

**UNIVERZITA KARLOVA V PRAZE**

**1. LÉKAŘSKÁ FAKULTA  
ÚSTAV DĚDIČNÝCH METABOLICKÝCH PORUCH**

Studijní obor: **Molekulární a buněčná biologie, genetika a virologie**



Autor: **Mgr. Viktor Stránecký**

**Současné metody analýzy genomu a jejich využití v hledání  
genetických příčin nemocí**

**Current methods of genome analysis and their use in identification  
of genetic determinants of human diseases**

Dizertační práce

Vedoucí práce: **doc. Ing. Stanislav Kmoch CSc.**

Místo a rok vypracování: **Praha, 2015**

## Prohlášení

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

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

V Praze, 29.04.2015

Viktor Stránecký

Podpis

Identifikační záznam:

STRÁNECKÝ, Viktor. *Současné metody analýzy genomu a jejich využití v hledání genetických příčin nemocí [Current methods of genome analysis and their use in identification of genetics determinants of human diseases]* Praha, 2015. 115 s. Dizertační práce. Univerzita Karlova v Praze, 1. lékařská fakulta, Ústav dědičných metabolických poruch. Vedoucí práce Kmoch, Stanislav.

## Abstrakt

Studium vzácných onemocnění je vhodným přístupem pro nalezení genetické a molekulární podstaty lidských znaků a výrazně napomohlo k identifikaci genů, objasnění jejich funkce a přispělo k charakterizaci funkce metabolických drah a buněčných procesů. V průběhu posledních 30-ti let byla vazebná analýza nejúspěšnějším přístupem k hledání genů podmiňujících Mendelovská onemocnění a přispěla k identifikaci řady genů, přesto podstata mnohých onemocnění zůstává stále neznámá. Nové metody studia lidského genomu, zejména technologie DNA čipů, masivně paralelní sekvenování (next generation sequencing) a metody analýzy takto získaných dat, představují způsob jak efektivně identifikovat příčinu geneticky podmíněných onemocnění na základě přímého pozorování mutací v celém genomu postižených jedinců. Tyto metody nahrazují tradiční způsob identifikace genů reprezentovaný vazebnou analýzou a sekvenováním kandidátních genů a stávající se standardním přístupem pro objasnění molekulární podstaty onemocnění. V této práci popisují možnosti studia vzácných genetických podmíněných onemocnění a výsledky dosažené s využitím těchto postupů - identifikaci genů podmiňujících mukopolysacharidózu typ IIIC (*TMEM76*), izolovaný defekt ATP syntázy (*TMEM70*), Rotorův syndrom (*SLCO1B3* a *SLCO1B1*), autozomálně dominantní ANCL (*DNAJC5*) a GAPO syndrom (*ANTXR1*).

## Klíčová slova

vzácná onemocnění, genetické mapování, technologie DNA čipů, exomové sekvenování, neuronální ceroidní lipofuscinóza, Rotorův syndrom, izolovaný defekt ATP syntázy, mukopolysacharidóza typu IIIC, GAPO syndrom

## **Abstract**

The study of rare genetic diseases presents unique opportunity to uncover the genetic and molecular basis of human traits and greatly helped to the identification of genes, to the elucidation of their function and to the characterization of metabolic pathways and cellular processes. Over the past decades, linkage analysis has been appropriate approach to search for the genes causing Mendelian diseases and contributed to the identification of many genes, but the genetic cause of many diseases remains unknown. New methods of studying the human genome, microarray technology and massively parallel sequencing (next generation sequencing), represent a way to efficiently identify the cause of genetically determined diseases, based on direct observation of mutations in the genome of affected individuals. These techniques replaced the traditional method of disease gene identification represented by linkage analysis and sequencing of candidate genes and have become the standard approach to elucidate the molecular basis of diseases. In this work, i describe the the results achieved by using these methods - identification of the genes underlying mucopolysaccharidosis type IIIC, isolated defect of ATP synthase, Rotor syndrome, autosomal dominant ANCL and GAPO syndrome.

## **Key words**

rare diseases, genetic mapping, microarray technology, exome sequencing, neuronal ceroid lipofuscinosis, Rotor syndrome, isolated defect of ATP synthase , GAPO syndrome

## Poděkování

Rád bych poděkoval především Standovi Kmochovi a Martině Živné za to, že mě umožnili poznat úžasný svět molekulární genetiky a všem svým spolupracovníkům z Ústavu dědičných metabolických poruch za pomoc, cenné rady, skvělou spolupráci a vždy přítomnou dobrou náladu. Poděkování patří zejména Kateřině Hodaňové, Haně Hartmannové, Lence Piherové, Veronice Barešové, Lence Noskové, Petrovi Vyleťalovi, Anně Přistoupilové, Aleně Čížkové-Vrbacké, Evě Oliveriusové a dalším.

Nemenší poděkování patří také všem mým blízkým, zejména za trpělivost při mém studiu.

Finanční podporu pro projekty zmíněné v této práci poskytly následující grantové agentury a granty: grantová agentura České republiky: 303/03/H065, 303/07/0781, 305/08/H037, grantová agentura Ministerstva zdravotnictví ČR: NR8069-3, NR8069-1, 1A/8239-3, NT13116-4/2012; grantová agentura Univerzity Karlovy: 54/20320827/05, 250051; programy Univerzity Karlovy: PRVOUK-P24/LF1/3, UNCE 204011 a SVV 260148/2015; Ministerstva školství, mládeže a tělovýchovy MSM0021620806, AV0Z50110509 a 1M6837805002; Evropský fond pro regionální rozvoj: CZ.1.05/1.1.00/02.0109.

## Obsah

Abstrakt.....	3
Obsah .....	6
Seznam zkratk .....	8
Část I. Úvod .....	9
Cíl dizertační práce.....	12
Část II. Technologie a metody .....	13
Technologie DNA čipů .....	13
Analýza dat - DNA čipy .....	13
Sekvenační technologie .....	16
Masivně paralelní sekvenování.....	16
Roche 454 GS FLX.....	17
Applied Biosystems SOLiD.....	17
Illumina .....	18
Metody obohacení DNA.....	18
Analýza sekvenačních dat .....	20
Metody hledání genů podmiňujících dědičná onemocnění .....	22
Vazebná analýza.....	23
Homozygotní mapování.....	24
Asociační analýza .....	24
Analýza počtu změn kopií DNA.....	25
Exomové sekvenování.....	25
Část III. Studium genetické podstaty vzácných onemocnění .....	27
Rotorův syndrom .....	27
Deficit ATP syntázy.....	29
Mukopolysacharidóza typu IIIC.....	30
Autozomálně dominantní ANCL.....	30
GAPO syndrom.....	32
Část IV. Výsledky .....	33
Závěr .....	34
Část V. Přílohy .....	35
Literatura .....	36
Kopie publikovaných prací.....	44

Příloha 1a Rotor-type hyperbilirubinaemia has no defect in the canalicular bilirubin export pump .....	44
Příloha 1b Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver.....	44
Příloha 2a Development of a human mitochondrial oligonucleotide microarray (h-MitoArray) and gene expression analysis of fibroblast cell lines from 13 patients with isolated F1FO ATP synthase deficiency .....	44
Příloha 2b TMEM70 mutations cause isolated ATP synthase deficiency and neonatal mitochondrial encephalomyopathy.....	44
Příloha 3 Mutations in TMEM76* cause mucopolysaccharidosis IIIC (Sanfilippo C syndrome) .....	44
Příloha 4 Mutations in DNAJC5, encoding cysteine-string protein alpha, cause autosomal-dominant adult-onset neuronal ceroid lipofuscinosis .....	44
Příloha 5 Mutations in ANTXR1 cause GAPO syndrome .....	44

## Seznam zkratek

ANCL	adultní forma neuronální ceroidní lipofuscinózy
ATP	adenosintrifosfát
BAC	bacterial artificial chromosome, umělý bakteriální chromosom
BAM	binary sequence alignment map
bp	base pairs, páry bazí
BWA	Burrows-Wheeler Aligner
CADD	Combined Annotation Dependent Depletion
cDNA	komplementární deoxyribonukleová kyselina
CGH	komparativní genomová hybridizace
cM	centimorgan
CNV	copy number variation, varianty v počtu kopií DNA
CSP $\alpha$	cysteine-string protein alpha
DJS	Dubin Johnsonův syndrom
DNA	deoxyribonukleová kyselina
EST	expressed sequence tags
EVS	Exome Variant Server
ExAC	Exome Aggregation Consortium
GATK	Genome Analysis Toolkit
gDNA	genomová deoxyribonukleová kyselina
GERP	Genomic Evolutionary Rate Profiling
HGMD	Human Gene Mutation Database
Indel	inzerce/delece
IKEM	Institut klinické a experimentální medicíny
LOD	logaritmus poměru pravděpodobnosti rekombinace
LOWESS	locally weighted scatterplot smoothing
MPSIIIC	mukopolysacharidóza typu IIIC
mtDNA	mitochondriální deoxyribonukleová kyselina
NCL	neuronální ceroidní lipofuscinózy
NGS	next-generation sequencing, sekvenování nové generace
OMIM	Online Mendelian Inheritance in Man
OTC	ornithine transcarbamylase
PCR	polymerase chain reaction, polymerázová řetězová reakce
PolyPhen	Polymorphism Phenotyping
RFLP	restriction fragment length polymorphism, polymorfismy v délce restrikčních fragmentů
RMA	Robust Multi-array Average
RNA	ribonukleotidová kyselina
RS	Rotorův syndrom
SNP	single nukleotide polymorphism, jednonukleotidový polymorfismus
STR	short tandem repeat, krátká tandemové repetice
TDT	transmission disequilibrium test, test nerovnováhy přenosu
ÚDMP	Ústav dědičných metabolických poruch
VCF	variant call format
VFN	Všeobecná fakultní nemocnice



## Část I. Úvod

Skutečnost, že některé znaky rostlin a živočichů se přenášejí z předků na potomky, je lidstvu známa již od starověku. Záznamy o chovu koní za účelem vylepšení některých jejich schopností pocházejí již z období Sumerské říše, tedy z 3.-2. tisíciletí př. n. l.. Ve starověkém Egyptě lidé ručně opylovali datle, aby zvýšili výnosy a podpořili jejich požadované charakteristiky.

První přenos onemocnění na potomky byl popsán již v talmudickém textu z 5. století n.l., v němž autoři v souvislosti s krvácivými projevy při rituální obřízce chlapců, popisují symptomy a charakteristiky přenosu onemocnění, *Rafei D'ma*, dnes známého jako hemofilie (Rosner F., 1995). Prvotní poznatky a mylné představy o zákonitostech dědičnosti byly v průběhu historie vyvracovány a nahrazovány poznatky novými. Až ve druhé polovině 19. století byly učiněny objevy, které významným způsobem přispěly k rozvoji znalostí o zákonitostech dědičnosti, tak jak je známe dnes. Zásadními se v tomto ohledu staly experimenty J.G. Mendela. V roce 1866 publikoval opat brněnského augustiniánského kláštera, J. G. Mendel, výsledky svých pokusů s hrachorem pod názvem: *“Pokusy s rostlinnými hybridy”* a na základě hodnocení sledovaných znaků rodičovských rostlin a hybridů v dalších generacích formuloval základní pravidla dědičnosti.

Již téměř zapomenutá práce J. G. Mendela byla znovuobjevena v roce 1900 prakticky současně třemi vědci; Hugo De Vriesem, Carlem Corrensem a Ericem von Tschermakem. Až toto *“znovuobjevení”* je považováno za událost, jež vedla k založení nového vědního oboru, genetiky. Autorem tohoto pojmu je William Bateson. Bateson přispěl k rozvoji genetiky nejen svým experimentálním výzkumem, ale také hledáním souvislostí mezi výsledky základního výzkumu dědičnosti na zvířatech a rostlinách a množstvím klinických záznamů o dědičných onemocněních lidí, jež byla na začátku 20. století k dispozici. Při svých experimentech s hrachorem objevil W. Bateson a R. Punnett výjimky, které neodpovídaly Mendelovým závěrům. Bateson tyto výjimky přisoudil genové vazbě, kterou exaktně vysvětlil až T. H. Morgan. Morgan, identifikoval u octomilky skupiny genů, které v rámci skupiny buď nerekombinují, nebo rekombinují omezeně. Odhalil tak funkci chromozomů při přenosu genetické informace a potvrdil, že geny lokalizované na stejném chromozómu jsou ve vazbě (Morgan T. H., 1910; Morgan T. H., 1911; Morgan T. H., 1911).

Na začátku 20. století bylo popsáno i první dědičné metabolické onemocnění, alkaptonurie. Archibald Garrod, identifikoval řadu případů tohoto onemocnění a zároveň si

povšimnul, že onemocnění se objevuje s typickým opakujícím se vzorcem dědičnosti. Znalosti z oblasti biochemie umožnily Garrodovi popsat první metabolické onemocnění člověka, u kterého byla prokázána platnost mendelovských zákonů dědičnosti (Garrod A. E., 1902). V první polovině 20. století bylo učiněno několik dalších významných objevů. H. J. Muller objevil schopnost rentgenových paprsků indukovat bodové mutace u octomilky a v roce 1944 O. Avery prokázal, že DNA je genetickým materiálem buněk.

Důležitým milníkem bylo určení přesné struktury molekuly DNA J. Watsonem a F. Crickem v roce 1953 (Watson J. D. and Crick F. H., 1953). I přes tento objev však zatím nebylo zřejmé, jakým způsobem řídí DNA syntézu proteinů a jakou roli v celém procesu hraje RNA. Definice centrálního dogmatu molekulární biologie popisujícího směr přenosu genetické informace z DNA do RNA a pak do proteinu (Crick F. H., 1956), objev mRNA (Jacob F. and Monod J., 1961) a rozluštění genetického kódu (Nirenberg M. W. and Matthaei J. H., 1961) plně objasnilo mechanismus funkčního vyjádření genetické informace.

Řada dalších objevů, zejména pak objev restričních endonukleáz (Smith H. O. and Wilcox K. W., 1970), rekombinantní DNA (Jackson D. A. et al., 1972), klonování DNA (Cohen S. N. et al., 1973), sekvenování DNA (Sanger F. and Coulson A. R., 1975), polymerázové řetězové reakce (Mullis K. B. and Faloona F. A., 1987) a polymorfních genetických markerů (Kan Y. W. and Dozy A. M., 1978) vedla k rozvoji metod molekulární genetiky umožňujících manipulaci, studium funkce a vlastností DNA. Tyto metody a postupy pozičního klonování umožnily studium genetické podstaty onemocnění a patofyziologických procesů s nimi spojených. Využití postupů pozičního klonování vedlo k identifikaci mnoha genů podmiňujících vzácná (Gusella J. F. et al., 1983; Koenig M. et al., 1987; Tsui L. C. et al., 1985) i populačně častá onemocnění (Hall J. M. et al., 1990; Miki Y. et al., 1994).

Zásadním milníkem v oblasti lidské genetiky byl Projekt lidského genomu (Human Genome Project) zahájený v roce 1990 s cílem určit přesnou sekvenci párů bází tvořících lidskou DNA a identifikovat všechny geny lidského genomu. První hrubá verze byla publikována v únoru 2001 (Consortium I. H. G. S., 2001) a k 50. výročí objevu struktury DNA byl projekt dokončen. Poznatky získané v rámci Projektu lidského genomu umožnily rozvoj mnoha nových technologií a metod analýzy geneticky podmíněných onemocnění. Znalost sekvence lidského genomu a popsaná variabilita (SNP) umožnila návrh genotypovacích DNA čipů a jejich následné využití při konstrukci haplotypové mapy lidského genomu

(International HapMap C., 2005) a studium genetické podstaty řady komplexních onemocnění pomocí celogenomových asociačních studií (Welter D. et al., 2014). Potřeba levného a vysoko-kapacitního sekvenování vedla k vývoji nových masivně paralelních sekvenačních technologií (Metzker M. L., 2010) a způsobila revoluci v možnostech studia populační variability lidského genomu (Genomes Project C. et al., 2012) a s ní spojených onemocnění. Technologie DNA čipů, NGS sekvenování a metody analýzy takto získaných dat představují nový způsob, kterým mohou být studována geneticky podmíněná onemocnění. Nahrazují tradiční způsob identifikace genů reprezentovaný vazebnou analýzou a sekvenováním kandidátních genů a stávají se standardním přístupem pro objasnění molekulární podstaty onemocnění (Bamshad M. J. et al., 2011).

Zavedení technologií DNA čipů, NGS sekvenování, postupů analýzy takto získaných dat a jejich využití při studiu molekulární podstaty vzácných onemocnění bylo hlavním cílem mé dizertační práce.

## **Cíl dizertační práce**

Hlavním cílem mé dizertační práce bylo zavedení technologie DNA čipů, NGS sekvenování, metod analýzy a interpretace takto získaných dat umožňujících jejich využití při studiu molekulární podstaty vzácných onemocnění. Dílčími cíli bylo:

- 1) Rozvoj technologie vlastních oligonukleotidových čipů pro studium genové exprese a změn genové dávky ve studiu genetických příčin vzácných nemocí
- 2) Využití SNP čipů pro celogenomové genotypování, genetické mapování a analýzu změn genové dávky
- 3) Využití komerčně dostupných DNA čipů pro studium genové exprese
- 4) Zavedení a využití metod masivně paralelního sekvenování ve studiu genetických příčin vzácných nemocí a DNA diagnostice

## Část II. Technologie a metody

### Technologie DNA čipů

DNA čipy jsou relativně novou technologií pro kvalitativní a kvantitativní analýzu nukleových kyselin, která je používána od poloviny 90. let minulého století. Jejich hlavní výhodou je možnost vyšetření velkého množství úseků nukleových kyselin v jednom experimentu.

V principu jsou DNA čipy malé pevné nosiče, na kterých mohou být imobilizovány v přesných pozicích až miliony oligonukleotidů (dříve i BAC nebo EST). Oligonukleotidy (probes) jsou navrženy tak, aby specificky hybridizovaly s komplementárními sekvencemi analyzované nukleové kyseliny (target). Po hybridizaci je detekován a kvantifikován fluorescenční signál pro každou z detekčních prób. Intenzita signálu odpovídá množství navázané analyzované nukleové kyseliny. DNA čipy byly nejprve využívány pro analýzu genové exprese (Schena M. et al., 1995), genotypování mitochondriálního genomu (Chee M. et al., 1996) a komparativní genomovou hybridizaci (CGH) (Solinas-Toldo S. et al., 1997). Možnosti jejich použití byly limitovány neznalostí referenčních sekvencí umožňující návrh detekčních sond. Teprve dokončení projektu lidského genomu umožnilo využití DNA čipů i pro jiné aplikace – vazebnou analýzu (Matsuzaki H. et al., 2004), celogenomové asociační studie (Klein R. J. et al., 2005), obohacení DNA (Hodges E. et al., 2007) a řadu dalších.

V současnosti se pro přípravu používají různé technologie, které se liší použitým nosičem a způsobem přípravy detekčních sond (viz. Obrázek 1) (Hughes T. R. et al., 2001; Michael K. L. et al., 1998; Pease A. C. et al., 1994).

### Analýza dat - DNA čipy

Proces analýzy a zpracování dat z DNA čipů se skládá z řady kroků, které obvykle zahrnují analýzu obrazu, odstranění nespecifického hybridizačního signálu, logaritmickou transformaci, normalizaci a statistické vyhodnocení. Používané metody jsou rozdílné pro jednotlivé platformy a aplikace. Pro zpracování dat existuje mnoho komerčních i volně dostupných programů (Koschmieder A. et al., 2012). Nejpoužívanější platformou pro analýzu všech typů DNA čipů jsou však tzv. balíčky z projektu Bioconductor (Gentleman R. C.

et al., 2004) využívající statistický programovací jazyk R. V následující části jsou popsány jednotlivé kroky zpracování dat.

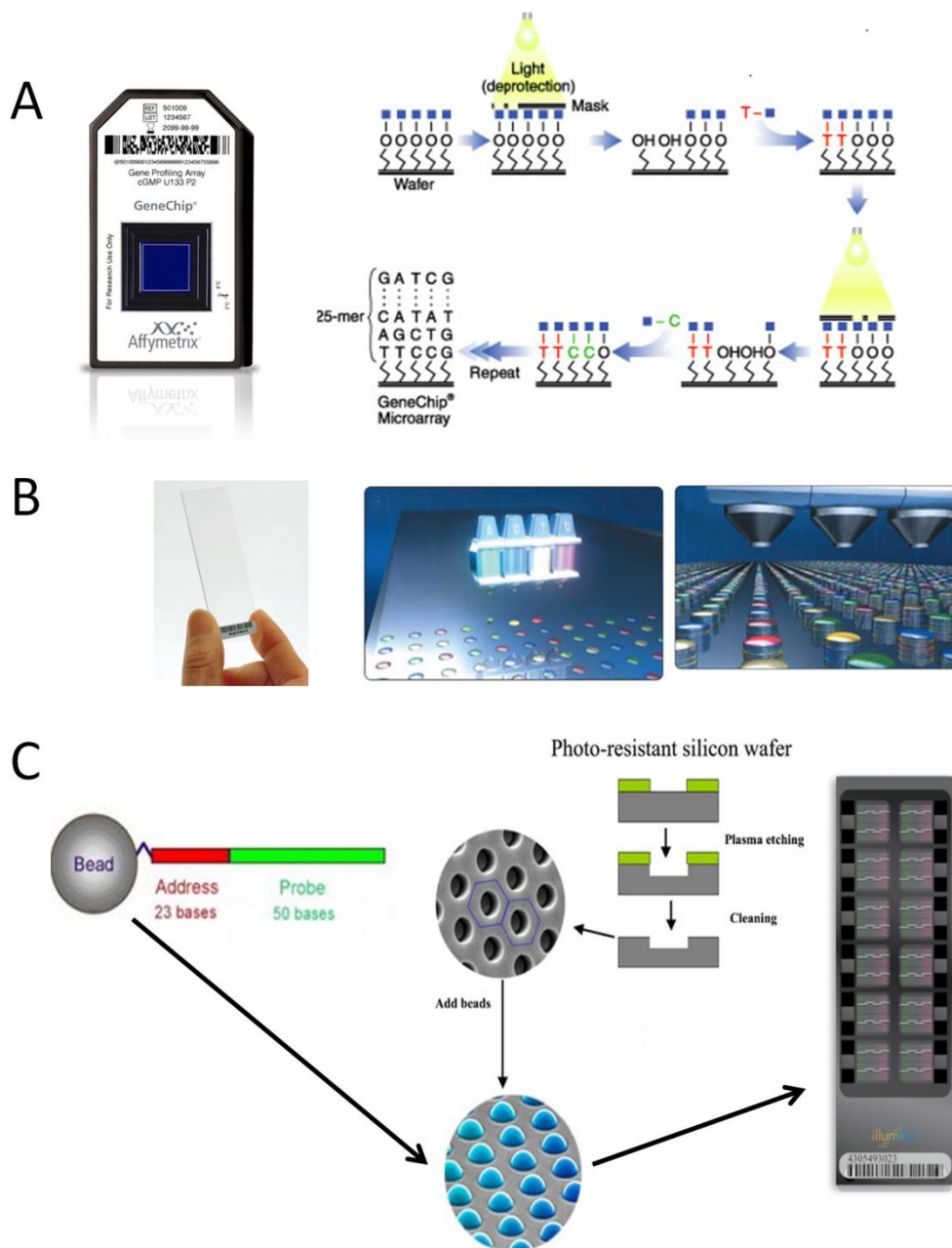
Analýza obrazu - primárním výstupem ze scanneru jsou obrazová data ve formátu TIFF, které je nutné převést na numerické hodnoty. Nejprve jsou proto identifikovány v obraze oblasti odpovídající jednotlivým detekčním sondám a nalezené intenzity signálu jsou převedeny na numerické hodnoty.

Odstranění nespecifického hybridizačního signálu - získané intenzity jsou součtem specifického a nespecifického signálu. Proto většina DNA čipů obsahuje tzv. negativní sondy, které umožňují kvantifikaci nespecifického hybridizačního signálu. Předpokládá se, že nespecifický hybridizační signál je aditivní a proto je od získaných hodnot odečten.

Logaritmická transformace - data jsou transformována na  $\log_2$  hodnoty. Tato transformace stabilizuje rozptyl a zároveň jsou data převedena do normálního rozdělení, které je předpokladem pro většinu běžných statistických testů.

Normalizace - před statistickým vyhodnocením dat je nutné minimalizovat technickou variabilitu, která může být způsobena například zpracováním vzorku, různým množstvím hybridizovaného vzorku nebo rozdílným nastavením scanneru. K tomuto účelu se používají metody normalizace dat. Nejčastěji používané metody jsou kvantilová, RMA a LOWESS normalizace (Bolstad B. M. et al., 2003; Smyth G. K. and Speed T., 2003).

Statistické vyhodnocení - pro analýzu genové exprese jsou používány běžné statistické testy (t-test, ANOVA) nebo jejich modifikace (Wettenhall J. M. and Smyth G. K., 2004). Vhodný test je vybrán v závislosti na počtu porovnávaných stavů a vztahu vzorků (párový nebo nepárový test). Získané hodnoty pravděpodobností jsou pro omezení falešně pozitivních výsledků upraveny s využitím metod korekce pro mnohonásobné testování (Hsueh H. M. et al., 2003). Interpretaci takto nalezeného seznamu rozdílně exprimovaných genů mohou usnadnit metody analýzy genového obohacení, například s využitím databází genové ontologie nebo databáze metabolických drah (Huang da W. et al., 2009).



**Obrázek 1.** Technologie přípravy DNA čipů

**A.** DNA čip Affymetrix - postupná fotolitografická syntéza 25-bázových oligonukleotidů se světelnou deprotekcí za použití systému specifických masek na křemíkovém nosiči (modifikováno podle [www.affymetrix.com](http://www.affymetrix.com))

**B.** DNA čip Agilent - na mikroskopickém sklíčku jsou postupně syntetizovány 60-bázové oligonukleotidy pomocí bezkontaktních tiskových hlav (modifikováno podle <http://www.genomics.agilent.com>)

**C.** DNA čip Illumina - syntetizované 73-bázové oligonukleotidy s 23-bázovou značkou, umožňující identifikaci, jsou navázány na silikátové kuličky a náhodně naneseny do jamek na křemíkové destičce (modifikováno podle [www.illumina.com](http://www.illumina.com))

## **Sekvenační technologie**

Možnost využití dideoxynukleotidů pro terminaci elongace sekvenovaného řetězce (Sanger F. et al., 1977) představovala zásadní mezník v historii DNA sekvenování. Tato myšlenka umožnila dohromady s objevem polymerázové řetězové reakce (Mullis K. B. and Faloona F. A., 1987) vývoj automatického Sangerova sekvenování (Ansorge W. et al., 1987; Smith L. M. et al., 1986), které se stalo nejpoužívanější DNA sekvenační technologií pro téměř dalších 20 let. Během tohoto období byla optimalizována pro čtení delších DNA fragmentů a pro vyšší kapacitu. V současnosti tato technologie umožňuje simultánní sekvenování až 384 fragmentů s maximální délkou čtení až tisíc párů bází (bp). Automatické Sangerovo sekvenování bylo hlavní technologií využívanou v rámci Projektu lidského genomu "Human Genome Project" zahájeném v roce 1990 s cílem identifikovat tři miliardy párů bází tvořících lidský genom. První výsledky tohoto projektu byly publikovány za deset let od zahájení projektu (Venter J. C. et al., 2001) a projekt byl dokončen po dalších třech letech (International Human Genome Sequencing C., 2004). Projekt lidského genomu vyžadoval rozsáhlé a ekonomicky náročné sekvenování a ve svém důsledku vedl nejen k identifikaci lidského genomu, ale také k vývoji nových sekvenačních technologií. Tyto nové masivně paralelní sekvenační technologie (NGS), umožňují efektivní a rychlé sekvenování celých genomů. Jejich rozvoj způsobil revoluci v možnostech studia variability lidského genomu jednotlivců (Levy S. et al., 2007; Wheeler D. A. et al., 2008) a následně i celých populací (Siva N., 2008). V rámci tohoto projektu bylo možné během pouhých 4 let analyzovat kompletní sekvenci genomu 1092 jednotlivců ze 14 různých populací (Genomes Project C. et al., 2012). Současné verze sekvenačních přístrojů umožňují v rámci jednoho běhu analyzovat až 12 lidských genomů. Jádru výzkumu se proto přesouvá od získávání sekvenčních dat k problematice jejich analýzy, interpretace, ukládání a zálohování.

### **Masivně paralelní sekvenování**

Pokrok ve vývoji sekvenačních technologií umožňuje efektivně a dostupně generovat velké množství sekvenačních dat. Nové sekvenační technologie označované jako "next generation sequencing" poskytují možnost analyzovat celé genomy. Principem těchto metod je postup, který umožňuje náhodně rozmístit a poté amplifikovat jednotlivé molekuly komplexní směsi DNA. S využitím polymerázy nebo ligázy jsou v opakujících se krocích do analyzovaných



řetězců DNA inkorporovány komplementární báze a tento proces je monitorován detekcí fluorescenčního nebo jiného signálu. Toto umožňuje generovat statisíce až miliardy sekvenačních čtení s délkou 75 až 1000 bází. Tato data jsou obvykle namapována na referenční sekvenci, nalezené rozdíly anotovány a dále interpretovány. Přestože cena sekvenování dramaticky poklesla, stále není dostatečně nízká, aby bylo možné provádět celogenomové sekvenování v rozsáhlých genetických studiích s dostatečnou statistickou silou. Efektivnějším přístupem je zaměřit se pouze na relevantní genomické oblasti a analyzovat větší množství vzorků. Vývoj nových sekvenačních technologií způsobil revoluci v postupech, jakými jsou prováděny sekvenační analýzy a klinické genetické testování. Tato sekce není zamýšlena jako přesný popis každé z nových sekvenačních technologií, ale spíše jako úvod poskytující odpovídající kontext pro využívané metody.

### **Roche 454 GS FLX**

Prvním dostupným přístrojem nové generace byl sekvenátor 454 GS využívající technologii pyrosekvenování (Ronaghi M. et al., 1996), uvedený na trh v roce 2005 (Margulies M. et al., 2005). Molekuly sekvenované DNA knihovny jsou navázány na kuličky a podrobeny emulznímu PCR (Dressman D. et al., 2003), poté jsou kuličky imobilizovány v sekvenační destičce. V každé jamce sekvenační destičky se nachází právě jedna kulička, to umožňuje detekovat signál z jedné výchozí molekuly DNA. Poté jsou v opakujících se sekvenačních cyklech přidávány jednotlivé deoxynukleotidy. Uvolněný pyrofosfát je detekován luciferázovou reakcí. Aktuální verze sekvenátoru 454 FLX umožňuje generovat až milion čtení s průměrnou délkou 700 bp.

### **Applied Biosystems SOLiD**

Technologie firmy Applied Biosystems uvedená v roce 2007 je založena na sekvenačním postupu využívajícím ligázu (Shendure J. et al., 2005). Stejně jako v případě 454 je pro amplifikaci jednotlivých molekul DNA knihovny použita metoda emulzního PCR. Po amplifikaci jsou kuličky nesoucí templát na 3' konci modifikovány tak, aby byla umožněna vazba s nosičem. Sekvenační reakce probíhá na mikroskopickém sklíčku, na které jsou náhodně kovalentně navázány kuličky nesoucí templát. V průběhu každého sekvenačního cyklu dojde k navázání fluorescenčně značeného oktadimetru. Fluorescenční značka je

specifická pro první dvě báze, po detekci signálu je odstraněna fluorescenční značka přítomná na posledních dvou bazích oktameru. Po dosažení nastaveného počtu sekvenačních cyklů je proces ještě čtyřikrát opakován s použitím dalších sekvenačních primerů tak, že přítomnost každé báze je detekována dvěma překrývajícími se oktamery. Díky tomu, že každá pozice analyzovaného fragmentu je charakterizována dvěma fluorescenčními signály, je zajištěna vyšší přesnost určení báze v dané pozici. Aktuální verze sekvenátoru SOLiD produkují až 1.6 miliardy sekvenačních čtení o maximální délce 75 bp.

## **Illumina**

Technologie firmy Illumina, uvedená v roce 2006, je založena na sekvenování syntézou s využitím reverzibilně terminovaných a značených deoxynukleotidů. Sekvenování probíhá v sekvenační cele, na jejímž povrchu jsou navázány oligonukleotidy komplementární k adaptorům nalogovaným na sekvenovaných fragmentech. Po navázání fragmentů je provedena tzv. můstková amplifikace, dochází tak ke vzniku kolonií (clusters) z každého navázaného fragmentu (Fedurco M. et al., 2006).

Sekvenační reakce je zahájena navázáním sekvenačního primeru na templát. Poté je přidána směs značených deoxynukleotidů a dojde k inkorporaci první báze. Po každém cyklu je s využitím laseru osvícen celý povrch flowcely a zaznamenán fluorescenční signál. Následuje odstranění terminátoru a fluorescenční značky. Délka čtení je určena počtem sekvenačních cyklů a jednotlivé báze emitovanou vlnovou délkou fluorescenčního signálu. Aktuální verze sekvenátorů Illumina produkují až 3 miliardy párových čtení s maximální délkou 250 bp.

## **Metody obohaceni DNA**

I když lze sekvenovat celé lidské genomy, často je efektivnější zaměřit se pouze na specifické oblasti zájmu. Tento fakt vedl k rozvoji technologií umožňujících sekvenovat pouze vybrané oblasti genomu (Mamanova L. et al., 2010).

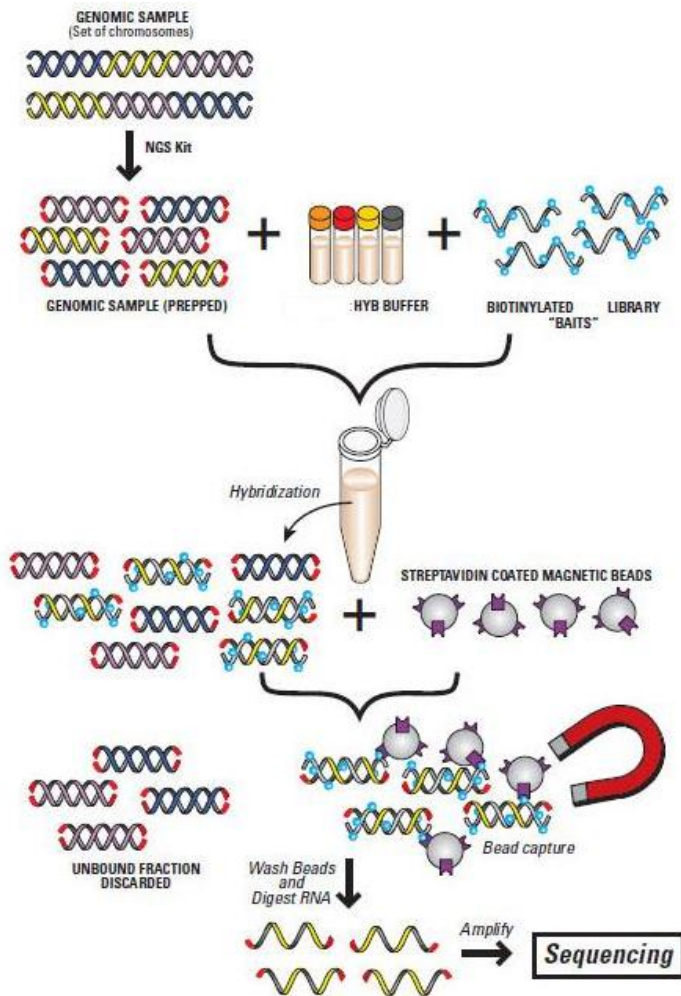
Standardní molekulární technikou pro selektivní amplifikaci genomových oblastí je polymerázová řetězová reakce (PCR). PCR amplikony jsou vhodným vstupem pro NGS sekvenování zejména v případě, že počet amplifikovaných oblastí je malý, nicméně se

vzrůstající velikostí oblastí je tento přístup neefektivní. Tento nedostatek je možné obejít využitím multiplexního PCR a automatizované technologie (Porreca G. J., 2007; Tewhey R., 2009).

Další metody, které byly vyvinuty specificky pro potřeby NGS sekvenování lze obecně rozdělit do dvou kategorií na hybridizační a enzymatické.

Hybridizační metody využívají oligonukleotidové sondy, které jsou komplementární k oblastem zájmu. Postup obohacení gDNA knihoven založený na DNA čípech (Albert T. J. et al., 2007) byl později nahrazen snadnějším přístupem, kdy se obohacení provádí v roztoku. Řešení založené na obohacení v roztoku mají řadu výhod oproti metodě založené na DNA čípech, zejména schopnost zachytit větší oblasti, možnost automatizace, menší pracnost a časovou náročnost (Bainbridge M. N. et al., 2010). Principem této metody je využití biotinem značených RNA nebo DNA sond (baits), které se hybridizují s gDNA knihovnou. Po hybridizaci jsou k roztoku přidány magnetické kuličky pokryté streptavidinem, na které se naváží biotinem značené sondy nesoucí fragmenty zachycené DNA. Pomocí magnetu je komplex zachycen na stěně zkumavky a nenavázaná DNA odmyta. Zachycená DNA je eluována z kuliček a připravena pro sekvenování (viz. Obrázek 2). Řešení založená na hybridizaci jsou dominantní metodou pro cílené exomové sekvenování.

Enzymatické metody využívají pro obohacení DNA molekulární inverzní sondy (MIPS) (Turner E. H. et al., 2009), jejich výhodou je malé vstupní množství DNA a vysoká specifita obohacení, nevýhodou jsou vysoké náklady.



**Obrázek 2.** Princip obohacení DNA hybridizací v roztoku – fragmentovaná DNA s naligovanými sekvenčními adaptory je hybridizována s knihovnou biotinem značených sond komplementárních k vybraným oblastem. Po hybridizaci jsou k roztoku přidány magnetické kuličky pokryté streptavidinem na které se naváží biotinem značené sondy nesoucí fragmenty zachycené DNA. Pomocí magnetu je komplex zachycen na stěně zkumavky a nenavázaná DNA odmyta. Zachycená obohacená DNA je eluována z kuliček a připravena pro sekvenování. (modifikováno podle [www.genomics.agilent.com](http://www.genomics.agilent.com))

## Analýza sekvenčních dat

Výstupem ze sekvenátoru jsou obvykle data ve formátu FASTQ obsahující jednotlivá čtení a kvalitu přečtených bazí.

Analýza sekvenčních dat je řadou na sebe navazujících kroků, které zahrnují kontrolu kvality, přiřazení sekvenčních čtení k referenční sekvenci, nalezení odchylek oproti referenční sekvenci a jejich funkční anotaci.

Analýza sekvenačních dat není jasně definovaný proces, existuje velké množství používaných nástrojů např. BWA (Li H. and Durbin R., 2009), Samtools (Li H. et al., 2009), GATK (McKenna A. et al., 2010) a výsledky se mohou v závislosti na použitých programech a parametrech analýzy podstatným způsobem odlišovat (O'Rawe J. et al., 2013). Je proto důležité chápat principy a omezení jednotlivých kroků analýzy dat.

Prvním krokem je kontrola kvality, kdy jsou ze sekvenačních dat odstraněny čtení s nízkou kvalitou a kontaminace, například sekvence adaptorů používaných v průběhu přípravy sekvenační knihovny.

Poté jsou jednotlivá čtení přiřazena k referenční sekvenci, tento krok představuje výpočetně nejnáročnější část zpracování sekvenčních dat. Výstupem tohoto kroku jsou data ve formátu BAM. BAM formát je definován tak, aby obsahoval veškeré informace nezbytné pro další zpracování dat (Li H. et al., 2009). Pro každý fragment je zaznamenána přesná genetická pozice včetně kvality mapování a informace o nalezených rozdílech oproti referenční sekvenci.

Následuje řada kroků, které umožňují správnou detekci variant - odstranění PCR duplikátů, zarovnání okolo inzercí/delecí a recalibrace kvality bází (Van der Auwera G. A. et al., 2013).

Posledním krokem je genotypování. V průběhu tohoto kroku jsou nalezeny všechny odchylky přítomné v analyzovaném vzorku oproti referenční sekvenci. Tyto změny jsou definovány genomickou pozicí a nalezenou sekvenční změnou. Výstupem tohoto kroku je VCF soubor, který je seznamem všech nalezených variant a jejich kvalit (Danecek P. et al., 2011). Takto získaný seznam variant je následně funkčně anotován (Wang K. et al., 2010).

## Metody hledání genů podmiňujících dědičná onemocnění

Propojení genotypu s fenotypem je jedním z hlavních cílů genetiky. Unikátní možnosti studia v tomto ohledu nabízí výzkum vzácných geneticky podmíněných onemocnění.

Vzácná geneticky podmíněná onemocnění jsou skupinou převážně monogenních onemocnění, která dle definice postihují méně než 1 osobu z 2000 v Evropské Unii nebo méně než 1 osobu z 1250 ve Spojených státech (Remuzzi G. and Garattini S., 2008). Počet vzácných onemocnění je odhadován na více než 7000 (McKusick V. A., 2007). Přibližně u poloviny z nich je genetická příčina stále neznámá a pouze pro 5% z těchto onemocnění existuje v současnosti účinná léčba (Rohn J., 2013). Identifikace genů odpovědných za tato onemocnění umožňuje molekulární diagnostiku, prenatální diagnostiku a představuje první krok k lepšímu porozumění fyziologické funkce genů, proteinů a spojených biologických procesů, které je nezbytné pro vývoj léčiv.

Metody používané k objasnění genetické podstaty lidských onemocnění a dalších znaků spojených se zdravím jedince často vychází ze zjednodušujícího rozdělení na monogenní, vzácná a komplexní, populačně častá onemocnění. Nejpoužívanější metodou v poslední době bylo genetické mapování pomocí vazebné analýzy, které není závislé na jakékoliv předchozí znalosti biologie nebo funkce, a místo toho je založeno čistě na sledování dědičnosti studovaných znaků ve spojení s genetickými markery. S pomocí vazebné analýzy a sekvenování kandidátních genů byla odhalena molekulární podstata řady známých fenotypů s předpokládanou mendelovskou dědičností. Například geny pro cystickou fibrózu (Tsui L. C. et al., 1985), Huntigtonovu chorobu (Gusella J. F. et al., 1983) a diabetes mellitus (Bell G. I. et al., 1984).

Na druhém konci spektra celogenomové asociační studie identifikovaly velké množství oblastí přispívajících ke vzniku komplexních onemocnění (Welter D. et al., 2014), bohužel prakticky ve všech případech tyto oblasti vysvětlují pouze malou část pozorované heritability studovaných znaků (Manolio T. A. et al., 2009). Navíc i velká část populačně častých onemocnění, u kterých byla dříve předpokládána složitá multifaktoriální dědičnost, je nyní považována za heterogenní skupinu vzácných onemocnění (McClellan J. and King M. C., 2010).

Existuje mnoho faktorů, které komplikují možnost využití tradičních technik, například pouze malé množství pacientů dostupných k analýze, snížená penetrance onemocnění, heterogenita a snížená reprodukční schopnost postižených jedinců (Antonarakis S. E. and Beckmann J. S., 2006). Obejít tyto problémy je možné s využitím sekvenačního přístupu, kdy lze mutace identifikovat přímo prostřednictvím sekvenování (Ng S. B. et al., 2010). Nicméně až do nedávné doby, toto bylo velmi náročné na zdroje a obecně nemožné provést ve velkém měřítku nebo na úrovni celého genomu.

Sekvenování celého genomu nebo exomu pacienta také odstraňuje potřebu vybírat kandidátní geny pro sekvenování pozičním mapováním a zjednodušuje proces identifikace z dvou-stupňového (poziční mapování následované Sangerovým sekvenováním) na jednostupňový (exomové nebo celogenomové sekvenování). S rozvojem NGS se těžiště přesouvá od identifikace k interpretaci variant, jsou identifikovány desetitisíce variant, ale pouze jedna nebo dvě vysvětlují onemocnění. Postup filtrace variant proto představuje zásadní krok v identifikaci příčinných genů.

V následující části jsou popsány principy využívaných metod.

### **Vazebná analýza**

Vazba je tendence lokusů se dědit společně, protože nejsou odděleny rekombinací během meiózy díky malé vzájemné vzdálenosti. Cílem vazebné analýzy je identifikovat chromozomální oblast, která segreguje se sledovaným fenotypem v jedné nebo více rodinách. To se provádí pomocí celogenomového genotypování pravidelně rozložených genetických markerů se známou chromozomální pozicí a následnou počítačovou analýzou. V průběhu počítačové analýzy je pro každý marker a všechny analyzované jedince vypočtena celková pravděpodobnost jako podíl pravděpodobnosti, že dva lokusy jsou ve vazbě (rekombinační frakce =  $\theta$ ) a pravděpodobnosti, že ve vazbě nejsou (rekombinační frakce = 0,5). Tento poměr udává pravděpodobnost vazby a logaritmus tohoto poměru se označuje jako LOD skóre (Morton N. E., 1955). Na základě konvence se hodnota LOD skóre větší než 3 považuje za důkaz vazby a hodnota menší než -2 za vyloučení vazby. Markery, které nerekombinují se studovaným fenotypem díky vzájemné blízkosti s hledaným genem, vymezují kandidátní oblast obsahující hledaný gen. Protože počet rekombinací je v rámci rodiny omezen, výsledná vazebná oblast má obvykle velikost 1- 10 cM.

## **Homozygotní mapování**

Homozygotní mapování je dalším postupem pro lokalizaci genů podmiňujících vzácná recesivní onemocnění (Lander E. S. and Botstein D., 1987). Tento přístup předpokládá příbuznost rodičů postiženého jedince. Principem metody je hledání dlouhých autozygotních oblastí (identical by descent, pocházejících od společného předka), které pravděpodobně obsahují genetickou variantu podmiňující fenotyp. Homozygotní mapování lze provádět s využitím genotypovacích čipů s vysokou hustotou nebo genotypů získaných pomocí exomového sekvenování.

## **Asociační analýza**

Asociační analýza je přístupem pro mapování genů, který přímo testuje vztah mezi konkrétní alelou, genotypem nebo haplotypem a studovaným znakem (onemocněním). Asociační studie jsou vhodným nástrojem pro posouzení kandidátních genů, upřesnění oblastí definovaných vazebnou analýzou a díky dostupnosti DNA čipů s vysokou hustotou také pro celogenomové mapování oblastí spojených s populačně častými, komplexními onemocněními (Cardon L. R. and Bell J. I., 2001). V typickém uspořádání, případ-kontrola, je porovnávána alelická frekvence určitého genetického markeru mezi skupinou nepříbuzných kontrol a skupinou nepříbuzných jedinců nesoucích studovaný znak. Tyto dvě skupiny musí být srovnatelné z hlediska etnického původu. Pokud je nalezena asociace mezi zkoumanou genetickou variantou a onemocněním, lze předpokládat, že tato varianta nějakým způsobem souvisí s onemocněním, nebo je ve vazebné nerovnováze s příčinnou mutací. Nevýhodou asociační analýzy je množství falešně pozitivních výsledků, v důsledku různého populačního rozvrstvení porovnávaných skupin (Cardon L. R. and Palmer L. J., 2003). Výhodou naopak schopnost detekovat geny s relativně malým příspěvkem ke studovanému onemocnění.

Variantou kombinující přístup vazebné analýzy a asociace je test nerovnováhy přenosu (TDT)(Ewens W. J. and Spielman R. S., 1995). Tento test srovnává frekvenci přenosu nebo nepřenosu daného genetického markeru na postižené potomky od heterozygotních rodičů. V případě, že testovaný marker zvyšuje riziko onemocnění, je přenášen na postižené potomky se zvýšenou frekvencí. Výhodou tohoto přístupu je odolnost proti populačnímu rozvrstvení, nevýhodou může být obtížnost shromáždit dostatečný počet rodin k provedení analýzy, zejména u chorob s pozdním nástupem.



## **Analýza počtu změn kopií DNA**

Delece a duplikace, typy strukturních variant s velikostí větší než 50bp jsou označovány jako CNV (copy number variants) (MacDonald J. R. et al., 2014). CNV významným způsobem přispívají ke genetické variabilitě populace – v současné době bylo identifikováno více než 350 tis. CNV ovlivňujících přibližně 9.5% lidského genomu (Zarrei M. et al., 2015), mohou být zděděny nebo vznikají de-novo v průběhu meotického dělení. CNV jsou asociovány s řadou patologických stavů, jako jsou například schizofrenie (Malhotra D. and Sebat J., 2012), autismus (Pinto D. et al., 2014), Crohnova choroba (Wellcome Trust Case Control C. et al., 2010) a mnoha dalších. Informace o CNV z různých projektů jsou shromažďovány v databázích. Příkladem jsou Database of Genomic Variants (Lafrate A. J. et al., 2004) obsahující varianty nalezené v kontrolních souborech a databáze DECIPHER obsahující varianty nalezené v souborech pacientů (Firth H. V. et al., 2009).

Identifikace CNV je možná s využitím genotypovacích nebo CGH čipů, pomocí kterých je možné v závislosti na použité platformě spolehlivě detekovat změny větší než 10kb (Haraksingh R. R. et al., 2011). Další, zejména v poslední době využívanou metodou je celogenomové sekvenování, které oproti DNA čipům umožňuje přesné určení pozice a je vhodné i pro identifikaci velmi malých změn (Mills R. E. et al., 2011).

## **Exomové sekvenování**

Nedávný pokrok v sekvenačních technologiích zásadně změnil způsob, jakým jsou identifikovány geny podmiňující neznámá onemocnění. Přestože je možné sekvenovat celé lidské genomy, analýza těchto dat je velmi náročná. U vzácných Mendelovských onemocnění se předpokládá, že mutace mají velký efekt a proto se unikátně vyskytují pouze u pacientů nebo s velmi malou frekvencí v populaci, jsou lokalizovány v oblastech genomu kódujících proteiny a přímo ovlivňují funkci proteinu kódovaného mutovaným genem (Ng S. B. et al., 2010). Efektivním přístupem proto je zaměřit se pouze na oblasti genomu kódující proteiny (exom). Exomové sekvenování je proces, ve kterém jsou analyzovány všechny oblasti kódující proteiny v celém genomu (Ng S. B. et al., 2009). V současnosti existuje mnoho komerčně dostupných řešení pro přípravu exomových sekvenačních knihoven, které se liší velikostí cílené oblasti (Clark M. J. et al., 2011). Některé obsahují pouze kódující

oblasti - exony, jiné i další funkčně významné elementy např. miRNA nebo nepřekládané oblasti genů.

Výsledkem exomového sekvenování jsou desetitisíce variant, je proto důležité zvolit přístup umožňující efektivně vybrat kandidátní varianty. Počet kandidátních variant je možné výrazně omezit i správným výběrem vzorků pro analýzu (Cheung C. Y. et al., 2014). Například pro dominantní onemocnění je vhodným přístupem vybrat jedince, které odděluje největší počet meióz.

Analýza exomových dat je založena na filtrování nalezených variant, které se obvykle provádí dle kvality genotypu (např. počet nezávislých čtení podporujících variantu, procento čtení obsahující variantu), efektu varianty na sekvenci proteinu, populační frekvenci varianty (1000Genomes, EVS, ExAC, dbSNP), přítomnosti varianty v databázi patogenních variant (HGMD (Stenson P. D. et al., 2009), ClinVar (Landrum M. J. et al., 2014)), genomické pozici varianty - pokud jsou k dispozici výsledky vazebné analýzy nebo homozygotního mapování, evoluční konzervovanosti (PhyloP (Pollard K. S. et al., 2010), GERP (Cooper G. M. et al., 2005)), predikce škodlivosti varianty (SIFT (Ng P. C. and Henikoff S., 2001), PolyPhen (Adzhubei I. A. et al., 2010), CADD (Kircher M. et al., 2014)), předpokládaného modelu dědičnosti onemocnění (segregace varianty v rámci rodiny a u postižených jedinců), exprese genů v tkáních postižených studovaným onemocněním a relevanci známé funkce genů ke studovanému onemocnění. Parametry pro filtrování je nutné nastavit podle předpokládaného modelu dědičnosti a prevalence studovaného onemocnění. Nesprávné nastavení filtračních kroků může odstranit i příčinnou variantu.

Úspěšnost exomového sekvenování při identifikaci genů podmiňujících Mendelovská onemocnění se pohybuje okolo 60% (Gilissen C. et al., 2012). Hlavní nevýhodou exomového sekvenování jsou především technická omezení (problematické pokrytí GC bohatých a sekvenčně nespécifických oblastí) a omezení daná principem této metody (závislost na definici oblastí v použitém kitu, nemožnost detekovat nekódující varianty, omezená možnost detekce CNV).

## Část III. Studium genetické podstaty vzácných onemocnění

### Rotorův syndrom

Rotorův syndrom (RS, OMIM#237450) je typem vzácné dědičné konjugované hyperbilirubinémie spojené s koproporfyriinurií a sníženým jaterním vstřebáváním mnoha diagnostických látek včetně cholelscintigrafických radiofarmak (Fretzayas A. M. et al., 1994). Svůj název nese po filipínském lékaři Arturovi Bellezovi Rotorovi, který syndrom s velmi vzácnou prevalencí popsal již v roce 1948 (A. B. Rotor L. M., A. Florentin, 1948). RS je autozomálně recesivní onemocnění, které je klinicky podobné dalšímu typu vrozené hyperbilirubinémie, Dubin-Johnsonovu syndromu (DJS, OMIM#237500). Hlavním rozdílem oproti DJS je nepřítomnost pigmentových deposit v hepatocytech. Bilirubin je látka vznikající rozkladem hemu a jeho metabolismus byl dosud popisován jako jednosměrný proces skládající se ze dvou kroků. Nejprve je nekonjugovaný bilirubin přenesen do hepatocytu, kde se konjuguje s glukuronovou kyselinou za pomoci glukuronosyltransferázy a následně je vyloučen do žluče.

RS jsme začali studovat v roce 2006 ve spolupráci s Institutem klinické a experimentální medicíny. Vzhledem k dosud nejasné molekulární podstatě RS a jeho podobnosti s DJS byla nejprve testována hypotéza, že RS by mohl být alelickou variantou DJS, který je způsoben mutacemi v genu *ABCC2* (Paulusma C. C. et al., 1997). Proto byla u dvou pacientů provedena mutační analýza genu *ABCC2* s využitím Sangerova sekvenování s negativním výsledkem. Také imunohistochemické nálezy neukázaly žádný rozdíl oproti zdravým kontrolám. Pro vyloučení velkých delecí v genu *ABCC2*, které nejsou detekovatelné Sangerovým sekvenováním, byl navržen DNA čip pro komparativní genomovou hybridizaci (CGH). Pomocí CGH nebyly nalezeny žádné změny v počtu kopií u všech exonů genu *ABCC2*. Tyto výsledky vyloučily možnost, že RS je alelickou variantou DJS (příloha 1a).

Ve studiu RS jsme dále pokračovali genotypováním 11 pacientů z 8 rodin s využitím DNA čipu Affymetrix SNP 6.0. Homozygotní mapování definovalo u všech pacientů jedinou společnou oblast na chromozomu 12 přítomnou na třech různých haplotypech. Souběžně provedená analýza počtu změn kopií odhalila dvě změny ve stejné oblasti, homozygotní delecí části genu *SLCO1B3* přítomnou na haplotypu R1 a homozygotní delecí v oblasti genu *SLCO1B3*, *SLCO1B7* a *SLCO1B1* na haplotypu R2. Následná sekvenační analýza odhalila

homozygotní mutace v genu *SLCO1B3* u haplotypu R1 a v genech *SLCO1B3* a *SLCO1B1* u haplotypu R3. U všech probandů byly nalezeny delece nebo mutace postihující obě alely genů *SLCO1B3* a *SLCO1B1*. Absence proteinů OAPT1B3 a OATP1B1, které jsou kódovány geny *SLCO1B3* a *SLCO1B1* byla potvrzena imunohistochemickým barvením jaterních biopsií pacientů. Tyto výsledky potvrdily, že RS je způsoben kompletním defektem obou alel genů *SLCO1B1* a *SLCO1B3*.

Proteiny OATP1B1 a OATP1B3 lokalizované na sinusoidní membráně hepatocytu jsou hlavními jaterními transportéry převážně organických aniontů ale i dalších látek. Bylo prokázáno, že jejich substráty je řada endogenních, ale i exogenních látek, jako jsou například žlučové kyseliny, steroidní sulfáty, thyroïdní hormony, konjugovaný bilirubin, statiny, paclitaxel, rifampicin a mnoho dalších (Hagenbuch B. and Gui C., 2008; Niemi M. et al., 2011).

Nezávisle byl skupinou z The Netherlands Cancer Institute v Amsterdamu studován transport bilirubinu na myším modelu s deficiencí proteinů Oatp1a/1b, Abcc3 a Abcc2, myších homologů lidských proteinů OATP1B1, OATP1B3, ABCC3 a ABCC2. U *Slco1a/1b*<sup>-/-</sup> myši byly pozorovány zvýšené hodnoty bilirubinu v plazmě, ty jsou významně sníženy u *Slco1a/1b; Abcc3*<sup>-/-</sup> myši, přičemž bylo dokázáno, že Abcc3 protein odpovídá z největší části za zvýšené hodnoty bilirubinu v plazmě. Z výsledků práce vyplývá, že Abcc3 transportuje konjugovaný bilirubin z hepatocytů zpět do krve a proteiny Oatp1a a Oatp1b transportují tento bilirubin z krve zpět do hepatocytů. Transgenní exprese lidského OATP1B3 nebo OATP1B1 proteinu v *Slco1a/1b*<sup>-/-</sup> myši vede k normalizaci hladin bilirubinu v plazmě. Tím bylo potvrzeno, že oba lidské proteiny OATP1B3 i OATP1B1 transportují konjugovaný bilirubin z plazmy zpět do hepatocytu. Tyto výsledky ukazují, že exkreční dráha bilirubinu není jednosměrným transportem bilirubinu z krve do hepatocytu a následně do žluče. Ale že část konjugovaného bilirubinu je z hepatocytu vyloučena pomocí transportéru ABCC3 do krve, odkud je zpětně reabsorbována pomocí OATP1B3 a OATP1B1 a následně je vyloučena do žluče pomocí transportéru ABCC2.

Propojení výsledků těchto studií umožnilo objasnit příčinu a projevy Rotorova syndromu a definovat nový mechanismus transportu bilirubinu v játrech.

