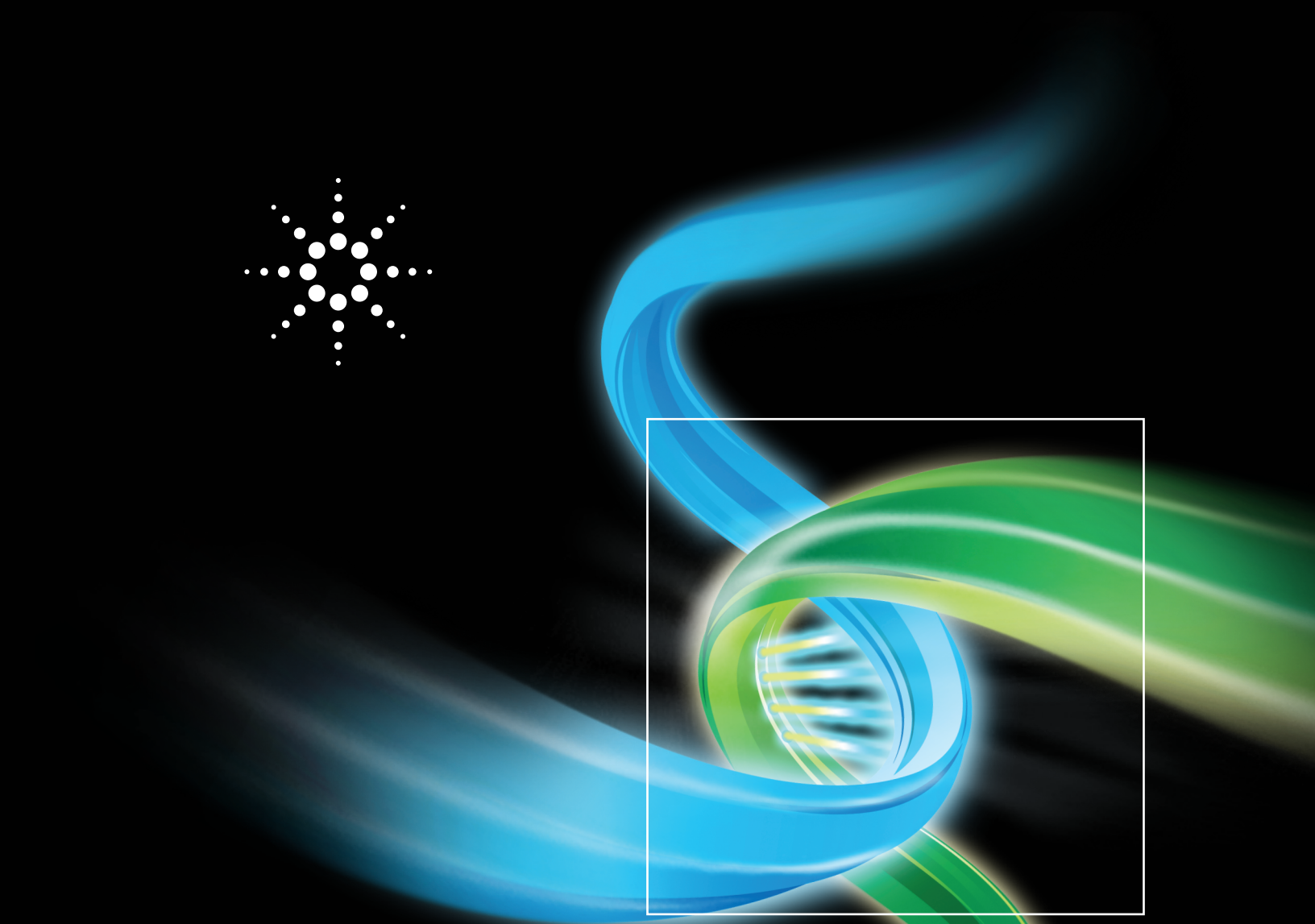


Accelerating
Scientific Discovery

SureSelect
Publication Compendium

January 2011

The Measure of Confidence

Agilent Technologies

Introduction

To the Geneticist Community,

We are very proud to share with you this compendium of several key publications utilizing Agilent's SureSelect Target Enrichment System. Since we launched the SureSelect platform in early 2009, scientists all over the world have been using our target enrichment technology followed by next-generation sequencing to enable breakthrough discoveries — from improving data analysis in linkage studies to more precisely identifying mutations associated with cancer and other diseases.

Our in-solution hybridization capture technology, which enables scientists to easily enrich genomic regions of interest prior to sequencing, was first developed in collaboration with the Broad Institute of MIT and Harvard University and subsequently published in Nature Biotechnology. Less than two years later, there are already more than 25 publications from groups that have been using this technology, and the number continues to grow. We've included here some of these key studies.

This is just a sample of how Agilent's SureSelect portfolio can help your research speed into breakthrough findings. Stay tuned, as we are increasing our product portfolio to expand your opportunities for targeted scientific discovery.

We make it, you make it happen!

Sincerely,
The Agilent SureSelect group

Highlights from the Compendium

Mendelian Diseases Discovery

Linkage studies have been very popular in the field of Mendelian diseases, but although those studies point out specific genomic regions associated with the diseases, they do not allow the identification of the causal mutation.

Researchers have shown that whole exome sequencing is a very effective approach for discovering the causal mutations for Mendelian diseases. By using SureSelect Human All-Exon kits, compatible with all major next-generation sequencing platforms, scientists can capture and sequence up to 50Mb of exon sequences making it easier to find the causal mutation. This can be done *by sequencing the exome of as little as one single affected sample,* as done in the studies from Byun et al., 2010 and Haack et al., 2010. Up to now, SureSelect has enabled research into the discovery of the causes of almost 10 different Mendelian disorders, such as Sensenbrenner syndrome (Gilissen et al., 2010), Complex I deficiency (Haack et al., 2010), Robin's syndrome (Johnston et al., 2010), Mabry syndrome (Krawitz et al., 2010), and Terminal Osseous Dysplasia (Sun et al., 2010).

Exome sequencing has also enabled scientists to identify de novo mutations that cause diseases such as mental retardation (Vissers et al., 2010) and Schinzel-Giedion syndrome (Hoischen et al., 2010).

Unlocking the Cancer Genome

Targeted resequencing is also being successfully applied in the field of cancer research. Vlierberghe et al. (2010) used Agilent's SureSelect Human X Chromosome kit to identify genetic mutations and deletions associated with T-cell acute lymphoblastic leukemia, an aggressive hematological malignancy with an increased incidence in males. Jones et al. (2010) used SureSelect to identify mutations associated with ovarian clear cell carcinoma.

The Power of Custom Designs

Walsh et al. (2010) developed a custom SureSelect assay to capture, sequence and detect mutations in 21 genes relevant for predisposition to breast and ovarian cancers, including BRCA1 and BRCA2. In this study, published in PNAS, they show that this assay enables the identification of single-nucleotide substitutions, small insertions and deletions and large structural variations.

The development of custom designs for disease research go beyond cancer, as we can see in the studies from Bonnefond et al. (2010), focused on neonatal diabetes mellitus, and from Shearer et al. (2010), who developed a design to investigate hereditary hearing loss.

Contents

nature genetics

A *de novo* paradigm for mental retardation

Lisenka E L M Vissers[1,2], Joep de Ligt[1,2], Christian Gilissen[1], Irene Janssen[1], Marloes Steehouwer[1], Petra de Vries[1], Bart van Lier[1], Peer Arts[1], Nienke Wieskamp[1], Marisol del Rosario[1], Bregje W M van Bon[1], Alexander Hoischen[1], Bert B A de Vries[1], Han G Brunner[1,3] & Joris A Veltman[1,3]

The per-generation mutation rate in humans is high. *De novo* mutations may compensate for allele loss due to severely reduced fecundity in common neurodevelopmental and psychiatric diseases, explaining a major paradox in evolutionary genetic theory. Here we used a family based exome sequencing approach to test this *de novo* mutation hypothesis in ten individuals with unexplained mental retardation. We identified and validated unique non-synonymous *de novo* mutations in nine genes. Six of these, identified in six different individuals, are likely to be pathogenic based on gene function, evolutionary conservation and mutation impact. Our findings provide strong experimental support for a *de novo* paradigm for mental retardation. Together with *de novo* copy number variation, *de novo* point mutations of large effect could explain the majority of all mental retardation cases in the population.

Recent studies[1,2] have indicated that humans have an exceptionally high per-generation mutation rate of between 7.6×10^{-9} and 2.2×10^{-8}. An average newborn is calculated to have acquired 50 to 100 new mutations in his or her genome, resulting in approximately 0.86 new amino-acid–altering mutations[2]. Spontaneous germline mutations can have serious phenotypic consequences when they affect functionally relevant bases in the genome. In fact, their occurrence may explain why diseases with a severely reduced fecundity remain frequent in the human population, especially when the mutational target is large and comprised of many genes. This would explain a major paradox in the evolutionary genetic theory of mental disorders[3,4]. In agreement with this hypothesis, *de novo* copy number variations (CNVs) are a known cause of schizophrenia, autism and mental retardation[5,6]. Much less is known about the frequency and impact of *de novo* point mutations in these common diseases. Whole genome or exome sequencing now permits the study of these mutations and their role in disease in a systematic genome-wide manner. This approach has recently been used to identify causative genes in several rare syndromes[1,7–10]. In addition, targeted resequencing of the coding exons of the X chromosome revealed nine genes associated with X-linked forms of mental retardation[11], showing the strength of these analyses in common diseases. In this study, we used a family based whole-exome–sequencing approach to test the *de novo* mutation hypothesis in an unselected cohort of individuals with mental retardation.

We sequenced the exomes of ten case-parent trios. All cases, eight males and two females, had moderate to severe mental retardation and a negative family history. Clinical evaluation did not lead to a syndromic or etiologic diagnosis (**Supplementary Note**). Prior cytogenetic analysis showed normal chromosomes, and array-based genomic profiling did not reveal *de novo* or other CNVs associated with mental retardation. In addition, fragile X syndrome was excluded by *FMR1* repeat expansion analysis. On average, we obtained 3.1 Gb of mappable sequence data per individual after exome enrichment (37 Mb of genomic sequence targeting ~18,000 genes) and sequencing on one quarter of a SOLiD sequencing slide (Online Methods and **Supplementary Table 1**). Color space reads were mapped to the reference genome. On average, 79.6% of the bases originated from the targeted exome, with 90% of the targeted exons covered at least ten times. The median exon coverage was 42-fold, indicating that the majority of variants present in each exome could be robustly detected using a custom bioinformatic analysis pipeline (**Supplementary Fig. 1**).

On average, we identified 21,755 genetic variants per individual with high confidence (**Table 1** and **Supplementary Fig. 2**). We developed an automated prioritization scheme to systematically identify all candidate dominant *de novo* mutations in each affected individual (**Fig. 1**). We first excluded all nongenic, intronic and synonymous variants other than those occurring at canonical splice sites. This first step reduced the number of candidates to an average of 5,640 non-synonymous and canonical splice site variants per affected individual. We further reduced this number to 143 by excluding all known, likely benign, variants by comparison with data from dbSNP database v130 and our in-house variant database. Next, we used the exome data from each case's parents to exclude all remaining inherited variants. This resulted in an average of five (with a range of two to seven) candidate *de novo* non-synonymous mutations per affected individual (**Table 1**).

For all 51 candidate mutations (**Supplementary Table 2**), we performed Sanger sequencing to (i) validate the mutations observed in the probands and (ii) validate the absence of the mutations in the parental DNA. Thirty-eight candidates could not be validated in the proband (covered by a median of five variant reads in the exome sequencing experiment), but 13 candidates could be validated

[1]Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [2]These authors contributed equally to this work. [3]These authors jointly directed this work. Correspondence should be addressed to H.G.B. (h.brunner@antrg.umcn.nl) or J.A.V. (j.veltman@antrg.umcn.nl).

Received 11 August; accepted 18 October; published online 14 November 2010; doi:10.1038/ng.712

Table 1 Overview of all variants detected per proband and impact of the prioritization steps for selecting candidate non-synonymous *de novo* mutations

Trio	1	2	3	4	5	6	7	8	9	10	Average
High-confidence variant calls	20,810	21,658	21,338	22,647	17,694	22,333	21,369	22,658	24,085	22,962	21,755
After exclusion of nongenic, intronic and synonymous variants	5,556	5,665	5,691	5,991	4,607	5,567	5,716	5,628	5,985	5,994	5,640
After exclusion of known variants	165	159	157	155	120	136	120	149	96	171	143
After exclusion of inherited variants	4	7	3	7	7	2	2	6	6	7	5

(covered by a median of 17 variant reads). Parental analysis validated the *de novo* occurrence for 9 of these 13 mutations, detected in seven different individuals (**Table 2** and **Supplementary Figs. 3** and **4**). We did not identify these mutations in a total of 1,664 control chromosomes, nor did we see other likely pathogenic mutations identified in the affected genes in these control chromosomes, indicating that the population frequency of these types of *de novo* mutations in these genes will be lower than 0.22% (power = 0.95, $\alpha = 0.05$). Eight of the *de novo* mutations were present in a heterozygous state on the autosomes and one was present in a hemizygous state on the X chromosome. All *de novo* mutations occurred in different genes, including two genes recently implicated in mental retardation (**Table 2**). In addition to using a dominant disease model, we also analyzed the data for recessive forms of mental retardation. In the affected male of trio 10, we identified a maternally inherited non-synonymous variant in *JARID1C* (**Table 2**), which is a well-described X-linked mental retardation gene[12]. Subsequent analysis of this variant in DNA obtained from the affected individual's grandparents indicated that the mutation had occurred *de novo* in the mother of this proband. No conclusive evidence for autosomal recessive inheritance, either homozygous or compound heterozygous, was obtained for the other affected individuals.

Next, we evaluated the function of each mutated gene in relation to the disorder (**Table 2**). Three genes do not seem to play a role in biological pathways linked to mental retardation. *BPIL3* is involved in the innate immune response[13], whereas *PGA5* is involved in protease activity in the stomach[14]. The function of *ZNF599* is currently unknown. For the six other genes affected by *de novo* mutations, functional evidence suggests a role in mental retardation. Two mutations occurred in genes (*RAB39B* and *SYNGAP1*) that, when disrupted, are known to cause mental retardation (**Table 2**)[15,16]. For the remaining four mutated genes, evidence for a causal link with mental retardation is provided by model organisms and protein-protein interaction studies. *DYNC1H1* encodes a cytoplasmic dynein that acts as a motor for intracellular retrograde axonal transport. Heterozygous *Dync1h1*$^{+/-}$ mutant mice exhibit sensory neuropathy[17], and studies in zebrafish have shown the importance of *dync1h1* in correct nuclear positioning. Mislocalization of nuclei in the vertebrate central nervous system is likely to result in profound patterning defects and severely compromised function[18]. Notably, *DYNC1H1* interacts with *PAFAH1B1*, the gene associated with type I lissencephaly, which involves gross disorganization of the neurons within the cerebral cortex[19]. *YY1* encodes the ubiquitously expressed transcription factor yin-yang 1 and directs histone deacetylases and histone acetyltransferases, implicating chromatin remodeling as its main function. Complete ablation of *Yy1* in mice results in early embryonic lethality, whereas *Yy1* heterozygous mice display growth retardation, neurulation defects and brain abnormalities[20]. Recent studies show that YY1 interacts directly with MECP2; *MECP2* is mutated in Rett syndrome[21]. *DEAF1* encodes a transcription factor that regulates the 5-HT1A receptor in the human brain. Mutations in the *Drosophila DEAF1* ortholog result in early embryonic arrest, suggesting an essential role for the gene in early development[22]. Additional evidence is provided by *Deaf1*-deficient mice, which show neural tube defects including exencephaly[23]. Finally, *CIC* is a member of the HMG-box transcription factor superfamily, which is associated with neuronal and glial development of the nervous system. *CIC* is predominantly and transiently expressed in immature granule cells of the cerebellum, hippocampus and neocortex, suggesting a critical role in central nervous system development[24].

We next examined the evolutionary conservation of affected nucleotides (using the phyloP score), as well as the potential of the *de novo* mutations to affect the structure or function of the resulting proteins (using the Grantham score; **Table 2**). All *de novo* missense mutations and the inherited X-linked mutation were included in this analysis; no Grantham scores were available for the additional nonsense and frameshift mutations. Of note, *de novo* mutations in genes with a functional link to mental retardation showed a higher phyloP (mean, 4.7) and Grantham score (mean, 135) than mutations in genes without such a functional indication (mean phyloP score, −0.5 and mean Grantham score, 38). We also compared these scores to those for all non-synonymous variants in the dbSNP database as well as those in the Human Gene Mutation Database (HGMD). The distribution of phyloP scores and Grantham scores differed markedly between dbSNP and the HGMD (Online Methods and **Supplementary Fig. 5**). The four mutations in genes functionally linked to mental retardation all showed higher probability values for being observed in HGMD

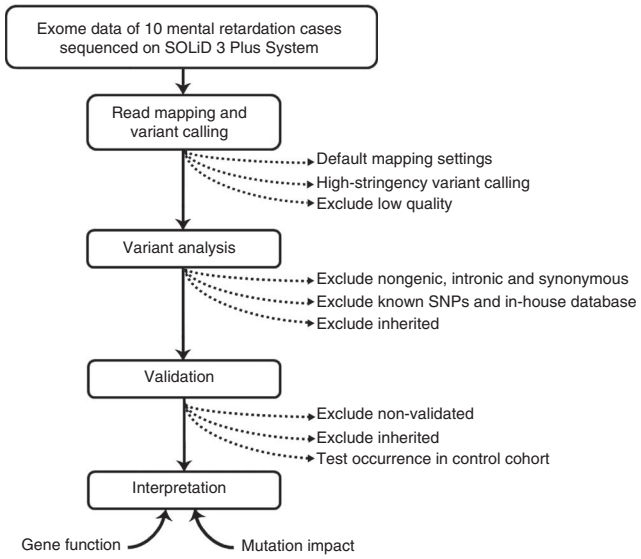


Figure 1 Experimental work flow for detecting and prioritizing sequence variants. For all ten mental retardation trios, prioritization of variants observed in the probands was based on selection for non-synonymous changes of high quality only and exclusion of all variants previously observed in healthy individuals, together with those variants that were inherited from an unaffected parent. Interpretation of *de novo* variants was based on gene function and the impact of the mutation.

Table 2 Overview of all *de novo* variants identified by exome sequencing in ten individuals with unexplained mental retardation

Gene	Trio	Sex[a]	NM number	cDNA level change	Protein level change	PhyloP score	Grantham score	Probability of being observed in dbSNP[b]	Probability of being observed in HGMD[b]	Gene function
***De novo* mutations**										
DYNC1H1	1	M	NM_001376	c.11465A>C	p.His3822Pro	5.5	77	0.20	0.80	Retrograde axonal transporter; interacts with *PAFAH1B1* (mutation of which causes lissencephaly, a neurodevelopmental disorder)
ZNF599	1	M	NM_001007248	c.532C>T	p.Leu187Phe	−1.5	22	1.00	2.65×10^{-4}	Unknown
RAB39B	2	M	NM_171998	c.557G>A	p.Trp186X	4.8	–	–	–	Known X-linked mental retardation gene
YY1	3	M	NM_003403	c.1138G>T	p.Asp380Tyr	6.9	160	2.27×10^{-6}	1.00	Ubiquitously expressed transcription factor; mouse knockdown results in growth retardation, neurulation defects and brain abnormalities; interacts with *MECP2*, a known mental retardation gene
BPIL3	3	M	NM_174897	c.887G>A	p.Arg269His	0.5	29	0.97	0.03	Innate immune response
PGA5	4	F	NM_014224	c.1058T>C	p.Val353Ala	0.7	64	0.84	0.16	Precursor of pepsin
DEAF1	5	M	NM_021008	c.683T>G	p.Ile228Ser	4.9	142	0.01	0.99	Transcription factor; regulator of 5-HT1A receptor in the brain; mouse knockout causes neural tube defects
CIC	6	M	NM_015125	c.1474C>T	p.Arg492Trp	2.6	101	0.46	0.54	Granule cell development in central nervous system
SYNGAP1	8	F	NM_006772	c.998_999del	p.Val333AlafsX	3.3	–	–	–	Known autosomal dominant mental retardation gene
X-linked inherited mutations										
JARID1C	10	M	NM_001146702	c.1919G>A	p.Cys640Tyr	5.1	194	2.09×10^{-6}	1.00	Known X-linked mental retardation gene

[a]Sex of proband, with M for male and F for female. [b]Visual representation of probabilities are provided in **Supplementary Figure 5**. Grantham scores for nonsense (in *RAB39B*) and frameshift mutations (in *SYNGAP1*) could not be calculated.

(mean, 0.83) than for being observed in dbSNP (mean, 0.17). The three mutations in genes without a functional link to mental retardation showed an average probability of 0.94 for being observed in dbSNP and an average probability of 0.06 for being observed in HGMD (**Table 2**). Additionally, the inherited *JARID1C* mutation showed a probability of 1.00 for being in HGMD versus 2.09×10^{-6} for being in dbSNP.

This analysis of the mutated nucleotides and their impact on gene function strongly supports pathogenicity for six of the nine *de novo* mutations. Importantly, these six mutations occurred in genes with a functional link to mental retardation, two of which are known mental retardation genes. In contrast, three *de novo* variants in genes without a functional link did not appear to significantly affect protein function. Moreover, we identified a maternally inherited mutation in a known X-linked mental retardation gene that arose *de novo* in the proband's mother. Although we have not provided individual functional tests to prove causality, these data collectively provide strong evidence for a major role of *de novo* mutations in mental retardation. The identification of recurrent mutations in these genes in unrelated cases would provide additional proof for disease causality, but this may require the evaluation of thousands of affected individuals. The identification of subtle CNVs encompassing (part of) these genes may also provide additional proof for disease causality, as was shown recently for mutations in X-linked mental retardation genes[25]. As of yet, no such CNVs have been reported, nor have we found such CNVs in our diagnostic cohort of ~4,500 individuals with mental retardation (data not shown).

The discovery of nine *de novo* non-synonymous mutations in this cohort of ten affected individuals is concordant with the recently estimated background mutation rate of 0.86 amino-acid–altering mutations per newborn in controls[2], but it will be important to compare this result to similar data from healthy control trios when available. Notably, after applying the same systematic filtering approach and Sanger sequencing, we could only validate a single *de novo* synonymous mutation, which occurred in *GRIN1* (c.351C>T, seen in trio 10). This base pair is not conserved through evolution (phyloP score = −3.2) and does not seem to alter splicing, suggesting that this mutation is an unlikely candidate for causing mental retardation. Of note, the individual carrying this mutation also carries the *JARID1C* mutation. The observed ratio of non-synonymous to synonymous *de novo* mutations is far greater than would be expected for protein-coding genes under purifying selection and indicates that many of these mutations will result in a reproductive disadvantage. In contrast, the average non-synonymous to synonymous ratio reported in dbSNP for the six genes with predicted pathogenic mutations is significantly lower than that of the three genes with mutations reflecting the background mutation rate (Fisher's Exact test, $P = 0.0016$), which is to be expected for disease genes in the normal population.

In summary, our results suggest that *de novo* mutations are a major cause of unexplained mental retardation. These mutations can readily be identified using a family based exome sequencing approach and require only limited follow-up by Sanger sequencing. Our findings have implications for preventive and diagnostic strategies in mental retardation. Systematic genome-wide resequencing in parent-child trios may uncover further examples of this *de novo* paradigm for other human neurodevelopmental disorders.

URLs. 1000 Genomes Project, http://www.1000genomes.org; dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/; HGMD, http://www.hgmd.cf.ac.uk/ac/index.php; R, http://www.r-project.org/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. The genomic reference sequence for *DYNC1H1* can be found under the GenBank accession number NM_001376; for *ZNF599* under NM_001007248; for *RAB39B* under NM_171998; for *YY1* under NM_003403; for *BPIL3* under NM_174897; for *PGA5* under NM_014224; for *DEAF1* under NM_021008; for *CIC* under NM_015125; for *SYNGAP1* under NM_006772; for *JARID1C* under NM_001146702; and for *GRIN1* under NM_021569.2.

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

ACKNOWLEDGMENTS

We thank R. de Reuver and J. Hehir-Kwa for bioinformatics support in data analysis and personnel from the Sequencing Facility of our department for timely completion of Sanger sequencing of validation experiments. This work was funded in part by grants from The Netherlands Organization for Health Research and Development (ZonMW grants 916-86-016 to L.E.L.M.V., 917-66-36 and 911-08-025 to J.A.V. and 917-86-319 to B.B.A.d.V.), the EU-funded TECHGENE project (Health-F5-2009-223143 to J.d.L. and J.A.V.) and the AnEUploidy project (LSHG-CT-2006-37627 to A.H., B.W.M.v.B., H.G.B., B.B.A.d.V. and J.A.V.).

AUTHOR CONTRIBUTIONS

J.A.V., L.E.L.M.V. and H.G.B. conceived the project and planned the experiments. B.B.A.d.V. and B.W.M.v.B. performed sample collection and reviewed phenotypes. L.E.L.M.V., A.H., I.J., M.S., P.d.V., B.v.L. and P.A. performed next-generation sequencing experiments using a custom pipeline set up by C.G. and A.H. J.d.L. and C.G. analyzed and interpreted the data with support from N.W. and M.d.R. L.E.L.M.V., P.d.V., I.J. and M.S. performed validation experiments. L.E.L.M.V., J.d.L. and J.A.V. prepared the first draft of the manuscript. All authors contributed to the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Roach, J.C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636–639 (2010).
2. Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. USA* **107**, 961–968 (2010).
3. Keller, M.C. & Miller, G. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav. Brain Sci.* **29**, 385–404 (2006).
4. Uher, R. The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Mol. Psychiatry* **14**, 1072–1082 (2009).
5. Cook, E.H. Jr. & Scherer, S.W. Copy-number variations associated with neuropsychiatric conditions. *Nature* **455**, 919–923 (2008).
6. de Vries, B.B. *et al.* Diagnostic genome profiling in mental retardation. *Am. J. Hum. Genet.* **77**, 606–616 (2005).
7. Ng, S.B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010).
8. Lupski, J.R. *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* **362**, 1181–1191 (2010).
9. Hoischen, A. *et al. De novo* mutations of *SETBP1* cause Schinzel-Giedion syndrome. *Nat. Genet.* **42**, 483–485 (2010).
10. Sobreira, N.L. *et al.* Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. *PLoS Genet.* **6**, e1000991 (2010).
11. Tarpey, P.S. *et al.* A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat. Genet.* **41**, 535–543 (2010).
12. Jensen, L.R. *et al.* Mutations in the *JARID1C* gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am. J. Hum. Genet.* **76**, 227–236 (2005).
13. Mulero, J.J. *et al.* Three new human members of the lipid transfer/lipopolysaccharide binding protein family (LT/LBP). *Immunogenetics* **54**, 293–300 (2002).
14. Taggart, R.T. *et al.* Relationships between the human pepsinogen DNA and protein polymorphisms. *Am. J. Hum. Genet.* **38**, 848–854 (1986).
15. Giannandrea, M. *et al.* Mutations in the small GTPase gene *RAB39B* are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. *Am. J. Hum. Genet.* **86**, 185–195 (2010).
16. Hamdan, F.F. *et al.* Mutations in *SYNGAP1* in autosomal nonsyndromic mental retardation. *N. Engl. J. Med.* **360**, 599–605 (2009).
17. Chen, X.J. *et al.* Proprioceptive sensory neuropathy in mice with a mutation in the cytoplasmic Dynein heavy chain 1 gene. *J. Neurosci.* **27**, 14515–14524 (2007).
18. Tsujikawa, M., Omori, Y., Biyanwila, J. & Malicki, J. Mechanism of positioning the cell nucleus in vertebrate photoreceptors. *Proc. Natl. Acad. Sci. USA* **104**, 14819–14824 (2007).
19. Tai, C.Y., Dujardin, D.L., Faulkner, N.E. & Vallee, R.B. Role of dynein, dynactin, and CLIP-170 interactions in LIS1 kinetochore function. *J. Cell Biol.* **156**, 959–968 (2002).
20. He, Y. & Casaccia-Bonnefil, P. The Yin and Yang of YY1 in the nervous system. *J. Neurochem.* **106**, 1493–1502 (2008).
21. Forlani, G. *et al.* The MeCP2/YY1 interaction regulates ANT1 expression at 4q35: novel hints for Rett syndrome pathogenesis. *Hum. Mol. Genet.* **19**, 3114–3123 (2010).
22. Veraksa, A., Kennison, J. & McGinnis, W. DEAF-1 function is essential for the early embryonic development of *Drosophila*. *Genesis* **33**, 67–76 (2002).
23. Hahm, K. *et al.* Defective neural tube closure and anteroposterior patterning in mice lacking the LIM protein LMO4 or its interacting partner Deaf-1. *Mol. Cell. Biol.* **24**, 2074–2082 (2004).
24. Lee, C.J. *et al.* CIC, a member of a novel subfamily of the HMG-box superfamily, is transiently expressed in developing granule neurons. *Brain Res. Mol. Brain Res.* **106**, 151–156 (2002).
25. Whibley, A.C. *et al.* Fine-scale survey of X chromosome copy number variants and indels underlying intellectual disability. *Am. J. Hum. Genet.* **87**, 173–188 (2010).

ONLINE METHODS

Subjects. Ten individuals with unexplained moderate to severe mental retardation (with normal karyotypes and genomic profiles obtained using 250K SNP arrays) were selected for exome sequencing (**Supplementary Note**). Family history for mental retardation was negative for all cases. Nongenic causes for mental retardation, including pre-, peri- and post-natal infection and perinatal injury, were excluded. DNA was obtained from peripheral blood from the ten probands as well as from their unaffected parents. DNA isolation was performed using QIAamp DNA Mini Kit (QIAGEN), according to the instructions of the manufacturer. This study was approved by the Medical Ethics Committee of the Radboud University Nijmegen Medical Centre, and all participants signed written informed consent.

Library generation. Exome enrichment required 3 μg of genomic DNA, and an AB SOLiD Optimized SureSelect Human Exome Kit (Agilent) was used for enrichment, containing the exonic sequences of ~18,000 genes and covering a total of ~37 Mb of genomic sequence, as specified by the company. We followed the manufacturer's instructions (version 1.5) for enrichment with a minor modification, which was the reduction of the number of post-hybridization ligation-mediated PCR cycles from 12 cycles to 9 cycles.

