bainbridge MN, Haan E, Corbett M, Gardner A, et al. (2015). Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. Mol. Psychiatry 20, 176–182.

Merson TD, Dixon MP, Collin C, Rietze RL, Bartlett PF, et al. (2006). The transcriptional coactivator Querkopf controls adult neurogenesis. J. Neurosci. 26, 11359–11370.

Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, et al. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am. J. Hum. Genet. 86, 749–764.

Moss J, Howlin P. (2009). Autism spectrum disorders in genetic syndromes: implications for diagnosis, intervention and understanding the wider autism spectrum disorder population. J. Intellect. Disabil. Res. 53, 852–873.

Nevado J, Mergener R, Palomares-Bralo M, Souza KR, Vallespín E, et al. (2014). New microdeletion and microduplication syndromes: A comprehensive review. Genet. Mol. Biol. 37, 210–219.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461, 272–276.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. Nat. Genet. 42, 30–35.

Nielsen J, Rasmussen K. (1976). Autosomal reciprocal translocations and 13/14 translocations: a population study. Clin. Genet. 10, 161–177.

Pelz J, Arendt V, Kunze J. (1996). Computer assisted diagnosis of malformation syndromes: an evaluation of three databases (LDDB, POSSUM, and SYNDROC). Am. J. Med. Genet. 63, 257-267.

Pinkel D, Straume T, Gray JW. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc. Natl. Acad. Sci. U. S. A. 83, 2934–2938.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, et al. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet. 20, 207–211.

Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, et al. (2006). Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. Am. J. Med. Genet. A. 140, 2063–2074.

Rauen KA. (2013). The RASopathies. Annu. Rev. Genomics Hum. Genet. 14, 355–369.

Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, et al. (2015). Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. Value Health.18, 906–914.

Ropers HH. (2008). Genetics of intellectual disability. Curr. Opin. Genet. Dev. 18, 241–250.

Ropers HH. (2010). Genetics of early onset cognitive impairment. Annu. Rev. Genomics Hum. Genet. 11, 161–187.

Sanger F, Nicklen S, Coulson AR. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467.

Schluth-Bolard C, Delobel B, Sanlavill, D, Boute O, Cuisset JM, et al. (2009). Cryptic genomic imbalances in de novo and inherited apparently balanced chromosomal rearrangements: array CGH study of 47 unrelated cases. Eur. J. Med. Genet. 52, 291–296.

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, et al. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 30, e57.

Sellner LN, Taylor GR. (2004). MLPA and MAPH: new techniques for detection of gene deletions. Hum. Mutat. 23, 413–419.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, et al. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308–311.

Sismani C, Kitsiou-Tzeli S, Ioannides M, Christodoulou C, Anastasiadou V, et al. (2008). Cryptic genomic imbalances in patients with de novo or familial apparently balanced translocations and abnormal phenotype. Mol. Cytogenet. 1, 15.

Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH. (2003). Mutations in RAI1 associated with Smith-Magenis syndrome. Nat. Genet. 33, 466–468.

Slavotinek AM. (2008). Novel microdeletion syndromes detected by chromosome microarrays. Hum. Genet. 124, 1–17.

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, et al. (1997). Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes. Chromosomes Cancer 20, 399–407.

South ST, Brothman AR. (2011). Clinical laboratory implementation of cytogenomic microarrays. Cytogenet. Genome Res. 135, 203–211.

Speicher MR, Carter NP. (2005). The new cytogenetics: blurring the boundaries with molecular biology. Nat. Rev. Genet. 6, 782–792.

Speir ML, Zweig AS, Rosenbloom KR, Raney BJ, Paten B, et al. (2016). The UCSC Genome Browser database: 2016 update. Nucleic Acids Res. 44, D717-D725.

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, et al. (2015). An integrated map of structural variation in 2,504 human genomes. Nature 526, 75–81.

Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, et al. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat. Genet. 39, 319–328.

Tabet AC, Verloes A, Pilorge M, Delaby E, Delorme R, et al. (2015). Complex nature of apparently balanced chromosomal rearrangements in patients with autism spectrum disorder. Mol. Autism 6, 19.

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, et al. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337, 64–69.

Tharapel AT, Tharapel SA, Bannerman RM. (1985). Recurrent pregnancy losses and parental chromosome abnormalities: a review. Br. J. Obstet. Gynaecol. 92, 899–914.

Vandeweyer G, Kooy RF. (2013). Detection and interpretation of genomic structural variation in health and disease. Expert Rev. Mol. Diagn. 13, 61–82.

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. (2014). Ten years of next-generation sequencing technology. Trends Genet. 30, 418–426.

Vissers LELM, Stankiewicz P. (2012). Microdeletion and microduplication syndromes. Methods Mol. Biol. 838, 29–75.

Vissers LELM, Gilissen C, Veltman JA. (2016). Genetic studies in intellectual disability and related disorders. Nat. Rev. Genet. 17, 9–18.

Vulto-van Silfhout AT, van Ravenswaaij CMA, Hehir-Kwa JY, Verwiel ETP, Dirks R, et al. (2013). An update on ECARUCA, the European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations. Eur. J. Med. Genet. 56, 471–474.

Wang DG, Fan JB, Siao CJ, Berno A, Young P, et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280, 1077–1082.

Warner G, Moss J, Smith P, Howlin P. (2014). Autism characteristics and behavioural disturbances in ~ 500 children with Down's syndrome in England and Wales. Autism Res. 7, 433–441.

Weise A, Mrasek K, Ewers E, Mkrtchyan H, Kosyakova N, et al. (2009). Diagnostic applications of fluorescence in situ hybridization. Expert Opin. Med. Diagn. 3, 453–460.

Weise A, Mrasek K, Klein E, Mulatinho M, Llerena JC, et al. (2012). Microdeletion and microduplication syndromes. J. Histochem. Cytochem. 60, 346–358.

Winter RM, Baraitser M, Douglas JM. (1984). A computerised data base for the diagnosis of rare dysmorphic syndromes. J. Med. Genet. 21, 121–123.

Yates A, Akanni W, Amode MR, Barrell D, Billis K, et al. (2016). Ensembl 2016. Nucleic Acids Res. 44, D710-D716.

Yilmaz R, Beleza-Meireles A, Price S, Oliveira R, Kubisch C, et al. (2015). A recurrent synonymous KAT6B mutation causes Say-Barber-Biesecker/Young-Simpson syndrome by inducing aberrant splicing. Am. J. Med. Genet. A. 167A, 3006–3010.

Zahir F, Friedman JM. (2007). The impact of array genomic hybridization on mental retardation research: a review of current technologies and their clinical utility. Clin. Genet. 72, 271–287.

Zarrei M, MacDonald JR, Merico D, Scherer SW. (2015). A copy number variation map of the human genome. Nat. Rev. Genet. 16, 172–183.

Zollino M, Marangi G, Ponzi E, Orteschi D, Ricciardi, S, et al. (2015). Intragenic KANSL1 mutations and chromosome 17q21.31 deletions: broadening the clinical spectrum and genotype-phenotype correlations in a large cohort of patients. J. Med. Genet. 52, 804–814.