## Deficit ATP syntázy

Mitochondriální ATP syntáza je klíčovým enzymem mitochondriálního energetického metabolismu katalyzujícím syntézu ATP v procesu oxidativní fosforylace. ATP syntáza je proteinový komplex složený z 16 typů podjednotek (Collinson I. R. et al., 1996), dvě z těchto jednotek jsou kódovány mtDNA (*ATP6, ATP8*) a zbytek jadernou DNA. Mitochondriální onemocnění spojená s izolovaným deficitem ATP syntázy jaderného původu (OMIM#604273) jsou charakterizována snížením množství enzymu pod 30% spojeným se ztrátou syntetické i hydrolytické aktivity (Houstek J. et al., 1999). Onemocnění se projevuje již v novorozeneckém nebo kojeneckém období, nejčastějšími příznaky jsou laktátová acidóza, hypertrofická kardiomyopatie, poškození CNS a 3-metylglutakonová acidurie (Sperl W. et al., 2006). S cílem identifikovat gen podmiňující onemocnění jsme navrhli vlastní čip pro studium genové exprese h-MitoArray obsahující celkem 1632 převážně mitochondriálních genů a využili jej ke studiu genové exprese ve fibroblastech pacientů s popsaným defektem ATP syntázy. Porovnání expresních profilů, funkční anotace a metoda genového obohacení rozdělila pacienty do tří specifických skupin, kandidátní gen však nebyl nalezen (příloha 2a). Proto jsme ve studiu dále pokračovali a pomocí genotypovacích čipů Affymetrix analyzovali 8 postižených jedinců a 13 jejich rodičů nebo nepostižených sourozenců celkem z 6 rodin. Pomocí homozygotního mapování byla nalezena jediná společná oblast na chromozomu 8 o velikosti přibližně 1 Mb obsahující celkem 7 genů. Současně byla analyzována genová exprese v patientských a kontrolních fibroblastech s využitím DNA čipu Agilent 44k. Propojení výsledků těchto analýz definovalo kandidátní gen - *TMEM70*, který se nacházel ve sdílené homozygotní oblasti a zároveň měl významně sníženou expresi oproti kontrolním vzorkům. Sekvenační analýzou tohoto genu byla nalezena homozygotní mutace 317-2A > G (NM017866) vedoucí k aberantnímu sestřihu a ztrátě transkriptu. Následně byla shodná mutace identifikována u 23 z 25 dostupných pacientů. Funkční význam této mutace byl potvrzen komplementační studií. Vnesení wt formy genu *TMEM70* do patientských fibroblastů vedlo ke zvýšení množství ATP syntázy a obnovení funkce respiračního řetězce. Fylogenetická analýza potvrdila přítomnost homologů genu *TMEM70* u vyšších eukaryot a rostlin, ne však u kvasinek a hub. Bylo tak prokázáno, že *TMEM70* se účastní biogeneze ATP syntázy u vyšších eukaryot a jeho defekt je relativně častou příčinou mitochondriálních onemocnění, zvláště v romské populaci.

## **Mukopolysacharidóza typu IIIC**

Mukopolysacharidózy (MPS) patří do skupiny stádavých lyzozomálních onemocnění, jejichž příčinou je deficit enzymů katalyzujících degradaci glykosaminoglykanů. Mukopolysacharidóza typu IIIC (MPSIIIC, Sanfilippo syndrom C, OMIM #252930) je autozomálně recesivní onemocnění způsobené deficitem enzymu alfa-glukosamin N-acetyltransferázy (Klein U. et al., 1978), které se projevuje převážně postižením centrálního nervového systému. Gen pro N-acetyltransferázu byl již dříve mapován do 8.3 cM oblasti na chromozómu 8 (Ausseil et al. 2004). Pro upřesnění již dříve reportované kandidátní oblasti na chromozómu 8 byla provedena vazebná analýza s využitím STR markerů u 5 pacientů ze 4 nepříbuzných rodin a jejich 49 rodinných příslušníků diagnostikovaných na základě vyšetření aktivity N-acetyltransferázy. Výsledkem byla kandidátní oblast o velikosti 2.6 cM obsahující 32 známých nebo predikovaných genů. Následně byla provedena expresní analýza těchto genů v leukocytech pacientů s využitím připraveného DNA čipu. *HGSNAT* (dříve *TMEM76*) byl vybrán jako kandidátní gen na základě známých vlastností enzymu (předpokládaná velikost proteinu, přítomnost transmembránových domén) a snížené exprese tohoto genu u pacientů. S využitím sekvenační analýzy na rozšířeném souboru pacientů byly nalezeny v genu *HGSNAT* 4 nesmyslné mutace, 11 mutací měnících smysl, 2 mutace způsobují posun čtecího rámce, 6 sestřihových mutací a jedna rozsáhlá delece. Funkční význam genu *HGSNAT* byl potvrzen komplementační studií.

## **Autozomálně dominantní ANCL**

Neuronální ceroidní lipofuscinózy (NCL) jsou heterogenní skupinou vzácných dědičně podmíněných neurodegenerativních onemocnění, jejichž společným znakem je na buněčné úrovni stádání autofluorescentního materiálu (lipofuscinu) v lyzozomech neuronů CNS a v periferních tkáních. Mezi charakteristické projevy onemocnění patří progresivní porucha zraku, epilepsie, parkinsonismus a zhoršení kognitivních funkcí vedoucí k demenci. Podle věku nástupu onemocnění jsou NCL děleny na infantilní, pozdně infantilní, časně juvenilní, juvenilní a adultní formy (Mole S. E. et al., 2011). Léčba žádné z forem NCL není v současnosti dostupná, jedinou možností je prevence onemocnění s využitím postupů prenatální a preimplantační diagnostiky. Ve spojení s NCL bylo doposud popsáno dvanáct genů (*PPT1, TPP1, CLN3, CLN5, CLN6, MFSD8, CLN8, CSTD, CTSF, GRN, ATP13A2, KCTD7*). Genetická

podstata autozomálně dominantní adultní formy neuronální ceroidní lipofuscinózy (ANCL) (CLN4B, OMIM#162350) nebyla zatím objasněna.

Pro studium molekulární podstaty autozomálně dominantní ANCL byla využita kombinace metod vazebné analýzy, expresní analýzy, analýzy počtu změn kopií a exomového sekvenování. Výsledky vazebné analýzy nejprve definovaly pět kandidátních oblastí na chromosomech 1, 4, 15, 20 a 22. Paralelně provedená analýza změn počtu kopií u 7 pacientů neodhalila žádné CNV větší než 10kb segregující s onemocněním. Následně byla s cílem identifikovat varianty ovlivňující množství transkripce provedena analýza genové exprese v leukocytech 4 pacientů a kontrol. Výsledkem této analýzy byl seznam 2131 rozdílně exprimovaných genů, z nichž se 65 nacházelo v oblastech definovaných vazebnou analýzou. Vzhledem ke stále velkému počtu kandidátních genů bylo provedeno exomové sekvenování jednoho pacienta na sekvenátoru SOLiD 4. Pomocí exomového sekvenování bylo identifikováno u tohoto pacienta celkem 957 unikátních variant nepřítomných v populačních databázích (dbSNP, 1000Genomes). Propojením výsledků vazebné analýzy, expresní analýzy a exomového sekvenování byla nalezena heterozygotní mutace v genu *DNAJC5* c.346\_348delCTC (p. Leu116del). Segregace této mutace byla u dalších postižených členů rodiny ověřena Sangerovým sekvenováním. Díky spolupráci s Rare NCL Gene Consortium byla sekvenační analýzou následně nalezena stejná mutace u dalšího nepříbuzného pacienta a zároveň identifikovaná druhá varianta c.344T>G(p.Leu115Arg) u 3 dalších nepříbuzných pacientů.

*DNAJC5* kóduje cysteine-string protein alpha (CSP $\alpha$ ), evolučně konzervovaný membránový protein lokalizovaný v synaptických membránách neuronů (Tobaben S. et al., 2001). Jeho mutace vedou u modelových organismů k neurodegeneraci a zkrácení délky života (Schmitz F. et al., 2006; Zinsmaier K. E. et al., 1994). Význam nalezených variant na funkci proteinu byl ověřen pomocí *in-silico* analýzy, kdy nalezené mutace snižují hydrofobicitu a palmitoylaci proteinu, studií v tkáňových kulturách neuronálních buněk byla zjištěna změna lokalizace mutovaného proteinu a imunohistochemické barvení ukázalo snížené množství nebo absenci proteinu v šedé hmotě mozkové kůry pacientů. Prokázali jsme tak, že mutace v genu *DNAJC5* jsou příčinou autozomálně dominantní formy ANCL.

## **GAP0 syndrom**

Gapo syndrom (OMIM#230740) je velmi vzácné autozomálně recesivní onemocnění. Název syndromu je zkratkou anglických slov popisujících hlavní projevy syndromu - **G**rowth retardation (růstovou retardaci), **A**lopecia (plešatost), **P**pseudoanodontia (porucha prořezávání zubů) a **O**ptic atrophy (atrofie optického nervu) (Tipton R. E. and Gorlin R. J., 1984). Doposud bylo celosvětově popsáno pouze okolo 40 pacientů. Většina postižených tímto syndromem pochází z příbuzenských svazků.

Ve spolupráci s Klinikou dětského a dorostového lékařství jsme měli možnost studovat již dříve reportovanou rodinu s jedním postiženým potomkem (Baxova A. et al., 1997). S využitím DNA čipu Affymetrix SNP 6.0 bylo provedeno genotypování celé rodiny. Analýza počtu kopií neidentifikovala žádnou delecii nebo amplifikaci větší než 10 kb, která by odpovídala předpokládanému modelu dědičnosti. Pomocí homozygotního mapování byly nalezeny dvě rozsáhlé homozygotní oblasti na chromozomu 2 a 4, obsahující 114 a 29 genů. Vzhledem k velikosti nalezených oblastí bylo provedeno exomové sekvenování celé rodiny. Analýza exomových dat identifikovala tři kandidátní mutace odpovídající autozomálně recesivnímu modelu dědičnosti, z nichž se pouze jedna nacházela v homozygotní oblasti. Nalezená homozygotní mutace v genu *ANTXR1* (c.C505>T; p.R169X) nebyla přítomna v žádné z populačních databází. Sekvenační analýzou genu *ANTXR1* dalšího dostupného pacienta byla nalezena mutace (c.C262>T; p.R88X) a rekurence byla dále potvrzena díky mezinárodní spolupráci analýzou dalších dvou rodin (c.C262>T; p.R88X a sestřihová mutace c.1435–12A>G). *ANTXR1* (dříve TEM8, tumor endothelial marker 8) je transmembránový glykoprotein typu I lokalizovaný na plazmatické membráně, který byl původně popsán jako nádorově specifický endoteliální marker, jehož exprese je zvýšená během procesu nádorové angiogeneze (St Croix B. et al., 2000). Krátce poté, byl nezávisle identifikován jako receptor pro toxin *Bacillus anthracis* (ATR) (Bradley K. A. et al., 2001). Mezi jeho funkce patří zprostředkování interakce buňky s komponentami extracelulární matrix (Hotchkiss K. A. et al., 2005), vazba ligandů k aktinovému cytoskeletu (Yang M. Y. et al., 2011) a regulace buněčné adheze (Werner E. et al., 2006). Funkční význam nalezených mutací byl potvrzen podstatně sníženým množstvím transkriptu, nepřítomností proteinu ve fibroblastech pacientů a barvení phalloidinem také prokázalo výrazné změny v síti aktinových vláken cytoskeletu fibroblastů.



## Část IV. Výsledky

Tato dizertační práce představuje možnosti využití nových genomických technik ve studiu genetické podstaty řady vzácných onemocnění a jejich úspěšnou aplikaci.

Předkládanými výsledky jsou:

1. Objasnění genetické podstaty Rotorova syndromu (*SLCO1B1* a *SLCO1B3*) a popsání nového mechanismu transportu bilirubinu v játrech s využitím vlastních oligonukleotidových čipů, homozygotního mapování a analýzy změn počtu kopií DNA (příloha 1a a 1b).
2. Identifikace genu odpovědného za izolovaný deficit ATP syntázy (*TMEM70*) pomocí vlastního čipu H-MitoArray, analýzy genové exprese, vazebné analýzy a homozygotního mapování (příloha 2a a 2b).
3. Identifikace genu podmiňujícího mukopolysacharidózu typu IIIC (*TMEM76*) s využitím vazebné analýzy a analýzy genové exprese na vlastních DNA čipech (příloha 3).
4. Objasnění genetické podstaty adultní formy autozomálně dominantní neuronální ceroidní lipofuscinózy (*DNAJC5*) s využitím kombinace vazebné analýzy, analýzy genové exprese, analýzy změn počtu kopií DNA a exomového sekvenování (příloha 4).
5. Objasnění genetické podstaty GAPO syndromu (*ANTXR1*) s využitím analýzy změn počtu kopií DNA, homozygotního mapování a exomového sekvenování (příloha 5).

## Závěr

Má práce významně přispěla k zavedení technologie masivně paralelního sekvenování, k rozvoji technologie DNA čipů, zavedení postupů bioinformatické analýzy a interpretace takto získaných dat v Ústavu dědičných metabolických poruch. Tyto metody a postupy jsou dnes základním přístupem pro studium molekulární podstaty geneticky podmíněných onemocnění a DNA diagnostiku známých onemocnění.

1. Zavedení těchto metod a postupů umožnilo jejich úspěšné využití ve více než 20 projektech a vedlo k objasnění genetické molekulární příčiny MPSIIIC (příloha 3), deficitu ATP syntázy (příloha 2a a 2b), Rotorova syndromu (příloha 1a a 1b), Kufsovy choroby (příloha 4), GAPO syndromu (příloha 5), X-vázané familiární kardiomyopatie (Hartmannova H. et al., 2013), poruchy glykosylace (Park E. J. et al., 2014) a dětské slepoty (Kmocho S. et al., 2015). S jejich využitím byl také studován mechanismus leukemogeneze (Takacova S. et al., 2012), hypercholesterolemie (Kolarova H. et al., 2014), příčina mitochondriálních onemocnění (Vondrackova A. et al., 2014), příčina dědičné hemochromatózy (Neroldova M. et al., 2015) a příčina deficitu OTC (Storkanova G. et al., 2013).

2. Tyto metody a postupy představují univerzální technologickou platformu pro studium genetických příčin nemocí v ČR. V současnosti jsou tyto postupy využívány ve spolupráci s IKEM, VFN, FN Motol a Fyziologickým ústavem AV v řadě dalších projektů – studiu vzácných nemocí, familiálních kardiomyopatií, rakoviny prsu, příčin statinové myopatie, genetické komponenty násilného chování, familiálních nefropatií, neurodegenerativních onemocnění a mitochondriálních onemocnění.

2. Identifikace kauzálních genů umožnila diagnostiku a prevenci onemocnění s využitím postupů prenatální a preimplantační diagnostiky.

3. Souhrnná data tvoří základ české populační databáze variant pro efektivní analýzu dat produkovaných masivně paralelním sekvenováním.

4. Získané zkušenosti umožnily vývoj kitů pro obohacení DNA o vybrané oblasti a zavedení metod DNA diagnostiky založené na NGS pro mitochondriální, metabolická a onkologická onemocnění.

## Část V. Přílohy

### Seznam příloh

#### Příloha 1a

Hřebíček, M., Jirásek, T., Hartmannová, H., Nosková, L., **Stránecký, V.**, Ivánek, R., Kmoč, S., Cebecauerová, D., Vítek, L., Mikulecký, M., Subhanová, I., Hozák, P., Jirsa, M. Rotor-type hyperbilirubinaemia has no defect in the canalicular bilirubin export pump, 2007, *Liver International*, 27 (4), pp. 485-491.

#### Příloha 1b

Van De Steeg, E\*, **Stránecký, V\***, Hartmannová, H., Nosková, L., Hřebíček, M., Wagenaar, E., Van Esch, A., De Waart, D.R., Oude Elferink, R.P.J., Kenworthy, K.E., Sticová, E., Al-Edreesi, M., Knisely, A.S., Kmoč, S., Jirsa, M., Schinkel, A.H. Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver, 2012, *Journal of Clinical Investigation*, 122 (2), pp. 519-528.

#### Příloha 2a

Čížková, A., **Stránecký, V.**, Ivánek, R., Hartmannová, H., Nosková, L., Piherová, L., Tesařová, M., Hansíková, H., Honzík, T., Zeman, J., Paul, J., Sperl, W., Mayr, J.A., Seneca, S., Houštěk, J., Kmoč, S. Development of a human mitochondrial oligonucleotide microarray (h-MitoArray) and gene expression analysis of fibroblast cell lines from 13 patients with isolated F1F0 ATP synthase deficiency, 2008, *BMC Genomics*, 9, art. no. 38.

#### Příloha 2b

Čížková, A., **Stránecký, V.**, Mayr, J.A., Tesařová, M., Havlíčková, V., Paul, J., Ivánek, R., Kuss, A.W., Hansíková, H., Kaplanová, V., Vrbacký, M., Hartmannová, H., Nosková, L., Honzík, T., Drahot, Z., Magner, M., Hejzlarová, K., Sperl, W., Zeman, J., Houštěk, J., Kmoč, S. TMEM70 mutations cause isolated ATP synthase deficiency and neonatal mitochondrial encephalomyopathy, 2008, *Nature Genetics*, 40 (11), pp. 1288-1290.

#### Příloha 3

Hřebíček, M., Mrázová, L., Seyrantepe, V., Durand, S., Roslin, N.M., Nosková, L., Hartmannová, H., Ivánek, R., Čížková, A., Poupětová, H., Sikora, J., Uřínová, J., **Stránecký, V.**, Zeman, J., Lepage, P., Roquis, D., Verner, A., Ausseil, J., Beesley, C.E., Maire, I., Poorthuis, B.J.H.M., Van De Kamp, J., Van Diggelen, O.P., Wevers, R.A., Hudson, T.J., Fujiwara, T.M., Majewski, J., Morgan, K., Kmoč, S., Pshezhetsky, A.V. Mutations in TMEM76\* cause mucopolysaccharidosis IIIC (Sanfilippo C syndrome), 2006, *American Journal of Human Genetics*, 79 (5), pp.807-819.

#### Příloha 4

Nosková, L\*, **Stránecký, V\***, Hartmannová, H., Přistoupilová, A., Barešová, V., Ivánek, R., Hlčková, H., Jahnová, H., Van Der Zee, J., Staropoli, J.F., Sims, K.B., Tyneš, J., Van Broeckhoven, C., Nijssen, P.C.G., Mole, S.E., Elleder, M., Kmoč, S. Mutations in DNAJC5, encoding cysteine-string protein alpha, cause autosomal-dominant adult-onset neuronal ceroid lipofuscinosis, 2011, *American Journal of Human Genetics*, 89 (2), pp. 241-252.

#### Příloha 5

**Stránecký, V.**, Hoischen, A., Hartmannová, H., Zaki, M.S., Chaudhary, A., Zudaire, E., Nosková, L., Barešová, V., Přistoupilová, A., Hodaňová, K., Sovová, J., Hůlková, H., Piherová, L., Hehir-Kwa, J.Y., De Silva, D., Senanayake, M.P., Farrag, S., Zeman, J., Martásek, P., Baxová, A., Afifi, H.H., St. Croix, B., Brunner, H.G., Temtamy, S., Kmoč, S. Mutations in ANTXR1 cause GAPO syndrome, 2013, *American Journal of Human Genetics*, 92 (5), pp. 792-799.

## Literatura

- A. B. Rotor L. M., A. Florentin, 1948, *Familial non-hemolytic jaundice with direct van den Bergh reaction*. Acta medica Philippina. **5**: p. 37-49.
- Adzhubei I. A., Schmidt S., Peshkin L., et al., 2010, *A method and server for predicting damaging missense mutations*. Nat Methods. **7**(4): p. 248-249.
- Albert T. J., Molla M. N., Muzny D. M., et al., 2007, *Direct selection of human genomic loci by microarray hybridization*. Nat Methods. **4**(11): p. 903-905.
- Ansorge W., Sproat B., Stegemann J., et al., 1987, *Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis*. Nucleic Acids Res. **15**(11): p. 4593-4602.
- Antonarakis S. E. and Beckmann J. S., 2006, *Mendelian disorders deserve more attention*. Nat Rev Genet. **7**(4): p. 277-282.
- Bainbridge M. N., Wang M., Burgess D. L., et al., 2010, *Whole exome capture in solution with 3 Gbp of data*. Genome Biol. **11**(6): p. R62.
- Bamshad M. J., Ng S. B., Bigham A. W., et al., 2011, *Exome sequencing as a tool for Mendelian disease gene discovery*. Nat Rev Genet. **12**(11): p. 745-755.
- Baxova A., Kozlowski K., Obersztyn E., et al., 1997, *GAP0 syndrome (Radiographic clues to early diagnosis)*. Radiol Med. **93**(3): p. 289-291.
- Bell G. I., Horita S., and Karam J. H., 1984, *A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus*. Diabetes. **33**(2): p. 176-183.
- Bolstad B. M., Irizarry R. A., Astrand M., et al., 2003, *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias*. Bioinformatics. **19**(2): p. 185-193.
- Bradley K. A., Mogridge J., Mourez M., et al., 2001, *Identification of the cellular receptor for anthrax toxin*. Nature. **414**(6860): p. 225-229.
- Cardon L. R. and Bell J. I., 2001, *Association study designs for complex diseases*. Nat Rev Genet. **2**(2): p. 91-99.
- Cardon L. R. and Palmer L. J., 2003, *Population stratification and spurious allelic association*. Lancet. **361**(9357): p. 598-604.
- Clark M. J., Chen R., Lam H. Y., et al., 2011, *Performance comparison of exome DNA sequencing technologies*. Nat Biotechnol. **29**(10): p. 908-914.
- Cohen S. N., Chang A. C., Boyer H. W., et al., 1973, *Construction of biologically functional bacterial plasmids in vitro*. Proc Natl Acad Sci U S A. **70**(11): p. 3240-3244.
- Collinson I. R., Skehel J. M., Fearnley I. M., et al., 1996, *The F1F0-ATPase complex from bovine heart mitochondria: the molar ratio of the subunits in the stalk region linking the F1 and F0 domains*. Biochemistry. **35**(38): p. 12640-12646.

- Consortium I. H. G. S., 2001, *Initial sequencing and analysis of the human genome*. Nature. **409**(6822): p. 860-921.
- Cooper G. M., Stone E. A., Asimenos G., et al., 2005, *Distribution and intensity of constraint in mammalian genomic sequence*. Genome Res. **15**(7): p. 901-913.
- Danecek P., Auton A., Abecasis G., et al., 2011, *The variant call format and VCFtools*. Bioinformatics. **27**(15): p. 2156-2158.
- Dressman D., Yan H., Traverso G., et al., 2003, *Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations*. Proc. Natl. Acad. Sci. USA. **100**: p. 8817-8822.
- Ewens W. J. and Spielman R. S., 1995, *The transmission/disequilibrium test: history, subdivision, and admixture*. Am J Hum Genet. **57**(2): p. 455-464.
- Fedurco M., Romieu A., Williams S., et al., 2006, *BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies*. Nucleic Acids Res. **34**(3): p. e22.
- Firth H. V., Richards S. M., Bevan A. P., et al., 2009, *DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources*. Am J Hum Genet. **84**(4): p. 524-533.
- Fretzayas A. M., Garoufi A. I., Moutsouris C. X., et al., 1994, *Cholescintigraphy in the diagnosis of Rotor syndrome*. J Nucl Med. **35**(6): p. 1048-1050.
- Garrod A. E., 1902, *About Alkaptonuria*. Med Chir Trans. **85**: p. 69-78.
- Genomes Project C., Abecasis G. R., Auton A., et al., 2012, *An integrated map of genetic variation from 1,092 human genomes*. Nature. **491**(7422): p. 56-65.
- Gentleman R. C., Carey V. J., Bates D. M., et al., 2004, *Bioconductor: open software development for computational biology and bioinformatics*. Genome Biol. **5**(10): p. R80.
- Gilissen C., Hoischen A., Brunner H. G., et al., 2012, *Disease gene identification strategies for exome sequencing*. Eur J Hum Genet. **20**(5): p. 490-497.
- Gusella J. F., Wexler N. S., Conneally P. M., et al., 1983, *A polymorphic DNA marker genetically linked to Huntington's disease*. Nature. **306**(5940): p. 234-238.
- Hagenbuch B. and Gui C., 2008, *Xenobiotic transporters of the human organic anion transporting polypeptides (OATP) family*. Xenobiotica. **38**(7-8): p. 778-801.
- Hall J. M., Lee M. K., Newman B., et al., 1990, *Linkage of early-onset familial breast cancer to chromosome 17q21*. Science. **250**(4988): p. 1684-1689.
- Haraksingh R. R., Abyzov A., Gerstein M., et al., 2011, *Genome-wide mapping of copy number variation in humans: comparative analysis of high resolution array platforms*. PLoS One. **6**(11): p. e27859.
- Hartmannova H., Kubanek M., Sramko M., et al., 2013, *Isolated X-Linked Hypertrophic Cardiomyopathy Caused by a Novel Mutation of the Four-and-a-Half LIM Domain 1 Gene*. Circ Cardiovasc Genet.

- Hodges E., Xuan Z., Balija V., et al., 2007, *Genome-wide in situ exon capture for selective resequencing*. *Nat Genet.* **39**(12): p. 1522-1527.
- Hotchkiss K. A., Basile C. M., Spring S. C., et al., 2005, *TEM8 expression stimulates endothelial cell adhesion and migration by regulating cell-matrix interactions on collagen*. *Exp Cell Res.* **305**(1): p. 133-144.
- Houstek J., Klement P., Floryk D., et al., 1999, *A novel deficiency of mitochondrial ATPase of nuclear origin*. *Hum Mol Genet.* **8**(11): p. 1967-1974.
- Hsueh H. M., Chen J. J., and Kodell R. L., 2003, *Comparison of methods for estimating the number of true null hypotheses in multiplicity testing*. *J Biopharm Stat.* **13**(4): p. 675-689.
- Huang da W., Sherman B. T., and Lempicki R. A., 2009, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists*. *Nucleic Acids Res.* **37**(1): p. 1-13.
- Hughes T. R., Mao M., Jones A. R., et al., 2001, *Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer*. *Nat Biotechnol.* **19**(4): p. 342-347.
- Chee M., Yang R., Hubbell E., et al., 1996, *Accessing genetic information with high-density DNA arrays*. *Science.* **274**(5287): p. 610-614.
- Cheung C. Y., Marchani Blue E., and Wijsman E. M., 2014, *A statistical framework to guide sequencing choices in pedigrees*. *Am J Hum Genet.* **94**(2): p. 257-267.
- lafrate A. J., Feuk L., Rivera M. N., et al., 2004, *Detection of large-scale variation in the human genome*. *Nat Genet.* **36**(9): p. 949-951.
- International HapMap C., 2005, *A haplotype map of the human genome*. *Nature.* **437**(7063): p. 1299-1320.
- International Human Genome Sequencing C., 2004, *Finishing the euchromatic sequence of the human genome*. *Nature.* **431**(7011): p. 931-945.
- Jackson D. A., Symons R. H., and Berg P., 1972, *Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli*. *Proc Natl Acad Sci U S A.* **69**(10): p. 2904-2909.
- Jacob F. and Monod J., 1961, *Genetic regulatory mechanisms in the synthesis of proteins*. *J Mol Biol.* **3**: p. 318-356.
- Kan Y. W. and Dozy A. M., 1978, *Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation*. *Proc Natl Acad Sci U S A.* **75**(11): p. 5631-5635.
- Kircher M., Witten D. M., Jain P., et al., 2014, *A general framework for estimating the relative pathogenicity of human genetic variants*. *Nat Genet.* **46**(3): p. 310-315.
- Klein R. J., Zeiss C., Chew E. Y., et al., 2005, *Complement factor H polymorphism in age-related macular degeneration*. *Science.* **308**(5720): p. 385-389.

- Klein U., Kresse H., and von Figura K., 1978, *Sanfilippo syndrome type C: deficiency of acetyl-CoA:alpha-glucosaminide N-acetyltransferase in skin fibroblasts*. Proc Natl Acad Sci U S A. **75**(10): p. 5185-5189.
- Kmoch S., Majewski J., Ramamurthy V., et al., 2015, *Mutations in PNPLA6 are linked to photoreceptor degeneration and various forms of childhood blindness*. Nat Commun. **6**: p. 5614.
- Koenig M., Hoffman E. P., Bertelson C. J., et al., 1987, *Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals*. Cell. **50**(3): p. 509-517.
- Kolarova H., Tesarova M., Svecova S., et al., 2014, *Lipoprotein Lipase Deficiency: Clinical, Biochemical and Molecular Characteristics in Three Patients with Novel Mutations in the LPL Gene*. Folia Biol (Praha). **60**(5): p. 235-243.
- Koschmieder A., Zimmermann K., Trissl S., et al., 2012, *Tools for managing and analyzing microarray data*. Brief Bioinform. **13**(1): p. 46-60.
- Lander E. S. and Botstein D., 1987, *Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children*. Science. **236**(4808): p. 1567-1570.
- Landrum M. J., Lee J. M., Riley G. R., et al., 2014, *ClinVar: public archive of relationships among sequence variation and human phenotype*. Nucleic Acids Res. **42**(Database issue): p. D980-985.
- Levy S., Sutton G., Ng P. C., et al., 2007, *The diploid genome sequence of an individual human*. PLoS Biol. **5**(10): p. e254.
- Li H. and Durbin R., 2009, *Fast and accurate short read alignment with Burrows-Wheeler transform*. Bioinformatics. **25**(14): p. 1754-1760.
- Li H., Handsaker B., Wysoker A., et al., 2009, *The Sequence Alignment/Map format and SAMtools*. Bioinformatics. **25**(16): p. 2078-2079.
- MacDonald J. R., Ziman R., Yuen R. K., et al., 2014, *The Database of Genomic Variants: a curated collection of structural variation in the human genome*. Nucleic Acids Res. **42**(Database issue): p. D986-992.
- Malhotra D. and Sebat J., 2012, *CNVs: harbingers of a rare variant revolution in psychiatric genetics*. Cell. **148**(6): p. 1223-1241.
- Mamanova L., Coffey A. J., Scott C. E., et al., 2010, *Target-enrichment strategies for next-generation sequencing*. Nat Methods. **7**(2): p. 111-118.
- Manolio T. A., Collins F. S., Cox N. J., et al., 2009, *Finding the missing heritability of complex diseases*. Nature. **461**(7265): p. 747-753.
- Margulies M., Egholm M., Altman W. E., et al., 2005, *Genome sequencing in microfabricated high-density picolitre reactors*. Nature. **437**(7057): p. 376-380.
- Matsuzaki H., Loi H., Dong S., et al., 2004, *Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array*. Genome Res. **14**(3): p. 414-425.

- McClellan J. and King M. C., 2010, *Genetic heterogeneity in human disease*. Cell. **141**(2): p. 210-217.
- McKenna A., Hanna M., Banks E., et al., 2010, *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data*. Genome Res. **20**(9): p. 1297-1303.
- McKusick V. A., 2007, *Mendelian Inheritance in Man and its online version, OMIM*. Am J Hum Genet. **80**(4): p. 588-604.
- Metzker M. L., 2010, *APPLICATIONS OF NEXT-GENERATION SEQUENCING Sequencing technologies - the next generation*. Nature Reviews Genetics. **11**(1): p. 31-46.
- Michael K. L., Taylor L. C., Schultz S. L., et al., 1998, *Randomly ordered addressable high-density optical sensor arrays*. Anal Chem. **70**(7): p. 1242-1248.
- Miki Y., Swensen J., Shattuck-Eidens D., et al., 1994, *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1*. Science. **266**(5182): p. 66-71.
- Mills R. E., Walter K., Stewart C., et al., 2011, *Mapping copy number variation by population-scale genome sequencing*. Nature. **470**(7332): p. 59-65.
- Mole S. E., Williams R. E., and Goebel H. H., 2011, *The neuronal ceroid lipofuscinoses (Batten disease)*. 2nd ed, Oxford: Oxford University Press. xxx, 444 p.
- Morgan T. H., 1910, *Sex Limited Inheritance in Drosophila*. Science. **32**(812): p. 120-122.
- Morgan T. H., 1911, *Chromosomes and Associative Inheritance*. Science. **34**(880): p. 636-638.
- Morgan T. H., 1911, *The Origin of Five Mutations in Eye Color in Drosophila and Their Modes of Inheritance*. Science. **33**(849): p. 534-537.
- Morton N. E., 1955, *Sequential tests for the detection of linkage*. Am J Hum Genet. **7**(3): p. 277-318.
- Mullis K. B. and Faloona F. A., 1987, *Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction*. Methods Enzymol. **155**: p. 335-350.
- Neroldova M., Frankova S., Stranecky V., et al., 2015, *Hereditary haemochromatosis caused by homozygous HJV mutation evolved through paternal disomy*. Clin Genet. **87**(1): p. 96-98.
- Ng P. C. and Henikoff S., 2001, *Predicting deleterious amino acid substitutions*. Genome Res. **11**(5): p. 863-874.
- Ng S. B., Turner E. H., Robertson P. D., et al., 2009, *Targeted capture and massively parallel sequencing of 12 human exomes*. Nature. **461**(7261): p. 272-276.
- Ng S. B., Bigham A. W., Buckingham K. J., et al., 2010, *Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome*. Nat Genet. **42**(9): p. 790-793.
- Ng S. B., Nickerson D. A., Bamshad M. J., et al., 2010, *Massively parallel sequencing and rare disease*. Hum Mol Genet. **19**(R2): p. R119-124.
- Niemi M., Pasanen M. K., and Neuvonen P. J., 2011, *Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake*. Pharmacol Rev. **63**(1): p. 157-181.



- Nirenberg M. W. and Matthaei J. H., 1961, *The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides*. Proc Natl Acad Sci U S A. **47**: p. 1588-1602.
- O'Rawe J., Jiang T., Sun G., et al., 2013, *Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing*. Genome Med. **5**(3): p. 28.
- Park E. J., Grabinska K. A., Guan Z., et al., 2014, *Mutation of Nogo-B receptor, a subunit of cis-prenyltransferase, causes a congenital disorder of glycosylation*. Cell Metab. **20**(3): p. 448-457.
- Paulusma C. C., Kool M., Bosma P. J., et al., 1997, *A mutation in the human canalicular multispecific organic anion transporter gene causes the Dubin-Johnson syndrome*. Hepatology. **25**(6): p. 1539-1542.
- Pease A. C., Solas D., Sullivan E. J., et al., 1994, *Light-generated oligonucleotide arrays for rapid DNA sequence analysis*. Proc Natl Acad Sci U S A. **91**(11): p. 5022-5026.
- Pinto D., Delaby E., Merico D., et al., 2014, *Convergence of genes and cellular pathways dysregulated in autism spectrum disorders*. Am J Hum Genet. **94**(5): p. 677-694.
- Pollard K. S., Hubisz M. J., Rosenbloom K. R., et al., 2010, *Detection of nonneutral substitution rates on mammalian phylogenies*. Genome Res. **20**(1): p. 110-121.
- Porreca G. J., 2007, *Multiplex amplification of large sets of human exons*. Nat. Methods. **4**: p. 931-936.
- Remuzzi G. and Garattini S., 2008, *Rare diseases: what's next?* Lancet. **371**(9629): p. 1978-1979.
- Rohn J., 2013, *Billions spent on rare diseases*. Nat Biotech. **31**(5): p. 368-368.
- Ronaghi M., Karamohamed S., Pettersson B., et al., 1996, *Real-time DNA sequencing using detection of pyrophosphate release*. Anal Biochem. **242**(1): p. 84-89.
- Rosner F., 1995, *Medicine in the Bible and the Talmud : selections from classical Jewish sources*. Augm. ed. The library of Jewish law and ethics, Yeshiva University Press, New York.
- Sanger F. and Coulson A. R., 1975, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase*. Journal of Molecular Biology. **94**(3): p. 441-448.
- Sanger F., Nicklen S., and Coulson A. R., 1977, *DNA sequencing with chain-terminating inhibitors*. Proc Natl Acad Sci U S A. **74**(12): p. 5463-5467.
- Shendure J., Porreca G. J., Reppas N. B., et al., 2005, *Accurate multiplex polony sequencing of an evolved bacterial genome*. Science. **309**(5741): p. 1728-1732.
- Schena M., Shalon D., Davis R. W., et al., 1995, *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. Science. **270**(5235): p. 467-470.
- Schmitz F., Tabares L., Khimich D., et al., 2006, *CSPalpha-deficiency causes massive and rapid photoreceptor degeneration*. Proc Natl Acad Sci U S A. **103**(8): p. 2926-2931.
- Siva N., 2008, *1000 Genomes project*. Nat Biotechnol. **26**(3): p. 256.

- Smith H. O. and Wilcox K. W., 1970, *A restriction enzyme from Hemophilus influenzae. I. Purification and general properties*. J Mol Biol. **51**(2): p. 379-391.
- Smith L. M., Sanders J. Z., Kaiser R. J., et al., 1986, *Fluorescence detection in automated DNA sequence analysis*. Nature. **321**(6071): p. 674-679.
- Smyth G. K. and Speed T., 2003, *Normalization of cDNA microarray data*. Methods. **31**(4): p. 265-273.
- Solinas-Toldo S., Lampel S., Stilgenbauer S., et al., 1997, *Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances*. Genes Chromosomes Cancer. **20**(4): p. 399-407.
- Sperl W., Jesina P., Zeman J., et al., 2006, *Deficiency of mitochondrial ATP synthase of nuclear genetic origin*. Neuromuscul Disord. **16**(12): p. 821-829.
- St Croix B., Rago C., Velculescu V., et al., 2000, *Genes expressed in human tumor endothelium*. Science. **289**(5482): p. 1197-1202.
- Stenson P. D., Ball E. V., Howells K., et al., 2009, *The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics*. Hum Genomics. **4**(2): p. 69-72.
- Storkanova G., Vlaskova H., Chuzhanova N., et al., 2013, *Ornithine carbamoyltransferase deficiency: molecular characterization of 29 families*. Clin Genet. **84**(6): p. 552-559.
- Takacova S., Slany R., Bartkova J., et al., 2012, *DNA damage response and inflammatory signaling limit the MLL-ENL-induced leukemogenesis in vivo*. Cancer Cell. **21**(4): p. 517-531.
- Tewhey R., 2009, *Microdroplet-based PCR enrichment for large-scale targeted sequencing*. Nat. Biotechnol. **27**: p. 1025-1031.
- Tipton R. E. and Gorlin R. J., 1984, *Growth retardation, alopecia, pseudo-anodontia, and optic atrophy--the GAPO syndrome: report of a patient and review of the literature*. Am J Med Genet. **19**(2): p. 209-216.
- Tobaben S., Thakur P., Fernandez-Chacon R., et al., 2001, *A trimeric protein complex functions as a synaptic chaperone machine*. Neuron. **31**(6): p. 987-999.
- Tsui L. C., Buchwald M., Barker D., et al., 1985, *Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker*. Science. **230**(4729): p. 1054-1057.
- Turner E. H., Lee C., Ng S. B., et al., 2009, *Massively parallel exon capture and library-free resequencing across 16 genomes*. Nat Methods. **6**(5): p. 315-316.
- Van der Auwera G. A., Carneiro M. O., Hartl C., et al., 2013, *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline*. Curr Protoc Bioinformatics. **11**(1110): p. 11 10 11-11 10 33.
- Venter J. C., Adams M. D., Myers E. W., et al., 2001, *The sequence of the human genome*. Science. **291**(5507): p. 1304-1351.

- Vondrackova A., Vesela K., Kratochvilova H., et al., 2014, *Large copy number variations in combination with point mutations in the TYMP and SCO2 genes found in two patients with mitochondrial disorders*. Eur J Hum Genet. **22**(3): p. 431-434.
- Wang K., Li M., and Hakonarson H., 2010, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data*. Nucleic Acids Res. **38**(16): p. e164.
- Watson J. D. and Crick F. H., 1953, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. Nature. **171**(4356): p. 737-738.
- Wellcome Trust Case Control C., Craddock N., Hurles M. E., et al., 2010, *Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls*. Nature. **464**(7289): p. 713-720.
- Welter D., MacArthur J., Morales J., et al., 2014, *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations*. Nucleic Acids Res. **42**(Database issue): p. D1001-1006.
- Werner E., Kowalczyk A. P., and Faundez V., 2006, *Anthrax toxin receptor 1/tumor endothelium marker 8 mediates cell spreading by coupling extracellular ligands to the actin cytoskeleton*. J Biol Chem. **281**(32): p. 23227-23236.
- Wettenhall J. M. and Smyth G. K., 2004, *limmaGUI: a graphical user interface for linear modeling of microarray data*. Bioinformatics. **20**(18): p. 3705-3706.
- Wheeler D. A., Srinivasan M., Egholm M., et al., 2008, *The complete genome of an individual by massively parallel DNA sequencing*. Nature. **452**(7189): p. 872-876.
- Yang M. Y., Chaudhary A., Seaman S., et al., 2011, *The cell surface structure of tumor endothelial marker 8 (TEM8) is regulated by the actin cytoskeleton*. Biochim Biophys Acta. **1813**(1): p. 39-49.
- Zarrei M., MacDonald J. R., Merico D., et al., 2015, *A copy number variation map of the human genome*. Nat Rev Genet. **16**(3): p. 172-183.
- Zinsmaier K. E., Eberle K. K., Buchner E., et al., 1994, *Paralysis and early death in cysteine string protein mutants of Drosophila*. Science. **263**(5149): p. 977-980.

## **Kopie publikovaných prací**

## CLINICAL STUDIES

## Rotor-type hyperbilirubinaemia has no defect in the canalicular bilirubin export pump

Martin Hřebíček<sup>1</sup>, Tomáš Jirásek<sup>2</sup>, Hana Hartmannová<sup>1</sup>, Lenka Nosková<sup>1</sup>, Viktor Stránecký<sup>1,3</sup>, Robert Ivánek<sup>4,1</sup>, Stanislav Kmoch<sup>1,3</sup>, Dita Cebecauerová<sup>5</sup>, Libor Vitek<sup>6</sup>, Miroslav Mikulecký<sup>7</sup>, Iva Subhanová<sup>6</sup>, Pavel Hozák<sup>4</sup> and Milan Jirsa<sup>5,6</sup>

1 Institute of Inherited Metabolic Diseases, Charles University 1st Faculty of Medicine, Prague, Czech Republic

2 Department of Pathology, Charles University 3rd Faculty of Medicine, Prague, Czech Republic

3 Center for Applied Genomics, Charles University 1st Faculty of Medicine, Prague, Czech Republic

4 Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic

5 Institute for Clinical and Experimental Medicine, Prague, Czech Republic

6 Institute of Clinical Biochemistry and Laboratory Diagnostics, Charles University 1st Faculty of Medicine, Prague, Czech Republic

7 1<sup>st</sup> Medical Clinic, Teaching Hospital, Comenius University, Bratislava, Slovak Republic

### Keywords

ABCC2 – Dubin–Johnson syndrome – hereditary jaundice – MRP2 – Rotor syndrome – UGT1A1

### Correspondence

Milan Jirsa, MD, PhD, Laboratory of Experimental Hepatology, Institute for Clinical and Experimental Medicine, Vídeňská 1958/9, 140 21 Praha 4–Křč, Czech Republic.  
Tel: +420 261 362 773  
Fax: +420 241 721 666  
e-mail: milan.jirsa@medicon.cz

Received 28 August 2006

accepted 22 December 2006

DOI:10.1111/j.1478-3231.2007.01446.x

### Abstract

**Background:** The cause of Rotor syndrome (RS), a rare-familial conjugated hyperbilirubinaemia with normal liver histology, is unclear. We hypothesized that RS can be an allelic variant of Dubin–Johnson syndrome, caused by mutation in *ABCC2*, and investigated *ABCC2* (gene) and ABCC2 (protein) in two patients with RS. **Methods:** A 57-year-old male presented with a 5-year history of predominantly conjugated hyperbilirubinaemia (170 µmol/l). Urinary porphyrin excretion was increased; cholescintigraphy revealed no chromoexcretion. A 68-year-old male presented with lifelong conjugated hyperbilirubinaemia (85 µmol/l). Bromosulphophthalein elimination was typical for RS. Both patients had histologically normal liver, without pigment. *ABCC2* expression was investigated by confocal fluorescence microscopy. *ABCC2* was sequenced from genomic DNA and cDNA, and exon deletions/duplications were sought by comparative genomic hybridization on a custom micro-array. **Results:** In both patients, *ABCC2* was expressed unremarkably at the apical membrane of hepatocytes and no sequence alterations were found in 32 exons, adjacent intronic regions and the promoter region of *ABCC2*. **Conclusions:** Rotor-type hyperbilirubinaemia is not an allelic variant of *ABCC2* deficiency.

The bilirubin excretory pathway consists of two steps: conjugation of unconjugated bilirubin with glucuronic acid, catalyzed by uridine diphosphate-glucuronosyl transferase 1A1 (UGT1A1), and secretion of conjugated bilirubin into bile via ABCC2, the canalicular bilirubin export pump. Mutations in *ABCC2*, encoding ABCC2, are known to cause Dubin–Johnson syndrome (DJS, OMIM No. 237500) (1), a rare benign predominantly conjugated hyperbilirubinaemia with typical deposits of melanin-like pigment within hepatocyte lysosomes. Rotor syndrome (RS, OMIM No. 237450) represents another form of hereditary jaundice with predominantly conjugated hyperbilirubinaemia. Unlike most patients with DJS, patients with RS have no abnormal hepatic pigmentation. Total porphyrin excretion in urine is increased and the ratio of coproporphyrin isomers I:III is lower than

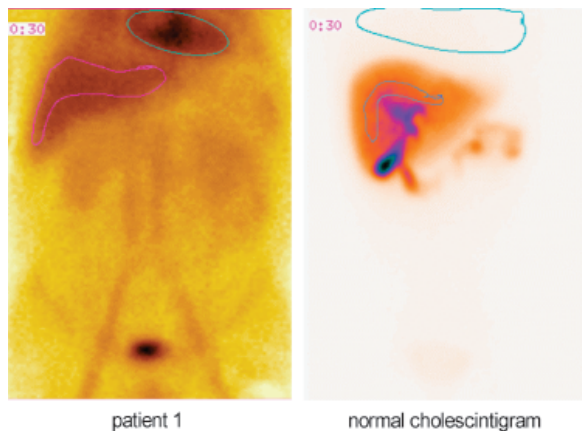
in DJS (2). Unlike patients with DJS, patients with RS exhibit marked retention of bromosulphophthalein (BSP) after injection (3). Neither the liver nor the biliary tree is visualized by cholescintigraphy in RS (4, 5). The molecular basis of RS is unknown.

The definition of RS as a pathophysiological entity distinct from DJS is based on the differences in BSP clearance and total urinary coproporphyrin output. Other features of both disorders such as liver pigmentation and visualization of the gallbladder by cholescintigraphy are less specific (6), making the diagnosis of RS difficult to establish. We hypothesized that the phenotypic differences between RS and DJS do not exclude the possibility that RS and DJS are allelic variants. In our study, we investigated the potential role of *ABCC2* as a candidate gene responsible for Rotor-type hyperbilirubinaemia in two affected subjects.

## Patients and methods

### Case 1

A 57-year-old male had scleral icterus since birth. Generalized jaundice appeared for the first time at age 7 years and was diagnosed as posthepatic. At age 18 years, he was judged unfit for compulsory military service because of jaundice. He was hospitalized, aged 52 years, owing to an acute occlusion of the central retinal artery. His total serum bilirubin concentration was 170  $\mu\text{mol/l}$ , with a direct-reacting fraction of 120  $\mu\text{mol/l}$ . Blood counts were within the normal ranges. Serum concentrations of aspartate aminotransferase (AST), alanine transaminase (ALT) and  $\gamma$ -glutamyl transferase activities and of  $\alpha_1$ -antitrypsin, copper, ceruloplasmin and bile salts were normal. No serologic evidence of infection with hepatitis A, B and C viruses or with other hepatotropic viruses was found. Urinary coproporphyrin output ranged between 80 and 500  $\mu\text{g}/24\text{ h}$  (normal < 200  $\mu\text{g}/24\text{ h}$ ); isomer I represented 57% of total coproporphyrin. Except for multiple gallstones, which were asymptomatic, no pathological changes of the liver and biliary tree were observed on ultrasonography and endoscopic retrograde cholangiopancreatography (ERCP). In contrast,  $^{99\text{m}}\text{Tc}$ -HIDA cholescintigraphy revealed no uptake of the radionuclide by the liver, and the bile ducts and gallbladder were not visualized (Fig. 1). No abnormality was found on microscopy of a percutaneous needle liver biopsy specimen. Jaundice has persisted, with biochemical documentation, for the subsequent 5 years, with serum concentrations of total and direct bilirubin oscillating around 170 and 120  $\mu\text{mol/l}$  respectively.



**Fig. 1.** Cholescintigraphy of patient 1 (left). In contrast to the normal situation (right), the radionuclide was retained in the circulation and no uptake of  $^{99\text{m}}\text{Tc}$ -HIDA by the liver and visualization of bile ducts and gallbladder was observed.

The family history of the patient was of potential interest because his daughter and one of his four sisters had Gilbert syndrome.

### Case 2

Our second patient (7), a 68-year-old male with no family history of liver disease, had scleral jaundice since childhood. Clinical and laboratory investigations on military conscription aged 19 years revealed predominantly direct hyperbilirubinaemia with no other clinical-biochemistry evidence of hepatobiliary injury and without serologic evidence of viral hepatitis. One year later, the patient was re-examined in the Central Military Hospital (Bratislava, Slovak Republic) for fatigue, lack of appetite, weight loss, abdominal pain, dark urine and jaundice. Aside from jaundice, chronic tonsillitis and skin rash, the physical examination was unremarkable. Total serum bilirubin concentrations ranged from 41 to 121  $\mu\text{mol/l}$ ; direct-reacting bilirubin represented 53–72%. Total urinary coproporphyrin output was repeatedly increased, but coproporphyrin isomers were not quantitatively differentiated. Other laboratory values were within the normal ranges. BSP retention 30 min after administration of 2 mg/kg was 77.5% (normal < 10%). The gallbladder was repeatedly not visualized by oral cholecystography. Exploratory laparotomy revealed a normal appearance of the liver, bile ducts and gallbladder. Normal architecture without any pigment or signs of cholestasis was found on microscopy of a biopsy specimen of liver. RS was diagnosed (7). During the subsequent 48 years, the patient's liver was assessed twice at surgery as normal in contour and color (ureterotomy for ureterolithiasis, age 44 years; cholecystectomy for gallstones, age 56 years). Persistent jaundice with predominantly conjugated hyperbilirubinaemia (total serum bilirubin ranging between 42 and 170  $\mu\text{mol/l}$  with seasonal fluctuations) has been repeatedly documented.

### Histology

Liver from Patient 1 was fixed in formalin; liver from Patient 2 was fixed in formalin or in Carnoy's fixative. All specimens were embedded in paraffin. Sections were stained with haematoxylin/eosin and with periodic acid–Schiff technique.

### Antibodies

Mouse anti-ABCC2 monoclonal antibody (clone M<sub>2</sub>III-6) was purchased from Kamiya (Seattle, WA, USA). Rabbit polyclonal anti-human carcinoembryonic antigen antibody, which recognizes

carcinoembryonic antigen-related adhesion molecule 1 (CEACAM1) on bile canaliculi of the human liver, was purchased from DAKO (Glostrup, Denmark) together with the EnVision Peroxidase Kit and LSAB+Kit. Fluorescein isothiocyanate-conjugated donkey anti-rabbit antibody and Cy5-conjugated goat anti-mouse antibody were obtained from Jackson (West Grove, PA, USA).

### Confocal laser scanning microscopy (CLSM)

For immunohistochemical procedures, 5 µm sections cut from formalin-fixed (Patient 1) and Carnoy's-fixed (Patient 2), paraffin-embedded tissue samples were deparaffinized and treated with 2.7% hydrogen peroxide and 0.1% sodium azide. Double immunolabeling was performed as described (8). The slides were observed in a CLSM Leica TCS SP (Leica Microsystems, Wetzlar, Germany). Simultaneous excitation with an argon-krypton laser (wavelength 488 nm) for fluorescein isothiocyanate and a helium-neon laser (wavelength 633 nm) for Cy5 was used. Sections incubated without primary antibodies were used as negative controls. Sections of a liver-biopsy specimen assessed as exhibiting minimal changes, obtained from an adult patient, were used as positive controls. Sections of a liver-biopsy specimen from a patient with proven ABCC2 deficiency (8) served as a negative control.

### Mutation analysis

Written informed consents were obtained from the patients and family members before genetic investigation and skin biopsy. *ABCC2* was analyzed by direct sequencing of polymerase chain reaction (PCR) products amplified from genomic DNA extracted from peripheral blood leukocytes. All exons and the 1500-bp-long promoter region were amplified by PCR using reported intronic oligonucleotide primers (8). All amplicons were gel-purified, extracted with QIA quick spin columns (Qiagen, Hilden, Germany) and sequenced on an automated fluorescent DNA sequencer (AlfExpress, Amersham-Pharmacia, Uppsala, Sweden).

Analysis of copy number changes caused by exon deletions or duplications was performed by comparative hybridization of the patient's and control male genomic DNAs. As probes, the micro-array contained PCR amplified products representing each *ABCC2* exon and 5'-aminomethylated 40-mer oligonucleotides corresponding to specific regions of *ABCC2* exons. Oligonucleotide sequences were designed using Oligo-picker software and purchased from Illumina (San Diego, CA, USA). The PCR products (100 ng/µl) and oligonucleotides (20 µM) in 3 × SSC were printed in

triplicate on aminosilane-modified slides and immobilized by standard techniques that combined baking and UV cross-linking. The slides were pretreated by baking at 80 °C; after UV cross-linkage, they were then washed twice in 0.1% sodium dodecyl sulfate (SDS) for 2 min, twice in 0.2 × SSC for 2 min and four times in MilliQ water, followed by denaturation in boiling water for 2 min. Prehybridization was performed in prehybridization buffer (6 × SSC, 0.5% SDS, 1% bovine serum albumin). Genomic DNA was extracted by the phenol/chloroform method from peripheral blood leukocytes, fragmented with *MboI* restriction endonuclease (New England Biolabs, Ipswich, MA, USA) and labelled using Cy3-AP3-dUTP or Cy5-AP3-dUTP (Amersham Biosciences, Piscataway, NJ, USA). Patient samples, control samples and 5 µg of human Cot-1 (Invitrogen, Carlsbad, CA, USA) were combined and dissolved in hybridization buffer (50% formamide, 6 × SSC, 0.5% SDS, 5 × Denhardt's). Hybridizations were performed at 37 °C in an ArrayIt Hybridization Cassette chamber (TeleChem International, Sunnyvale, CA, USA). Patient and control samples were analyzed in a dye swap mode with two replicates of each mode. The hybridized slides were scanned using a GenePix 4200A scanner (Axon Instrument, Union City, CA, USA) with photomultiplier gains adjusted to obtain highest-intensity unsaturated images. Data analysis was performed in the R statistical environment (version 2.2.1) using the Linear models for Microarray data package LIMMA 2.2.0, which is part of the Bioconductor project ([www.bioconductor.org](http://www.bioconductor.org)) (9). Raw data were processed using lowess normalization and movingmin. Correlation between three duplicate spots per gene in each array was used to increase robustness. The linear model was fitted for each exon given from a series of arrays using lmFit function. The empirical Bayes method (10) was used to rank differential expression signal-fold changes of individual gene exons using eBayes function.

*ABCC2* mRNA was isolated from cultured skin fibroblasts and subjected to reverse transcription. Overlapping ≈ 800 bp fragments were amplified by nested PCR from cDNA (see Table 1 for primer sequences), gel purified and sequenced.

Mutations in *UGT1A1* known to be associated with Gilbert syndrome in Caucasians were detected as described (8).

## Results

### Histology

Routine light microscopy of haematoxylin/eosin-stained sections of liver from both patients found no

**Table 1.** Sequences of PCR primer pairs used to amplify the overlapping fragments of *ABCC2* from cDNA

Pair number	Forward primer	Reverse primer
1	5'-TAGAAGAGTCTTCGTCCAGACGCAG-3'	5'-AGTGCCCGCCTGGCTTTC-3'
2	5'-TTCTGAAAGGCTACAAGCGTCCTC-3'	5'-ATTGGGATTACAAGCACCATCACC-3'
3	5'-AACTTCATGCACATGCTGTGGTC-3'	5'-CCTTTATGGTGCCATTCTGAATCC-3'
4	5'-AATCCTCCTTGATATCAGCCATGC-3'	5'-TCAAGGAGTTTCTCAGGGACTTCAG-3'
5	5'-ACAGCTTTCGTCGAACACTTAGCC-3'	5'-GGATAACTGGCAAACCTGATACGG-3'
6	5'-ACCATCATCGTCATTCTTGG-3'	5'-TGTGAAAGGGTCGAGATTCATCC-3'
7	5'-ATATTGCTCCATTGGGCTCCAC-3'	5'-TGGGTAGTAGGTTTCATGGGTGTTCC-3'

abnormalities (Fig. 2). Pigment accumulation, the characteristic histomorphological feature of DJS, was not detected. No autofluorescence was observed on fluorescence microscopy.

### ABCC2 protein expression

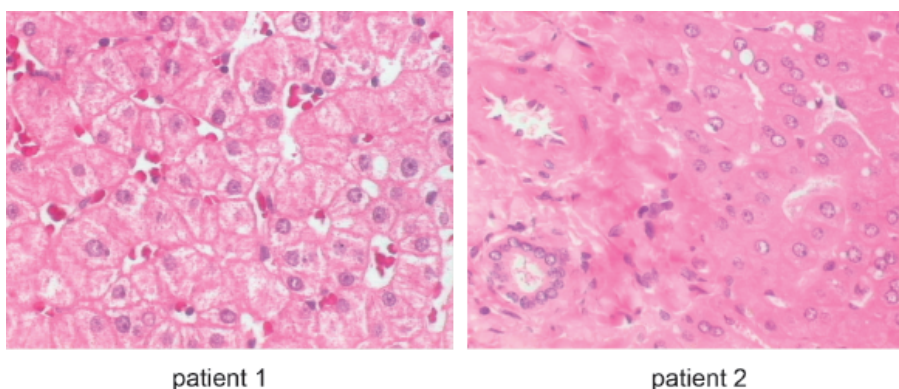
Immunohistochemical staining showed linear marking for CEACAM1 in the canalicular membrane of hepatocytes in both our patients, as well as in the positive control. CLSM with double immunofluorescence staining confirmed the localization of ABCC2 in the canalicular membrane of hepatocytes (Fig. 3). In contrast, no ABCC2 immunostaining was observed in sections of a liver-biopsy specimen from a patient with proven ABCC2 deficiency (negative control).