SOLiD sequencing. The enriched exome libraries were subsequently used for emulsion PCRs, following the manufacturer's instructions (Life Technologies), based on a library concentration of 1 picomolar (pM) (version March 2010). For each sample, one-quarter of a sequencing slide (Life Technologies) was used on a SOLiD 3 Plus System.

Mapping of variants. Color space reads were mapped to the hg18 reference genome with the SOLiD bioscope software v1.2, which utilizes an iterative mapping approach. Single-nucleotide variants were subsequently called by the diBayes algorithm[26] using high stringency settings, requiring calls on each strand. Small insertions and deletions were detected using the SOLiD Small Indel Tool. We assumed a binomial distribution with a probability of 0.5 of sequencing the variant allele at a heterozygous position. Under this assumption, at least ten reads are required to obtain a 99% probability that at least two reads contain the variant allele. Variants and indels were selected using strict quality control settings, which included the presence of at least four unique variant reads (that is, having different start sites), as well as the variant being present in at least 15% of all reads. All called variants and indels were combined and annotated using a custom analysis pipeline (resulting in HCDiff files for each individual).

Custom bioinformatic analysis pipeline. All variants reported in the HCDiff files were filtered to ensure an optimal prioritization process. For this, we first excluded all nongenic, intronic (other than canonical splice sites) and synonymous variants, reducing the number of variants to an average of 5,640 per individual. Second, all known variants were excluded by comparison with data from dbSNP v130 as well as from our in-house variant database. At the time of this study, this in-house database contained variants from (i) 78 in-house performed 'exomes', contributing 515,480 variants, and (ii) the 1000 Genomes Project (see URLs) and published data from various other studies[27–29], contributing 3,059,835 variants, thereby bringing the number of variants in the in-house database to 3,525,278. Of note, if the variant observed in the proband occurred at a genomic position known in dbSNP v130, but the change present was different in the two (for example, A/C in dbSNP but A/T in the proband), the variant was not excluded from analysis. The filtering step using this data further reduced the average number of variants to 143 per proband.

Next, for a dominant model of disease, we used the exome data from accompanying parents to exclude all inherited variants. This step further reduced the number of potential *de novo* variants to an average of 33 per proband. As not all variants identified in the exomes of the probands may have been sequenced at sufficient coverage in the parental samples, we checked all remaining variants in the exome data from the accompanying parents. In brief, even if only a single read showed the variant allele in one of the parental exome samples, the variant was excluded for validation in the proband. Simultaneously, we checked all remaining potential *de novo* indels for annotation differences in each child-parent trio and excluded those that were found to be identical variants in both parent and child. After this final check, an average of five potential *de novo* variants per proband remained for further validation.

To evaluate the presence of recessive mutations, variant filtering was essentially performed as described above, with the main difference being that uniquely inherited parental variants were not excluded here. The remaining variants were evaluated for the presence of compound heterozygous variants, as well as variants that were present in >80% of all reads. Subsequently, parental exome data were used for segregation analysis of the variants identified.

dbSNP and HGMD. To explore the pathogenicity of our *de novo* variants, the genomic evolutionary conservation score (phyloP) and the amino-acid change (Grantham) were compared to those scores present in dbSNP (build 130) and the HGMD (see URLs). All non-synonymous changes reported in dbSNP and HGMD were retrieved, and overlap between databases was removed from both datasets. In addition, non-synonymous variants in dbSNP with an OMIM disease entry, suggestive for a Mendelian phenotype, were omitted from the dbSNP dataset.

Next, quadratic discriminant analysis[30] was performed on these two datasets to determine the significance of the phyloP and Grantham scores as discriminating factors. Statistical tests were performed using the R statistics package (see URLs). The assumption of normality in the data required for the model was determined using Lilliefors (Kolmogorov-Smirnov) normality testing[31]:

PhyloP $D = 0.0626$, $P < 2.2 \times 10^{-16}$; Grantham $D = 0.0828$, $P < 2.2 \times 10^{-16}$; PhyloP × Grantham $D = 0.1395$, $P < 2.2 \times 10^{-16}$. D represents the maximum absolute difference between the empirical and hypothetical cumulative distribution function.

The combination of both scores together yielded the highest power to discriminate the two datasets, and as such, the combined value was used to calculate probabilities for our *de novo* variants to be observed in either database.

Validation experiments. Validation and *de novo* testing for candidate *de novo* mutations was performed using standard Sanger sequencing approaches. Primers were designed to surround the candidate mutation, and PCR reactions were performed using RedTaq Readymix PCR reaction mix (Sigma-Aldrich). Primer sequences and PCR conditions are available upon request. For all *de novo* mutations identified, an additional control cohort of 75 ethnically matched controls was tested for the presence of the same mutation by Sanger sequencing. Together with the results from 679 control individuals from the 1000 Genomes Project as well as the 78 'exomes' present in our in-house database, the control cohort for the *de novo* mutations encompassed 1,664 control chromosomes.

26. Marth, G.T. *et al.* A general approach to single-nucleotide polymorphism discovery. *Nat. Genet.* **23**, 452–456 (1999).
27. Ng, S.B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
28. Pushkarev, D., Neff, N.F. & Quake, S.R. Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* **27**, 847–852 (2009).
29. Wang, J. *et al.* The diploid genome sequence of an Asian individual. *Nature* **456**, 60–65 (2008).
30. Venables, W.N. & Ripley, B.D. *Modern Applied Statistics with S* (Springer, 4th edn., New York, New York, USA, 2002).
31. Lilliefors, H. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* **62**, 399–402 (1967).

nature genetics

De novo mutations of *SETBP1* cause Schinzel-Giedion syndrome

Alexander Hoischen[1,14], Bregje W M van Bon[1,14], Christian Gilissen[1,14], Peer Arts[1], Bart van Lier[1], Marloes Steehouwer[1], Petra de Vries[1], Rick de Reuver[1], Nienke Wieskamp[1], Geert Mortier[2], Koen Devriendt[3], Marta Z Amorim[4], Nicole Revencu[5], Alexa Kidd[6], Mafalda Barbosa[7], Anne Turner[8], Janine Smith[9], Christina Oley[10], Alex Henderson[11], Ian M Hayes[12], Elizabeth M Thompson[13], Han G Brunner[1], Bert B A de Vries[1] & Joris A Veltman[1]

Schinzel-Giedion syndrome is characterized by severe mental retardation, distinctive facial features and multiple congenital malformations; most affected individuals die before the age of ten. We sequenced the exomes of four affected individuals (cases) and found heterozygous *de novo* variants in *SETBP1* in all four. We also identified *SETBP1* mutations in eight additional cases using Sanger sequencing. All mutations clustered to a highly conserved 11-bp exonic region, suggesting a dominant-negative or gain-of-function effect.

Schinzel-Giedion syndrome (MIM%269150) is a highly recognizable syndrome (**Fig. 1a**) characterized by severe mental retardation, distinctive facial features, multiple congenital malformations (including skeletal abnormalities, genitourinary and renal malformations, and cardiac defects) and a higher-than-normal prevalence of tumors, notably neuroepithelial neoplasia[1,2]. In almost all subjects, the disease phenotype occurs sporadically, suggesting heterozygous *de novo* mutations in a single gene as the underlying mechanism. Rare recurrences of this syndrome may be due to gonadal mosaicism. Traditional disease-gene identification approaches have so far failed to identify the gene associated with this disease or those responsible for the majority of this class of rare sporadic disorder. Microarray-based copy number variation screening has been successful for a number of disorders[3], but this method may fail unless the underlying disease mechanism is haploinsufficiency. Recently, whole-exome sequencing was shown to be effective for disease-gene identification[4] and was successfully used to determine the genetic basis of Miller syndrome, a recessive Mendelian disorder[5].

We sequenced the exomes (37 Mb of genomic sequence, targeting ~18,000 genes) of four unrelated individuals with Schinzel-Giedion syndrome to a mean coverage of 43-fold (**Supplementary Table 1**, **Supplementary Figs. 1** and **2**). The exomes of all four individuals were enriched using the SureSelect human exome kit (Agilent) and were subsequently sequenced using one quarter of a SOLiD sequencing slide (Life Technologies). A total of 2.7–3.0 gigabases of mappable sequence data were generated per individual, with 65–72% of bases mapping to the targeted exome (**Supplementary Table 1**). On average, 85% of the exome was covered at least tenfold, and 21,800 genetic variants were identified per individual, including 5,351 nonsynonymous changes. A number of prioritization steps were applied to reduce this number and to identify the potentially pathogenic mutations, similar to the methods used in previous studies[4,5] (**Supplementary Table 2**). A comparison with the NCBI dbSNP build 130 as well as with recently released SNP data from other groups and in-house SNP data (see **Supplementary Note**) showed that >95% of all variants investigated here were previously reported SNPs and cannot explain a genetically dominant disease. We focused on the 12 genes for which all four individuals studied carried variants and found that only two genes showed variants at different genomic positions, strengthening the likelihood that these variants are causative and not simply unidentified SNPs. One of these two candidate genes, *CTBP2,* was excluded from further analysis because it contained numerous variants found during different in-house exome sequencing experiments (data not shown), which may be due to highly homologous sequences from other genomic loci.

The second candidate was *SETBP1*, which encodes SET binding protein 1. Validation of all four variants in this gene by Sanger sequencing confirmed that these variants were indeed present in a heterozygous state in all four affected individuals (**Supplementary Fig. 3**). Moreover, we tested the DNA of the parents of the affected individuals, which showed that all mutations occurred *de novo*. Using Sanger sequencing, we also identified *SETBP1* mutations in eight out of nine additional individuals with a clinical diagnosis of Schinzel-Giedion syndrome. In total, all 13 affected individuals fulfilled previously suggested diagnostic criteria[2] (**Table 1** and **Supplementary Table 3**); all are of European descent, living in various regions: Europe (n = 7), New Zealand (n = 3), Australia (n = 2) and the United States (n = 1). For six of the eight follow-up cases, parental DNA was available, and the mutations present in the affected individuals were again shown to have occurred

[1]Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [2]Centre for Medical Genetics, Antwerp University Hospital, Antwerp, Belgium. [3]Centre for Human Genetics, Leuven University Hospital, Leuven, Belgium. [4]Serviço de Genética de Coimbra, Hospital Pediátrico de Coimbra, Coimbra, Portugal. [5]Centre for Human Genetics, Cliniques Universitaires St. Luc, Université Catholique de Louvain, Brussels, Belgium. [6]Canterbury Health Laboratories, Christchurch Hospital, Christchurch, New Zealand. [7]Centro de Genética Médica Doutor Jacinto Magalhães, Instituto Nacional de Saúde Doutor Ricardo, Porto, Portugal. [8]Department of Medical Genetics, Sydney Children's Hospital, Sydney, Australia. [9]Department of Clinical Genetics, The Children's Hospital at Westmead, Sydney, Australia. [10]Clinical Genetics Unit, Birmingham Women's Hospital, Birmingham, UK. [11]Institute of Human Genetics, Newcastle upon Tyne Hospitals National Health Service Foundation Trust, Newcastle upon Tyne, UK. [12]Northern Regional Genetics Service, Auckland, New Zealand. [13]South Australian Clinical Genetics Service, South Australian Pathology, Women's and Children's Hospital, North Adelaide, South Australia, Australia. [14]These authors contributed equally to this work. Correspondence should be addressed to J.A.V. (j.veltman@antrg.umcn.nl).

Received 9 February; accepted 8 April; published online 2 May 2010; doi:10.1038/ng.581

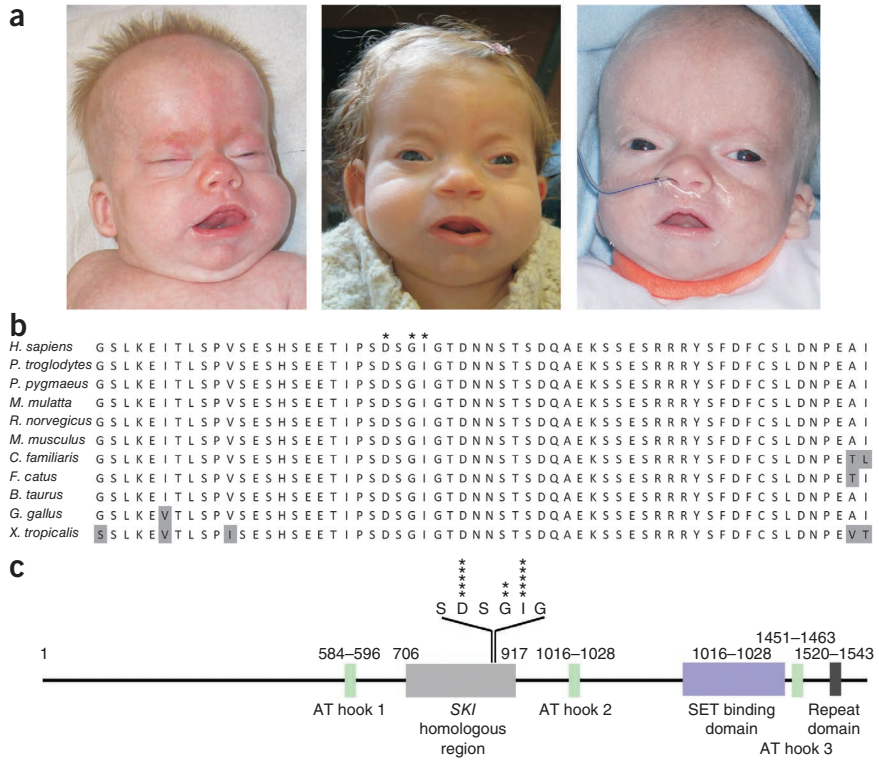


Figure 1 Identification of *SETBP1* mutations in individuals with Schinzel-Giedion syndrome. Photos showing individuals 1, 3 and 4 (from left to right) (**a**). Note the characteristic facial features, prominent forehead, bi-temporal narrowing, mid-face retraction, hypertelorism, deep groove under the eyes, short upturned nose and low-set abnormal ears. We obtained written consent to publish photographs of all individuals shown. Evolutionary conservation of the SETBP1 region that harbors amino acid residues affected in Schinzel-Giedion syndrome (marked by asterisks above human protein sequence) (**b**). All affected residues (Asp868, Gly870 and Ile871) are completely conserved throughout evolution. Residues identical in all sequences are shown as black on a white background; different amino acids are shown as black on a gray background. Schematic overview of the SETBP1 protein (**c**). Known and predicted protein domains are shown in relation to the mutation cluster observed in 12 individuals with Schinzel-Giedion syndrome. Affected amino acid positions are marked with asterisks; each asterisk represents a single case.

de novo. Parentage was confirmed by simple tandem repeat marker analysis in all available case-parent trios. None of the *SETBP1* mutations were found in the 188 control chromosomes. The presence of a second mutation in *SETBP1* was excluded by sequencing all of the coding exons of this gene in the cases. The altered base pairs seen here are located among the most highly conserved positions in the human genome (having phyloP 44 species conservation scores of up to 7.129). This is in agreement with our earlier observation that base pair conservation can be used for variant prioritization[6]. Notably, all changes occurred in a genomic stretch of only 11 nucleotides (chromosome position 40,789,905–40,789,915 at nucleotides 2,602–2,612), affecting three of four consecutive amino acids (868–871: aspartate, serine, glycine and isoleucine). Two mutations were recurrent: G2602A and T2612C, present in four and five cases, respectively.

There are two indications that the mutations seen here may either result in a gain-of-function effect or have a dominant-negative effect. First, a gain-of-function mechanism has been described for other syndromes, such as Noonan syndrome[7], that have very similar mutation clustering in *PTPN11* and *SOS1*; the *FGFR3* mutation clustering in achondroplasia also results in a gain-of-function effect. Second, the phenotype of individuals with partial chromosome 18q deletions which affect *SETBP1* (ref. 8) does not resemble Schinzel-Giedion syndrome.

SETBP1 shows ubiquitous expression[9], which is consistent with the multisystemic defects in individuals with Schinzel-Giedion syndrome. Relatively little is known about the function of SETBP1 except that it binds to SET domains. This domain is named after three *Drosophila* proteins involved in epigenetic processes, Su(var), E(z) and trithorax[10], and is known to be involved in the methylation of lysine residues on histone tails. We note, however, that all observed mutations are located outside of the SET interacting domain and that they do not affect the DNA binding domains. The protein sequence in which the mutations cluster is highly conserved throughout evolution (**Fig. 1b**). A glycosaminoglycan attachment site was predicted at this position in the SETBP1 protein (amino acids 815–818: serine, glycine, isoleucine and glycine)[11]. For one type of glycosaminoglycan, the keratan sulfates, an important role in bone development has been described[12]. Of note, skeletal malformations are one of the major features of Schinzel-Giedion syndrome. Notably, the mutation cluster examined here overlaps with a region that shows homology to the oncogene *SKI* (located at amino acids 706–917; **Fig. 1c**). It has been speculated that *SETBP1* might be involved in the regulation of Ski-Ski homodimer and/or Ski-SnoN heterodimer formation, which both cause cellular transformation[13]. High levels of *SETBP1* expression have been described in cancer cell lines, and *SETBP1* has been identified in a specific pediatric acute T-cell lymphoblastic leukemia[14] as a chromosomal translocation partner of *NUP98*. Whether this relates to the elevated risk for neuroepithelial neoplasia and sacrococcygeal teratomas in individuals with Schinzel-Giedion syndrome (**Supplementary Table 3**) requires further investigation.

In conclusion, our study represents the first elucidation, to our knowledge, of a dominant Mendelian disorder using exome sequencing, illustrating the potential of this technique for disease-gene identification. It is highly likely that the mutations examined here cause disease through either a dominant-negative or a gain-of-function effect. Exome sequencing is particularly useful for identifying these types of mutations for which no other genome-wide approach is applicable. One limitation of this approach is the current inability of exome sequencing to reliably identify structural genomic variation associated with disease. We therefore recommend combining exome sequencing with microarray-based copy number variation detection. *De novo* mutations and structural variations may be a frequent cause of sporadic conditions with reduced fecundity, such as congenital malformations, mental retardation and psychiatric disorders.

Table 1 Major clinical findings in 13 individuals with Schinzel-Giedion syndrome and summary of mutations identified in exon 4 of *SETBP1*

Individual ID	1	2	3	4	5	6	7	8	9	10	11	12	13
Gender	F	M	F	M	M	F	F	F	M	F	M	M	M
Neurodevelopmental anomalies													
Developmental delay	+	+	+	+	+	+	+	+	+	+	u*	+	u*
Seizures	+	+	+	+	+	+	+	+	+	+	–	+	+
Vision impairment	+	+	+	u	u	+	+	+	+	u	+	–	u
Hearing impairment	+	+	+	u	u	+	–	+	+	u	+	+	+
Craniofacial features													
Large anterior fontanelle	–	+	u	+	+	+	–	+	+	+	+	+	+
Prominent forehead	+	+	+	+	+	+	+	+	+	–	+	+	+
Mid-face retraction	+	+	+	+	+	+	+	+	+	+	+	+	+
Hypertelorism	+	+	+	+	+	+	+	+	+	+	–	+	+
Short, upturned nose	+	+	+	+	+	+	+	+	+	+	+	+	+
Low-set ears	+	+	+	+	+	+	+	+	+	+	+	+	–
Structural anomalies													
Genital	+	+	+	+	+	+	+	+	+	+	+	+	+
Hydronephrosis or vesicoureteral reflux	+	+	+	+	+	+	+	+	+	+	+	+	+
Cardiac defect	–	+	+	–	–	+	+	+	+	–	+	–	–
Characteristic skeletal malformations	u	u	+	+	+	+	+	+	+	+	+	+	+
Choanal stenosis	+	–	+	–	–	+	–	–	–	–	–	+	–
Mutations identified													
Genomic position (bp) (hg18, NCBI36)	40,785,915	40,785,905	40,785,905	40,785,912	40,785,905	40,785,905	40,785,906	40,785,911	40,785,915	40,785,915	40,785,915	40,785,915	none
cDNA position	T2612C	G2602A	G2602A	G2609A	G2602A	G2602A	A2603C	G2608A	T2612C	T2612C	T2612C	T2612C	–
Protein consequences	I871T	D868N	D868N	G870D	D868N	D868N	D868A	G870S	I871T	I871T	I871T	I871T	–
PolyPhen prediction	Possibly damaging	Possibly damaging	Possibly damaging	Possibly damaging	Possibly damaging	Possibly damaging	Probably damaging	Possibly damaging	Possibly damaging	Possibly damaging	Possibly damaging	Possibly damaging	–
SIFT	Not tolerated	Not tolerated	Not tolerated	Not tolerated	Not tolerated	Not tolerated	Not tolerated	Not tolerated	Not tolerated	Not tolerated	Not tolerated	Not tolerated	–
De novo	Yes	Yes	Yes	Yes	Yes	Yes	u	Yes	Yes	u	Yes	Yes	–
phyloP cons. score (44 species)	5.13	7.13	7.13	7.13	7.13	7.13	5.65	7.13	5.13	5.13	5.13	5.13	–

U, unknown; u*, subject died shortly after birth; SIFT (sorting intolerant from tolerant), amino acid substitution prediction method; not tolerated, variants predicted to have a phenotypic effect.
Table adapted from ref. 15.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

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

ACKNOWLEDGMENTS

We thank the subjects and their parents for participation in this study. We thank S. Keijzers-Vloet, J. de Ligt, N. Leijsten and personnel from the Sequencing Facility Nijmegen for technical assistance and I. Saleem for referring patient 3 to us. This study was financially supported by the Netherlands Organization for Health Research and Development (ZonMW grants 917-66-36 and 911-08-025 to J.A.V. and 917-86-319 to B.B.A.d.V.), the EU-funded TECHGENE project (Health-F5-2009-223143 to P.A. and J.A.V.) and the AnEUploidy project (LSHG-CT-2006-37627 to A.Hoischen, B.W.M.v.B., H.G.B., B.B.A.d.V. and J.A.V.).

AUTHOR CONTRIBUTIONS

A. Hoischen, B.W.M.v.B., C.G., H.G.B., B.B.A.d.V. and J.A.V. conceived the project and planned the experiments. B.W.M.v.B., H.G.B. and B.B.A.d.V. performed review of phenotypes and sample collection. G.M., K.D., M.Z.A., N.R., A.K., M.B., A.T., J.S., C.O., A. Henderson, I.M.H. and E.M.T. clinically characterized the Schinzel-Giedion syndrome cases and collected blood samples. A. Hoischen, P.A. and B.v.L. performed next-generation sequencing experiments. M.S. and P.d.V. performed validation experiments. C.G., R.d.R. and N.W. analyzed and interpreted the data. A. Hoischen, B.W.M.v.B., C.G. and J.A.V. prepared the draft manuscript. All authors contributed to the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Schinzel, A. & Giedion, A. *Am. J. Med. Genet.* **1**, 361–375 (1978).
2. Lehman, A.M. *et al. Am. J. Med. Genet. A.* **146A**, 1299–1306 (2008).
3. Vissers, L.E. *et al. Nat. Genet.* **36**, 955–957 (2004).
4. Ng, S.B. *et al. Nature* **461**, 272–276 (2009).
5. Ng, S.B. *et al. Nat. Genet.* **42**, 30–35 (2010).
6. Hoischen, A. *et al. Hum. Mutat.* **31**, 494–499 (2010).
7. Tartaglia, M. & Gelb, B.D. *Annu. Rev. Genomics Hum. Genet.* **6**, 45–68 (2005).
8. Buysse, K. *et al. Am. J. Med. Genet. A.* **146A**, 1330–1334 (2008).
9. Su, A.I. *et al. Proc. Natl. Acad. Sci. USA* **101**, 6062–6067 (2004).
10. Varez-Venegas, R. & Avramova, Z. *Gene* **285**, 25–37 (2002).
11. Krueger, R.C. Jr., Fields, T.A., Hildreth, J. & Schwartz, N.B. *J. Biol. Chem.* **265**, 12075–12087 (1990).
12. Nieduszynski, I.A. *et al. Biochem. J.* **271**, 243–245 (1990).
13. Minakuchi, M. *et al. Eur. J. Biochem.* **268**, 1340–1351 (2001).
14. Panagopoulos, I. *et al. Br. J. Haematol.* **136**, 294–296 (2007).
15. Minn, D. *et al. Am. J. Med. Genet.* **109**, 211–217 (2002).

Exome Sequencing Identifies *WDR35* Variants Involved in Sensenbrenner Syndrome

Christian Gilissen,[1,3] Heleen H. Arts,[1,3] Alexander Hoischen,[1,3] Liesbeth Spruijt,[1] Dorus A. Mans,[1] Peer Arts,[1] Bart van Lier,[1] Marloes Steehouwer,[1] Jeroen van Reeuwijk,[1] Sarina G. Kant,[2] Ronald Roepman,[1] Nine V.A.M. Knoers,[1] Joris A. Veltman,[1] and Han G. Brunner[1,*]

Sensenbrenner syndrome/cranioectodermal dysplasia (CED) is an autosomal-recessive disease that is characterized by craniosynostosis and ectodermal and skeletal abnormalities. We sequenced the exomes of two unrelated CED patients and identified compound heterozygous mutations in *WDR35* as the cause of the disease in each of the two patients independently, showing that it is possible to find the causative gene by sequencing the exome of a single sporadic patient. With RT-PCR, we demonstrate that a splice-site mutation in exon 2 of *WDR35* alters splicing of RNA on the affected allele, introducing a premature stop codon. WDR35 is homologous to TULP4 (from the Tubby superfamily) and has previously been characterized as an intraflagellar transport component, confirming that Sensenbrenner syndrome is a ciliary disorder.

Cranioectodermal dysplasia (CED; MIM 218330), also known as Sensenbrenner syndrome, is an autosomal-recessive disease that is characterized by sagittal craniosynostosis and facial, ectodermal, and skeletal anomalies.[1,2] A proportion of cases have nephronophthisis, hepatic fibrosis, retinitis pigmentosa, and brain anomalies.[3] This phenotype shows remarkable overlap with the ciliopathies, a spectrum of disorders associated with dysfunction of the cilium, a microtubule-based organelle that protrudes from the membrane in many vertebrate cell types.[4] Furthermore, it has recently been shown that defects in the ciliary gene *IFT122* (intraflagellar transport 122; MIM 606045) can be a cause of CED.[5]

Here we report on two unrelated Sensenbrenner patients with remarkably similar phenotypes (Figure 1; Table 1). These cases were previously screened diagnostically by using Affymetrix 250k arrays. Because neither pathogenic copy-number variants nor large homozygous regions were identified, we decided to use a different approach to identify the cause of disease in these two patients, hereafter referred to as patients 1 and 2. The current study was approved by the Medical Ethics Committee of the Radboud University Nijmegen Medical Centre. Written informed consent to participate in the study was obtained for both patients (and all other ciliopathy patients described in this paper), as well as informed consent to publish clinical photos for patients 1 and 2.

We applied a genome-wide approach and sequenced the exomes (targeting ~18,000 genes) of both patients. We obtained 3.6 Gb and 3.4 Gb of mappable sequence data per patient by using a SureSelect human exome kit (Agilent, Santa Clara, CA, USA) in combination with one quarter of a SOLiD sequencing slide (Life Technologies, Carlsbad, CA, USA). Color space reads were mapped to the hg18 reference genome with SOLiD BioScope software version 1.0, which utilizes an iterative mapping approach. In total, 89% and 86% of bases originated from the targeted exome, resulting in a mean coverage of 67- and 59-fold (see Table S1 available online). Single-nucleotide variants were subsequently called by the DiBayes algorithm with a conservative call stringency. The DiBayes SNP caller requires at least two variant reads to call a SNP. We assumed a binomial distribution with probability 0.5 of sequencing the variant allele at a heterozygous position. At least ten reads are then required to obtain a 99% probability of having at least two reads containing the variant allele. More than 89% of the targeted exons were covered more than ten times. Small insertions and deletions were detected by using the SOLiD Small InDel Tool. Called SNP variants and indels were then combined and annotated by using a custom analysis pipeline.

On average, 12,736 genetic variants were identified per patient in the coding regions or the canonical dinucleotide of the splice sites, including 5,657 nonsynonymous changes (Table S2). A prioritization scheme was applied to identify the pathogenic mutation in each patient separately, similar to a recent study.[6] We excluded known dbSNP130 variants as well as variants from our in-house variant database, reducing the number of candidates by more than 98%. The in-house database consists of data from in-house exome resequencing projects of patients with rare syndromes (548,103 variants), the 1000 Genomes Project, and published data from various studies[7–9] (2,535,563 variants). For a recessive disease, it is possible that heterozygous variants found in healthy individuals have been reported as benign polymorphisms within dbSNP or our internal variant database. However, given the rare incidence of CED, it is almost impossible

[1]Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, 6525 GA Nijmegen, The Netherlands; [2]Department of Clinical Genetics, Leiden University Medical Center, 2333 ZC Leiden, The Netherlands
[3]These authors contributed equally to this work
*Correspondence: h.brunner@antrg.umcn.nl
DOI 10.1016/j.ajhg.2010.08.004.