### DNA analysis

In Patient 1, sequence analysis of *ABCC2* disclosed heterozygosity for the known synonymous polymorphism 3972C/T (I1324I, GenBank dbSNP rs3740066) in exon 28 (11) and for the polymorphism –1023G/A (GenBank dbSNP rs7910642). The sequence of cDNA corresponded with the genomic

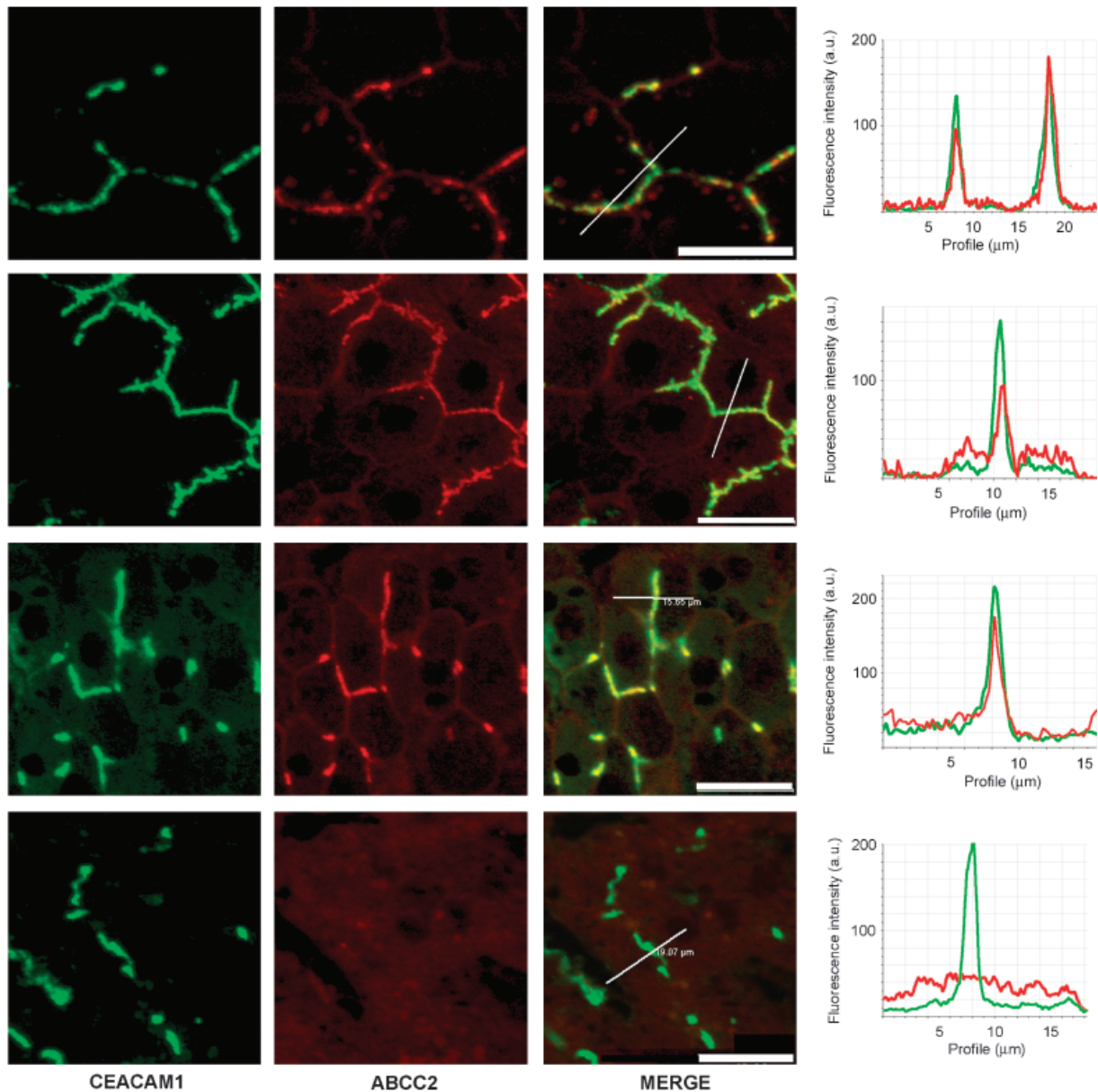
sequence. Heterozygosity for the 3972C/T polymorphism indicated that both alleles of the gene were expressed at the RNA level in cultured skin fibroblasts. A heterozygous polymorphism A(TA)<sub>7</sub>TAA of the TATAA-box and a heterozygous polymorphism –3263T/G in the phenobarbital-responsive enhancer module were detected in the promoter region of the *UGT1A1* gene. Family analysis revealed homozygosity for both variants, A(TA)<sub>7</sub>TAA and –3263G, in the proband's daughter, indicating that both heterozygous polymorphisms are located on the same chromosome in the proband, Patient 1. Finally, Patient 1 was homozygous for the wild-type 211G allele.

Sequence analysis of *ABCC2* in Patient 2 revealed homozygosity for two known polymorphisms: 3972C/T (I1324I) in the protein-coding region and –24C/T (GenBank dbSNP rs717620) in the 5'-untranslated region (11). As in Patient 1, the sequence of cDNA corresponded with the genomic sequence. In contrast to Patient 1, the expression of individual alleles could not be addressed because no heterozygous sequence variation was present in the transcribed mRNA. The patient was found to be homozygous for the wild-type



**Fig. 2.** Liver histology of Patient 1 (centrizonal area, needle biopsy taken at the age of 52 years) and Patient 2 (periportal area, liver excision performed in 1958 at the age of 20 years). Normal histology and cytology with no liver pigment was found in both specimens. The absence of lipopigment in hepatocytes in Patient 1 at the age of 52 is uncommon. Haematoxylin&eosin, original magnification  $\times 400$ .





**Fig. 3.** Confocal laser scanning micrographs with double immunofluorescence staining for CEACAM1 (left column) and ABCC2 (middle column). In the control liver as well as in both samples from the patients with RS, ABCC2 and CEACAM1 colocalize with the canalicular membrane of hepatocytes (right column – yellow color, and graphs). In the liver of a patient with proven ABCC2 deficiency (bottom slides), ABCC2 protein is absent both at canalicular membranes and in the cytoplasm of hepatocytes, whereas expression of CEACAM1 is not affected. Bar = 20  $\mu\text{m}$ , original magnification  $\times 400$ .

alleles A(TA)<sub>6</sub>TAA and – 3263T of the *UGT1A1* gene promoter and for the wild-type 211G allele in the first exon of *UGT1A1*.

Comparative genomic hybridization to a custom micro-array revealed no significant copy number changes in any of 32 exons of *ABCC2* in either Patient 1 or 2.

## Discussion

We attempted to investigate the role of *ABCC2* in two subjects with Rotor-type hyperbilirubinaemia; to our knowledge, ours is the first such attempt. Normal expression and localization of *ABCC2* on the canalicular membrane of hepatocytes ruled out the most

common findings seen in DJS, when mutations in both alleles of *ABCC2* abolish *ABCC2* expression. Normal localization of *ABCC2* excluded potential defects in *RDX*, encoding radixin, a cytoskeletal protein essential in anchoring *ABCC2* to the canalicular membrane (12). The immunohistologic findings, however, did not exclude two possibilities. Firstly, a mutation-impairing function but not expression of *ABCC2* might be present in at least one allele of *ABCC2*. Secondly, a mutation in the regulatory region of *ABCC2* might decrease but not abolish *ABCC2* expression. Both possibilities were checked by sequence analysis of the coding and promoter regions of *ABCC2*. In addition to conventional mutational screening, a thorough search for less common types of mutations – exon deletions and duplications – was performed; none was found. Finally, a contribution of Gilbert syndrome to the unconjugated fraction of elevated serum bilirubin was excluded by *UGT1A1* genotyping.

The rationale for investigation of *ABCC2* and *ABCC2* in RS, which is considered a disorder of hepatic bilirubin storage (13), can be questioned. Significant reduction in the actual or apparent capacity of hepatocytes to store unconjugated bilirubin, unconjugated BSP and indocyanine green (which does not undergo conjugation) can result from decreased concentrations of 'ligandin' in cytosol of hepatocytes; from decreased hepatocellular uptake of unconjugated bilirubin, BSP and indocyanine green; and from occupation of binding sites of intracellular 'ligandin' – proteins belonging mainly to the  $\alpha$ -class of the glutathione-S-transferase family (GST- $\alpha$ ).

The hypothesis that RS is a primary disorder of hepatic bilirubin storage is supported by the kinetics of selected anionic dyes and by the immunohistological findings of Abei M et al. (14). To the best of our knowledge, these findings have not been confirmed by quantitative analysis of GST- $\alpha$  isoenzymes or by mutational analysis of the corresponding *GSTA1-5* genes. A compensatory upregulation of GST- $\alpha$  family proteins in the liver has been detected in *Gsta4* null mice; no jaundice has been observed in these animals (15). If the same is true for single gene defects in *GSTA1* or *GSTA2* (less likely in *GSTA3* and *GSTA5* because their expression is low), such defects cannot be expected to reduce substantially the total concentration of GST- $\alpha$  family proteins in the liver.

Defective uptake of unconjugated bilirubin and other organic anions is not compatible with predominantly conjugated hyperbilirubinaemia in RS.

Thirdly, reduction of hepatic bilirubin storage capacity can be caused by occupation of the binding sites of

'ligandin'. Retention of 'ligandin'-binding substrates in cytosol may arise from changes in affinity of mutated *ABCC2* for a subgroup of *ABCC2* substrates. *ABCC2* can be mutated in such a way that retained nonbilirubin substrates do not upregulate expression of *ABCC3* in compensation for impaired function of *ABCC2*. Mutations affecting the affinity of *ABCC2* for various substrates have been documented by Ito et al. (16–18). Our presented results exclude this possibility in these patients.

Alternatively, 'ligandin'-binding substrates may be retained in cytosol of hepatocytes owing to deficient transport of polar conjugates from the endoplasmic reticulum (ER) to the cytoplasm. Conjugation takes place on the luminal aspect of ER (19, 20). Transport of highly polar conjugates is likely mediated by an ATP-independent permease specific for conjugates of bilirubin (and possibly other substrates) with glucuronic acid (19, 21). Production of bilirubin glucuronides in the ER lumen and activity of the canalicular bilirubin export pump *ABCC2*, which keeps cytoplasmic concentrations of bilirubin glucuronides low, constitute the driving force for translocation of conjugated bilirubin across the ER membrane. In the case of impaired export, conjugated bilirubin should be retained in the ER lumen and secreted into the plasma via exocytosis. Increased intra-ER concentrations of conjugated bilirubin (and perhaps other glucuronides) may decrease the rates of conjugation of the corresponding substrates. Unconjugated substrates retained in cytosol (owing to deficient export of conjugates from the ER to the cytoplasm) bind to GST- $\alpha$  and decrease hepatic storage capacity for unconjugated bilirubin, unconjugated BSP and indocyanine green as well as the transport maximum for BSP.

In conclusion, we have shown that RS is not an allelic variant of DJS. The discovery of the molecular background of RS would be helpful in differentiating among forms of hereditary conjugated jaundice.

## Acknowledgements

This study was supported by grant MZO 00023001 from the Institute for Clinical and Experimental Medicine. The authors from Institute of Inherited Metabolic Diseases were supported by grant 1A/8239-3 from the Grant Agency of the Ministry of Health of the Czech Republic; their institutional support was provided by grant MSM 0021620806 from the Ministry of Education of the Czech Republic. The authors thank Lucie Budisova for technical assistance, MVDr. Zorka Novotna for the culture of primary fibroblasts and A.S. Knisely, M.D., and Prof. MUDr. Milan Elleder, Dr.Sc., for their comments and help with the preparation of the manuscript.

## References

1. Paulusma CC, Kool M, Bosma PJ, *et al.* A mutation in the human canalicular multispecific organic anion transporter gene causes the Dubin–Johnson syndrome. *Hepatology* 1997; **25**: 1539–42.
2. Wolkoff AW, Wolpert E, Pascasio FN, Arias IM. Rotor's syndrome. A distinct inheritable pathophysiologic entity. *Am J Med* 1976; **60**: 173–9.
3. Wolpert E, Pascasio FM, Wolkoff AW, Arias IM. Abnormal sulfobromophthalein metabolism in Rotor's syndrome and obligate heterozygotes. *N Engl J Med* 1977; **296**: 1099–101.
4. Bar-Meir S, Baron J, Seligson U, Gottesfeld F, Levy R, Gilat T. 99mTc-HIDA cholescintigraphy in Dubin–Johnson and Rotor syndromes. *Radiology* 1982; **142**: 743–6.
5. Fretzayas AM, Garoufi AI, Moutsouris CX, Karpathios TE. Cholescintigraphy in the diagnosis of Rotor syndrome. *J Nucl Med* 1994; **35**: 1048–50.
6. Fretzayas AM, Stavrinadis CS, Koukoutsakis PM, Sinaniotis CA. Diagnostic approach of Rotor syndrome with cholescintigraphy. *Clin Nucl Med* 1997; **22**: 635–6.
7. Mikulecky M. Das atypische Dubin–Johnsonsche syndrom. *Gastroenterologia* 1960; **94**: 201–26.
8. Cebecauerova D, Jirasek T, Budisova L, *et al.* Dual hereditary jaundice: simultaneous occurrence of mutations causing Gilbert's and Dubin–Johnson syndrome. *Gastroenterology* 2005; **129**: 315–20.
9. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, eds. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. New York: Springer, 2005; 397–420.
10. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; **3**: Article3.
11. Itoda M, Saito Y, Soyama A, *et al.* Polymorphisms in the ABCC2 (cMOAT/MRP2) gene found in 72 established cell lines derived from Japanese individuals: an association between single nucleotide polymorphisms in the 5'-untranslated region and exon 28. *Drug Metab Dispos* 2002; **30**: 363–4.
12. Kikuchi S, Hata M, Fukumoto K, *et al.* Radixin deficiency causes conjugated hyperbilirubinaemia with loss of Mrp2 from bile canalicular membranes. *Nat Genet* 2002; **31**: 320–5.
13. Chowdhury NR, Arias IM, Wolkoff AW, Chowdhury JR. Disorders of bilirubin metabolism. In: Arias IM, ed. *The Liver. Biology and Pathobiology*. Philadelphia: Lippincott Williams & Wilkins, 2001; 290–309.
14. Abei M, Matsuzaki Y, Tanaka N, Osuga T, Adachi Y. Defective hepatic glutathione S-transferase in Rotor's syndrome. *Am J Gastroenterol* 1995; **90**: 681–2.
15. Engle MR, Singh SP, Czernik PJ, *et al.* Physiological role of mGSTA4-4, a glutathione S-transferase metabolizing 4-hydroxynonal: generation and analysis of mGsta4 null mouse. *Toxicol Appl Pharmacol* 2004; **194**: 296–308.
16. Ito K, Suzuki H, Sugiyama Y. Charged amino acids in the transmembrane domains are involved in the determination of the substrate specificity of rat Mrp2. *Mol Pharmacol* 2001; **59**: 1077–85.
17. Ito K, Suzuki H, Sugiyama Y. Single amino acid substitution of rat MRP2 results in acquired transport activity for taurocholate. *Am J Physiol Gastrointest Liver Physiol* 2001; **281**: G1034–43.
18. Ito K, Oleschuk CJ, Westlake C, Vasa MZ, Deeley RG, Cole SP. Mutation of Trp1254 in the multispecific organic anion transporter, multidrug resistance protein 2 (MRP2) (ABCC2), alters substrate specificity and results in loss of methotrexate transport activity. *J Biol Chem* 2001; **276**: 38108–14.
19. Jansen PL, Mulder GJ, Burchell B, Bock KW. New developments in glucuronidation research: report of a workshop on “glucuronidation, its role in health and disease”. *Hepatology* 1992; **15**: 532–44.
20. Radomska-Pandya A, Czernik PJ, Little JM, Battaglia E, Mackenzie PI. Structural and functional studies of UDP-glucuronosyltransferases. *Drug Metab Rev* 1999; **31**: 817–99.
21. Csala M, Staines AG, Banhegyi G, Mandl J, Coughtrie MW, Burchell B. Evidence for multiple glucuronide transporters in rat liver microsomes. *Biochem Pharmacol* 2004; **68**: 1353–62.



# Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver

Evita van de Steeg,<sup>1</sup> Viktor Stránecký,<sup>2,3</sup> Hana Hartmannová,<sup>2,3</sup> Lenka Nosková,<sup>3</sup> Martin Hřebíček,<sup>3</sup> Els Wagenaar,<sup>1</sup> Anita van Esch,<sup>1</sup> Dirk R. de Waart,<sup>4</sup> Ronald P.J. Oude Elferink,<sup>4</sup> Kathryn E. Kenworthy,<sup>5</sup> Eva Sticová,<sup>6</sup> Mohammad al-Edreesi,<sup>7</sup> A.S. Knisely,<sup>8</sup> Stanislav Kmoch,<sup>2,3</sup> Milan Jirsa,<sup>6</sup> and Alfred H. Schinkel<sup>1</sup>

<sup>1</sup>Division of Molecular Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands. <sup>2</sup>Center for Applied Genomics and

<sup>3</sup>Institute of Inherited Metabolic Diseases, Charles University of Prague, First Faculty of Medicine, Prague, Czech Republic.

<sup>4</sup>Tytgat Institute for Liver and Intestinal Research, Academic Medical Center, Amsterdam, The Netherlands.

<sup>5</sup>Department of Drug Metabolism and Pharmacokinetics, GlaxoSmithKline, Ware, United Kingdom. <sup>6</sup>Institute for Clinical and Experimental Medicine, Prague, Czech Republic. <sup>7</sup>Department of Pediatrics, Saudi Aramco Dhahran Health Centre, Dhahran, Saudi Arabia. <sup>8</sup>Institute of Liver Studies, King's College Hospital, London, United Kingdom.

**Bilirubin, a breakdown product of heme, is normally glucuronidated and excreted by the liver into bile. Failure of this system can lead to a buildup of conjugated bilirubin in the blood, resulting in jaundice. The mechanistic basis of bilirubin excretion and hyperbilirubinemia syndromes is largely understood, but that of Rotor syndrome, an autosomal recessive disorder characterized by conjugated hyperbilirubinemia, coproporphyrinuria, and near-absent hepatic uptake of anionic diagnostics, has remained enigmatic. Here, we analyzed 8 Rotor syndrome families and found that Rotor syndrome was linked to mutations predicted to cause complete and simultaneous deficiencies of the organic anion transporting polypeptides OATP1B1 and OATP1B3. These important detoxification-limiting proteins mediate uptake and clearance of countless drugs and drug conjugates across the sinusoidal hepatocyte membrane. OATP1B1 polymorphisms have previously been linked to drug hypersensitivities. Using mice deficient in *Oatp1a/1b* and in the multispecific sinusoidal export pump *Abcc3*, we found that *Abcc3* secretes bilirubin conjugates into the blood, while *Oatp1a/1b* transporters mediate their hepatic reuptake. Transgenic expression of human OATP1B1 or OATP1B3 restored the function of this detoxification-enhancing liver-blood shuttle in *Oatp1a/1b*-deficient mice. Within liver lobules, this shuttle may allow flexible transfer of bilirubin conjugates (and probably also drug conjugates) formed in upstream hepatocytes to downstream hepatocytes, thereby preventing local saturation of further detoxification processes and hepatocyte toxic injury. Thus, disruption of hepatic reuptake of bilirubin glucuronide due to coexisting OATP1B1 and OATP1B3 deficiencies explains Rotor-type hyperbilirubinemia. Moreover, OATP1B1 and OATP1B3 null mutations may confer substantial drug toxicity risks.**

## Introduction

Rotor syndrome (RS; OMIM #237450) is a rare, benign hereditary conjugated hyperbilirubinemia, also featuring coproporphyrinuria and strongly reduced liver uptake of many diagnostic compounds, including cholescintigraphic tracers (1–6). RS is an autosomal recessive disorder that clinically resembles another conjugated hyperbilirubinemia, the Dubin-Johnson syndrome (DJS; OMIM #237500) (7, 8). In both RS and DJS, mild jaundice begins shortly after birth or in childhood. There are no signs of hemolysis, and routine hematologic and clinical-biochemistry test results are normal, aside from the primarily conjugated hyperbilirubinemia. RS is, however, distinguishable from DJS by several criteria (1, 2, 9, 10): (a) it lacks the hepatocyte pigment deposits typical of DJS; (b) in

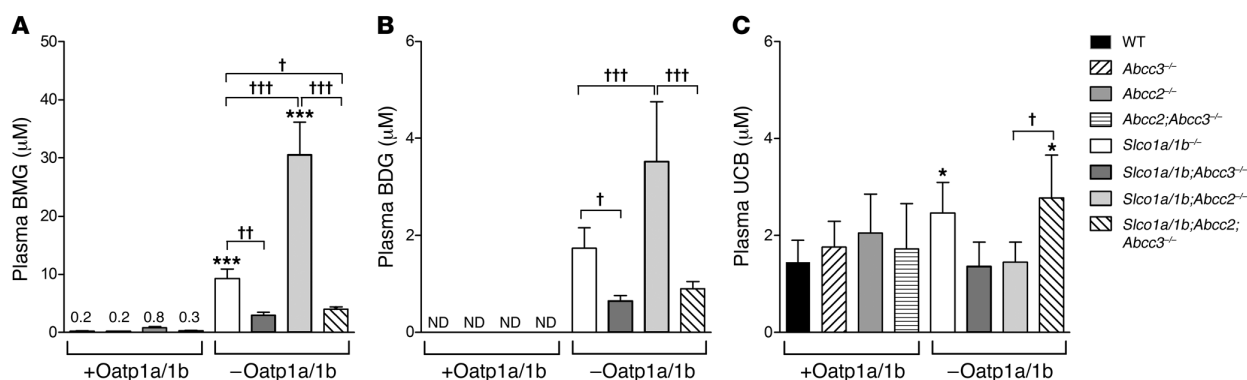
RS, but not DJS, there is delayed plasma clearance of unconjugated bromsulphthalein (BSP), an anionic diagnostic dye, and no conjugated BSP appears in plasma (4); (c) the liver in RS is scarcely visualized on <sup>99m</sup>Tc-N[2,6-dimethylphenyl-carbamoylmethyl] iminodiacetic acid (<sup>99m</sup>Tc-HIDA) cholescintigraphy, with slow liver uptake, persistent visualization of the cardiac blood pool, and prominent kidney excretion (5); and (d) total urinary excretion of coproporphyrins is greatly increased in RS, with coproporphyrin I being the predominant isomer (11).

DJS is caused by mutations affecting ABCC2/MRP2, a canalicular bilirubin glucuronide and xenobiotic export pump, thus disrupting bilirubin glucuronide excretion into bile (7, 8). Excretion of bilirubin glucuronides is then redirected into plasma by the action of ABCC3/MRP3, a homolog of ABCC2 that is present in the sinusoidal membrane and is upregulated in DJS (12, 13). The molecular mechanism of DJS is in line with the generally accepted paradigm of normal hepatic bilirubin excretion, according to which a unidirectional elimination pathway is postulated: first, uptake of unconjugated bilirubin (UCB) from blood into hepatocytes; subse-

**Authorship note:** Evita van de Steeg and Viktor Stránecký, and Milan Jirsa and Alfred H. Schinkel contributed equally to this work.

**Conflict of interest:** The research group of Alfred H. Schinkel receives revenues from commercial distribution of some of the mouse strains used in this study.

**Citation for this article:** *J Clin Invest.* 2012;122(2):519–528. doi:10.1172/JCI59526.



**Figure 1** Increased plasma bilirubin glucuronide in *Slco1a/1b*<sup>-/-</sup> mice is in part dependent on *Abcc3*. (A) BMG, (B) BDG, and (C) UCB levels in plasma of male wild-type, *Abcc3*<sup>-/-</sup>, *Abcc2*<sup>-/-</sup>, *Abcc2*<sup>-/-</sup>*Abcc3*<sup>-/-</sup>, *Slco1a/1b*<sup>-/-</sup>, *Slco1a/1b*;*Abcc3*<sup>-/-</sup>, *Slco1a/1b*;*Abcc2*<sup>-/-</sup>, and *Slco1a/1b*;*Abcc2*;*Abcc3*<sup>-/-</sup> mice (*n* = 4–7). +*Oatp1a/1b* denotes strains possessing *Oatp1a/1b* proteins, and –*Oatp1a/1b* denotes strains lacking *Oatp1a/1b* proteins. Data are mean ± SD. \**P* < 0.05, \*\*\**P* < 0.001 compared with wild-type mice. Bracketed comparisons: †*P* < 0.05, ††*P* < 0.01, †††*P* < 0.001. ND, not detectable; detection limit was 0.1 µM.

quent glucuronidation; and finally, secretion of bilirubin glucuronide into bile via *ABCC2*. Individuals with RS, however, lack *ABCC2* mutations (14), and the mechanistic basis of RS is unknown.

Organic anion transporting polypeptides (OATPs, genes: *SLCOs*) contain 12 plasma membrane–spanning domains and mediate sodium-independent cellular uptake of highly diverse compounds, including bilirubin glucuronide, bile acids, steroid and thyroid hormones, and numerous drugs, toxins, and their conjugates (15, 16). Human OATP1B1 and OATP1B3 localize to the sinusoidal membrane of hepatocytes and mediate the liver uptake of, among other compounds, many drugs (15–19). Various SNPs in *SLCO1B1* cause reduced transport activity and altered plasma and tissue levels of statins, methotrexate, and irinotecan in patients, potentially resulting in life-threatening toxicities (20–24).

In a *Slco1a/1b*<sup>-/-</sup> mouse model recently generated by our group, the importance of *Oatp1a/1b* proteins in hepatic uptake and clearance of drugs was confirmed, but the mice also displayed marked conjugated hyperbilirubinemia (25). We therefore hypothesized that sinusoidal Oatps in the normal, healthy mouse liver function in tandem with the sinusoidal efflux transporter *Abcc3* to mediate substantial hepatic secretion and reuptake of bilirubin glucuronides and other conjugated compounds (25).

Here we describe how a combination of functional studies in mice to address this hypothesis and independent genetic studies in humans has resulted in elucidation of the genetic and mechanistic basis of Rotor syndrome.

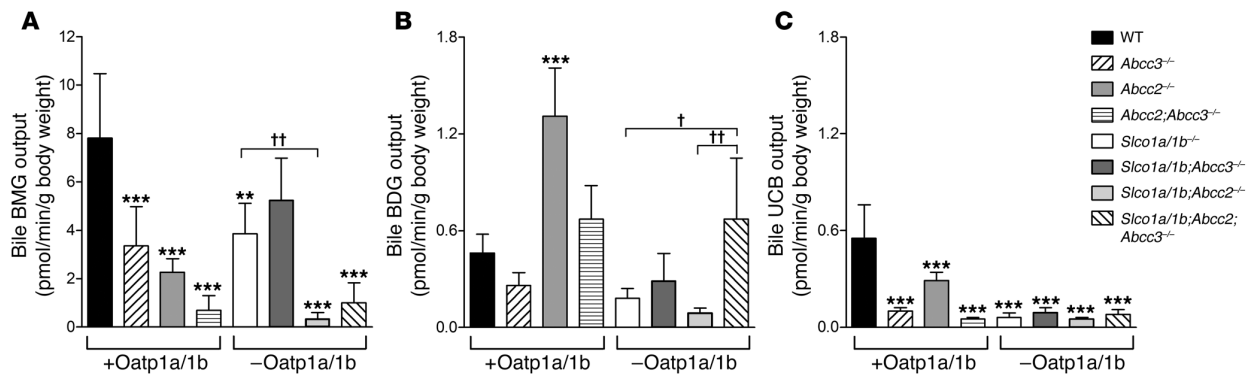
## Results

To test our hypothesis regarding the involvement of *Abcc3* in the sinusoidal cycling of bilirubin glucuronides, and to assess a possible interplay with *Abcc2*, we generated *Slco1a/1b*<sup>-/-</sup>*Abcc3*<sup>-/-</sup> (*Slco1a/1b*;*Abcc3*<sup>-/-</sup>), *Slco1a/1b*<sup>-/-</sup>*Abcc2*<sup>-/-</sup> (*Slco1a/1b*;*Abcc2*<sup>-/-</sup>), and *Slco1a/1b*<sup>-/-</sup>*Abcc2*<sup>-/-</sup>*Abcc3*<sup>-/-</sup> (*Slco1a/1b*;*Abcc2*;*Abcc3*<sup>-/-</sup>) mice by crossbreeding of existing strains. All strains were fertile, with normal life spans and body weights. As previously found for *Abcc2*<sup>-/-</sup> and *Abcc2*<sup>-/-</sup>*Abcc3*<sup>-/-</sup> mice (26, 27), liver weights of *Slco1a/1b*;*Abcc2*<sup>-/-</sup> and *Slco1a/1b*;*Abcc2*;*Abcc3*<sup>-/-</sup> mice were significantly increased (~30% and ~50%, respectively) compared with wild-type

mice (data not shown). Quantitative RT-PCR analysis of functionally relevant uptake and efflux transporters in liver, kidney, and intestine of the single and combination knockout strains revealed only some modest expression changes (Supplemental Table 1 and Supplemental Results; supplemental material available online with this article; doi:10.1172/JCI59526DS1). Hepatic UDP-glucuronosyltransferase 1a1 (*Ugt1a1*) expression was not significantly altered in any of the strains.

Importantly, the markedly increased plasma bilirubin mono-glucuronide (BMG) and bilirubin diglucuronide (BDG) levels observed in *Slco1a/1b*<sup>-/-</sup> mice were substantially reduced in *Slco1a/1b*;*Abcc3*<sup>-/-</sup> mice, demonstrating that *Abcc3* is necessary for most of this increase (Figure 1, A and B). Plasma BMG levels in *Slco1a/1b*;*Abcc2*<sup>-/-</sup> mice, even further increased owing to strongly reduced biliary BMG excretion (Figure 2, A and B), were similarly decreased in *Slco1a/1b*;*Abcc2*;*Abcc3*<sup>-/-</sup> mice (Figure 1, A and B). Thus, *Abcc3* secretes bilirubin glucuronides back into blood, and *Oatp1a/1b* proteins mediate their efficient hepatic reuptake, thereby together establishing a sinusoidal liver–blood shuttling loop. The incomplete reversion of plasma bilirubin glucuronide levels in the *Oatp1a/1b*/*Abcc3*-deficient strains (Figure 1, A and B) suggests that additional sinusoidal exporter(s), e.g., *Abcc4* (28), can partly take over the sinusoidal bilirubin glucuronide extrusion function of *Abcc3*.

The biliary output of bilirubin glucuronides in the single and combination knockout mice showed that, as long as *Oatp1a/1b* was functional, *Abcc3* improved the efficiency of biliary bilirubin glucuronide excretion, even though it transports its substrates initially from liver to blood, not bile (Figure 2, A and B, strains +*Oatp1a/1b*). This suggests that, within liver lobules, the bilirubin glucuronide extruded by *Abcc3* in upstream hepatocytes is efficiently taken up in downstream hepatocytes via *Oatp1a/1b* and then excreted into bile. The resulting relief of possible saturation of (or competition for) biliary excretion in the upstream hepatocytes may explain why the overall biliary excretion is enhanced by this transfer to downstream hepatocytes. However, when *Oatp1a/1b* was absent, *Abcc3* instead decreased biliary bilirubin glucuronide excretion (Figure 2, strains –*Oatp1a/1b*) and

**Figure 2**

In the presence of *Oatp1a/1b*, but not in its absence, *Abcc3* enhances biliary excretion of bilirubin glucuronides. (A) BMG, (B) BDG, and (C) UCB output in bile of male wild-type, *Abcc3*<sup>-/-</sup>, *Abcc2*<sup>-/-</sup>, *Abcc2*<sup>-/-</sup>*Abcc3*<sup>-/-</sup>, *Slco1a/1b*<sup>-/-</sup>, *Slco1a/1b*;*Abcc3*<sup>-/-</sup>, *Slco1a/1b*;*Abcc2*<sup>-/-</sup>, and *Slco1a/1b*;*Abcc2*<sup>-/-</sup>*Abcc3*<sup>-/-</sup> mice. Bile collected during the first 15 minutes after gall bladder cannulation was analyzed. +*Oatp1a/1b* denotes strains possessing *Oatp1a/1b* proteins, and -*Oatp1a/1b* denotes strains lacking *Oatp1a/1b* proteins. Data are shown as mean ± SD (*n* = 4–7). \*\**P* < 0.01, \*\*\**P* < 0.001 compared with wild-type mice. Bracketed comparisons: †*P* < 0.05, ††*P* < 0.01.

redirected excretion toward urine via the increased plasma bilirubin glucuronide levels (Supplemental Figure 1). Obviously, in the absence of *Oatp1a/1b*-mediated hepatic reuptake, *Abcc3* activity can only decrease hepatocyte levels of bilirubin glucuronide in upstream and downstream hepatocytes alike, and will therefore reduce overall biliary excretion. Thus, both components of the *Abcc3* and *Oatp1a/1b* shuttling loop are necessary to improve hepatobiliary excretion efficiency.

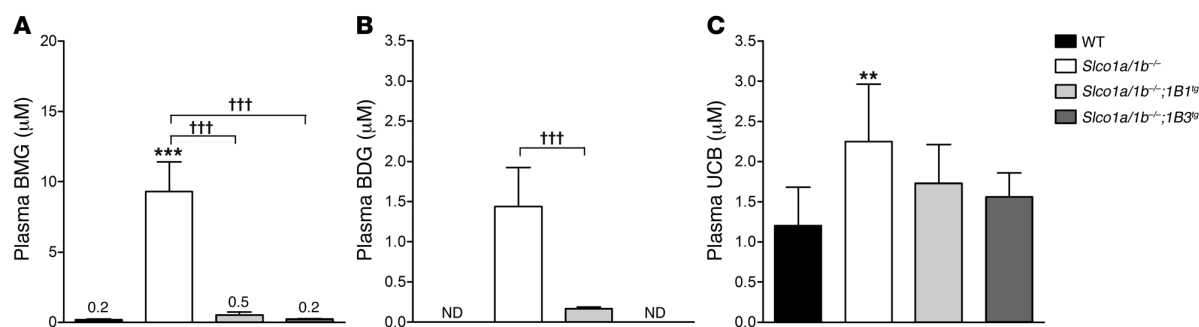
Human hepatocytes express only two *OATP1A/1B* proteins at the sinusoidal membrane, *OATP1B1* and *OATP1B3* (15). To test whether these could mediate the identified *Oatp1a/1b* functions, and in a liver-specific manner, we generated *Slco1a/1b*<sup>-/-</sup> mice with liver-specific expression of either human *OATP1B1* or *OATP1B3*. Liver-specific expression was obtained using an apoE promoter (29). These strains were viable and fertile, and displayed normal life spans and body weights. Liver levels of transgenic *OATP1B1* and *OATP1B3* proteins were similar to those seen in pooled human liver samples (data not shown). Both of the transgenic rescue strains displayed a virtually complete reversal of the increases in plasma and urine levels of BMG and BDG seen in *Slco1a/1b*<sup>-/-</sup> mice (Figure 3, A and B, and Supplemental Figure 2). This indicates that both human *OATP1B1* and *OATP1B3* effectively reabsorb bilirubin glucuronides from plasma into the liver, in line with their demonstrated *in vitro* role in bilirubin glucuronide uptake (30). The modest (~1.8-fold) increase in plasma UCB in *Slco1a/1b*<sup>-/-</sup> mice was also reduced in the rescue strains (Figure 3C), suggesting an ancillary role of these proteins in hepatic UCB uptake.

These findings collectively raised the question as to whether humans with a severe deficiency in *OATP1B1* and *OATP1B3*, possibly leading to a conjugated hyperbilirubinemia, might exist. A literature search suggested RS as a candidate inborn metabolic disorder. A search for RS subjects by part of the present group led to collaboration with another team already working on mapping of the RS gene(s).

In an unbiased approach, scanning the whole genome, we mapped the genomic candidate intervals for RS in 11 RS index subjects from 8 different families, 4 Central European (CE1–CE4), 3 Saudi-Arabian (A1–A3), and 1 Filipino (P1) (Figure 4A and Supplemen-

tal Table 2). Homozygosity mapping identified a single genomic region on chromosome 12 for which 8 tested index subjects and no healthy siblings or parents were homozygous (Figure 4B), suggesting inheritance of both alleles from a common ancestor. Three distinct homozygous haplotypes (R1–R3) segregated with RS: R1 in families CE1, CE2, and CE4; R2 in families CE3, A1, A2, and A3; and R3 in family P1 (Figure 4B; for genotyping details, see Methods). Intersection of these haplotypes defined a candidate genomic region spanning the *SLCO1C1*, *SLCO1B3*, *SLCO1B1*, *SLCO1A2*, and *IAPP* genes (Figure 4B). A parallel genome-wide copy number analysis detected a homozygous deletion within the *SLCO1B3* gene in the R1 haplotype and a homozygous approximately 405-kb deletion encompassing *SLCO1B3* and *SLCO1B1* and the *LST-3TM12* pseudogene in the R2 haplotype (Figure 4B and Supplemental Figure 3).

Sequence analysis revealed predictably pathogenic mutations affecting both *SLCO1B3* and *SLCO1B1* in each of the haplotypes (Figure 4, B–D, Table 1, Supplemental Figure 3, and Supplemental Table 3). In the R1 haplotype, a 7.2-kb deletion removes exon 12 of *SLCO1B3*, encoding amino acids 500–560 of *OATP1B3* (702 aa long) and introduces a frameshift and premature stop codon, thus removing the C-terminal 3 transmembrane domains. Furthermore, a nonsense mutation in exon 13, c.1738C→T, introduces a premature stop codon (p.R580X) in R1-linked *OATP1B1* (691 aa long), removing the C-terminal one-and-a-half transmembrane domains. The 405-kb R2 deletion encompasses exons 3–15 of *SLCO1B3* (sparing only a small N-terminal region) and the whole of *SLCO1B1*, but not *SLCO1A2*. The R3 haplotype harbors a splice donor site mutation, c.1747+1G→A, in intron 13 of *SLCO1B3*. If *SLCO1B3* is still yielding functional mRNA, this would truncate *OATP1B3* after amino acid 582, deleting the C-terminal one-and-a-half transmembrane domains. A nonsense mutation, c.757C→T, in exon 8 of R3-linked *SLCO1B1* introduces a premature stop (p.R253X), truncating *OATP1B1* before the C-terminal 7 transmembrane domains. All of these mutations would severely disrupt or annihilate proper protein expression and function. Moreover, they all showed consistent autosomal recessive segregation with the RS phenotype in the investigated families (Table 1). No

**Figure 3**

Increased plasma bilirubin glucuronide in *Slco1a/1b*<sup>-/-</sup> mice is reversed by human OATP1B1 and OATP1B3. (A) BMG, (B) BDG, and (C) UCB levels in plasma of male wild-type and *Slco1a/1b*<sup>-/-</sup> mice, and of the derived OATP1B1- and OATP1B3-transgenic strains (*Slco1a/1b*<sup>-/-</sup>;1B1<sup>tg</sup> and *Slco1a/1b*<sup>-/-</sup>;1B3<sup>tg</sup>, respectively) ( $n = 5-8$ ). Data are mean  $\pm$  SD. \*\* $P < 0.01$ , \*\*\* $P < 0.001$  compared with wild-type mice. Bracketed comparisons: ††† $P < 0.001$ . Detection limit was 0.1  $\mu$ M.

*SLCO1A2* sequence variation was found in probands representing the 3 haplotypes, rendering involvement of OATP1A2 in RS unlikely. The severity of the identified mutations affecting *SLCO1B3* and *SLCO1B1* and their strict cosegregation with the RS phenotype indicate that RS is caused by co-inherited complete functional deficiencies in both OATP1B3 and OATP1B1.

The severity of the mutations was independently supported by immunohistochemical studies of the sparse RS liver biopsy material available. Given their sparseness, immunostaining of these liver biopsies was performed using one antibody recognizing the N terminus of both OATP1B1 and OATP1B3 (31). This revealed absence of detectable staining in probands representing each haplotype (Figure 5). In controls, basolateral membranes of centrilobular hepatocytes stained crisply, as previously reported (31). Thus, the *SLCO1B1* and *SLCO1B3* mutations in each haplotype result in absence of a detectable signal for OATP1B protein in the liver.

In family A2, a heterozygous splice donor site mutation, c.481+1G→T, in intron 5 of *SLCO1B1* would result in dysfunctional RNA or protein. Its co-occurrence with the 405-kb R2 deletion in two asymptomatic family members (Table 1) indicates that a single functional *SLCO1B3* allele can prevent RS.

A search for copy number variations (CNVs) in existing databases and CNV genotyping of more than 2,300 individuals from various populations (see Supplemental Results) revealed additional heterozygous small and large deletions predicted to disrupt *SLCO1B1* or *SLCO1B3* function, including several approximately 400-kb deletions similar or identical to the R2 haplotype-linked deletion. One individual without jaundice, heterozygous for the R1 haplotype-associated c.1738C→T (p.R580X) mutation in *SLCO1B1*, was also homozygous for the R1 haplotype-associated deletion in *SLCO1B3*. Thus, a single functional *SLCO1B1* allele can also prevent RS. Combined with the findings in family A2 described above, this demonstrates that only a complete deficiency of both alleles of *SLCO1B1* and *SLCO1B3* will result in RS.

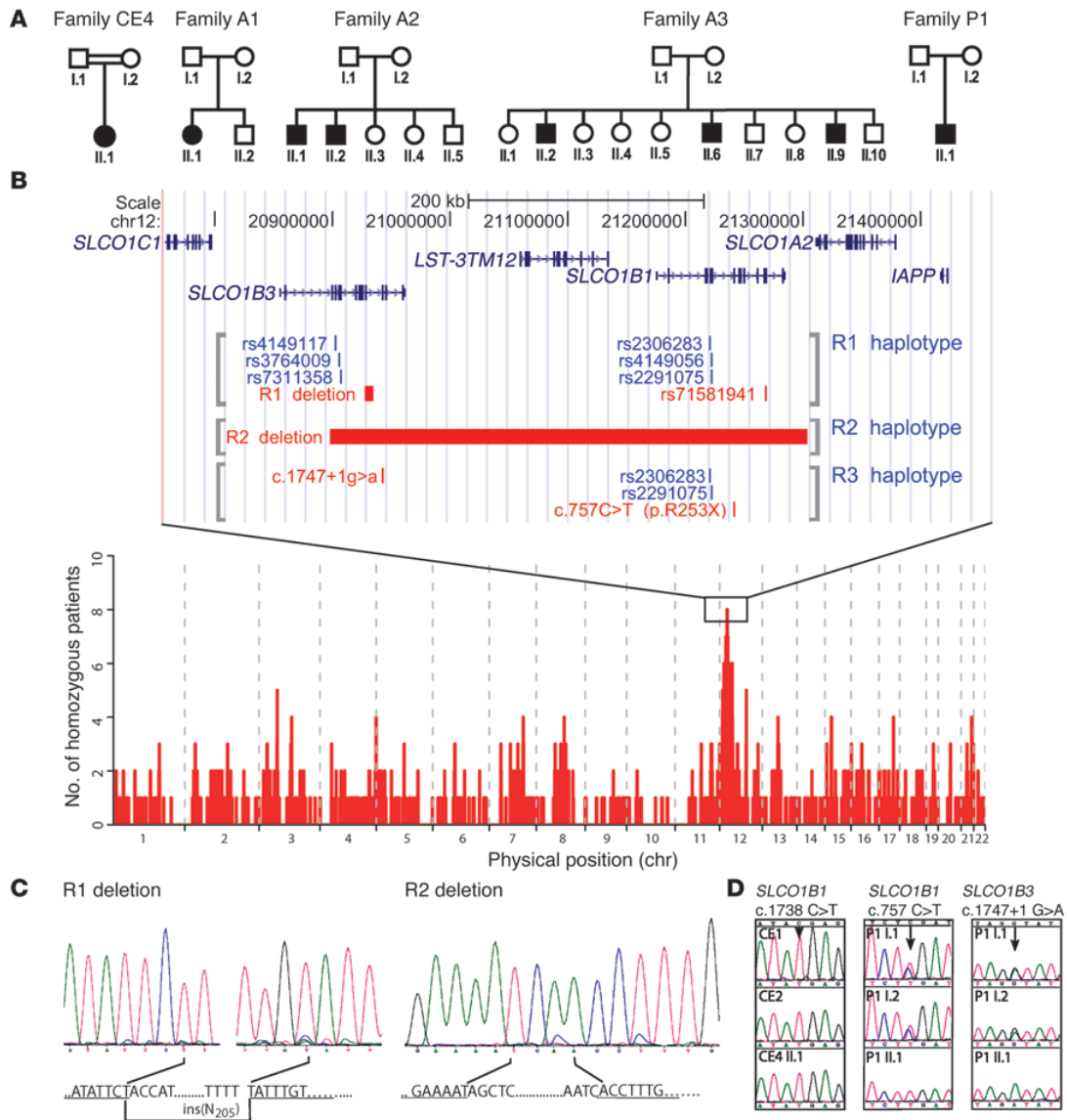
## Discussion

We demonstrate here that RS is an obligate two-gene disorder, caused by a complete deficiency of the major hepatic drug uptake transporters OATP1B1 and OATP1B3. We further identified individuals with a complete deficiency of either OATP1B1 or OATP1B3, which was not recognizable by obvious jaundice.

In spite of the documented important functions of especially OATP1B1 in drug detoxification, apparently such deficiencies are compatible with relatively normal life.

Using *Oatp1a/1b*-knockout mice, which can, retrospectively, be considered to be a partial model for RS, we showed that *Abcc3* is an important factor for the RS-like conjugated hyperbilirubinemia. Our data imply that in the normal human liver *ABCC3*, OATP1B1, and OATP1B3 may form a liver-blood shuttling loop for bilirubin glucuronide, similar to that driven by *Oatp1a/1b* and *Abcc3* in the mouse (Figure 6). A substantial fraction of bilirubin conjugated in hepatocytes is secreted back into the blood by *ABCC3* and subsequently reabsorbed in downstream hepatocytes by OATP1B1 and OATP1B3. In RS this reuptake is hampered, causing increased plasma bilirubin glucuronide levels and jaundice. The flexible “hepatocyte hopping” afforded by this loop facilitates efficient detoxification, presumably by circumventing saturation of further detoxification processes in upstream hepatocytes, including excretion into bile. Indeed, we could show that, counterintuitively, but in accordance with the hepatocyte hopping model, loss of *Abcc3* in mice resulted in decreased biliary excretion of bilirubin glucuronide, as long as *Oatp1a/1b* was present (Figure 2). This process likely also enhances hepatic detoxification of numerous drugs and drug conjugates (e.g., glucuronide, sulfate, and glutathione conjugates) transported by OATP1B1/3 and *ABCC3*. Moreover, this principle may also apply to other saturable hepatocyte detoxifying processes, such as phase I and phase II metabolism, as long as the substrate compounds involved are transported by *ABCC3* and OATP1B proteins. Additional sinusoidal efflux and uptake transporters (e.g., *ABCC4*, OATP2B1, NTCP) will further widen the scope of compounds affected by this hepatocyte hopping process. Results obtained with the *Slco1a/1b;Abcc3*-knockout mice indeed show that in addition to *Abcc3* there must be other sinusoidal efflux processes for bilirubin glucuronides. Preventing accumulation of drug glucuronides may be particularly important, since protein adduction by acyl-glucuronides is a well-established cause of drug (hepato)toxicity (32).

One should exercise caution when extrapolating mouse data to humans, and the individual *Oatp1a/1b* proteins are not straightforward orthologs of human OATP1B1 and OATP1B3. However, there is a strong analogy between the bilirubin phenotypes of *Oatp1a/1b*-knockout mice and human Rotor subjects. Moreover, the hepatic transgenic expression of human



**Figure 4**

RS families display deficiencies in *SLCO1B1* and *SLCO1B3*. **(A)** Pedigrees of the investigated families. Black symbols denote RS index subjects. Parents in family CE4 had a documented common ancestor. Families CE1–CE3 (only single individuals analyzed) are not shown. **(B)** Homozygosity regions in 8 RS index subjects and overview of detected mutations and polymorphisms. The genome map shows number and location of overlapping homozygosity regions in RS index subjects, gene content of the top candidate region on chromosome 12, and the genotypes forming all 3 identified RS haplotypes. Mutations crucial for RS are shown in red. chr, chromosome. **(C)** Sequences and electropherograms of the R1 and R2 deletion breakpoints. **(D)** Pathogenic point mutations in R1 and R3 haplotypes. Electropherograms indicate the c.1738C→T (p.R580X) mutation in *SLCO1B1* in probands CE1, CE2, and CE4 II.1 and the c.757C→T (p.R253X) and c.1747+1G→A mutations in *SLCO1B1* and *SLCO1B3*, respectively, in family P1.

OATP1B1 or OATP1B3 resulted in virtually complete rescue of the Oatp1a/1b-knockout phenotype for bilirubin handling (Figure 3). This strongly supports that the principles governing bilirubin handling by Oatp1a/1b in mouse liver also apply to OATP1B1 and OATP1B3 in human liver.

Analogous to the mouse data for Oatp1a/1b (25), the extensive glucuronidation of bilirubin in Rotor subjects suggests that OATP1B1 and/or OATP1B3 are not strictly essential for uptake of UCB into

the liver. Passive transmembrane diffusion is one likely candidate to take over this process, in hepatocytes and probably many other cell types as well (e.g., ref. 33), but we do not exclude that additional uptake transporters (perhaps OATP2B1) can also contribute to UCB uptake. However, OATP1B1 and/or OATP1B3 probably do contribute to hepatic UCB uptake, since in RS subjects a significant increase in plasma UCB is usually observed and reduced clearance of UCB has been reported (34, 35). Moreover, polymorphisms in *SLCO1B1*





**Table 1**  
Mutations in *SLCO1B* genes detected in RS subjects and their family members

Subject	Family status	Haplotype R1-linked mutations		Haplotype R2-linked mutations		Haplotype R3-linked mutations	
		<i>SLCO1B3</i> 7.2-kb deletion	<i>SLCO1B1</i> c.1738C→T (p.R580X) rs71581941	<i>SLCO1B</i> locus 405-kb deletion	<i>SLCO1B1</i> c.481+1G→T splice site mutation	<i>SLCO1B3</i> c.1747+1G→A splice site mutation	<i>SLCO1B1</i> c.757C→T (p.R253X)
<b>CE1</b>	<b>Proband</b>	del/del	T/T				
<b>CE2</b>	<b>Proband</b>	del/del	T/T				
CE4 I.1	Father	del/WT	T/C				
CE4 I.2	Mother	del/WT	T/C				
<b>CE4 II.1</b>	<b>Proband</b>	del/del	T/T				
<b>CE3</b>	<b>Proband</b>			del/del	—/—		
A1 I.1	Father			del/WT	—/G		
A1 I.2	Mother			del/WT	—/G		
<b>A1 II.1</b>	<b>Proband</b>			del/del	—/—		
A1 II.2	Brother			WT/WT	G/G		
A2 I.1	Father			del/WT	—/G		
A2 I.2	Mother			del/WT	—/T		
<b>A2 II.1</b>	<b>Brother</b>			del/del	—/—		
<b>A2 II.2</b>	<b>Proband</b>			del/del	—/—		
A2 II.3	Sister			del/WT	—/T		
A2 II.4	Sister			del/WT	—/G		
A2 II.5	Brother			WT/WT	G/T		
A3 I.1	Father			del/WT	—/G		
A3 I.2	Mother			del/WT	—/G		
A3 II.1	Sister			WT/WT	G/G		
<b>A3 II.2</b>	<b>Proband</b>			del/del	—/—		
A3 II.3	Sister			del/WT	—/G		
A3 II.4	Sister			del/WT	—/G		
A3 II.5	Sister			del/WT	—/G		
<b>A3 II.6</b>	<b>Brother</b>			del/del	—/—		
A3 II.7	Brother			del/WT	—/G		
A3 II.8	Sister			del/WT	—/G		
<b>A3 II.9</b>	<b>Brother</b>			del/del	—/—		
A3 II.10	Brother			WT/WT	G/G		
P1 I.1	Father					G/A	C/T
P1 I.2	Mother					G/A	C/T
<b>P1 II.1</b>	<b>Proband</b>					A/A	T/T

Boldface indicates index subjects with RS ( $n = 11$ ; 8 probands, 3 affected siblings); 405-kb deletion (assembly NCBI36/hg18) — g.(20898911)\_(21303509)del(CA)ins; 7.2-kb deletion (assembly NCBI36/hg18) — g.(20927077)\_(20934292)del(N205)ins. WT, wild-type sequence, i.e., sequence from which all exons of *SLCO1B1* and *SLCO1B3* could be amplified. Genotypes for all empty entries were wild-type in sequence and/or heterozygous or homozygous for the large haplotype R2-linked deletion as predicted.

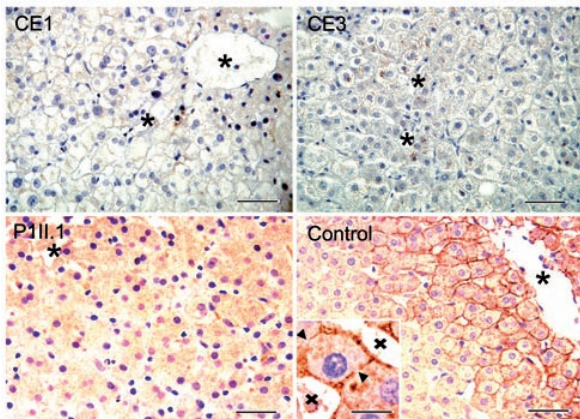
and *SLCO1B3* have been associated with increased serum UCB levels (36, 37). There was also a significant, nearly 2-fold increase in plasma UCB in the *Sco1a/1b*<sup>-/-</sup> mice, and this was partially reversed by both human OATP1B1 and OATP1B3 expression (Figure 3C).

It should be noted that UGT1A1-mediated glucuronidation may also occur in extrahepatic tissues, for instance, colon (38), and we cannot exclude that some of the bilirubin glucuronide observed in RS plasma has resulted from such extrahepatic glucuronidation, possibly enhanced by the increased plasma UCB levels. It seems unlikely, however, that all bilirubin glucuronide in RS subjects would derive from extrahepatic glucuronidation. This would require a complete block of hepatic UCB uptake (due to the OATP1B1 and OATP1B3 deficiency), but at the same time require efficient uptake of UCB into UGT1A1-containing extrahepatic cells (e.g., colonocytes) that do not normally express OATP1B1 and OATP1B3, and certainly not in Rotor subjects. If UCB transmembrane diffusion can do this efficiently, it is hard to see why this would not

mediate substantial uptake into the liver as well. Only if hepatic diffusion uptake is negligible (which seems physically unlikely) and an unknown efficient UCB uptake system would function in colonocytes (and not in liver), could one envisage such a situation. On balance, this seems rather implausible.

Elucidation of OATP1B1 and OATP1B3 deficiency as the cause of RS can also readily explain the other diagnostic traits of the disorder. Absence of OATP1B1/3-mediated liver uptake would cause the decreased plasma clearance of anionic diagnostic dyes such as indocyanine green and BSP, an excellent substrate of OATP1B1 and OATP1B3 (15), and the greatly reduced or delayed visualization of the liver by anionic cholestigraphic radiotracers such as <sup>99m</sup>Tc-HIDA and <sup>99m</sup>Tc-mebrofenin (5, 6). <sup>99m</sup>Tc-mebrofenin, for instance, is efficiently transported by both OATP1B1 and OATP1B3 (39).

The markedly increased urinary excretion of coproporphyrins, and the increased preponderance of isomer I over III in urine of RS subjects, could be simply explained by reduced (re)uptake of



**Figure 5**  
Liver expression of OATP1B proteins in RS subjects and control. With an anti-OATP1B/3 antibody, basolateral membrane immunostaining of hepatocytes in centrilobular areas was intense in control. Asterisks indicate central veins, arrowheads bile canaliculi, and crosses sinusoids. OATP1B proteins were not detectable in RS subjects CE1 (haplotype R1), CE3 (haplotype R2), and P1 II.1 (haplotype R3). Scale bars: 25  $\mu\text{m}$  (original magnification of CE1 and CE3,  $\times 400$ ; original magnification of P1 II.1 and control,  $\times 200$ ); inset: 5  $\mu\text{m}$  (original magnification,  $\times 1,000$ ).

these compounds into the liver, partly shifting the excretion route from hepatobiliary/fecal to urinary, especially for isomer I. Coproporphyrin I and III thus most likely are transported substrates of OATP1B1 and OATP1B3. Indeed, interaction of several porphyrins with OATP1B1 has recently been demonstrated (40).

Phenotypic abnormalities in RS subjects are surprisingly moderate. Perhaps OATP1B1 and OATP1B3 functions are partly taken over by other sinusoidal uptake transporters, such as OATP2B1. Nevertheless, since even reduced-activity OATP1B1 polymorphisms can result in life-threatening drug toxicities (20–24, 41, 42), such risks are likely increased substantially in RS subjects. Their evident jaundice, however, may have been a warning sign for physicians to prescribe drugs with caution.

The obligatory deficiency in two different, medium-sized genes explains the rarity of RS, with a roughly estimated frequency of about 1 in  $10^6$ , although it might be several-fold lower or higher in different populations. Complete deficiency of either OATP1B1 or OATP1B3 alone will occur much more frequently but will not cause jaundice. For instance, the p.R580X mutation in OATP1B1 occurred at an allele frequency of 0.008 (3 of 354) in a Japanese population (43), suggesting that about 1 in 14,000 individuals in this population would be homozygous for this full-deficiency mutant. Such individuals might demonstrate idiosyncratic hypersensitivity to OATP1B1 substrate drugs, including statins or irinotecan. Similarly, in the present study we identified a non-jaundiced individual homozygously deficient for *SLCO1B3* in our CNV screening of approximately 2,300 individuals, in line with a non-negligible incidence of fully OATP1B3-deficient individuals.

Some drugs, such as high-dose cyclosporine A, can transiently increase plasma levels of conjugated bilirubin without evoking other markers for liver damage (44, 45). Until now, such increases were thought to be primarily mediated by inhibition of ABCC2 as the main biliary excretion factor for bilirubin glucuronide.

However, given the insights from the present study, direct inhibition of OATP1B1 and/or OATP1B3 by the applied drug may be an additional or even the main cause of such drug-induced conjugated hyperbilirubinemia. This might for instance apply to cyclosporine A, rifampin, rifamycin SV, or other drugs that are established inhibitors of OATP1B proteins (23). Moreover, heterozygous carriers of the various full-deficiency mutations in OATP1B1/3 might be more susceptible to such inhibitory effects. This also applies to drug-drug interactions mediated through OATP1B1/3 inhibition.

The molecular mechanism we identified in RS may also underlie a similar disorder called hepatic uptake and storage syndrome, or conjugated hyperbilirubinemia type III (OMIM %237550) (46). This hypothesis can now be tested by mutational analysis of OATP1B1 and OATP1B3 in the only reported family to date. Furthermore, a mutant strain of Southdown sheep has also been described as displaying a similar hepatic uptake and storage syndrome (46), and it would not surprise us if these animals would likewise have a deficiency of one or more hepatic sinusoidal OATPs. The observation that mutant Southdown sheep, like the *Slco1a/1b*<sup>-/-</sup> mice (25), also display strongly reduced clearance of (unconjugated) cholic acid, but not of (conjugated) taurocholic acid (47), further supports this idea.

Collectively, our findings explain the genetic and molecular basis of RS. The demonstration of an *Abcc3*<sup>-</sup>, OATP1B1<sup>-</sup>, and OATP1B3-driven detoxification-enhancing liver-blood shuttling loop in mice and, by implication, most likely also in humans challenges the view of one-way excretion from blood through liver to bile of bilirubin and drugs detoxified by conjugation. Furthermore, the identified full-deficiency alleles of *SLCO1B1* and *SLCO1B3* may contribute to various “idiosyncratic” drug hypersensitivities.

## Methods

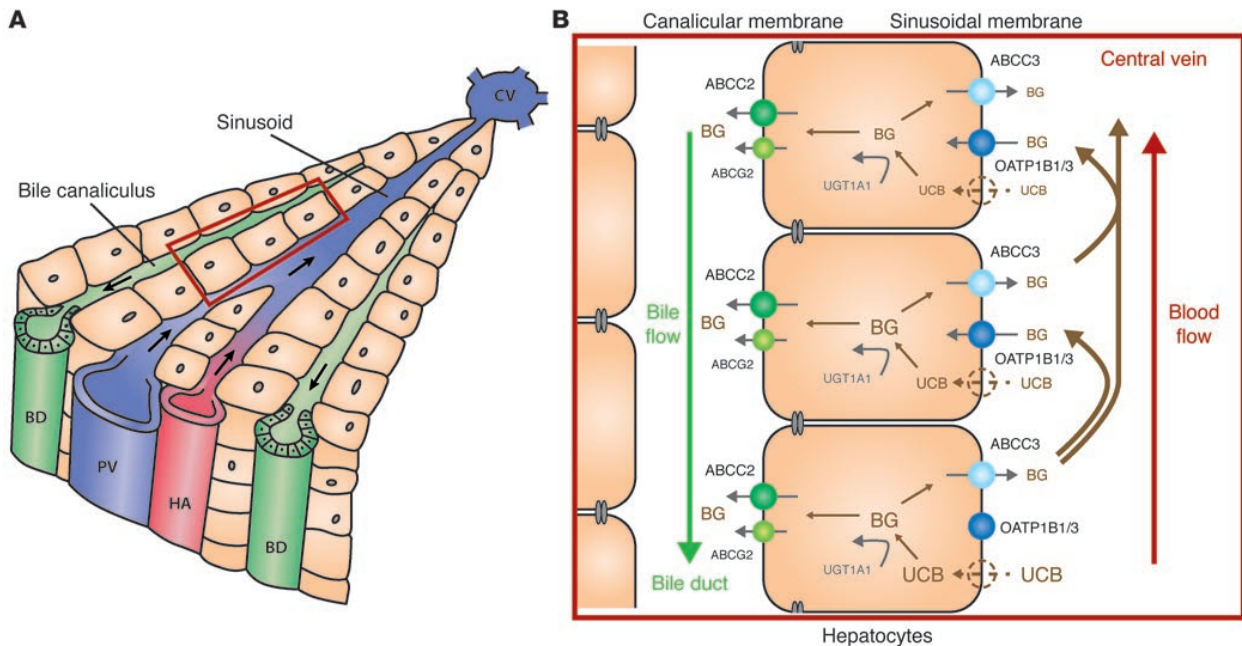
**Mouse strains and conditions.** Mice were housed and handled according to institutional guidelines complying with Dutch legislation. *Slco1a/1b*<sup>-/-</sup>, *Abcc2*<sup>-/-</sup>, *Abcc3*<sup>-/-</sup>, and *Abcc2*<sup>-/-</sup>*Abcc3*<sup>-/-</sup> mice have been described (25–27, 48). Human OATP1B1 transgenic mice have been described (29), and human OATP1B3 transgenic mice were generated in an analogous manner, using an apoE promoter to obtain liver-specific expression of the transgene. Each transgene was crossed back into an *Slco1a/1b*<sup>-/-</sup> background to obtain the corresponding humanized rescue strains. Routine mouse conditions and analyses of mouse samples are described in Supplemental Methods.

**Western blot analysis.** Isolation of crude membrane fractions from mouse liver, kidney, and small intestine and Western blotting were as described previously (29). For detection of *Abcc2* and *Abcc3* primary antibodies, M<sub>2</sub>III-5 (dilution 1:1,000) and M<sub>3</sub>-18 (dilution 1:25) were used, respectively. For detection of transgenic OATP1B1 and OATP1B3 in mouse liver, the rabbit polyclonal antibodies ESL and SKT, provided by D. Keppler (Deutsches Krebsforschungszentrum, Heidelberg, Germany) were used (17, 18).

**RNA isolation, cDNA synthesis, and RT-PCR.** RNA isolation from mouse liver, kidney, and small intestine and subsequent cDNA synthesis and RT-PCR were as described previously (49). Specific primers (QIAGEN) were used to detect expression levels of *Slco1a1*, *Slco1a4*, *Slco1a6*, *Slco1b2*, *Slco2b1*, *Slc10a1*, *Slc10a2*, *Abcc2-4*, *Abcb1a*, *Abcb1b*, *Abcb11*, *Abcg2*, *Osta*, *Ostb*, and *Ugt1a1*.

**Analysis of bilirubin in mouse plasma, bile, and urine.** Gallbladder cannulations and collection of bile and urine in male mice of the various strains ( $n = 4-7$ ) as well as bilirubin detection were as described (25, 50, 51). For details, see Supplemental Methods.

**RS families.** We examined 11 RS index subjects (8 probands, 3 siblings of probands) of 8 families and 21 clinically healthy members of 5 of these 8 families. Family members of 3 probands (CE1–CE3) were not available.



**Figure 6**

Hepatocyte hopping distributes the biliary excretion load of bilirubin glucuronides across the liver lobule. **(A)** Schematic of liver lobule. Hepatocytes are organized around portal tracts, with branches of the portal vein (PV), hepatic artery (HA), and bile ducts (BD). The PV and HA deliver nutrient- and oxygen-rich blood, respectively, which flows through the sinusoids toward the central vein (CV). Basolateral (sinusoidal) membranes of hepatocytes are flushed with perisinusoidal plasma. Bile flows in the opposite direction toward bile ducts through canaliculi lined by canalicular membranes of hepatocytes. **(B)** Hepatocyte hopping cycle. UCB enters the hepatocytes via passive diffusion and/or transporters, which may include OATP1B1 and/or OATP1B3 in non-Rotor subjects. Conjugation with glucuronic acid by UGT1A1 to bilirubin glucuronides (BG) takes place in endoplasmic reticulum. BG is secreted into bile mainly by ABCC2. ABCG2 also can contribute to this process. Even under physiological conditions, a substantial fraction of the intracellular BG is rerouted by ABCC3 to the blood, from which it can be taken up by downstream hepatocytes via OATP1B1/3 transporters. This flexible off-loading of BG to downstream hepatocytes prevents saturation of biliary excretion capacity in upstream hepatocytes. Relative type sizes of UCB and BG represent local concentrations. Schematic modified, with permission, from ref. 54.

Families CE1–CE4 are of mixed Central European descent by family report. Three families (A1–A3) are Saudi Arabs, and one family (P1) is from the Philippines. Central European families were ascertained at the Institute for Clinical and Experimental Medicine, Prague, and Saudi Arab and Filipino families at the Saudi Aramco Dhahran Health Center. Medical histories were obtained by referring consultants. Subjects CE1 and CE2 were reported as case 1 and case 2, respectively (14).

**ABCC2 mutation screening.** ABCC2 mutation screening was performed in 8 probands representing all studied families as described previously (14).

**Genotyping.** Genotyping was performed using Affymetrix GeneChip Mapping 6.0 Arrays (Affymetrix) according to the manufacturer’s protocol. Raw feature intensities were extracted from Affymetrix GeneChip Scanner 3000 7G images using GeneChip Control Console Software 2.01. Individual SNP calls were generated using Affymetrix Genotyping Console Software 3.02. Details of the experiment and individual genotyping data are available at the GEO repository (<http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE33733.

**Multipoint nonparametric and parametric linkage analysis.** Multipoint nonparametric and parametric linkage analysis along with determination of the most likely haplotypes was performed with version 1.1.2 of Merlin software (52). Parametric linkage was carried out assuming an autosomal recessive mode of inheritance with a 1.00 constant, age-independent penetrance, 0.00 phenocopy rate, and 0.0001 frequency of disease allele. Results

were visualized in version 1.032 of HaploPainter software (53) and in version 2.9.2 of R-project statistical software (<http://www.r-project.org/>).

**Homozygosity mapping.** Extended homozygosity regions were identified in Affymetrix Genotyping Console Software version 3.02 using the algorithm comparing values from the user’s sample set and SNP-specific distributions derived from a reference set of 200 ethnically diverse individuals. Distribution of extended homozygosity regions in affected and healthy individuals was analyzed and visualized using custom R-script.

**Copy number changes.** Copy number changes were identified in Affymetrix Genotyping Console Software version 3.02. Data from both SNP and copy number probes were used to identify copy number aberrations compared with built-in reference. Only regions larger than 10 kb containing at least 5 probes were reported.

**Quantitative PCR.** Quantitative PCR was carried out in duplicate on a LightCycler 480 System (Roche Applied Science). Data were analyzed by LightCycler 480 Software, release 1.5.0. Absolute quantification was used to determine copy number status of a given fragment in analyzed samples. Genomic positions of the analyzed fragments and control genes, corresponding primer sequences, and Universal ProbeLibrary probes used for amplification and quantitation are provided in Supplemental Table 4.

**Mutation analysis.** Long-range PCR products encompassing the genomic regions of deletion breakpoint boundaries were gel-purified and sequenced using a primer walking approach. DNA sequencing of PCR products and



genomic fragments covering 1 kb of the promoter regions and all of the exons, with their corresponding exon-intron boundaries, of *SLCO1B1*, *SLCO1B3*, and *SLCO1A2* was performed. For details, see Supplemental Methods. Confirmation and segregation of both identified copy number changes and missense mutations in the families, as well as frequency of the mutations in a control population of mixed European descent, were assessed by PCR, PCR-RFLP, and direct sequencing of corresponding genomic DNA fragments. For primer sequences, see Supplemental Table 4.

**Histology and immunohistochemistry.** Archival liver biopsy specimens were available from 5 unrelated RS index subjects (probands, families CE1, CE2, CE3, and P1; brother [A3 II.9] of proband from family A3). Sections of paraffin-embedded material (formalin or Carnoy solution fixative; 4–6  $\mu\text{m}$  thick) were routinely stained with hematoxylin and eosin and periodic acid–Schiff techniques. For OATP1B1, OATP1B3, and ABCC2 immunostaining, routine techniques were applied (see Supplemental Methods). OATP1B1 and OATP1B3 detection was performed with a primary mouse anti-OATP1B antibody (clone mMDQ, GeneTex; recognizing the N terminus of both OATP1B1 and OATP1B3), 1:100 dilution, overnight at 4°C (31).

**Statistics.** One-way ANOVA followed by Tukey's multiple comparison test was used to assess statistical significance of differences between data sets. Results are presented as mean  $\pm$  SD. Differences were considered statistically significant when  $P$  was less than 0.05.

**Study approval.** All mouse studies were ethically reviewed and carried out in accordance with European directive 86/609/EEC and Dutch legislation and the GlaxoSmithKline policy on the Care, Welfare and Treatment of Laboratory Animals. Experiments were approved by the Animal Experimentation Committee (DEC) of the Netherlands Cancer Institute. Invest-

igations involving humans were approved by the Institutional Review Boards of the Institute for Clinical and Experimental Medicine, Prague, Czech Republic, and the Saudi Aramco Dhahran Health Centre, with written informed consent received from participants or their guardians, and conducted according to Declaration of Helsinki principles.

## Acknowledgments

The human study was supported in part by the Ministry of Education of the Czech Republic (projects MSM0021620806 and 1M6837805002) and by the Institute for Clinical and Experimental Medicine (MZO 00023001). The mouse work was supported in part by grant S2918 from GlaxoSmithKline and grant NKI 2007-3764 from the Dutch Cancer Society. The authors thank L. Budišová and M. Boučková for technical assistance and L. Vitek, M. Mikulecký, J. Horák, and A. Šuláková for referring patients CE1–CE4.

Received for publication June 16, 2011, and accepted in revised form November 30, 2011.

Address correspondence to: Alfred H. Schinkel, Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. Phone: 31.20.5122046; Fax: 31.20.6961383; E-mail: a.schinkel@nki.nl. Or to: Milan Jirsa, Department of Experimental Medicine, Institute for Clinical and Experimental Medicine, Videnská 1958/9, 140 00 Prague 4 – Krč, Czech Republic. Phone: 420.261362773; Fax: 420.241721666; E-mail: miji@ikem.cz.

1. Chowdhury JR, Chowdhury NR, Jansen PLM. Bilirubin metabolism and its disorders. In: Boyer TD, Wright TL, Manns MP, Zakim D, eds. *Zakim and Boyer's Hepatology. A Textbook of Liver Diseases*. Vol. 2. Philadelphia, Pennsylvania, USA: Saunders Elsevier; 2006:1449–1474.
2. Chowdhury JR, Wolkoff AW, Chowdhury NR, Arias IM. Hereditary jaundice and disorders of bilirubin metabolism. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The Metabolic and Molecular Bases of Inherited Disease*. Vol. 2. New York, New York, USA: McGraw Hill; 2001:3063–3101.
3. Rotor AB, Manahan L, Florentin A. Familial nonhemolytic jaundice with direct van den Berg reaction. *Acta Med Phil*. 1948;5:37–49.
4. Wolpert E, Pascasio FM, Wolkoff AW, Arias IM. Abnormal sulfbromophthalein metabolism in Rotor's syndrome and obligate heterozygotes. *N Engl J Med*. 1977;296(19):1099–1101.
5. Bar-Meir S, Baron J, Seligson U, Gottesfeld F, Levy R, Gilat T. 99mTc-HIDA cholescintigraphy in Dubin-Johnson and Rotor syndromes. *Radiology*. 1982; 142(3):743–746.
6. LeBouthillier G, Morais J, Picard M, Picard D, Chartrand R, Pommier G. Scintigraphic aspect of Rotor's disease with Technetium-99m-mebrofenin. *J Nucl Med*. 1992;33(8):1550–1551.
7. Kartenbeck J, Leuschner U, Mayer R, Keppler D. Absence of the canalicular isoform of the MRP gene-encoded conjugate export pump from the hepatocytes in Dubin-Johnson syndrome. *Hepatology*. 1996;23(5):1061–1066.
8. Paulusma CC, et al. A mutation in the human canalicular multispecific organic anion transporter gene causes the Dubin-Johnson syndrome. *Hepatology*. 1997;25(6):1539–1542.
9. Nowicki MJ, Poley JR. The hereditary hyperbilirubinemia. *Baillieres Clin Gastroenterol*. 1998; 12(2):355–367.
10. Strassburg CP. Hyperbilirubinemia syndromes (Gilbert-Meulengracht, Crigler-Najjar, Dubin-Johnson, and Rotor syndrome). *Best Pract Res Clin Gastroenterol*. 2010;24(5):555–571.
11. Wolkoff AW, Wolpert E, Pascasio FN, Arias IM. Rotor's syndrome. A distinct inheritable pathophysiological entity. *Am J Med*. 1976;60(2):173–179.
12. König J, Rost D, Cui Y, Keppler D. Characterization of the human multidrug resistance protein isoform MRP3 localized to the basolateral hepatocyte membrane. *Hepatology*. 1999;29(4):1156–1163.
13. Lee YM, et al. Identification and functional characterization of the natural variant MRP3-Arg1297His of human multidrug resistance protein 3 (MRP3/ABCC3). *Pharmacogenetics*. 2004;14(4):213–223.
14. Hrebicek M, et al. Rotor-type hyperbilirubinemia has no defect in the canalicular bilirubin export pump. *Liver Int*. 2007;27(4):485–491.
15. Hagenbuch B, Gui C. Xenobiotic transporters of the human organic anion transporting polypeptide (OATP) family. *Xenobiotica*. 2008;38(7–8):778–801.
16. Hagenbuch B, Meier PJ. Organic anion transporting polypeptides of the OATP/SLC21 family: phylogenetic classification as OATP/SLCO superfamily, new nomenclature and molecular/functional properties. *Pflügers Arch*. 2004;447(5):653–665.
17. König J, Cui Y, Nies AT, Keppler D. A novel human organic anion transporting polypeptide localized to the basolateral hepatocyte membrane. *Am J Physiol Gastrointest Liver Physiol*. 2000;278(1):G156–G164.
18. König J, Cui Y, Nies AT, Keppler D. Localization and genomic organization of a new hepatocellular organic anion transporting polypeptide. *J Biol Chem*. 2000;275(30):23161–23168.
19. Abe T, et al. LST-2, a human liver-specific organic anion transporter, determines methotrexate sensitivity in gastrointestinal cancers. *Gastroenterology*. 2001;120(7):1689–1699.
20. Takane H, et al. Life-threatening toxicities in a patient with UGT1A1\*6/\*28 and SLC01B1\*15/\*15 genotypes after irinotecan-based chemotherapy. *Cancer Chemother Pharmacol*. 2009; 63(6):1165–1169.
21. Link E, et al. SLC01B1 variants and statin-induced myopathy – a genome-wide study. *N Engl J Med*. 2008; 359(8):789–799.
22. Treviño LR, et al. Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J Clin Oncol*. 2009;27(35):5972–5978.
23. Kalliokoski A, Niemi M. Impact of OATP transporters on pharmacokinetics. *Br J Pharmacol*. 2009; 158(3):693–705.
24. König J, Seithel A, Gradhand U, Fromm MF. Pharmacogenomics of human OATP transporters. *Naunyn Schmiedebergs Arch Pharmacol*. 2006; 372(6):432–443.
25. Van de Steeg E, et al. Organic anion transporting polypeptide 1a/1b-knockout mice provide insights into hepatic handling of bilirubin, bile acids and drugs. *J Clin Invest*. 2010;120(8):2942–2952.
26. Vlaming ML, et al. Carcinogen and anticancer drug transport by Mrp2 in vivo: studies using Mrp2 (Abcc2) knockout mice. *J Pharmacol Exp Ther*. 2006; 318(1):319–327.
27. Vlaming ML, et al. Impact of Abcc2 (Mrp2) and Abcc3 (Mrp3) on the in vivo elimination of methotrexate and its main toxic metabolite 7-hydroxymethotrexate. *Clin Cancer Res*. 2008;14(24):8152–8160.
28. Rius M, Nies AT, Hummel-Eisenbeiss J, Jedlitschky G, Keppler D. Cotransport of reduced glutathione with bile salts by MRP4 (ABCC4) localized to the basolateral hepatocyte membrane. *Hepatology*. 2003; 38(2):374–384.
29. Van de Steeg E, et al. Methotrexate pharmacokinetics in transgenic mice with liver-specific expression of human OATP1B1 (SLCO1B1). *Drug Metab Dispos*. 2009;37(1):1–5.
30. Cui Y, König J, Leier I, Buchholz U, Keppler D. Hepatic uptake of bilirubin and its conjugates by the human organic anion transporter SLC21A6. *J Biol Chem*. 2001;276(13):9626–9630.
31. Cui Y, et al. Detection of the human organic anion transporters SLC21A6 (OATP2) and SLC21A8 (OATP8) in liver and hepatocellular carcinoma. *Lab Invest*. 2003;83(4):527–538.
32. Zhou S, Chan E, Duan W, Huang M, Chen YZ.



- Drug bioactivation, covalent binding to target proteins and toxicity relevance. *Drug Metab Rev.* 2005; 37(1):41–213.
33. Zucker SD, Goessling W, Hoppin AG. Unconjugated bilirubin exhibits spontaneous diffusion through model lipid bilayers and native hepatocyte membranes. *J Biol Chem.* 1999;274(16):10852–10862.
  34. Kawasaki H, Kimura N, Irisa T, Hirayama C. Dye clearance studies in Rotor's syndrome. *Am J Gastroenterol.* 1979;71(4):380–388.
  35. Fedeli G, et al. Impaired clearance of cholephilic anions in Rotor syndrome. *Z Gastroenterol.* 1983; 21(5):228–233.
  36. Zhang W, et al. OATP1B1 polymorphism is a major determinant of serum bilirubin level but not associated with rifampicin-mediated bilirubin elevation. *Clin Exp Pharmacol Physiol.* 2007;34(12):1240–1244.
  37. Sanna S, et al. Common variants in the SLCO1B3 locus are associated with bilirubin levels and unconjugated hyperbilirubinemia. *Hum Mol Genet.* 2009; 18(14):2711–2718.
  38. Strassburg CP, Manns MP, Tukey RH. Expression of the UDP-glucuronosyltransferase 1A locus in human colon. Identification and characterization of the novel extrahepatic UGT1A8. *J Biol Chem.* 1998; 273(15):8719–8726.
  39. Ghibellini G, Leslie EM, Pollack GM, Brouwer KL. Use of tc-99m mebrofenin as a clinical probe to assess altered hepatobiliary transport: integration of in vitro, pharmacokinetic modeling, and simulation studies. *Pharm Res.* 2008;25(8):1851–1860.
  40. Campbell SD, Lau WF, Xu JJ. Interaction of porphyrins with human organic anion transporting polypeptide 1B1. *Chem Biol Interact.* 2009;182(1):45–51.
  41. Morimoto K, Oishi T, Ueda S, Ueda M, Hosokawa M, Chiba K. A novel variant allele of OATP-C (SLCO1B1) found in a Japanese patient with pravastatin-induced myopathy. *Drug Metab Pharmacokin.* 2004;19(6):453–455.
  42. Tirona RG, Kim RB. Pharmacogenomics of organic anion-transporting polypeptides (OATP). *Adv Drug Deliv Rev.* 2002;54(10):1343–1352.
  43. Kim SR, et al. Genetic variations and frequencies of major haplotypes in SLCO1B1 encoding the transporter OATP1B1 in Japanese subjects: SLCO1B1\*17 is more prevalent than \*15. *Drug Metab Pharmacokin.* 2007;22(6):456–461.
  44. Yahanda AM, et al. Phase I trial of etoposide with cyclosporine as a modulator of multidrug resistance. *J Clin Oncol.* 1992;10(10):1624–1634.
  45. List AF, et al. Phase I/II trial of cyclosporine as a chemotherapy-resistance modifier in acute leukemia. *J Clin Oncol.* 1993;11(9):1652–1660.
  46. Dhumeaux D, Berthelot P. Chronic hyperbilirubinemia associated with hepatic uptake and storage impairment. A new syndrome resembling that of mutant southdown sheep. *Gastroenterology.* 1975; 69(4):988–993.
  47. Engelking LR, Gronwall R. Bile acid clearance in sheep with hereditary hyperbilirubinemia. *Am J Vet Res.* 1979;40(9):1277–1280.
  48. Zelcer N, et al. Mice lacking Mrp3 (Abcc3) have normal bile salt transport, but altered hepatic transport of endogenous glucuronides. *J Hepatol.* 2006; 44(4):768–775.
  49. Van Waterschoot RA, et al. Midazolam metabolism in cytochrome P450 3A knockout mice can be attributed to up-regulated CYP2C enzymes. *Mol Pharmacol.* 2008;73(3):1029–1036.
  50. Van Herwaarden AE, et al. The breast cancer resistance protein (Bcrp1/Abcg2) restricts exposure to the dietary carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine. *Cancer Res.* 2003; 63(19):6447–6452.
  51. Spivak W, Carey MC. Reverse-phase h.p.l.c. separation, quantification and preparation of bilirubin and its conjugates from native bile. Quantitative analysis of the intact tetrapyrroles based on h.p.l.c. of their ethyl anthranilate azo derivatives. *Biochem J.* 1985;225(3):787–805.
  52. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin — rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002;30(1):97–101.
  53. Thiele H, Nurnberg P. HaploPainter: a tool for drawing pedigrees with complex haplotypes. *Bioinformatics.* 2005;21(8):1730–1732.
  54. Van de Steeg E, Iusuf D, Schinkel AH. Physiological and pharmacological functions of OATP1A/1B transporters: insights from knockout and transgenic mice. In: Van de Steeg E. *Physiological and pharmacological functions of OATP1A/1B transporters.* PhD thesis, University of Utrecht, Enschede, the Netherlands: Gildeprint drukkerijen; 2010:9–37.

Research article

Open Access

## Development of a human mitochondrial oligonucleotide microarray (h-MitoArray) and gene expression analysis of fibroblast cell lines from 13 patients with isolated F<sub>1</sub>F<sub>0</sub> ATP synthase deficiency

Alena Čížková<sup>1,2,5</sup>, Viktor Stránecký<sup>1,2</sup>, Robert Ivánek<sup>1,2,4</sup>,  
Hana Hartmannová<sup>1,2</sup>, Lenka Nosková<sup>2</sup>, Lenka Piherová<sup>1,2</sup>,  
Markéta Tesařová<sup>1,3</sup>, Hana Hansíková<sup>1,3</sup>, Tomáš Honzík<sup>3</sup>, Jiří Zeman<sup>1,3</sup>,  
Petr Divina<sup>4</sup>, Andrea Potocká<sup>1,5</sup>, Jan Paul<sup>1,5</sup>, Wolfgang Sperl<sup>6</sup>,  
Johannes A Mayr<sup>6</sup>, Sara Seneca<sup>7</sup>, Josef Houšťek<sup>1,5</sup> and Stanislav Kmoch\*<sup>1,2</sup>

Address: <sup>1</sup>Center for Applied Genomics, 1st Faculty of Medicine, Charles University, Prague, Czech Republic, <sup>2</sup>Institute of Inherited Metabolic Disorders, 1st Faculty of Medicine, Charles University, Prague, Czech Republic, <sup>3</sup>Department of Pediatrics, 1st Faculty of Medicine, Charles University, Prague, Czech Republic, <sup>4</sup>Institute of Molecular Genetics, Academy of Science of the Czech Republic, Prague, Czech Republic, <sup>5</sup>Department of Bioenergetics, Institute of Physiology, Academy of Science of the Czech Republic, Prague, Czech Republic, <sup>6</sup>Department of Pediatrics, Paracelsus Medical University, Salzburg, Austria and <sup>7</sup>Center of Medical Genetics, Free University Brussels, Brussels, Belgium

Email: Alena Čížková - acizk@LF1.cuni.cz; Viktor Stránecký - vstra@LF1.cuni.cz; Robert Ivánek - ivanek@img.cas.cz;  
Hana Hartmannová - hhart@LF1.cuni.cz; Lenka Nosková - Inosk@LF1.cuni.cz; Lenka Piherová - Lenka.Piherova@LF1.cuni.cz;  
Markéta Tesařová - Marketa.Tesarova@LF1.cuni.cz; Hana Hansíková - HHansikova@seznam.cz; Tomáš Honzík - HonzikT@seznam.cz;  
Jiří Zeman - jzem@LF1.cuni.cz; Petr Divina - divina@img.cas.cz; Andrea Potocká - potockaa@biomed.cas.cz; Jan Paul - paulj@biomed.cas.cz;  
Wolfgang Sperl - w.sperl@salk.at; Johannes A Mayr - h.mayr@salk.at; Sara Seneca - sara.seneca@az.vub.ac.be;  
Josef Houšťek - houstek@biomed.cas.cz; Stanislav Kmoch\* - skmoch@LF1.cuni.cz

\* Corresponding author

Published: 25 January 2008

Received: 6 September 2007

BMC Genomics 2008, 9:38 doi:10.1186/1471-2164-9-38

Accepted: 25 January 2008

This article is available from: <http://www.biomedcentral.com/1471-2164/9/38>

© 2008 Čížková et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** To strengthen research and differential diagnostics of mitochondrial disorders, we constructed and validated an oligonucleotide microarray (h-MitoArray) allowing expression analysis of 1632 human genes involved in mitochondrial biology, cell cycle regulation, signal transduction and apoptosis. Using h-MitoArray we analyzed gene expression profiles in 9 control and 13 fibroblast cell lines from patients with F<sub>1</sub>F<sub>0</sub> ATP synthase deficiency consisting of 2 patients with mt9205ΔTA microdeletion and a genetically heterogeneous group of 11 patients with not yet characterized nuclear defects. Analysing gene expression profiles, we attempted to classify patients into expected defect specific subgroups, and subsequently reveal group specific compensatory changes, identify potential phenotype causing pathways and define candidate disease causing genes.

**Results:** Molecular studies, in combination with unsupervised clustering methods, defined three subgroups of patient cell lines – M group with mtDNA mutation and N1 and N2 groups with nuclear defect. Comparison of expression profiles and functional annotation, gene enrichment and pathway analyses of differentially expressed genes revealed in the M group a transcription profile suggestive of synchronized suppression of mitochondrial biogenesis and G1/S arrest. The N1 group showed elevated expression of complex I and reduced expression of complexes III, V, and V-type

ATP synthase subunit genes, reduced expression of genes involved in phosphorylation dependent signaling along MAPK, Jak-STAT, JNK, and p38 MAP kinase pathways, signs of activated apoptosis and oxidative stress resembling phenotype of premature senescent fibroblasts. No specific functionally meaningful changes, except of signs of activated apoptosis, were detected in the N2 group. Evaluation of individual gene expression profiles confirmed already known *ATP6/ATP8* defect in patients from the M group and indicated several candidate disease causing genes for nuclear defects.

**Conclusion:** Our analysis showed that deficiency in the ATP synthase protein complex amount is generally accompanied by only minor changes in expression of ATP synthase related genes. It also suggested that the site (mtDNA vs nuclear DNA) and the severity (ATP synthase content) of the underlying defect have diverse effects on cellular gene expression phenotypes, which warrants further investigation of cell cycle regulatory and signal transduction pathways in other OXPHOS disorders and related pharmacological models.

---

## Background

Mitochondria generate most of the cellular energy in the form of ATP, regulate cellular redox state, cytosolic concentration of  $Ca^{2+}$ , are a source of endogenous reactive oxygen species, and integrate many of the signals for initiating apoptosis. By means of retrograde signaling mitochondria communicate all these events to the nucleus and thus modulate nuclear gene expression and cell cycle.

In humans, mitochondrial dysfunction leads to a vast array of pathologies, and hundreds of diseases result from various defects of mitochondrial biogenesis and maintenance, respiratory chain complexes, or individual mitochondrial proteins [1].

The most frequent group of mitochondrial diseases results from genetic defects of the oxidative phosphorylation system (OXPHOS) [2]. OXPHOS defects form a highly diverse group of diseases that affect primarily energy demanding tissues, such as the central nervous system, heart, and skeletal muscles. Their prevalence is estimated as at least 1:5000 [3]. About half of the OXPHOS defects result from mtDNA mutations [4]. Diseases resulting from mtDNA mutations usually show maternal mode of inheritance and variable penetrance of the disease phenotype, reflecting levels of mtDNA heteroplasmy and threshold effects in affected tissues. Remaining OXPHOS defects result from mutations in genes encoded in nuclear DNA. The majority of the nuclear encoded diseases are inherited as autosomal recessive traits and produce severe and usually fatal phenotypes in infants [5]. Up to now, mutations in approximately 50 nuclear genes have been identified, but most of nuclear genetic defects remain unknown and can involve any of approximately 1000 mitochondria related genes [6]. These genes play an essential role in the assembly or maintenance of individual OXPHOS complexes, in maintenance of mtDNA integrity, and mitochondrial biogenesis.

Diagnostic process of OXPHOS defects requires a combination of biochemical, enzymatic, immunohistochemical and molecular biology methods. To distinguish between isolated and combined OXPHOS deficiencies, the diagnostic process starts with measurements of selected mitochondrial enzyme activities and activities of individual OXPHOS complexes. The diagnostic procedure continues with analysis of OXPHOS complex protein composition. The origin of the molecular defect (mtDNA vs ncDNA) is often apparent from clinical presentation and family history. If not, it can be determined by using transmitochondrial cybrid cell analysis. Final steps in the diagnosis represent mutation analysis either in mtDNA or in nuclear encoded candidate genes in accordance with observed clinical and biochemical phenotypes. The diagnostic process is experimentally demanding and time-consuming and in majority of cases leads only to biochemical diagnosis. The molecular basis of the disease, especially in nuclear encoded defects, mostly remains unknown.

Identification of nuclear gene defects in OXPHOS deficiencies requires combination of positional cloning, functional complementation, and candidate gene analysis. Application of these "standard" procedures is however greatly hampered by limited number of affected patients, complexity and overlap of observed diseases phenotypes, difficulties in measurement of biochemical phenotypes *in vitro*, and by the existence of many candidate nuclear genes [7].

Another method having potential to contribute to differential diagnosis and research of OXPHOS defects relies on gene expression profiling. This type of analysis has a potential to provide information on putative diseases subtypes [8], suggest candidate disease causing genes [7,9,10], reveal pathogenic mechanism of the disease [11] and define specific gene expression profiles usable in future disease class prediction [12].

One of the possibilities for long term studies selectively targeted to mitochondrial gene expression analysis involves development and application of a focused microarray interrogating set of all known and hypothetical human mitochondrial genes, and several human mitochondria focused microarrays were prepared recently [13-16]. All these microarray platforms were based on PCR amplified probes prepared from selected IMAGE consortium cDNA clones. This approach however poses a number of technical obstacles. High rate of miss-annotation and contamination in the commercially distributed subset of the IMAGE Consortium cDNA clone collection [17] requires resequencing of individual clone inserts, and subsequent PCR preparation of individual probes is laborious and time consuming. Given these difficulties it has become very attractive to use sets of oligonucleotide probes that obviate much of the probe preparation work. Since the yield of long oligonucleotides has improved and cost has fallen recently, the current trend in preparation of low density, tailor-made microarrays favours oligonucleotide microarrays [18].

In this paper, we describe development and validation of a focused oligonucleotide microarray for expression profiling of human mitochondria related genes – "h-MitoArray" and report gene expression analysis of fibroblast cell lines from 9 controls and 13 patients with isolated deficiency of F<sub>1</sub>F<sub>0</sub> ATP synthase caused either by microdeletion of mtDNA encoded ATP6 gene [19,20] or by mutation of unknown nuclear genes [21,22].

## Results

### **Microarray design and preparation**

For microarray preparation we selected genes coding for known or predicted mitochondrial proteins, genes known to be involved in cell cycle growth and regulation, and genes involved in apoptosis and free radical metabolism.

The final set contained 1632 genes, of which 992 are "mitochondrial" genes, 42 lysosomal genes, 277 genes are associated with apoptosis, and 321 are "oncogenes". For normalization and background correction we included 146 human "housekeeping" genes, 10 *Arabidopsis* genes and 32 blanks. Full list of selected genes with corresponding symbols, accession and LocusLink codes is provided [see Additional file 1]. Functional annotation of selected genes and comparison of the gene content against whole human genome set is provided [see Additional file 2].

### **Microarray validation**

Hybridization properties and performance of designed oligonucleotide probes and control features placed on h-MitoArray were tested by hybridization of fluorescently labeled panomers and fluorescently labeled cDNA prepared from a pool of total RNA isolated from several cell

lines (test RNA). Gene expression signal was detected in > 77% of 1820 elements when fluorescently labeled cDNA pool was used.

Following comparison of various labeling strategies and optimization of hybridization conditions a series of self-to-self experiments was performed using test RNA. Data analysis showed acceptable reproducibility with Pearson correlation coefficient ranging 0.987 – 0.991.

### **Gene expression analysis in ATP synthase deficient fibroblasts**

Fluorescent cDNA probes labeled with Cy5 were prepared from 13 patient and 9 control cell lines and were hybridized to common reference cDNA probe labeled with Cy3 in two technical replicates for each sample. Following data acquisition, transformation, normalization and replicate averaging, gene expression signals were obtained for 1264 genes. Ratios of Log<sub>2</sub> sample gene intensities against Log<sub>2</sub> common reference gene intensities (M) were calculated and are provided [see Additional file 3]. Calculated ratios of individual patient Log<sub>2</sub> gene intensities against the Log<sub>2</sub> of average of controls gene intensities (M) are provided [see Additional file 4].

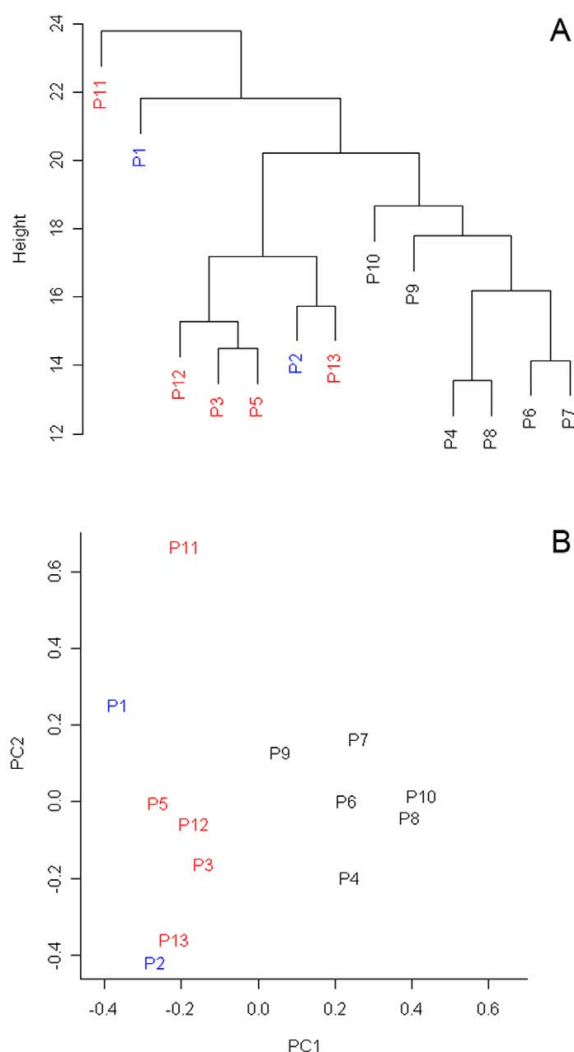
### **Principal component analysis**

To assess overall data quality and visualize relations between analyzed samples, we removed from the original data set 47 genes showing low expression variability (based on criteria  $|M_{min}; M_{max}| \leq 0.58$ , less than 1-fold change across all the samples) and subjected resulting data set to principal components analysis. Visual inspection of resulting plots showed no gross differences among the individual samples but suggested that several samples from nuclear defect patients group might be distinct from the others (Figure 1B).

### **Hierarchical clustering**

To reveal gene expression changes, survey variation in patient samples, and better interpret the results of principal component analysis (PCA), gene expression signals from individual patient samples were compared to average of gene expression signals from all controls. Hierarchical clustering of all gene ratios across all patient samples was performed using Euclidean distance metrics and average linkage clustering algorithm. Resulting expression map (not shown) and sample dendrogram shown in Figure 1A defined, in agreement with previous PCA, two distinct subgroups of patients with nuclear defect, (N1 and N2 group) which were considered in subsequent gene expression comparisons and functional evaluations.





**Figure 1**  
**Results of unsupervised clustering methods. A)** Dendrogram resulting from two-dimensional hierarchical clustering of all genes across all patient samples performed using Euclidean distance metrics and average linkage clustering algorithm. **B)** Two-dimensional PCA plot of all expression data showing the separation of samples forming N1 group. Patients from M, N1 and N2 groups are shown in blue, black and red, respectively.

#### Overall gene expression changes triggered by ATP synthase deficiency

Comparison of gene expression patterns between ATP synthase deficient and control fibroblast cell lines was performed in R statistical environment as described in methods. This analysis revealed 78 genes to be differentially expressed at adjusted  $P < 0.01$  significance level [see

Additional file 5]. Detailed inspection of expression map and evaluation of individual gene expression profiles showed, that although defined as significant, majority of the identified genes was not uniformly altered across all the patient samples.

#### Identification of subgroup specific gene expression profiles

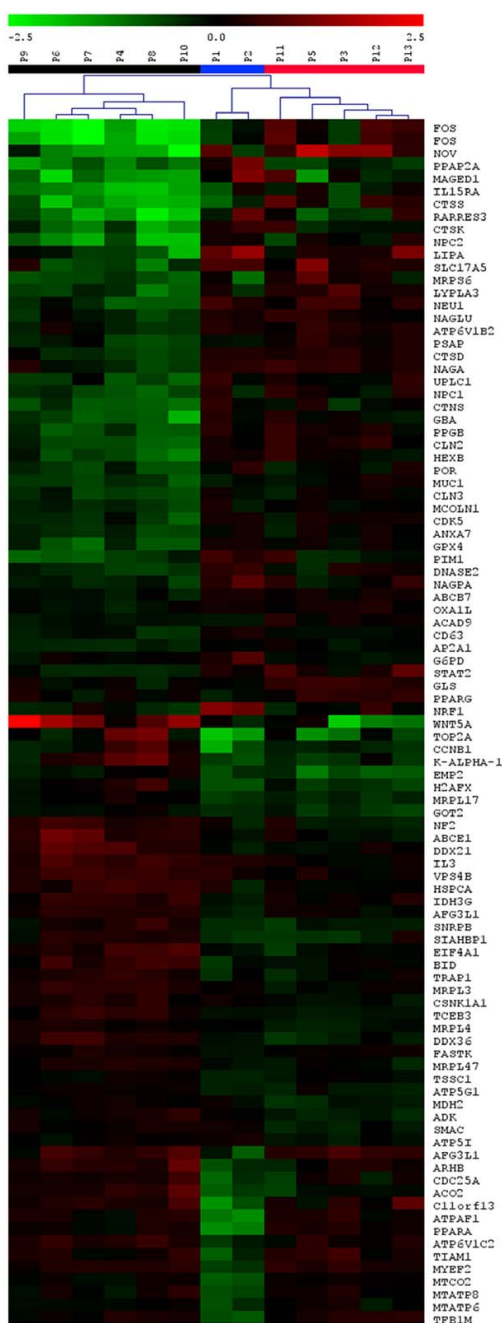
To identify the subgroup specific gene expression changes, the subgroups of patients defined by a mutation of the *MTATP6* gene of the mtDNA (M group), PCA and hierarchical clustering (N1 and N2 groups) were compared. ANOVA analysis performed in MeV software revealed 97 genes to be differentially expressed at unadjusted  $P < 0.01$  (Figure 2), [see Additional file 6].

Inspection of resulting data showed that the M group was specifically characterized by reduced expression of mitochondria encoded ATP synthase subunit genes *MTATP6*, *MTATP8*, nuclear encoded ATP synthase assembly factor *ATPAF1*, cytochrome *c* oxidase subunit II gene *MTCO2*, mitochondrial transcription factors *TFAM* and *TFB1M*, peroxisome proliferator-activated receptor alpha (*PPARA*), regulatory genes *H2AFX*, *CCNB1*, *C11orf13* (*RASSF7*), *TPR* and *ACO2*. This was accompanied by induction of *NRF1*.

The N1 group was characterized by reduced expression of genes involved in cell growth, differentiation and transduction pathways (*FOS*, *NOV*, *MAGED1*, *IL15RA*, *RARRES3*, *CTSK*, *UPLC1*, *PIM1*), mitochondrial proteosynthesis (*MRPS5*), lysosomal metabolism and function (cathepsins *S*, *K* and *D*, *GBA*, *PPGB*, *NPC*, *CLN2*, *FUCA1*, *HEXB*), protein transport (*AP2A1*), protein phosphorylation (*CDK5*, *PPAP2A*), hydrolase activity (*LIPA*, *LYPLA3*), reactive oxygen species metabolism (*GPX4*) and membrane transport (*SLC17A5*, *CTNS*). This was accompanied by elevated expression of several cell cycle regulatory genes such *WNT5A*, *IL3*, *CSNK1A1*, *BID*, *EIF4A1*, and *ACO2*.

The N2 group showed reduced expression of *WNT5A*, *EMP2*, *ADK*, *MDH2*, *SMAC* and elevated expression of *PPARG* and *GLS*. Extent and range of detected changes were much less than that observed in M and N1 groups.

Following ANOVA analysis, which revealed only inter-group specific differences, a list of group specific gene expression changes was obtained by comparison between defined patient subgroups and controls in R statistical environment as described in Methods. The analysis revealed 61, 215, and 54 genes to be differentially expressed at adjusted  $P < 0.01$  in the M, N1 and N2 groups, respectively. In addition to the above mentioned genes revealed by ANOVA, we found in the M group elevated expression of *mitofusin* and coordinately reduced



**Figure 2**  
**Differentially expressed genes defined by ANOVA analysis.** Heatmap of genes detected as differentially expressed between defined patient groups using ANOVA analysis and unadjusted  $P < 0.01$  significance level. The results are shown as  $\text{Log}_2$  ratio of relative gene expression signal in each patient sample to average of this of control samples. Ratio values are represented as the pseudo-color whose scale is shown in corresponding lookup picture.

expression of genes regulating G1/S phase transitions (*E2F1*, *MYC*, *CDC2*, *GAS1*, *CCNA2*, *CCNB*, *CDK2*, *CDC25A*, *PCNA*), thymidine metabolism (*TK*, *TYMS*) and DNA topology (*H2AFX*, *TOP2*, *LMN2*). In the N1 group, we observed reduced expression of genes regulating cell growth and signaling (*JUNB*, *MAPK3*, *WT1*, *CEBPA*, *CEBPB*) and lysosomal metabolism. We found elevated expression in genes involved in apoptosis (*FAS*, *CYTC*, *SMAC*, *IGFBP3*). In the N2 group, we found signs of started apoptosis (*SMAC*, *CASP8*). Group specific gene lists with expression values and corresponding P-statistics are provided [see Additional file 7, 8, 9].

### Biological consequences of identified gene expression changes

To reveal biological consequences and to identify pathways potentially involved in the pathogenesis of the studied defects, we extracted from original expression data for each of the three defined groups all genes found to be differentially expressed at unadjusted  $P < 0.05$  and showing expression change  $|M| > 0.2$ . Resulting expression datasets were uploaded into the DAVID database [23] and gene enrichment analysis was performed against h-MitoArray gene list. Results are provided in Table 1.

As the enrichment analysis suggested group specific dysregulation of several metabolic and signaling pathways, we further uploaded identical datasets into KEGGArray software (KEGG pathway databases – Kyoto Encyclopedia of Genes and Genomes) and inspected gene expression changes in all the indicated pathways.

In the M group, generally reduced expression was observed in cell cycle regulation (Figure 3), Krebs cycle (*OGDH*, *IDH1*, *ACO2*) and gluconeogenesis (*ALDOA*, *LDHA*, *PGAM1*) pathways. With an exception of *MTATP6*, *MTATP8* and *MTCOX2*, no multiple changes in OXPHOS system, valine, leucine, isoleucine, lysine,  $\beta$ -oxidation and MAP kinase pathway were observed. Reduced expression of *CytC* and *NFKB* and elevated expression of *FAS* were detected in the apoptotic pathway. In contrast to the N1 group, elevated expression of genes involved in N-glycan and heparan sulfate was detected.

In the N1 group, the analysis revealed elevated expression of several complex I subunit genes (*ND1*, *ND2*, *ND4*, *ND4L*, *Ndufs1*, *Ndufv2*, *Nufa9*, *Ndufb9* and *Ndufa10*) and generally reduced expression of complex IV (*COX4*, *COX5A*, *COX6A*, *COX6B*, *COX6C* and *COX15*) and complex V subunit genes (*ATPAF1*, *ATP5G2*) in OXPHOS system. Generally reduced expression of V-type ATP synthase subunit genes was observed. Elevated transcription activity was found along valine, leucine, isoleucine, lysine and fatty acid  $\beta$ -oxidation pathways. Elevated expression of *FGF*, *FGFR*, *Ras* and *PKC* and reduced expression of *Raf1*,

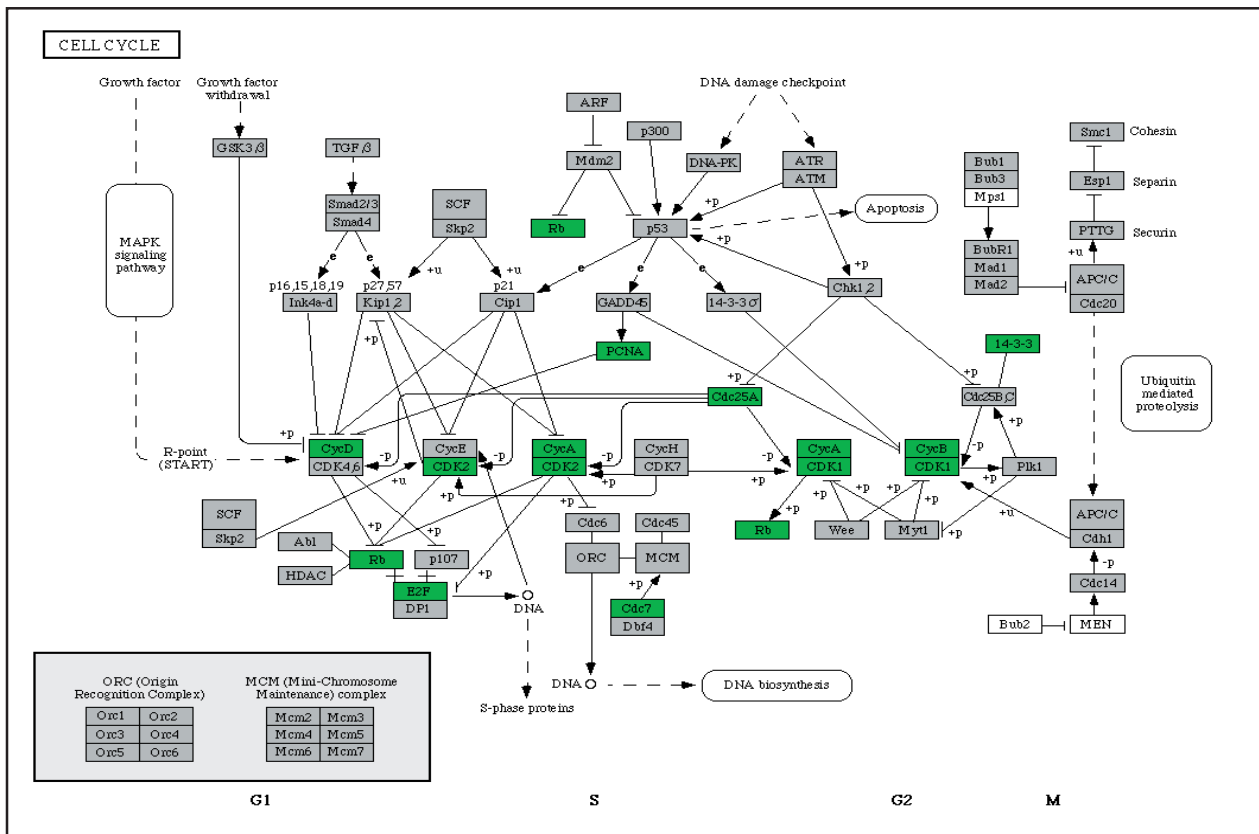
**Table 1: Functional annotation of defined patient subgroups.**

<b>M</b>		<b>N1</b>		<b>N2</b>				
category	n	p	category	n	p			
<b>DAVID IDs</b>								
	258			383	238			
<b>Biological processes</b>								
	230			344	203			
DNA replication	13	5E-3	endosome transport	7	1E-3	development	38	9E-3
taxis	9	7E-3	vacuole organization and biogenesis	7	9E-3	reactive oxygen species metabolism	6	1E-2
carbohydrate metabolism	25	9E-3	response to chemical stimuli	23	1E-2	response to oxidative stress	5	3E-2
negative regulation of biological processes	26	1E-2	regulation of enzyme activity	20	3E-2	dephosphorylation	6	2E-2
nucleic acid metabolism	69	2E-2	vesicle mediated transport	18	3E-2	intracellular protein transport	18	3E-2
<b>Molecular function</b>								
	238			347	211			
DNA binding	39	2E-2	protein dimerization activity	14	2E-2	protein domain specific binding	6	3E-2
protein dimerization activity	10	5E-2	hydrolase activity on glycosyl bonds	12	5E-2	GTPase activity	7	5E-2
nucleic acid binding	54	5E-2						
<b>Cellular component</b>								
	285			342	195			
chromosome	11	2E-3	vacuole	44	2E-8	chromosome	8	4E-2
chromatin	7	9E-3	lytic vacuole	39	2E-7			
nucleus	70	6E-3	lysosome	39	1E-7	lytic vacuole	17	4E-2
lytic vacuole	20	2E-3	extracellular region	32	2E-2	lysosome	17	4E-2
lysosome	20	2E-3	endosome	8	4E-2	non-membrane bound organelle	28	4E-2
<b>KEGG pathway</b>								
	122			185	125			
N-glycan degradation	5	2E-2	antigen processing	9	2E-3	Toll-like receptor signaling	10	2E-2
hematopoietic cell lineage	8	3E-2	glycosphingolipid metabolism	7	2E-2	glycosylaminoglycan degradation	6	2E-2
			hematopoietic cell lineage	10	4E-2			
<b>Biocarta pathway</b>								
	68			92	57			
cyclins and cell cycle regulation	9	2E-2	role of ERB2 in signal transduction	9	5E-3	activation of Src	4	3E-2
			IL 3 signaling pathway	7	1E-2	phospholipid signaling intermediates	5	4E-2
			IL 6 signaling pathway	8	1E-2			
			Erk and PI-3 kinase pathway	7	2E-2			
			signaling pathway from G-protein families	7	3E-2			

"n", number of genes involved in the corresponding annotation category; p, modified Fisher exact p-value of the gene enrichment for each category.

*MEF1*, *ERK*, *Elk1* and *FOS* were found in classical MAP kinase pathway (Figure 4A). Reduced expression of *IL1*, *IL1R*, *AKT*, *Elk1*, *GADD153* and *JunD* with elevated expression of *p53*, *p38* and *Evi1* were found in JNK and p38 MAP kinase pathways (Figure 4A). Reduced expression of *STAT*, *CPB*, *Pim-1*, *AKT* and *BclXL* and elevated expression of

*IL2/3* and *IL3R* were found in Jak-STAT signaling pathway. Elevated expression of *Bid* and *CytC* with reduced expression of *Bcl-2/XL* and *CASP9* were detected in the apoptotic pathway. General decrease in expression of genes involved in N-glycan, glycosylaminoglycan, and ganglioside degradation was found. In conjunction with 3-meth-



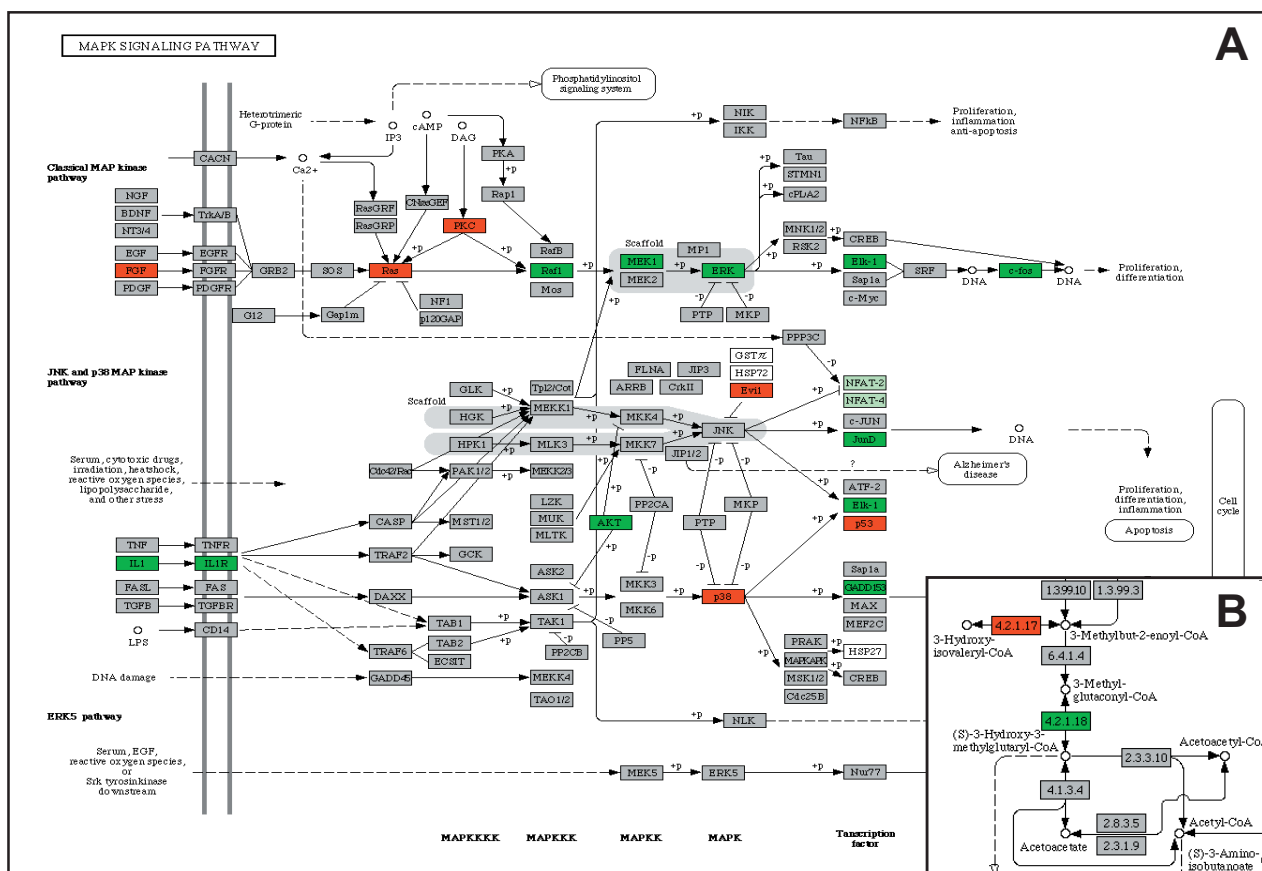
**Figure 3**  
**Gene expression changes detected in selected pathways in M group.** General changes in cell cycle pathway detected in patients with mtDNA mutation (M group) using KEGGArray software.

ylglutaconic aciduria, which is a characteristic biochemical feature of the patients from this group, inspection of leucine degradation pathway showed moderately reduced expression of 3-methylglutaconyl-CoA hydratase gene, *AUH*, (Figure 4B). Although the extent of the *AUH* expression changes neither directly implicates the deficiency of 3-methylglutaconyl-CoA hydratase nor explains 3-methylglutaconic aciduria present in these patients, it is possible that such changes might be much more pronounced and have functional effects during metabolic stress and/or in metabolically active tissues. In the **N2 group**, reduced expression of *GRB2*, *RAS* and *ERK* and elevated expression of *FOS*, *JUND* and *Evi1* was found in MAP kinase pathway. This was accompanied by elevated expression of genes involved in N-glycan, glycosylaminoglycan and ganglioside degradation. No multiple changes in the apoptotic and valine, leucine, isoleucine, lysine, β-oxidation degradation pathways were found. All mentioned pathways and gene expression changes identified by KEGGArray software are provided [see Additional file 10 and 11].

*Identification of patient specific gene expression profiles and definition of candidate disease causing genes*

To get specific information on patient mitochondrial genome expression, we extracted and clustered gene expression data for all 37 mtDNA genes. Resulting mitochondrial genome expression map (Figure 5A) reflects relative mitochondrial DNA amount with generally elevated expression in P11, P3, P10 and P6 and generally reduced expression in P2, P4 and P8. Specific gene expression changes were detected in P1 and P2, where the expression map revealed reduced amount of *MTATP6*, *MTATP8* and *MTCOXII* transcript reflecting disease causing microdeletion of *MTATP6*, and in P12 with specifically reduced expression of *tRNAGly*.

To obtain the information on patient specific ATP synthase complex expression, we extracted and clustered gene expression data for all of its structural genes and assembly factors. Resulting expression map is provided in Figure 5B. In P1 and P2, it shows reduced expression of mitochondrial subunits *MTATP6*, *MTATP8* and also of *ATPAF1*.



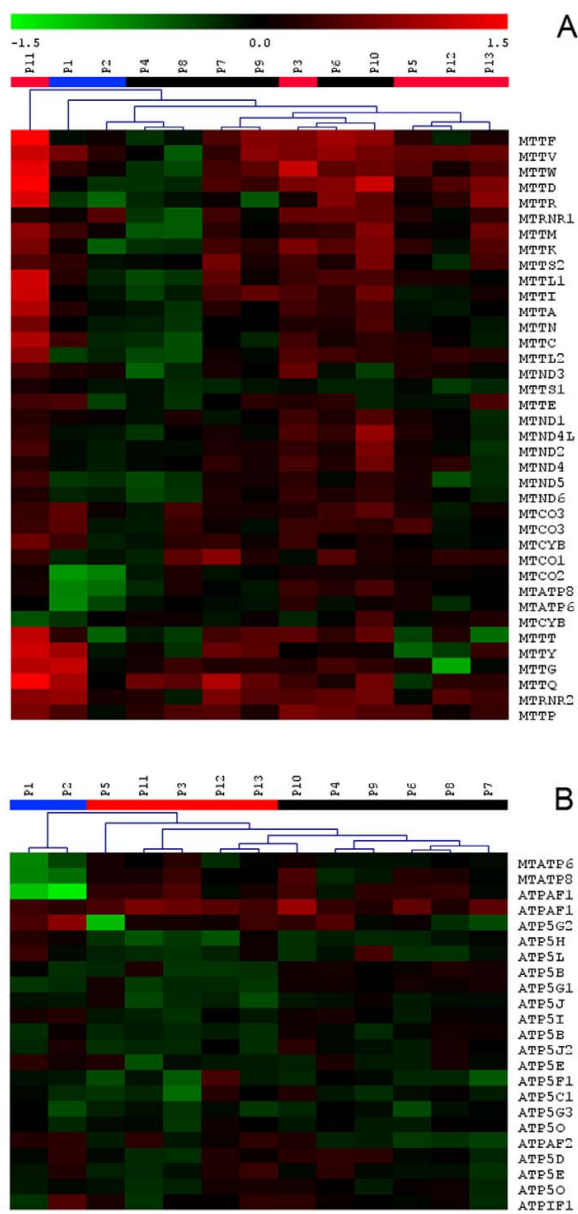
With the exception of reduced expression of *ATP5G2* in P5 and maybe also *ATP5C1* in P3 no additional subgroup and/or patient specific profile were found.