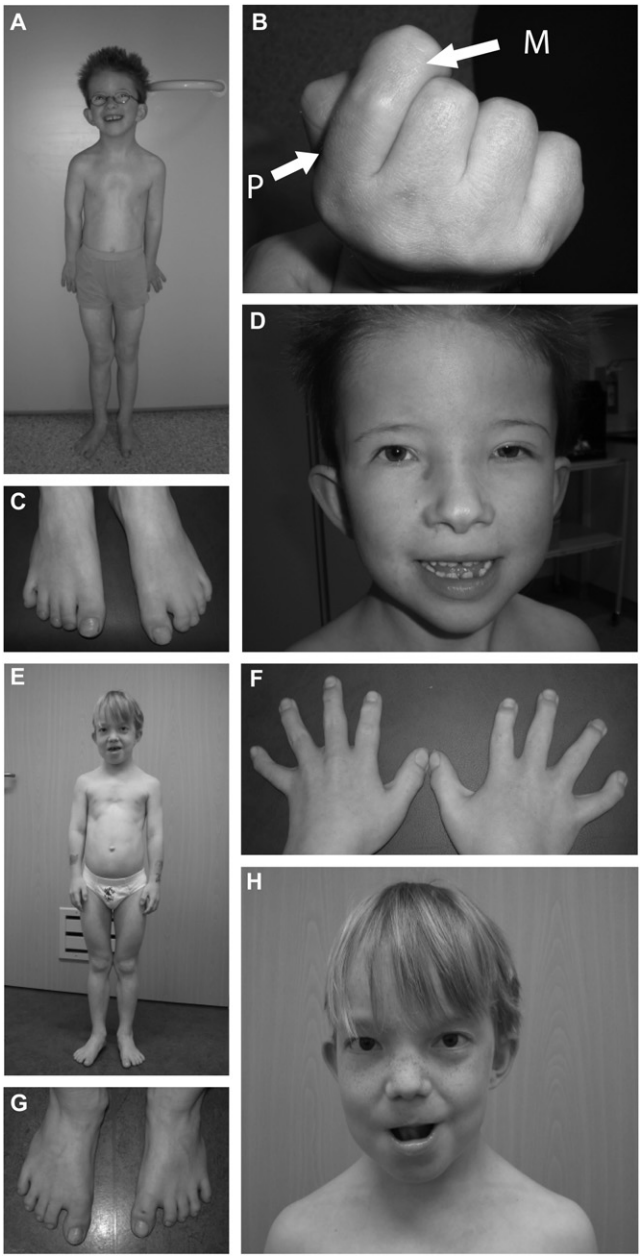


Figure 1. Two Patients with Sensenbrenner Syndrome for Whom Exome Sequencing Was Performed
(A) Patient 1: small thorax, pectus excavatum, rhizomelic shortening of limbs.
(B) Shortening of proximal second phalanx. The proximal phalanx is indicated by "P"; the middle phalanx is indicated by "M."
(C) Syndactyly 2-3 right foot, 2-3-4 left foot.
(D) Hypertelorism, unilateral ptosis of left eye, low-set ears, everted lower lip.
(E) Patient 2: small thorax, pectus excavatum, rhizomelic shortening of limbs.
(F) Short, broad hands.
(G) Bilateral sandal gap between first and second toe.
(H) Hypertelorism, low-set simple ears, thin hair.

that heterozygous mutations occur at a reasonable frequency in the healthy population, and thus it is unlikely that they have been included in dbSNP.

Under the assumption of an autosomal-recessive disease model, we found three candidate genes with compound heterozygous variants in patient 1 (*FLG*, *MFRP*, and *WDR35*). The inheritance of the variants in the three candidate genes was determined by Sanger sequencing, showing that the two *WDR35* and the two *MFRP* variants were inherited from different parents (Table S3). Based on evolutionary conservation score,[10] both variants in *WDR35* ranked at the top position among all variants of the three candidate genes (Table S3). Moreover, *WDR35* was the only candidate with a ciliary function according to the Ciliary Proteome database[11] (cutoff e value 30; 2,127 entries). In patient 2, we identified four candidate genes that harbored two or more variants. Sanger validation excluded two of these as candidate genes because the variants were inherited from a single parent. The remaining candidates, *USH2A* (MIM 608400) and *WDR35*, both had a putative ciliary function. The two conserved variants in *USH2A* were both inherited paternally, whereas a third nonconserved variant was not inherited paternally, which makes *USH2A* an unlikely candidate for CED. Furthermore, this patient had no signs of retinitis pigmentosa (MIM 608000) or Usher syndrome (MIM 276901). The *WDR35* variants were inherited from both parents and affected base pairs with high evolutionary conservation (Table S3). In conclusion, we independently identified *WDR35* as the most likely candidate disease gene in both patients.

In patient 1, we identified a canonical splice-site mutation 2 bp upstream of exon 2 (c.25-2A>G [p.I9TfsX7]) and a missense mutation in exon 17 (c.1877A>G [p.E626G]) (Figure 2; Figure S1). With RT-PCR, we demonstrated that splicing of *WDR35* RNA (derived from Epstein-Barr virus cell lines) was indeed altered in patient 1 compared to an unrelated control individual (Figure 2). Sequencing of the bands revealed that the affected allele contained a 58 bp insertion that introduced a premature stop codon. The missense mutation in exon 17 was predicted to be "probably damaging" by PolyPhen.[12] The mutated amino acid is highly conserved up to insects and nematodes (Figure S2). Because the *C. elegans* protein WDR35 ortholog (IFTA-1) is most distantly related to the human WDR35 protein (only 44.2% similar and 28.4% identical), the conservation of the glutamine provides a strong indication of the importance of this amino acid for WDR35/IFTA-1 function.

In patient 2, a deletion of a C nucleotide in exon 25 predicts a frameshift and a premature stop (1:c.2891 delC [p.P964LfsX15]) (Figure S1). On the second allele, a substitution in exon 23 (c.2623G>A [p.A875T]) leads to the change of a highly conserved alanine to a threonine (Figures S1 and S2). The amino acid substitution was classified as "potentially damaging" by PolyPhen. It is remarkable that the variants in both patients are a combination of a missense and a truncating mutation. In ciliary diseases, phenotypical severity is often determined by the combination of missense and nonsense mutations.[13] This phenomenon could be used in the prioritization of

Table 1. Clinical Details of Two Patients with Cranioectodermal Dysplasia

Patient	1	2
Mutations DNA	c.1877A>G, c.25-2A>G	1:c.2891delT, c.2623G>A
Mutations protein	p.E626G, p.I9TfsX7	p.P964LfsX15, p.A875T
Age at diagnosis	7	9
Height	<2.5 standard deviations	<2.5 standard deviations
Dolichocephaly	+	+
Craniosynostosis	surgically corrected at age 1	surgically corrected at age 1
Frontal bossing	+	+
Macrocephaly	–	–
Sparse, fine hair	–	+
Narrow palpebral fissure	+	+
Telecanthus	+	+
Hypermetropia	+	–
Nystagmus	–	–
Ptosis	unilateral	–
Hypertelorism	+	+
Strabism	+	–
Low-implanted ears	+	+
Simple ears	+	+
Everted lower lip	+	+
Micrognathia	+	–
Widely spaced teeth	+	+
Hypoplastic teeth	+	+
Fused teeth	+	+
Short neck	+	+
Narrow thorax	+	+
Pectus excavatum	+	+
Short limbs	+	+
Brachydactyly	+	+
Webbing of fingers	+	+
Postaxial polydactyly	+	–
Restricted flexion of fingers	+	–
Syndactyly 2-3 (fourth) toe	+	–
Bilateral sandal gap	–	+
Joint laxity	+	+
Inguinal hernia (bilateral)	+	+
Renal disease	–	–
Hepatic disease	–	–
Recurrent lung infections	+	–
Intelligence	normal	normal
Behavior	happy, friendly	happy, friendly

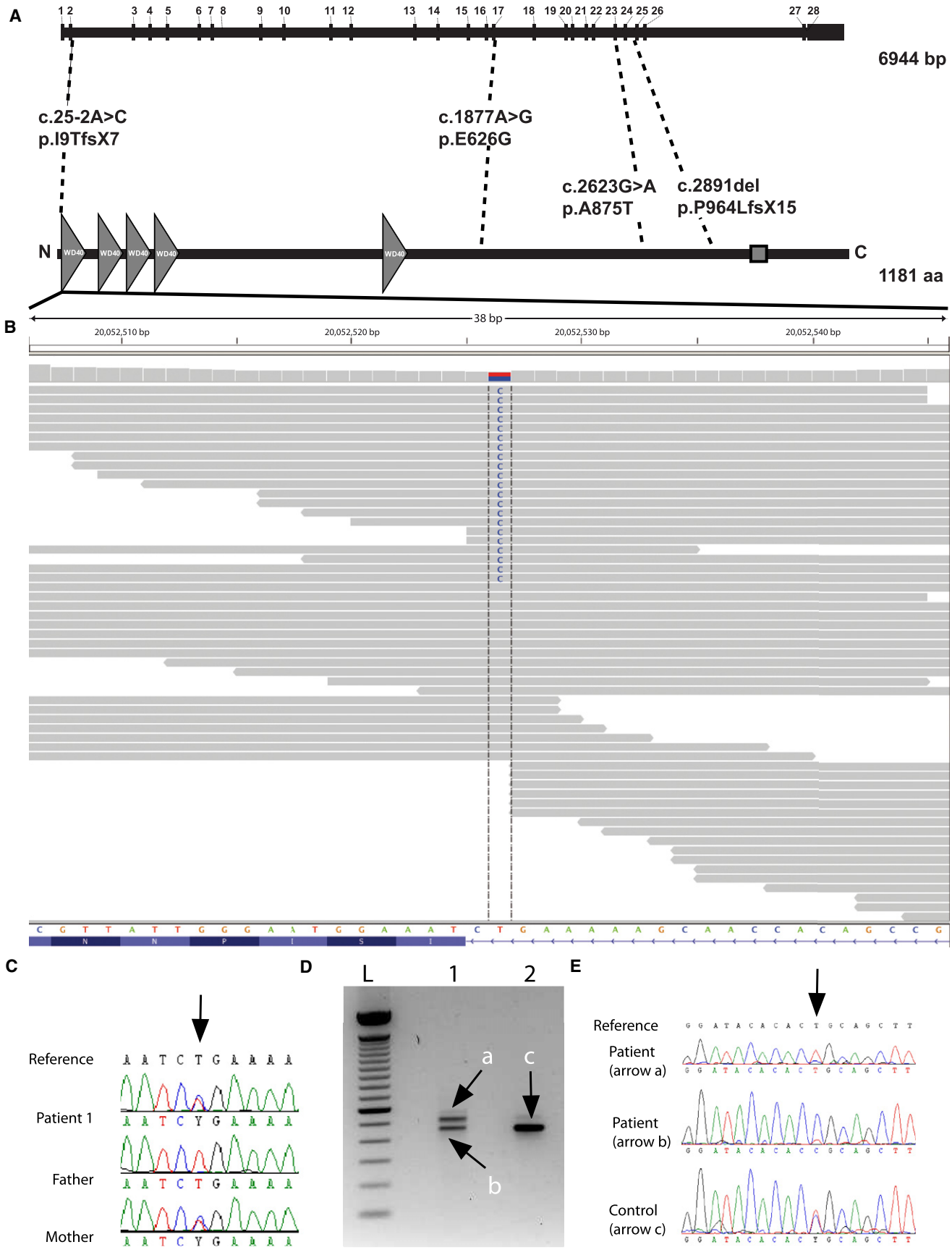
variants from exome sequencing of other (ciliary) diseases, which in this case would have immediately identified *WDR35*.

None of the four identified variants in *WDR35* were detected in 210 control alleles, indicating that the identified variants are uncommon in the Dutch population from which the patients originated, further supporting the thought that the *WDR35* variants are pathogenic. Furthermore, complete loss of *WDR35* function leads to a severe short-rib polydactyly syndrome, another ciliopathy (P.J. Lockhart, personal communication).

WDR35 contains 28 coding exons that encode at least four known protein isoforms (as determined in Ensembl). The WDR35 protein, which is part of the WD-repeat protein family, was first characterized in the green alga *Chlamydomonas reinhardtii*[14]. The *Chlamydomonas* ortholog (IFT121) is part of the intraflagellar transport complex A, together with at least five other proteins. One of them, IFT122/WDR10, also contains N-terminal WD repeats and has recently been shown to be involved in CED as well.[5] Like other IFT-A proteins, WDR35 is instrumental for retrograde IFT (from the ciliary tip to the basal body) in mice.[15] Studies in *Drosophila* and *C. elegans* also have demonstrated that these species' WDR35 orthologs (Oseg4 and IFTA-1, respectively) localize to the cilium and act in IFT.[16] It is thus most likely that cilium dysfunction due to disrupted (retrograde) IFT is underlying the CED phenotype of patients with mutations in *WDR35*.

To evaluate whether mutations in *WDR35* are a common cause of CED, we performed mutation analysis in six additional CED patients. These patients presented with additional clinical phenotypes and did not show the striking phenotypic similarity observed between patients 1 and 2. In three of these patients, mutations in *IFT122* had been excluded, whereas in two other patients, linkage regions were identified that did not contain *IFT122* (or *IFT121*). We did not find any causative mutations in *WDR35* in these patients. Thus, only 25% of our cohort (2 out of 8 CED patients) carried mutations in *WDR35*, further confirming that CED is a genetically heterogeneous disorder similar to other ciliopathies.[5,13]

Our results are consistent with the previously demonstrated importance of IFT in bone development; mutations in *IFT80* (MIM 611177) and *DYNC2H1* (MIM 603297) have been associated with Jeune syndrome (MIM 208500), a disorder with significant clinical overlap with CED.[17,18] Based on this and the fact that many ciliary disease genes are associated with multiple ciliopathy syndromes,[13] we screened *WDR35* for mutations in 15 Jeune syndrome patients. No mutations were found in these patients,


A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
6944 bp
c.25-2A>C
p.I9TfsX7
c.1877A>G
p.E626G
c.2623G>A
p.A875T
c.2891del
p.P964LfsX15
N
WD40
WD40
WD40
WD40
WD40
C
1181 aa
B
38 bp
20,052,510 bp
20,052,520 bp
20,052,530 bp
20,052,540 bp
C
Reference
Patient 1
Father
Mother
D
L
1
2
a
c
b
E
Reference
Patient
(arrow a)
Patient
(arrow b)
Control
(arrow c)

indicating that mutations in *WDR35* are not a major cause of Jeune syndrome. So far, Sensenbrenner syndrome is the only ciliopathy that includes a craniosynostosis phenotype. Because conditional IFT knockout mouse models indicate that Sonic and Indian Hedgehog signaling is regulated by cilia during skeletal development, we hypothesize that disrupted ciliary Hedgehog signaling due to disrupted IFT is involved in the skeletal features, including craniosynostosis, in our CED patients. *Wdr35* knockout mice do indeed display Hedgehog defects during limb development.[15]

By using several bioinformatic tools (i.e., the phylogenetics tree database TreeFam, Gene Tree from Ensembl, and NCBI BLAST), we found that TULP4 (a member of the Tubby superfamily) is homologous to WDR35. Although little is known about the function of TULP4, it is of interest that the Tubby family member Tulp3 is known to modulate Shh signaling during early embryonic development of the mouse,[19] like WDR35[15] and other IFT proteins. Moreover, because the phenotype of *Tulp3* knockout mice as well as other mutants from the Tubby family shares features with the phenotypes of CED and other ciliopathies, we conclude that *TULP4* is an excellent candidate gene for such disorders.[19–23] Together, these findings also suggest that the IFT proteins and some members of the Tubby family of proteins, in particular TULP4, are functionally related.

In our study, the availability of two independent cases with a strikingly similar phenotype was obviously very useful for identifying the causative gene. For both individual patients, we identified only a small number of candidate disease genes. In each of the two cases, this was further reduced to a single candidate (*WDR35*) by segregation analysis and additional evidence on evolutionary conservation and the ciliome database. The finding that *WDR35* is mutated in a family with short-rib polydactyly syndrome (MIM 263510) confirms that *WDR35* is indeed involved in "skeletal" ciliopathies (P.J. Lockhart, personal communication). Interestingly, both patients presented with sagittal craniosynostosis, a birth defect characterized by premature closure of the skull sutures that occurs in 1.5 per 10.000 newborns. Familial recurrences and occasional concordant twins indicate the presence of genetic factors underlying sagittal synostosis, but these remain largely unknown. The involvement of *WDR35* in this phenotype might provide insight into the underlying biology of sagittal craniosynostosis.

It should be noted that theoretically we could have missed additional variants that are relevant for the disease. Only high-quality whole-genome sequencing could have fully excluded disease-related genomic variants or modifiers. However, we believe that the combined genetic and functional data undoubtedly show the involvement of *WDR35* in CED. In summary, our data indicate that for a rare recessive condition, it is possible to find the causative gene by sequencing the exome of a single sporadic patient. This is consistent with the results of a recent study of Perrault syndrome (MIM 233400) that also found the causative gene by sequencing the exome of a single affected individual.[24]

Supplemental Data

Supplemental Data include two figures and five tables and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

We thank the Sensenbrenner families for their participation. We also thank E.M. Bongers and P.L. Beales for supplying the additional CED patients, B.C. Hamel for the Jeune syndrome patients, and N. Wieskamp for bioinformatics support. We thank P. Mill, I. Jackson, and P.J. Lockhart for sharing their unpublished data with us. This research was supported by grants from the Dutch Kidney Foundation (KJPB09.009 to H.H.A.), the Netherlands Organization for Scientific Research (NWO/ZonMw Vidi 91786396 to R.R.), the Netherlands Organization for Health Research and Development (ZonMw 917-66-363 and 911-08-025 to J.A.V.), the European Union-funded TECHGENE project (Health-F5-2009-223143 to P.A. and J.A.V.), SYSCILIA (Health-F5-2010-241955 to R.R.), and the AnEUploidy project (LSHG-CT-2006-37627 to A.H., H.G.B., and J.A.V.).

Received: July 2, 2010
Revised: August 2, 2010
Accepted: August 4, 2010
Published online: September 2, 2010

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes project, http://www.1000genomes.org/
BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi
Ciliaproteome V3.0, http://v3.ciliaproteome.org/cgi-bin/index.php

Figure 2. Splice-Site Mutation in Patient 1
(A) Gene and protein structure of *WDR35*. WD domains are indicated by triangles; the box indicates a low-complexity region.
(B) Sequencing reads showing the heterozygous splice-site mutation at the splice-acceptor site of exon 2 as well as a nearby exonic polymorphism (rs1060742).
(C) Maternal inheritance of the splice-site mutation in patient 1 shown by Sanger sequencing.
(D) Effect of the splice-site mutation on the RNA shows two different products of equal intensity in lane 1 (523 bp and 465 bp for the upper and lower band, respectively). Lane 2 shows the product of an unrelated control (465 bp).
(E) Sequence of the two WDR35 RT-PCR products from Figure 2D, showing that the polymorphism (rs1060742) is predominantly present in the normal spliced product as a C (Figure 2D, lane 1, arrow a), whereas this is a T in the alternatively spliced product (Figure 2D, lane 1, arrow b). The sequence of the RT-PCR product of an unrelated individual with the same heterozygous SNP (rs1060742) is shown as a control (Figure 2D, lane 3, arrow c).

Ensembl Gene Tree, http://www.ensembl.org/Homo_sapiens/Gene/Compara_Tree?g=ENSG00000118965
IGV browser, http://www.broadinstitute.org/igv
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/
TreeFam, http://www.treefam.org/
UCSC Genome Browser, http://genome.ucsc.edu/

References

1. Sensenbrenner, J.A., Dorst, J.P., and Owens, R.P. (1975). New syndrome of skeletal, dental and hair anomalies. Birth Defects Orig. Artic. Ser. *11*, 372–379.
2. Levin, L.S., Perrin, J.C., Ose, L., Dorst, J.P., Miller, J.D., and McKusick, V.A. (1977). A heritable syndrome of craniosynostosis, short thin hair, dental abnormalities, and short limbs: Cranioectodermal dysplasia. J. Pediatr. *90*, 55–61.
3. Amar, M.J., Sutphen, R., and Kousseff, B.G. (1997). Expanded phenotype of cranioectodermal dysplasia (Sensenbrenner syndrome). Am. J. Med. Genet. *70*, 349–352.
4. Baker, K., and Beales, P.L. (2009). Making sense of cilia in disease: The human ciliopathies. Am. J. Med. Genet. C Semin. Med. Genet. *151C*, 281–295.
5. Walczak-Sztulpa, J., Eggenschwiler, J., Osborn, D., Brown, D.A., Emma, F., Klingenberg, C., Hennekam, R.C., Torre, G., Garshasbi, M., Tzschach, A., et al. (2010). Cranioectodermal dysplasia, Sensenbrenner syndrome, is a ciliopathy caused by mutations in the IFT122 gene. Am. J. Hum. Genet. *86*, 949–956.
6. 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.
7. Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. Nature *461*, 272–276.
8. Pushkarev, D., Neff, N.F., and Quake, S.R. (2009). Single-molecule sequencing of an individual human genome. Nat. Biotechnol. *27*, 847–852.
9. Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Zhang, J., et al. (2008). The diploid genome sequence of an Asian individual. Nature *456*, 60–65.
10. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121.
11. Gherman, A., Davis, E.E., and Katsanis, N. (2006). The ciliary proteome database: An integrated community resource for the genetic and functional dissection of cilia. Nat. Genet. *38*, 961–962.
12. Ramensky, V., Bork, P., and Sunyaev, S. (2002). Human non-synonymous SNPs: Server and survey. Nucleic Acids Res. *30*, 3894–3900.
13. Cardenas-Rodriguez, M., and Badano, J.L. (2009). Ciliary biology: Understanding the cellular and genetic basis of human ciliopathies. Am. J. Med. Genet. C Semin. Med. Genet. *151C*, 263–280.
14. Cole, D.G. (2003). The intraflagellar transport machinery of Chlamydomonas reinhardtii. Traffic *4*, 435–442.
15. Mill, P., Hall, E., Keighren, M., Lawson, K., and Jackson, I. (2009). Wdr35 is required for mammalian ciliogenesis and Hh responsiveness. Mech. Dev. *126* (*Suppl. 1*), S265.
16. Avidor-Reiss, T., Maer, A.M., Koundakjian, E., Polyanovsky, A., Keil, T., Subramaniam, S., and Zuker, C.S. (2004). Decoding cilia function: Defining specialized genes required for compartmentalized cilia biogenesis. Cell *117*, 527–539.
17. Beales, P.L., Bland, E., Tobin, J.L., Bacchelli, C., Tuysuz, B., Hill, J., Rix, S., Pearson, C.G., Kai, M., Hartley, J., et al. (2007). IFT80, which encodes a conserved intraflagellar transport protein, is mutated in Jeune asphyxiating thoracic dystrophy. Nat. Genet. *39*, 727–729.
18. Dagoneau, N., Goulet, M., Geneviève, D., Sznajer, Y., Martinovic, J., Smithson, S., Huber, C., Baujat, G., Flori, E., Tecco, L., et al. (2009). DYNC2H1 mutations cause asphyxiating thoracic dystrophy and short rib-polydactyly syndrome, type III. Am. J. Hum. Genet. *84*, 706–711.
19. Cameron, D.A., Pennimpede, T., and Petkovich, M. (2009). Tulp3 is a critical repressor of mouse hedgehog signaling. Dev. Dyn. *238*, 1140–1149.
20. Coleman, D.L., and Eicher, E.M. (1990). Fat (fat) and tubby (tub): Two autosomal recessive mutations causing obesity syndromes in the mouse. J. Hered. *81*, 424–427.
21. Ohlemiller, K.K., Hughes, R.M., Mosinger-Ogilvie, J., Speck, J.D., Grosof, D.H., and Silverman, M.S. (1995). Cochlear and retinal degeneration in the tubby mouse. Neuroreport *6*, 845–849.
22. Ikeda, S., Shiva, N., Ikeda, A., Smith, R.S., Nusinowitz, S., Yan, G., Lin, T.R., Chu, S., Heckenlively, J.R., North, M.A., et al. (2000). Retinal degeneration but not obesity is observed in null mutants of the tubby-like protein 1 gene. Hum. Mol. Genet. *9*, 155–163.
23. Norman, R.X., Ko, H.W., Huang, V., Eun, C.M., Abler, L.L., Zhang, Z., Sun, X., and Eggenschwiler, J.T. (2009). Tubby-like protein 3 (TULP3) regulates patterning in the mouse embryo through inhibition of Hedgehog signaling. Hum. Mol. Genet. *18*, 1740–1754.
24. Pierce, S.B., Walsh, T., Chisholm, K.M., Lee, M.K., Thornton, A.M., Fiumara, A., Opitz, J.M., Levy-Lahad, E., Klevit, R.E., and King, M.C. (2010). Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault syndrome. Am. J. Hum. Genet. *87*, 282–288.

tolerance to freezing (e.g., tomato, maize, and rice), suggesting that SFR2 function is not restricted to freezing protection (*9*). There is a large overlap in the mechanisms required for freezing tolerance and dehydration because of water deficit or high salinity (*3*, *4*), and SFR2 orthologs in these species may act on membrane stabilization during other abiotic stresses that cause cellular dehydration. Indeed, infiltration of *Arabidopsis* leaves with any osmotically active compound tested induced GGGT activity (table S1).

Our hypothesis for the requirement of SFR2-dependent galactolipid remodeling in freezing tolerance centers on the prevention of membrane fusion from the formation of non-bilayer H_{II}-type structures brought about by dehydration. During dehydration, non-bilayer structures are formed at the interface of apposed membranes and are believed to initiate at the chloroplast envelope membranes during freezing (*1*, *2*, *8*). This results in fusion between bilayers (*27*), particularly when membranes are enriched in glycerolipid species with relatively small head groups (e.g., MGDG and phosphatidylethanolamine), because these show a higher propensity for transition to the H_{II} phase in vitro. Previously, *sfr2* mutants showed extensive chloroplast and tonoplast rupture in leaves during freezing recovery, which was proposed to arise from fusion of destabilized membranes (*9*). Here, we have shown that SFR2 partially converts MGDG to DGDG and oligogalactolipids, the latter of which are not prone to form H_{II} phases in vitro. In addition to the prevention of non-bilayer–type structures, accumulation of oligogalactolipids results in an increased average thickness of the head-group domain of the bilayers and an increase in the localized concentration of hydroxyl groups per unit surface area, which enhance the repulsive hydration force between apposed bilayers during freeze-induced dehydration (*28*). Together, these factors could promote a sufficient distance between apposed bilayers that prevents membrane fusion. The overall phenomenon is analogous to the UDP-glucose–dependent modulation of mono- to diglucolipid ratios observed in *Acheloplasma laidlawii* under different abiotic stress conditions (*29*). One distinction is that DAG is produced by SFR2 and is further metabolized to TAG and possibly other lipid species, thereby preventing the accumulation of DAG, which can form nonlamellar phases. In this regard, the action of SFR2 provides a mechanism by which polar membrane lipids and excess membrane are removed by conversion to nonpolar lipids (e.g., TAGs) to accommodate a shrinking organelle after freezing or, more generally, osmotic stress.

References and Notes

1. P. L. Steponkus, *Annu. Rev. Plant Physiol.* **35**, 543 (1984).
2. M. Uemura, R. A. Joseph, P. L. Steponkus, *Plant Physiol.* **109**, 15 (1995).
3. M. F. Thomashow, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 571 (1999).
4. T. Hirayama, K. Shinozaki, *Plant J.* **61**, 1041 (2010).
5. S. Fowler, M. F. Thomashow, *Plant Cell* **14**, 1675 (2002).
6. D. Cook, S. Fowler, O. Fiehn, M. F. Thomashow, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15243 (2004).
7. M. Uemura, G. Warren, P. L. Steponkus, *Plant Physiol.* **131**, 1800 (2003).
8. P. L. Steponkus, M. Uemura, R. A. Joseph, S. J. Gilmour, M. F. Thomashow, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14570 (1998).
9. N. Fourrier *et al.*, *Plant J.* **55**, 734 (2008).
10. R. McKown, G. Kuroki, G. Warren, *J. Exp. Bot.* **47**, 1919 (1996).
11. G. Thorlby, N. Fourrier, G. Warren, *Plant Cell* **16**, 2192 (2004).
12. Z. Y. Du, S. Xiao, Q. F. Chen, M. L. Chye, *Plant Physiol.* **152**, 1585 (2010).
13. W. Li *et al.*, *J. Biol. Chem.* **283**, 461 (2008).
14. A. van Besouw, J. F. Wintermans, *Biochim. Biophys. Acta* **529**, 44 (1978).
15. C. Benning, H. Ohta, *J. Biol. Chem.* **280**, 2397 (2005).
16. P. Dörmann, I. Balbo, C. Benning, *Science* **284**, 2181 (1999).
17. A. A. Kelly, J. E. Froehlich, P. Dörmann, *Plant Cell* **15**, 2694 (2003).
18. Materials and methods are available as supporting material on *Science* Online.
19. E. Heinz, in *Advances in Lipid Methodology - Three*, W. W. Christy, Ed. (Oily, Dundee, UK, 1996), chap. 5, pp. 211–332.
20. M. L. Sinnott, *Chem. Rev.* **90**, 1171 (1990).
21. A. J. Dorne, M. A. Block, J. Joyard, R. Douce, *FEBS Lett.* **145**, 30 (1982).
22. J. Browse, L. Kunst, S. Anderson, S. Hugly, C. Somerville, *Plant Physiol.* **90**, 522 (1989).
23. J. W. Heemskerk, G. Bögemann, G. M. Wintermans, *Biochim. Biophys. Acta* **754**, 181 (1983).
24. A. D. Hill, P. J. Reilly, *Biopolymers* **89**, 1021 (2008).
25. A. R. Marques, P. M. Coutinho, P. Videira, A. M. Fialho, I. Sá-Correia, *Biochem. J.* **370**, 793 (2003).
26. G. Perugino, A. Trincone, M. Rossi, M. Moracci, *Trends Biotechnol.* **22**, 31 (2004).
27. D. P. Siegel, R. M. Epand, *Biophys. J.* **73**, 3089 (1997).
28. J. Wolfe, G. Bryant, *Cryobiology* **39**, 103 (1999).
29. A. Wieslander, L. Rilfors, G. Lindblom, *Biochemistry* **25**, 7511 (1986).

30. We thank S. Marshall and T. Durret for help with proton-NMR and mass spectrometry analyses, respectively, and the laboratory of M. Thomashow for use of equipment and advice on the plant freezing experiments conducted in this work. This work was supported in parts by a grant from the U.S. Department of Energy (DE-FG02-98ER20305) and Michigan Agricultural Experiment Station.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1191803/DC1
Materials and Methods
Figs. S1 to S6
Table S1
References

3 May 2010; accepted 4 August 2010
Published online 26 August 2010;
10.1126/science.1191803
Include this information when citing this paper.

Frequent Mutations of Chromatin Remodeling Gene *ARID1A* in Ovarian Clear Cell Carcinoma

Siân Jones,[1] Tian-Li Wang,[2] Ie-Ming Shih,[3] Tsui-Lien Mao,[4] Kentaro Nakayama,[5] Richard Roden,[3] Ruth Glas,[6] Dennis Slamon,[6] Luis A. Diaz Jr.,[1] Bert Vogelstein,[1] Kenneth W. Kinzler,[1]* Victor E. Velculescu,[1]* Nickolas Papadopoulos[1]*

Ovarian clear cell carcinoma (OCCC) is an aggressive human cancer that is generally resistant to therapy. To explore the genetic origin of OCCC, we determined the exomic sequences of eight tumors after immunoaffinity purification of cancer cells. Through comparative analyses of normal cells from the same patients, we identified four genes that were mutated in at least two tumors. *PIK3CA*, which encodes a subunit of phosphatidylinositol-3 kinase, and *KRAS*, which encodes a well-known oncoprotein, had previously been implicated in OCCC. The other two mutated genes were previously unknown to be involved in OCCC: *PPP2R1A* encodes a regulatory subunit of serine/threonine phosphatase 2, and *ARID1A* encodes adenine-thymine (AT)–rich interactive domain–containing protein 1A, which participates in chromatin remodeling. The nature and pattern of the mutations suggest that *PPP2R1A* functions as an oncogene and *ARID1A* as a tumor-suppressor gene. In a total of 42 OCCCs, 7% had mutations in *PPP2R1A* and 57% had mutations in *ARID1A*. These results suggest that aberrant chromatin remodeling contributes to the pathogenesis of OCCC.