To define potential candidate disease genes, we finally compared gene expression data of individual patients with a group of controls in R statistical environment as described in methods, and searched for genes showing significantly reduced expression and having known function either in ATP synthase biogenesis, mitochondrial protein trafficking or mitochondrial biogenesis. In P1 and P2, we detected reduced expression of ATP synthase structural subunits *MTATP6*, *MTATP8* and also of *ATPAF1*. In P3 we detected reduced expression of *ATP5C1* and *ATP5O*. In P4 and P8 we detected reduced expression of *TOM7*. Mitochondrial carrier homolog 1 (*C. elegans*) (*MTCH1*) transcript was reduced in P6, P10, P11 and P12.

In P10 we detected reduced expression of mitochondrial elongation factor *EFG1* and *TOM22*. In P11 we found reduced expression of *TIM23*, *TIM8* and *TOM34* homologs, *ATP5H* and *ATP5E*. In P12 we found reduced expression of *mitofusin* and *ATP5H*. The lists of all the differentially expressed genes are shown [see Additional file 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24].

**Confirmation of the hybridization results**

To rule-out platform specific bias, we re-analyzed all RNA samples from the N1 and control groups using the same common reference RNA on Agilent 44 k arrays. We used available annotations and extracted from the Agilent data gene expression values for the genes identified as significantly ( $P < 0.05$ ), differentially expressed in the N1 group on our platform. Correlation coefficient of expression values of 102 identified genes was 0.925.



**Figure 5**  
**Two-dimensional hierarchical clustering of patient samples. A)** Expression matrix of all 37 mtDNA encoded genes. **B)** Expression matrix of structural and assembly factor genes involved in ATP synthase complex biogenesis. Selected genes were clustered across all patient samples using Euclidean distance metrics and average linkage clustering algorithm.

**Correlation of expression data with available RT-PCR and Western blot results**

ATP synthase deficiency of nuclear genetic origin is characterized at the protein level by pronounced decrease of

the individual subunits and the mature ATP synthase protein complex amounts (Table 2). However, our data in patient cell lines, in agreement with previous Q-PCR analyses, did not show pronounced alterations in ATP synthase subunits or of ATP synthase-specific assembly factors mRNA levels that could explain it easily. Only in the M group, the data showed decrease of *MTATP6* and *MTATP8* mRNA levels which correspond with previously performed Northern blot and Q-PCR analysis showing that this mutation affects processing of *ATP8/ATP6/COX-III* polycistronic transcript and results in decreased levels and/or stability of mature *ATP8/ATP6* mRNA [19,24]. Many of mitochondrial diseases are associated with compensatory changes in the cellular content of mitochondria and/or the content of one or more OXPHOS complexes. Western blot analysis of fibroblasts with ATP synthase deficiency has previously shown increased mitochondrial content of complex I and complex III [25]. In agreement with this observation, our data showed elevated expression of complex I subunit genes in N1 group. Expression of complex III subunit genes was however decreased. Parallel analyses of the fibroblasts with nuclear ATP synthase defects used in this study revealed variable changes in fibroblast COX and/or SDH specific content (Table 2). These changes were not associated with generally elevated expression of COX and SDH subunit genes. Detailed inspection of individual gene expression profiles [see Additional file 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24] however suggested that elevated expression of *COX7A2L* and *SDHA* may correlate with this observation (P3, P5, P6, P8, P10).

**Discussion**

**Platform selection and evaluation**

In our work, we attempted to set up an experimental platform which will allow in a cost-effective way prospective gene expression analysis of cell lines and tissues from patients with various genetically determined OXPHOS defects. We considered availability of biological materials for the analysis, estimated the number of informative genes, evaluated gene content of commercially available microarrays and took into account instrumentation availability and platform related running costs. Since cultured skin fibroblasts are the most accessible, relatively well standardized, and multiple analysis amendable source for gene expression analysis especially in nuclear encoded OXPHOS defects, we estimated (based on Gene Expression Omnibus database data), that in fibroblasts, reliable expression signal may be obtained for approximately 6000 (HG-U95 array) to 10000 (HG-U133 Plus 2.0 Array) genes, of which only part may be meaningful to detect and understand anticipated changes in mitochondrial biology and related basic cell responses. In addition, we also evaluated representation of mitochondria encoded genes on available whole genome arrays. We found that

**Table 2: Clinical, biochemical and molecular description of patients (P1 – P13).**

Patient (group)	Phenotype	Biochemical data	Genetic defect	ATPase (% of C)	SDH (% of C)	COX (% of C)	Ref.
P1 (M)	PMR, encephalomyopathy, spastic quadraparesis, microcephalia,	lactate: 1.0–3.4 3 MGA: <15	mt9205ΔTA	*80–120	120–200	80–120	[19]
P2 (M)	transient lactic acidosis, nystagmus, GR	lactate: 3.9–10	mt9205ΔTA	*80–120	80–120	80–120	[20]
P3 (N2)	PMR, HCMP, hypotonia, peripheral neuropathy,	lactate: 1.4–10 3 MGA: 133–281	ncDNA, unknown	<30	120–200	120–200	[21]
P4 (N1)	Fatal lactic acidosis, HCMP	lactate: 30–36	ncDNA, unknown	<30	120–200	80–120	[82]
P5 (N2)	PMR, HCMP, hypotonia, dysmorphism, microcephaly	lactate: 1.6–8 3 MGA: 22–225	ncDNA, unknown	<30	>200	>200	[21]
P6 (N1)	PMR, HCMP, hypotonia, dysmorphism, microcephaly	lactate: 3.6–4.5 3 MGA: 28–260	ncDNA, unknown	<30	>200	>200	NR
P7 (N1)	PMR, HCMP, hypotonia, dysmorphism, microcephaly, epilepsy	lactate: 2.2–6.0 3 MGA: 28–161	ncDNA, unknown	<30	80–120	120–200	NR
P8 (N1)	PMR, hypotonia, dysmorphism, microcephaly	lactate: 3.6–6.7 3 MGA: 56–252	ncDNA, unknown	<30	120–200	>200	NR
P9 (N1)	PMR, hypotonia, dysmorphism, microcephaly	lactate: 2.2–10 3 MGA: 62–150	ncDNA, unknown	<30	>200	>200	[21]
P10 (N1)	PMR, hypotonia, dysmorphism, microcephaly	lactate: 1.4–4.6 3 MGA: 64–270	ncDNA, unknown	<30	120–200	120–200	NR
P11 (N2)	PMR, hypotonia, GR, HCMP dysmorphism, microcephaly	lactate: 1.5–8.2 3 MGA: 34–254	ncDNA, unknown	<10	80–120	80–120	[21]
P12 (N2)	PMR, hypotonia, HCMP	lactate: 2–6.0 3 MGA: 115–460	ncDNA, unknown	<10	80–120	80–120	[25]
P13 (N2)	PMR, GR, microcephaly, mild spasticity, hepatopathy	lactate: 1.2–3.9 3 MGA: 37–132	ncDNA, unknown	<30	120–200	120–200	[21]

Patient assignment to groups is based on DNA sequencing data (M) and results of PCA and hierarchical clustering (N1, N2). PMR – psychomotor retardation, HCMP – hypertrophic cardiomyopathy, GR – growth retardation, lactate – blood lactate (mmol/l), 3 MGA – 3-methylglutaconic aciduria (mg/g creatinine). ATPase (complex V), SDH (complex II) and COX (complex IV) represent enzyme protein content in fibroblast homogenates quantified by SDS PAGE/WB as in [19], using specific primary antibodies (MitoSciences, OR), Alexa Fluor® 680-labeled secondary antibodies and an Odyssey® Infrared Imaging System (LI-COR Biotechnology, Lincoln, NE). Data are presented as % of control values. \* Decreased content of subunit a (ATP6). NR means not reported.

(at the time of project planning) no complete coverage of mitochondrial tRNA, rRNA and OXPHOS structural subunits have been available on() Affymetrix HG\_U95Av2 Array (contained just *TRNC*, *TRNY* and *TRNS1*), HG-U133 Plus 2.0 Array (no tRNAs, rRNAs and *ND1*, *ND4L*, *CYTB*) and Agilent 44 k Array (no tRNAs, rRNAs and *ND4L*). Considering this data and also available instrumentation, we then decided to construct focused an oligonucleotide microarray and employ competitive two-color hybridization approach with common reference experimental design.

Selected gene content allows gene expression analysis of the entire mitochondrial genome and almost all of "mitochondria" related genes in context of key DNA synthesis, growth response, regulatory and apoptotic genes. Hybridization signal was obtained from 78% of the designed oligonucleotides. Vast majority of the oligonucleotides giving no hybridization signal were designed to detect regulatory genes and transcription factor transcripts probably not transcribed in the analyzed materials. Interestingly, we detected hybridization signals for almost all mitochondrial tRNA and rRNA probes which is, in respect to

oligo-dT labeling strategy, suggestive that all those transcript are also at least partially polyadenylated [26].

#### **Gene expression analysis in patients with defect of $F_1F_0$ ATP synthase**

In the work presented herein, we analyzed and compared gene expression profiles in fibroblast cell lines from 9 control individuals and 13 patients with biochemically proven but genetically heterogeneous  $F_1F_0$  ATP synthase deficiency. We aimed to identify gene expression changes indicating how affected cells react to and compensate for the common biochemical defect, use gene expression data to assign patients into already defined and/or putative disease subgroups, identify candidate disease causing genes, and define potential pathogenetic mechanisms associated with the disease.

The magnitude of observed expression changes was moderate with only several dozens of genes exceeding 2-fold changes. Comparing all the patient cell lines with all control cell lines, we have not identified any common and meaningful gene expression changes attributable to ATP synthase deficiency *per se*. It has been suggested recently that the degree and compartmentalization of ATP deple-

tion may be defect specific and may thus have also specific biological consequences [27]. Our data support this view.

Cell lines with mtDNA mutation (M\_group) showed gene expression changes suggestive of suppressed mitochondrial biogenesis and metabolism characterized by down regulation of *TFAM* and *TFB1M*, master regulators of mitochondrial transcription, accompanied by reduced expression of other mitochondria encoded transcripts (*MTCO2*, *MTATP6*, *MTATP8*, and *MTND6*), reduced expression of *ATPAF1*, *E2F1*, *ACO2* (component of mitochondria to nucleus retrograde pathway) and *PPARA*. This "mitochondria silencing" activity seems to be sensed and counterbalanced by elevated expression of *NRF1*, which is however not accompanied by expression changes of any NRF-1 target and/or coactivator genes [28,29]. Inhibition of mitochondrial biogenesis is synchronized with reduced expression of genes regulating the G1/S phase transition (*E2F1*, *MYC*, *Rb*, *CycA*, *CycD*, *CDK2*, *Cdc7*, *Cdc25A*, *PCNA*) and associated thymidine metabolism (*TK*, *TYMS*) [30]. We interpret this gene expression pattern as an ATP depletion mediated G1/S arrest [31] associated with synchronized replication arrest of mitochondrial genome [32] and repression of NRF-1 activity [33]. Our observations are quite similar to that made in *Drosophila* mutants, in which low ATP levels lead to arrest in the G1 phase without affecting cellular differentiation and cell viability [34,35]. Furthermore, our observation conforms to the view that mitochondria co-regulate cell cycle progression and that this regulation is executed not only at posttranscriptional [34] but also at transcriptional level.

The N1\_group differed from M group in that it showed very minor signs of mitochondrial response suggested only by slightly elevated expression of *PPGC-1*, *TFAM*, *TFB2M* and *ACO2*. More significant and distinct changes were however observed in signal transduction pathways regulating mitochondrial oxidative phosphorylation [36]. The gene expression portrait, reduced expression of many transcription factors and cytokines regulating cell growth and differentiation (*FOS* [37], *JUNB* and *MAPK3* [38], *CEBPA* and *CEBPB*, *CXCL1* and *CXCL2* [39]), elevated expression of *IGFBP3* [40] and *CAV2* [41], together with activated apoptosis (*BCL2L1*, *SMAC*, *CYCS*, *FAF1*), signs of oxidative stress (*TR2*) [42] and general decrease in lysosomal activities [43], resemble characteristic signs of senescent fibroblasts [44,45]. However all the cell lines

from the N1 group have originated from very young donors, all but one were in their early passages and all showed the same passage frequency of 5–6 days, (Table 3). It has been shown that inhibition of oxidative phosphorylation may play an active role in the process of cellular senescence in human fibroblasts [46], and that changes in transcription activity may be governed by changes in protein phosphorylation [47]. We therefore interpret the observed gene expression pattern as accelerated stress induced premature senescence phenotype resulting from impaired oxidative phosphorylation and profoundly reduced ATP availability for critical energy-dependent cellular processes. Our explanation of N1 cellular phenotype is the following. Mitochondrial ATP synthesis is markedly decreased in fibroblasts derived from patients with nuclear DNA-related disorders but only variably so in patients with mtDNA mutations [48]. ATP depletion is sensed by AMP-activated protein kinase which acts as a metabolic sensor or "fuel gauge" that monitors cellular AMP and ATP levels [49]. Once activated, the enzyme switches off ATP-consuming anabolic pathways and switches on ATP-producing catabolic pathways [50], such as fatty acid oxidation (elevated expression of *ECH1*, *ECHS1*, *ETFDH*, *CABC1*) and amino acid catabolism. Despite this compensatory effort, mitochondrial ATP depletion persists due to intrinsic ATP synthase defect, activation of AMPK persist and leads to accelerated p53-dependent cellular senescence [51]. AMPK activity also leads to decrease of HuR cytosolic translocation, which influences the mRNA-stabilizing function of HuR [52] and diminishes the expression and half-lives of HuR target transcripts, such as *FOS* [53] or *CDKN1A* [54] which also leads to the premature senescence phenotype [55]. ATP availability probably modulates cytoplasmic translocation and recruitment of other RNA-binding proteins stabilizing various mRNAs [56]. In this context it is interesting that we have detected reduced expression (or transcript abundance) of two RNA-binding protein genes *CUGBP1* and *AUH*. *CUGBP1* affects translation of *CDKN1A* [57] and *CEBPB* [58], and our data show decrease in those two transcripts as well. *AUH* stabilizes *FOS* and other immediate early mRNA's [59], and its deficiency is also causing methylglutaconic aciduria [60], a characteristic biochemical phenotype observed specifically in this group of nuclear encoded ATP synthase deficient patients [21] (Table 2).

**Table 3: Growth characteristics of the fibroblast cell lines.**

	patients										controls											
	1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9
passage number	17	19	15	4	20	9	6	6	28	12	28	17	12	22	22	14	27	16	17	11	16	13
passage frequency (days)	5	5	6	6	6	6	6	6	5	6	6	7	7	9	3	3	4	4	5	4	4	7



Resulting transcriptional silencing and other ATP depletion mediated disturbances of intracellular signal transduction cascades lead thus to premature senescence phenotype, reduced proteasome activity and accumulation of oxidized proteins, which may explain observed discrepancies between gene expression and Western blot data. Patients forming this group are of common ethnic origin, and this is suggestive that common genetic defect may underlie this specific gene expression profile.

The **N2 group** showed neither signs of mitochondria response observed in the M group, nor signs of premature senescence observed in the N1 group. Expression profile is suggestive of partly activated apoptosis (SMAC) and disturbances of intracellular signaling transduction cascades (down regulation of several cytokines, early genes, and regulatory proteins). However, all these changes were not uniformly present in all cell lines, which together with variability in clinical and biochemical data is suggestive of further genetic heterogeneity within this group of patients.

#### **Selection of candidate disease causing genes**

As gene expression changes may be used for selection of candidate disease causing genes [9,61], we evaluated group specific and individual gene expression profiles. This approach was successful in both patients from the M group, in whom detected alterations clearly indicated the involvement of ATP6/ATP8/COXIII transcript. In other patients we first focused on expression of ATP synthase subunits. Inspection of this expression profile (Figure 5B) suggested involvement of ATP synthase assembly factor ATPAF2 in several patients from N1 group. Mutation of ATPAF2 has been found in the case with ATP synthase deficiency [62] and this warrant sequence analysis of this gene in this group of patients. Other candidate genes may be ATP5G2, the expression of which is decreased in P5 and possibly also ATP5C1 found lowered in P3. From other genes, no clear candidates for immediate sequence analysis may be defined yet. However, more focused interpretation will be possible once candidate disease genomic intervals are defined by ongoing linkage studies.

#### **Conclusion**

We designed, produced, and validated an oligonucleotide microarray focused on expression profiling of human mitochondria related genes, and searched for gene expression changes in genetically heterogeneous group of 13 patients with F<sub>1</sub>F<sub>0</sub> ATP synthase deficiency. The analysis classified patients into three distinct groups and suggested that site (mtDNA vs nucleus) and severity (residual content of ATP synthase) of underlying biochemical defect have diverse effects on cell gene expression phenotype. Comparisons with controls, between defined groups and among individual patient cell lines did not show any uni-

form transcription changes explaining pronounced decrease in ATP synthase content and alterations of the other OXPHOS complexes observed at the protein level. The analysis nevertheless confirmed the already known and indicated candidate disease causing genes, and suggested that defects in ATP synthesis lead to deregulation of signal transduction pathways and affect mitochondrial and nuclear DNA replication. These may be important pathogenic mechanisms involved not only in F<sub>1</sub>F<sub>0</sub> ATP synthase deficiency but also in other OXPHOS defects. Observed gene expression changes therefore warrant further investigation of major cell cycle regulatory and signal transduction pathways in other OXPHOS disorders and pharmacological models. Full potential of the constructed h-MitoArray platform will be further revealed in ongoing positional cloning studies in herein analyzed patients and in gene expression studies in other groups of OXPHOS deficient cell lines.

#### **Methods**

##### **Database of human mitochondrial genes**

Lists of "mitochondrial" and "mitochondria related" genes were extracted and merged from various public databases such as Mitomap [63], Mitop [64], Migenes [65], Mitoproteom [66], Molecular Signature Database [67], OMIM, RefSeq and Unigene sections at NCBI [68], Gene Ontology database [69] and UniProt resource [70]. Full annotation of selected genes has been obtained and deposited in a locally installed database BASE [71].

##### **Microarray preparation**

For each of the selected 1632 genes, a single 5'-aminomodified 40-mer oligonucleotide was designed using Oligopicker software [72]. Blast searches were performed with each candidate probe to exclude possibility of cross hybridization with homologous genes prior to the synthesis of oligonucleotide probes. Synthesized oligonucleotides, Generi Biotech (Czech Republic) and Illumina (San Diego, CA), were resuspended at 20 μM concentration in 3 × SSC, printed in triplicates on aminosilane modified slides, and immobilised by standard technique using combination of baking and UV cross-link as previously described [61]. Qualities of arrays from individual printing series were assessed using fluorescently labelled panomers (Invitrogen, Carlsbad, CA).

##### **Mixed RNA for microarray validation**

As a standard for microarray optimisation, standardization and validation total RNA was isolated from HeLa G, ECV 304, 293, U 937, JURKAT and A 301 cell lines using the TRIZOL solution (Invitrogen, Carlsbad, CA). Isolated RNA samples were pooled, and aliquots were stored at -80 °C until the analysis.

**Reference RNA preparation**

As a common reference RNA for gene expression studies, total RNA from cultured HeLa cells was chosen. Total RNA was extracted as above. Concentration was determined spectrophotometrically at A 260 by NanoDrop (NanoDrop Technologies, Wilmington, DE) and quality was checked on Agilent 2100 bioanalyser – RNA Lab-On-a-Chip (Agilent Technologies, Santa Clara, CA). Aliquots of isolated RNA were stored at -80°C until the analysis.

**Control group**

Selected control fibroblasts cell lines were used repeatedly in previous diagnostic biochemical tests and showed no signs of any mitochondrial or other metabolic defect.

**Patients**

Fibroblast cell lines from 13 patients were used in this study. All the patients showed major clinical symptoms associated with OXPHOS defect. Biochemical diagnosis of ATP synthase deficiency was based on absence or significant decrease of mature ATP synthase complex and of its subunits in electrophoretic analysis of OXPHOS complexes in cultured fibroblasts and other available tissues [73]. Mitochondrial genome sequencing performed in all patients revealed disease causing mitochondrial DNA mutations in two patients (P1, P2, M group) [19]. Molecular basis of defect in the other patients has not yet been defined. Relevant clinical, biochemical and molecular data and references on individual patients included in this study are provided in Table 2.

**Cell culturing**

Growth characteristics of the cell lines used in this study are provided in Table 3. Skin fibroblasts were cultured in the Dulbecco's modified Eagle's medium supplemented by 10% fetal calf serum, 20 mM HEPES pH 7.5, 0.2% NaHCO<sub>3</sub> and gentamycin 0.02 mg/ml at 37°C in a 5% CO<sub>2</sub> humidified atmosphere. For experiments, confluent cell were harvested using 0.05% trypsin and 0.02% EDTA. Detached cells were diluted in ice-cold culture medium, sedimented by centrifugation (600 g) and washed twice in phosphate buffered saline (140 mM NaCl, 5.4 mM KCl, 8 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.4 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.2).

**RNA preparation, cDNA labeling and hybridization**

Total RNA was extracted from cultured cells and QC controlled as described above.

Five µg of total RNA was reverse transcribed and labeled by Array 900 Expression Detection Kit (Genisphere, Hatfield, PA) according to the manufacturer protocol. The slides were pretreated by baking at 80°C, UV cross-linked and washed twice in 0.1% SDS for 2 minutes, twice in 0.2 × SSC for 2 min, four times in MilliQ water, followed by

denaturation in boiling water for 2 minutes. Prehybridization was performed using hybridization buffer (Genisphere, Hatfield, PA) according to the manufacturer protocol. All hybridizations were performed in humid hybridization chamber, ArrayIt Hybridization Cassette chamber (TeleChem International, Sunnyvale, CA).

**Microarray scanning**

The hybridized slides were scanned with GenePix 4200A scanner (Axon Instruments, Union City, CA) with PMT gains adjusted to obtain highest intensity unsaturated images. GenePix Pro software (Axon Instruments, Union City, CA) was used for image analysis of the TIFF files, as generated by the scanner.

**Experimental setup and data normalization**

All 13 patient samples and 9 controls were hybridized to common reference (HeLa cell lines) in two replicates of each sample. All arrays were hybridized with a Cy5-labeled sample cDNA and a Cy3-labeled reference cDNA.

Expression data were obtained using GenPix Pro software. Comparative microarray analysis was performed according to MIAME guidelines [74]. Normalization was performed in R statistical environment [75] using Limma package [76] which is part of the Bioconductor project [77]. Raw data from individual arrays were processed using Loess normalization and normexp background correction. Gquantile function was used for normalization between arrays. The correlation between 3 replicate spots per gene on each array was used to increase the robustness. Linear model was fitted for each gene given a series of arrays using lmFit function. The empirical Bayes method was used to rank differential expression of genes using eBayes function. Multiple testing correction was performed using Benjamini & Hochberg method [78].

**Quality control**

Variation among feature replicates on the array was calculated by conversion of raw data to log-ratios. Data were further normalized using Loess function. Features with less than double background intensity ( $A < 8.5$ ) were removed. For each feature on the array the deviation from the mean computed as the difference between the ratio of the feature and the mean of the set of feature replicates was calculated. Standard deviation of the error distribution using all of the replicates was calculated and converted to coefficient of variability using equation.

$$CV = \sqrt{\exp[(\ln 2 * SD)^2] - 1}$$

The variability between the duplicate spots ranged from 8.1% to 27.5%. Arrays with variability higher than 18% were removed from the analysis.

**Statistical analysis**

Principal component analysis, hierarchical clustering, ANOVA and SAM analyses were performed in TIGR Multiexperiment Viewer (MeV), version 4.0 [79], available [80]. Significant gene expression changes between defined subgroups were identified using t-test in R statistical environment [75]. Applied parameters are provided in corresponding result sections.

**Functional annotation**

Functional annotation and pathway enrichment analysis was performed in DAVID (The Database for Annotation, Visualization and Integrated Discovery [23]). Visualization of gene expression changes along affected pathways was performed in KEGGArray software (KEGG pathway databases – Kyoto Encyclopedia of Genes and Genomes) [81].

**Data accession**

Description of h-MitoArray platform and gene expression data reported in this study are stored and available in Gene Expression Omnibus repository under accessions GPL5150 and GSE8648.

**Ethics**

The project was approved by the Scientific Ethics Committee of the 1st Faculty of Medicine of Charles University of Prague under reference NR/8069-3. Patient participation in the project was made on a voluntary basis after oral and written information and consent according to the Helsinki V Declaration.

**Authors' contributions**

A.Č. tested, compared and optimized labeling and hybridization condition, performed all the RNA sample isolations, cDNA labeling, microarray hybridizations, data acquisition, and contributed to data analysis and interpretation. V.S. performed all data analysis in R and MeV environments. R.I. designed oligonucleotide probes, updated functional annotation of selected gene set, and participated in data analysis. H.H., L.N. a L.P. set and kept BASE database, optimized methods for microarray manufacturing and prepared all microarrays used in this study. M.T, H. Han., T.H., J.Z., J.A.M., W.S., A.P., J.P. and J.H. performed all relevant clinical, biochemical, and molecular investigations in studied patients, and together with S.S. provided patient and control cell lines. S.K. conceived and coordinated the study, was involved together with J.H. in data analysis, result interpretation and manuscript preparation.

**Additional material****Additional file 1**

*Annotation of h-MitoArray. List of selected genes with corresponding symbols, accession and LocusLink codes. Probes showing hybridization signal with fluorescently labeled panomers and fluorescently labeled cDNA are presented as "signal detected".*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S1.xls>]

**Additional file 2**

*Functional annotation of the h-MitoArray. Functional annotation of selected genes and comparison of the h-MitoArray gene content against whole human genome reference set.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S2.xls>]

**Additional file 3**

*Expression matrix – samples compared to common reference. Ratios of Log<sub>2</sub> sample gene intensities against Log<sub>2</sub> gene intensities of common reference.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S3.xls>]

**Additional file 4**

*Expression matrix – patients compared to controls. Ratios of individual patient Log<sub>2</sub> gene intensities against the average of the Log<sub>2</sub> controls intensities.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S4.xls>]

**Additional file 5**

*Differentially expressed genes between all patients and controls. List of genes detected as differentially expressed between all studied ATP synthase deficient and control fibroblast cell lines at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S5.xls>]

**Additional file 6**

*Characterization of defined patient groups using ANOVA analysis. List of genes detected as differentially expressed between defined patient groups using ANOVA analysis and unadjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S6.xls>]

**Additional file 7**

*Lists of differentially expressed genes in M group. Lists of genes detected as differentially expressed in M group, when compared to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S7.xls>]

**Additional file 8**

*Lists of differentially expressed genes in N1 group. Lists of genes detected as differentially expressed in N1 group, when compared to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S8.xls>]

**Additional file 9**

*Lists of differentially expressed genes in N2 group. Lists of genes detected as differentially expressed in N2 group, when compared to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S9.xls>]

**Additional file 10**

*Pathway analysis in M group. Pathways and gene expression changes identified by KEGGArray software in M group.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S10.PDF>]

**Additional file 11**

*Pathway analysis in N1 group. Pathways and gene expression changes identified by KEGGArray software in N1 group.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S11.PDF>]

**Additional file 12**

*Differentially expressed genes in patient P1. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S12.xls>]

**Additional file 13**

*Differentially expressed genes in patient P2. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S13.xls>]

**Additional file 14**

*Differentially expressed genes in patient P3. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S14.xls>]

**Additional file 15**

*Differentially expressed genes in patient P4. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S15.xls>]

**Additional file 16**

*Differentially expressed genes in patient P5. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S16.xls>]

**Additional file 17**

*Differentially expressed genes in patient P6. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S17.xls>]

**Additional file 18**

*Differentially expressed genes in patient P7. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S18.xls>]

**Additional file 19**

*Differentially expressed genes in patient P8. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S19.xls>]

**Additional file 20**

*Differentially expressed genes in patient P9. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S20.xls>]

**Additional file 21**

*Differentially expressed genes in patient P10. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S21.xls>]

**Additional file 22**

*Differentially expressed genes in patient P11. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S22.xls>]

**Additional file 23**

Differentially expressed genes in patient P12. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S23.xls>]

**Additional file 24**

Differentially expressed genes in patient P13. Lists of genes detected as differentially expressed in individual patients comparing to controls at adjusted  $P < 0.01$  significance level.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-38-S24.xls>]

**Acknowledgements**

This work was supported by grant NR8069-3 from the Grant Agency of the Ministry of Health of the Czech Republic. Further support was provided by grants 303/03/H065 and 303/07/0781 from the Grant Agency of the Czech Republic, 54/203208 27/05 from the Grant Agency of the Charles University of Prague and the Czech-Austrian Bilateral Cooperation Project (Kontakt 2006/3). Institutional support was provided by Ministry of Education of Czech Republic grants IM6837805002, AVOZ 501 10509 and MSM0021620806.

**References**

- McFarland R, Taylor RW, Turnbull DM: **Mitochondrial disease – its impact, etiology, and pathology.** *Curr Top Dev Biol* 2007, **77**:113-155.
- DiMauro S, Schon EA: **Mitochondrial respiratory-chain diseases.** *N Engl J Med* 2003, **348**:2656-2668.
- DiMauro S: **Mitochondrial DNA medicine.** *Biosci Rep* 2007, **27**:5-9.
- Chinnery PF: **Searching for nuclear-mitochondrial genes.** *Trends Genet* 2003, **19**:60-62.
- Shoubridge EA: **Nuclear gene defects in respiratory chain disorders.** *Semin Neurol* 2001, **21**:261-267.
- Calvo S, Jain M, Xie X, Sheth SA, Chang B, Goldberger OA, Spinazzola A, Zeviani M, Carr SA, Mootha VK: **Systematic identification of human mitochondrial disease genes through integrative genomics.** *Nat Genet* 2006, **38**:576-582.
- Thorburn DR, Sugiana C, Salemi R, Kirby DM, Worgan L, Ohtake A, Ryan MT: **Biochemical and molecular diagnosis of mitochondrial respiratory chain disorders.** *Biochim Biophys Acta* 2004, **1659**:121-128.
- Slonim DK: **From patterns to pathways: gene expression data analysis comes of age.** *Nat Genet* 2002, **32**(Suppl):502-508.
- Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, Delmonte T, Villeneuve A, Sladek R, Xu F, Mitchell GA, Morin C, Mann M, Hudson TJ, Robinson B, Rioux JD, Lander ES: **Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics.** *Proc Natl Acad Sci USA* 2003, **100**:605-610.
- Kirby DM, Salemi R, Sugiana C, Ohtake A, Parry L, Bell KM, Kirk EP, Boneh A, Taylor RW, Dahl HH, Ryan MT, Thorburn DR: **NDUFS6 mutations are a novel cause of lethal neonatal mitochondrial complex I deficiency.** *J Clin Invest* 2004, **114**:837-845.
- Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC: **PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.** *Nat Genet* 2003, **34**:267-273.
- Chen JJ: **Key aspects of analyzing microarray gene-expression data.** *Pharmacogenomics* 2007, **8**:473-482.
- Alesci S, Manoli I, Michopoulos VJ, Brouwers FM, Le H, Gold PW, Blackman MR, Rennert OM, Su YA, Chrousos GP: **Development of a human mitochondria-focused cDNA microarray (hMitChip) and validation in skeletal muscle cells: implications for pharmaco- and mitogenomics.** *Pharmacogenomics J* 2006, **6**:333-342.
- Kerstann KW, Procaccio VF, Yen HC, Hosseini SH, Golik PZ, Wallace DC: **Microarray Analysis of Human Mitochondrial Disease Patients.** *Am J Hum Genet* 2000, **67**:271.
- Van Der Westhuizen FH, Van Den Heuvel LP, Smeets R, Veltman JA, Pfundt R, Van Kessel AG, Ursing BM, Smeitink JA: **Human mitochondrial complex I deficiency: investigating transcriptional responses by microarray.** *Neuropediatrics* 2003, **34**:14-22.
- Bai X, Wu J, Zhang Q, Alesci S, Manoli I, Blackman MR, Chrousos GP, Goldstein AL, Rennert OM, Su YA: **Third-generation human mitochondria-focused cDNA microarray and its bioinformatic tools for analysis of gene expression.** *Biotechniques* 2007, **42**:365-375.
- Halgren RG, Fielden MR, Fong CJ, Zacharewski TR: **Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones.** *Nucleic Acids Res* 2001, **29**:582-588.
- Holloway AJ, van Laar RK, Tothill RW, Bowtell DD: **Options available – from start to finish – for obtaining data from DNA microarrays II.** *Nat Genet* 2002, **32**(Suppl):481-489.
- Jesina P, Tesarova M, Fornuskova D, Vojtkova A, Pecina P, Kaplanova V, Hansikova H, Zeman J, Houstek J: **Diminished synthesis of subunit a (ATP6) and altered function of ATP synthase and cytochrome c oxidase due to the mtDNA 2 bp microdeletion of TA at positions 9205 and 9206.** *Biochem J* 2004, **383**:561-571.
- Seneca S, Abramowicz M, Lissens W, Muller MF, Vamos E, de Meirleir L: **A mitochondrial DNA microdeletion in a newborn girl with transient lactic acidosis.** *J Inher Metab Dis* 1996, **19**:115-118.
- Sperl W, Jesina P, Zeman J, Mayr JA, Demeirleir L, VanCoster R, Pickova A, Hansikova H, Houstkova H, Krejcek Z, Koch J, Smet J, Muss W, Holme E, Houstek J: **Deficiency of mitochondrial ATP synthase of nuclear genetic origin.** *Neuromuscul Disord* 2006, **16**:821-829.
- Houstek J, Pickova A, Vojtkova A, Mracek T, Pecina P, Jesina P: **Mitochondrial diseases and genetic defects of ATP synthase.** *Biochim Biophys Acta* 2006, **1757**:1400-1405.
- The Database for Annotation, Visualization and Integrated Discovery (DAVID) 2007 [<http://david.abcc.ncifcrf.gov/>]
- Chrzanowska-Lightowlers ZM, Temperley RJ, Smith PM, Seneca SH, Lightowlers RN: **Functional polypeptides can be synthesized from human mitochondrial transcripts lacking termination codons.** *Biochem J* 2004, **377**:725-731.
- Mayr JA, Paul J, Pecina P, Kurnik P, Forster H, Fotschl U, Sperl W, Houstek J: **Reduced respiratory control with ADP and changed pattern of respiratory chain enzymes as a result of selective deficiency of the mitochondrial ATP synthase.** *Pediatr Res* 2004, **55**:988-994.
- Slomovic S, Laufer D, Geiger D, Schuster G: **Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark.** *Mol Cell Biol* 2005, **25**:6427-6435.
- Gajewski CD, Yang L, Schon EA, Manfredi G: **New insights into the bioenergetics of mitochondrial disorders using intracellular ATP reporters.** *Mol Biol Cell* 2003, **14**:3628-3635.
- Kelly DP, Scarpulla RC: **Transcriptional regulatory circuits controlling mitochondrial biogenesis and function.** *Genes Dev* 2004, **18**:357-368.
- Scarpulla RC: **Nuclear control of respiratory gene expression in mammalian cells.** *J Cell Biochem* 2006, **97**:673-683.
- Hu CM, Chang ZF: **Mitotic control of dTTP pool: a necessity or coincidence?** *J Biomed Sci* 2007.
- Gemin A, Sweet S, Preston TJ, Singh G: **Regulation of the cell cycle in response to inhibition of mitochondrial generated energy.** *Biochem Biophys Res Commun* 2005, **332**:1122-1132.
- Martinez-Diez M, Santamaria G, Ortega AD, Cuezva JM: **Biogenesis and Dynamics of Mitochondria during the Cell Cycle: Significance of 3'UTRs.** *PLoS ONE* 2006, **1**:e107.
- Wang C, Li Z, Lu Y, Du R, Katiyar S, Yang J, Fu M, Leader JE, Quong A, Novikoff PM, Pestell RG: **Cyclin D1 repression of nuclear res-**

- piratory factor I integrates nuclear DNA synthesis and mitochondrial function.** *Proc Natl Acad Sci USA* 2006, **103**:11567-11572.
34. Mandal S, Guptan P, Owusu-Ansah E, Banerjee U: **Mitochondrial regulation of cell cycle progression during development as revealed by the tenured mutation in Drosophila.** *Dev Cell* 2005, **9**:843-854.
  35. Liao TS, Call GB, Guptan P, Cespedes A, Marshall J, Yackle K, Owusu-Ansah E, Mandal S, Fang QA, Goodstein GL, Kim W, Banerjee U: **An efficient genetic screen in Drosophila to identify nuclear-encoded genes with mitochondrial function.** *Genetics* 2006, **174**:525-533.
  36. Boneh A: **Regulation of mitochondrial oxidative phosphorylation by second messenger-mediated signal transduction mechanisms.** *Cell Mol Life Sci* 2006, **63**:1236-1248.
  37. Seshadri T, Campisi J: **Repression of c-fos transcription and an altered genetic program in senescent human fibroblasts.** *Science* 1990, **247**:205-209.
  38. Chalmers CJ, Gilley R, March HN, Balmanno K, Cook SJ: **The duration of ERK1/2 activity determines the activation of c-Fos and Fra-1 and the composition and quantitative transcriptional output of AP-1.** *Cell Signal* 2007, **19**:695-704.
  39. Limatola C, Mileo AM, Giovannelli A, Vacca F, Ciotti MT, Mercanti D, Santoni A, Eusebi F: **The growth-related gene product beta induces sphingomyelin hydrolysis and activation of c-Jun N-terminal kinase in rat cerebellar granule neurones.** *J Biol Chem* 1999, **274**:36537-36543.
  40. Moerman EJ, Thweatt R, Moerman AM, Jones RA, Goldstein S: **Insulin-like growth factor binding protein-3 is overexpressed in senescent and quiescent human fibroblasts.** *Exp Gerontol* 1993, **28**:361-370.
  41. Park WY, Park JS, Cho KA, Kim DI, Ko YG, Seo JS, Park SC: **Up-regulation of caveolin attenuates epidermal growth factor signaling in senescent cells.** *J Biol Chem* 2000, **275**:20847-20852.
  42. Lee SR, Kim JR, Kwon KS, Yoon HW, Levine RL, Ginsburg A, Rhee SG: **Molecular cloning and characterization of a mitochondrial selenocysteine-containing thioredoxin reductase from rat liver.** *J Biol Chem* 1999, **274**:4722-4734.
  43. Cuervo AM, Dice JF: **How do intracellular proteolytic systems change with age?** *Front Biosci* 1998, **3**:d25-43.
  44. Cristofalo VJ, Lorenzini A, Allen RG, Torres C, Tresini M: **Replicative senescence: a critical review.** *Mech Ageing Dev* 2004, **125**:827-848.
  45. Shelton DN, Chang E, Whittier PS, Choi D, Funk WD: **Microarray analysis of replicative senescence.** *Curr Biol* 1999, **9**:939-945.
  46. Stockl P, Hutter E, Zwierschke W, Jansen-Durr P: **Sustained inhibition of oxidative phosphorylation impairs cell proliferation and induces premature senescence in human fibroblasts.** *Exp Gerontol* 2006, **41**:674-682.
  47. Whitmarsh AJ, Davis RJ: **Regulation of transcription factor function by phosphorylation.** *Cell Mol Life Sci* 2000, **57**:1172-1183.
  48. Shepherd RK, Checcharelli N, Naini A, De Vivo DC, DiMauro S, Sue CM: **Measurement of ATP production in mitochondrial disorders.** *J Inherit Metab Dis* 2006, **29**:86-91.
  49. Hardie DG, Hawley SA, Scott JW: **AMP-activated protein kinase – development of the energy sensor concept.** *J Physiol* 2006, **574**:7-15.
  50. Zong H, Ren JM, Young LH, Pypaert M, Mu J, Birnbaum MJ, Shulman GI: **AMP kinase is required for mitochondrial biogenesis in skeletal muscle in response to chronic energy deprivation.** *Proc Natl Acad Sci USA* 2002, **99**:15983-15987.
  51. Jones RG, Plas DR, Kubek S, Buzzai M, Mu J, Xu Y, Birnbaum MJ, Thompson CB: **AMP-activated protein kinase induces a p53-dependent metabolic checkpoint.** *Mol Cell* 2005, **18**:283-293.
  52. Wang W, Fan J, Yang X, Furer-Galban S, Lopez de Silanes I, von Kobbe C, Guo J, Georas SN, Fougelle F, Hardie DG, Carling D, Gorospe M: **AMP-activated kinase regulates cytoplasmic HuR.** *Mol Cell Biol* 2002, **22**:3425-3436.
  53. Wang W, Yang X, Cristofalo VJ, Holbrook NJ, Gorospe M: **Loss of HuR is linked to reduced expression of proliferative genes during replicative senescence.** *Mol Cell Biol* 2001, **21**:5889-5898.
  54. Wang W, Furneaux H, Cheng H, Caldwell MC, Hutter D, Liu Y, Holbrook N, Gorospe M: **HuR regulates p21 mRNA stabilization by UV light.** *Mol Cell Biol* 2000, **20**:760-769.
  55. Wang W, Yang X, Lopez de Silanes I, Carling D, Gorospe M: **Increased AMP:ATP ratio and AMP-activated protein kinase activity during cellular senescence linked to reduced HuR function.** *J Biol Chem* 2003, **278**:27016-27023.
  56. Chen HH, Xu J, Safarpour F, Stewart AF: **LMO4 mRNA stability is regulated by extracellular ATP in F11 cells.** *Biochem Biophys Res Commun* 2007, **357**:56-61.
  57. Iakova P, Wang GL, Timchenko L, Michalak M, Pereira-Smith OM, Smith JR, Timchenko NA: **Competition of CUGBP1 and calreticulin for the regulation of p21 translation determines cell fate.** *Embo J* 2004, **23**:406-417.
  58. Timchenko LT, Salisbury E, Wang GL, Nguyen H, Albrecht JH, Hershey JW, Timchenko NA: **Age-specific CUGBP1-eIF2 complex increases translation of CCAAT/enhancer-binding protein beta in old liver.** *J Biol Chem* 2006, **281**:32806-32819.
  59. Nakagawa J, Waldner H, Meyer-Monard S, Hofsteenge J, Jenö P, Moroni C: **AUH, a gene encoding an AU-specific RNA binding protein with intrinsic enoyl-CoA hydratase activity.** *Proc Natl Acad Sci USA* 1995, **92**:2051-2055.
  60. L I, Loupatty FJ, Ruiter JP, Duran M, Lehnert W, Wanders RJ: **3-Methylglutaconic aciduria type I is caused by mutations in AUH.** *Am J Hum Genet* 2002, **71**:1463-1466.
  61. Hrebicek M, Mrazova L, Seyrantepe V, Durand S, Roslin NM, Noskova L, Hartmannova H, Ivanek R, Cizkova A, Poupetova H, Sikora J, Urinovska J, Stranecky V, Zeman J, Lepage P, Roquis D, Verner A, Arsoni J, Beesley CE, Maire I, Poorthuis BJ, van de Kamp J, van Diggelelen OP, Wevers RA, Hudson TJ, Fujiwara TM, Majewski J, Morgan K, Kmoch S, Pshzhetsky AV: **Mutations in TMEM76\* cause mucopolysaccharidosis IIIC (Sanfilippo C syndrome).** *Am J Hum Genet* 2006, **79**:807-819.
  62. De Meirleir L, Seneca S, Lissens W, De Clercq I, Eyskens F, Gerlo E, Smet J, Van Coster R: **Respiratory chain complex V deficiency due to a mutation in the assembly gene ATP12.** *J Med Genet* 2004, **41**:120-124.
  63. **MITOMAP – A human mitochondrial genome database** [<http://www.mitomap.org>]
  64. **Mitochondrial Proteome, Database for mitochondrial-related genes, proteins and diseases** [<http://www.mitop2.de>]
  65. **MiGenes Database** [<http://www.pharm.stonybrook.edu/migenes>]
  66. **MitoProteome Database** [<http://www.mitoproteome.org>]
  67. **Molecular Signatures Database** [[http://www.broad.mit.edu/gsea/msigdb/msigdb\\_index.html](http://www.broad.mit.edu/gsea/msigdb/msigdb_index.html)]
  68. **National Center for Biotechnology Information** [<http://www.ncbi.nlm.nih.gov/>]
  69. **Gene Ontology** [<http://www.godatabase.org/dev/database>]
  70. **UniProt – the universal protein resource** [<http://www.expasy.org/>]
  71. **BASE – BioArray Software Environment** [<http://base.thep.lu.se>]. <http://base.img.cas.cz>
  72. **OligoPicker** [<http://pga.mgh.harvard.edu/oligopicker/>]
  73. Klement P, Nijtmans LG, Van den Bogert C, Houstek J: **Analysis of oxidative phosphorylation complexes in cultured human fibroblasts and amniocytes by blue-native-electrophoresis using mitoplasts isolated with the help of digitonin.** *Anal Biochem* 1995, **231**:218-224.
  74. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M: **Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.** *Nat Genet* 2001, **29**:365-371.
  75. **The R Project for Statistical Computing** [<http://www.r-project.org/>]
  76. Smyth GK: *Limma: linear models for microarray data* New York: Springer; 2005.
  77. **BIOCONDUCTOR – open source software for bioinformatics** [<http://www.bioconductor.org/>]
  78. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing.** *Journal of the Royal Statistical Society* 1995, **B**, **57**:289-300.
  79. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovskiy I, Liu Z, Vinsavich A, Trush V, Quackenbush J: **TM4: a free, open-source system for microarray data management and analysis.** *Biotechniques* 2003, **34**:374-378.
  80. **TM4 – microarray software suite** [<http://www.tm4.org/>]

81. **KEGG: Kyoto Encyclopedia of Genes and Genomes** [<http://www.genome.jp/kegg/>]
82. Houstek J, Klement P, Floryk D, Antonicka H, Hermanska J, Kalous M, Hansikova H, Houtkova H, Chowdhury SK, Rosipal T, Kmoch S, Stratilova L, Zeman J: **A novel deficiency of mitochondrial ATPase of nuclear origin.** *Hum Mol Genet* 1999, **8**:1967-1974.

## *TMEM70* mutations cause isolated ATP synthase deficiency and neonatal mitochondrial encephalomyopathy

Alena Čížková<sup>1,2</sup>, Viktor Stránecký<sup>1</sup>, Johannes A Mayr<sup>3</sup>, Markéta Tesařová<sup>4</sup>, Vendula Havlíčková<sup>2</sup>, Jan Paul<sup>2</sup>, Robert Ivánek<sup>1</sup>, Andreas W Kuss<sup>5</sup>, Hana Hansíková<sup>4</sup>, Vilma Kaplanová<sup>2</sup>, Marek Vrbacký<sup>2</sup>, Hana Hartmannová<sup>1</sup>, Lenka Nosková<sup>1</sup>, Tomáš Honzík<sup>4</sup>, Zdeněk Drahota<sup>2</sup>, Martin Magner<sup>4</sup>, Kateřina Hejzlarová<sup>2</sup>, Wolfgang Sperl<sup>3</sup>, Jiří Zeman<sup>4</sup>, Josef Houštěk<sup>2</sup> & Stanislav Kmoch<sup>1</sup>

**We carried out whole-genome homozygosity mapping, gene expression analysis and DNA sequencing in individuals with isolated mitochondrial ATP synthase deficiency and identified disease-causing mutations in *TMEM70*. Complementation of the cell lines of these individuals with wild-type *TMEM70* restored biogenesis and metabolic function of the enzyme complex. Our results show that *TMEM70* is involved in mitochondrial ATP synthase biogenesis in higher eukaryotes.**

Mitochondrial ATP synthase, a key enzyme of mitochondrial energy provision, catalyzes synthesis of ATP during oxidative phosphorylation. ATP synthase is a 650-kDa protein complex composed of 16 types of subunits; 6 form the globular  $F_1$  catalytic part and 10 form the transmembraneous  $F_0$  part with two connecting stalks<sup>1</sup>. Two mammalian ATP synthase subunits, ATP6 and ATP8, are encoded by mtDNA; all the others are encoded by nuclear DNA. Biogenesis of ATP synthase is a stepwise process requiring a concerted action of assembly factors. Several of these factors have been described in yeast (for example, ATP10, ATP11, ATP12, ATP22, ATP23 and FMC1)<sup>2</sup>, but only three have been found in mammals—homologs of  $F_1$ -specific factors ATP11 and ATP12 (refs. 2–4) essential for assembly of  $F_1$  subunits  $\alpha$  and  $\beta$ , and a homolog of the  $F_0$ -related ATP23 with unclear function in mammals<sup>5</sup>.

Inherited disorders of ATP synthase belong to most deleterious mitochondrial diseases, which typically affect the pediatric population<sup>6</sup>. Maternally transmitted ATP synthase disorders are caused by heteroplasmic mutations of *MT-ATP6* (ref. 7) and rarely of *MT-ATP8* (ref. 8). These defects impair the energetic function of the  $F_0$  proton channel and thus prevent ATP synthesis, although the rate of ATP hydrolysis and the concentration of the enzyme complex remain largely unchanged. In contrast, ATP synthase defects of nuclear genetic

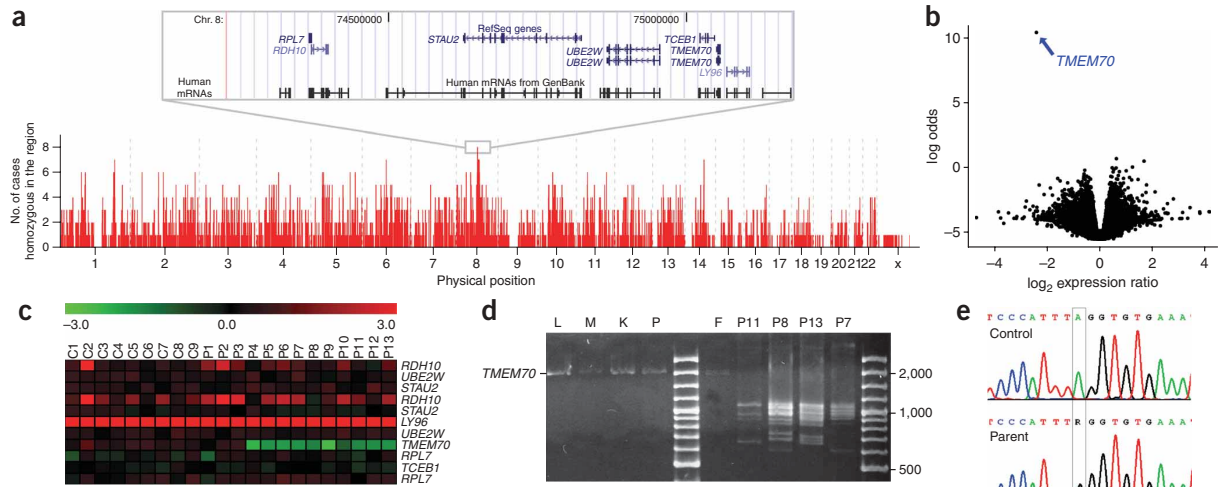
origin (MIM604273) are characterized by selective decrease of ATP synthase concentrations (to <30%) and a profound loss of both synthetic and hydrolytic activities<sup>9</sup>. Most affected individuals show neonatal lactic acidosis, hypertrophic cardiomyopathy and/or variable central nervous system involvement and 3-methylglutaconic aciduria. The disease outcome is severe, and half of affected individuals die in early childhood<sup>10</sup>. During the last decade, an increasing number of affected individuals, mostly of Roma (Gypsy) ethnic origin, have been reported<sup>10–13</sup>, but a mutation affecting the  $F_1$ -specific factor ATP12 was only found in one case<sup>11</sup>. To identify the genetic defect in the other affected individuals with isolated deficiency of ATP synthase we used Affymetrix GeneChip Mapping 250K arrays and genotyped eight index affected individuals, their healthy siblings and parents from six families (Supplementary Methods and Supplementary Fig. 1 online) and performed linkage analysis (Supplementary Fig. 2 online) and homozygosity mapping (Fig. 1a and Supplementary Fig. 3 online). To prioritize candidate genes, we intersected the mapping information with Agilent 44K array gene expression data<sup>13</sup>. This analysis illuminated a single gene, *TMEM70*, as it has previously been localized in a top-candidate region on chromosome 8 (Fig. 1a), showed reduced transcript amount in fibroblast cell lines from affected individuals (Fig. 1b,c,d) and encodes what has been characterized as a mitochondrial protein<sup>14</sup>. Through sequence analysis of genomic DNA (Supplementary Table 1 online), we identified in affected individuals a homozygous substitution, 317-2A>G, located in the splice site of intron 2 of *TMEM70* (NM-017866; Fig. 1e), which leads to aberrant splicing and loss of *TMEM70* transcript (Fig. 1b,d). We carried out PCR-RFLP analysis in investigated families and proved autosomal recessive segregation of the mutation, as all the affected individuals were homozygous, all parents were heterozygous and unaffected siblings showed either the wild-type or heterozygous genotype. We screened for the 317-2A>G mutation among 25 individuals with low ATP synthase content being studied in our institutions, and found 23 who were homozygous for the mutation (Supplementary Table 2 online). In an additional single heterozygous individual, P27, we identified on the second allele the frameshift mutation 118\_119insGT (Supplementary Fig. 4 online), which encodes a truncated *TMEM70* protein, Ser40CysfsX11. We did not find any mutation in affected individual P3, in whom *TMEM70* transcript amount was also unchanged (Fig. 1c). We did not find any of the identified mutations in 100 control individuals.

To prove that *TMEM70* is necessary for the biogenesis of the ATP synthase, we carried out RT-PCR analysis of several human tissues (Fig. 1d) and found no evidence of distinct *TMEM70* splicing variants

<sup>1</sup>Institute of Inherited Metabolic Disorders, Charles University of Prague, First Faculty of Medicine, Prague 12808, Czech Republic. <sup>2</sup>Department of Bioenergetics, Institute of Physiology, Academy of Science of the Czech Republic, Prague 14220, Czech Republic. <sup>3</sup>Department of Pediatrics, Paracelsus Medical University, Salzburg A5020, Austria. <sup>4</sup>Department of Pediatrics, Charles University of Prague, First Faculty of Medicine, Prague 12808, Czech Republic. <sup>5</sup>Max Planck Institute for Molecular Genetics, Berlin 14195, Germany. Correspondence should be addressed to J.H. (houstek@biomed.cas.cz) or S.K. (skmoch@f1.cuni.cz).

Received 3 April; accepted 28 August; published online 26 October 2008; doi:10.1038/ng.246

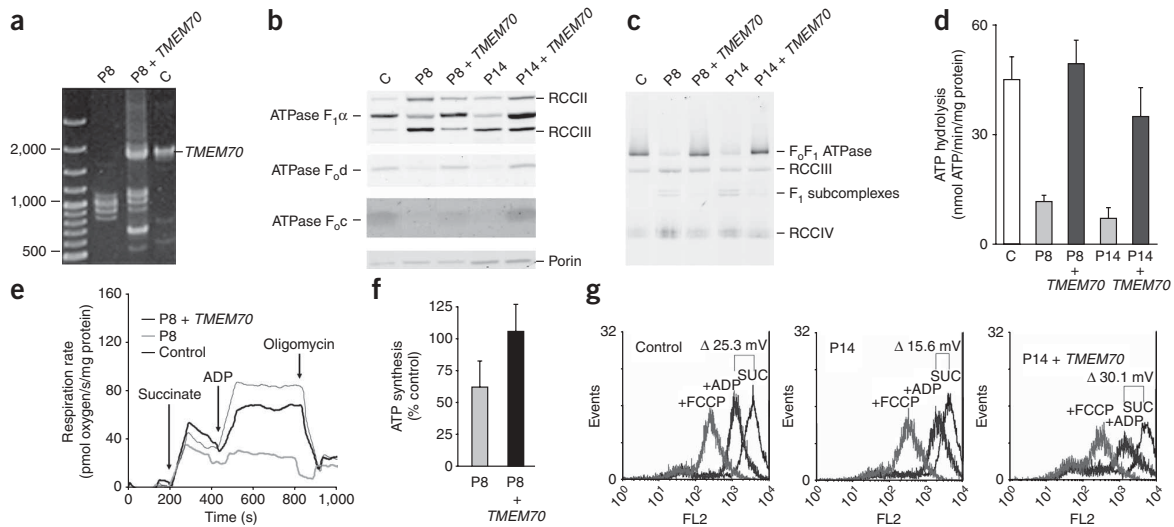




**Figure 1** Positional cloning of *TMEM70*. (a) A genome map showing the number and location of overlapping homozygosity regions identified in eight index individuals with ATP synthase deficiencies. Physical position and gene content of the top-candidate region on chromosome 8 are shown above. (b) Gene expression changes between the case and control fibroblast cell lines. The logarithm of the probability that the gene is differentially expressed (log odds) is plotted as a function of the logarithm of the expression ratio ( $\log_2$  expression ratio) between the case and control samples. (c) Expression matrix of the genes located in the candidate region showing reduced *TMEM70* transcript amount in all but one case (P3) with a nuclear defect. Normal *TMEM70* transcript amount is present in cases P1 and P2, as these individuals had the mt9205 $\Delta$ TA microdeletion and not the *TMEM70* mutation. The results are shown as  $\log_2$  ratios of gene expression signal in each sample to that of a common reference sample. (d) *TMEM70* cDNA analysis. L, M, K, P and F lanes show presence of a single RT-PCR product in control human liver, muscle, kidney, pancreas and fibroblasts, respectively. Abnormal RT-PCR products were observed in fibroblasts from affected individuals (lanes P11, P8, P13 and P7). (e) Chromatograms of *TMEM70* genomic DNA sequence showing 317-2A>G substitution.

reported in genomic databases. We cloned *TMEM70* cDNA into the pEF-DEST51 expression vector and transfected skin fibroblast cell lines of several affected individuals (Fig. 2a). We found that trans-

ected cells increased the amount of both  $F_1$  and  $F_0$  structural subunits of ATP synthase (Fig. 2b) and produced normal concentrations of the full size, assembled ATP synthase complex (Fig. 2c). Consequently,



**Figure 2** *TMEM70* complementation of ATP synthase deficiency. (a) *TMEM70* cDNA is present after transfection. (b) SDS-PAGE protein blot of fibroblasts shows a specific increase of the content of ATP synthase subunits relative to the respiratory chain complexes and porin. (c) BN-PAGE protein blot of mitochondria shows increase of the full-size assembled ATP synthase 650-kDa complex relative to respiratory chain complexes. (d) Oligomycin-sensitive ATP synthase hydrolytic activity is restored. (e) ADP stimulation is enhanced in digitonin-permeabilized cells. (f) Analysis of ATP formation shows restoration of mitochondrial ATP synthesis. (g) TMRM cytofluorometric measurements in permeabilized cells show restoration of the ADP-induced drop of mitochondrial membrane potential at state 4. Data in d and f are shown as mean  $\pm$  s.d.;  $n = 3$ .

the vector restored oligomycin-sensitive ATP hydrolysis (Fig. 2d), ADP-stimulated respiration (Fig. 2e), mitochondrial ATP synthesis (Fig. 2f) and ADP-induced decrease of mitochondrial membrane potential (Fig. 2g).

TMEM70 contains the conserved domain DUF1301 and two putative transmembrane regions. Using phylogenetic analysis, we found *TMEM70* homologs in genomes of multicellular eukaryotes and plants, but not in yeast and fungi (Supplementary Fig. 5 online). This indicates that the evolution of TMEM70 may be an important factor accounting for differences in the ATP synthase assembly process in higher eukaryotes, yeast and bacteria<sup>2,3</sup>.

We have identified TMEM70 as a protein involved in the biogenesis of the ATP synthase in higher eukaryotes and shown that its defect is relatively frequent among individuals, particularly Romanians, with mitochondrial energy provision disorders. Existence of the prevalent mutation and co-occurrence of cases with severe and milder phenotypes, probably representing varying quality and functionality of individual nonsense-mediated RNA decay systems, open a way for investigation of translational bypass therapy in this group of individuals.

Note: Supplementary information is available on the Nature Genetics website.

#### ACKNOWLEDGMENTS

This study was supported by grants from Ministry of Education of Czech Republic (1M6837805002, AV0Z 50110509, MSM0021620806, Kontakt 14/2006), GAČR (305/08/H037), OeNB 12568, Päd. Forschungsverein and PMU Salzburg (06/04/022). We thank R.Gallyová, Š.Rosipal, V.Smolka, A.Hlavatá, P.Freisinger, M.Huemer and O.Bodamer, who provided samples from affected individuals for this study, and D. Seelow for bioinformatic support.

#### AUTHOR CONTRIBUTIONS

A.C., H.Hartmannová and L.N. carried out DNA and gene expression analysis and TMEM70 cloning. V.S. and R.I. were responsible for genotyping, gene expression analysis and bioinformatics. J.A.M. carried out biochemical diagnosis and DNA analysis. A.W.K. did genotyping and homozygosity mapping. M.T. and H.Hansková carried out biochemical diagnosis, cell culturing and transfections. V.H., J.P. and V.K. carried out transfections, complementation studies, ELFO/WB analysis and bioinformatics. M.V., Z.D. and K.H. were responsible for functional studies. T.H. and M.M. were responsible for family ascertainment and sample collection, and J.Z. and W.S. handled diagnosis and clinical characterization. S.K. and J.H. initiated and coordinated the study and wrote the manuscript.

Published online at <http://www.nature.com/naturegenetics/>  
Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Collinson, I.R., Skehel, J.M., Fearnley, I.M., Runswick, M.J. & Walker, J.E. *Biochemistry* **35**, 12640–12646 (1996).
- Ackerman, S.H. & Tzagoloff, A. *Prog. Nucleic Acid Res. Mol. Biol.* **80**, 95–133 (2005).
- Pickova, A., Potocky, M. & Houstek, J. *Proteins* **59**, 393–402 (2005).
- Wang, Z.G., White, P.S. & Ackerman, S.H. *J. Biol. Chem.* **276**, 30773–30778 (2001).
- Zeng, X., Neupert, W. & Tzagoloff, A. *Mol. Biol. Cell* **18**, 617–626 (2007).
- Houstek, J. *et al. Biochim. Biophys. Acta* **1757**, 1400–1405 (2006).
- Schon, E.A., Santra, S., Pallotti, F. & Girvin, M.E. *Semin. Cell Dev. Biol.* **12**, 441–448 (2001).
- Jonckheere, A. *et al. J. Med. Genet.* **45**, 129–133 (2007).
- Houstek, J. *et al. Hum. Mol. Genet.* **8**, 1967–1974 (1999).
- Sperl, W. *et al. Neuromuscul. Disord.* **16**, 821–829 (2006).
- De Meirleir, L. *et al. J. Med. Genet.* **41**, 120–124 (2004).
- Mayr, J.A. *et al. Pediatr. Res.* **55**, 988–994 (2004).
- Cizkova, A. *et al. BMC Genomics* **9**, 38 (2008).
- Calvo, S. *et al. Nat. Genet.* **38**, 576–582 (2006).

## Mutations in *TMEM76*\* Cause Mucopolysaccharidosis IIIC (Sanfilippo C Syndrome)

Martin Hřebíček, Lenka Mrázová, Volkan Seyrantepe, Stéphanie Durand, Nicole M. Roslin, Lenka Nosková, Hana Hartmannová, Robert Ivánek, Alena Čížková, Helena Poupětová, Jakub Sikora, Jana Uřinová, Viktor Stránecký, Jiří Zeman, Pierre Lepage, David Roquis, Andrei Verner, Jérôme Ausseil, Clare E. Beesley, Irène Maire, Ben J. H. M. Poorthuis, Jiddeke van de Kamp, Otto P. van Diggelen, Ron A. Wevers, Thomas J. Hudson, T. Mary Fujiwara, Jacek Majewski, Kenneth Morgan, Stanislav Kmoch,<sup>†</sup> and Alexey V. Pshezhetsky

Mucopolysaccharidosis IIIC (MPS IIIC, or Sanfilippo C syndrome) is a lysosomal storage disorder caused by the inherited deficiency of the lysosomal membrane enzyme acetyl-coenzyme A:α-glucosaminide N-acetyltransferase (*N*-acetyltransferase), which leads to impaired degradation of heparan sulfate. We report the narrowing of the candidate region to a 2.6-cM interval between *D8S1051* and *D8S1831* and the identification of the transmembrane protein 76 gene (*TMEM76*), which encodes a 73-kDa protein with predicted multiple transmembrane domains and glycosylation sites, as the gene that causes MPS IIIC when it is mutated. Four nonsense mutations, 3 frameshift mutations due to deletions or a duplication, 6 splice-site mutations, and 14 missense mutations were identified among 30 probands with MPS IIIC. Functional expression of human *TMEM76* and the mouse ortholog demonstrates that it is the gene that encodes the lysosomal *N*-acetyltransferase and suggests that this enzyme belongs to a new structural class of proteins that transport the activated acetyl residues across the cell membrane.

Heparan sulfate is a polysaccharide found in proteoglycans associated with the cell membrane in nearly all cells. The lysosomal membrane enzyme, acetyl-coenzyme A (CoA):α-glucosaminide *N*-acetyltransferase (*N*-acetyltransferase) is required to *N*-acetylate the terminal glucosamine residues of heparan sulfate before hydrolysis by the α-*N*-acetyl glucosaminidase. Since the acetyl-CoA substrate would be rapidly degraded in the lysosome,<sup>1</sup> *N*-acetyltransferase employs a unique mechanism, acting both as an enzyme and a membrane channel, and catalyzes the transmembrane acetylation of heparan sulfate.<sup>2</sup> The mechanism by which this is achieved has been the topic of considerable investigation, but, for many years, the isolation and cloning of *N*-acetyltransferase has been hampered by its low tissue content, instability, and hydrophobic nature.<sup>3–5</sup>

Genetic deficiency of *N*-acetyltransferase causes mucopolysaccharidosis IIIC (MPS IIIC [MIM 252930], or Sanfilippo syndrome C), a rare autosomal recessive lysosomal disorder of mucopolysaccharide catabolism.<sup>6–8</sup> MPS IIIC is clinically similar to other subtypes of Sanfilippo syn-

drome.<sup>9</sup> Patients manifest symptoms during childhood with progressive and severe neurological deterioration causing hyperactivity, sleep disorders, and loss of speech accompanied by behavioral abnormalities, neuropsychiatric problems, mental retardation, hearing loss, and relatively minor visceral manifestations, such as mild hepatomegaly, mild dwarfism with joint stiffness and biconvex dorsolumbar vertebral bodies, mild coarse faces, and hypertrichosis.<sup>7</sup> Most patients die before adulthood, but some survive to the 4th decade and show progressive dementia and retinitis pigmentosa. Soon after the first 3 patients with MPS IIIC were described by Kresse et al.,<sup>6</sup> Klein et al.<sup>8,10</sup> reported a similar deficiency in 11 patients who had received the diagnosis of Sanfilippo syndrome, therefore suggesting that the disease is a relatively frequent subtype. The birth prevalence of MPS IIIC in Australia,<sup>11</sup> Portugal,<sup>12</sup> and the Netherlands<sup>13</sup> has been estimated to be 0.07, 0.12, and 0.21 per 100,000, respectively.

The putative chromosomal locus of the MPS IIIC gene was first reported in 1992. By studying two siblings who received the diagnosis of MPS IIIC and had an apparently

From the Institute for Inherited Metabolic Disorders (M.H.; L.M.; L.N.; H.H.; R.I.; A.Č.; H.P.; J.S.; J.U.; V. Stránecký; J.Z.; S.K.) and Center for Applied Genomics (R.I.; A.Č.; V. Stránecký; J.Z.; S.K.), Charles University 1st School of Medicine, and Institute of Molecular Genetics, Academy of Sciences of the Czech Republic (R.I.), Prague; Hôpital Sainte-Justine and Départements de Pédiatrie (V. Seyrantepe; S.D.; J.A.; A.V.P.) and Biochimie (A.V.P.), Université de Montréal, and Research Institute of the McGill University Health Centre (N.M.R.; T.J.H.; T.M.F.; K.M.), McGill University and Genome Quebec Innovation Centre (P.L.; D.R.; A.V.; T.J.H.; J.M.), and Departments of Human Genetics (T.J.H.; T.M.F.; J.M.; K.M.), Medicine (T.J.H.; T.M.F.; K.M.), and Anatomy and Cell Biology (A.V.P.), McGill University, Montreal; Biochemistry, Endocrinology & Metabolism Unit, UCL Institute of Child Health, London (C.E.B.); Hôpital Debrousse, Lyon, France (I.M.); Department of Medical Biochemistry, Academic Medical Center UVA (B.J.H.M.P.), and Department of Clinical Genetics, VU University Medical Center (J.v.d.K.), Amsterdam; Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands (O.P.v.D.); and Laboratory of Pediatrics and Neurology, University Medical Center, Nijmegen, The Netherlands (R.A.W.)

Received June 8, 2006; accepted for publication August 8, 2006; electronically published September 8, 2006.

Address for correspondence and reprints: Dr. Alexey V. Pshezhetsky, Service de Génétique Médicale, Hôpital Sainte-Justine, 3175 Côte Sainte-Catherine, Montreal, Quebec H3T 1C5, Canada. E-mail: alexei.pchejetski@umontreal.ca

\* Footnote added in proof: the gene name has been changed to *HGSNAT*.

<sup>†</sup> S.K. has led the Prague team.

*Am. J. Hum. Genet.* 2006;79:807–819. © 2006 by The American Society of Human Genetics. All rights reserved. 0002-9297/2006/7905-0004\$15.00

balanced Robertsonian translocation, Zaremba et al.<sup>14</sup> suggested that the mutant gene may be located in the pericentric region of either chromosome 14 or chromosome 21, but no further confirmation of this finding was provided. Previously, we performed a genomewide scan on 27 patients with MPS IIIC and 17 unaffected family members, using 392 highly informative microsatellite markers with an average interspacing of 10 cM. For chromosome 8, the scan showed an apparent excess of homozygosity in patients compared with their unaffected relatives.<sup>15</sup> Additional genotyping of 38 patients with MPS IIIC for 22 markers on chromosome 8 identified 15 consecutive markers (from *D8S1051* to *D8S2332*) in an 8.3-cM interval for which the genotypes of affected siblings were identical in state. A maximum multipoint LOD score of 10.6 was found at marker *D8S519*, suggesting that this region includes the locus for MPS IIIC.<sup>15</sup> Recently, localization of the MPS IIIC causative gene on chromosome 8 was confirmed by microcell-mediated chromosome transfer in cultured skin fibroblasts of patients with MPS IIIC.<sup>16</sup>

Here, we report the results of linkage analyses that narrowed the candidate region for MPC IIIC to a 2.6-cM interval between *D8S1051* and *D8S1831* and the identification of the *TMEM76* gene, located within the candidate region, as the gene that codes for the lysosomal *N*-acetyltransferase and, when mutated, is responsible for MPS IIIC.

## Material and Methods

### Families

In Montreal, 33 affected individuals and 35 unaffected relatives comprising 15 families informative for linkage were genotyped. The families came from Europe, North Africa, and North America. An additional 27 affected individuals and 9 unaffected relatives in uninformative pedigrees, as well as 40 controls, were also genotyped. Eleven of these families and the controls have been reported elsewhere.<sup>15</sup> In addition, 54 individuals from four MPS IIIC-affected families from the Czech Republic were studied in Prague (fig. 1). One family had two affected brothers, whereas the remaining three families each had one affected individual. The families came from various regions of the Czech Republic and were not related within the four most-recent generations. The diagnosis for affected individuals was confirmed by the measurement of *N*-acetyltransferase activity in cultured skin fibroblasts or white blood cells.

### Genotyping

The samples in Montreal were genotyped for 22 microsatellite markers in the pericentromeric region of chromosome 8 spanning 8.9 cM on the Rutgers map, version 2.0.<sup>17</sup> The genotyping was performed as described by Mira et al.<sup>18</sup> at the McGill University and Genome Quebec Innovation Centre on an ABI 3730xl DNA Analyzer platform (Applied Biosystems). Alleles were assigned using Genotyper, version 3.6 (Applied Biosystems). The random-error model of SimWalk2, version 2.91,<sup>19,20</sup> was used to detect potential genotyping errors, with an overall error rate of 0.025. Nine genotypes for which the posterior probability of being incorrect was >0.5 were removed before subsequent analyses. In

addition, nine genotypes for one marker in one family were removed because of a suspected microsatellite mutation. The samples from the Czech Republic were genotyped in Prague for 18 microsatellite markers in an 18.7-cM region that includes the 8.9-cM region mentioned above. The genotyping was performed on an LI-COR IR2 sequencer by use of Saga genotyping software (Li-Cor) as described elsewhere.<sup>21</sup> Genotypes were screened for errors by use of the PedCheck program.<sup>22</sup>

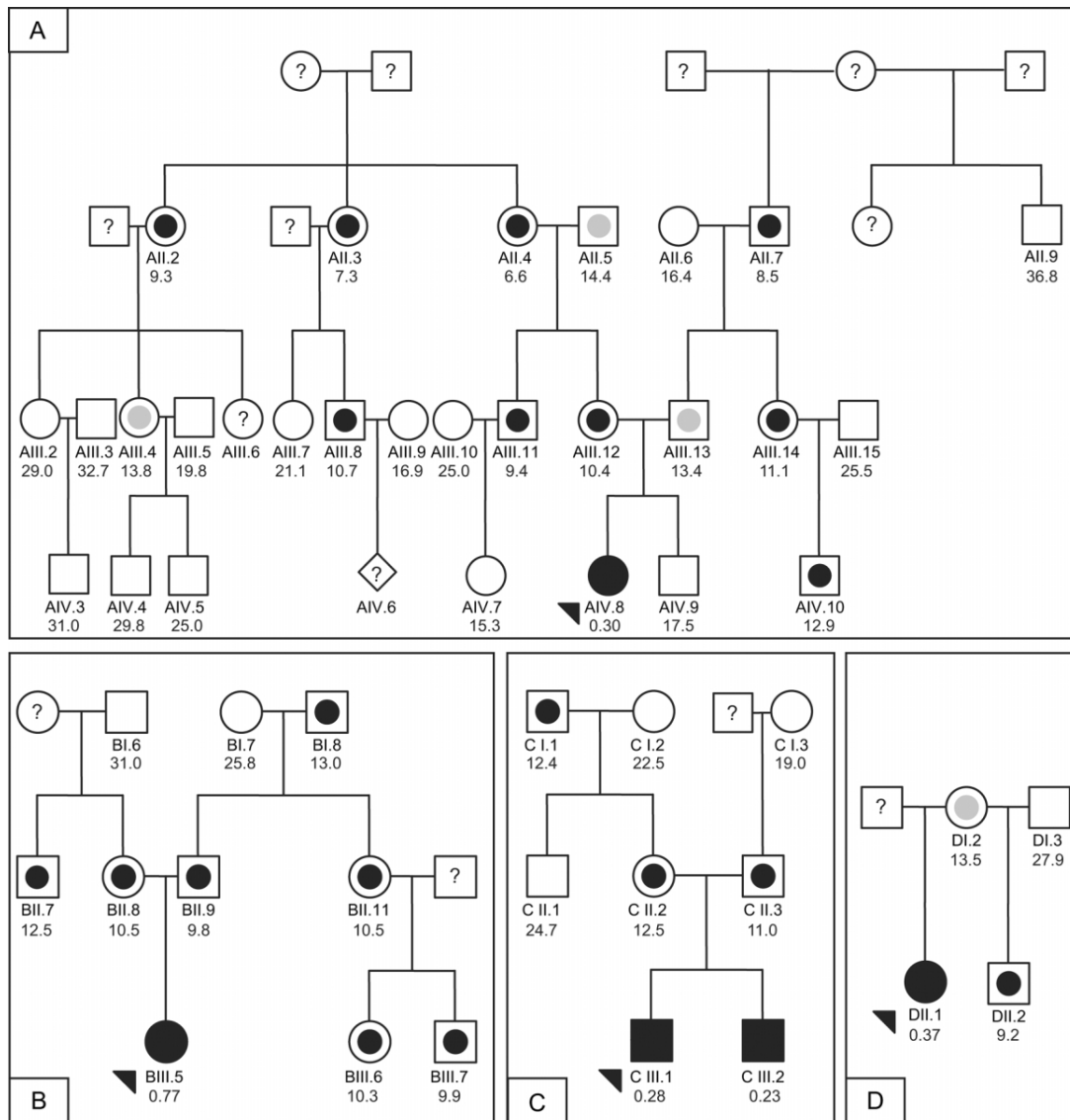
### Linkage Analysis

For the families genotyped in Montreal, multipoint linkage analysis was performed using the Markov chain–Monte Carlo (MCMC) method implemented in SimWalk2, version 2.91,<sup>19</sup> since one pedigree was too large to be analyzed by exact computation. A fully penetrant autosomal recessive parametric model was used with a disease-allele frequency of 0.0045. Marker-allele frequencies were estimated by counting alleles in the available parents of patients with MPS IIIC and in control individuals. To check the consistency of the results, the MCMC analysis was repeated four times.

*N*-acetyltransferase activity was measured in all participants of the four families from the Czech Republic.<sup>23</sup> Individuals were classified as affected, carriers, or unaffected on the basis of the results of this assay. Mean affected and carrier activities were determined from the five affected individuals and their seven obligate heterozygote parents, respectively, whereas the mean control activity was determined from a sample of 89 unrelated individuals. Four individuals were unable to be classified because their values were within 2 SDs of the means of both the control and carrier groups. Multipoint linkage analysis was performed using a codominant model with a penetrance of 0.99 and a phenocopy rate of 0.01, to account for the possibility of misclassification or genotyping errors. The same disease-allele frequency of 0.0045 was used. Marker-allele frequencies were estimated by counting all genotyped individuals. Exact multipoint linkage analysis was run on 18 microsatellite markers by use of Allegro 1.2c,<sup>24</sup> which was also used to infer haplotypes.

### Gene-Expression Analysis

For each of 32 genes located in the candidate interval, a single 5'-amino-modified 40-mer oligonucleotide probe (Illumina) was spotted in quadruplicate on aminosilane-modified microscopic slides and was immobilized using a combination of baking and UV cross-linking. Total RNA (250–1,000 ng) from white blood cells of two patients with MPS IIIC (patients AIV.8 and BIII.5) and four healthy individuals were amplified using the SenseAmp plus RNA Amplification Kit (Genisphere) and were reverse transcribed using 300 ng of poly(A)-tailed mRNA. Reverse transcription and microarray detection were done using the Array 900 Expression Detection Kit (Genisphere) according to the manufacturer's protocol. The two patient samples and four control samples were analyzed in dye-swap mode, in two replicates of each mode. The hybridized slides were scanned with a GenePix 4200A scanner (Molecular Devices), with photomultiplier gains adjusted to obtain the highest-intensity unsaturated images. Data analysis was performed in the R statistical environment (The R Project for Statistical Computing, version 2.2.1) by use of the Linear Models for Microarray Data package (Limma, version 2.2.0).<sup>25</sup> Raw data were processed using loess normalization and a moving minimum background correction on individual arrays and quantile



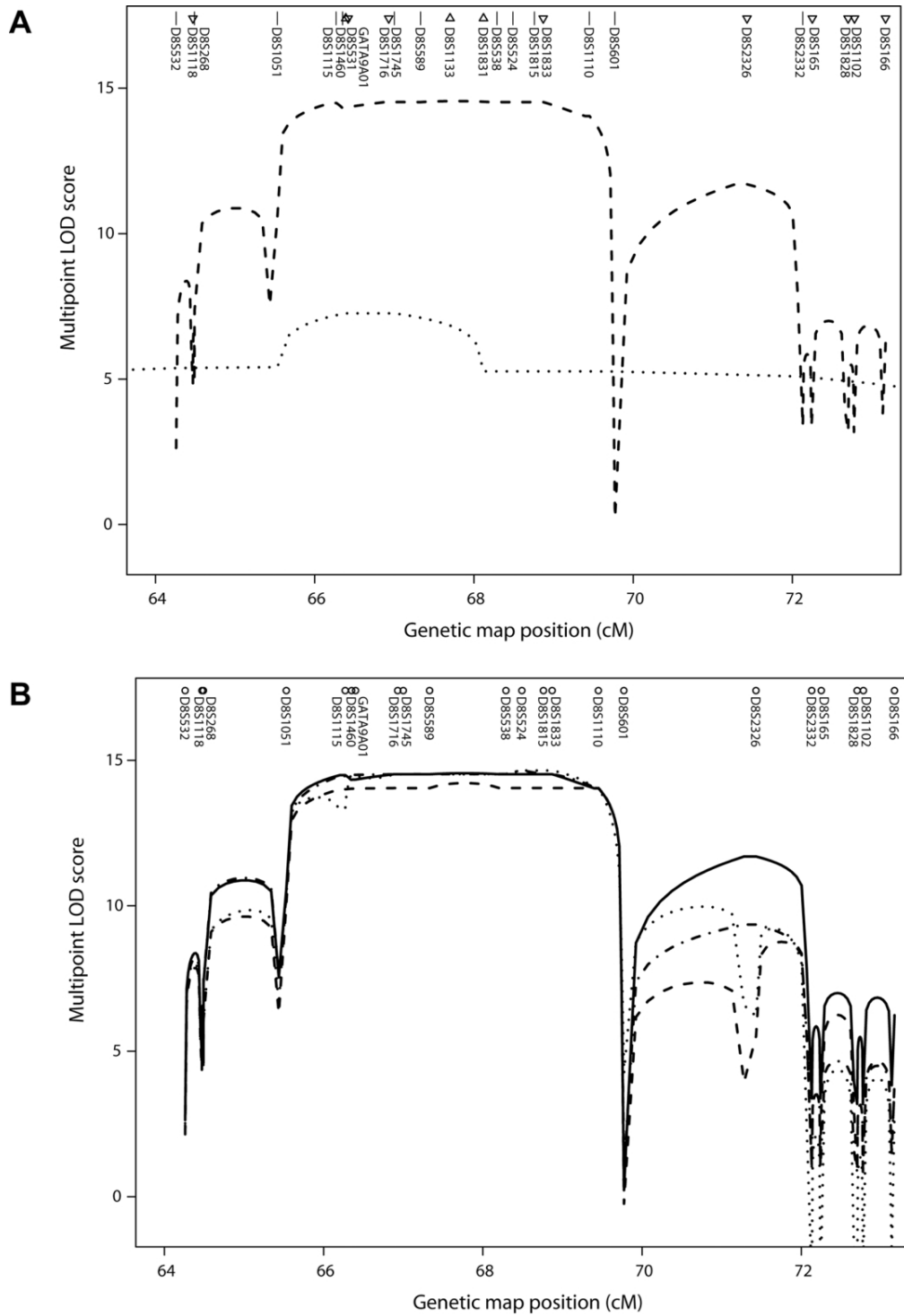
**Figure 1.** Four families from the Czech Republic used in the linkage and mutation analyses. Fully blackened symbols indicate individuals with MPS IIIC; arrowheads indicate probands. Measurements in seven obligate heterozygotes from these pedigrees (mean  $\pm$  SD 11.6  $\pm$  1.5 nmol/h/mg) and 89 controls not known to be related to members of the pedigree (mean  $\pm$  SD 24.4  $\pm$  5.7 nmol/h/mg) were used to establish *N*-acetyltransferase activity ranges for heterozygotes (symbols with blackened inner circle) and normal homozygotes (open symbols). An individual was assigned to a class if his or her enzyme activity was within 2 SDs of the class, unless the value was within the overlap of the upper end of the obligate heterozygotes and the lower end of the controls. Individuals with values within the open interval 13.0–14.6 nmol/h/mg were classified as unknown (symbols with gray inner circle). A symbol with a question mark (?) indicates that no material was available for the enzyme assay. DNA was available for individuals with ID numbers, and *N*-acetyltransferase activity measurements in white blood cells are shown below the ID numbers.

normalization between arrays. The correlation between four duplicate spots per gene on each array was used to increase the robustness. A linear model was fitted for each gene given a series of arrays by use of the lmFit function. The empirical Bayes method<sup>26</sup> was used to rank the differential expression of genes by use of the eBayes function. Correction for multiple testing was performed using the Benjamini and Hochberg false-discovery–

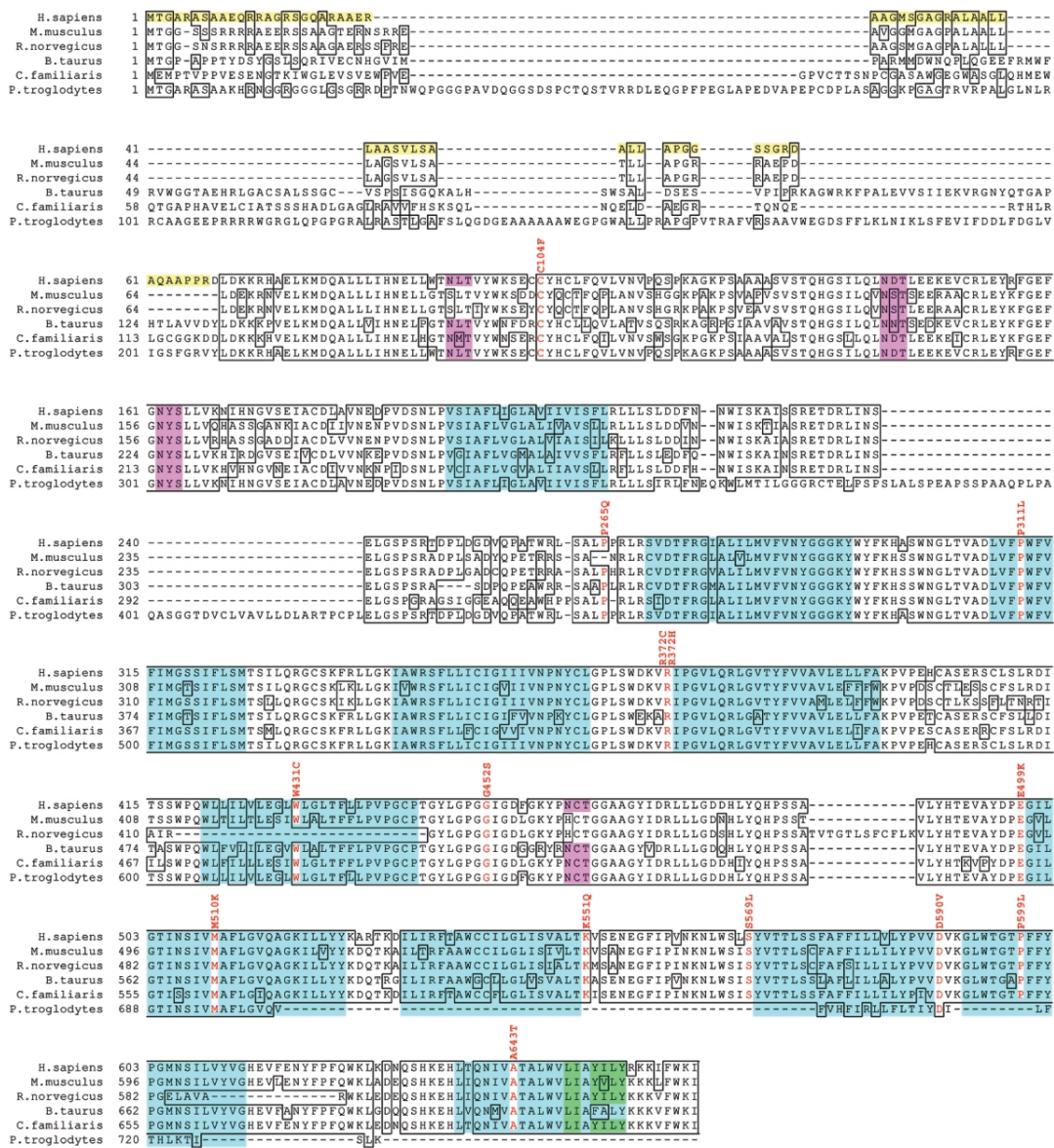
rate method.<sup>27</sup> We considered genes to be differentially expressed if the adjusted *P* value was <.01.

#### DNA and RNA Isolation and Sequencing

Cultured skin fibroblasts from patients with MPS IIIC and normal controls were obtained from cell depositories (Hôpital Debrousse,



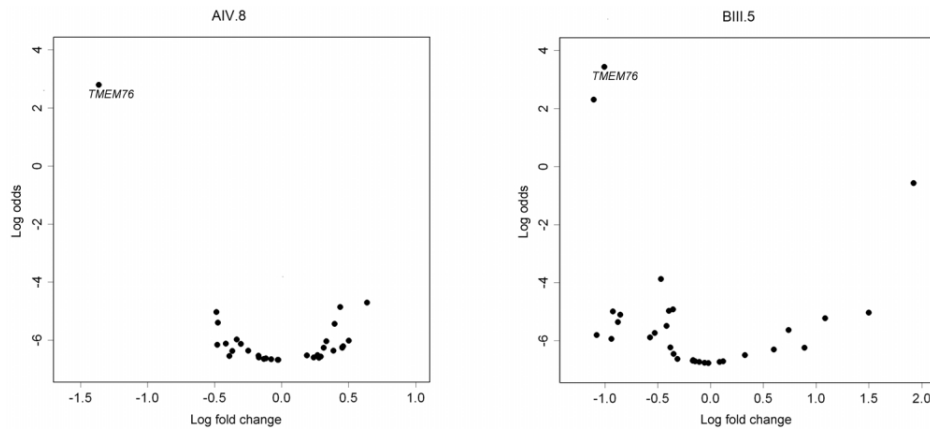
**Figure 2.** Multipoint linkage analysis of MPS IIIC on chromosome 8. *A*, Multipoint LOD scores in an 8.9-cM interval from two sets of families. Symbols above the marker names indicate the map position. Marker names are listed in the correct order but may be displaced from the symbols for visibility. The dashed line is based on families genotyped in Montreal, and the dotted line on families genotyped in Prague. Straight lines next to marker names indicate that the markers were typed in both data sets. Triangles pointing down indicate markers typed only in the Montreal data set, and triangles pointing up indicate markers typed only in the Prague data set. For the Montreal data, the SimWalk2 run with the highest likelihood is shown. *TMEM76* lies between *D8S1115* and *D8S1460*, and, according to the March 2006 freeze of the human genome sequence from the University of California–Santa Cruz Genome Browser,<sup>30</sup> the order is *D8S1115*–(500 kb)–*TMEM76*–(800 kb)–centromere–(200 kb)–*D8S1460*. *B*, Multipoint LOD scores from the Montreal data from four runs of SimWalk2, version 2.91,<sup>19</sup> showing the variation between runs.



**Figure 3.** Predicted amino acid sequence of the TMEM76 protein. Amino acid sequence alignment of *Homo sapiens* TMEM76 with orthologs from *Mus musculus* (cloned sequence), *Canis familiaris* (GenBank accession number XP\_539948.2), *Bos taurus* (XP\_588978.2), *Rattus norvegicus* (XP\_341451.2), and *Pan troglodytes* (XP\_519741.1) by use of BLAST. All cDNA sequences are predicted except the sequence for *M. musculus*. The identical residues are boxed, the residues with missense mutations in patients with MPS IIIC are shown in red, and the amino acid changes are indicated above the sequence. The first 67 aa of the human sequence shown as black on yellow comprise the predicted signal peptide. The predicted transmembrane domains in the human sequence are shown as black on turquoise. The topology model<sup>5-7</sup> strongly predicts that the N-terminus is inside the lysosome and the C-terminus is outside. Four predicted N-glycosylation sites are shown as black on pink, and the predicted motifs for the lysosomal targeting, as black on green.

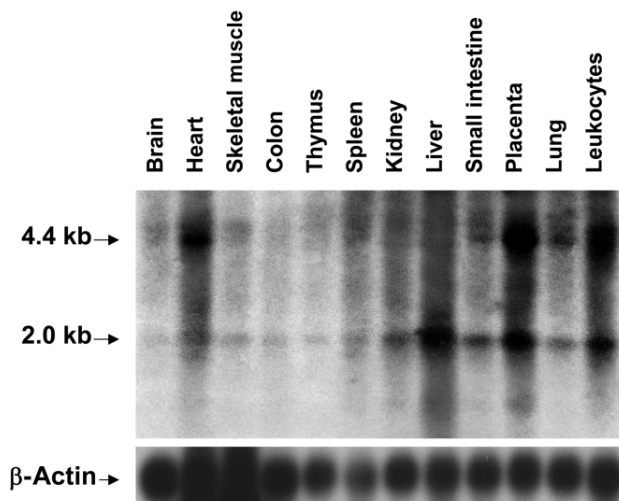
France; NIGMS Human Genetic Mutant Cell Repository; Montreal Children's Hospital, Canada; and Department of Clinical Genetics, Erasmus Medical Center, The Netherlands). Blood samples from patients with MPS IIIC, their relatives, and controls were collected with ethics approval from the appropriate institutional review boards. DNA from blood or cultured skin fibroblasts was extracted using the PureGene kit (Gentra Systems). Total RNA

from cultured skin fibroblasts and pooled tissues (spleen, liver, kidney, heart, lung, and brain) of a C57BL/6J mouse was isolated using Trizol (Invitrogen), and first-strand cDNA synthesis was prepared with SuperScript II (Invitrogen). DNA fragments containing *TMEM76* exons and adjacent regions (~40 bp from each side; primer sequences are shown in appendix A) were amplified by PCR from genomic DNA and were purified with Montage PCR96



**Figure 4.** Volcano plot of genes located within the MPS IIIC candidate region, showing significantly reduced expression of the *TMEM76* gene in white blood cells of two patients with MPS IIIC: AIV.8 and BIII.5. The natural logarithm of the probability that the gene is differentially expressed (Log odds) is plotted as a function of the logarithm of the gene-expression  $\log_2$  fold change (Log fold change) between the patient and control samples.

filter plates (Millipore). Each sequencing reaction contained 2  $\mu$ l of purified PCR product, 5.25  $\mu$ l of H<sub>2</sub>O, 1.75  $\mu$ l of 5 $\times$  sequencing buffer, 0.5  $\mu$ l of 20  $\mu$ M primer, and 0.5  $\mu$ l of Big Dye Terminator v3.1 (all from Applied Biosystems). In Montreal, PCR products were analyzed using an ABI 3730xl DNA Analyzer (Applied Biosystems). In Prague, PCR products were analyzed on an ALFexpress DNA sequencer (Pharmacia), as described elsewhere.<sup>28</sup> Included in the sequencing analysis were 30 probands with MPS IIIC who were considered unrelated and 105 controls. The controls were unrelated CEPH individuals, and amplified DNAs were combined in pools of two before sequencing.



**Figure 5.** Northern-blot analysis of *TMEM76* mRNA in human tissues. A 12-lane blot containing 1  $\mu$ g of poly A<sup>+</sup> RNA per lane from various adult human tissues was hybridized with a [<sup>32</sup>P]-labeled 220-bp cDNA fragment corresponding to exons 8–10 of the *TMEM76* gene or  $\beta$ -actin, as described in the Material and Methods section.

#### Northern Blotting

A 12-lane multiple-tissue northern blot containing 1  $\mu$ g of poly A<sup>+</sup> RNA per lane from various human tissues (BD Biosciences Clontech) was hybridized with the 220-bp cDNA fragment corresponding to exons 8–10 of the human *TMEM76* gene or the entire cDNA of human  $\beta$ -actin labeled with [<sup>32</sup>P]-dCTP by random priming with the MegaPrime labeling kit (Amersham). Prehybridization of the blot was performed at 68°C for 30 min in ExpressHyb (Clontech). The denatured probes were added directly to the prehybridization solution and were incubated at 68°C for 1 h. The blots were washed twice for 30 min at room temperature with 2 $\times$  sodium chloride–sodium citrate (SSC) solution and 0.05% SDS and once for 40 min at 50°C with 0.1 $\times$  SSC and 0.1% SDS and were exposed to a BioMax film for 48 h.

#### Mouse and Human *TMEM76* cDNA Cloning

Mouse coding sequence was amplified by PCR (forward primer 5'-GAATTCATGACGGGCGGGTTCGAGC-3'; reverse primer 5'-ATATGTCGACGATTTTCCAAAACAGCTTC-3') and was cloned into pCMV-Script, pCMV-Tag4A (Stratagene), and pEGFP-N3 (BD Biosciences Clontech) vectors by use of the *Eco*RI and *Sal*I restriction sites of the primers. The cloned sequence was identical to GenBank accession number AK152926.1, except that an "AT" was needed to complete an alternate ATG initiation codon. GenBank accession number AK149883.1 provides what we consider to be the complete clone and encodes a 656-aa protein. The GenBank sequences differ by 1 aa and three silent substitutions.

A 1,907-bp fragment of the human *TMEM76* cDNA (nt +75 to +1992) was amplified using Platinum High Fidelity *Taq* DNA polymerase (Invitrogen), a sense primer with an *Hind*III site (5'-AAGCTTGGCGGGCGGGCATGAG-3'), and an antisense primer with an *Sal*I site (5'-GTCGACCTCAGTGGGAGCCATCAGATTTT-3') and was cloned into pCMV-Script expression vector (Stratagene). Since high GC content (85%) of the 5' region of human *TMEM76* cDNA prevented its amplification by PCR, a synthetic 186-bp codon-optimized double-stranded oligonucleotide fragment (5'-AAGCTTATGACCGGAGCGAGGGCAAGCGCCGCG-



AACAAAGAAGAGCCGGACGGTCCGGCCAGGCTAGGGCCGC-AGAGCGAGCTGCTGGCATGTCAGGTGCAGGGCGCGCACTTGCCGCTTGCTGCTCGCCGCGAGTGTGCTGAGCGTGCCCTCCTGGCTCCCGGAGGCTCTCCGGGCGGGAC-3') corresponding to nt +1 to +186 of human *TMEM76* cDNA was purchased from BioS&T. A 177-bp 5' fragment was combined with rest of the cDNA by use of *HindIII* and *SapI* sites. The cloned sequence is identical to GenBank accession number XM\_372038.4 from nt 131 to nt 1946, except for the presence of SNP *rs1126058*.

### Cell Culture and Transfection

Skin fibroblasts and COS-7 cells were cultured in Eagle's minimal essential medium (Invitrogen) supplemented with 10% (v/v) fetal bovine serum (Invitrogen) and were transfected with the full-size mouse *Tmem76* (*Hgsnat*) coding sequence subcloned into pCMV-Script, pCMV-Tag4A, and pEGFP-N3 vectors or with the full-size human *TMEM76* coding sequence subcloned into pCMV-Script vector by use of Lipofectamine Plus (Invitrogen) according to the manufacturer's protocol. The cells were harvested 48 h after transfection, and *N*-acetyltransferase activity was measured in the homogenates of *TMEM76*-transfected and mock-transfected cells (i.e., transfected with only the cloning vector).

### Enzyme Assay

*N*-acetyltransferase enzymatic activity was measured using the fluorogenic substrate 4-methylumbelliferyl  $\beta$ -D-glucosaminide (Moscerdam) as described elsewhere.<sup>23</sup> Protein concentration was measured according to the method of Bradford.<sup>29</sup> This assay was used for the activity measurements in cultured skin fibroblasts or white blood cells from patients and all participating members of the Czech families and for the functional expression experiments.

### Confocal Microscopy

To establish colocalization of the tagged protein with the lysosomal compartment, the skin fibroblasts expressing mouse *TMEM76*-EGFP were treated with 50 nM LysoTracker Red DND-99 dye (Molecular Probes), were washed twice with ice-cold PBS, and were fixed with 4% paraformaldehyde in PBS for 30 min. Slides were studied on an LMS 510 Meta inverted confocal microscope (Zeiss).

## Results

### Linkage Analysis

Previously, we performed a genomewide linkage study that indicated that the locus for MPS IIIC is mapped to an 8.3-cM interval in the pericentromeric region of chromosome 8.<sup>15</sup> To reduce this interval, we genotyped the families from that study as well as newly obtained MPS IIIC-affected families for 22 microsatellite markers (Montreal data). Linkage analysis under an autosomal recessive model resulted in LOD scores >14 in the 4.2-cM region spanning *D8S1051* to *D8S601*, which included the centromere (fig. 2). The results of multiple MCMC runs showed consistent trends. Linkage was also performed in four families from the Czech Republic by use of an autosomal codominant model (Prague data). For these data, linkage analysis produced a maximum LOD score of 7.8 at 66.4 cM at *D8S531* and reduced the linked region for

the Montreal data to a 2.6-cM interval between *D8S1051* and *D8S1831*. This region was defined by inferred recombinants at *D8S1051* in one family in each of the Montreal and Prague data sets, and a recombinant at *D8S1831* in an additional family in the Prague data set. This interval contains 32 known or predicted genes and ORFs.

### Identification of a Candidate Gene

On the basis of our previous studies that defined the molecular properties of the lysosomal *N*-acetyltransferase,<sup>31</sup> we searched the candidate region for a gene encoding a protein with multiple transmembrane domains and a molecular weight of ~100 kDa, which allowed us to exclude the majority of the genes in the region. In contrast, the predicted protein product of the *TMEM76* gene has multiple putative transmembrane domains. The predicted coding region in GenBank accession number XM\_372038.4 was extended by 28 residues at the 5' end on the basis of the transcript in GenBank accession number DR000652.1 (which includes 14 of the 28 residues), examination of the genomic sequence in NT\_007995.14, and comparison with mouse sequence AK149883.1. We predict that the modified *TMEM76* contains 18 exons, corresponding to an ORF of 1,992 bp, and codes for a 73-kDa protein. A comparison of human *TMEM76* with five vertebrate orthologs is shown in figure 3. Furthermore, of all the genes present in the candidate interval, only *TMEM76* showed a statistically significant reduction of the transcript level in the cells of two patients with MPS IIIC (AIV.8 and BIII.5; adjusted *P* values < .001) in the custom oligonucleotide-based microarray assay (fig. 4). Further, we showed that both patients carried nonsense mutations presumably causing mRNA decay (R534X and L349X; see table 2).

### Analysis of the *TMEM76* Transcript by Northern Blot and RT-PCR

Northern-blot analysis identified two major *TMEM76* transcripts of 4.5 and 2.1 kb ubiquitously expressed in various human tissues (fig. 5). The highest expression was detected in leukocytes, heart, lung, placenta, and liver, whereas the gene was expressed at a much lower level in the thymus, colon, and brain, which is consistent with the expression patterns of lysosomal proteins. Consistent with the northern-blot results, a full-length 4.5-kb cDNA containing 1,992 bp of coding sequence and two polyadenylation signals as well as two shorter transcripts were amplified by RT-PCR from the total RNA of normal human skin fibroblasts, white blood cells, and skeletal muscle. In one transcript, exons 9 and 10 were spliced out, leading to an in-frame deletion of 64 aa, which contains the predicted transmembrane domains III and IV. Most likely, this transcript does not encode an active enzyme, since it was also detected in the RNA of two patients with MPS IIIC (patients CIII.1 and CIII.2) who had almost complete loss of

*N*-acetyltransferase activity. Another transcript lacked exons 3, 9, and 10.

The deduced amino acid sequence predicts 11 transmembrane domains and four potential *N*-glycosylation sites (fig. 3), consistent with the molecular properties of lysosomal *N*-acetyltransferase.<sup>31</sup> The first 67 aa may comprise the signal peptide, with length and composition resembling those of lysosomal proteins. According to the predictions made by empirical computer algorithms,<sup>32-34</sup> the C-terminus of the *TMEM76* protein is exposed to the cytoplasm and contains conserved Tyr-X-X-Θ and Leu-Leu sequence motifs involved in the interaction with the adaptor proteins responsible for the lysosomal targeting of membrane proteins.<sup>35</sup>

#### Mutation Analysis

We identified 27 *TMEM76* mutations in the DNA of 30 MPS IIIC-affected families (table 1) that were not found in DNA from 105 controls. Among the identified mutations, there were 4 nonsense mutations, 14 missense mu-

tations, 3 predicted frameshift mutations due to deletions or duplications, and 6 splice-site mutations. All the missense mutations occur at residues conserved among five species with the most homologous *TMEM76* sequences (fig. 3), except for P265Q, which is not conserved in the mouse, and W431C, which is not conserved in the rat. There were three instances of two mutations on the same allele that were found in patients who were homozygous, and these are designated as complex mutations in table 1. cDNA sequencing of one of the patients homozygous for the splice-site mutation in intron 2 and a missense mutation (P265Q) demonstrated that the splice-site mutation disrupts the consensus splice-site sequence between exon 2 and intron 2 and causes exon 2 skipping and a frameshift (not shown).

Consanguinity was reported in 4 of the 13 families in which the patients were homozygous for *TMEM76* mutations: the two Moroccan families, the French family with two missense mutations (W431C and A643T), and the Turkish family with the splice-site mutation in intron

**Table 1. Mutations in *TMEM76* Identified in Patients from 30 Families with MPS IIIC**

Mutation Group and Mutation <sup>a</sup>	Predicted Effect on Protein	No. of Alleles	Location in <i>TMEM76</i>
Nonsense mutations:			
c.1031G→A	p.W344X	2	Exon 10
c.1046T→G	p.L349X	1	Exon 10
c.1234C→T	p.R412X	8	Exon 12
c.1600C→T	p.R534X	1	Exon 15
Missense mutations:			
c.311G→T	p.C104F	1	Exon 2
c.932C→T	p.P311L	3	Exon 9
c.1114C→T	p.R372C	3	Exon 11
c.1115G→A	p.R372H	1	Exon 11
c.1354G→A	p.G452S	2	Exon 13
c.1495G→A	p.E499K	3	Exon 14
c.1529T→A	p.M510K	1	Exon 14
c.1706C→T	p.S569L	4	Exon 17
c.1769A→T	p.D590V	1	Exon 17
c.1796C→T	p.P599L	1	Exon 17
Frameshift mutations:			
c.1118_1133del	p.I373SfsX3	1	Exon 11
c.1420_1456dup	p.V488GfsX22	1	Exon 13
c.1834delG	p.V612SfsX16	1	Exon 18
Splice-site mutations:			
c.202+1G→A	p.L69EfsX32 <sup>b</sup>	1	Intron 1
c.577+1G→A	p.P193HfsX20 <sup>b</sup>	1	Intron 4
c.935+5G→A	p.F313X	1	Intron 9
c.1334+1G→A	p.G446X <sup>b</sup>	1	Intron 12
c.1810+1G→A	p.S567NfsX14	2	Intron 17
Complex mutations:			
c.[318+1G→A; 794C→A]	p.[D68VfsX19; P265Q]	6	Intron 2; exon 7
c.[577+1G→A; 1650A→C]	p.[P193HfsX20; K551Q]	2	Intron 4; exon 16
c.[1293G→T; 1927G→A]	p.[W431C; A643T]	2	Exon 12; exon 18

<sup>a</sup> Mutation names were assigned according to the guidelines of the Human Genome Variation Society and on the basis of the cDNA sequence from GenBank accession number NT\_007995.14, except that the first exon includes 84 nt 5' of the stated ATG initiation codon. Thus, +1 corresponds to the A of the ATG at nt 13315945 (instead of nt 13316029).

<sup>b</sup> The mutations were named under the assumption that no exon skipping takes place; cDNA sequencing was not done.

17. The two Moroccan families were not known to be related to each other or to the Spanish patient homozygous for the same mutations (table 2). The parents of the French patient are second cousins in two ways (see family F1 in the work of Ausseil et al.<sup>15</sup>).

The splice-site mutation in the above-mentioned Turkish family disrupts the consensus splice-site sequence between exon 17 and intron 17 and causes exon 17 skipping and a frameshift in all transcripts, as detected by sequencing of multiple RT-PCR clones (not shown). The two affected siblings in this family (family F8 in the work of Ausseil et al.<sup>15</sup>) had a severe form of MPS IIIC and showed almost complete loss of *N*-acetyltransferase activity in cultured skin fibroblasts. Among other severely affected patients with MPS IIIC, a patient of French origin was homozygous for a nonsense mutation (W344X) in exon 10, which may result in the synthesis of a truncated protein or RNA decay. A patient of Polish origin was a compound heterozygote for a 37-bp duplication in exon 13 and a missense mutation (S569L) in exon 17 (table 2). The duplication results in a frameshift, whereas the substitution of a strictly conserved small polar Ser for a bulky hydrophobic Leu may have a significant structural impact (fig. 3).

The five patients from four Czech families are all compound heterozygotes for eight different mutations (table 2). Five of the eight mutations are predicted to result in truncated products (three nonsense mutations, one 16-bp

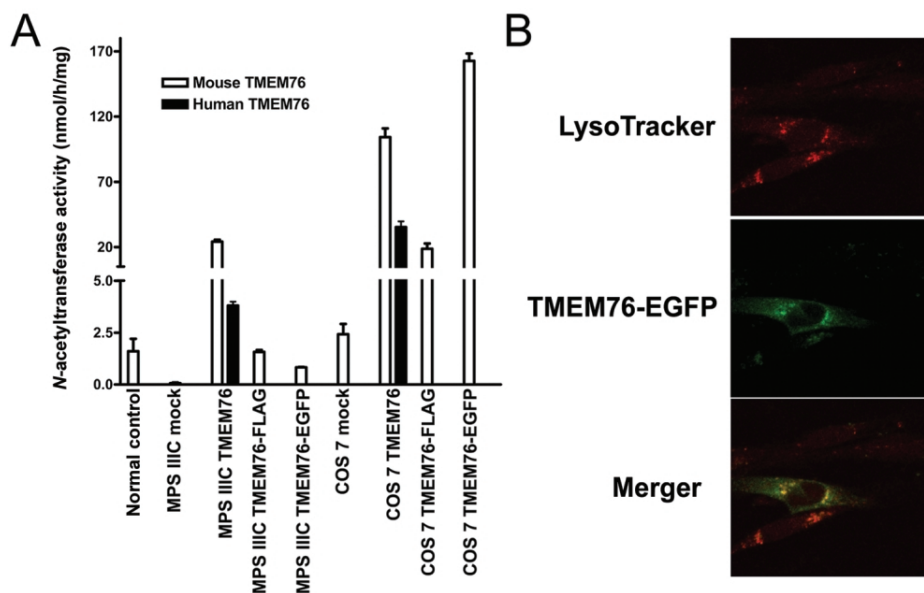
deletion, and one splice-site mutation leading to the inclusion of 89 bases from the 5' end of intron 9 and the splicing out of exon 10 in the transcript, and the remaining three are missense mutations affecting residues conserved among multiple species and located either in the predicted transmembrane regions (fig. 3) or in their close vicinity, suggesting that they may have a serious structural impact. In the Czech families, the mutations completely segregated with reduced enzyme activity. That is, all individuals assigned to be heterozygotes on the basis of the enzyme assay as well as the four individuals who were within 2 SD of the lower end of the controls (symbols with gray inner circle in fig. 1) were found to carry *TMEM76* mutations.

#### Functional Expression Studies

The fibroblast cell line from a patient homozygous for a splice-site mutation in intron 17 with negligible *N*-acetyltransferase activity was transfected with plasmids containing human *TMEM76* cDNA or cDNA of the mouse ortholog of *TMEM76* carrying a FLAG tag on the C-terminus or of a fusion protein of mouse *TMEM76* with enhanced green fluorescent protein (EGFP). All constructs increased the *N*-acetyltransferase activity in the mutant fibroblast cells to approximately normal level (fig. 6A). Significant increase in activity was also observed in transfected COS-7 cells, confirming that the *TMEM76* protein

**Table 2. *TMEM76* Predicted Mutations in Proband from 30 Families with MPS IIIC**

Patient Group and Mutation 1	Mutation 2	No. of Patients	Geographic Origin of Patient(s)
Patients from Czech families:			
p.I373SfsX3	p.R534X	1	Czech Republic
p.L349X	p.M510K	1	Czech Republic
p.F313X	p.R412X	1	Czech Republic
p.R372H	p.P599L	1	Czech Republic
Patients homozygous for <i>TMEM76</i> mutations:			
p.[D68VfsX19; P265Q]	p.[D68VfsX19; P265Q]	3	Morocco, Morocco, and Spain
p.[P193HfsX20; K551Q]	p.[P193HfsX20; K551Q]	1	France
p.P311L	p.P311L	1	United Kingdom
p.W344X	p.W344X	1	France
p.R372C	p.R372C	1	United Kingdom
p.R412X	p.R412X	2	Turkey and Poland
p.[W431C; A643T]	p.[W431C; A643T]	1	France
p.G452S	p.G452S	1	Canada
p.E499K	p.E499K	1	Canada
p.S567NfsX14	p.S567NfsX14	1	Turkey
Patients compound heterozygous for <i>TMEM76</i> mutations:			
p.C104F	...	1	Belarus
p.E499K	p.D590V	1	France
p.P193HfsX20	p.R412X	1	Canada
p.P311L	p.R372C	1	France
p.R412X	...	1	Poland
p.R412X	p.G446X	1	Poland
p.S569L	...	2	France and Portugal
p.S569L	p.L69EfsX32	1	United States
p.V488GfsX22	p.S569L	1	Poland
p.V612SfsX16	...	1	Finland
Families with no mutations identified to date	...	2	North Africa and Portugal



**Figure 6.** Functional expression of human and mouse *TMEM76* protein. *A*, The full-size human and mouse *TMEM76* coding sequences subcloned into pCMV-Script, pCMV-Tag4A, and pEGFP-N3 vectors were expressed in COS-7 cells and in cultured skin fibroblasts from a patient with MPS IIIC. The cells were harvested 48 h after transfection, and *N*-acetyltransferase activity was measured in the homogenates of *TMEM76*-transfected and mock-transfected fibroblast or COS-7 cells by use of the artificial fluorometric substrate 4-methylumbelliferyl- $\beta$ -D-glucosaminide.<sup>23</sup> Values represent means  $\pm$  SD of four independent experiments. *B*, The intracellular localization of *TMEM76* was studied by expressing the fusion protein of the mouse *TMEM76* with EGFP. Before fixation, the cells were treated for 45 min with 50 nM lysosomal marker, LysoTracker Red DND-99 dye. Slides were analyzed on an LMS 510 Meta confocal microscope (Zeiss). Magnification  $\times$  1000. The image was randomly selected from 30 studied panels, all of which showed a similar localization of *TMEM76*-EGFP. The fluorescence of EGFP was not quenched as it would have been if the fluorophore had been exposed to the acidic lysosomal microenvironment, confirming that the C-terminus of *TMEM76* faces the cytoplasmic side of the lysosomal membrane.

by itself has *N*-acetyltransferase activity. Confocal fluorescent microscopy shows that *TMEM76*-EGFP (fig. 6B) or *TMEM76*-FLAG (not shown) peptides are targeted in human fibroblasts to cytoplasmic organelles, colocalizing with the lysosomal-endosomal marker LysoTracker Red.

## Discussion

Degradation of heparan sulfate occurs within the lysosomes by the concerted action of a group of at least eight enzymes: four sulfatases, three exo-glycosidases, and one *N*-acetyltransferase, which work sequentially at the terminus of heparan sulfate chains, producing free sulfate and monosaccharides. The inherited deficiencies of four enzymes involved in the degradation of heparan sulfate cause four subtypes of MPS III: MPS IIIA (heparan *N*-sulfatase deficiency [MIM 252900]), MPS IIIB ( $\alpha$ -*N*-acetylglucosaminidase deficiency [MIM 252920]), MPS IIIC (acetyl-CoA: $\alpha$ -glucosaminide acetyltransferase deficiency), and MPS IIID (*N*-acetylglucosamine 6-sulfatase deficiency [MIM 252940]). Since the clinical phenotypes of all these disorders are similar, precise diagnosis relies on the determination of enzymatic activities in patients' cultured skin fibroblasts or leukocytes. The biochemical defect in MPS IIIC was identified 30 years ago as a deficiency of an en-

zyme that transfers an acetyl group from cytoplasmically derived acetyl-CoA to terminal  $\alpha$ -glucosamine residues of heparan sulfate within the lysosomes, resulting in the accumulation of heparan sulfate. Therefore, for identification of the molecular basis of this disorder, we used two complementary approaches. First, we performed a partial purification of human and mouse lysosomal *N*-acetyltransferase, which suggested that the enzyme has properties of an oligomeric transmembrane glycoprotein, with an  $\sim$ 100-kDa polypeptide containing the enzyme active site.<sup>31</sup> Second, by linkage analysis, we narrowed the locus for MPC IIIC to a 2.6 cM-interval (*D8S1051-D8S1831*) and, third, compared the level of transcripts of the genes present in the candidate region between normal control cells and those from patients with MPS IIIC. Thus, an integrated bioinformatic search and gene-expression analysis both pinpointed a single gene, *TMEM76*, which encodes a 73-kDa protein with predicted multiple transmembrane domains and glycosylation sites. DNA mutation analysis showed that patients with MPS IIIC harbor *TMEM76* mutations incompatible with the normal function of the predicted protein, whereas expression of human *TMEM76* and the mouse ortholog proved that the protein has *N*-acetyltransferase activity and lysosomal lo-

calization, providing evidence that *TMEM76* is the gene that codes for the lysosomal *N*-acetyltransferase.

The *TMEM76* protein does not show a structural similarity to any known prokaryotic or eukaryotic *N*-acetyltransferases or to other lysosomal proteins, on the basis of sequence homology searches. Thus, we think that it belongs to a new structural class of proteins capable of transporting the activated acetyl residues across the cell membrane. Moreover, *TMEM76* shares homology with a conserved family of bacterial proteins COG4299 (uncharacterized protein conserved in bacteria) (Entrez Gene GeneID 138050). All 146 members of this family are predicted proteins from diverse bacterial species, including Proteobacteria, Cyanobacteria, and Deinococci. Since many of these bacteria are capable of synthesizing heparan sulfate and other structurally related glycosaminoglycans and perform reactions of transmembrane acetylation, it is tempting to speculate that this activity may also be performed by the proteins of the COG4299 family. Previous studies suggested two contradictory mechanisms of transmembrane acetylation. Bame and Rome<sup>2,36,37</sup> proposed that it is performed via a ping-pong mechanism. First, the acetyl group of acetyl-CoA is transferred to an His residue in the active site of the enzyme. This induces a conformational change that results in the translocation of the protein domain containing the acetylated residue to the lysosome, where the acetyl residue is transferred to the glucosamine residue of heparan sulfate. In contrast, Meikle et al.<sup>38</sup> were unable to demonstrate any specific acetylation of the lysosomal membranes and proposed an alternative mechanism that involved the formation of a tertiary complex of the enzyme, acetyl-CoA, and heparan sulfate. Identification of *N*-acetyltransferase as a 73-kDa protein with multiple transmembrane domains, together with our previous data that showed that *N*-acetyltransferase is acetylated by [<sup>14</sup>C]acetyl-CoA in the absence of glucosamine,<sup>31</sup> strongly supports the ping-pong mechanism of transmembrane acetylation.

For 23 of the 30 probands included in this study for mutation analysis, *TMEM76* mutations were identified in both alleles. Five probands were heterozygous for a missense mutation, with a second mutation yet to be identified. In two probands from North Africa and Portugal,

we did not identify any mutations in the coding regions or immediate flanking regions. These patients are homozygous for the microsatellite markers throughout the entire MPS IIIC locus and may be homozygous for a yet-to-be-identified *TMEM76* mutation; however, we cannot formally exclude defects in other genes. Additional studies have been initiated to search for mutations in the introns and promoter regions. The patients with MPS IIIC with the identified frameshift and nonsense mutations all have a clinically severe early-onset form. The almost complete deficiency of *N*-acetyltransferase activity in cultured skin fibroblasts from these patients is consistent with the predicted protein truncations and/or nonsense-mediated mRNA decay. Further expression studies are necessary to confirm the impact of the identified substitutions of the conserved amino acids on enzyme activity. Nevertheless, the identification of the lysosomal *N*-acetyltransferase gene which, when mutated, accounts for the molecular defect in patients with MPS IIIC sets the stage for DNA-based diagnosis and genotype-phenotype correlation studies and marks the end of the gene-discovery phase for lysosomal genetic enzymopathies.