Ovarian cancers are a heterogeneous group of diseases with distinct clinicopathological and molecular features (*1*). Among them, OCCCs, which account for 10% of epithelial ovarian cancers, is one of the most aggressive types because, unlike the high grade–serous type, it is refractory to standard platinum-based chemotherapy. Previous morphological and molecular studies have indicated that OCCC develops in a stepwise fashion from a common disease progenitor state, in many cases endometriosis, and then proceeds to frank malignancy (*2*–*6*). Activating mutations in *PIK3CA* (*7*) and genomic amplification of chr20q13.2 (*8*) are the most common molecular genetic alterations so far identified in OCCC.

To explore the genetic basis of this tumor type, we have determined the sequences of the ~18,000 protein-encoding genes listed in the RefSeq data-

[1]Ludwig Center for Cancer Genetics and Therapeutics and Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA. [2]Department of Gynecology and Obstetrics and Oncology, Johns Hopkins Medical Institutes, Baltimore, MD 21231, USA. [3]Department of Pathology, Oncology, Gynecology, and Obstetrics, Johns Hopkins Medical Institutes, Baltimore, MD 21231, USA. [4]Department of Pathology, National Taiwan University College of Medicine, Taipei 100, Taiwan. [5]Department of Gynecology and Obstetrics, Shimane University School of Medicine, Izumo, Shimane 6938501, Japan. [6]Division of Hematology/Oncology, David Geffen School of Medicine at the University of California, Los Angeles, CA 99095, USA.

*To whom correspondence should be addressed. E-mail: npapado1@jhmi.edu (N.P.); kinzlke@jhmi.edu (K.W.K.); velculescu@jhmi.edu (V.E.V.)

base in tumors from eight patients (table S1). Because these tumors are composed of a mixture of cancer and stromal cells, we purified the cancer cells using epithelial cell target antibodies attached to magnetic beads (*9*). Staining of the cells bound to the beads revealed that >90% of them were OCCC cells. This procedure thereby maximized the sensitivity of the sequencing analyses by eliminating most of the contaminating normal cells (containing normal genomes) from the sample. DNA from the purified cells, as well as from normal cells obtained from the blood or uninvolved tissues of the same patients, were used to generate libraries suitable for massively parallel sequencing by synthesis (*9*). After capture of the coding sequences of the targeted genes with a SureSelect Enrichment System, the DNA was sequenced using an Illumina GAIIx platform. The average coverage of each base in the targeted regions was 84-fold, and 92.7% of these bases were represented in at least 10 reads (table S2).

Using stringent criteria for analysis of these data (*9*), we identified 268 somatic mutations in 253 genes among the eight tumors. The range of mutations per tumor was 13 to 125 alterations. Of these, 237 (88%) mutations were confirmed by Sanger sequencing (table S3). The tumor with 125 mutations (OCC06PT) was from a patient with recurrent disease that had previously been treated with chemotherapy. Excluding OCC06PT, there was an average of 20 mutations per tumor (tables S2 and S3). The mutation spectrum was enriched for C to T transitions at 5′-CG base pairs, similar to those of other tumors whose exomes have been sequenced (*10*–*14*). Only four genes were mutated in more than one of the eight tumors studied: *PIK3CA*, *KRAS*, *PPP2R1A,* and *ARID1A*. The mutations in each of these four genes, and their somatic nature, were confirmed by Sanger sequencing of the DNA from the tumor and normal tissues of the corresponding patients (examples in Fig. 1). The sequences of these four genes were then determined in the tumor and normal tissues of an additional 34 OCCC cases using polymerase chain reaction amplification and Sanger sequencing with the primers listed in table S4. In total, *PIK3CA*, *KRAS*, *PPP2R1A*, and *ARID1A* mutations were identified in 40%, 4.7%, 7.1%, and 57% of the 42 tumors, respectively (Table 1).

Cancer cell lines with mutations in genes involved in cancer development provide valuable tools for further research. To this end, we extended the analysis of these four genes in seven OCCC cell lines that were derived from tumors independent of those described above. Five of the seven cell lines had mutations in *ARID1A* (nine mutations total), three had *PPP2R1A* mutations, one had a *KRAS* mutation, and four had *PIK3CA* mutations (table S5).

The nature of the somatic mutations in tumors can often be used to classify them as oncogenes or tumor suppressor genes (*15*). In particular, all bona fide oncogenes are mutated recurrently (that is, at the same codon or clustered at a few adjacent codons in different tumors), and the mutations are nearly always missense. In contrast, all bona fide tumor suppressor genes are mutated at a variety of positions throughout the coding region of the gene, and the mutations often truncate the encoded protein through production of a stop codon by a base substitution, an out-of-frame insertion or deletion (indel), or a splice site mutation. Moreover, tumor suppressor gene mutations generally affect both alleles, whereas mutations in oncogenes commonly affect only one allele.

Based on this logic, we can speculate about the likely function of the four genes in OCCC. *PIK3CA* and *KRAS* are well-studied oncogenes, and the 19 mutations identified in OCCC were heterozygous and clustered; fourteen of the 17 mutations in PIK3CA were at codons 542, 545, 546, or 1047, whereas both mutations in KRAS were at codon 12 (Table 1). The three mutations in *PPP2R1A* were similarly heterozygous and clustered, suggesting that it functions, when mutated, as an oncogene (Table 1). In contrast, the 32 mutations in *ARID1A* were distributed throughout the coding region, and all were predicted to truncate the protein through a base substitution resulting in a stop codon (9 mutations) or an out-of-frame insertion or deletion (23 mutations) (Table 1). In 11 of the 24 tumors with *ARID1A* mutations, both *ARID1A* alleles were affected through either a mutation in one allele and loss of heterozygosity of the other allele, or through two mutations that were presumably biallelic. Thus, we hypothesize that ARID1A functions as a tumor suppressor gene and that somatic mutations inactivate the gene product.

The serine/threonine protein phosphatase PP2A represents a family of holoenzymes with various activities. The holoenzyme contains a core composed of a heterodimer of a catalytic subunit (PPP2CA or PPP2CB) and a constant regulatory subunit (PPP2R1A or PPP2R1B). PPP2R1A serves as a scaffold to coordinate the interaction of the core enzyme with one of more than 15 regulatory subunits to form the heterotrimeric holoenzyme (*16*, *17*). Somatic mutations in *PPP2R1A* are not listed in the Cancer Gene Census of the Catalogue of Somatic Mutations in Cancer (COSMIC) database, although a few alterations in this gene have been previously reported (*18*). Functional studies have shown that PP2A is involved in the control of cell growth and division. Specifically, this protein is required for proper chromosome segregation through its interactions with Bub1 and Sgo1 (*19*). The two arginine residues that were somatically mutated in OCCC are highly conserved and reside within one of the Huntington, elongation factor3, PP2A, TOR (HEAT) domains of PPP2R1A that are involved in binding regulatory subunits.

The protein encoded by *ARID1A* can bind to AT-rich DNA sequences and is a component of the adenosine triphosphate–dependent chromatin modeling complex switch/sucrose-nonfermentable (SWI/SNF). The SWI/SNF chromatin-remodeling complex mobilizes nucleosomes and functions as

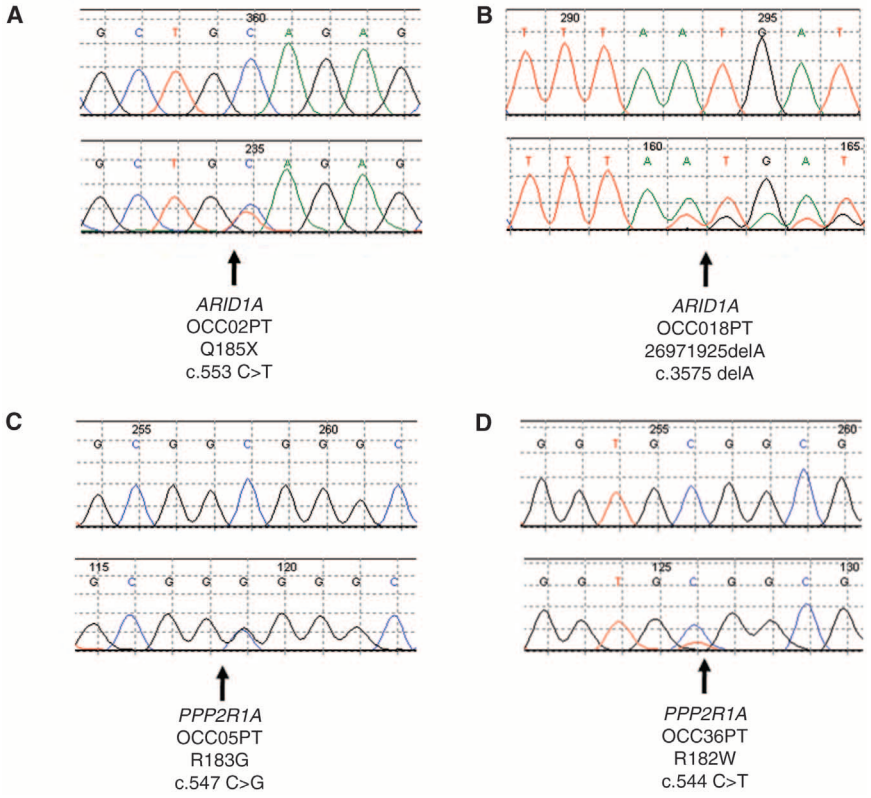


Fig. 1. Sequence chromatograms showing somatic *ARID1A* and *PPP2R1A* mutations. The lower panels (**C** and **D**) show the tumor, and the upper panels (**A** and **B**) show the matched normal control.

Table 1. Mutations in *ARID1A*, *KRAS*, *PIK3CA*, and *PPP2R1A* in human ovarian clear cell carcinomas.*

Sample†	Gene	Transcript accession	Nucleotide (genomic)‡	Nucleotide (cDNA)	Amino acid (protein)	Mutation type
OCC01PT	*ARID1A*	CCDS285.1	g.chr1:26972561_26972562insA	c.3854_3855insA	fs	Indel
OCC02PT	*ARID1A*	CCDS285.1	g.chr1:26896034C>T	c.553C>T	p.Q185X	Nonsense
OCC02PT	*ARID1A*	CCDS285.1	g.chr1:26978879-26978880dupGT g.chr1:26972009_26972034delTGATGGGGCG	c.5903_5904dupGT c.3659_3684delTGATGGGGCG	fs	Indel
OCC03PT	*ARID1A*	CCDS285.1	CATGTCCTATGAGCCA (hom)	CATGTCCTATGAGCCA	fs	Indel
OCC07PT	*ARID1A*	CCDS285.1	g.chr1:26896066C>A	c.585C>A	p.Y195X	Nonsense
OCC08PT	*ARID1A*	CCDS285.1	g.chr1:26970389delC	c.3391delC	fs	Indel
OCC10PT	*ARID1A*	CCDS285.1	g.chr1:26972790_26972792dupGCA (hom)	c.4001_4002dupGCA (hom)	fs	Indel
OCC10PT	*ARID1A*	CCDS285.1	g.chr1:26979804_26979805delTG (hom)	c.6828_6829delTG(hom)	fs	Indel
OCC11PT	*ARID1A*	CCDS285.1	g.chr1:26930334_26930335insCCTAC	c.1451_1455insCCTAC	fs	Indel
OCC13PT	*ARID1A*	CCDS285.1	g.chr1:26974233_26974234insTGGC	c.4926_4927insTGGC	fs	Indel
OCC14PT	*ARID1A*	CCDS285.1	g.chr1:26972886_26972887_delTT (hom)	c.4011_4012delTT (hom)	fs	Indel
OCC15PT	*ARID1A*	CCDS285.1	g.chr1:26973940G>A	c.4635G>A	p.W1545X	Nonsense
OCC15PT	*ARID1A*	CCDS285.1	g.chr1:26978178T>A	c.5202T>A	p.Y1734X	Nonsense
OCC16PT	*ARID1A*	CCDS285.1	g.chr1:26895967_26895973delCGCCGCC (hom)	c.486_492delCGCCGCC (hom)	fs	Indel
OCC18PT	*ARID1A*	CCDS285.1	g.chr1:26971925delA	c.3575delA	fs	Indel
OCC20PT	*ARID1A*	CCDS285.1	g.chr1:26970221delG	c.3223delG	fs	Indel
OCC22PT	*ARID1A*	CCDS285.1	g.chr1:26979694dupG	c.6718dupG	fs	Indel
OCC23PT	*ARID1A*	CCDS285.1	g.chr1:26896379_2689637980_insCGTC	c.898_899insCGTC	fs	Indel
OCC23PT	*ARID1A*	CCDS285.1	g.chr1:26979686_26979687insT	c.6710_6711insT	fs	Indel
OCC24PT	*ARID1A*	CCDS285.1	g.chr1:26930542C>T	c.1663C>T	p.Q555X	Nonsense
OCC27PT	*ARID1A*	CCDS285.1	g.chr1:26896263_26896272delCGTCGTCTTC	c.782_791delCGTCGTCTTC	fs	Indel
OCC27PT	*ARID1A*	CCDS285.1	g.chr1:26971984_26971994delCAGCCCAGTAT	c.3634_3644delCAGCCCAGTAT	fs	Indel
OCC30PT	*ARID1A*	CCDS285.1	g.chr1:26931823C>T	c.1873C>T	p.Q625X	Nonsense
OCC32PT	*ARID1A*	CCDS285.1	g.chr1:26960135C>T	c.2122C>T	p.Q708X	Nonsense
OCC34PT	*ARID1A*	CCDS285.1	g.chr1:26931754G>T	c.1804G>T	p.E602X	Nonsense
OCC34PT	*ARID1A*	CCDS285.1	g.chr1:26979678delT	c.6702delT	fs	Indel
OCC36PT	*ARID1A*	CCDS285.1	g.chr1:26928932T>G	c.1341T>G	p.Y447X	Nonsense
OCC36PT	*ARID1A*	CCDS285.1	g.chr1:26971613delC	c.3442delC	fs	Indel
OCC39PT	*ARID1A*	CCDS285.1	g.chr1:26896364dupC	c.883dupC	fs	Indel
OCC39PT	*ARID1A*	CCDS285.1	g.chr1:26965434delC	c.2868delC	fs	Indel
OCC41PT	*ARID1A*	CCDS285.1	g.chr1:26931831delT	c.1881delT	fs	Indel
OCC42PT	*ARID1A*	CCDS285.1	g.chr1:26960479_26960488delCGGCCACCCA	c.2179_2188delCGGCCACCCA	fs	Indel
OCC04PT	*KRAS*	CCDS8703.1	g.chr12:25289551C>T	c.35G>A	p.G12D	Missense
OCC05PT	*KRAS*	CCDS8703.1	g.chr12:25289551C>G	c.35G>C	p.G12A	Missense
OCC01PT	*PIK3CA*	CCDS43171.1	g.chr3:180418788C>A	c.1636C>A	p.Q546K	Missense
OCC02PT	*PIK3CA*	CCDS43171.1	g.chr3:180418776G>A	c.1624G>A	p.E542K	Missense
OCC06PT	*PIK3CA*	CCDS43171.1	g.chr3:180418785G>A	c.1633G>A	p.E545K	Missense
OCC08PT	*PIK3CA*	CCDS43171.1	g.chr3:180418785G>A	c.1633G>A	p.E545K	Missense
OCC09PT	*PIK3CA*	CCDS43171.1	g.chr3:180434779A>T	c.3140A>T	p.H1047L	Missense
OCC10PT	*PIK3CA*	CCDS43171.1	g.chr3:180434779A>G	c.3140A>G	p.H1047R	Missense
OCC11PT	*PIK3CA*	CCDS43171.1	g.chr3:180418777A>T	c.1625A>T	p.E542V	Missense
OCC13PT	*PIK3CA*	CCDS43171.1	g.chr3:180434779A>G	c.3140A>G	p.H1047R	Missense
OCC15PT	*PIK3CA*	CCDS43171.1	g.chr3:180410152C>G	c.1221C>G	p.C407W	Missense
OCC20PT	*PIK3CA*	CCDS43171.1	g.chr3:180434779A>G	c.3140A>G	p.H1047R	Missense
OCC22PT	*PIK3CA*	CCDS43171.1	g.chr3:180434779A>G	c.3140A>G	p.H1047R	Missense
OCC23PT	*PIK3CA*	CCDS43171.1	g.chr3:180399648_180399649insCCTCAA	c.341_342insCCTCAA	N114_R115insLN	Indel
OCC27PT	*PIK3CA*	CCDS43171.1	g.chr3:180399638A>G	c.331A>G	p.K111E	Missense
OCC30PT	*PIK3CA*	CCDS43171.1	g.chr3:180434779A>G	c.3140A>G	p.H1047R	Missense
OCC35PT	*PIK3CA*	CCDS43171.1	g.chr3:180418776G>A	c.1624G>A	p.E542K	Missense
OCC36PT	*PIK3CA*	CCDS43171.1	g.chr3:180418785G>A	c.1633G>A	p.E545K	Missense
OCC42PT	*PIK3CA*	CCDS43171.1	g.chr3:180434779A>G	c.3140A>G	p.H1047R	Missense
OCC05PT	*PPP2R1A*	CCDS12849.1	g.chr19:57407794C>G	c.547C>G	p.R183G	Missense
OCC07PT	*PPP2R1A*	CCDS12849.1	g.chr19:57407794C>T	c.547C>T	p.R183W	Missense
OCC36PT	*PPP2R1A*	CCDS12849.1	g.chr19:57407791C>T	c.544C>T	p.R182W	Missense

*Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. †Samples OCC01 to OCC08 were used for the initial (discovery) screen for mutations. ‡Coordinates refer to the human reference genome hg18 release (NCBI 36.1, March 2006).

a regulator of gene expression and chromatin dynamics. ARID1A is one of the two mutually exclusive ARID1 subunits of the SWI/SNF complex and is thought to provide specificity to this complex (*20*). Changes in chromatin can influence the epigenetic regulation of many genes, inducing those that play a role in cancer (*20*–*22*). Indeed, functional studies have implicated ARID1A in the ability of the SWI/SNF complex to inhibit cell growth (*23*). No mutations of *ARID1A* are listed in the Cancer Gene Census of the COSMIC database, but chromosomal translocations that involve this gene have been identified in a human breast cancer and a human lung cancer cell line (*24*). Knock-down of ARID1A in a leukemia cell line confers resistance to Fas-mediated apoptosis (*25*).

The results of this study emphasize two themes in modern cancer genetics. The first is that specific tumor types are characterized by mutations in "communal cancer genes" like *KRAS* and *PIK3CA*, as well as in "restricted cancer genes" like *PPP2R1A* and *ARID1A*. The communal cancer genes are involved in a variety of cancers and have been extensively studied. Restricted cancer genes have been shown to contribute to specific types of leukemias and sarcomas, mainly through translocations (e.g., the *ABL* oncogene in chronic myelogenous leukemia and *EWS* fused to a gene encoding an ETS transcription factor family member in Ewing's sarcoma). With the advent of whole exome sequencing, we are beginning to see similar specificity with respect to point mutations [e.g., *IDH1* in gliomas (*14*) and *GNAQ* in uveal melanomas (*26*)].

The second theme is that mutations of chromatin-modifying genes are characteristic of certain tumor types. Recent examples include the *JARID1C* gene, also known as lysine (K)–specific demethylase 5C (*KDM5C*), in renal cell cancers (*27*), *SMARCA4/BRG1* (SWI/SNF related, matrix-associated, actin-dependent regulator of chromatin, subfamily a, member 4) in lung cancers (*28*), and now *ARID1A* in OCCC. Epigenetic changes in cancers, including methylation of deoxycytidine residues in DNA and a variety of covalent modifications of chromatin proteins, have been extensively studied (*20*–*22*, *29*). Interestingly, however, the reason(s) that DNA methylation and chromatin are different in cancer cells than in normal cells is completely unknown. Similarly, the relationship between the genetic alterations that unequivocally drive tumorigenesis and the epigenetic changes that are so widespread in tumor genomes has not been defined. Discovery of tumor suppressor genes such as *ARID1A* that are mutated in cancers bridge this gap, because they are likely to directly lead to epigenetic changes in cancer cells through specific modifications of chromatin proteins. They additionally provide a potential approach to determine which of the numerous epigenetic changes in cancers confer a selective growth advantage and which are simply "passengers" that do not play a causal role. The identification of the genes whose expression is specifically modulated by *ARID1A* inactivation will be the next crucial step in this line of research.

References and Notes

1. K. R. Cho, I.-M. Shih, *Annu. Rev. Pathol.* **4**, 287 (2009).
2. M. Erzen, S. Rakar, B. Klancnik, K. Syrjänen, *Gynecol. Oncol.* **83**, 100 (2001).
3. M. Fukunaga, K. Nomura, E. Ishikawa, S. Ushigome, *Histopathology* **30**, 249 (1997).
4. R. T. Marquez *et al.*, *Clin. Cancer Res.* **11**, 6116 (2005).
5. N. Sato *et al.*, *Cancer Res.* **60**, 7052 (2000).
6. E. Veras *et al.*, *Am. J. Surg. Pathol.* **33**, 844 (2009).
7. K. T. Kuo *et al.*, *Am. J. Pathol.* **174**, 1597 (2009).
8. K. T. Kuo *et al.*, *Clin. Cancer Res.* **16**, 1997 (2010).
9. Materials and methods are available as supporting material on *Science* Online.
10. C. Greenman *et al.*, *Nature* **446**, 153 (2007).
11. T. Sjöblom *et al.*, *Science* **314**, 268 (2006).
12. L. D. Wood *et al.*, *Science* **318**, 1108 (2007).
13. S. Jones *et al.*, *Science* **321**, 1801 (2008).
14. D. W. Parsons *et al.*, *Science* **321**, 1807 (2008).
15. B. Vogelstein, K. W. Kinzler, *Nat. Med.* **10**, 789 (2004).
16. P. J. Eichhorn, M. P. Creyghton, R. Bernards, *Biochim. Biophys. Acta* **1795**, 1 (2009).
17. Y. Shi, *Cell* **139**, 468 (2009).
18. G. A. Calin *et al.*, *Oncogene* **19**, 1191 (2000).
19. Z. Tang *et al.*, *Dev. Cell* **10**, 575 (2006).
20. J. I. Wu, J. Lessard, G. R. Crabtree, *Cell* **136**, 200 (2009).
21. B. Weissman, K. E. Knudsen, *Cancer Res.* **69**, 8223 (2009).
22. F. Banine *et al.*, *Cancer Res.* **65**, 3542 (2005).
23. N. G. Nagl Jr., X. Wang, A. Patsialou, M. Van Scoy, E. Moran, *EMBO J.* **26**, 752 (2007).
24. J. Huang, Y. L. Zhao, Y. Li, J. A. Fletcher, S. Xiao, *Genes Chromosomes Cancer* **46**, 745 (2007).
25. B. Luo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20380 (2008).
26. C. D. Van Raamsdonk *et al.*, *Nature* **457**, 599 (2009).
27. G. L. Dalgliesh *et al.*, *Nature* **463**, 360 (2010).
28. P. P. Medina *et al.*, *Hum. Mutat.* **29**, 617 (2008).
29. P. A. Jones, S. B. Baylin, *Cell* **128**, 683 (2007).

30. We thank Y. He for technical advice and M. Whalen, J. Ptak, and N. Silliman for expert technical assistance. Funded by the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation, American Association for Cancer Research Stand Up To Cancer–Dream Team Translational Cancer Research Grant, The Virginia and D. K. Ludwig Fund for Cancer Research, U.S. Department of Defense grant OC0400600 (T.-L.W.), and NIH grants CA121113, CA57345, CA129080, CA103937 (IMS), and CA122581 (R.R.). N.P., B.V., L.D., V.V., and K.W.K. are members of the Scientific Advisory Board of Inostics, a company that is developing technologies for the molecular diagnosis of cancer. N.P., B.V., L.D., V.V., and K.W.K. also own stock in Inostics. The authors are entitled to a share of the royalties received by the university on sales of products related to genes described in this manuscript. The terms of these arrangements are being managed by the university in accordance with their conflict of interest policies. V.V., B.V., K.W.K., and Johns Hopkins University also own stock in Genzyme and Exact Sciences, which are managed under similar arrangements.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1196333/DC1
Materials and Methods
Tables S1 to S5
References

10 August 2010; accepted 26 August 2010
Published online 8 September 2010;
10.1126/science.1196333
Include this information when citing this paper.

Histone H3 Thr-3 Phosphorylation by Haspin Positions Aurora B at Centromeres in Mitosis

Fangwei Wang,[1] Jun Dai,[1] John R. Daum,[2] Ewa Niedzialkowska,[3] Budhaditya Banerjee,[3] P. Todd Stukenberg,[3] Gary J. Gorbsky,[2] Jonathan M. G. Higgins[1]*

Aurora B is a component of the chromosomal passenger complex (CPC) required for correct spindle-kinetochore attachments during chromosome segregation and for cytokinesis. The chromatin factors that recruit the CPC to centromeres are unknown, however. Here we show that phosphorylation of histone H3 threonine 3 (H3T3ph) by Haspin is necessary for CPC accumulation at centromeres and that the CPC subunit Survivin binds directly to H3T3ph. A nonbinding Survivin-D70A/D71A mutant does not support centromeric CPC concentration, and both Haspin depletion and Survivin-D70A/D71A mutation diminish centromere localization of the kinesin MCAK and the mitotic checkpoint response to taxol. Survivin-D70A/D71A mutation and microinjection of H3T3ph-specific antibody both compromise centromeric Aurora B functions but do not prevent cytokinesis. Therefore, H3T3ph generated by Haspin positions the CPC at centromeres to regulate selected targets of Aurora B during mitosis.

The chromosomal passenger complex (CPC)—which contains Aurora B, inner centromere protein (INCENP), Survivin, and Borealin—is found on chromosome arms in prophase, concentrates at inner centromeres during prometaphase, and transfers to the central spindle at anaphase (*1*). Aurora B phosphorylates several substrates at these locations, including histone H3 Ser^{10} (H3S10ph) on chromosome arms, mitotic centromere-associated kinesin (MCAK) at inner centromeres, centromere protein–A Ser^{7} (CENP-AS7ph) at outer centromeres, and the KNL1/Mis12 complex/Ndc80 network at kinetochores (*1*–*7*). Current models suggest that centromeric Aurora B responds to lack of tension across sister kinetochores that are incorrectly attached to the spindle. Bipolar kinetochore attachment forces may pull kinetochore substrates away from inner centromeric Aurora B, which would lead to substrate dephosphorylation, selective stabilization of correct microtubule attachments, and eventually satisfaction of the spindle checkpoint (fig. S1) (*8*).

nature genetics

PHF6 mutations in T-cell acute lymphoblastic leukemia

Pieter Van Vlierberghe[1–3,21], Teresa Palomero[1,4,21], Hossein Khiabanian[5], Joni Van der Meulen[2], Mireia Castillo[4], Nadine Van Roy[2], Barbara De Moerloose[6], Jan Philippé[7], Sara González-García[8], María L Toribio[8], Tom Taghon[7], Linda Zuurbier[3], Barbara Cauwelier[9], Christine J Harrison[10], Claire Schwab[10], Markus Pisecker[11], Sabine Strehl[11], Anton W Langerak[12], Jozef Gecz[13,14], Edwin Sonneveld[15], Rob Pieters[3,15], Elisabeth Paietta[16], Jacob M Rowe[17], Peter H Wiernik[16], Yves Benoit[6], Jean Soulier[18], Bruce Poppe[2], Xiaopan Yao[19], Carlos Cordon-Cardo[4], Jules Meijerink[3], Raul Rabadan[5], Frank Speleman[2,22] & Adolfo Ferrando[1,4,20,22]

Tumor suppressor genes on the X chromosome may skew the gender distribution of specific types of cancer[1,2]. T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematological malignancy with an increased incidence in males[3]. In this study, we report the identification of inactivating mutations and deletions in the X-linked plant homeodomain finger 6 (*PHF6*) gene in 16% of pediatric and 38% of adult primary T-ALL samples. Notably, *PHF6* mutations are almost exclusively found in T-ALL samples from male subjects. Mutational loss of *PHF6* is importantly associated with leukemias driven by aberrant expression of the homeobox transcription factor oncogenes *TLX1* and *TLX3*. Overall, these results identify *PHF6* as a new X-linked tumor suppressor in T-ALL and point to a strong genetic interaction between *PHF6* loss and aberrant expression of TLX transcription factors in the pathogenesis of this disease.

T-ALL is an aggressive malignancy in which multiple genetic defects collaborate in the transformation of T-cell progenitors[4,5]. Notably, T-ALL has a threefold higher incidence in males[3], whereas other immature hematological tumors such as precursor B-lineage ALL are equally frequent in males and females[3].

To identify a possible X-linked tumor suppressor in T-ALL, we performed an X-chromosome-targeted mutational analysis in tumor DNA samples from 12 males with T-ALL. For each sample, we performed in-solution DNA capture of 7,674 regions encompassing 3,045,708 nucleotides corresponding to 5,215 X-chromosome exons using the Agilent Sure Select oligonucleotide capture system[6]. DNA samples enriched for X-chromosome exons were then analyzed by next-generation sequencing using the SOLiD 3 platform from Applied Biosystems. This analysis identified 66 candidate previously uncharacterized nonsynonymous single-nucleotide variants and 7 positions with high confidence calls for containing complex variants such as insertions or deletions (**Fig. 1a**). Dideoxynucleotide DNA sequencing of PCR products encompassing affected exons confirmed the presence of 92% (61/66) of these single-nucleotide variants and 57% (4/7) of the more complex variants, including 2 insertions and 2 deletions (**Supplementary Tables 1** and **2**). Sequence analysis of paired DNA samples obtained at the time of clinical remission showed that most of these variants corresponded to previously unreported germline polymorphisms. However, and most notably, we also identified three somatically acquired changes corresponding to two nonsynonymous single-nucleotide substitutions (A902G T300A and A990G H330R) and a frameshift-creating insertion of five nucleotides (124_125insAGGCA, H43fs) in the *PHF6* gene (**Fig. 1a**).

In a complementary approach we analyzed X-chromosome array comparative genome hybridization (array-CGH) data from 246 primary T-ALL samples (179 male and 67 female) in a multicenter setting. These analyses revealed the presence of recurrent deletions in chromosomal band Xq26 in 8 out of 246 (~3%) T-ALL samples (**Table 1**). For three del(X)(q26)-positive T-ALL samples, we performed array-CGH analysis against the corresponding remission material, which showed that these Xq26 deletions were somatically acquired leukemia-associated genetic events (**Table 1**). Reanalysis of all eight del(X)(q26)-positive T-ALL samples on a custom

[1]Institute for Cancer Genetics, Columbia University Medical Center, New York, New York, USA. [2]Center for Medical Genetics, Ghent University Hospital, Ghent, Belgium. [3]Department of Pediatric Oncology/Hematology, Erasmus MC, Rotterdam, The Netherlands. [4]Department of Pathology, Columbia University Medical Center, New York, New York, USA. [5]Center for Computational Biology and Bioinformatics, Columbia University, New York, New York, USA. [6]Department of Pediatric Hemato-Oncology, Ghent University Hospital, Ghent, Belgium. [7]Department of Clinical Chemistry, Immunology and Microbiology, Ghent University Hospital, Ghent, Belgium. [8]Centro de Biología Molecular "Severo Ochoa", Consejo Superior de Investigaciones Científicas (CSIC), Universidad Autónoma de Madrid (UAM), Madrid, Spain. [9]Department of Hematology, Hospital St-Jan, Bruges, Belgium. [10]Leukaemia Research Cytogenetics Group, Northern Institute for Cancer Research, Newcastle University, Newcastle, UK. [11]Children's Cancer Research Institute, St. Anna Kinderkrebsforschung, Vienna, Austria. [12]Department of Immunology, Erasmus MC, Rotterdam, The Netherlands. [13]Department of Genetics and Molecular Pathology, University of Adelaide, Adelaide, Australia. [14]Department of Pediatrics, University of Adelaide, Adelaide, Australia. [15]On behalf of the Dutch Childhood Oncology Group (DCOG), The Hague, The Netherlands. [16]Montefiore Medical Center North, Bronx, New York, USA. [17]Rambam Medical Center and Technion, Israel Institute of Technology, Haifa, Israel. [18]Hematology Laboratory APHP, INSERM U944, Hôpital Saint Louis, Paris, France. [19]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [20]Department of Pediatrics, Columbia University Medical Center, New York, New York, USA. [21]These authors contributed equally to this work. [22]These authors jointly directed this work. Correspondence should be addressed to A.F. (af2196@columbia.edu).

Received 11 November 2009; accepted 3 February 2010; published online 14 March 2010; doi:10.1038/ng.542

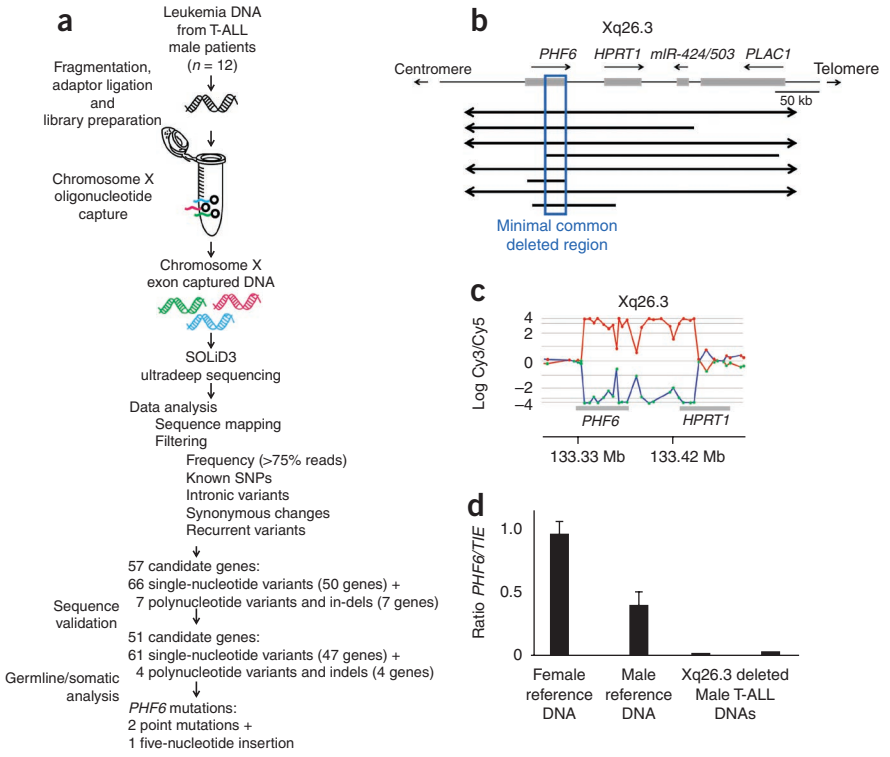


Figure 1 Next-generation sequencing and microarray-based comparative genomic hybridization (array-CGH) analysis of the X chromosome identifies *PHF6* mutations in human T-cell acute lymphoblastic leukemia (T-ALL). (**a**) Overview of mutation screening approach of the human X-chromosome exome in a panel of tumor DNA samples from 12 males with T-ALL using oligonucleotide sequence capture and next-generation sequencing with SOLiD 3. After filtering and confirmation of high-throughput sequencing data, analysis of corresponding remission DNA samples led to the identification of three somatically acquired changes in *PHF6*. (**b**) Schematic overview of the recurrent genomic deletions involving chromosomal band Xq26.3 in eight human T-ALL samples. Specific genes located in Xq26.3 are shown. (**c**) Detailed view of a representative oligonucleotide array-CGH plot of leukemia DNA/control DNA ratios (blue trace) versus the dye-swap experiment (red trace) in an individual harboring an Xq26.3 deletion. (**d**) DNA quantitative PCR analysis of *PHF6* copy number dose in female and male reference genomic DNAs and two primary samples from males with T-ALL harboring Xq26.3 deletions.

high-resolution X-chromosome oligonucleotide array (**Fig. 1b**,**c**) narrowed down the common minimally deleted region to an area of 80 kb containing the *PHF6* gene. Consistently, quantitative PCR analysis confirmed loss of the *PHF6* locus in the del(X)(q26)-positive cases (**Fig. 1d**). The convergent findings of our X-chromosome exon mutation analysis and analysis of copy number alterations by array-CGH thus identified *PHF6* as a new tumor suppressor gene mutated and deleted in T-ALL.

PHF6 encodes a plant homeodomain (PHD) factor containing four nuclear localization signals and two imperfect PHD zinc-finger domains[7] with a proposed role in controlling gene expression[7]. Notably, inactivating mutations in *PHF6* cause Börjeson-Forssman-Lehmann syndrome (MIM301900), a relatively uncommon type of X-linked familial syndromic mental retardation that has not been associated with increased incidence of T-ALL[7–9]. Quantitative RT-PCR analysis demonstrated ubiquitous expression of *PHF6* transcripts in human tissues, with highest levels of expression in thymus, ovary and thyroid, and moderate levels of expression in spleen, testes and adipose tissue (**Supplementary Fig. 1**). Consistent with these results, PHF6 was readily detected by immunohistochemistry in mouse thymus (**Supplementary Fig. 1**). Finally, quantitative RT-PCR analysis of human thymocyte populations at different stages of development showed variable levels of *PHF6* expression, with marked upregulation of *PHF6* transcripts in CD4/CD8 double-positive cells (**Supplementary Fig. 1**).

Mutational analysis of *PHF6* in an extended panel of pediatric and adult T-ALL primary samples identified truncating or missense mutations in *PHF6* in 38% (16/42) of adult and ~16% (14/89) of pediatric T-ALL samples (**Fig. 2a** and **Table 1**). In all available cases (7/30), analysis of matched buccal and/or bone marrow remission genomic DNA confirmed the somatic origin of *PHF6* mutations (4/21 frameshift mutations and 3/9 missense mutations) (**Fig. 2b** and **Table 1**). Finally, no mutations in *PHF6* were identified in DNA samples from B-lineage ALL ($n = 62$), suggesting that mutational loss of *PHF6* in lymphoid tumors could be restricted to T-ALL.

Nonsense and frameshift mutations accounted for 70% (21/30) of all *PHF6* mutations identified in our series and were evenly distributed throughout the gene. Missense mutations accounted for the remaining 30% (9/30) of *PHF6* lesions and recurrently involved codon C215 and the second zinc-finger domain of the protein (**Fig. 2a**). DNA sequence analysis of *PHF6* in a panel of 15 well-characterized T-ALL cell lines (**Supplementary Table 3**) showed the presence of truncating mutations in *PHF6* in the DND41, HPB-ALL and T-ALL1 cell lines. Protein blot analysis and immunohistochemical staining demonstrated robust expression and nuclear localization of PHF6 in *PHF6* wild-type tumors and complete loss of PHF6 protein in T-ALL cell lines harboring mutations in *PHF6* (**Fig. 2c**,**d**).

PHD finger-containing proteins have been implicated in numerous cellular functions, including transcriptional regulation and in some instances as specialized reader modules that recognize the methylation status of histone lysine residues[10]. In addition, PHF6 has been reported to be phosphorylated during mitosis[11] and by the ATM and ATR kinases upon DNA damage[12], which suggests a dynamic regulation of PHF6 during cell cycle and DNA repair. Consistent with this notion, short hairpin RNA (shRNA) knockdown of PHF6 resulted in increased levels of phosphorylated H2AX (γ-H2AX), a post-translational modification associated with the presence of DNA double-strand breaks[13] (**Fig. 2e**).

Sex determination in humans is controlled by differential representation of the X and Y chromosomes, with the presence of an XY pair in males and two copies of the X chromosome in females. The presence of numerous genes in the nonautosomal region of the X chromosome could result in a genetic imbalance between male and female cells, which is compensated for by random chromosomal inactivation of one copy of the X chromosome in female cells[14]. However, allelic-expression analysis has shown that some genes can escape X-chromosome inactivation in certain tissues[1,2,15]. To test the possibility that *PHF6* could escape X-chromosome inactivation in T-ALL cells, we performed allelic-expression analysis of a silent SNP (rs17317724) located in the 3′ UTR of *PHF6* in lymphoblasts from three informative samples from females with T-ALL. In each of these samples, *PHF6* was monoallelically expressed, suggesting that biallelic

Table 1 Characteristics of 38 primary T-ALL samples showing *PHF6* inactivation

ID	Sex	Age	WBC ($\times 10^9$ l^{-1})	Immuno-phenotype	Genetic subtype	*NOTCH1*	Type of alteration	Deletion size or predicted *PHF6* protein lesion	Germline or somatic
1	M	Ped	77	Cortical	*TLX3*	Mut	Deletion	0.55 Mb	NA
2	M	Ped	46	Pre-T	*TLX3*	WT	Deletion	0.23 Mb	NA
3	M	Ped	31	Pre-T	*TLX3*	NA	Deletion	1.50 Mb	NA
4	M	Ped	2	Pre-T	Unknown	NA	Deletion	0.27 Mb	NA
5	M	Ped	NA	Cortical	*HOXA*	NA	Deletion	1.90 Mb	Somatic
6	M	Ped	NA	Cortical	Unknown	NA	Deletion	0.20 Mb	Somatic
7	M	Ped	NA	Cortical	*TLX1*	NA	Deletion	0.08 Mb	Somatic
8	M	Adult	NA	Pre-T	Unknown	NA	Deletion	0.11 Mb	NA
9	M	Ped	185	Cortical	*TLX3*	WT	Nonsense	G122X	NA
10	M	Ped	417	Pre-T	*TLX3*	Mut	Nonsense	R116X	NA
11	F	Ped	280	Pre-T	*TLX1*	Mut	Frameshift	F172fs	NA
12	M	Ped	405	Pre-T	*TLX3*	Mut	Frameshift	Y303fs	NA
13	M	Ped	159	Pre-T	*TLX1*	Mut	Nonsense	K235X	NA
14	M	Ped	500	Pre-T	*TLX3*	Mut	Frameshift	A41fs	NA
15	M	Ped	347	Cortical	*HOXA*	Mut	Nonsense	K274X	NA
16	M	Ped	129	Cortical	Unknown	Mut	Frameshift	D333fs	NA
17	M	Ped	174	Cortical	*TLX3*	Mut	Nonsense	R225X	NA
18	M	Ped	27	Cortical	*TLX1*	WT	Nonsense	R116X	NA
19	M	Ped	310	Cortical	*TAL1*	Mut	Missense	C283R	NA
20	M	Ped	189	Cortical	*TAL1*	Mut	Frameshift	C28fs	NA
21	M	Adult	170	Cortical	*TLX3*	WT	Frameshift	H44fs	NA
22	M	Adult	21	Cortical	*TLX1*	Mut	Frameshift	H43fs	Somatic
23	M	Adult	NA	Pre-T	*TLX3*	Mut	Frameshift	T98fs	Somatic
24	M	Adult	14	Pre-T	*TLX3*	WT	Frameshift	Y105fs	NA
25	M	Adult	28	Mature	*TLX3*	WT	Frameshift	S158fs	NA
26	M	Adult	NA	Cortical	*TLX3*	Mut	Missense	C215Y	NA
27	M	Adult	NA	Pre-T	Unknown	Mut	Missense	C215F	NA
28	M	Adult	31	Cortical	Unknown	Mut	Missense	C215Y	NA
29	M	Adult	NA	Mature	*TLX3*	Mut	Missense	T300A	Somatic
30	M	Adult	21	Cortical	*TLX1*	WT	Missense	A311P	NA
31	M	Adult	30	Mature	Unknown	WT	Missense	C280Y	NA
32	M	Adult	23	Cortical	Unknown	WT	Missense	H329R	Somatic
33	M	Ped	NA	NA	*TAL1*	NA	Nonsense	R257X	Somatic
34	M	Ped	NA	NA	*TLX1*	NA	Frameshift	S191fs	Somatic
35	M	Adult	NA	NA	*TLX1*	NA	Missense	C215F	Somatic
36	M	Adult	NA	NA	*TLX1*	NA	Nonsense	Y303X	NA
37	M	Adult	NA	NA	Unknown	NA	Nonsense	R274X	NA
38	M	Adult	NA	NA	Unknown	NA	Frameshift	H135fs	NA

Mb, megabases; Mut, mutated; NA, not available; Ped, pediatric; WT, wild-type; X, stop codon.

expression of *PHF6* is not commonly found in T-ALL (**Fig. 3a**). Most notably, we found that *PHF6* mutations are almost exclusively found in samples from males with T-ALL. *PHF6* mutations were present in 32% (29/92) of males and in only ~2.5% (1/39) of females ($P < 0.001$; **Fig. 3b** and **Supplementary Table 4**). Moreover, all eight *PHF6* deletions identified by array-CGH analysis were found in samples from males with T-ALL, and each of the three cell lines with mutations in *PHF6* were derived from males with T-ALL.

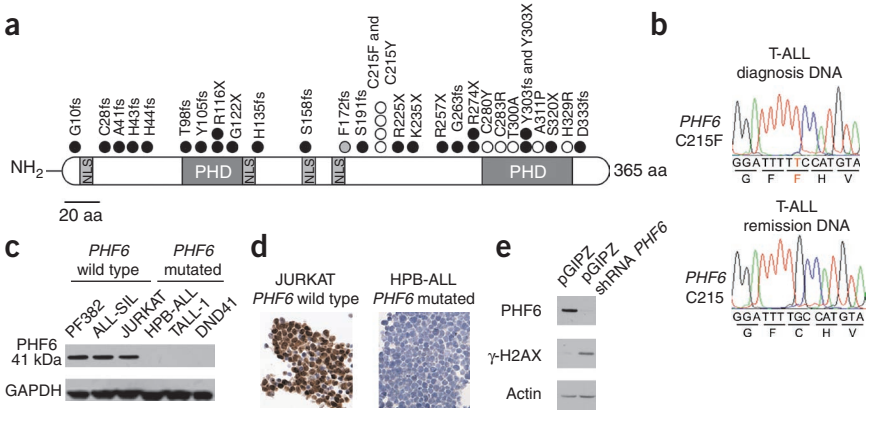


Figure 2 *PHF6* mutations and expression in T-cell acute lymphoblastic leukemia (T-ALL) lymphoblasts. (**a**) Structure of the PHF6 protein, including four nuclear localization signals (NLSs) and two imperfect plant homeodomain (PHD) zinc-finger domains. Overview of all *PHF6* mutations identified in primary T-ALL samples and T-ALL cell lines. Filled circles represent nonsense and frameshift mutations, whereas missense mutations are depicted as open circles. The circle filled in gray indicates the mutation identified in a sample from a female with T-ALL. aa, amino acids. (**b**) Representative DNA sequencing chromatograms of paired diagnosis and remission genomic T-ALL DNA samples showing a somatic mutation in exon 7 of *PHF6*. (**c**) Protein blot analysis of T-ALL cell lines revealed complete loss of PHF6 protein expression in the *PHF6*-mutated T-ALL cell lines. (**d**) PHF6 immunostaining in the Jurkat and HPB-ALL, wild-type and mutant T-ALL cell lines, respectively. (**e**) Protein blot analysis of PHF6 and γ-H2AX expression in HEK293T cells upon PHF6 short hairpin RNA knockdown. Actin concentrations are shown as a loading control.

Immunohistochemical analysis of PHF6 expression in wild-type primary T-ALL samples showed positive PHF6 immunostaining ($n = 5$; three males and two females), whereas cases with *PHF6*-truncating mutations ($n = 4$) (**Fig. 3c**) or a point mutation in C215 (C215F) were negative for PHF6 protein expression (**Fig. 3c**). In contrast, primary T-ALL cells harboring a *PHF6* point mutation in the PHD2 domain (T300A) were positive for PHF6 protein expression (**Fig. 3c**). Overall, these results suggest that truncating mutations and point mutations in C215 impair PHF6 expression, whereas amino acid substitutions in the PHD2 domain of PHF6 may selectively impair the tumor suppressor function of this protein.

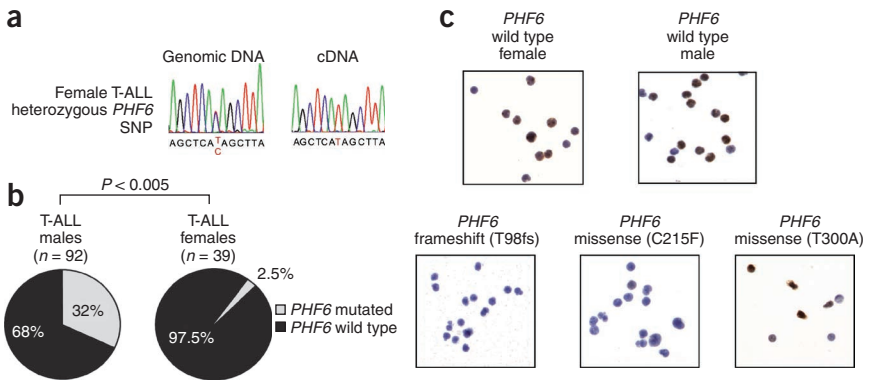


Figure 3 *PHF6* expression in T-cell acute lymphoblastic leukemia (T-ALL) lymphoblasts. (**a**) Sequence analysis of paired genomic DNA and complementary DNA samples shows monoallelic expression of *PHF6* SNP rs17317724 in lymphoblasts from a wild-type *PHF6* female with T-ALL. (**b**) Differential distribution of *PHF6* mutations in samples from males and females with T-ALL. (**c**) Immunohistochemical analysis of PHF6 expression in wild-type and mutant T-ALL lymphoblasts.

Leukemic transformation of immature thymocytes is the result of a multistep process involving numerous genetic abnormalities, which can be associated with different clinical features, including age and prognosis. Notably, *PHF6* mutations were significantly more prevalent in adult T-ALL patients (16/42; 38%) than in pediatric patients (14/89; 16%) ($P = 0.005$; **Fig. 4a**). Detailed genetic information was available for T-ALL cases treated in Dutch Childhood Oncology Group (DCOG) clinical trials ($n = 65$) (**Supplementary Table 5**). In this cohort, *PHF6* mutations were significantly associated with the aberrant expression of *TLX1* and *TLX3* ($P < 0.005$; **Fig. 4b** and **Supplementary Table 5**), two related oncogenes activated by chromosomal translocations in T-ALL[16–18]. No significant associations were observed between *PHF6* mutations and *NOTCH1*, *FBXW7* or *PTEN* mutations in either pediatric ($n = 65$) or adult ($n = 34$) T-ALL cohorts (**Supplementary Tables 5** and **6**). Overall survival in *PHF6* wild-type children with T-ALL treated on DCOG protocols[19] was 65% (33/51) vs. 71% (10/14) for *PHF6*-mutated cases (log-rank $P = 0.71$) (**Fig. 4c**). Overall survival in *PHF6* wild-type adults with T-ALL treated in the Eastern Cooperative Oncology Group ECOG-2993 clinical trial was 36% (7/12) vs. 58% (8/22) for *PHF6*-mutated samples (log-rank $P = 0.24$) (**Fig. 4d**).

Overall, these results identify *PHF6* as a new X-linked tumor suppressor gene and imply a specific interaction between the oncogenic programs activated by aberrant expression of TLX transcription factors and mutational loss of *PHF6* in the pathogenesis of T-ALL.

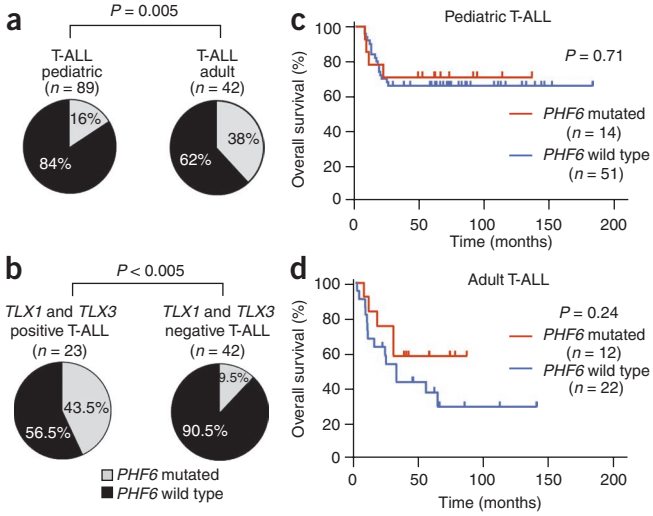


Figure 4 Clinical and biological characteristics associated with *PHF6* mutations in T-cell acute lymphoblastic leukemia (T-ALL). (**a**) Frequencies of *PHF6* mutations in pediatric and adult T-ALL samples. (**b**) Differential distribution of *PHF6* mutations in *TLX1/TLX3*-positive and *TLX1/TLX3*-negative T-ALL samples. (**c**) Kaplan-Meier curve of overall survival in pediatric T-ALL cases from Dutch Childhood Oncology Group trials ALL7, ALL8 and ALL9 with and without *PHF6* mutations. (**d**) Kaplan-Meier survival curve in adult T-ALL cases with and without mutations in *PHF6* treated in Eastern Cooperative Oncology Group clinical trial ECOG2993.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

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

ACKNOWLEDGMENTS

This study was supported by the Fund for Scientific Research (FWO) Flanders (postdoctoral grants to P.V.V. and T.T., PhD grant to J.V.d.M., senior clinical investigator award to B.P. and project grants G.0198.08 and G.0869.10N to F.S.); the GOA-UGent (grant no. 12051203); the IWT-Vlaanderen (SBO grant no. 060848); the Children Cancer Fund Ghent (F.S.); Leukemia Research UK (C.J.H.); the Stichting Kinderen Kankervrij (KiKa; grant no. KiKa 2007-012 to L.Z.); the Belgian Program of Interuniversity Poles of Attraction; the Belgian Foundation against Cancer; the Austrian Ministry of Science and Research (GEN-AU Child, GZ 200.136/1-VI/1/2005 to S.S.), the US National Library of Medicine (1R01LM010140-01 to R.R. and H.K.); the ECOG and DCOG tumor banks; grants from Spain's Plan Nacional (BFU 2007-60990 and PlanE2009-0110 to M.L.T.), Comunidad de Madrid (S-SAL0304-2006 to M.L.T.), Fundación MM (M.L.T.), Instituto de Salud Carlos III (RECAVA RD06/0014/1012 to M.L.T.), an Institutional Grant from the Fundación Ramón Areces (M.L.T.), the Alex's Lemonade Stand Foundation Young Investigator Award (T.P.); a US Northeast Biodefense Center ARRA award (U54-AI057158 to R.R.); the US National Institutes of Health (R01CA120196 and R01CA129382 to A.F.); the Rally across America Foundation (A.F); the Swim across America Foundation (A.F.); the Golfers against Cancer Foundation (A.F.); and a Leukemia and Lymphoma Society Scholar Award (A.F.). We thank the Pediatric Cardiosurgery Units from Centro Especial Ramón y Cajal and Ciudad Sanitaria La Paz (Madrid, Spain) for thymus samples.

AUTHOR CONTRIBUTIONS

P.V.V. performed array-CGH and mutation analysis of *PHF6* and wrote the manuscript. T.P. performed exon capture and next-generation sequencing of T-ALL samples and wrote the manuscript. H.K. analyzed next-generation sequencing data. J.V.d.M. performed additional array-CGH analysis and *PHF6* mutation screening in T-ALL and BCP-ALL samples. T.T., N.V.R. and A.W.L. performed experiments. M.C. and C.C.-C. performed and analyzed histological and immunohistochemical staining. J.P. collaborated on *PHF6* mutation screening

in BCP-ALL samples. C.J.H. and C.S. collaborated on additional screening for genomic *PHF6* deletions in T-ALL. Y.B., B.D.M. and B.C. collaborated on the *PHF6* mutation screening. R.P., M.P., S.S. and J.S. collaborated on the multicenter array-CGH study. S.G.-G. and M.L.T. performed the isolation of T-cell progenitor cells for expression analysis of *PHF6*. X.Y. performed survival analysis of ECOG T-ALL patients. J.G. provided critical reagents and discussion. E.S. provided samples and correlative clinical data from DCOG. E.P., J.M.R. and P.H.W. provided samples and correlative clinical data from ECOG. J.M. and L.Z. collaborated on the multicenter array-CGH study and *PHF6* mutation analysis, provided molecular data on the characterization of T-ALL and performed survival analysis of *PHF6* mutations in the DCOG series. R.R. designed and directed the analysis of next-generation sequencing results. F.S. and B.P. designed the studies and directed research. A.F. designed the studies, directed research and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Carrel, L., Cottle, A.A., Goglin, K.C. & Willard, H.F. A first-generation X-inactivation profile of the human X chromosome. *Proc. Natl. Acad. Sci. USA* **96**, 14440–14444 (1999).
2. Carrel, L. & Willard, H.F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400–404 (2005).
3. Goldberg, J.M. *et al.* Childhood T-cell acute lymphoblastic leukemia: the Dana-Farber Cancer Institute acute lymphoblastic leukemia consortium experience. *J. Clin. Oncol.* **21**, 3616–3622 (2003).
4. Aifantis, I., Raetz, E. & Buonamici, S. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nat. Rev. Immunol.* **8**, 380–390 (2008).
5. Pui, C.H., Robison, L.L. & Look, A.T. Acute lymphoblastic leukaemia. *Lancet* **371**, 1030–1043 (2008).
6. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).
7. Lower, K.M. *et al.* Mutations in *PHF6* are associated with Borjeson-Forssman-Lehmann syndrome. *Nat. Genet.* **32**, 661–665 (2002).
8. Borjeson, M., Forssman, H. & Lehmann, O. An X-linked, recessively inherited syndrome characterized by grave mental deficiency, epilepsy, and endocrine disorder. *Acta Med. Scand.* **171**, 13–21 (1962).
9. Turner, G. *et al.* The clinical picture of the Borjeson-Forssman-Lehmann syndrome in males and heterozygous females with *PHF6* mutations. *Clin. Genet.* **65**, 226–232 (2004).
10. Baker, L.A., Allis, C.D. & Wang, G.G. PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat. Res.* **647**, 3–12 (2008).
11. Dephoure, N. *et al.* A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci. USA* **105**, 10762–10767 (2008).
12. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160–1166 (2007).
13. Lowndes, N.F. & Toh, G.W. DNA repair: the importance of phosphorylating histone H2AX. *Curr. Biol.* **15**, R99–R102 (2005).
14. Payer, B. & Lee, J.T. X chromosome dosage compensation: how mammals keep the balance. *Annu. Rev. Genet.* **42**, 733–772 (2008).
15. Carrel, L. & Willard, H.F. Heterogeneous gene expression from the inactive X chromosome: an X-linked gene that escapes X inactivation in some human cell lines but is inactivated in others. *Proc. Natl. Acad. Sci. USA* **96**, 7364–7369 (1999).
16. Ferrando, A.A. *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1**, 75–87 (2002).
17. Soulier, J. *et al.* HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). *Blood* **106**, 274–286 (2005).
18. Van Vlierberghe, P. *et al.* The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia. *Blood* **111**, 4668–4680 (2008).
19. van Grotel, M. *et al.* The outcome of molecular-cytogenetic subgroups in pediatric T-cell acute lymphoblastic leukemia: a retrospective study of patients treated according to DCOG or COALL protocols. *Haematologica* **91**, 1212–1221 (2006).

ONLINE METHODS

Clinical samples and cell lines. Leukemic DNA and cryopreserved lymphoblast samples were provided by collaborating institutions in the United States (ECOG), The Netherlands (DCOG), France (Hôpital Saint-Louis, Paris), Austria (Children's Cancer Research Institute, St. Anna Kinderkrebsforschung, Vienna) and Belgium (Department of Pediatric Hemato-Oncology, Ghent University Hospital, Ghent; Department of Hematology, Hospital St.-Jan, Bruges). All samples were collected in clinical trials with informed consent and under the supervision of local institutional review board committees. The subjects' parents or their legal guardians provided informed consent to use leftover material for research purposes according to the Declaration of Helsinki. T-cell phenotype was confirmed by flow cytometry. Survival analysis was performed in pediatric T-ALL samples from DCOG trials ALL7, ALL8 and ALL9 (ref. 19) and from adults with T-ALL treated in the ECOG2993 clinical trial[20].