#### Acknowledgments

We thank the patients, their families, and the Czech Society for Mucopolysaccharidosis, for participating in our study, and members of the sequencing and genotyping facilities at the McGill University and Genome Quebec Innovation Centre, for their technical support. We also acknowledge Nina Gusina, Joe Clarke, and Tony Rugar, for providing cell lines from patients with MPS IIIC; Mila Ashmarina, Milan Elleder, J. Loredó-Osti, and Johanna Rommens, for helpful discussions; Karine Landry, for technical support; and Maryssa Canuel, for help with confocal microscopy. The Montreal study was supported by operating grants from the Sanfilippo Children's Research Foundation (to A.V.P.) and by the Canadian Networks of Centres of Excellence Program—the Mathematics of Information Technology and Complex System network (to K.M.). The Prague study was supported by grants NR8069-1 and 1A/8239-3 from the Grant Agency of the Ministry of Health of the Czech Republic. Institutional support was provided by Ministry of Education of the Czech Republic grant MSM0021620806. A.V.P. is a National Investigator of the Fonds de la Recherche en Santé du Québec.

**Table A1. Exon-Flanking Primers Used for PCR Amplification and Sequencing of the Exons in the Human *TMEM76* Gene**

Primer	Sequence (5'→3')
TMEM76_Exon1_F	CTCCCGAAGACAAACACTCC
TMEM76_Exon1_R	GCGAAGTCGCAGCAACAGC
TMEM76_Exon2_F	AAGCTTTTGAGAAGCACTACTGG
TMEM76_Exon2_R	GAAGGGCTTTAGACATGAGAGC
TMEM76_Exon3_F	GGAAAAGTCATGTCAGGATCTCC
TMEM76_Exon3_R	GAATAATACATGTTCTGGGTACG
TMEM76_Exon4_F	TTATTCTGCCTCCATGATATTAGC
TMEM76_Exon4_R	CTACAGAAAAGCGTCATGGACTGC
TMEM76_Exon5_F	GGAAATTCAGCATGAGAATATAGC
TMEM76_Exon5_R	GCCACTTGAGGGTGACAGC
TMEM76_Exon6_F	GAATATGAGCTTAAATTTTATTTCC
TMEM76_Exon6_R	TTAGGAATACGGGAGCTACAACC
TMEM76_Exon7_F	CAAAATGAAATTTACCCCTTAGC
TMEM76_Exon7_R	ACATCCAAGAAATCCTTCTAGC
TMEM76_Exon8_F	CCTTCTTTTCCATAGCAAACC
TMEM76_Exon8_R	GCTCTGTGAAGGACGTATATAAGC
TMEM76_Exon9_F	CCCTGGGTTTACTTTTCTATACC
TMEM76_Exon9_R	CCAGCATCATCTGAAAAACAGG
TMEM76_Exon10_F	GGGGCTATATTCTGAACTCTTCC
TMEM76_Exon10_R	ACCTGAGATGGAGGAATTGC
TMEM76_Exon11_F	CTGGGATGAGAGGAGAAGTCC
TMEM76_Exon11_R	ACTGAAGCCAGGAGTGAGG
TMEM76_Exon12_F	CCTTCTATTGCATTAGTTCACC
TMEM76_Exon12_R	GAGAATCTCTGACTCGAGACC
TMEM76_Exon13_F	TTTTATTCTGTCCCTCTGTTCG
TMEM76_Exon13_R	CACCTCTGAAAGCCTGAGTTCC
TMEM76_Exon14_F	TTGGTCTAGGAGCTGTTGTACG
TMEM76_Exon14_R	CCATAGCACAAAGAGAGAATATGC
TMEM76_Exon15_F	TCTTTGTCAGGTAGTTAAGACAGTGG
TMEM76_Exon15_R	GTGAAGGAAAGGAATTTTAGC
TMEM76_Exon16_F	ACAAGTTTCAGCCCTCTCTACG
TMEM76_Exon16_R	GTGGAGGAGACGTTTCAGTGC
TMEM76_Exon17_F	ATGCTGAAATTGGATTTGTCC
TMEM76_Exon17_R	ACCAAGGATGCTCCAGAGG
TMEM76_Exon18_F	AGTAGCCAACAATGGAAGTGC
TMEM76_Exon18_R	GAGCCGTGTCACAGTTAAC

NOTE.—For bidirectional sequencing on the ALFexpress DNA sequencer, all primers have the universal overhang synthesized on the 5' end (AATACGACTCACTATAG for forward [F] primers and CAGGAAACAGCTATGAC for reverse [R] primers).

### Web Resources

Accession numbers and URLs for data presented herein are as follows:

BLAST, <http://www.ncbi.nlm.nih.gov/blast/> (used to identify ortholog protein sequences)

Entrez Gene, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene> (for GeneID 138050)

GenBank, <http://www.ncbi.nlm.nih.gov/Genbank/> (for accession numbers AK152926.1, AK149883.1, DR000652.1, XM\_372038.4, NT\_007995.14, XP\_539948.2, XP\_588978.2, XP\_341451.2, and XP\_519741.1)

Human Genome Variation Society, <http://www.hgvs.org/>

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/> (for MPS IIIA, IIIB, IIIC, and IID)

### References

- Rome LH, Hill DF, Bame KJ, Crain LR (1983) Utilization of exogenously added acetyl coenzyme A by intact isolated lysosomes. *J Biol Chem* 258:3006–3011
- Bame KJ, Rome LH (1985) Acetyl-coenzyme A:α-glucosaminide N-acetyltransferase: evidence for a transmembrane acetylation mechanism. *J Biol Chem* 260:11293–11299
- Pohlmann R, Klein U, Fromme HG, von Figura K (1981) Localisation of acetyl-CoA: α-glucosaminide N-acetyltransferase in microsomes and lysosomes of rat liver. *Hoppe Seylers Z Physiol Chem* 362:1199–1207
- Hopwood JJ, Freeman C, Clements PR, Stein R, Miller AL (1983) Cellular location of N-acetyltransferase activities toward glucosamine and glucosamine-6-phosphate in cultured human skin fibroblasts. *Biochem Int* 6:823–830
- Meikle PJ, Whittle AM, Hopwood JJ (1995) Human acetyl-coenzyme A:α-glucosaminide N-acetyltransferase: kinetic characterization and mechanistic interpretation. *Biochem J* 308:327–333
- Kresse H, von Figura K, Bartsocas C (1976) Clinical and biochemical findings in a family with Sanfilippo disease, type C. *Clin Genet* 10:364
- Bartsocas C, Grobe H, van de Kamp JJ, von Figura K, Kresse H, Klein U, Giesberts MA (1979) Sanfilippo type C disease: clinical findings in four patients with a new variant of mucopolysaccharidosis III. *Eur J Pediatr* 130:251–258
- Klein U, Kresse H, von Figura K (1978) Sanfilippo syndrome type C: deficiency of acetyl-CoA:α-glucosaminide N-acetyltransferase in skin fibroblasts. *Proc Natl Acad Sci USA* 75: 5185–5189
- Sanfilippo SJ, Podosin R, Langer LO Jr, Good RA (1963) Mental retardation associated with acid mucopolysacchariduria (heparitin sulfate type). *J Pediatr* 63:837–838
- Klein U, van de Kamp JJP, von Figura K, Pohlmann R (1981) Sanfilippo syndrome type C: assay for acetyl-CoA:α-glucosaminide N-acetyltransferase in leukocytes for detection of homozygous and heterozygous individuals. *Clin Genet* 20:55–59
- Meikle PJ, Hopwood JJ, Clague AE, Carey WF (1999) Prevalence of lysosomal storage disorders. *JAMA* 281:249–254
- Pinto R, Caseiro C, Lemos M, Lopes L, Fontes A, Ribeiro H, Pinto E, Silva E, Rocha S, Marcao A, Ribeiro I, Lacerda L, Ribeiro G, Amaral O, Sa Miranda MC (2004) Prevalence of lysosomal storage diseases in Portugal. *Eur J Hum Genet* 12: 87–92
- Poorthuis BJ, Wevers RA, Kleijer WJ, Groener JE, de Jong JG, van Weely S, Niezen-Koning KE, van Diggelen OP (1999) The frequency of lysosomal storage diseases in The Netherlands. *Hum Genet* 105:151–156
- Zaremba J, Kleijer WJ, Juijms JG, Poorthuis B, Fidzianska E, Glogowska I (1992) Chromosomes 14 and 21 as possible candidates for mapping the gene for Sanfilippo disease type IIIC. *J Med Genet* 29:514
- Ausseil J, Loredó-Osti JC, Verner A, Darmond-Zwaig C, Maire I, Poorthuis B, van Diggelen OP, Hudson TJ, Fujiwara TM, Morgan K, Pshchetsky AV (2004) Localization of a gene for mucopolysaccharidosis IIIC to chromosome region 8p11-8q11. *J Med Genet* 41:941–945

16. Seyrantepe V, Tihy F, Pshezhetsky AV (2006) The microcell-mediated transfer of human chromosome 8 restores the deficient *N*-acetyltransferase activity in skin fibroblasts of mucopolysaccharidosis type IIIC patients. *Hum Genet* 120:293–296
17. Kong X, Murphy K, Raj T, He C, White PS, Matisse TC (2004) A combined linkage-physical map of the human genome. *Am J Hum Genet* 75:1143–1148
18. Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, Mai CP, Nguyen TH, Nguyen NB, Pham XK, Sarno EN, Alter A, Montpetit A, Moraes ME, Moraes JR, Dore C, Gallant CJ, Lepage P, Verner A, Van De Vosse E, Hudson TJ, Abel L, Schurr E (2004) Susceptibility to leprosy is associated with *PARK2* and *PACRG*. *Nature* 427:636–640
19. Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker sharing statistics. *Am J Hum Genet* 58:1323–1337
20. Sobel E, Papp JC, Lange K (2002) Detection and integration of genotyping errors in statistical genetics. *Am J Hum Genet* 70:496–508
21. Hodanova K, Majewski J, Kublova M, Vyletal P, Kalbacova M, Stiburkova B, Hulkova H, Chagnon YC, Lanouette CM, Marinaki A, Frys JP, Venkat-Raman G, Knoch S (2005) Mapping of a new candidate locus for uromodulin-associated kidney disease (UAKD) to chromosome 1q41. *Kidney Int* 68:1472–1482
22. O'Connell JR, Weeks DE (1998) PedCheck: a program for identifying genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259–266
23. Voznyi YV, Karpova EA, Dudukina TV, Tsvetkova IV, Boer AM, Janse HC, van Diggelen OP (1993). A fluorimetric enzyme assay for the diagnosis of Sanfilippo disease C (MPS III C). *J Inher Metab Dis* 16:465–472
24. Gudbjartsson DF, Jonasson K, Frigge M, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 25:12–13
25. Smyth GK (2005) Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds) *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer, New York, pp 397–420
26. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:article 3
27. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300
28. Knoch S, Hartmannova H, Stiburkova B, Krijt J, Zikanova M, Sebesta I (2000) Human adenylosuccinate lyase (ADSL), cloning and characterization of full-length cDNA and its isoform, gene structure and molecular basis for ADSL deficiency in six patients. *Hum Mol Genet* 9:1501–1513
29. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254
30. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, et al (2006) The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* 34:D590–D598
31. Ausseil J, Landry K, Seyrantepe V, Trudel S, Mazur A, Lapointe F, Pshezhetsky AV (2006) An acetylated 120-kDa lysosomal transmembrane protein is absent from mucopolysaccharidosis IIIC fibroblasts: a candidate molecule for MPS IIIC. *Mol Genet Metab* 87:22–31
32. Khsay RY, Gao G, Liao L (2005) An improved hidden Markov model for transmembrane protein detection and topology prediction and its applications to complete genomes. *Bioinformatics* 21:1853–1858
33. Jensen LJ, Gupta R, Blom N, Devos D, Tamames J, Kesmir C, Nielsen H, Staerfeldt HH, Rapacki K, Workman C, Andersen CA, Knudsen S, Krogh A, Valencia A, Brunak S (2002) Prediction of human protein function from post-translational modifications and localization features. *J Mol Biol* 319:1257–1265
34. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4:1633–1649
35. Bonifacino JS, Traub LM (2003) Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu Rev Biochem* 72:395–447
36. Bame KJ, Rome LH (1986a) Acetyl-coenzyme A: $\alpha$ -glucosaminide *N*-acetyltransferase: evidence for an active site histidine residue. *J Biol Chem* 261:10127–10132
37. Bame KJ, Rome LH (1986b) Genetic evidence for transmembrane acetylation by lysosomes. *Science* 233:1087–1089
38. Meikle PJ, Whittle AM, Hopwood JJ (1995) Human acetyl-coenzyme A: $\alpha$ -glucosaminide *N*-acetyltransferase: kinetic characterization and mechanistic interpretation. *Biochem J* 308:327–333

# Mutations in *DNAJC5*, Encoding Cysteine-String Protein Alpha, Cause Autosomal-Dominant Adult-Onset Neuronal Ceroid Lipofuscinosis

Lenka Nosková,<sup>1,2,9</sup> Viktor Stránecký,<sup>1,2,9</sup> Hana Hartmannová,<sup>1,2</sup> Anna Přistoupilová,<sup>1,2</sup> Veronika Barešová,<sup>1,2</sup> Robert Ivánek,<sup>1,2</sup> Helena Hůlková,<sup>1</sup> Helena Jahnová,<sup>1</sup> Julie van der Zee,<sup>3,4</sup> John F. Staropoli,<sup>5</sup> Katherine B. Sims,<sup>5</sup> Jaana Tyynelä,<sup>6</sup> Christine Van Broeckhoven,<sup>3,4</sup> Peter C.G. Nijssen,<sup>7</sup> Sara E. Mole,<sup>8</sup> Milan Elleder,<sup>1,2</sup> and Stanislav Kmoch<sup>1,2,\*</sup>

Autosomal-dominant adult-onset neuronal ceroid lipofuscinosis (ANCL) is characterized by accumulation of autofluorescent storage material in neural tissues and neurodegeneration and has an age of onset in the third decade of life or later. The genetic and molecular basis of the disease has remained unknown for many years. We carried out linkage mapping, gene-expression analysis, exome sequencing, and candidate-gene sequencing in affected individuals from 20 families and/or individuals with simplex cases; we identified in five individuals one of two disease-causing mutations, c.346\_348delCTC and c.344T>G, in *DNAJC5* encoding cysteine-string protein alpha (CSP $\alpha$ ). These mutations—causing a deletion, p.Leu116del, and an amino acid exchange, p.Leu115Arg, respectively—are located within the cysteine-string domain of the protein and affect both palmitoylation-dependent sorting and the amount of CSP $\alpha$  in neuronal cells. The resulting depletion of functional CSP $\alpha$  might cause in parallel the presynaptic dysfunction and the progressive neurodegeneration observed in affected individuals and lysosomal accumulation of misfolded and proteolysis-resistant proteins in the form of characteristic ceroid deposits in neurons. Our work represents an important step in the genetic dissection of a genetically heterogeneous group of ANCLs. It also confirms a neuroprotective role for CSP $\alpha$  in humans and demonstrates the need for detailed investigation of CSP $\alpha$  in the neuronal ceroid lipofuscinoses and other neurodegenerative diseases presenting with neuronal protein aggregation.

## Introduction

The neuronal ceroid lipofuscinoses (NCLs) are a heterogeneous group of inherited neurodegenerative disorders with an incidence of between 1 and 30 per 100,000. Common findings in the NCLs are an accumulation of autofluorescent storage material in neural and peripheral tissues and neurodegeneration. Although mutations in eight genes—*CLN1* (*PPT1* [MIM 256730]), *CLN2* (*TPP1* [MIM 204500]), *CLN3* (MIM 204200), *CLN5* (MIM 256731), *CLN6* (MIM 601780), *CLN7* (*MFSN8* [MIM 610951]), *CLN8* (MIM 600143), and *CLN10* (*CTSD* [MIM 610127])—have been identified in autosomal-recessive childhood and juvenile NCLs<sup>1</sup> and recently also in autosomal-recessive adult-onset NCL (Kufs disease [MIM 204300])<sup>2</sup>, the genetic and molecular basis of adult-onset NCL with dominant inheritance (Parry type [MIM 162350]) remains unknown.

Autosomal-dominant adult-onset neuronal ceroid lipofuscinosis (ANCL) was first described in a family of British descent from New Jersey, USA (Parry disease),<sup>3</sup> and in a second family reported in Spain.<sup>4</sup> More recently, a large

American family with English ancestry (UCL563 in this study),<sup>5</sup> another family from Alabama, USA (UCL562),<sup>6,7</sup> and a third family from the Netherlands (N1)<sup>8</sup> were presented. Common characteristics of affected individuals included generalized seizures, movement disorders, cognitive deterioration, and progressive dementia; the age of onset varied between 25 and 46 years.

In this work we describe a Czech family (P1) with autosomal-dominant ANCL in whom, by using a combination of linkage mapping, gene-expression analysis, and exome sequencing, we identified a unique heterozygous mutation in *DNAJC5* encoding cysteine-string protein alpha (CSP $\alpha$  [MIM 611203]; information on CSP $\alpha$  is accessible in the National Center for Biotechnology Information [NCBI] Gene Entrez database under GeneID 54968). The same or a second heterozygous *DNAJC5* mutation was found in four additional unrelated ANCL families that, together with altered palmitoylation-dependent sorting of mutant proteins in a cellular model and a reduced amount of CSP $\alpha$  in neuronal cells of affected individuals, confirmed the causality of CSP $\alpha$  mutations in autosomal-dominant ANCL.

<sup>1</sup>Institute for Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague, 120 00 Prague, Czech Republic; <sup>2</sup>Center for Applied Genomics, First Faculty of Medicine, Charles University in Prague, 120 00 Prague, Czech Republic; <sup>3</sup>Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, B-2610 Antwerp, Belgium; <sup>4</sup>Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, B-2610 Antwerp, Belgium; <sup>5</sup>Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA; <sup>6</sup>Institute of Biomedicine/Biochemistry and Developmental Biology, University of Helsinki, 00014 Helsinki, Finland; <sup>7</sup>Department of Neurology, St. Elisabeth Hospital, 5022 Tilburg, The Netherlands; <sup>8</sup>MRC Laboratory for Molecular Cell Biology, Institute of Child Health and Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK

<sup>9</sup>These authors contributed equally to this work

\*Correspondence: skmoch@lf1.cuni.cz

DOI 10.1016/j.ajhg.2011.07.003. ©2011 by The American Society of Human Genetics. All rights reserved.



## Material and Methods

### Subjects

The Czech family (P1) was ascertained at the Institute of Inherited Metabolic Disorders in Prague. Some families were described earlier—an American family from USA with English ancestry UCL563,<sup>5</sup> a family from Alabama, USA (UCL562),<sup>6,7</sup> and one from the Netherlands (N1)<sup>8</sup>. Previously unpublished data from families from the USA, France, the Netherlands, Belgium, Poland, Austria, Italy, and Germany were collected under the auspices of the Rare NCL Gene Consortium by Sara Mole. Enzyme assay or analysis of known genes in which mutations lead to NCL had excluded these mutations as the cause in some but not all subjects. Diagnosis of ANCL disease is very challenging, partly because of its rarity but also because for some cases it can only be verified by finding the characteristic pathology in the brain, and not all affected individuals undergo this procedure. The cases included here were diagnosed by clinicians in different countries over two decades. Because full documentation was not always accessible, some medical histories could not be reviewed. However, we chose to test as many likely cases as possible and to fully report negative findings. Investigations were approved by participating centers' institutional review boards and were conducted according to the Declaration of Helsinki principles. Written, informed consent was obtained from all subjects.

### Genotyping and Linkage Analysis

Genomic DNA was isolated by standard technology. We genotyped DNA samples by using Affymetrix GeneChip Mapping 10K 2.0 arrays (Affymetrix, Santa Clara, CA) according to the manufacturer's protocol at the microarray core facility of the Institute of Molecular Genetics in Prague. We extracted raw feature intensities from the Affymetrix GeneChip Scanner 3000 7G images by using the GeneChip operating Software (GCOS) 1.4 and generated individual SNP calls by using Affymetrix Genotyping Analysis Software (GTTYPE) 4.1.

We carried out multipoint parametric linkage analysis along with a determination of the most likely haplotypes by using affected-only analysis under the assumption of an autosomal-dominant mode of inheritance with a 0.99 constant, age-independent penetrance, 0.01 phenocopy rate, and 0.001 frequency of disease allele; the analysis was performed with version 1.1.2 of Merlin software.<sup>9</sup> The results were visualized in the version 1.032 of the HaploPainter software<sup>10</sup> and in version 2.9.2 of R-project statistical software.

### Gene-Expression Analysis

We isolated leucocytes from freshly drawn blood by using a standard erythrocyte lysis protocol and isolated total RNA from freshly isolated cells by using TRIZOL solution (Invitrogen, Carlsbad, CA). RNA concentration was determined spectrophotometrically at A260 nm by NanoDrop (NanoDrop Technologies), and quality was checked on an Agilent 2100 Bioanalyser (Agilent Technologies). Aliquots of isolated RNA were stored at  $-80^{\circ}\text{C}$  until analysis. Expression analysis was performed on the Illumina HumanRef-8\_V2 BeadChip at the microarray core facility of the Institute of Molecular Genetics in Prague. Hybridized slides were scanned on an Illumina BeadArray Reader, and bead level data were summarized by Illumina BeadStudio Software v3. Bead summary data were imported into R-project statistical software v.2.9.2 and normalized with the quantile method in the Lumi package. Differ-

ential gene-expression analysis was performed with the Limma package and the lmFit function. A multiple testing correction was performed with the Benjamini and Hochberg method. Database for Annotation, Visualization and Integrated Discovery version 6.7 (DAVID) was used for functional annotation. Details on the experiment and raw expression data are available at the Gene Expression Omnibus (GEO) repository under accession GSE30369.

### Copy-Number Analysis

DNA samples from seven individuals of family P1 (II.2, IV.1, IV.2, IV.3, IV.4, IV.7, and IV.8) were genotyped with Affymetrix GeneChip Mapping 6.0 array (Affymetrix, Santa Clara, CA) at the microarray core facility of the Institute of Molecular Genetics in Prague according to the manufacturer's protocol. Raw feature intensities were extracted from the Affymetrix GeneChip Scanner 3000 7G images with the GeneChip Control Console Software 2.01. We generated individual SNP calls by using Affymetrix Genotyping Console Software 3.02. Copy-number changes were identified in Affymetrix Genotyping Console Software (GTC version 3.02). We used data from both SNP and copy-number probes to identify copy-number aberrations relative to a built-in reference. Only regions larger than 10 Kb and containing at least five probes were reported.

### Exome Sequencing

We performed DNA enrichment by using 3  $\mu\text{g}$  of DNA from individual IV.7 and the SureSelect All Exome kit (Agilent, Santa Clara, USA) according to the manufacturer's protocol. DNA sequencing was performed on the captured DNA library with one-quarter of a SOLiD 4 slide (Applied Biosystems, Carlsbad, USA) at CeGaT (Tubingen, Germany). We aligned reads in color space to the reference genome (hg19) by using NovoalignCS version 1.01 (Novocraft, Malaysia) with the default parameters. Sequence variants in the analyzed sample were identified with the SAMtools package (version 0.1.8).<sup>11</sup> The high-confidence variants list (SNP quality > 100 and indel quality > 50) was annotated with the SeattleSeq Annotation server (hg19). Sequence variants that were not annotated in the dbSNP or 1000 Genomes databases were prioritized for further analysis.

### DNA Sequencing and Mutation Analysis

All exons and corresponding exon-intron boundaries of *DNAJC5* (NM\_025219.2), encoding CSP $\alpha$ , were amplified by PCR from genomic DNA of the probands and sequenced with version 3.1 Dye Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) with electrophoresis on an ABI 3500XL Avant Genetic Analyzer (Applied Biosystems). Data were analyzed with Sequencing Analysis software. Segregation of the candidate mutations was assessed by PCR and direct sequencing of the corresponding genomic DNA fragments. Primer sequences are available in Table S1, available online.

### Homozygosity-Haplotype Analysis

*DNAJC5* genomic fragments containing multiple SNPs with high-heterozygosity values were amplified by PCR from genomic DNA of probands and sequenced as described above. Genotypes for individual SNPs were obtained, and homozygous haplotypes were defined as described recently.<sup>12</sup> We compared the resulting homozygous haplotypes across individuals to determine whether

the chromosomal segments around the same identified mutations could be identical by descent.

### Bioinformatic Analysis of the Cysteine-String Domain

Hydrophobicity of the wild-type and mutant cysteine-string domains were analyzed with a Kyte-Doolittle algorithm available at ExPasy server. Potential effects of detected mutations on CSP $\alpha$  palmitoylation were assessed with the prediction program CSS-Palm 2.0. Obtained hydrophobicity values and palmitoylation score values were exported for each of the sequences and plotted with an Excel function. We assessed possible impacts of the p.Leu115Arg substitution on the structure and function of CSP $\alpha$  by using SIFT and PolyPhen-2 servers.

### CSP $\alpha$ -Expression Vectors

*DNAJC5/CSP $\alpha$*  cDNA were amplified by RT-PCR from a control and an affected individuals' leucocytes with primers incorporating a *BspEI* site at the 5' end of PCR products. Resulting PCR products were first cloned into pCR4 TOPO vector (Invitrogen) and, after sequencing verification, these were further subcloned in frame into a pEGFP-C1 vector with *BspEI* and *Apal* restriction sites. The initiating methionine codon was removed from *DNAJC5/CSP $\alpha$*  in all enhanced green fluorescent protein (EGFP)-CSP $\alpha$  constructs.

### Transient Expression of EGFP-CSP $\alpha$

pEGFP-CSP $\alpha$  constructs were transfected into CAD-2A2D5 (CAD5) cells derived from Cath.a-differentiated (CAD) cells (provided by Sukhvir Mahal, The Scripps Research Institute, Jupiter, FL, USA). One day before transfection,  $8 \times 10^4$  cells/cm<sup>2</sup> were seeded with OptiMEM medium (OptiMEM; Invitrogen) containing 9% BGS (HyClone, Logan, UT), 90 units penicillin/ml, and 90 g of streptomycin/ml. Cells were transfected by either 0.8  $\mu$ g or 4.5  $\mu$ g of plasmid constructs with Lipofectamine 2000 (Invitrogen) in serum and antibiotics free OptiMEM medium according to the manufacturer's protocol. Transfection experiments were performed in more than five replicates.

### Immunofluorescence Analysis

Cells were fixed 24 hr after transfection with 4% paraformaldehyde, permeabilized in 0.1% TRITON, washed, blocked with 5% bovine serum albumin (BSA), and incubated for 1 hr at 37°C with anti-protein disulfide isomerase (PDI) mouse monoclonal IgG1 (Stressgen, San Diego, CA) for endoplasmic reticulum (ER) localization, anti-GS28 mouse IgG1 (Stressgen, San Diego, CA) for Golgi localization, and anti-GFP rabbit polyclonal IgG (Abcam) for EGFP-CSP $\alpha$  detection. For fluorescence detection, corresponding species-specific secondary antibodies Alexa Fluor 488 and Alexa Fluor 555 (Molecular Probes, Invitrogen, Paisley, UK) were used. Nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI). Prepared slides were mounted in fluorescence mounting medium Immu-Mount (Shandon Lipshaw, Pittsburgh, PA) and analyzed by confocal microscopy.

### Image Acquisition and Analysis

XYZ images sampled according to Nyquist criterion were acquired with a TE2000E C1si laser scanning confocal microscope, a Nikon PlanApo objective (40 $\times$ , N.A.1.30), 488 nm and 543 nm laser lines, and 515  $\pm$  15 nm and 590  $\pm$  15 nm band-pass filters. Images were deconvolved with the classic maximum likelihood restoration algorithm in Huygens Professional software (SVI, Hilversum,

The Netherlands).<sup>13</sup> Colocalization maps employing single pixel overlap coefficient values ranging from 0–1<sup>14</sup> were created with Huygens Professional software. The resulting overlap coefficient values are presented as pseudocolor (the scale is shown in the corresponding lookup tables).

### Immunoblot Analysis

#### *Transfected CAD5 Cells*

Cells were harvested in PBS; centrifuged at 500 g for 7 min; and resuspended in 10 mM Tris, 10 mM KCl, 2 mM EDTA, 4% glycerol, 1 mM DTT, and Complete Protease Inhibitor Cocktail (Roche); homogenized by sonication followed by centrifugation at 20,000 g for 15 min at 4°C; and assessed for protein content in the supernatant with the Bradford assay.

#### *Brain Homogenates*

Frozen autopsy materials were homogenized under liquid nitrogen; dissolved in 10 mM Tris, 10 mM KCl, 2 mM EDTA, 4% glycerol, 1 mM DTT, and Complete Protease Inhibitor Cocktail (Roche); centrifuged at 20,000 g for 15 min at 4°C; and assessed for protein content in the supernatant with the Bradford assay. Homogenate aliquots corresponding to 30  $\mu$ g of total protein in brain homogenates or 20  $\mu$ g of total protein in CAD5 cells were resolved on 12% SDS-PAGE under nonreducing or reducing conditions and transferred to the polyvinylidene fluoride (PVDF) membrane. Membranes were blocked by 5% BSA and 0.05% Tween 20 in PBS. CSP $\alpha$  or CSP $\alpha$ -EGFP protein was visualized by incubation with rabbit CSP antibody (Stressgen) at 1: 500 in 5% BSA and 0.05% Tween 20 in PBS for 90 min or rabbit GFP antibody (Abcam) at 1:5000 in 5% BSA and 0.05% Tween 20 in PBS for 90 min, followed by incubation with goat anti-rabbit HRP (Pierce) at 1:10000 in 0.05% Tween 20 in PBS for 60 min and detection by SuperSignal West Femto Maximum Sensitivity Substrate (Pierce). For depalmitoylation studies, samples were depalmitoylated prior to SDS-PAGE by treatment with neutral 1 M hydroxylamine or 1 M Tris as a control for 20 hr at room temperature.

### Immunohistochemical and Histochemical Studies

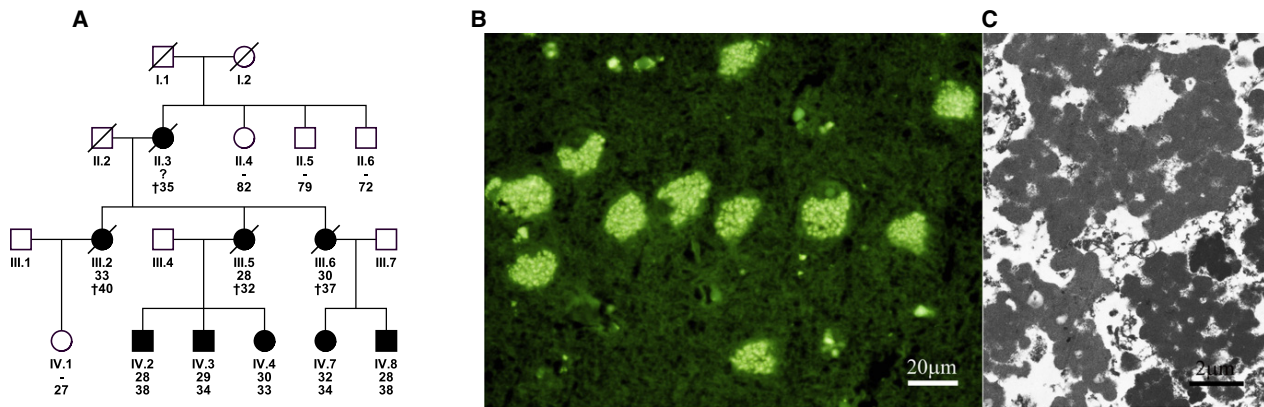
Formaldehyde-fixed brain samples were analyzed. Immunodetection of CSP $\alpha$  on paraffin sections was performed with rabbit CSP antibody (Stressgen; diluted 1:750 in 5% BSA) in PBS. Synaptic regions were detected with monoclonal mouse IgG1 synaptobrevin antibody (Sigma, Saint Louis, USA; diluted 1:8000 in 5% BSA) in PBS, which was applied after heat-induced epitope retrieval at pH 6.0. Detection of the bound primary antibody was achieved with Dako EnVision + TM Peroxidase Rabbit kit (Dako, Glostrup, Denmark) with 3,3'-diaminobenzidine as substrate. The specificity of the antigen detection was always ascertained by omitting of the primary antibody-binding step.

Stored ceroid material was best detected because of its prominent autofluorescence via filter block with an excitation wavelength of 400–440 nm (fluorescence microscope Nikon E800, filter block BV-2A).

## Results

### Clinical Observations and Biochemical Findings

The diagnosis of ANCL in family P1 (Figure 1A) was based on clinical presentation and examination of proband III.6, who presented at age 30 with myoclonic epilepsy, generalized tonic-clonic seizures, and progressive cognitive



**Figure 1. Pedigree and Neuropathology Findings in Family P1**

(A) Pedigree of the Czech family. Black symbols denote affected individuals; open symbols denote unaffected individuals. Age of onset is shown above current age or age of death (indicated by †). (B) Epifluorescence. Hippocampal pyramidal neurons with prominent lysosomal storage of autofluorescent material representing the general neurolysosomal storage pattern in the brain cortex. The autofluorescence was demonstrated with the filter block with an excitation wavelength of 400–440 nm. (C) Electron micrograph. GROD-type ultrastructure of the storage lysosomes.

deterioration with depression; these symptoms were followed by progressive motor neurological symptoms leading to death at age 37 years. There was normal activity of palmitoyl-protein-thioesterase 1 (PPT1) in leucocytes. Neuropathological examination of postmortem brain tissue showed characteristic neurolysosomal storage of autofluorescent material with ultrastructural appearance corresponding to granular osmiophilic deposits (GRODs) (Figures 1B and 1C). A skin biopsy was free of lysosomal storage at the ultrastructural level. An affected status in other family members was assigned if a very similar clinical course starting with myoclonic and/or generalized tonic-clonic seizures followed after 1–2 years by progressive cognitive deterioration and depressive symptomatology. All affected individuals showed generalized epileptic discharges in electroencephalograms and manifested brainstem and central pyramidal neurological symptomatology in the later period of disease. Other ANCL families analyzed in this study are described in Table 1. Previously unpublished families and cases with mutation in CSP $\alpha$  are described in more detail below.

The proband of family UCL328 was a male of European descent and in good health until his first generalized tonic-clonic seizure at age 34. This was followed by evidence of progressive confusion and dementia as well as more frequent, medically refractory generalized seizures. Long-term electroencephalography showed generalized periodic epileptiform discharges superimposed on a background of diffuse low-amplitude, high-frequency activity consistent with a dementing process. A brain MRI at age 38 showed prominence of cortical sulci and cerebellar folds and mild enlargement of the lateral ventricles consistent with diffuse cerebral and cerebellar atrophy. Concurrent neuropsychiatric testing showed a verbal IQ of 77, a performance IQ of 71, and a full-scale IQ of 73. Regression of gross and fine motor skills began at age 40, and there was ensuing

evidence of ataxia and myoclonus. By age 45, the proband was wheelchair-bound and required nursing-home care. Visual function was normal. A frontal lobe brain biopsy revealed numerous neurons containing homogeneous eosinophilic material with a golden-brown hue. The pigmented material stained intensely by the periodic acid-Schiff reaction and was found to be autofluorescent. Ultrastructural examination showed multiple neurons distended by granular osmiophilic deposits. There was no family history of seizures, early-onset dementia, or other neurologic abnormality.

The proband of family UCL519 is one of at least five similarly affected individuals over three generations with apparent autosomal-dominant inheritance. He showed obsessive behavior starting in his mid-20s, and the first seizure occurred when he was in his early 30s. His speech regressed, his short-term memory became impaired, and he had difficulty in walking without an aid. No further details are available.

#### Identification of CSP $\alpha$ Mutation in Family P1 by a Combination of Linkage Analysis, Copy-Number Analysis, Gene-Expression Analysis, and Exome Sequencing

To map the disease locus, we used Affymetrix GeneChip Mapping 10K v2.0 arrays, genotyped all available and informative family members, and performed linkage analysis. We identified five candidate regions with positive LOD scores on chromosomes 1, 4, 15, 20, and 22 (Figure 2A). In parallel, we used Affymetrix GeneChip Mapping 6.0 array, genotyped seven individuals, and assessed copy-number changes; we found no indication for a potentially disease-causing deletion or duplication.

To identify a mutation that affected the amount of transcript, we compared gene-expression profiles in leucocytes isolated from four affected individuals to those from four

**Table 1. ANCL Families Analyzed in This Study**

Family No.	Mutation in CSP $\alpha$	Country	Diagnosis	References
P1	p.Leu116del	Czech Republic	ANCL, autosomal dominant	
N1	p.Leu115Arg	The Netherlands	ANCL, autosomal dominant	8,31,32
UCL563	p.Leu115Arg	USA	ANCL, autosomal dominant	5
UCL328	p.Leu115Arg	USA, French-Canadian	Kufs	
UCL519	p.Leu116del	USA	Kufs, autosomal dominant	
UCL417	–	France	Kufs, autosomal dominant	
UCL562	–	USA	Kufs, autosomal dominant	6,7
UCL572	–	USA/Italy	Kufs, autosomal dominant?	
UCL327	–	USA	Kufs, with ALS in extended family	
UCL385	–	Belgium	Kufs Type A or atypical juvenile NCL, autosomal recessive	
UCL403	–	France	Kufs Type B, autosomal recessive	
UCL450	–	Poland	variant juvenile or ANCL, autosomal recessive (heterozygous change in <i>CLCN6</i> already known)	33
UCL472	–	Germany	variant juvenile or ANCL	34
UCL482	–	The Netherlands	ANCL	
UCL508	–	USA	Kufs	
UCL520	–	USA	Kufs	
UCL522	–	USA	Kufs	
UCL545	–	Netherlands	Kufs	
UCL568	–	Austria	Kufs	
UCL571	–	Netherlands	Kufs	

Diagnosis is provided as reported by referring clinician. In all cases there was no visual failure, and no distinction was made according the mode of inheritance, if apparent.

age-matched controls by using Illumina HumanRef-8v2 Expression BeadChips. This analysis identified a set of 2131 differentially expressed genes, of which 65 were localized within candidate regions identified by linkage analysis (Figure 2B and Table S2). At the same time, we analyzed gene-expression changes by using gene-enrichment analysis and found that the identified profiles indicated significant dysregulation of spliceosome, upregulation of many components of respiratory chain complexes, altered expression of genes active in pathways involved in neurodegenerative diseases, and accelerated proteolysis (Table 2 and Figures S1–S7).

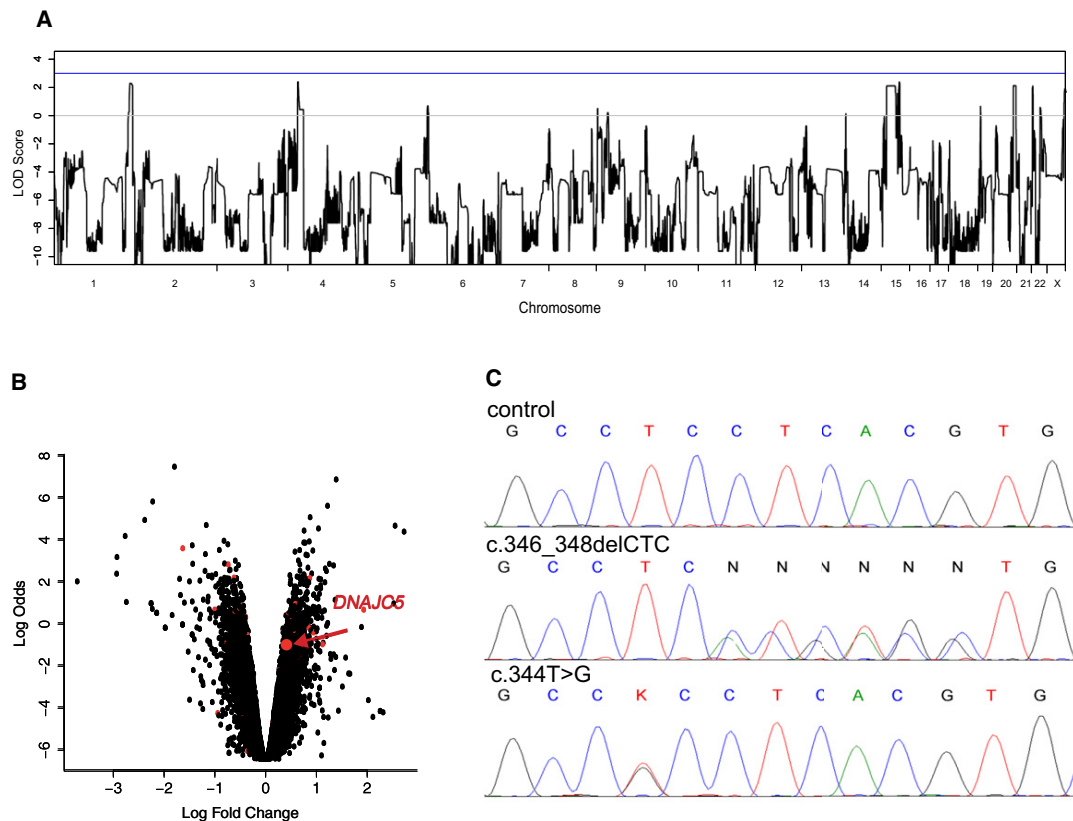
To directly identify possible disease-causing mutation(s) among the candidate genes defined by this combination of linkage analysis and gene-expression profiling, we performed exome sequencing in individual IV.7. From the sequencing run we obtained 94.7 M sequencing reads, of which we were able to map 50.2 M on the human genome reference sequence. After removing PCR generated duplicate reads (23.6 M), we obtained 26.6 M unique reads, of which 19.5 M (73.3%) mapped on a targeted exome sequence and were 92% covered at least once. When the sequence of the proband was compared to the reference sequence, 22,617 single nucleotide variants (SNP

quality > 100) and 2604 indels (indel quality > 50) were revealed in the proband, of which 957 (617 SNPs and 340 indels) were novel (e.g., were not present in the dbSNP and 1000 Genomes databases).

We intersected the results of exome sequencing with the mapping information and the gene-expression changes, and this analysis illuminated a single gene, *DNAJC5*, encoding the protein CSP $\alpha$ , located in the candidate region on chromosomal region 20q13.33, (*DNAJC5* hg19 coordinates chr20:62526518–62565394) and showing a significant increase in transcript levels in affected individuals' leucocytes (Figure 2B), and had a unique heterozygous mutation c.346\_348delCTC (p.Leu116del) compatible with autosomal-dominant inheritance of the disease (Table 3).

#### CSP $\alpha$ Mutations Segregate with ANCL in Additional Families

Through sequence analysis of *DNAJC5* genomic DNA, we found consistent segregation of the c.346\_348delCTC mutation with the ANCL phenotype within the Czech family P1 (Figure 2C). Moreover, among 20 additional ANCL families and/or simplex cases tested (Table 1), we identified the same mutation in a previously unreported



**Figure 2. Identification of *DNAJC5* Mutations**

(A) A whole-genome parametric linkage analysis showing candidate regions reaching the theoretical maximum LOD scores of 2.1 attainable in this family on chromosomes 1 (1: 233,697,529–249,250,621), 4 (4: 23,561,661–28,920,119), 15 (15: 39,049,915–61,382,423; 65,139,935–67,296,086; 71,515,415–78,819,152), 20 (20:53,448,624–63,025,520), and 22 (22: 1–21,982,248). All coordinates refer to hg19.

(B) Gene-expression changes in leucocytes from four affected individuals compared to those of four controls. The logarithm of the probability that the gene is differentially expressed (log odds) is plotted as a function of the logarithm of the gene-expression fold change (log fold change) between the patient and control samples. Differentially expressed genes located in the candidate regions are shown as red dots, and *DNAJC5* is specifically indicated. The list of differentially expressed genes located within the linked regions is, together with log fold changes and corresponding t test values, p-values and adjusted p-values, provided in Supplemental Data.

(C) Chromatograms of *DNAJC5* genomic DNA sequences showing identified heterozygous mutations. (Upper panel) Sequence of an unaffected individual, (middle panel) sequence showing heterozygous mutation c.346\_348delCTC in the proband from family P1, and (lower panel) sequence showing heterozygous mutation c.344T>G in the proband from family N1.

American family, UCL519, and a second heterozygous mutation (c.344T>G [p.Leu115Arg]) (Figure 1D) segregating with the phenotype in the Dutch family N1<sup>8</sup> and the American family UCL563<sup>5</sup> and present in a previously unreported simplex case UCL328. Mutations were found in all 14 affected individuals (five Czech, six Dutch, and one in each of the other pedigrees) across these five families and were absent in all seven unaffected siblings (two Czech, six Dutch, and one from American family UCL563) from whom DNA was available for testing. In addition to this, the identified mutations were absent in 200 control samples of European descent and were not present in the dbSNP or 1000 Genomes databases.

Haplotypes segregating with ANCL phenotype in Czech family P1 and Dutch family N1 were obtained from genotypes generated with Affymetrix GeneChip Mapping 10K

v2.0 arrays and are shown in Figures S8 and S9. For simplex cases, phased haplotypes could not be obtained. To reveal whether probands carrying the same mutation might be distantly related and share a mutation-carrying chromosomal segment from a common ancestor, we examined homozygosity haplotypes across the *DNAJC5* genomic region (Table S3). The c.346\_348delCTC (p.Leu116del) mutations in families P1 and UCL519 are present on two distinct haplotypes, indicating that these families are probably not related and that the mutations appeared independently. The mutations c.344T>G (p.Leu115Arg) are also present on two distinct haplotypes, one in UCL328 and one shared by family N1 and UCL563. This mutation therefore probably also appeared independently in two different lineages, but it is possible that families N1 and UCL563 are identical by descent.

**Table 2. Functional Annotation of Gene-Expression Changes and KEGG Pathways Defined by Gene-Enrichment Analysis**

Term	Count	%	p Value	Population Hits	Population Total	Fold Enrichment	FDR
hsa03040: spliceosome	41	2.48	$2.3 \times 10^{-10}$	126	5085	2.92	$2.9 \times 10^{-7}$
hsa05016: Huntington disease	46	2.78	$7.7 \times 10^{-8}$	180	5085	2.29	$9.5 \times 10^{-5}$
hsa05010: Alzheimer disease	43	2.60	$8.5 \times 10^{-8}$	163	5085	2.37	$1.1 \times 10^{-4}$
hsa05012: Parkinson disease	35	2.11	$7.0 \times 10^{-7}$	128	5085	2.45	$8.7 \times 10^{-4}$
hsa00190: oxidative phosphorylation	35	2.11	$1.0 \times 10^{-6}$	130	5085	2.41	$1.3 \times 10^{-3}$
hsa00520: amino sugar and nucleotide sugar metabolism	13	0.79	$2.4 \times 10^{-3}$	44	5085	2.65	$2.9 \times 10^0$
hsa03050: proteasome	12	0.73	$1.2 \times 10^{-2}$	47	5085	2.29	$1.4 \times 10^1$
hsa04120: ubiquitin mediated proteolysis	25	1.51	$1.5 \times 10^{-2}$	137	5085	1.64	$1.7 \times 10^1$
hsa04662: B cell receptor signaling pathway	16	0.97	$1.7 \times 10^{-2}$	75	5085	1.91	$1.9 \times 10^1$
hsa04621: NOD-like receptor signaling pathway	14	0.85	$1.7 \times 10^{-2}$	62	5085	2.03	$1.9 \times 10^1$
hsa00052: galactose metabolism	8	0.48	$2.0 \times 10^{-2}$	26	5085	2.76	$2.2 \times 10^1$
hsa03010: ribosome	17	1.03	$2.9 \times 10^{-2}$	87	5085	1.75	$3.0 \times 10^1$

### Identified Mutations Affect Palmitoylation-Dependent Sorting and the Amount of CSP $\alpha$ in Neuronal Cells

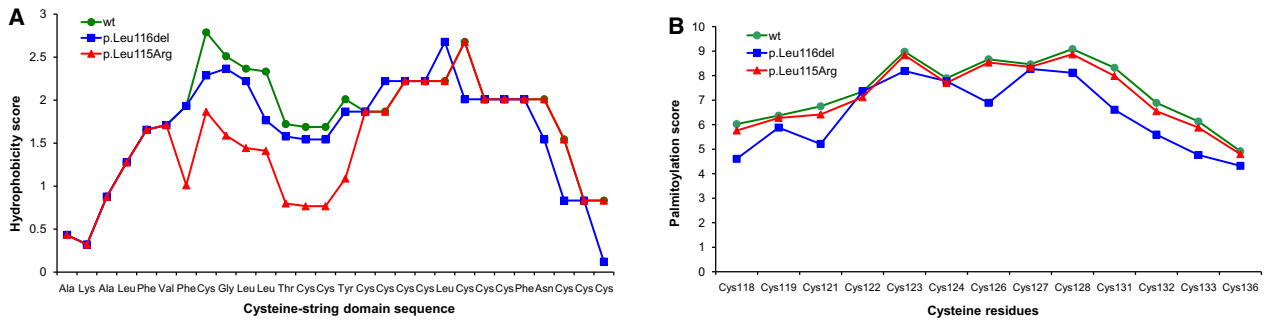
Both identified mutations affect conserved dileucine residues located in the cysteine-string domain implicated in palmitoylation and membrane trafficking of CSP $\alpha$ <sup>16</sup>. Using

in silico analysis, we found that p.Leu115Arg is predicted to decrease the hydrophobicity of the cysteine-string domain that is needed for initial binding of CSP $\alpha$  to the endoplasmic reticulum (ER) (Figure 3A), whereas p.Leu116del probably affects the efficiency of palmitoylation of adjacent cysteine residues (Figure 3B). SIFT analysis

**Table 3. Exome Sequencing and a List of High-Confidence Novel Coding Variants Revealed by Exome Sequencing**

Chromosome	Position	Reference Base	Sample Alleles	Function Genome Variation Server	Amino Acids	Protein Position	Gene List
<b>Single nucleotide variants</b>							
1	235,715,488	C	C/T	missense	ARG.GLN	50/76	<i>GNG4</i>
1	236,987,512	C	C/T	synonymous	none	286/1266	<i>MTR</i>
1	247,835,885	G	C/G	synonymous	none	153/308	<i>OR13G1</i>
15	43,552,700	G	G/T	missense	HIS.ASN	30/721	<i>TGM5</i>
15	43,900,153	C	C/T	synonymous	none	1234/1776	<i>STRC</i>
15	45,028,847	G	G/T	utr-5	none	NA	<i>TRIM69</i>
15	59,500,166	A	A/G	missense	ILE.VAL	343/382	<i>MYO1E</i>
15	65,555,518	A	A/G	synonymous	none	220/324	<i>PARP16</i>
15	66,857,721	C	C/T	utr-5	none	NA	<i>LCTL</i>
15	75,116,809	G	A/G	missense	VAL.MET	481/527	<i>LMAN1L</i>
20	60,884,827	G	A/G	synonymous	none	3631/3696	<i>LAM5</i>
22	20,097,643	C	C/T	utr-3	none	NA	<i>DGCR8</i>
22	21,138,487	C	C/T	synonymous	none	373/500	<i>SERPIND1</i>
<b>Indels</b>							
4	25,678,161	TGC	-TGC	coding	none	NA	<i>SLC34A2</i>
20	62,562,227	CTC	-CTC	coding	none	NA	<i>DNAJC5</i>

All coordinates refer to hg19. SNP quality > 100 and indel quality > 50. Only Variants located within the linkage candidate regions and not present in dbSNP or 1000 Genomes databases are shown.



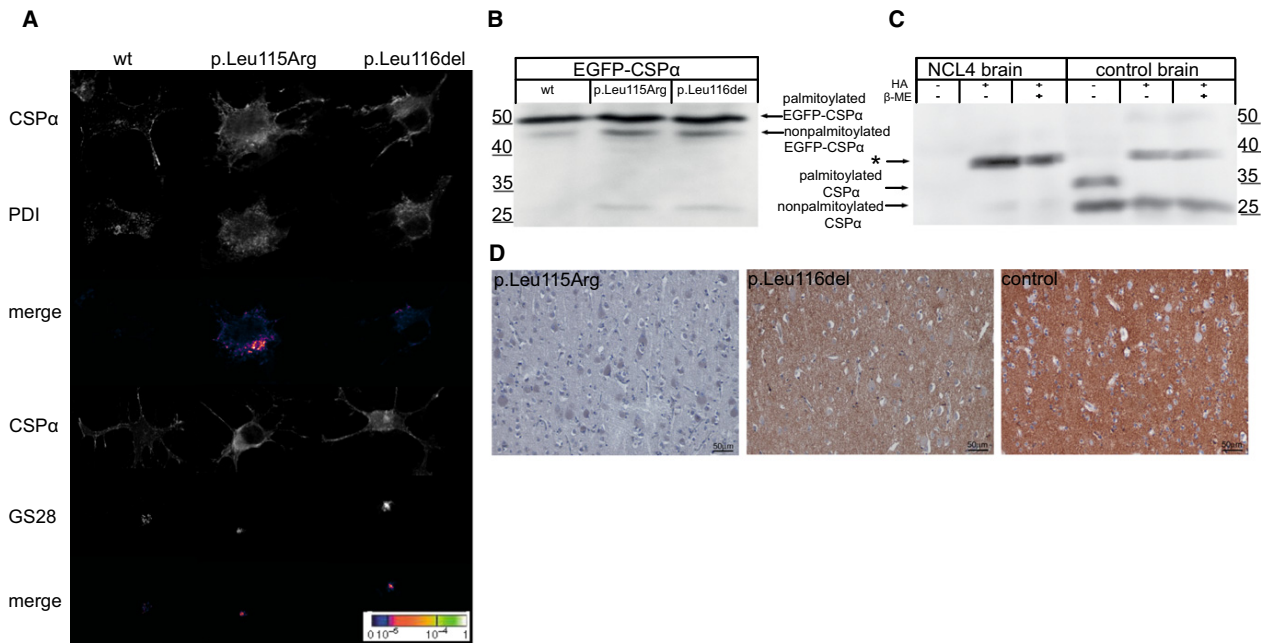
**Figure 3. In Silico Analysis of Properties of the Cysteine-String Domain**

(A) p.Leu115Arg mutation decreases the hydrophobicity of this domain, which is needed for initial binding to the ER. (B) The p.Leu116del mutation decreases the palmitoylation score, that is, the confidence that cysteine residues adjacent to Leu116 might be efficiently palmitoylated.

(score = 0.00) predicted that the p.Leu115Arg mutation affects protein function, and analysis with PolyPhen (overall score = 0.782; sensitivity = 0.85; and specificity = 0.93) predicted that it is possibly damaging. No such predictions can be obtained for the identified deletion p.Leu116del.

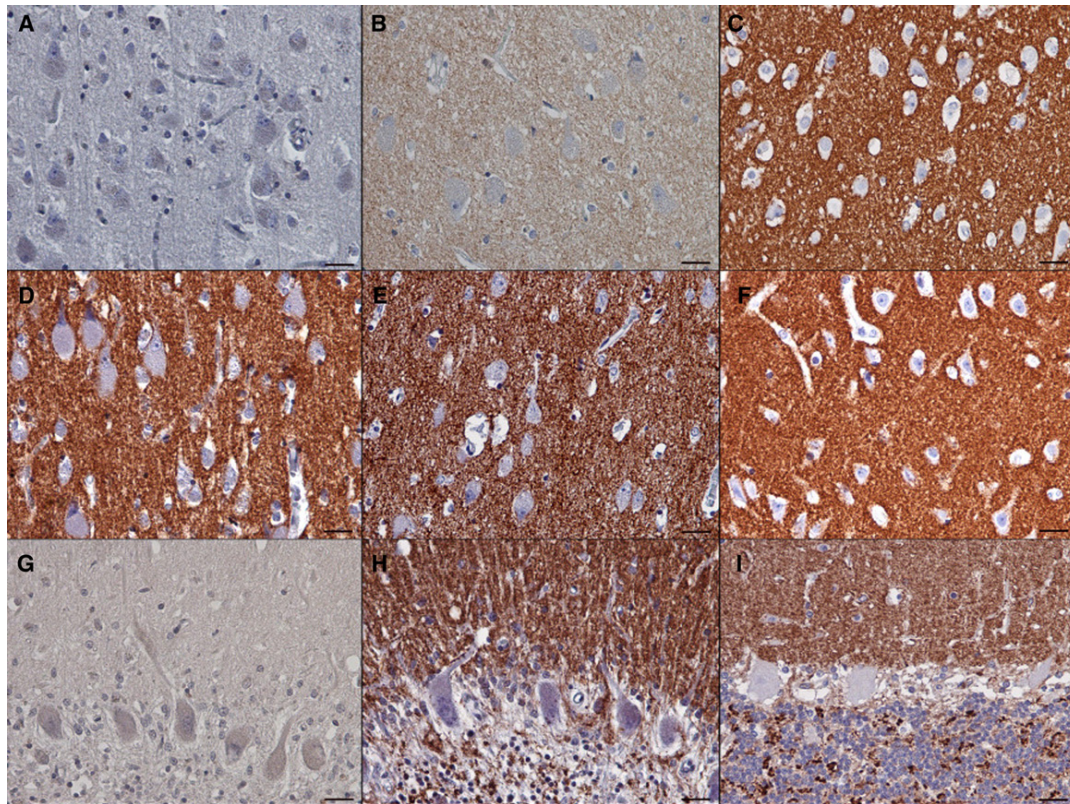
To study an effect of the identified mutations, we transiently expressed wild-type EGFP-tagged CSP $\alpha$  or mutant protein containing either p.Leu115Arg or p.Leu116del in

CAD5 neuronal cells. Using immunofluorescence analysis, we found wild-type EGFP-CSP $\alpha$  predominantly at the plasma membrane, whereas both mutated proteins showed diffuse intracellular staining and abnormal colocalization with markers for the ER and Golgi apparatus (Figure 4A). In addition, using immunoblot analysis of transfected cell lysates, we found that both mutated proteins were less efficiently palmitoylated than the wild-type protein (Figure 4B).



**Figure 4. Characterization of Mutated CSP $\alpha$**

(A) Immunofluorescence analysis of transiently expressed EGFP-CSP $\alpha$  proteins in CAD5 cells showing prominent membrane localization of wild-type CSP $\alpha$  compared to the diffuse cytoplasmic staining and marked colocalization of mutated CSP $\alpha$  with endoplasmic reticulum represented by PDI and Golgi apparatus represented by Golgi-SNARE of 28 kDa (GS28). (B) Immunoblot analysis of transiently expressed EGFP-CSP $\alpha$  proteins showing higher levels of nonpalmitoylated protein precursors for mutant proteins compared the wild-type (wt) protein. (C) Immunoblot analysis of brain homogenates showing no soluble CSP $\alpha$  and the marked presence of CSP $\alpha$ -containing beta-mercaptoethanol ( $\beta$ -ME)-resistant aggregate (indicated by the asterisk) released upon hydroxylamine (HA) treatment in the affected individual (NCL4) compared to the brain homogenates of the control. (D) Immunohistochemistry analysis of CSP $\alpha$  in gray matter of the cerebral cortex showing, at a low field, a significant decrease of CSP $\alpha$  in affected individuals compared to the strong CSP $\alpha$  staining in the age-matched control.