Jurkat and PF382 cells were obtained from the American Type Culture Collection. The ALL-SIL, HPB-ALL and T-ALL-1 cell lines were from the DSMZ repository (the German national resource center for biological material). The DND41 cell line was a gift from A.T. Look (Dana-Farber Cancer Institute). T-ALL cell lines were cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum, 100 U/ml penicillin G and 100 μg/ml streptomycin at 37 °C in a humidified atmosphere under 5% CO_2.

X chromosome exome capture and next-generation sequencing. Libraries of synthetic biotinylated RNA oligonucleotides (baits) targeting the X-chromosome exons were obtained from Agilent Technologies. The targeted region includes 5,217 exons for a total of 3 megabases and is designed to capture 85% of the exons on the human X chromosome. Fragment libraries using 2–4 μg of genomic DNA as starting material were prepared following the SOLiD standard library preparation protocol with some modifications, including the use of shortened adaptors and a precapture amplification for six cycles. A total of 500 ng of captured library were hybridized with the baits for 42 h, washed and eluted using the protocol provided by Agilent. The resulting captured DNA was amplified using Herculase II Fusion DNA polymerase (Agilent) for 18 cycles. Enrichment in the targeted regions was calculated by real-time PCR quantification of single exons within four X-chromosome loci (*ARSF*, *OTC*, *NAP1L3* and *SOX3*), which showed an average enrichment of 400- to 1,200-fold for the different loci across the 12 different samples. After library quantification by real-time PCR, the amplified captured libraries were subjected to emulsion PCR and sequenced following standard SOLiD 3 protocols by depositing 10–15 million beads per sample using an eight-region mask.

Data analysis SOLiD3 ultradeep sequencing. A reference genome of the captured regions was created based on the March 2006 human reference sequence (hg18). To map the sequence data into this reference genome, we used the SHRiMP algorithm with its default parameters[21]. SOLiD platform employs a two-base encoding system, where a single variation in the color space solely indicates a sequencing error and two consecutive variations in the color space point to a base change in the nucleotide-space. In our analysis we included only the reads with a maximum number of two color-space mismatches that are also uniquely mapped to the reference genome. An average 90.1% of the reference genome is covered in the 12 samples, where the mean depth is 42 per base. Less restrictive filtering increases the false-positive rate of candidate genomic variants without improving the coverage to any great extent. We found 66 candidates of exonic nonsynonymous single-nucleotide variation by requiring each variation to be reported in a minimum 75% of at least three reads mapping to its position. These candidates exclude the previously reported SNPs in the human genome. Using ParMap, an algorithm specifically developed to identify small deletions and insertions (along with their nucleotide sequence) through statistical analysis of partially mapped reads[22], we identified seven candidates of such complex variations. Validation of the next-generation sequencing results was done by Sanger sequencing of PCR amplified exons. Overall, 89% of all previously uncharacterized candidate variants were confirmed.

Microarray-based comparative genomic hybridization. Analysis of X-chromosome array-CGH data was performed in a multicenter setting. Depending on the institution of origin, array-CGH analysis was performed using an oligonucleotide array-CGH platform[18,23] (Agilent) or a tiling path BAC array-CGH platform[24]. To determine the exact size of the recurrent Xq26 deletions, we reanalyzed all eight del(X)(q26)-positive T-ALL cases using a custom high-resolution X-chromosome oligonucleotide array with an average resolution of 3 kb according to the manufacturer's instructions (Agilent). Slides were scanned in a 2565AA DNA microarray scanner (Agilent). Microarray images were analyzed using Feature Extraction software (Agilent), and the data were subsequently imported into array-CGH Analytics software (Agilent).

Real-time quantification of DNA copy number. Chromosome Xq26.3 deletions were confirmed with real-time quantitative DNA PCR using the Fast SYBR Green Master Mix (Applied Biosystems) and the LightCycler 480 Real-Time PCR System (Roche Diagnostics) as described[25] using *TIE2* as control gene. Data were analyzed using the comparative ΔΔCT method (Applied Biosystems). The primers used for the quantitative PCR analysis of the *PHF6* locus are shown in **Supplementary Table 7**.

***PHF6* mutation analysis.** *PHF6* mutations were analyzed by PCR amplification of *PHF6* exons 2–10 followed by direct bidirectional DNA sequencing. The primers used for *PHF6* mutation detection are summarized in **Supplementary Table 7**.

Protein blot and immunohistochemistry. Protein blot analysis was performed using a rabbit polyclonal antibody specific to PHF6 (1:10,000; Novus Biologicals) recognizing a C-terminal (amino acids 315–365) epitope in PHF6; a mouse monoclonal antibody (1:1,000) recognizing an N-terminal (amino acids 120–140) PHF6 epitope[26]; an antibody specific to γ-H2AX (1:1,000; Cell Signaling Technologies) and a mouse monoclonal antibody specific to GAPD (1:1000; Santa Cruz Biotechnology) using standard procedures.

Immunohistochemistry analysis was performed as described following the standard avidin-biotin immunoperoxidase staining procedure using an N-terminal antibody to PHF6. Briefly, PHF6 immunostaining of formalin-fixed paraffin-embedded tissue sections was performed after heat-induced epitope retrieval in a microwave in citrate buffer, pH 6.0. Subsequently, sections were incubated in 10% normal goat serum for 30 min, followed by primary antibody incubation (amino acids 1–94 of rabbit polyclonal antibody to PHF6; Sigma Prestige Antibodies, dilution 1:100) overnight at 4 °C. Next, slides were incubated with biotinylated immunoglobulins specific to rabbit at a 1:1000 dilution (Vector Laboratories) for 30 min, followed by avidin-biotin peroxidase complexes at a 1:25 dilution (Vector Laboratories) for 30 min. Diaminobenzidine was used as the chromogen and hematoxylin as a nuclear counterstain.

Quantitative real-time PCR. Thymocyte populations were isolated from human thymi as described before[27]. Total RNA was extracted using the Trizol method (Invitrogen) following the manufacturer's instructions. Total RNA from 20 different normal human tissues was obtained from the FirstChoice Human Total RNA Survey Panel (Applied Biosystems). Complementary DNA (cDNA) was generated with the ThermoScript RT-PCR system (Invitrogen) and analyzed by quantitative real-time PCR using the SYBR Green RT-PCR Core Reagents kit (Applied Biosystems) and the 7300 Real-Time PCR System (Applied Biosystems). *PHF6* expression levels were calculated using *GAPDH* as a reference gene. Primers used for *PHF6* expression analysis are shown in **Supplementary Table 7**.

SNP genotyping. Genotyping of SNP rs17317724 located in the 3′ UTR region of *PHF6* was performed using a TaqMan SNP Genotyping Assay (Assay ID C__34812972_10, Applied Biosystems) according to manufacturer's instructions in DNA samples from females with wild-type *PHF6* T-ALL. Genotyping was confirmed by direct DNA sequencing of PCR products encompassing the 3′ UTR of *PHF6*. Allelic expression analysis was performed in cDNA samples from heterozygous females with T-ALL so as to evaluate monoallelic and biallelic *PHF6* expression. Before cDNA synthesis, RNA samples were treated with DNase I using the DNA-free DNase Treatment kit (Applied Biosystems) to remove any traces of genomic DNA. The *PHF6* 3′ UTR-specific primers used for amplification of SNP rs17317724 are also summarized in **Supplementary Table 7**.

***PHF6* shRNA knockdown.** We produced lentiviral particles driving the expression of a shRNA directed against *PHF6* (target sequence CAGAATTTGGAGACTTTGA) using the pGIPZ Lentiviral shRNAmir vector system (V2LHS_138602, Open Biosystems) as described[28]. We infected HEK293T cells with viral supernatants generated with pGIPZ PHF6 or an empty pGIPZ control using spinoculation in the presence of polybrene. PHF6 knockdown was evaluated by protein blot analysis at 72 h as described above.

Statistical analysis. Fisher's exact test was used to compare the frequency of *PHF6* mutations between clinical and genetic groups of T-ALL. Bar graphs represent mean values ± s.e.m. Therapeutic outcome was analyzed in leukemia patients treated in DCOG trials ALL7, ALL8 and ALL9 and in ECOG trial 2993 according to overall survival. Kaplan-Meier curves were used to assess survival, and differences between groups were compared by the log-rank test.

20. Marks, D.I. *et al.* T-cell acute lymphoblastic leukemia in adults: clinical features, immunophenotype, cytogenetics, and outcome from the large randomized prospective trial (UKALL XII/ECOG 2993). *Blood* **114**, 5136–5145 (2009).
21. Rumble, S.M. *et al.* SHRiMP: accurate mapping of short color-space reads. *PLOS Comput. Biol.* **5**, e1000386 (2009).
22. Khiabanian, H., Van Vlierberghe, P., Palomero, T., Ferrando, A.A. & Rabadan, R. ParMap, an algorithm for the identification of complex genomic variations in nextgen sequencing data. *Nature Precedings* published online, <http://hdl.handle.net/10101/npre.2010.4145.1> (12 January 2010).
23. Clappier, E. *et al.* The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. *Blood* **110**, 1251–1261 (2007).
24. Erdogan, F. *et al.* Impact of low copy repeats on the generation of balanced and unbalanced chromosomal aberrations in mental retardation. *Cytogenet. Genome Res.* **115**, 247–253 (2006).
25. Lahortiga, I. *et al.* Duplication of the *MYB* oncogene in T cell acute lymphoblastic leukemia. *Nat. Genet.* **39**, 593–595 (2007).
26. Voss, A.K. *et al.* Protein and gene expression analysis of *Phf6*, the gene mutated in the Borjeson-Forssman-Lehmann syndrome of intellectual disability and obesity. *Gene Expr. Patterns* **7**, 858–871 (2007).
27. González-García, S. *et al.* CSL-MAML-dependent Notch1 signaling controls T lineage-specific IL-7Rα gene expression in early human thymopoiesis and leukemia. *J. Exp. Med.* **206**, 779–791 (2009).
28. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).

Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing

A. Eliot Shearer[a,b,1], Adam P. DeLuca[c,d,1], Michael S. Hildebrand[a,1], Kyle R. Taylor[c,d], José Gurrola II[a], Steve Scherer[e,2], Todd E. Scheetz[c,d,f,2], and Richard J. H. Smith[a,b,g,2,3]

[a]Department of Otolaryngology, Head and Neck Surgery, University of Iowa, Iowa City, IA 52242; [b]Department of Molecular Physiology and Biophysics, University of Iowa Carver College of Medicine, Iowa City, IA 52242; [c]Department of Biomedical Engineering, University of Iowa, Iowa City, IA 52242; [d]Center for Bioinformatics and Computational Biology, University of Iowa, Iowa City, IA 52242; [e]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; [f]Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA 52242; and [g]Interdepartmental PhD Program in Genetics, University of Iowa, Iowa City, IA 52242

Edited* by Mary-Claire King, University of Washington, Seattle, WA, and approved October 14, 2010 (received for review August 31, 2010)

The extreme genetic heterogeneity of nonsyndromic hearing loss (NSHL) makes genetic diagnosis expensive and time consuming using available methods. To assess the feasibility of target-enrichment and massively parallel sequencing technologies to interrogate all exons of all genes implicated in NSHL, we tested nine patients diagnosed with hearing loss. Solid-phase (NimbleGen) or solution-based (SureSelect) sequence capture, followed by 454 or Illumina sequencing, respectively, were compared. Sequencing reads were mapped using GSMAPPER, BFAST, and BOWTIE, and pathogenic variants were identified using a custom-variant calling and annotation pipeline (ASAP) that incorporates publicly available in silico pathogenicity prediction tools (SIFT, BLOSUM, Polyphen2, and Align-GVGD). Samples included one negative control, three positive controls (one biological replicate), and six unknowns (10 samples total), in which we genotyped 605 single nucleotide polymorphisms (SNPs) by Sanger sequencing to measure sensitivity and specificity for SureSelect-Illumina and NimbleGen-454 methods at saturating sequence coverage. Causative mutations were identified in the positive controls but not in the negative control. In five of six idiopathic hearing loss patients we identified the pathogenic mutation. Massively parallel sequencing technologies provide sensitivity, specificity, and reproducibility at levels sufficient to perform genetic diagnosis of hearing loss.

deafness | genomics | Usher syndrome | diagnostics | next-generation sequencing

Hereditary sensorineural hearing loss (SNHL) is the most common sensory impairment in humans (1, 2). In developed countries, two-thirds of prelingual-onset SNHL is estimated to have a genetic etiology, of which ~70% is nonsyndromic hearing loss (NSHL). Eighty percent of NSHL is autosomal recessive nonsyndromic hearing loss (ARNSHL), ~20% is autosomal dominant (AD), and the remainder is composed of X-linked and mitochondrial forms (1, 3). To date, 134 deafness loci have been identified, and 32 recessive (DFNB), 23 dominant (DFNA) and 2 X-linked (DFNX) genes have been cloned; 8 genes are associated with both ARNSHL and ADNSHL (4).

Establishing a genetic diagnosis of NSHL is a critical component of the clinical evaluation of deaf and hard-of-hearing persons and their families. If a genetic cause of hearing loss is determined, it is possible to provide families with prognostic information, recurrence risks, and improved habilitation options. For persons diagnosed with Usher syndrome, preventative measures including sunlight protection and vitamin therapy can be implemented to minimize the rate of progression of retinitis pigmentosa (5). Most current genetic testing strategies for NSHL rely on a gene-specific Sanger sequencing approach. Because mutations in a single gene, *GJB2* (DFNB1), account for up to 50% of ARNSHL in many world populations (6), this approach has changed the evaluation of patients with presumed ARNSHL. However, the mutation frequency in other genes in persons with NSHL in outbred populations is unknown, making sequential gene screening problematic (7). The extreme heterogeneity of NSHL also makes serial sequencing approaches unfavorable in terms of efficiency and cost.

The advent of new technologies that target and enrich specific regions of the genome coupled with massively parallel sequencing offers an alternative approach to genetic testing for deafness. Although genetic diagnoses can also be made by whole genome sequencing (8) or targeted sequence capture of the entire exome (9, 10), these approaches are expensive, and time-consuming data analysis is required. Therefore, our aim was to develop and test a streamlined, comprehensive genetic diagnostic platform that targets only the 0.014% of the genome currently associated with NSHL.

Results

We elected to compare the two most widely used target enrichment approaches (NimbleGen solid-phase and SureSelect solution-based sequence capture) and two massively parallel sequencing technologies (454 GS FLX pyrosequencing and Illumina GA_{II} cyclic reversible termination sequencing) using genomic DNA from NSHL families. The optimized platform and diagnostic pipeline, which we refer to as OtoSCOPE (otologic sequence capture of pathogenic exons), targets the exons of all 54 known deafness genes (Table S1). The subjects were nine individuals with presumed NSHL (Table 1). Sanger sequencing-based genetic testing had been completed for *GJB2* and *SLC26A4* in all autosomal recessive (AR) cases. Positive controls had been genetically diagnosed by Sanger sequencing prior to this study.

We performed array and solution-based targeted capture of all exons of the 54 genes known to cause NSHL, also including Usher syndrome genes because in infants and children, Usher syndrome is not readily distinguishable from ARNSHL (*Materials and Methods* and Table S1). Some requested target regions were not covered by the NimbleGen and SureSelect designs (9.3% and 8.3%, respectively; Table 2) due to repetitive regions. These percentages are consistent with other reports focused on exon sequence capture (11, 12). However, the proportion of protein coding regions included in the SureSelect bait design was comparatively better than that covered by NimbleGen (97.7 and 93.6%, respectively; Table 2), reflecting different methods for defining repetitive regions and complementary oligonucleotide selection.

Author contributions: A.E.S., M.S.H., S.S., T.E.S., and R.J.H.S. designed research; A.E.S., A.P.D., M.S.H., and K.R.T. performed research; A.P.D., K.R.T., J.G., S.S., and T.E.S. contributed new reagents/analytic tools; A.E.S., A.P.D., M.S.H., K.R.T., J.G., T.E.S., and R.J.H.S. analyzed data; and A.E.S., M.S.H., and R.J.H.S. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]A.E.S., A.P.D., and M.S.H. contributed equally to this work.

[2]S.S., T.E.S., and R.J.H.S. contributed equally to this work.

[3]To whom correspondence should be addressed. E-mail: richard-smith@uiowa.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1012989107/-/DCSupplemental.

Table 1. Samples studied and diagnostic sequencing results by allele

Sample	Status	Hearing loss inheritance	Candidate variants	Gene	Nucleotide change	Protein change	Genomic coordinates (Hg19)	Sequencing reads–variant reads/total coverage (%)	Status
1	Positive control	ADNSHL	3	*COCH*	c.151C > T	p.P51S*	Chr14: 31,346,846	463/902 (51.3)	Pos Ctrl mutation
				EYA4	c.390C > A	p.Tyr130X	Chr6: 133,782,271	234/673 (34.77)	RO-segregation
				OTOF	c.4484G > C	p.R1495P	Chr2: 26,689,598	639/1,266 (50.5)	RO-segregation
2	Positive control	ARNSHL	1	*GJB2*	c.109G > A	p.V37I†	Chr13: 20,763,612	1,075/1,078 (99.7)	Pos Ctrl mutation
				GJB2	c.109G > A	p.V37I	Chr13: 20,763,612	1,075/1,078 (99.7)	Pos Ctrl mutation
3	Replicate of 2	ARNSHL	1	*GJB2*	c.109G > A	p.V37I	Chr13: 20,763,612	1,113/1,114 (99.9)	Pos Ctrl mutation
				GJB2	c.109G > A	p.V37I	Chr13: 20,763,612	1,113/1,114 (99.9)	Pos Ctrl mutation
4	Negative control	n/a	0	—	—	—	—	—	—
5	Idiopathic SNHL	ARNSHL	1	*STRC*	c.4057C > T	p.Q1353X	Chr15: 43,896,918	44/47 (93.6)	Causative mutation
				STRC	~100-kb deletion	—	Chr15: q15	—¶	Causative mutation
6	Idiopathic SNHL	ADNSHL	2	*MYO6*	c.862_865delACAA	p.D288DfsX17	Chr6: 76,554,660–76,554,663	109/306 (35.6)	Causative mutation
				MYO6	c.3667G > A	p.D1223N	Chr6: 76,624,538	427/890 (48.0)	RO-segregation
7	Idiopathic SNHL	ADNSHL	4	*MYO7A*	c.5156A > G	p.Y1719C	Chr11: 76,913,457	439/908 (48.3)	RO-segregation
				PCDH15	c.1039C > T	p.L347F	Chr10: 55,973,755	215/459 (46.8)	RO-segregation
				USH2A	c.6713A > C	p.E2238A	Chr1: 216,166,454	372/752 (49.47)	RO-segregation
				OTOF	c.3751T > G	p.C1251G	Chr2: 26,695,500	706/720 (98.1)	RO-segregation
8	Idiopathic SNHL	ADNSHL	2	*KCNQ4*	c.842T > C	p.L281S‡	Chr1: 41,285,554	348/893 (39.0)	Causative mutation
				WFS1	c.2327A > T	p.E776V	Chr4: 6,303,849	161/321 (50.1)	RO-segregation
9	Idiopathic SNHL	ADNSHL	2	*MYH14*	c.5893G > T	p.E1965X	Chr19: 50,812,952	587/1,255 (46.8)	Causative mutation
				MYH14	c.4193A > G	p.E1398G	Chr19: 50,784,876	413/1,334 (31.0)	RO-controls
10	Idiopathic SNHL	ARNSHL	2	*CDH23*	c.1096G > A	p.A366T§	Chr10: 73,377,112	2,233/4,529 (49.3)	Causative mutation
				CDH23	c.0.3293A > G	p.N1098S§	Chr10: 73,472,494	2,708/5,445 (49.7)	Causative mutation

Candidate variants follow inheritance pattern (if known) and are predicted deleterious by Polyphen2, SIFT, BLOSUM62, and align-GVGD. Candidate mutations were ruled out or in by segregation analysis and/or presence in controls. Causative mutations are predicted damaging, segregate with hearing loss in family, are not found in controls, and/or are known to cause hearing loss. NSHL, nonsyndromic hearing loss; ADNSHL, autosomal dominant nonsyndromic hearing loss; ARNSHL, autosomal recessive nonsyndromic hearing loss; Pos Ctrl, positive control; RO-segregation, mutation ruled out by segregation analysis; RO-controls, mutation ruled out by controls analysis.

*Previously reported (16).

†Previously reported (17).

‡Previously reported (18).

§Previously reported (19).

¶Deletion determined by sequencing reads analysis, confirmed by PCR and array-CGH, previously reported (ref. 20 and Fig. 1).

Table 2. Sequence capture performance results

	NimbleGen-454		SureSelect-Illumina	
	All regions (%)	Protein coding regions (%)	All regions (%)	Protein coding regions
Number of bases requested*	421,741 bp	187,017 bp	421,741 bp	187,017 bp
Bases covered by complementary oligonucleotides (%)	386,880 bp (91.7)	266,947 bp (93.6)	382,627 bp (90.7)	182,749 bp (97.7)
Mean sequence coverage in bases (%)	406,756 bp (96.4)	180,489 bp (96.5)	401,252 bp (95.1)	182,727 bp (97.7)
Mean sequence coverage at variant calling threshold in bases (%)†	403,873 bp (95.8)	179,468 bp (96.0)	385,101 bp (91.3)	178,230 bp (95.3)

Values given are a mean of all samples for each method.
*The same bases were requested for targeting using either NimbleGen or SureSelect and default repetitive regions were avoided in both methods.
†Variant calling threshold: for Illumina-SureSelect, 40× coverage and >30% of reads; for 454, ≥3× reads (multidirectional) and >30% of reads.

We paired NimbleGen solid phase sequence capture with 454 GS FLX pyrosequencing (NimbleGen-454 method) and SureSelect solution-based sequence capture with Illumina GAII sequencing (SureSelect-Illumina method) to determine the efficacy of these approaches for clinical diagnostics. Samples 1 and 2 were sequenced using both the NimbleGen-454 and SureSelect-Illumina to provide direct comparison of these methods, whereas samples 3–10 were examined only using SureSelect-Illumina.

The results of our study show that due to the disparity in sequence output between the two sequencing platforms (13), the overall sequence depth of coverage was on average 13-fold higher for the SureSelect-Illumina method as compared with the NimbleGen-454 method (903× and 71× depth of coverage, respectively (*Materials and Methods*); Table 3). However the percent of on-target reads (i.e., capture efficiency) was significantly lower ($P < 0.0001$) for the SureSelect-Illumina method (average 19.6%) than the NimbleGen-454 method (average 64.2%), which may reflect PCR-induced bias during Illumina library preparation, small target size, and sequencer oversaturation (Tables 2 and 3).

As a threshold for variant detection, we required ≥3 reads (multidirectional) for 454 and ≥40 reads for Illumina, with the added stipulation that the variant had to be present in at least 30% of reads to be considered high quality. Similar thresholds have been used in other studies (14) and reflect the need for high sensitivity in diagnostic testing (false positives are preferred over false negatives) (15). At these thresholds, 96.0% and 95.3% of protein coding bases were covered by the NimbleGen-454 and SureSelect-Illumina platforms, respectively (Table 2).

To validate the platforms as diagnostic tests we genotyped 605 highly heterozygous SNPs (average heterozygosity, 0.46; average 55 SNPs per sample) in the targeted regions by Sanger sequencing. Homozygous-alternative allele calls ($n = 149$; average 14 per sample) and heterozygous allele calls ($n = 199$; average 18 per sample) were used as true positive variants; homozygous-reference allele calls ($n = 257$; average 23 per sample) were used as true negative variants (simulated nonvariants). Variants not covered at variant-calling threshold for either platform were considered false negatives.

With both the NimbleGen-454 and SureSelect-Illumina methods, we identified the causative mutations in samples 1 and 2. Specificity was ∼98% for NimbleGen-454 and >99% for SureSelect-Illumina (Table 3). For the SureSelect-Illumina method, one false negative was found across all 10 samples (sensitivity 99.72%) reflecting high coverage depth of targeted regions. For the NimbleGen-454 method, a mean depth of coverage of 71× was associated with an increased false-negative rate (average two per sample) and a decrease in sensitivity (93.98%). This difference may be attributable to a relatively lower quality depth-of-coverage threshold for the NimbleGen-454 method (3× versus 40× for SureSelect-Illumina) as mean sequence coverage for protein-coding sequence was similar.

On the basis of data from samples 1 and 2, we analyzed the remaining eight samples using the SureSelect-Illumina method and a sequence analysis protocol we developed (*Materials and Methods*). In total, we detected causative mutations in the three positive controls and in five of six persons with idiopathic hearing loss (Table 1). Analysis of controls (samples 1–4) was completed in a blinded fashion with only the mode of inheritance known. Sample 1 was from a family segregating ADNSHL. Of three candidate variants, the known hearing loss mutation in *COCH* (c.151C > T, p.P51S, rs28938175; DFNA9) was the only variant that segregated with the phenotype in this family (16). In samples 2 and 3 (biological replicates), the only candidate variant determined in each sample was the known deafness mutation in *GJB2* (c.109G > A, p.V37I; DFNB1) (17). In sample 4, the negative control, none of the nine nonsynonymous/splice site/indel variations identified was predicted to be pathogenic.

Of the six persons with idiopathic hearing loss, two were from families segregating ARNSHL. Sample 5 was from a person with profound SNHL who was shown to carry a variant of unknown significance (VUS) in *STRC* (c.4057C > T, p.Q1353×; rs2614824; DFNB16) in 44/47 Illumina reads. This variant was not observed

Table 3. Sequencing results

Sequencing method	Capture method	Total sequencing reads	Uniquely mapping reads	% of uniquely mapping reads	Uniquely mapping reads on target (<1KB)	Mean coverage of targeted regions	Sens/spec	FN/FP
454	NimbleGen solid-phase	303,552	293,368	96.7%	194,155	71 X	93.98%/97.92%	2/0
Illumina	Agilent solution-based	41,234,307	33,693,349	81.3%	6,941,755	903 X	99.72%/>99%	0*/0

Illumina sequencing was performed using one sample per flow cell channel; 454 sequencing was performed using one sample per quarter of a four-way gasket. Values given are the mean of all samples for a given method: Illumina, 8 samples; 454, 2 samples. Sens, sensitivity; spec, specificity; FN, false negatives; FP, false positives.
*One FN was found across all ten samples.

in 100 ethnically matched controls (200 chromosomes). A hemizygous deletion of ~100 kb involving the *STRC*–ψ*STRC* region was also detected (Fig. 1) and independently verified by PCR and array-comparative genomic hybridization (CGH). Homozygosity for this contiguous gene deletion is known to cause autosomal recessive deafness–infertility syndrome (20), making this person a compound heterozygote for a novel point mutation in *STRC in trans* with a large contiguous gene deletion that includes *STRC* (DFNB16). Sample 10, from a 4-y-old cochlear implant recipient with no family history of hearing loss, was compound heterozygous for known USH1D mutations in *CDH23* (c.1096G > A, p.A366T; c.3293A > G, p.N1098S) (19).

The remaining four samples were from families segregating ADNSHL. For these samples, we also used, when possible, the computer program AudioGene (21) as a phenotypic filter to predict probable ADNSHL genotypes on the basis of audiometric criteria through audioprofiling. Sample 6 carried two candidate variants in *MYO6* (DFNA22). The missense mutation was ruled out by segregation analysis; however, the 4-bp deletion (c.862_865delACAA, p.D288DfsX17) segregated with the phenotype and was not found in 135 ethnically matched controls (270 chromosomes). Because the DFNA22 locus is not represented on AudioGene, phenotypic filtering was not applied; however, the phenotype was consistent with reported DFNA22 audioprofiles (22). Sample 8 carried two candidate variants—one mutation was ruled out by segregation analysis and the other is a known ADNSHL mutation [(DFNA2) c.842C > T, p.L281S] in *KCNQ4* (18). This finding was consistent with AudioGene analysis, which predicted DFNA2 as the most likely cause of hearing loss in this family. Sample 9 carried two candidate variants—both detected in *MYH14*—that segregated with the phenotype in the extended family. One mutation was ruled out by analysis of controls, whereas the other, a stop mutation (c.5893G > T, p.E1965×), was not found in 119 ethnically matched controls (238 chromosomes) and was consistent with the predicted DFNA4 audioprofile. Four candidate variants were detected in sample 7, all of which were ruled out by segregation analysis. AudioGene analysis predicted *KCNQ4* (DFNA2) as the most likely cause of hearing loss in this family; however, no variants in *KCNQ4* were identified in this analysis or by prior Sanger sequencing. This finding suggests that this family may segregate a novel genetic cause of ADNSHL, which is possible because causative genes have not been identified for more than half of the mapped ADNSHL loci (4).

Discussion

In this study, we identified NSHL mutations, three of which have not been reported, in eight of nine persons tested. Our sequencing results are consistent with other studies reporting utility of these technologies for diagnosis of other genetic diseases (11, 15, 23) and suggest that massively parallel sequencing is suitable for genetic testing of NSHL. A key concern for any diagnostic test is sensitivity, as it is critical that pathogenic mutations are not missed. This requirement could be viewed as a potential limitation of target-enrichment methodologies, as a significant portion of targeted bases in repetitive regions cannot be captured (Table 2). However, this limitation must be weighed against the decreased cost and time in simultaneously sequencing a large number of genes.

Multiplexing samples offers the opportunity to increase throughput, but to maintain read depth, capture efficiency must be increased (24, 25). Scalability and cost also must be considered for large-scale sequencing projects. Our results show that when adequate target coverage and depth of coverage are maintained, both the SureSelect-Illumina and NimbleGen-454 platforms offer relatively high specificity and sensitivity. However, the SureSelect-Illumina method is superior in terms of scalability, cost, and increased sensitivity. Our results compare very favorably with recent reports of high-throughput diagnostic tests for NSHL that rely on primer extension arrays (26) or targeted resequencing arrays (27), which are appealing due to low cost but are limited in terms of capacity (not all NSHL genes can be screened simultaneously) and therefore sensitivity.

In summary, we have demonstrated that OtoSCOPE has the potential to improve the efficiency of genetic testing for NSHL and Usher syndrome. Our results show that targeted capture plus massively parallel sequencing has a sensitivity and specificity comparable to Sanger sequencing. Comprehensive genetic screening for deafness using platforms like OtoSCOPE would allow clinicians to improve patient care by providing prognostic information and genetic counseling, and in cases like Usher syndrome, offer families preventative strategies to minimize the rate of progression of retinitis pigmentosa. Some of the novel habilitation options under

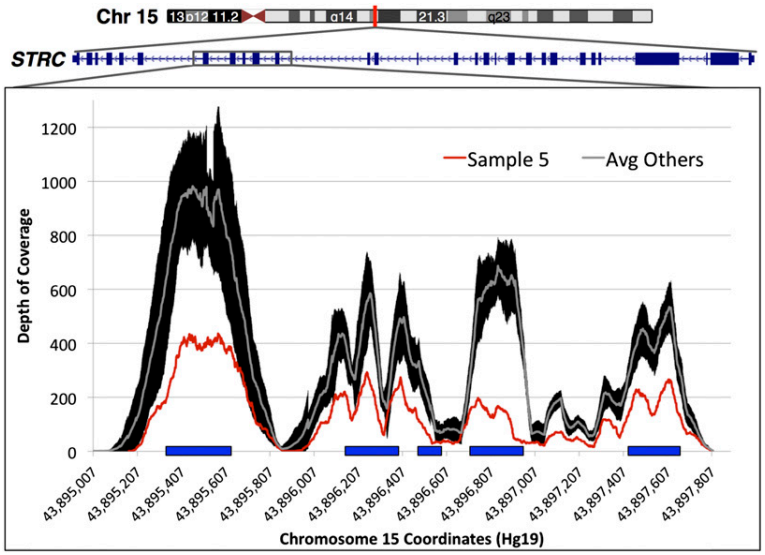


Fig. 1. Deletion analysis of sample 5 using massively parallel sequencing. Location on chromosome 15 is shown, with the highlighted region containing *STRC*; five exons are indicated by blue bars on *x* axis. Gray line is average sequencing depth of coverage for nine samples (all samples excluding sample 5); thick black line represents SD for these samples. Red line is depth of coverage for sample 5.

Table 4. Variant prioritization

Method	NimbleGen-454	SureSelect-Illumina
Variations in targeted genes at calling threshold*	1,008 (998–1,008)	876 (732–1112)
Variations with <1% or unknown allele frequency†	337 (335–339)	404 (297–507)
Synonymous variations	22 (17–26)	22 (18–25)
NS/SS/Indel variations	17 (15–18)	32 (21–40)
Individual-unique NS/SS/Indel variations‡	12 (9–14)	8 (3–14)
Predicted pathogenic variations	3 (2–4)	2 (0–3)
Candidate variants	2 (2)	2 (0–3)

Data given are means (range) (refer to Table S2 for detailed analysis). NS, nonsynonymous; SS, splice site.
*Variant calling threshold included only variants seen in 30% of reads and 40× Illumina depth of coverage or 3× multidirectional reads for 454.
†Allele frequency data taken from dbSNP130.
‡Individual unique variants excluded variants found in more than one sample in this study.

development to slow progression of hearing loss are also gene and even mutation specific (28), suggesting that comprehensive genetic testing will be an integral part of the care of deaf and hard-of-hearing patients in the future.

Materials and Methods

Patients. The subjects were nine individuals with presumed NSHL (Table 1) who provided informed consent for this study approved by the University of Iowa International Review Board. Detailed family histories, clinical evaluations, and audiograms were available for each patient.

Targeted Capture and DNA Sequencing. Genomic DNA (gDNA) was extracted from whole blood using standard procedures (29), quality assessed on an agarose gel and spectrophotometer, and quantified using the Qubit (Invitrogen) system. Three or 5 μg of gDNA were used for the SureSelect or NimbleGen capture methods, respectively.

Exons in all isoforms of the 54 known deafness genes were identified in the RefSeq and Ensembl databases using the University of California Santa Cruz table browser (http://genome.ucsc.edu/). An additional 50 bp of flanking intronic sequence were added to each exon and genomic intervals were merged using Galaxy software (http://galaxy.psu.edu). In total, we targeted 1,258 regions comprising 421,741 bp using both NimbleGen and SureSelect methods. The same genomic coordinates were sent to Roche for complementary oligonucleotide microarray design or uploaded to Agilent's eArray website for cRNA bait design.

DNA isolated using NimbleGen solid phase sequence capture was sequenced using 454 GS FLX pyrosequencing (454 Life Sciences); DNA obtained from SureSelect solution-based sequence capture was subjected to Illumina GA_{II} sequencing (Illumina). In both cases sequencing was performed according to manufacturer's protocols.

Sequence Analysis. Sequencing depth of coverage was defined as the number of sequencing reads, which had been filtered and mapped, per base. Average depth of coverage for each sequencing method was defined as the total depth of coverage per base, averaged over all requested bases. Depth of coverage per variant was defined as number of reads in which a variant was seen divided by the total depth of coverage for that base.

To prioritize a VUS, we developed a ranking algorithm that included the variant-calling threshold for each platform, location within the targeted region, type of change (nonsynonymous, splice-site, or frameshift deletion) and observed population frequency if the VUS was a reported SNP (Table 4). Other investigators have filtered variants on the basis of absence from the dbSNP database (9, 10, 14), a criterion we did not include as several mutations known to cause hearing loss have been assigned RefSNP (RS) numbers. Instead, we filtered variants on the basis of quantitative data from dbSNP130 such that any variant with an allele frequency >1% was considered a benign polymorphism (excluding known ARNSHL-associated variants of *GJB2*), whereas any variant with an allele frequency <1% (or of unknown frequency) was investigated further. In addition, we developed an in-house list of nonpathogenic nonsynonymous variants located within the targeted genes that can be excluded due to their presence in more than one non-replicate sample or the negative control (list available upon request).

Pathogenicity of a nonsynonymous, splice-site or insertion-deletion (indel) VUS was assessed using in silico mutation prediction software. We incorporated four algorithms (BLOSUM62, SIFT, PolyPhen2, and Align-GVGD) that have the highest positive predictive value (94.6%) when concordant (30) and we required concurrence in at least three of four prediction tools. We also prioritized gene variants in the context of inheritance pattern (i.e., genes known to cause ARNSHL or ADNSHL) (Table 4 and Table S1). A phenotypic filter was applied to ADNSHL using a computer learning algorithm we have developed called AudioGene, which predicts probable genotypes on the basis of audiometric criteria by constructing audioprofiles (20). Candidate mutations were verified by Sanger sequencing in the extended families and if unreported, excluded in ethnically matched controls.

ACKNOWLEDGMENTS. We thank the following individuals at the Baylor College of Medicine's Human Genome Sequencing Center for their invaluable contributions to this work: Dr. Richard Gibbs (director), Donna Muzny (director of operations), Dr. Yi Han, Dr. Min Wang, and Fiona Ongeri. We also thank Dr. Benjamin Darbro (University of Iowa) for assistance with the array CGH. Funding was provided by National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders Grant R01 DC02842 (to R.J.H.S.), a National Health and Medical Research Council Overseas Biomedical Postdoctoral Training Fellowship (to M.S.H.), NIH Predoctoral Research Fellowship T32 GM082729 (to A.P.D.), and a Doris Duke Clinical Research Fellowship (to A.E.S.).

1. Smith RJ, Bale JF, Jr., White KR, White KR (2005) Sensorineural hearing loss in children. *Lancet* 365:879–890.
2. Morton NE (1991) Genetic epidemiology of hearing impairment. *Ann N Y Acad Sci* 630:16–31.
3. Van Camp G, Willems PJ, Smith RJ (1997) Nonsyndromic hearing impairment: Unparalleled heterogeneity. *Am J Hum Genet* 60:758–764.
4. Van Camp G, Smith RJH (2010) Hereditary Hearing Loss Homepage. Available at http://www.hereditaryhearingloss.org. Accessed June 8, 2010.
5. Hamel C (2006) Retinitis pigmentosa. *Orphanet J Rare Dis* 1:40.
6. Green GE, et al. (1999) Carrier rates in the midwestern United States for GJB2 mutations causing inherited deafness. *JAMA* 281:2211–2216.
7. Hilgert N, Smith RJ, Van Camp G (2009) Forty-six genes causing nonsyndromic hearing impairment: Which ones should be analyzed in DNA diagnostics? *Mutat Res* 681: 189–196.
8. Lupski JR, et al. (2010) Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 362:1181–1191.
9. Ng S, et al. (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42:30–35.
10. Choi M, et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 106:19096–19101.
11. Tewhey R, et al. (2009) Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biol* 10:R116.
12. Gnirke A, et al. (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27:182–189.
13. Metzker ML (2010) Sequencing technologies: The next generation. *Nat Rev Genet* 11: 31–46.
14. Ng SB, et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461:272–276.
15. Harismendy O, et al. (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* 10:R32.
16. Hildebrand MS, et al. (2009) Mutation in the COCH gene is associated with superior semicircular canal dehiscence. *Am J Med Genet A* 149A:280–285.
17. Abe S, Usami S, Shinkawa H, Kelley PM, Kimberling WJ (2000) Prevalent connexin 26 gene (GJB2) mutations in Japanese. *J Med Genet* 37:41–43.
18. Talebizadeh Z, Kelley PM, Askew JW, Beisel KW, Smith SD (1999) Novel mutation in the KCNQ4 gene in a large kindred with dominant progressive hearing loss. *Hum Mutat* 14:493–501.

19. Oshima A, et al. (2008) Mutation profile of the CDH23 gene in 56 probands with Usher syndrome type I. *Hum Mutat* 29:E37–E46.
20. Zhang Y, et al. (2007) Sensorineural deafness and male infertility: A contiguous gene deletion syndrome. *J Med Genet* 44:233–240.
21. Hildebrand MS, et al. (2008) Audioprofile-directed screening identifies novel mutations in KCNQ4 causing hearing loss at the DFNA2 locus. *Genet Med* 10:797–804.
22. Hilgert N, et al. (2008) A splice-site mutation and overexpression of MYO6 cause a similar phenotype in two families with autosomal dominant hearing loss. *Eur J Hum Genet* 16:593–602.
23. Morgan JE, et al. (2010) Genetic diagnosis of familial breast cancer using clonal sequencing. *Hum Mutat* 31:484–491.
24. Mamanova L, et al. (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7:111–118.
25. Lee H, et al. (2009) Improving the efficiency of genomic loci capture using oligonucleotide arrays for high throughput resequencing. *BMC Genomics* 10:646.
26. Rodriguez-Paris J, et al. (2010) Genotyping with a 198 mutation arrayed primer extension array for hereditary hearing loss: Assessment of its diagnostic value for medical practice. *PLoS ONE* 5:e11804.
27. Kothiyal P, et al. (2010) High-throughput detection of mutations responsible for childhood hearing loss using resequencing microarrays. *BMC Biotechnol* 10:10.
28. Hildebrand MS, et al. (2008) Advances in molecular and cellular therapies for hearing loss. *Mol Ther* 16:224–236.
29. Grimberg J, et al. (1989) A simple and efficient non-organic procedure for the isolation of genomic DNA from blood. *Nucleic Acids Res* 17:8390.
30. Chan PA, et al. (2007) Interpreting missense variants: Comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). *Hum Mutat* 28:683–693.

nature genetics

Exome sequencing identifies *ACAD9* mutations as a cause of complex I deficiency

Tobias B Haack[1,2,11], Katharina Danhauser[1,2,11], Birgit Haberberger[1,2], Jonathan Hoser[3], Valentina Strecker[4], Detlef Boehm[5], Graziella Uziel[6], Eleonora Lamantea[7], Federica Invernizzi[7], Joanna Poulton[8], Boris Rolinski[9], Arcangela Iuso[1], Saskia Biskup[5], Thorsten Schmidt[3], Hans-Werner Mewes[3,10], Ilka Wittig[4], Thomas Meitinger[1,2], Massimo Zeviani[7] & Holger Prokisch[1,2]

An isolated defect of respiratory chain complex I activity is a frequent biochemical abnormality in mitochondrial disorders. Despite intensive investigation in recent years, in most instances, the molecular basis underpinning complex I defects remains unknown. We report whole-exome sequencing of a single individual with severe, isolated complex I deficiency. This analysis, followed by filtering with a prioritization of mitochondrial proteins, led us to identify compound heterozygous mutations in *ACAD9*, which encodes a poorly understood member of the mitochondrial acyl-CoA dehydrogenase protein family. We demonstrated the pathogenic role of the *ACAD9* variants by the correction of the complex I defect on expression of the wildtype ACAD9 protein in fibroblasts derived from affected individuals. *ACAD9* screening of 120 additional complex I–defective index cases led us to identify two additional unrelated cases and a total of five pathogenic *ACAD9* alleles.

Exome sequencing is a new powerful strategy to discover causative genes in rare Mendelian disorders[1], but despite its high sensitivity in detecting genetic variation[2], downstream filtering and interpretation of gene variants remain challenges. Successful application of exome sequencing has mainly been limited so far to cohorts of unrelated, affected individuals with a specific phenotype[3,4] and to individuals of consanguineous descent[5]. Here we demonstrate the efficacy of exome sequencing, in combination with a functional cell assay, to discover the molecular basis of a metabolic disorder.

Complex I deficiency is a frequent biochemical condition, accounting for about one-third of mitochondrial respiratory chain disorders[6]. Impairment of complex I activity compromises the transfer of electrons, derived from oxidative catabolism of carbohydrates and fatty acids, to coenzyme Q and downstream respiratory chain complexes (complex III and complex IV). Only a small proportion of complex I deficiencies have been associated with specific molecular abnormalities, usually in one of the genes, either mitochondrial or nuclear, encoding the structural subunits of the complex[7].

Table 1 Clinical, biochemical and genetic data of four individuals carrying mutations in *ACAD9*

Case, gender	Clinical phenotype (onset at birth)	Status	Tissue	Complex I[b]	Complex II and III[b]	Complex IV[b]	Complex V[b]	Predicted amino acid change
I:A[a], female	Cardiorespiratory depression, hypertrophic cardiomyopathy, encephalopathy, lactic acidosis	Died at age 46 days	Muscle Liver Fibroblasts	14% 1% 32%	n.r. n.r. n.r.	n.r. n.r. n.r.	52% 38% n.r.	 p.Phe44Ile p.Arg266Gln
I:B, male	Hypertrophic cardiomyopathy, mild exercise intolerance, persistent lactic acidosis	5 years old	Fibroblasts	38%	n.r.	n.r.	Slightly reduced	p.Phe44Ile p.Arg266Gln
II, female[c]	Hypertrophic cardiomyopathy, encephalomyopathy, lactic acidosis	Died at age 12 years	Muscle	13%	n.r.	n.r.	n.r.	p.Arg266Gln p.Arg417Cys
III, female	Hypertrophic cardiomyopathy, encephalopathy, lactic acidosis	Died at age 2 years	Muscle	26%	Increased	n.r.	n.r.	p.Ala326Pro p.Arg532Trp

[a]Index case in which mutations were originally identified by exome sequencing. [b]Activities are normalized to citrate synthase and expressed as the percentage of the lowest control value of the referring laboratory. [c]No alteration of β oxidation or radiolabeled palmitate or myristate was found in fibroblast cells. n.r., activity within the normal range.

[1]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [2]Institute of Human Genetics, Technische Universität München, Munich, Germany. [3]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Germany Research Center for Environmental Health, Neuherberg, Germany. [4]Molecular Bioenergetics, Medical School, Goethe-Universität Frankfurt, Frankfurt am Main, Germany. [5]CeGaT GmbH, Tübingen, Germany. [6]Unit of Child Neurology, Neurological Institute 'Carlo Besta'-Istituto Di Ricovero e Cura a Carattere Scientifico (IRCCS) Foundation, Milan, Italy. [7]Unit of Molecular Neurogenetics, Neurological Institute 'Carlo Besta'-IRCCS Foundation, Milan, Italy. [8]Nuffield Department of Obstetrics and Gynaecology, University of Oxford, The Women's Centre, John Radcliffe Hospital, Oxford, UK. [9]Städtisches Klinikum München GmbH, Department Klinische Chemie, Munich, Germany. [10]Chair of Genome Oriented Bioinformatics, Center of Life and Food Science, Freising-Weihenstephan, Technische Universität München, Munich, Germany. [11]These authors contributed equally to this work. Correspondence should be addressed to H.P. (prokisch@helmholtz-muenchen.de) or M.Z. (zeviani@istituto-besta.it).

Received 16 July; accepted 8 October; published online 7 November 2010; doi:10.1038/ng.706

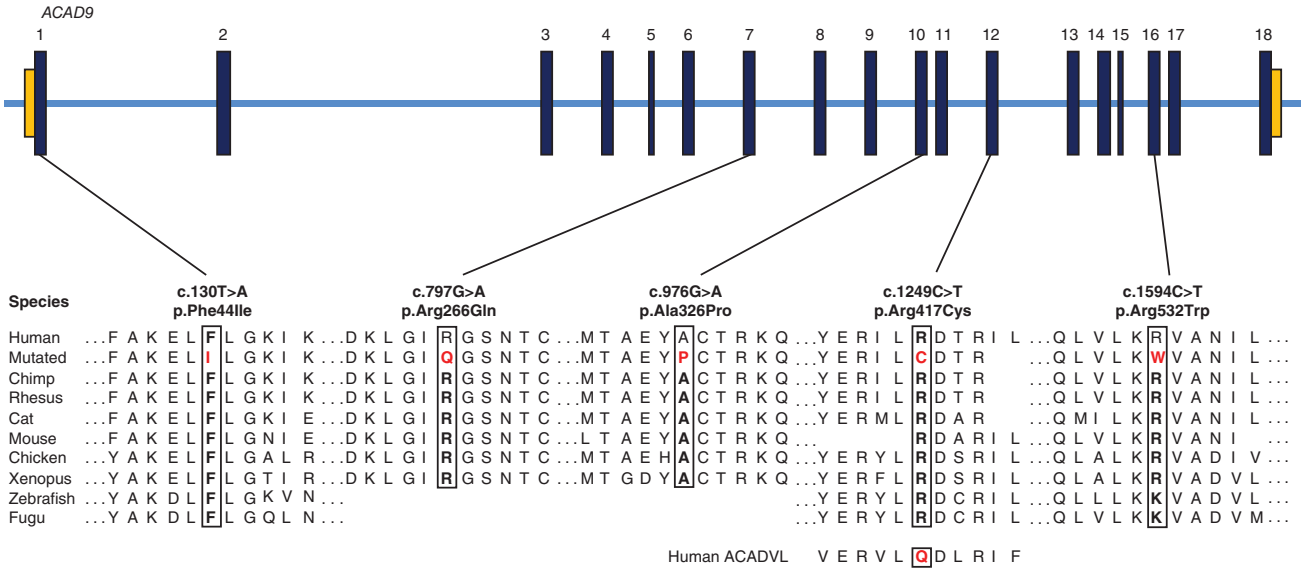


Figure 1 *ACAD9* gene structure and conservation of affected amino acid residues of identified mutations (shown in red).

By screening all of the 45 structural complex I subunits and 6 known assembly factors in 150 complex I–deficient index cases, we identified causative mutations in 30 individuals. We then selected four individuals who fulfilled the following criteria: (i) positive family history (18 out of 120 individuals), (ii) substantial reduction of complex I activity in tissues, including fibroblasts (defined as <70% of the lowest control value) (4 out of 18 individuals) and (iii) well-defined and documented clinical phenotype. In one individual, termed case I:A, exome sequence analysis allowed us to find the responsible gene (reported here), whereas investigation on the remaining three cases is still underway. Case I:A was born to healthy, non-consanguineous parents. Soon after birth, case I:A developed respiratory insufficiency, cardiac failure due to hypertrophic cardiomyopathy and lactic acidosis with severe encephalopathy. Profound complex I deficiency was detected in homogenates from this individual's liver, muscle and fibroblasts. A milder reduction was also found in ATPase activity (complex V), whereas other respiratory chain activities were normal (**Table 1**, for clinical and laboratory details, see **Supplementary Note**).

DNA from case I:A was used for exome capturing of 37.8 Mb of genomic sequence followed by sequencing on a SOLiD platform, which generated 3.65 Gb of high-quality sequence as single-end 50-bp reads. Of these reads, 67.7% mapped to the human genome (build hg18), with 72.2% of the mapped bases being on target. Altogether, we captured 165,637 exons with an average coverage of 47.6-fold, with 82.6% of the exome being covered at least 8-fold (**Supplementary Table 1**). A total of 14,167 potential single nucleotide variants (SNVs) were identified, including 4,928 splice-site and non-synonymous variants (SS-NSVs). Filtering against dbSNP, HapMap and in-house SNP data reduced this number to 533 new SS-NSVs (**Supplementary Table 2**). The proportion of known SS-NSVs identified (89.2%) was within the range of previous studies. However, the false positive discovery rate was high, as indicated by (i) the ratio of synonymous compared to non-synonymous variations among all (1:1.15), known (1:1.08) and new (1:2.52) changes, and (ii) transition over transversion ratios in all (2.81:1), known (3.21:1) and new (1.21:1) variants (**Supplementary Table 3**). A heterozygous change in *NDUFS2* (c.671C>T het; p.Ala224Val), also present in case I:A's father but not in her affected brother (case I:B), was considered as a non-pathogenic, non-causative variant in the index case. We next identified genes carrying homozygous (23) or at least two new heterozygous (25) SS-NSVs. Among the 48 corresponding proteins, only acyl-CoA dehydrogenase 9 (ACAD9) had a predicted mitochondrial localization (MitoP2 database)[8]. Sanger sequencing of the index case, her affected sibling and her parents confirmed the compound heterozygous state of the mutations, with each parent being heterozygous for one variant. High-resolution melting-curve analysis in 120 index cases with isolated complex I deficiency revealed two additional individuals carrying compound heterozygous variants (**Table 1**) and one case harboring a homozygous variant. Multiple sequence alignment of ACAD9 to mammalian orthologs (**Fig. 1**) and to other human ACAD proteins indicated that the altered positions are highly conserved. Notably, the *ACAD9* mutation c.1249C>T (p.Arg417Cys), found in case III, is identical to a mutation in the homologous region of ACADVL that was recently reported in an ACADVL-defective individual[9]. Of the six *ACAD9* variants, only the homozygous variant (c.976G>A,

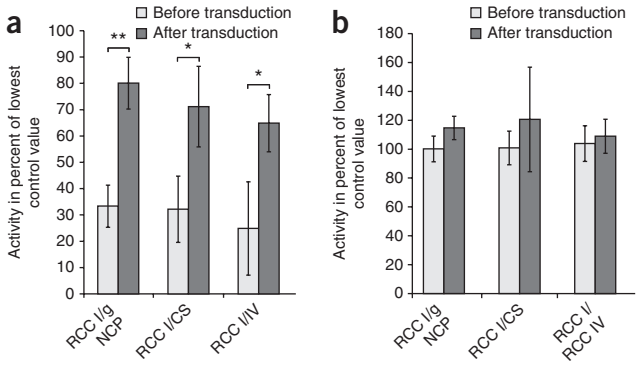


Figure 2 Cellular complementation experiment. We transduced case (**a**) and control (**b**) fibroblast cell lines with *ACAD9* overexpressing construct and we determined respiratory chain complex I, IV (RCCI and IV) and citrate synthase (CS) activities. The data are based on two independent transduction experiments for each cell line. Activities were determined at three different time points and expressed in percent of lowest control value ± standard deviation (s.d.). $*P < 0.05$, $**P < 0.01$. g NCP, gram non-collagen protein.

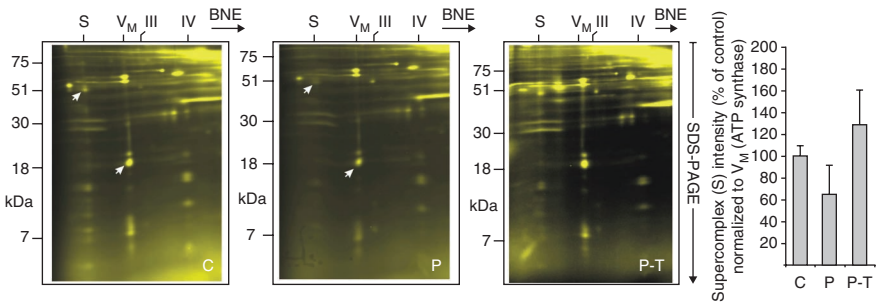


Figure 3 Complex I assembly in fibroblasts. Two-dimensional blue native–SDS-PAGE separation and quantification of fluorescent-labeled mitochondrial complexes from 10 mg patient (P), patient transduced with wildtype *ACAD9* (P-T) and control fibroblasts (C) are shown. Densitometric quantitation of supercomplex (S) fluorescence intensity was normalized to complex V (C, $n = 5$; P, $n = 3$; P-T, $n = 2$). Error bars indicate ± s.d. Gels from one typical experiment are shown as pseudocolors. Assignment of complexes: V_M, monomeric complex V or ATP synthase; III, complex III or cytochrome *c* reductase; IV, complex IV or cytochrome *c* oxidase; S, supercomplexes composed of respiratory chain complexes I, III and IV. Arrows indicate the subunits that were chosen for quantification.

p.Ala326Thr) was also found in the controls (14 out of 470 individuals). We therefore considered this variant as non-pathogenic. In addition, three heterozygous mutations of unknown pathogenicity were identified in three other cases (c.577C>T, p.Arg193Trp; c.701C>T, p.Ser234Phe; and c.907G>A, p.Gly303Ser). In order to test whether the mutations of the index case (case 1:A) were pathogenic, we performed complementation assays in complex I–defective fibroblasts. After transduction with a lentiviral vector expressing wildtype *ACAD9* complementary DNA (cDNA), the mutant cell lines showed a significant increase of complex I activity, whereas complex IV and citrate synthase activities remained unchanged (**Fig. 2**). This result supports the pathogenic relevance of the mutations found in *ACAD9*.

Next, we investigated whether the reduction of complex I activity was associated with reduced or abnormal complex I assembly. Blue native gel electrophoresis in combination with fluorescence labeling of mitochondrial proteins showed that the complex I holoenzyme was assembled in supercomplexes. The amount of complex I holoenzyme was substantially reduced (by 35%) in mutant cells, suggesting either complex I instability or impaired assembly. Transduction of wildtype *ACAD9* increased the amount of complex I holoenzyme in supercomplexes up to normal levels (**Fig. 3** and **Supplementary Table 4**).

All the sequence variants found in our cases are mis-sense mutations in conserved amino acid residues, suggesting impaired ACAD9 folding. Riboflavin, the vitamin precursor of the flavin adenine dinucleotide (FAD) moiety, which is the catalytic cofactor of ACADs, is known to foster their assembly and stability[10]. Therefore, we tested the effect of riboflavin in mutant cell cultures from cases I:A and 1:B. Supplementation of 5.3 μM of riboflavin for 3 days resulted in a significant increase of complex I activity (2.1-fold ($P < 0.05$) and 1.7-fold ($P < 0.05$), respectively; **Fig. 4**), whereas the activity of complex IV remained unchanged.

A function of ACAD9 in the oxidation of long-chain fatty acids has been shown *in vitro*, with an optimum activity toward unsaturated species[11,12]. However, as for other ACADs, including, for instance, ACADL, the physiological function of ACAD9 *in vivo* remains poorly understood. Three previously reported cases with metabolic and clinical features resembling a β-oxidation defect had low levels of *ACAD9* mRNA and protein but no mutations in *ACAD9* (ref. 13), a result that, albeit interesting, casts doubt on the primary role of *ACAD9* in the metabolic condition of these cases. On the contrary, we clearly demonstrated that mutations in *ACAD9* are associated with a mitochondrial disorder dominated by severe and generalized complex I deficiency. Notably, our cases had no biochemical abnormality of β-oxidation, based on both the absence of urinary and serum biomarkers (case I:B) and normal fatty acid β-oxidation activities in fibroblasts (case II).

How can ACAD9 impairment affect complex I activity and stability? A first hypothesis is that this protein plays a dual role, one role related to β-oxidation and the other to complex I stability. A dual role has indeed been demonstrated for the acyl carrier protein[14], which is in fact part of the mitochondrial fatty acid biosynthetic pathway and is also a protein subunit of complex I. A direct interaction of ACAD9 with two other complex I assembly factors, NDUFAF1 and ECSIT, has been shown recently[15]. A second hypothesis is that ACAD9 is involved in the biochemical dynamics of the lipid milieu in the inner mitochondrial membrane. Not only is complex I physically embedded in the inner mitochondrial membrane, but its stability and activity depend on the composition of the surrounding lipid compounds, particularly cardiolipin[16].

All of our four cases with mutations in *ACAD9* presented with cardiomyopathy, encephalopathy and lactic acidosis. Unlike in many mitochondrial disorders[6], we obtained a promising clinical response to a multivitamin scheme including daily riboflavin treatment (100 mg) in case I:B. Beneficial effects of riboflavin treatment have repeatedly been reported in individuals with mutations in genes encoding enzymes containing FAD and flavin mononucleotide (FMN)[17,18]; the identification of pathogenic mutations in *ACAD9*, encoding a FAD-containing flavoprotein, offers a rational, mechanistic explanation of our clinical observation. However, follow-up data of a larger cohort of individuals with *ACAD9* mutations is needed to establish the clinical efficacy of a supplementation with vitamins and cofactors.

In this study, we demonstrated that exome sequencing in combination with prioritization for mitochondrial proteins and cellular complementation assays, is effective in the elucidation of the

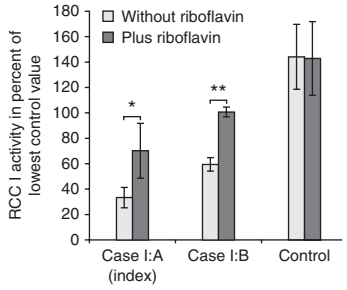


Figure 4 Effect of riboflavin treatment in fibroblasts. We treated case I:A, I:B and control fibroblast cell lines with riboflavin (5.3 μM for 72 h) and we determined respiratory chain complex I, IV (RCCI and IV) and citrate synthase activities. Measurements were performed in three independent experiments. Activities are expressed in percent of lowest control value ± s.d. $*P < 0.05$, $**P < 0.01$. g NCP, gram non-collagen protein.

molecular basis of complex I deficiency. This achievement is clearly important for the diagnosis and, eventually, the treatment of complex I individuals and offers a methodological paradigm that can be exploited in the near future for the definition of the gene repertoire involved in the formation, stability and activity of complex I.