### Figure 5. Brain Immunohistochemistry

(A–C) Detail of the CSP $\alpha$  staining in neuropil in the cerebral cortex that is absent in the individual with mutation p.Leu115Arg (A), decreased in the individual with mutation p.Leu116del (B), and strong in the age-matched control (C). Note the prominent neuronal storage, shown by large cell bodies, in both affected individuals.

(D–F) Staining pattern of the synaptic marker synaptobrevin in neuropil in the same regions in the individual with mutation p.Leu115Arg (D), in the individual with mutation p.Leu116del (E), and in the age-matched control (F).

(G and H) Cerebellar cortex of the case with p.Leu115Arg mutation. Similar to that in the cerebral cortex, CSP $\alpha$  staining is absent (G). This contrasts with a strong signal for synaptobrevin in the corresponding area in all three cerebellar cortical layers that are preserved adjacent to areas undergoing neurodegeneration (H).

(I) Strong CSP $\alpha$  staining in a control cerebellum. Note that the CSP $\alpha$  signal in the control (I) as well as the synaptobrevin signal in the individual with mutation p.Leu115Arg (H) are confined to the well defined synaptic regions (i.e., to the dendrites in the molecular layer, to the surface of the Purkinje cells, and to the synaptic glomeruli in the granular cell layer). The scale bars represent 25  $\mu$ m.

To correlate these observed effects with *in vivo*, we analyzed post-mortem brain specimens by immunoblotting. Although both palmitoylated and nonpalmitoylated CSP $\alpha$  were present in control brain lysates, we could not detect any CSP $\alpha$  in brain lysates from a Dutch case (family N1) with the p.Leu115Arg mutation. However, after chemical depalmitoylation, we detected a chemiluminescence signal, probably corresponding to an otherwise insoluble CSP $\alpha$ -containing aggregate, which appeared much stronger in brain lysate from the case, than in the control (Figure 4C). Using immunohistochemical staining of CSP $\alpha$  in paraffin-embedded brain sections, we consistently found an absence of CSP $\alpha$  staining in synaptic regions in both the cerebral and the cerebellar cortex of individuals with the p.Leu115Arg mutation and significantly reduced CSP $\alpha$  staining in the cerebral cortex of individuals with the p.Leu116del mutation when we compared these individuals to age-matched controls (Figures 4D and 5).

### Discussion

We carried out linkage mapping, gene-expression analysis, exome sequencing, and candidate-gene sequencing in affected individuals from 20 families and/or simplex cases of European descent suffering from autosomal-dominant adult-onset neuronal ceroid lipofuscinosis previously referred to as Parry disease. Using this approach, we identified in five of these families two recurrent mutations, c.346\_348delCTC (p.Leu116del) and c.344T>G (p.Leu115Arg), in *DNAJC5* encoding cysteine-string protein alpha (CSP $\alpha$ ). To prove their causality, we performed haplotype analysis, which revealed that the mutations had to appear independently in at least four lineages, and by using targeted genotyping of seven unaffected siblings and 200 control individuals as well as searching the 1000 Genome and dbSNP databases, we found that the mutations are exclusively present in 14 affected individuals.



CSP $\alpha$  is a highly conserved protein with no amino acid sequence variant found in humans so far. The identified mutations affect evolutionary conserved dileucine residues located in the cysteine-string domain that is implicated in palmitoylation and membrane targeting of CSP $\alpha$ .<sup>15–17</sup> Functional studies in transfected cell lines proved that these mutations affect palmitoylation and intracellular location of CSP $\alpha$  and thus decrease the level of the CSP $\alpha$  protein in the brain of affected individuals.

The molecular mechanisms underlying the dominant negative effect of the identified mutations on CSP $\alpha$  amounts in neuronal cells are not clear. It is known that CSP $\alpha$  forms detergent-resistant dimers<sup>18</sup> and that the presence of these dimers correlates with an inhibition of synapse formation and synaptic transmission.<sup>19</sup> Immunoblot analysis of brain lysate from affected individuals showed CSP $\alpha$  to be exclusively present in such an aggregate form. It is probable that the presence of mutant protein catalyzes accelerated aggregation and that the resulting aggregates will be composed equally of both mutant and wild-type proteins, and this will result in CSP $\alpha$  depletion. Another explanation of the dominant negative effect—nicely compatible with the observed lysosomal storage—would be a gradual accumulation of nondegradable CSP $\alpha$  aggregates in the lysosomal system. We followed this lead experimentally; however, we failed to identify CSP $\alpha$  in storage lysosomes by using immunohistochemistry analysis of fixed brain samples as well as in storage granules isolated from affected individuals' brains by using immunoblot analysis (data not shown).

CSP $\alpha$  associates with 70 kDa heat-shock cognate protein (Hsc70) and small glutamine-rich tetratricopeptide repeat domain protein (SGT) and forms an enzymatically active chaperone complex that is tethered to synaptic vesicles and ensures, in cooperation with other chaperones such as 40 kDa heat-shock protein (Hsp40),<sup>20</sup> 90 kDa heat-shock protein (Hsp90),<sup>21</sup> Hsc70 interacting protein (HIP)<sup>22</sup> and Hsp70 organizing protein (HOP),<sup>22</sup> correct conformation of many proteins essential for the functionality of synapses. It was shown that CSP $\alpha$  deletion causes progressive neurodegeneration and reduced life span in *Drosophila melanogaster*<sup>23</sup> and knockout mice.<sup>24,25</sup> Depletion of CSP $\alpha$  interferes with SNARE complex formation and has a profound effect on presynaptic vesicle release and synaptic function.<sup>19,24,26–29</sup> Thus, these CSP $\alpha$  mutations might lead to presynaptic dysfunction, explaining some of the neurological symptoms observed in affected individuals. In parallel, dysfunction of the CSP $\alpha$ /Hsc70/SGT chaperone complex might affect the folding quality of many client proteins and make them vulnerable to aggregation and degradation.<sup>30</sup> This could, in the long term, lead to lysosomal accumulation of misfolded and proteolysis-resistant proteins in the form of characteristic ceroid deposits in neurons.

Our finding of neurodegenerative disease caused by mutations in *DNAJC5* thus confirms a neuroprotective role for CSP $\alpha$  in humans and advocates detailed investigation of CSP $\alpha$  in the NCLs and other neurodegenerative diseases

presenting with neuronal protein aggregation. It is interesting that there is no visual failure in the cases reported here, in contrast to the rapid loss of vision in mice completely lacking CSP $\alpha$  function.<sup>25</sup>

In this study we were able to explain ~25% of ANCL cases tested, though not all were known to be autosomal-dominant and some could have been misdiagnosed. Those families that do not carry mutations in *DNAJC5* or other known NCL genes provide a resource for identification of further genes whose disruption causes late-onset NCL.

In conclusion, we believe that our work represents an important step in the genetic dissection of a genetically heterogeneous group of ANCLs. From a clinical perspective, and in the absence of specific biochemical markers, our finding, together with the recent identification of *CLN6* mutations in adult-onset recessive Kufs type A disease,<sup>2</sup> provide essential information allowing efficient DNA-based testing in families as well as simplex cases with ANCL presentation.

### Supplemental Data

Supplemental Data include nine figures and three tables and can be found with this article online at <http://www.cell.com/AJHG/>.

### Acknowledgments

This work was supported by the Grant Agency of Charles University of Prague (project 299911), the Ministry of Education of the Czech Republic (projects 1M6837805002 and MSM0021620806), Belgian Science Policy Office Interuniversity Attraction Poles program P6/43, Flemish Government Methusalem Excellence grant, Research Foundation Flanders (J.v.d.Z, postdoctoral fellowship), and the Batten Disease Support and Research Association. We thank clinical colleagues and families who contributed samples used in this study, especially John Morris and Joanne Porter (UCL563), David Sleat and the late Krystyna Wisniewski (UCL519), and Aristotle Siakotos (UCL328).

Received: May 3, 2011

Revised: July 4, 2011

Accepted: July 9, 2011

Published online: August 4, 2011

### Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, <http://www.1000genomes.org/>

CSS-Palm 2.0, <http://csspalm.biocuckoo.org/online.php>

DAVID, Database for Annotation, Visualization and Integrated

Discovery version 6.7, <http://david.abcc.ncifcrf.gov/>

dbSNP, <http://www.ncbi.nlm.nih.gov/projects/SNP/>

ExpASY, <http://expasy.org>

Gene Expression Omnibus, <http://www.ncbi.nlm.nih.gov/geo/>

GeneReviews, Mole, S.E., and Williams, R.E. (2010). Neuronal

Ceroid-Lipofuscinoses, [www.ncbi.nlm.nih.gov/books/NBK1428](http://www.ncbi.nlm.nih.gov/books/NBK1428)

Online Mendelian Inheritance in man (OMIM), <http://www.omim.org>

PolyPhen-2, <http://genetics.bwh.harvard.edu/pph2/>

## Accession Numbers

Gene-expression data are available at the Gene Expression Omnibus (GEO) repository under accession GSE30369.

## References

1. Mole, S.E., Williams, R.E., and Goebel, H.H. (2011). *The Neuronal Ceroid Lipofuscinoses (Batten Disease)* (Oxford: Oxford University Press).
2. Arsov, T., Smith, K.R., Damiano, J., Franceschetti, S., Canafoglia, L., Bromhead, C.J., Andermann, E., Vears, D.F., Cossette, P., Rajagopalan, S., et al. (2011). Kufs disease, the major adult form of neuronal ceroid lipofuscinosis, caused by mutations in CLN6. *Am. J. Hum. Genet.* *88*, 566–573.
3. Boehme, D.H., Cottrell, J.C., Leonberg, S.C., and Zeman, W. (1971). A dominant form of neuronal ceroid-lipofuscinosis. *Brain* *94*, 745–760.
4. Ferrer, I., Arbizu, T., Peña, J., and Serra, J.P. (1980). A golgi and ultrastructural study of a dominant form of Kufs' disease. *J. Neurol.* *222*, 183–190.
5. Josephson, S.A., Schmidt, R.E., Millsap, P., McManus, D.Q., and Morris, J.C. (2001). Autosomal dominant Kufs' disease: A cause of early onset dementia. *J. Neurol. Sci.* *188*, 51–60.
6. Burneo, J.G., Arnold, T., Palmer, C.A., Kuzniecky, R.I., Oh, S.J., and Faught, E. (2003). Adult-onset neuronal ceroid lipofuscinosis (Kufs disease) with autosomal dominant inheritance in Alabama. *Epilepsia* *44*, 841–846.
7. Sims, K.B., Cole, A.J., Sherman, J.C., Caruso, P.A., and Snuderl, M. (2011). Case records of the Massachusetts General Hospital. Case 8-2011. A 32-year-old woman with seizures and cognitive decline. *N. Engl. J. Med.* *364*, 1062–1074.
8. Nijssen, P.C., Brusse, E., Leyten, A.C., Martin, J.J., Teepen, J.L., and Roos, R.A. (2002). Autosomal dominant adult neuronal ceroid lipofuscinosis: Parkinsonism due to both striatal and nigral dysfunction. *Mov. Disord.* *17*, 482–487.
9. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* *30*, 97–101.
10. Thiele, H., and Nürnberg, P. (2005). HaploPainter: A tool for drawing pedigrees with complex haplotypes. *Bioinformatics* *21*, 1730–1732.
11. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
12. Jiang, H., Orr, A., Guernsey, D.L., Robitaille, J., Asselin, G., Samuels, M.E., and Dubé, M.P. (2009). Application of homozygosity haplotype analysis to genetic mapping with high-density SNP genotype data. *PLoS ONE* *4*, e5280.
13. Landmann, L. (2002). Deconvolution improves colocalization analysis of multiple fluorochromes in 3D confocal data sets more than filtering techniques. *J. Microsc.* *208*, 134–147.
14. Manders, E.M.M., Verbeek, F.J., and Aten, J.A. (1993). Measurement of Colocalization of Objects in Dual-Color Confocal Images. *Journal of Microscopy* *169*, 375–382.
15. Greaves, J., Salaun, C., Fukata, Y., Fukata, M., and Chamberlain, L.H. (2008). Palmitoylation and membrane interactions of the neuroprotective chaperone cysteine-string protein. *J. Biol. Chem.* *283*, 25014–25026.
16. Greaves, J., and Chamberlain, L.H. (2006). Dual role of the cysteine-string domain in membrane binding and palmitoylation-dependent sorting of the molecular chaperone cysteine-string protein. *Mol. Biol. Cell* *17*, 4748–4759.
17. Chamberlain, L.H., and Burgoyne, R.D. (1998). The cysteine-string domain of the secretory vesicle cysteine-string protein is required for membrane targeting. *Biochem. J.* *335*, 205–209.
18. Swayne, L.A., Blattler, C., Kay, J.G., and Braun, J.E. (2003). Oligomerization characteristics of cysteine string protein. *Biochem. Biophys. Res. Commun.* *300*, 921–926.
19. Xu, F., Proft, J., Gibbs, S., Winkfein, B., Johnson, J.N., Syed, N., and Braun, J.E. (2010). Quercetin targets cysteine string protein (CSPalpha) and impairs synaptic transmission. *PLoS ONE* *5*, e11045.
20. Gibbs, S.J., Barren, B., Beck, K.E., Proft, J., Zhao, X., Noskova, T., Braun, A.P., Artemyev, N.O., and Braun, J.E. (2009). Hsp40 couples with the CSPalpha chaperone complex upon induction of the heat shock response. *PLoS ONE* *4*, e4595.
21. Sakisaka, T., Meerlo, T., Matteson, J., Plutner, H., and Balch, W.E. (2002). Rab-alphaGDI activity is regulated by a Hsp90 chaperone complex. *EMBO J.* *21*, 6125–6135.
22. Rosales-Hernandez, A., Beck, K.E., Zhao, X., Braun, A.P., and Braun, J.E. (2009). RDJ2 (DNAJA2) chaperones neural G protein signaling pathways. *Cell Stress Chaperones* *14*, 71–82.
23. Zinsmaier, K.E., Eberle, K.K., Buchner, E., Walter, N., and Benzer, S. (1994). Paralysis and early death in cysteine string protein mutants of *Drosophila*. *Science* *263*, 977–980.
24. Fernández-Chacón, R., Wölfel, M., Nishimune, H., Tabares, L., Schmitz, F., Castellano-Muñoz, M., Rosenmund, C., Montesinos, M.L., Sanes, J.R., Schneggenburger, R., and Südhof, T.C. (2004). The synaptic vesicle protein CSP alpha prevents presynaptic degeneration. *Neuron* *42*, 237–251.
25. Schmitz, F., Tabares, L., Khimich, D., Strenzke, N., de la Villa-Polo, P., Castellano-Muñoz, M., Bulankina, A., Moser, T., Fernández-Chacón, R., and Südhof, T.C. (2006). CSPalpha-deficiency causes massive and rapid photoreceptor degeneration. *Proc. Natl. Acad. Sci. USA* *103*, 2926–2931.
26. Burgoyne, R.D., and Morgan, A. (2011). Chaperoning the SNAREs: A role in preventing neurodegeneration? *Nat. Cell Biol.* *13*, 8–9.
27. García-Junco-Clemente, P., Cantero, G., Gómez-Sánchez, L., Linares-Clemente, P., Martínez-López, J.A., Luján, R., and Fernández-Chacón, R. (2010). Cysteine string protein-alpha prevents activity-dependent degeneration in GABAergic synapses. *J. Neurosci.* *30*, 7377–7391.
28. Burré, J., Sharma, M., Tsetsenis, T., Buchman, V., Etherton, M.R., and Südhof, T.C. (2010). Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro. *Science* *329*, 1663–1667.
29. Sharma, M., Burré, J., and Südhof, T.C. (2011). CSPα promotes SNARE-complex assembly by chaperoning SNAP-25 during synaptic activity. *Nat. Cell Biol.* *13*, 30–39.
30. Johnson, J.N., Ahrendt, E., and Braun, J.E. (2010). CSPalpha: The neuroprotective J protein. *Biochem. Cell Biol.* *88*, 157–165.
31. Nijssen, P.C., Brekelmans, G.J., and Roos, R.A. (2009). Electroencephalography in autosomal dominant adult neuronal ceroid lipofuscinosis. *Clin. Neurophysiol.* *120*, 1782–1786.
32. Nijssen, P.C., Ceuterick, C., van Diggelen, O.P., Elleder, M., Martin, J.J., Teepen, J.L., Tyynelä, J., and Roos, R.A. (2003). Autosomal dominant adult neuronal ceroid lipofuscinosis: A

- novel form of NCL with granular osmiophilic deposits without palmitoyl protein thioesterase 1 deficiency. *Brain Pathol.* *13*, 574–581.
33. Poët, M., Kornak, U., Schweizer, M., Zdebik, A.A., Scheel, O., Hoelter, S., Wurst, W., Schmitt, A., Fuhrmann, J.C., Planells-Cases, R., et al. (2006). Lysosomal storage disease upon disruption of the neuronal chloride transport protein *CLIC-6*. *Proc. Natl. Acad. Sci. USA* *103*, 13854–13859.
34. Reif, A., Schneider, M.F., Hoyer, A., Schneider-Gold, C., Fallgatter, A.J., Roggendorf, W., and Pfulmann, B. (2003). Neuroleptic malignant syndrome in Kufs' disease. *J. Neurol. Neurosurg. Psychiatry* *74*, 385–387.

## Mutations in *ANTXR1* Cause GAPO Syndrome

Viktor Stránecký,<sup>1,9</sup> Alexander Hoischen,<sup>2,9</sup> Hana Hartmannová,<sup>1,9</sup> Maha S. Zaki,<sup>3</sup> Amit Chaudhary,<sup>4</sup> Enrique Zudaire,<sup>4</sup> Lenka Nosková,<sup>1</sup> Veronika Barešová,<sup>1</sup> Anna Přistoupilová,<sup>1</sup> Kateřina Hodaňová,<sup>1</sup> Jana Sovová,<sup>1</sup> Helena Hůlková,<sup>1</sup> Lenka Piherová,<sup>1</sup> Jayne Y. Hehir-Kwa,<sup>2</sup> Deepthi de Silva,<sup>5</sup> Manouri P. Senanayake,<sup>6</sup> Sameh Farrag,<sup>7</sup> Jiří Zeman,<sup>7</sup> Pavel Martásek,<sup>7</sup> Alice Baxová,<sup>8</sup> Hanan H. Afifi,<sup>3</sup> Brad St. Croix,<sup>4</sup> Han G. Brunner,<sup>2</sup> Samia Temtamy,<sup>3</sup> and Stanislav Kmoch<sup>1,\*</sup>

The genetic cause of GAPO syndrome, a condition characterized by growth retardation, alopecia, pseudoanodontia, and progressive visual impairment, has not previously been identified. We studied four ethnically unrelated affected individuals and identified homozygous nonsense mutations (c.262C>T [p.Arg88\*] and c.505C>T [p.Arg169\*]) or splicing mutations (c.1435-12A>G [p.Gly479Phefs\*119]) in *ANTXR1*, which encodes anthrax toxin receptor 1. The nonsense mutations predictably trigger nonsense-mediated mRNA decay, resulting in the loss of ANTXR1. The transcript with the splicing mutation theoretically encodes a truncated ANTXR1 containing a neopeptide composed of 118 unique amino acids in its C terminus. GAPO syndrome's major phenotypic features, which include dental abnormalities and the accumulation of extracellular matrix, recapitulate those found in *Antxr1*-mutant mice and point toward an underlying defect in extracellular-matrix regulation. Thus, we propose that mutations affecting ANTXR1 function are responsible for this disease's characteristic generalized defect in extracellular-matrix homeostasis.

GAPO syndrome (MIM 230740) is the acronym for a complex disorder characterized by growth retardation, alopecia, pseudoanodontia, and, in many but not all cases, progressive optic atrophy.<sup>1</sup> Although variations of these phenotypes have been associated with other syndromes, their combination is unique to individuals with GAPO syndrome, and more than 30 cases of various ethnic origins have been described.<sup>2-8</sup> Most of the cases are from consanguineous parents, and inheritance patterns within these families have suggested that the disease is inherited as an autosomal-recessive trait. Although affected individuals have no readily identifiable biochemical or endocrine abnormalities, histopathologic studies have revealed an abnormal accumulation of extracellular material,<sup>9,10</sup> and clinical presentation has shown predominant involvement of connective tissue (fibroblasts, chondrocytes, and osteoblasts), venous malformations, and heart, lung, and ocular abnormalities. These clinicopathologic changes point to a generalized defect in extracellular-matrix homeostasis. However, prior research has been unable to identify the genetic roots or reveal the basic molecular mechanisms responsible for GAPO syndrome.

To identify the genetic defect in GAPO syndrome, we performed genomic analysis in four unrelated and ethnically diverse families (Figure 1). The study was approved by institutional review boards, and the investigations were performed according to the Declaration of Helsinki principles. Adults provided informed consent, and the

affected child provided assent with parental consent. Consents to publish clinical photographs in scientific journals were also obtained.

We analyzed a previously reported Czech family trio<sup>11</sup> (called CZE1) with one affected child (II-1 [Figures 1A and 1B]) who died from a heart attack at the age of 19 years, a previously reported Egyptian family (EGY1) with one affected child (V-3 [Figures 1C and 1D])<sup>10</sup> who died from renal failure at the age of 12 years, and two recently identified cases in families from Egypt (EGY2) (VI-4 [Figures 1E and 1F]) and Sri Lanka (SRI1) (III-1 [Figures 1G and 1H]). All four cases demonstrated the major clinical hallmarks of GAPO syndrome as summarized in Table 1.

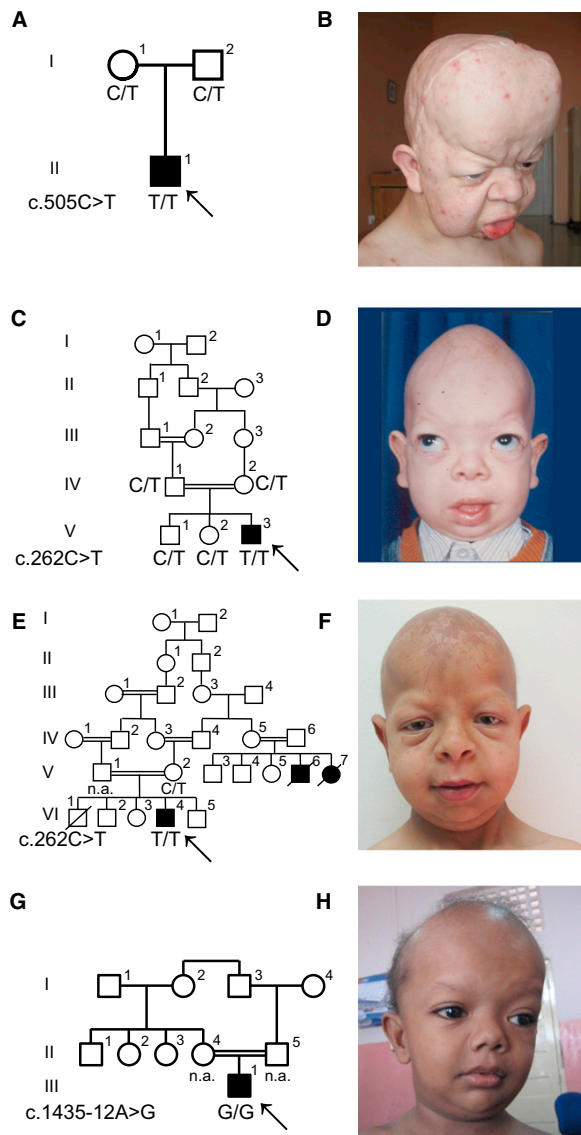
Participants provided venous blood samples, and genomic DNA was isolated with standard technology. We first genotyped genomic DNA from all three Czech family members (i.e., both unaffected parents and the affected child) by using Affymetrix GeneChip Mapping 6.0 Arrays. We used data from both SNP and copy-number probes and identified in all three individuals copy-number alterations relative to a built-in reference as previously described.<sup>12</sup> In our analysis of the Czech proband, II-1, no rare or potentially disease-causing deletion or amplification larger than 10 Kb was revealed to be compatible with an expected autosomal-recessive inheritance model. Because ~0.85% of the proband genome was found to be autozygous, we estimated that the parents might be fifth-degree relatives. Accordingly, when we used the Affymetrix Genotyping

<sup>1</sup>Institute for Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague, 120 00 Prague 2, Czech Republic; <sup>2</sup>Department of Human Genetics, Nijmegen Center for Molecular Life Sciences, Institute for Genetic and Metabolic Disease, Radboud University Nijmegen Medical Center, 6500 HC Nijmegen, the Netherlands; <sup>3</sup>Clinical Genetics Department, National Research Centre, Cairo 12311, Egypt; <sup>4</sup>Tumor Angiogenesis Section, Frederick National Laboratory for Cancer Research, Frederick, MD 21702-1201, USA; <sup>5</sup>Faculty of Medicine, University of Kelaniya, Ragama 11010, Sri Lanka; <sup>6</sup>Department of Paediatrics, Faculty of Medicine, University of Colombo, Colombo 08, Sri Lanka; <sup>7</sup>Department of Pediatrics, First Faculty of Medicine, Charles University in Prague, 120 00 Prague 2, Czech Republic; <sup>8</sup>Department of Medical Genetics, First Faculty of Medicine, Charles University in Prague, 120 00 Prague 2, Czech Republic

<sup>9</sup>These authors contributed equally to this work

\*Correspondence: [skmoch@lf1.cuni.cz](mailto:skmoch@lf1.cuni.cz)

<http://dx.doi.org/10.1016/j.ajhg.2013.03.023>. ©2013 by The American Society of Human Genetics. All rights reserved.



**Figure 1. Family Pedigrees, Segregation of the *ANTXR1* Mutations, and Facial Appearance of the Probands with GAPO Syndrome**

(A) Pedigree of the Czech family, CZE1.  
 (B) Facial appearance of the Czech proband, II-1, after neurosurgery due to a posthemorrhagic malacic lesion in the frontal cortex.  
 (C) Pedigree of Egyptian family EGY1.  
 (D) Facial appearance of Egyptian proband V-3.  
 (E) Pedigree of Egyptian family EGY2. Note extensive consanguinity and similarly affected relatives.  
 (F) Facial appearance of Egyptian proband VI-3.  
 (G) Pedigree of the Sri Lankan family, SRI1.  
 (H) Facial appearance of the Sri Lankan proband, III-1.  
 Black symbols denote affected individuals, and open symbols denote unaffected parents and siblings. "NA" indicates that DNA was not available for the investigation. The diagonal slash denotes deceased individuals. The arrows indicate the probands displayed in the corresponding pictures.

Console Software version 4.1 algorithm to compare values from the user's sample set and SNP-specific distributions derived from a reference set of 200 ethnically diverse individuals,<sup>12</sup> we identified in the proband sample two extended autozygous regions on chromosome 2 (chr2: 60,738,227–74,103,186) and chromosome 4 (chr4: 20,458,688–32,646,855), and they contained 114 and 29 genes, respectively (Figure S1A, available online).

To directly identify potential disease-causing mutations, we sequenced and analyzed the exomes of all three individuals from the Czech family as previously described.<sup>13</sup> In the resulting data set, we searched for variants that were either private or present in the internal exome database or in the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project Exome Variant Server with allele frequencies lower than 0.1% and whose genotypes were compatible with an expected autosomal-recessive model of the disease. This analysis revealed 121 candidate variants in proband II-1. However, the only relevant variant compatible with a recessive disorder was a homozygous nonsense mutation (c.505C>T [p.Arg169\*]) in *ANTXR1*, encoding anthrax toxin receptor 1, also known as tumor endothelial marker 8 (TEM8) (RefSeq accession number NM\_032208.2) (Table S1 and S2). This mutation is localized in one of the extended homozygous regions identified in the proband's genome and was inherited from both parents, who are heterozygous carriers. We confirmed the presence of the c.505C>T mutation in the parents and in the proband's genomic DNA by Sanger sequencing (Figure S2A). The identified mutation was not reported in dbSNP, 1000 Genomes, the Exome Variant Server, or an internal exome database (>120 exomes). It was absent in an additional 200 control samples that we analyzed with an XhoI-based restriction assay of the corresponding PCR-amplified genomic DNA fragments. To confirm the recurrence of *ANTXR1* mutations in another affected family, we sequenced *ANTXR1* genomic DNA of the proband (VI-4) from family EGY2 and identified a homozygous nonsense mutation (c.262C>T [p.Arg88\*]) (Figure S2B), which was also localized in an apparently autozygous region (according to the homozygous genotypes for common SNPs present across the analyzed *ANTXR1* genomic sequence and quantitative-PCR results verifying the presence of both mutated alleles; Figure S3) and was not reported in dbSNP, 1000 Genomes, the Exome Variant Server, an internal exome database, or 200 control samples analyzed with a BsaJI-based restriction assay performed on PCR-amplified genomic DNA fragments.

In parallel, DNA samples from two other cases (V-3 from family EGY1 and III-1 from family SRI1) were independently analyzed by exome sequencing performed essentially as above and as described previously.<sup>14–17</sup> As in a previous study,<sup>14</sup> autozygous regions were identified directly from the exome data of both samples. Strikingly, this resulted in two large overlapping regions of homozygosity on chromosome 2; the total overlap was a ~27 Mb region (chromosome 2: 43–70 Mb) containing 144 genes

**Table 1. Main Clinical Findings in the Four Studied Individuals with GAPO Syndrome**

Features	Cases			
	V-3 from EGY1	VI-4 from EGY2	II-1 from CZE1	III-1 from SRI1
<b>General</b>				
Age at evaluation (years)	3	10	18	4
Gender	male	male	male	male
Parental consanguinity (first or second cousins)	+	+	–	+
Family history of similarly affected case	–	+	–	–
Height	–2 SDs	–3.7 SDs	–4 SDs	–4 SDs
Head circumference	–2 SDs	+2 SDs	+1 SD	–2 SDs
Bone age	delayed	mild delay	normal for age	delayed
<b>Craniofacial</b>				
Plagiocephaly	+	+	+	–
Frontal bossing	+	+	+	+
Broad forehead	+	+	+	+
Enlarged persistent anterior fontanel	+	+	+	+
Widely spaced eyes	+	+	+	+
Epicantus	+	+	–	+
Depressed nasal bridge	+	+	+	+
Short nose	+	+	+	+
Long philtrum	+	+	+	+
Thick and anteverted nares	+	+	+	+
Thick lower lip	+	+	+	+
Micrognathia	+	+	+	+
Pseudoanodontia	+	+	+	+
<b>Skin and Hair</b>				
Sparse scalp hair (alopecia)	+	+	+	+
Scalp pigmented with scars and papules	–	+	–	–
Sparse eyebrows and eyelashes	+	+	+	+
<b>Ophthalmologic</b>				
Megalocornea	+	+	+	NR
Nystagmus	+	–	–	+
Esotropia	+	–	+	–
Shallow anterior chamber	–	+	+	NR
Bilateral engorged tortuous retinal vessels	–	+	+	+
Bilateral optic atrophy	+	–	+	+
VEP (abnormal pattern)	+	–	+	NR
<b>Other</b>				
Umbilical hernia	+	+	+	–
Hyperextensible joints	+	–	–	+
Mild webbing between fingers	–	+	–	–

(Continued on next page)

**Table 1. Continued**

Features	Cases			
	V-3 from EGY1	VI-4 from EGY2	II-1 from CZE1	III-1 from SRI1
Facial nerve palsy	+	–	–	–
MRI brain changes	bilateral high signal of deep white matter at deep parietal and occipital region and around the optic nerve	ND	ND	ND

Abbreviations are as follows: VEP, visual-evoked potential; NR, not recorded; and ND, not done.

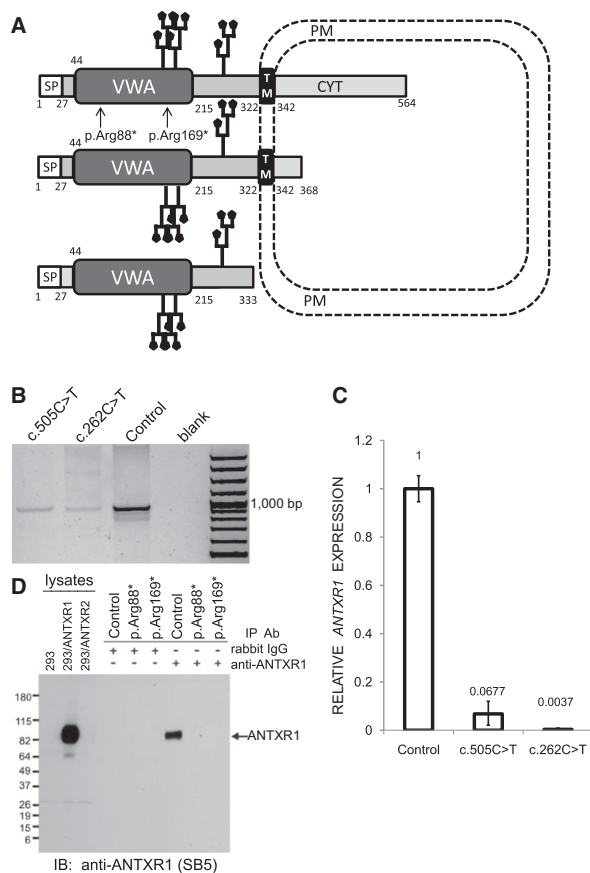
(Figures S1B and S1C). The only gene harboring private or rare homozygous coding or splice-site variants within this overlapping region was *ANTXR1*. For the EGY1 case (V-3), we identified the same nonsense mutation (c.262C>T [p.Arg88\*]) as for the EGY2 case (VI-4) (Figure S2C), whereas in the SRI1 case (III-1), we identified a substitution, c.1435–12A>G (Figure S2D). This latter variant is predicted by ESE finder<sup>18,19</sup> to generate an alternative strong splice acceptor site 11 nucleotides upstream of the last exon (Figure S4), and this would theoretically result in a frameshift of the complete reading frame of the last exon and proteosynthesis of a truncated *ANTXR1* containing a neopeptide composed of 118 unique amino acids in its C terminus (p.Gly479Phefs\*119) (Figure S5).

Two of the affected probands, V-3 and VI-4, from Egyptian families EGY1 and EGY2, respectively, harbor an identical c.262C>T [p.Arg88\*] mutation. To determine whether these two probands might be distantly related and share a mutated chromosomal segment from a common ancestor, we examined *ANTXR1* intragenic SNP haplotypes obtained by exome sequencing (for V-3 from EGY1) and Sanger sequencing (for VI-4 from EGY2). This revealed that the c.262C>T mutations are present on two distinct haplotypes, indicating that these mutations most likely developed independently or that these families share a very old ancestral allele (Figure S6). The c.262C nucleotide belongs to a CpG doublet, making deamination of the cytosine a possible explanation for the recurrence of the mutation.

*ANTXR1*, also called TEM8, was initially identified as one of the tumor endothelial markers (TEMs) that displays elevated protein levels during tumor angiogenesis.<sup>20,21</sup> Soon after its discovery, it was independently identified as the anthrax toxin receptor (ATR).<sup>22</sup> Several variants of human *ANTXR1* have been proposed to exist on the basis of the identification of rare alternative mRNA splice variants (Figure 2A). The biosynthesis of all known variants is driven by a common signal peptide (amino acids 1–27) and proceeds by cotranslational translocation in the endoplasmic reticulum. The full-length *ANTXR1* variant v1, (RefSeq NM\_032208.2) is by far the most prevalent transcript found in databases of cDNA and expressed sequence tags. It encodes a single-pass type 1 transmembrane glycoprotein that has a molecular weight of approximately 85 kDa and that is composed of a predicted N-terminal extracellular sequence (amino acids 28–322) containing a von Wille-

brand type A domain (amino acids 44–215), a transmembrane domain (amino acids 322–342), and large cytoplasmic domain (amino acids 343–564) (isoform 1). Variant v2 (RefSeq NM\_053034.2) encodes protein isoform 2, which is structurally similar to variant 1 but contains a much shorter cytoplasmic domain (amino acids 343–368).<sup>22</sup> Variant v3 (RefSeq NM\_018153.3) encodes protein isoform 3, which does not contain the transmembrane or cytoplasmic domains and is predicted to be secreted.<sup>23</sup> Two other transcript variants have recently been identified: v4 (GenBank accession number JX424838.1), potentially encoding membrane-bound protein isoform 4, which, compared to isoform 1, lacks 36 aa residues in its cytoplasmic domain, and v5 (GenBank JX424839.1), potentially encoding secreted protein isoform 5, which, compared to isoform 3, has an alternative C-terminal sequence.<sup>24</sup> Because detecting cDNA for alternative splice variants v2–v5 is difficult in that it requires as many as 60 cycles of nested PCR<sup>24</sup> and because similar conserved variants in other species have not yet been described, it is currently unclear whether they represent transcriptional noise caused by inappropriate splicing events and whether the encoded protein isoforms are produced at sufficient endogenous levels needed to impact biological function. However, the full-length *ANTXR1* isoform 1 has been shown to promote interaction between cells and various components of the extracellular matrix,<sup>25,26</sup> link extracellular ligands to the actin cytoskeleton,<sup>25,27</sup> and regulate cell spreading.<sup>28–30</sup>

In three of the four GAPO cases, the identified mutations introduce premature stop codons in *ANTXR1* mRNA, and in the fourth case (III-1 in SRI1), the mutation most likely results in a loss-of-function allele. From cases II-1 (CZE1) and VI-4 (EGY2), we studied skin fibroblast cell lines harboring the *ANTXR1* mutations encoding p.Arg169\* and p.Arg88\*. To characterize the molecular consequences of the identified mutations on *ANTXR1* mRNA expression, splicing, and stability, we isolated total RNA from two cases and control skin fibroblasts and performed RT-PCR and quantitative-PCR analyses. In fibroblasts from affected individuals, we found a single PCR product comparable in size to a control specimen (Figure 2B). Sanger sequencing demonstrated that the obtained PCR products corresponded to cDNA of the major transcript variant v1 of *ANTXR1* and independently confirmed the presence of the premature-stop-codon-encoding mutations previously identified in corresponding genomic DNA in affected



**Figure 2. Effects of the Identified *ANTXR1* Mutations**

(A) A schematic representation of ANTXR1 shows the protein structure, cellular topology, and location of the p.Arg88\* and p.Arg169\* substitutions. Two potential N-glycosylation sites are depicted. The numbers denote amino acid residues defining the boundaries of predicted ANTXR1 domains. Only ANTXR1 isoforms 1, 2, and 3 are depicted. Abbreviations are as follows: SP, signal peptide; VWA, Von Willebrand factor type A domain; TM, transmembrane domain; CYT, cytoplasmic domain; and PM, plasma membrane.

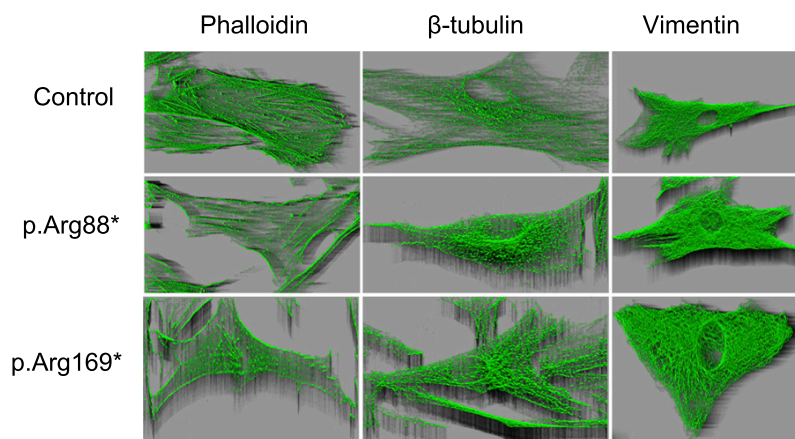
(B) ANTXR1 cDNA analysis. Total RNA was isolated from pellets from a cultured skin fibroblast cell line with the use of TRIZOL solution (Invitrogen). RNA concentrations were determined spectrophotometrically at A260 nm by NanoDrop (NanoDrop Technologies), and RNA quality was verified with an Agilent 2100 bioanalyser, RNA Lab-on-a-Chip (Agilent Technologies). The first-strand cDNA synthesis was carried out with an oligo-dT primer and SuperScript III Reverse Transcriptase (Life Technologies). ANTXR1 cDNA was PCR amplified from the synthesized first-strand cDNA with oligonucleotide primers designed to span and amplify all three ANTXR1 variants in parallel (Table S2). Lanes 1 and 2 show reduced amounts of RT-PCR products from cases with p.Arg169\* and p.Arg88\* substitutions, and lane 3 shows the cDNA amount obtained under identical conditions from a control cell line. C. Lane 4 is a negative control. Lane 5 is a 100 bp DNA ladder.

(C) Relative expression levels of ANTXR1 mRNA amounts normalized to glyceraldehydes-3-phosphate dehydrogenase (*GAPDH*) mRNA amounts in skin fibroblasts. Quantitative PCR was carried out on a StepOne Plus Real Time System (Applied Biosystems). The reactions were carried out in a 96-well plate in a 20  $\mu$ l reaction volume containing 10  $\mu$ l 2  $\times$  Maxima SYBR Green qPCR Master Mix (Thermo Scientific), 0.2  $\mu$ M forward and reverse primer, and

5 ng cDNA. Data were analyzed by StepOne Software v.2.0. The comparative Ct ( $\Delta\Delta$  Ct) method was used for normalizing target-gene mRNA to *GAPDH* mRNA. The relative amounts of the ANTXR1 cDNA were significantly reduced in cases compared to control samples. The means  $\pm$  SD of three experiments performed in triplicate are shown.

(D) Immunoblot (IB) analysis of immunoprecipitated (IP) total-protein extracts showing absence of ANTXR1 in cultured skin fibroblasts from cases with p.Arg88\* and p.Arg169\* substitutions. Cultured cells were lysed in TNT lysis buffer (50 mM Tris [pH 7.5], 75 mM NaCl, and 1% Triton X-100 plus complete protease inhibitor cocktail [Roche]) and clarified by centrifugation. Protein extracts were quantified with a BCA assay (Pierce), normalized, and immunoprecipitated with a rabbit monoclonal ANTXR1 antibody (clone 37). This rabbit monoclonal antibody was produced as part of a collaboration between Epitomics and the National Cancer Institute and will be described in more detail elsewhere. After immunoprecipitation using protein A agarose, protein extracts were separated by SDS-PAGE, transferred to a PDVF membrane (Millipore), and detected by immunoblotting with SB5 mouse monoclonal ANTXR1 antibodies followed by HRP-conjugated anti-mouse or anti-rabbit F(ab')<sub>2</sub> secondary antibodies (Jackson). Chemiluminescence was visualized with the ECL plus system (Amersham) according to the supplier's instructions. Lysates of 293 cells stably transfected with an empty vector (293), 293 cells stably expressing human ANTXR1 (293/ANTXR1), and 293 cells stably expressing ANTXR2 (293/ANTXR2) were used as negative, positive, and specificity controls, respectively. Equal protein amounts in the original lysates were immunoprecipitated in parallel with either control rabbit nonspecific IgG antibodies or rabbit ANTXR1 antibodies.





**Figure 3. Immunofluorescence Analysis of Cultured Skin Fibroblasts**

The cells were grown on 70 mm<sup>2</sup> glass chamber slides (Lab-Tek, Nalge Nunc International) for 48 hr. Then the cells were fixed with 4% paraformaldehyde in PBS, permeabilized in 0.1% TRITON, washed, blocked with 5% BSA in PBS, and incubated in a humidified chamber at 4°C overnight with mouse monoclonal β-tubulin antibody (Sigma) or Vimentin (V9) antibody (BioGenex). For fluorescence detection, species-specific secondary antibodies Alexa Fluor 488 or 555 (Molecular Probes, Invitrogen) were used. For actin staining, Alexa Fluor 488 Phalloidin (Molecular Probes, Invitrogen) was used. Slides were mounted in fluorescence mounting medium Immu-Mount (Shandon Lipshaw) and analyzed by confocal

microscopy. XYZ images were sampled according to Nyquist criterion with the Nikon TE2000E C1si laser-scanning confocal microscope with a Nikon PlanApo objective (60×, numerical aperture 1.40) and 488 and 543 laser lines. Images were restored with a classic maximum-likelihood restoration algorithm in the Huygens Professional Software (SVI, Hilversum, the Netherlands). Phalloidin staining demonstrated remarkable alterations in the actin cytoskeletal network in cell lines from GAPO cases with p.Arg169\* and p.Arg88\* substitutions. No abnormalities in microtubules or intermediate filaments were detected with β-tubulin or vimentin staining, respectively.

because *ANTXR1* has been previously shown to interact with actin. These studies revealed a striking reorganization of the actin cytoskeletal microfilaments specifically in the GAPO fibroblasts, but β-tubulin and vimentin staining of microtubules and intermediate filaments, respectively, were unaltered (Figure 3). This suggests that *ANTXR1*, a molecule mediating the coupling of extracellular ligands to the actin cytoskeleton, is crucial for actin assembly and that disruption of the actin network might be the major pathogenetic event leading to altered cell-adhesion properties and progressive extracellular-matrix buildup observed in individuals with GAPO syndrome.

Unfortunately, we did not have the opportunity to study mRNA processing, protein production, or the actin network in the fourth GAPO case with *ANTXR1* mutation c.1435–12A>G. In this case, the mutation theoretically affects splicing of *ANTXR1* mRNA and potentially encodes *ANTXR1* isoforms 1 and 2 with altered C-terminal cytoplasmic tails; if produced, these isoforms could potentially retain some biological functions. In addition to this, the mutation should not theoretically affect proteosynthesis of the secreted *ANTXR1* isoform 3. This could help to explain the evidently milder clinical presentation of this case compared to the other three cases with nonsense mutations.

*ANTXR1* is most highly produced in tumor endothelial cells and other tumor stromal cells, which might include both pericytes and fibroblasts.<sup>25,32,33</sup> *Antxr1*-mutant mice with targeted deletion of exon 13—encoding the transmembrane domain—are viable and progressively develop misaligned incisor teeth, and female mice are infertile.<sup>34</sup> In another mutant mouse model, complete *Antxr1* disruption due to removal of exon 1—encoding the start codon and signal peptide—leads to a moderate excess of extracellular matrix in many tissues, including the ovaries, uterus, skin, hair follicles, cranial sutures of the skull, and the peri-

odontal ligament of the incisors, resulting in dental dysplasia.<sup>31</sup> These features are consistent with the clinical presentation of individuals afflicted with GAPO syndrome. It is also notable that some of the individuals with GAPO syndrome have infantile hemangiomas,<sup>35</sup> which have been associated with germline heterozygosity for missense mutations in *ANTXR1*<sup>36</sup> and dysfunction of the complex formed by VEGFR2, β1 integrin, and *ANTXR1*.<sup>37</sup> Generalized extracellular-matrix-homeostasis defects observed in GAPO-syndrome-affected individuals with *ANTXR1* mutations are similar to those found in individuals with juvenile hyaline fibromatosis (MIM 228600) and infantile systemic hyalinosis (MIM 236490), which are allelic disorders caused by mutations in anthrax toxin receptor 2 (*ANTXR2*), also known as capillary morphogenesis gene 2 (*CMG2*), an *ANTXR1* homolog.<sup>38</sup>

We conclude that our data, together with recapitulation of many of the phenotypic features characteristic of GAPO syndrome in *Antxr1*-mutant mice and individuals with *ANTXR2* mutations, strongly suggest involvement of *ANTXR1* mutations in the generalized extracellular-matrix-homeostasis defect characteristic of this disease. From a clinical perspective, our finding provides essential information for DNA testing in other families. In addition, autopsy tissues and cultured skin fibroblasts from these affected individuals represent an interesting cellular model and potential resource for detailed studies on the pathogenesis of individual clinical symptoms present in GAPO syndrome and studies focused on *ANTXR1* functions in general.

#### Supplemental Data

Supplemental Data include six figures and three tables and can be found with this article online at <http://www.cell.com/AJHG>.

## Acknowledgments

This work was supported by Charles University institutional programs PRVOUK-P24/LF1/3, UNCE 204011, and SVV2012/2645; by the Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University (CZ.1.05/1.1.00/02.0109); by the European Regional Development Fund; by grant NT13116-4/2012 from the Ministry of Education and Ministry of Health of the Czech Republic; and by the intramural research program of the National Cancer Institute, National Institutes of Health, US Department of Health and Social Services. A.H. was supported by the Netherlands Organization for Health Research and Development (ZonMW 916-12-095). We thank clinical colleagues and families who contributed samples used in this study, as well as members of the Genomic Disorders Group Nijmegen for their technical support. A.C. and B.S.C. are coinventors of filed patents related to the development of *ANTXR1* antibodies for cancer therapy.

Received: November 9, 2012

Revised: January 30, 2013

Accepted: March 28, 2013

Published: April 18, 2013

## Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, <http://www.1000genomes.org/>  
ESE finder, [http://rulai.cshl.edu/cgi-bin/tools/ESE3/ese\\_finder.cgi?process=home](http://rulai.cshl.edu/cgi-bin/tools/ESE3/ese_finder.cgi?process=home)  
dbSNP, <http://www.ncbi.nlm.nih.gov/projects/SNP/>  
GenBank, <http://www.ncbi.nlm.nih.gov/genbank/>  
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, <http://evs.gs.washington.edu/EVS/>  
Online Mendelian Inheritance in Man (OMIM), <http://www.omim.org>  
RefSeq, <http://www.ncbi.nlm.nih.gov/RefSeq>

## References

1. Tipton, R.E., and Gorlin, R.J. (1984). Growth retardation, alopecia, pseudo-anodontia, and optic atrophy—the GAPO syndrome: report of a patient and review of the literature. *Am. J. Med. Genet.* *19*, 209–216.
2. Goloni-Bertollo, E.M., Ruiz, M.T., Goloni, C.B., Muniz, M.P., Valério, N.I., and Pavarino-Bertelli, E.C. (2008). GAPO syndrome: three new Brazilian cases, additional osseous manifestations, and review of the literature. *Am. J. Med. Genet. A.* *146A*, 1523–1529.
3. Sinha, R., Trikha, A., Laha, A., Raviraj, R., and Kumar, R. (2011). Anesthetic management of a patient with GAPO syndrome for glaucoma surgery. *Paediatr. Anaesth.* *21*, 910–912.
4. Nanda, A., Al-Ateeqi, W.A., Al-Khawari, M.A., Alsaleh, Q.A., and Anim, J.T. (2010). GAPO syndrome: a report of two siblings and a review of literature. *Pediatr. Dermatol.* *27*, 156–161.
5. Lei, S., Iyengar, S., Shan, L., Cherwek, D.H., Murthy, S., and Wong, A.M. (2010). GAPO syndrome: a case associated with bilateral interstitial keratitis and hypothyroidism. *Clin. Dysmorphol.* *19*, 79–81.
6. Castrillon-Oberndorfer, G., Seeberger, R., Bacon, C., Engel, M., Ebinger, F., and Thiele, O.C. (2010). GAPO syndrome associated with craniofacial vascular malformation. *Am. J. Med. Genet. A.* *152A*, 225–227.
7. Kocabay, G., and Mert, M. (2009). GAPO syndrome associated with dilated cardiomyopathy: an unreported association. *Am. J. Med. Genet. A.* *149A*, 415–416.
8. Demirgüneş, E.F., Ersoy-Evans, S., and Karaduman, A. (2009). GAPO syndrome with the novel features of pulmonary hypertension, ankyloglossia, and prognathism. *Am. J. Med. Genet. A.* *149A*, 802–805.
9. Wajntal, A., Koiffmann, C.P., Mendonça, B.B., Epps-Quaglia, D., Sotto, M.N., Rati, P.B., and Opitz, J.M. (1990). GAPO syndrome (McKusick 23074)—a connective tissue disorder: report on two affected sibs and on the pathologic findings in the older. *Am. J. Med. Genet.* *37*, 213–223.
10. Meguid, N.A., Afifi, H.H., Ramzy, M.I., Hindawy, A., and Temtamy, S.A. (1997). GAPO syndrome: first Egyptian case with ultrastructural changes in the gingiva. *Clin. Genet.* *52*, 110–115.
11. Baxova, A., Kozlowski, K., Obersztyń, E., and Zeman, J. (1997). GAPO syndrome (Radiographic clues to early diagnosis). *Radiol. Med. (Torino)* *93*, 289–291.
12. van de Steeg, E., Stránecký, V., Hartmannová, H., Nosková, L., Hřebíček, M., Wagenaar, E., van Esch, A., de Waart, D.R., Oude Elferink, R.P., Kenworthy, K.E., et al. (2012). Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver. *J. Clin. Invest.* *122*, 519–528.
13. Nosková, L., Stránecký, V., Hartmannová, H., Přistoupilová, A., Barešová, V., Ivánek, R., Hůlková, H., Jahnová, H., van der Zee, J., Staropoli, J.F., et al. (2011). Mutations in *DNAJC5*, encoding cysteine-string protein alpha, cause autosomal-dominant adult-onset neuronal ceroid lipofuscinosis. *Am. J. Hum. Genet.* *89*, 241–252.
14. Becker, J., Semler, O., Gilissen, C., Li, Y., Bolz, H.J., Giunta, C., Bergmann, C., Rohrbach, M., Koerber, F., Zimmermann, K., et al. (2011). Exome sequencing identifies truncating mutations in human *SERPINF1* in autosomal-recessive osteogenesis imperfecta. *Am. J. Hum. Genet.* *88*, 362–371.
15. Gilissen, C., Arts, H.H., Hoischen, A., Spruijt, L., Mans, D.A., Arts, P., van Lier, B., Steehouwer, M., van Reeuwijk, J., Kant, S.G., et al. (2010). Exome sequencing identifies *WDR35* variants involved in Sensenbrenner syndrome. *Am. J. Hum. Genet.* *87*, 418–423.
16. Hoischen, A., van Bon, B.W., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., de Vries, P., de Reuver, R., Wieskamp, N., Mortier, G., et al. (2010). De novo mutations of *SETBP1* cause Schinzel-Giedion syndrome. *Nat. Genet.* *42*, 483–485.
17. Hoischen, A., van Bon, B.W., Rodríguez-Santiago, B., Gilissen, C., Vissers, L.E., de Vries, P., Janssen, I., van Lier, B., Hastings, R., Smithson, S.F., et al. (2011). De novo nonsense mutations in *ASXL1* cause Bohring-Opitz syndrome. *Nat. Genet.* *43*, 729–731.
18. Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. (2003). ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* *31*, 3568–3571.
19. Smith, P.J., Zhang, C., Wang, J., Chew, S.L., Zhang, M.Q., and Krainer, A.R. (2006). An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum. Mol. Genet.* *15*, 2490–2508.
20. St Croix, B., Rago, C., Velculescu, V., Traverso, G., Romans, K.E., Montgomery, E., Lal, A., Riggins, G.J., Lengauer, C., Vogelstein, B., and Kinzler, K.W. (2000). Genes expressed in human tumor endothelium. *Science* *289*, 1197–1202.

21. Carson-Walter, E.B., Watkins, D.N., Nanda, A., Vogelstein, B., Kinzler, K.W., and St Croix, B. (2001). Cell surface tumor endothelial markers are conserved in mice and humans. *Cancer Res.* *61*, 6649–6655.
22. Bradley, K.A., Mogridge, J., Mourez, M., Collier, R.J., and Young, J.A. (2001). Identification of the cellular receptor for anthrax toxin. *Nature* *414*, 225–229.
23. Liu, S., and Leppla, S.H. (2003). Cell surface tumor endothelium marker 8 cytoplasmic tail-independent anthrax toxin binding, proteolytic processing, oligomer formation, and internalization. *J. Biol. Chem.* *278*, 5227–5234.
24. Vargas, M., Karamsetty, R., Leppla, S.H., and Chaudry, G.J. (2012). Broad expression analysis of human *ANTXR1*/*TEM8* transcripts reveals differential expression and novel splice variants. *PLoS ONE* *7*, e43174.
25. Nanda, A., Carson-Walter, E.B., Seaman, S., Barber, T.D., Stampfl, J., Singh, S., Vogelstein, B., Kinzler, K.W., and St Croix, B. (2004). *TEM8* interacts with the cleaved C5 domain of collagen alpha 3(VI). *Cancer Res.* *64*, 817–820.
26. Hotchkiss, K.A., Basile, C.M., Spring, S.C., Bonuccelli, G., Lisanti, M.P., and Terman, B.I. (2005). *TEM8* expression stimulates endothelial cell adhesion and migration by regulating cell-matrix interactions on collagen. *Exp. Cell Res.* *305*, 133–144.
27. Yang, M.Y., Chaudhary, A., Seaman, S., Dunty, J., Stevens, J., Elzarrad, M.K., Frankel, A.E., and St Croix, B. (2011). The cell surface structure of tumor endothelial marker 8 (*TEM8*) is regulated by the actin cytoskeleton. *Biochim. Biophys. Acta* *1813*, 39–49.
28. Werner, E., Kowalczyk, A.P., and Faundez, V. (2006). Anthrax toxin receptor 1/tumor endothelium marker 8 mediates cell spreading by coupling extracellular ligands to the actin cytoskeleton. *J. Biol. Chem.* *281*, 23227–23236.
29. Gu, J., Faundez, V., and Werner, E. (2010). Endosomal recycling regulates Anthrax Toxin Receptor 1/Tumor Endothelial Marker 8-dependent cell spreading. *Exp. Cell Res.* *316*, 1946–1957.
30. Verma, K., Gu, J., and Werner, E. (2011). Tumor endothelial marker 8 amplifies canonical Wnt signaling in blood vessels. *PLoS ONE* *6*, e22334.
31. Cullen, M., Seaman, S., Chaudhary, A., Yang, M.Y., Hilton, M.B., Logsdon, D., Haines, D.C., Tessarollo, L., and St Croix, B. (2009). Host-derived tumor endothelial marker 8 promotes the growth of melanoma. *Cancer Res.* *69*, 6021–6026.
32. Davies, G., Rmali, K.A., Watkins, G., Mansel, R.E., Mason, M.D., and Jiang, W.G. (2006). Elevated levels of tumour endothelial marker-8 in human breast cancer and its clinical significance. *Int. J. Oncol.* *29*, 1311–1317.
33. Chaudhary, A., and St Croix, B. (2012). Selective blockade of tumor angiogenesis. *Cell Cycle* *11*, 2253–2259.
34. Liu, S., Crown, D., Miller-Randolph, S., Moayeri, M., Wang, H., Hu, H., Morley, T., and Leppla, S.H. (2009). Capillary morphogenesis protein-2 is the major receptor mediating lethality of anthrax toxin in vivo. *Proc. Natl. Acad. Sci. USA* *106*, 12424–12429.
35. Goucha, S., Fazaa, B., Ezzine, N., Jaber, K., Elandaloussi, H., Abid, R., and Kamoun, M.R. (2000). [GAPO syndrome]. *Ann. Dermatol. Venereol.* *127*, 501–504.
36. Jinnin, M., Medici, D., Park, L., Limaye, N., Liu, Y., Boscolo, E., Bischoff, J., Vikkula, M., Boye, E., and Olsen, B.R. (2008). Suppressed NFAT-dependent VEGFR1 expression and constitutive VEGFR2 signaling in infantile hemangioma. *Nat. Med.* *14*, 1236–1246.
37. Nicolae, C., and Olsen, B.R. (2010). Unexpected matrix diseases and novel therapeutic strategies. *Cell Tissue Res.* *339*, 155–165.
38. Hanks, S., Adams, S., Douglas, J., Arbour, L., Atherton, D.J., Balci, S., Bode, H., Campbell, M.E., Feingold, M., Keser, G., et al. (2003). Mutations in the gene encoding capillary morphogenesis protein 2 cause juvenile hyaline fibromatosis and infantile systemic hyalinosis. *Am. J. Hum. Genet.* *73*, 791–800.