URLs. MitoP2, http://www.mitop.de:8080/mitop2/; Sequence Variant Analyzer (SVA), http://www.svaproject.org/; primer design, http://ihg2.helmholtz-muenchen.de/ihg/ExonPrimer.html; sequence alignment, http://www.mutationtaster.org/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

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

ACKNOWLEDGMENTS

We are indebted to the subjects and their families involved in the study. We gratefully acknowledge the support of R. Hellinger, A. Löschner, M. Bada and F. Carrara in genotyping and cell culture work. Thanks to C. Fratter, M. Pike and S. Olpin. T.M. and H.P. were supported by the Impulse and Networking Fund of the Helmholtz Association in the framework of the Helmholtz Alliance for Mental Health in an Ageing Society (HA-215), the German Federal Ministry of Education and Research (BMBF) funded German National Research Network (NGFNplus #01GS08134), German Network for Mitochondrial Disorders (mitoNET #01GM0862 and 01GM0867), German Center for Diabetes Research (DZD e.V.) and grant RF-INN-2007-634163 of the Italian Ministry of Health. H.-W.M., T.M. and H.P. were supported by the Systems Biology of Metabotypes (SysMBo #0315494A). I.W. was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 815, Project Z1 (Redox-Proteomics), by the German Network for Mitochondrial Disorders (mitoNET #01GM0863) and by the Cluster of Excellence 'Macromolecular Complexes' at the Goethe-University (EXC 115). M.Z. was supported by the Pierfranco and Luisa Mariani Foundation Italy, Fondazione Telethon Italy grant number GGP07019, Fondazione Giuseppe Tomasello–ONLUS and grant RF-INN-2007-634163 of the Italian Ministry of Health. J.P. was supported by the Medical Research Council (UK) and the National Commissioning Group.

AUTHOR CONTRIBUTIONS

Project planning: T.M., M.Z., H.P. Experimental design: H.P. Review of phenotypes and sample collection: G.U., E.L., F.I., J.P. and B.R. Mutation screening: T.B.H., D.B. and S.B. Data analysis: T.B.H., J.H., T.S., H.-W.M. and H.P. Cell biology experiments: K.D., B.H., V.S., I.W. and A.I. Manuscript writing: T.B.H., T.M., M.Z. and H.P. Critical revision of the manuscript: all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Ng, S.B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–276 (2009).
2. Hedges, D.J. *et al.* Exome sequencing of a multigenerational human pedigree. *PLoS ONE* **4**, e8232 (2009).
3. Ng, S.B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat. Genet.* **42**, 30–35 (2010).
4. Hoischen, A. *et al. De novo* mutations of *SETBP1* cause Schinzel-Giedion syndrome. *Nat. Genet.* **42**, 483–485 (2010).
5. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 19096–19101 (2009).
6. Triepels, R.H., Van Den Heuvel, L.P., Trijbels, J.M. & Smeitink, J.A. Respiratory chain complex I deficiency. *Am. J. Med. Genet.* **106**, 37–45 (2001).
7. Thorburn, D.R. *et al.* Biochemical and molecular diagnosis of mitochondrial respiratory chain disorders. *Biochim. Biophys. Acta* **1659**, 121–128 (2004).
8. Elstner, M., Andreoli, C., Klopstock, T., Meitinger, T. & Prokisch, H. The mitochondrial proteome database: MitoP2. *Methods Enzymol.* **457**, 3–20 (2009).
9. Gobin-Limballe, S. *et al.* Genetic basis for correction of very-long-chain acyl-coenzyme A dehydrogenase deficiency by bezafibrate in patient fibroblasts: toward a genotype-based therapy. *Am. J. Hum. Genet.* **81**, 1133–1143 (2007).
10. Saijo, T. & Tanaka, K. Isoalloxazine ring of FAD is required for the formation of the core in the Hsp60-assisted folding of medium chain acyl-CoA dehydrogenase subunit into the assembly competent conformation in mitochondria. *J. Biol. Chem.* **270**, 1899–1907 (1995).
11. Zhang, J. *et al.* Cloning and functional characterization of ACAD-9, a novel member of human acyl-CoA dehydrogenase family. *Biochem. Biophys. Res. Commun.* **297**, 1033–1042 (2002).
12. Ensenauer, R. *et al.* Human acyl-CoA dehydrogenase-9 plays a novel role in the mitochondrial beta-oxidation of unsaturated fatty acids. *J. Biol. Chem.* **280**, 32309–32316 (2005).
13. He, M. *et al.* A new genetic disorder in mitochondrial fatty acid beta-oxidation: ACAD9 deficiency. *Am. J. Hum. Genet.* **81**, 87–103 (2007).
14. Feng, D., Witkowski, A. & Smith, S. Down-regulation of mitochondrial acyl carrier protein in mammalian cells compromises protein lipoylation and respiratory complex I and results in cell death. *J. Biol. Chem.* **284**, 11436–11445 (2009).
15. Nouws, J. *et al.* Acyl-CoA dehydrogenase 9 is required for the biogenesis of oxidative phosphorylation complex I. *Cell Metab.* **12**, 283–294 (2010).
16. Ohtsuka, T., Nishijima, M., Suzuki, K. & Akamatsu, Y. Mitochondrial dysfunction of a cultured Chinese hamster ovary cell mutant deficient in cardiolipin. *J. Biol. Chem.* **268**, 22914–22919 (1993).
17. Ghezzi, D. *et al.* Severe X-linked mitochondrial encephalomyopathy associated with a mutation in apoptosis-inducing factor. *Am. J. Hum. Genet.* **86**, 639–649 (2010).
18. Gempel, K. *et al.* The myopathic form of coenzyme Q10 deficiency is caused by mutations in the electron-transferring-flavoprotein dehydrogenase (ETFDH) gene. *Brain* **130**, 2037–2044 (2007).

ONLINE METHODS

Cases and samples. For exome sequencing, we selected one out of two affected siblings of self-reported European ancestry with respiratory chain complex I deficiency. The index case was diagnosed at the Unit of Molecular Neurogenetics at the 'C. Besta' Neurological Institute–IRCCS Foundation.

Mutation screening was performed in a survey of index cases from various European centers. The biochemical diagnosis of complex I deficiency was mandatory for inclusion in the study. Written consent was obtained from all cases or their guardians, and the ethical committees of each referring center approved the study. Phenotypic data were collected from medical reports and the clinicians involved.

Targeted capture and exome sequencing. DNA was extracted from fibroblast cell lines of the index case using the AllPrep RNA/DNA Mini Kit (Qiagen). Targeted enrichment was performed with an ABI SOLiD optimized SureSelect human exome kit (Agilent) and 3 µg of genomic DNA. The kit is designed to enrich for 165,637 exons (~18,000 genes) covering a total of 37-Mb genomic sequences. The prepared exome library was used for electronic PCRs following the manufacturer's instructions (Life Technologies) based on a library concentration of 0.5 pM. Captured libraries were sequenced as single-end 50-bp reads. For each sample, one quad of a SOLiD sequencing slide (Life Technologies) was used.

Read mapping and variant analysis. Color space reads were mapped to the reference human genome (UCSC hg18) with the BLAT-like fast accurate search tool (BFAST v0.6.2a)[19]. SNVs were subsequently called using SAMTools (v0.1.7-18(r605))[20] with the following quality criteria: ≥×8 coverage, at least three reads indicating a sequence variation detected in at least 5% on both strands and a Phred-like consensus quality of ≥20 (ref. 3). Variants were defined as 'heterozygous' when ≥25% of all nucleotides at the position showed non-reference bases and as 'homozygous' according to the initial SAMTools classification (using default parameters).

For prioritization, we omitted known variants from HapMap[21] and dbSNP v130 (ref. 22) using the Sequence Variant Analyzer (SVA) version 1.01b. An additional 200 cSNVs detected in independent exome resequencing projects were excluded. We used SVA for the annotation of SNVs based on Ensembl and databases included in SVA. Based on a recessive model of inheritance[3], non-synonymous variants were filtered for genes with more than one variant. Further, we prioritized genes encoding proteins with a known or predicted mitochondrial localization according to the MitoP2 database[8].

***ACAD9* mutation screening.** PCR primers for the 18 *ACAD9* exons were designed using the UCSC ExonPrimer-Primer Design Program. Restriction of the PCR product size to a maximum of 500 bp resulted in 17 PCR amplicons. Primer sequences and PCR conditions are provided on request. Mutational screening in *ACAD9* in 120 index cases with isolated complex I deficiency was conducted on a LightScanner instrument (Idaho Technology) as described elsewhere[23]. In cases of altered melting curves compared with the average of multiple wildtypes, the exon was sequenced using the BigDye sequencing kit (Applied Biosystems). High-resolution melting-curve analysis assays were established for all exons except for exon 7, which was analyzed by Sanger sequencing. Only a heterozygous change in *NDUFS2* (c.671C>T het; p.Ala224Val) was identified. Molecular analysis of the mtDNA revealed neither new nor known pathogenic mutations, nor any evidence for mtDNA deletions or depletion.

Riboflavin treatment of fibroblast cell lines. Fibroblast cell lines were grown under standard conditions in high-glucose Dulbecco's Modified Eagle's Medium (DMEM, Invitrogen, 41966) supplemented with 10% of FBS (Invitrogen, 10270106), 1% penicillin/streptomycin (Invitrogen, 15070-063) and 200 µM uridine (Sigma, U3003), and cultivated at 37 °C in atmospheric oxygen and 5% CO_2. Riboflavin (Sigma, R9504) was added to the culture medium to a final concentration of 5.3 µM. Cells were collected for analysis of respiratory chain complex activity and gene expression analysis after 72 h of treatment.

We determined mitochondrial respiration in case I:B in digitonin-permeabilized fibroblasts using high-resolution respirometry (Oroboros Oxygraph-2k), and we used DatLab Software for data acquisition and analyses (Oroboros Instruments) as previously described[24]. Respiration rates were calculated as the time derivative of the oxygen concentration normalized for viable cells analyzed. Cell lines from case I:A were analyzed spectrophotometrically[25]. Three independent experiments were performed for each condition.

Lentiviral complementation. The feline immunodeficiency virus (FIV)–based lentiviral transduction was done according to the manufacturer's protocol (GeneCopoeia, Lenti-Pac FIV Expression Packaging Kit, User Manual Version III).

In brief, to produce the FIV particles, HEK293T cells were plated 2 days before transfection in high-glucose DMEM (Invitrogen, 41966) supplemented with 10% of heat-inactivated FBS (Invitrogen, 10270106) at 70% confluence. Without changing the medium, HEK293T cells were cotransfected using EndoFectin (Genecopoeia) with the packaging plasmids (GeneCopoeia, Lenti-Pac FIV) and the expression vector (Genecopoeia) containing the cDNA of *ACAD9* and a neomycin resistance gene at a 1:1 ratio. After 12 h, the medium was replaced with DMEM supplemented with 2% FBS and 1% penicillin/streptomycin. The supernatant containing the viral particles was collected 48 h after transfection, centrifuged at 500*g* for 10 min and filtered through a 0.45 µm filter. The fibroblasts (60–70% confluence) were than transduced using a 1:2 dilution with DMEM supplemented with 10% FBS, 1% penicillin/streptomycin (Invitrogen, 15070) and 200 µM uridine (Sigma, U3003). Forty-eight hours after the transduction, the fibroblasts were selected using 3 µl/ml G418 antibiotics (Sigma, G8168) for 2 weeks before performing biochemical analysis.

Complex I, complex IV and citrate synthase activities were determined spectrophotometrically as described[25] using a JASCO V-550 spectrophotometer. Overexpression of mutant *ACAD9* did not increase complex I activity.

Electrophoresis and fluorescence labeling of mitochondrial proteins. Standard protocols for solubilization of cells with digitonin, one-dimensional BNE (blue native electrophoresis)[26] and two-dimensional SDS-PAGE[27] were modified to improve isolation of mitochondrial complexes from human fibroblasts. Briefly, mitochondrial protein complexes from 10 mg fibroblasts (wet weight) were labeled with Fluorescein and solubilized with 5 µl digitonin (20%) in a buffer containing 50 mM NaCl, 50 mM imidazole, 2 mM aminohexanoic acid and 1 mM EDTA at pH 7. Following one-dimensional BNE and two-dimension SDS-PAGE, the two-dimension gels were scanned using a Typhoon scanner (GE Healthcare) to detect Fluorescein (excitation 488 nm, emission filter 520 nm). Detailed protocols for fluorescence labeling will be published elsewhere (Strecker, Wittig & Schägger, unpublished).

Sequence alignment. The multiple protein sequence alignment was generated using Mutation Taster[28] and the following protein sequences: human (ENST00000308982), chimpanse (ENSPTRG00000015362), rhesus (ENSMMUG00000006338), cat (ENSFCAG00000008660), mouse (ENSMUSG00000027710), chicken (ENSGALG00000005034), *Xenopus* (ENSXETG00000001719), zebrafish (ENSDARG00000055620) and fugu (ENSTRUG00000002716).

Statistical analysis. The data are presented as means ± s.d. of at least three independent experiments and analyzed by Student's *t*-test. *P* values less than 0.05 were considered to be statistically significant.

19. Homer, N., Merriman, B. & Nelson, S.F. BFAST: an alignment tool for large scale genome resequencing. *PLoS ONE* **4**, e7767 (2009).
20. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
21. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
22. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
23. Meisinger, C. *et al.* A genome-wide association study identifies three loci associated with mean platelet volume. *Am. J. Hum. Genet.* **84**, 66–71 (2009).
24. Kuznetsov, A.V. *et al.* Analysis of mitochondrial function *in situ* in permeabilized muscle fibers, tissues and cells. *Nat. Protoc.* **3**, 965–976 (2008).
25. Sgobbo, P., Pacelli, C., Grattagliano, I., Villani, G. & Cocco, T. Carvedilol inhibits mitochondrial complex I and induces resistance to H2O2-mediated oxidative insult in H9C2 myocardial cells. *Biochim. Biophys. Acta* **1767**, 222–232 (2007).
26. Wittig, I., Braun, H.P. & Schagger, H. Blue native PAGE. *Nat. Protoc.* **1**, 418–428 (2006).
27. Schägger, H. Tricine-SDS-PAGE. *Nat. Protoc.* **1**, 16–22 (2006).
28. Schwarz, J.M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).

OPEN ACCESS Freely available online

PLoS one

Molecular Diagnosis of Neonatal Diabetes Mellitus Using Next-Generation Sequencing of the Whole Exome

Amélie Bonnefond[1], Emmanuelle Durand[1], Olivier Sand[1], Franck De Graeve[1], Sophie Gallina[1], Kanetee Busiah[2], Stéphane Lobbens[1], Albane Simon[2], Christine Bellanné-Chantelot[3], Louis Létourneau[4], Raphael Scharfmann[2], Jérôme Delplanque[1], Robert Sladek[4], Michel Polak[2], Martine Vaxillaire[1], Philippe Froguel[1,5]*

1 CNRS-UMR-8199, Univ Lille Nord de France, UDSL, Lille, France, **2** Inserm-U845, Department of Pediatric Endocrinology, Necker Enfants Malades Hospital, Université Paris Descartes, Paris, France, **3** Department of Genetics, Pitié-Salpétrière Hospital, Paris, France, **4** Department of Human Genetics, Faculty of Medicine, McGill University, Montreal, and Genome Quebec Innovation Centre, Montreal, Canada, **5** Department of Genomics of Common Disease, School of Public Health, Imperial College London, Hammersmith Hospital, London, United Kingdom

Abstract

Background: Accurate molecular diagnosis of monogenic non-autoimmune neonatal diabetes mellitus (NDM) is critical for patient care, as patients carrying a mutation in *KCNJ11* or *ABCC8* can be treated by oral sulfonylurea drugs instead of insulin therapy. This diagnosis is currently based on Sanger sequencing of at least 42 PCR fragments from the *KCNJ11*, *ABCC8*, and *INS* genes. Here, we assessed the feasibility of using the next-generation whole exome sequencing (WES) for the NDM molecular diagnosis.

Methodology/Principal Findings: We carried out WES for a patient presenting with permanent NDM, for whom mutations in *KCNJ11*, *ABCC8* and *INS* and abnormalities in chromosome 6q24 had been previously excluded. A solution hybridization selection was performed to generate WES in 76 bp paired-end reads, by using two channels of the sequencing instrument. WES quality was assessed using a high-resolution oligonucleotide whole-genome genotyping array. From our WES with high-quality reads, we identified a novel non-synonymous mutation in *ABCC8* (c.1455G>C/p.Q485H), despite a previous negative sequencing of this gene. This mutation, confirmed by Sanger sequencing, was not present in 348 controls and in the patient's mother, father and young brother, all of whom are normoglycemic.

Conclusions/Significance: WES identified a novel *de novo ABCC8* mutation in a NDM patient. Compared to the current Sanger protocol, WES is a comprehensive, cost-efficient and rapid method to identify mutations in NDM patients. We suggest WES as a near future tool of choice for further molecular diagnosis of NDM cases, negative for chr6q24, *KCNJ11* and *INS* abnormalities.

Citation: Bonnefond A, Durand E, Sand O, De Graeve F, Gallina S, et al. (2010) Molecular Diagnosis of Neonatal Diabetes Mellitus Using Next-Generation Sequencing of the Whole Exome. PLoS ONE 5(10): e13630. doi:10.1371/journal.pone.0013630

Editor: Michael Nicholas Weedon, Peninsula Medical School, United Kingdom

Received July 19, 2010; **Accepted** September 30, 2010; **Published** October 26, 2010

Funding: AB is funded by a research fellowship from the Conseil Régional du Nord Pas de Calais (France) and the Centre National de la Recherche Scientifique (CNRS). This study was supported by the ANR-MRAR research program ("Programme Pluriannuel National de Recherche sur les Maladies Rares", ANR-07-MRAR-000, to MP), a transnational European research grant on Rare Diseases (ERANET 09 RARE 005, to MV and MP) and the European Union (Integrated Project EuroDia LSHM-CT-2006-518153 in the Framework Programme 6 [FP6] of the European-Community, to PF). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: p.froguel@imperial.ac.uk

Introduction

Neonatal diabetes mellitus (NDM) is a rare monogenic form of non-autoimmune diabetes which affects 1 in ~300,000 live births and is diagnosed before six months of age [1,2,3]. Approximately half of the NDM cases are transient (TNDM) but can ultimately relapse. In contrast, permanent NDM (PNDM) cases need continual treatment from diagnosis [1,2,3]. More than half of both forms of NDM cases have been elucidated, so far, and it appears that the genetic aetiologies of NDM are quite heterogeneous. Indeed, although the majority of TNDM cases have an abnormality in chromosome 6q24 and the other most frequent causes of NDM are missense mutations in the pancreatic β-cell K_{ATP} channel genes *KCNJ11* and *ABCC8*, and in the preproinsulin gene, NDM has been linked to numerous other genetic causes including point mutations in *GCK*, *GLIS3*, *EIF2AK3*, *PDX1*, *PTF1A*, *SLC2A2*, *HNF1B* or *FOXP3* [1,2,3].

Even if the presence of specific clinical features (*e.g.* relatively late age of onset, pancreas agenesis, developmental delay, renal failure, anaemia, thyroid disease, cardiac disorders…) or a family history of diabetes or consanguinity may suggest potential molecular aetiology(ies) for NDM, a molecular genetic diagnosis is crucial as it can predict the most appropriate treatment and genuinely improve quality of life [3]. The most striking example is seen for NDM patients with a mutation in the K_{ATP} channel genes, who can be treated effectively with oral sulfonylureas that directly bind the SUR1 regulatory subunit of the channel, rather than requiring life-long insulin therapy which usually provides poor glycemic contol [4,5,6].

Most developed countries offer DNA testing for NDM patients to establish a personal molecular genetic diagnosis for family counselling and to plan personalized pharmacotherapy. When severe hyperglycaemia is detected in a neonate, it is difficult to predict whether NDM will be transient or permanent. If the young patient does not have extrapancreatic features or a family history of diabetes (especially in a consanguineous context), it is suggested to first search for a chromosome 6q24 abnormality or for a *KCNJ11* mutation, as these NDM aetiologies are the most frequent, and then for mutations in *ABCC8* and *INS* if the first tests are negative [1,3]. As *KCNJ11*, *ABCC8* and *INS* altogether represent 42 coding exons, sequencing these genes using the standard Sanger protocol is obviously tedious, long and costly. If this first set of gene sequencing is negative, further molecular analysis of the other NDM genes is generally not performed. This current approach to molecular diagnosis of NDM provides only a limited sequencing of the known NDM genes and no assessment of possible modifier genetic loci elsewhere in the genome: a more comprehensive cost efficient methodology to scrutinize every new NDM case is necessary.

In the present study, we demonstrate the feasibility of next-generation whole exome sequencing (WES) for the molecular diagnosis of a patient with NDM without any extrapancreatic features or family history of diabetes. Despite previous negative Sanger sequencing of *ABCC8* by a hospital laboratory, we identified a novel non-synonymous mutation in this gene through WES. We show that this cutting-edge novel technology is more comprehensive, less labour intensive and thus cheaper for NDM diagnosis than standard sequencing protocols.

Results

The patient's clinical record reported that he has developed severe hyperglycemia, ketoacidosis and weight loss at two months of age. HLA typing showed neutral alleles for type 1 diabetes mellitus susceptibility. Pancreas ultrasound scan was normal and the patient did not show any specific extra-pancreatic clinical features. He was firstly treated with continuous subcutaneous insulin infusion during two years with rather low dose of insulin (<0.5 units/kg/day) for a pretty good metabolic control (A1C <8.5%). He was then switched with a basal-bolus scheme for technical adverse outcome. The patient is currently 20 years old. He is treated with 1.1 units/kg/day of insulin, with A1C values ranging between 8 and 9%. He always had an attention disorder and a learning disability without obvious motor symptoms or epilepsy.

After target enrichment, the whole exome DNA library from the patient was sequenced in 76 base-pairs (bp) paired-end reads, using two channels of the sequencing instrument. The WES generated 34,600,000 bp of nucleotide sequences mapping once to the targeted exome, which achieved a mean coverage of 122×of the target exome. WES identified 55,202 targeted DNA variants from the reference human genome sequence NCBI36/hg18, of which 4,463 were novel compared to the public database dbSNP130 and the eight HapMap exomes sequenced by Ng et al. [7] (Table 1). More specifically, we found 407 genes with one or more novel mutations including missense coding SNPs, gains of STOP codon and frameshift mutations (Table 1).

To validate our WES data and analysis protocol, the patient's DNA was genotyped using an Illumina Human1M-Duo array, which contains 8,500 single nucleotide polymorphisms (SNPs) located in exons captured by the Agilent SureSelect technology. WES identified 7,969 exomic SNPs present on the genotyping array (93.8%), and showed high concordance rates for homozygous and heterozygous calls (100% and 99.5% and respectively). By using only one channel of the sequencing instrument, the generated sequence achieved a mean coverage of 65×of the target genome and 88.9% of the exomic SNPs present on the genotyping array were identified. These results would suggest that our WES based on the use of two channels of the genome analyzer has a low

Table 1. Number of mutations identified through the WES analysis of DNA sample from the PNDM patient.

Sample		PNDM patient
Total targetted SNPs **(Novel*)**		55,202 **(4,463)**
Homozygous SNPs	Total SNPs **(Novel*)**	22,030 **(589)**
	Concordance* (%)	100
	Sensitivity† (%)	91.8
Heterozygous SNPs	Total SNPs (**Novel***)	33,172 **(3,872)**
	Concordance† (%)	99.5
	Sensitivity‡ (%)	95.1
Synonymous coding SNPs	Homozygous **(Novel*)**	3,262 **(14)**
	Heterozygous **(Novel*)**	4,780 **(277)**
Missense coding SNPs (a)	Homozygous **(Novel*)**	2,907 **(18)**
	Heterozygous **(Novel*)**	4,264 **(430)**
Gains of STOP codon (b)	Homozygous **(Novel*)**	16 **(0)**
	Heterozygous **(Novel*)**	30 **(7)**
Insertions or deletions (c)	Homozygous **(Novel*; Novel frameshift)**	1,748 **(951; 0)**
	Heterozygous **(Novel*; Novel frameshift)**	409 **(284; 0)**
Genes with one or more (a),(b) or (c) mutations **(Novel*)**		4,495 **(407)**

**Novel*: a novel mutation means that it is not present in the public database dbSNP130 and the eight HapMap exomes sequenced by Ng et al. [7].
†*Concordance*: % of similar allele assignment among exomic mutations detected on the Illumina Human1M-Duo array and those discovered by WES.
‡*Sensitivity*: % of exomic mutations present on the Illumina Human1M-Duo array that have been discovered by WES.
doi:10.1371/journal.pone.0013630.t001

false negative rate for detecting exomic mutations, which is necessary to achieve accurate molecular diagnosis for patients presenting with monogenic disorders.

In 2004, single-stranded Sanger sequencing analysis of the patient's DNA by a hospital unit showed no evidence for coding mutations in *ABCC8*. Also, no abnormalities in *KCNJ11*, *INS* and chromosome 6q24 were reported by the molecular diagnostic laboratory of Robert Debré Hospital (Paris, France). Since these assessments were performed six years ago, we checked again for mutations in these three genes as well as in all other genes known to be involved in monogenic forms of diabetes. Unexpectedly, we identified a novel heterozygous non-synonymous mutation c.1455G>C/p.Q485H in the 9th exon of *ABCC8*. Based on this startling result, we used a standard Sanger sequencing protocol, to study the patient's DNA sample available in the CNRS-UMR8199 unit (Lille, France) as well as the sample stored at the Robert Debré hospital (Paris, France). In both cases, we identified the p.Q485H mutation in *ABCC8*, confirming our WES results. In addition, retrospective re-examination of the data generated six years ago indicated that the p.Q485H mutation was present in the original *ABCC8* exon 9 sequences (chromogram). The *ABCC8* p.Q485H mutation was not found in 348 French nondiabetic individuals or in the patient's mother, father and young brother, all of whom are normoglycemic. The p.Q485H mutation affects an amino acid that is located in the transmembrane domain 1 (TMD1) of the ABCC8/SUR1 core; and that is highly conserved across species (Rhesus, Mouse, Dog, Rabbit, Elephant, Opossum, Platypus, Chicken, Lizard, Stickleback, X_Tropicalis, Tetraodon) according to the UCSC (NCBI/hg18) comparative genomics alignment pipeline (http://www.bx.psu.edu/miller_lab/). We evaluated the possible functional significance of the p.Q485H mutation by the PolyPhen-2 (Polymorphism Phenotyping v2) software which uses sequence- and structure-based criteria to predict the putative impact of point mutations on the structure and function of human proteins [8]: the p.Q485H mutation is predicted 'probably damaging' with a score of 0.999 (the score of 1 indicating the most damaging mutation). Following the identification of this *ABCC8* mutation in the patient's DNA, a switch from insulin to oral sulfonylurea treatment will be tried soon at Necker Hospital (Paris, France).

Discussion

In the present study, we demonstrate for the first time that WES can be seen as a relevant alternative for molecular diagnosis of NDM. Since an accurate molecular diagnosis for this condition can lead to very dramatic improvements in patient care, development of reliable and cost efficient methods for quick and accurate DNA analysis are of major interest.

Currently, patients with NDM are evaluated using Sanger sequencing, which is far more expensive per sequenced base-pairs than WES. Indeed, the standard sequencing of a single PCR fragment costs 67.50€ ($82) for the only French hospital laboratory specialized in NDM molecular diagnosis (Robert Debré hospital, Paris). This price includes consumables, equipment amortization, personnel salary and hospital overhead costs. Therefore, the total cost for the French National Insurance for the sequencing of *KCNJ11*, *ABCC8* and *INS* alone, which requires 42 PCR fragments is 2,835€ ($3,440; ~0.45€ or ~$0.55 per bp). In comparison, the all inclusive cost of WES for NDM, which will detect mutations in *KCNJ11*, *ABCC8* and *INS* as well as rarer genetic aetiologies of NDM, is currently 3,274€ ($4,146; <<0.001€ or $ per bp) per sample, by performing a sequencing on two channels and in 76 bp paired-end configuration (CNRS-UMR8199, Lille, France). Moreover, it is very likely that WES cost will fall in next months towards 2,000€ ($2,528) or even less.

We believe that the WES protocol is less labour intensive and time-consuming than the standard Sanger protocol for genetically heterogeneous disorders requiring several large genes to be screened. A WES run involving four DNA samples can be completed in two weeks, including the time required to analyse the data, which is comparable to the time required by current Sanger sequencing of *ABCC8* only with its 39 exons. The p.Q485H mutation was missed six years ago by the research assistant in charge of the sequence reading. Although we can assume that mutation detection bio-informatics tools were less efficient a few years ago (the hospital laboratory used the PhredPhrap software in 2004) and that current methods are more accurate, the Sanger protocol and specially the semi-automated reading of sequence traces is always laborious and demanding (thus expensive), and a double-check of sequence readings by two different persons is performed in several French diagnostic laboratories in order to avoid any errors in the mutation identification process.

WES method is not only a cost-effective tool for molecular diagnosis; it should be also seen as an excellent tool for further genetic research and identification of novel causal mutations. Indeed, in the French NDM cohort, half of PNDM cases are still not elucidated [9,10]. Classical linkage analyses are generally not successful as many NDM mutations occur *de novo* or are not fully penetrant. Most NDM genes have been found via candidate gene analyses but this approach has now reached its limits [11]. However, WES typically yields thousands of 'novel' genetic variants (*i.e.* not yet present in human genome variants databases). Therefore, the identification of truly causal variants would be strongly facilitated by the development of a high quality WES database of novel mutations found in both elucidated cases or in cases of unknown aetiology as well as in controls coming from same ethnicity. WES would also permit the identification of putative NDM modifier genes, a very challenging task for targeted gene analysis.

We are quite confident that the p.Q485H mutation is likely to be functional given the non ambiguous prediction of its putative damaging effect. In addition, the clinical data from the patient fit well with the features of PNDM linked to a *ABCC8* mutation (*e.g. de novo* mutation associated with very early-onset of the disease and attention disorder) [4,12].

We believe that other NDM patients should be assessed with the same protocol as DNA quality may change the WES accuracy. Also, our DNA capture was not totally perfect as we could miss approximately 6% of exomic SNPs present in the high-resolution oligonucleotide genotyping array. Furthermore, the exomic coverage was not homogeneous between NDM genes (Table 2), thus we could suspect that the WES accuracy would not be the same for all NDM genes. Therefore, despite high WES mean coverage and elevated rates of both concordance and sensitivity in mutation detection, it is also necessary to assess and to verify the homogeneity of the target capture, specially in the genes of interest that have to be screened for molecular diagnosis (Table 2).

Knowing that the capture technology is improving day after day (by enriching exomic loci poorly captured with the previous kits), our present study suggests that it will be possible to soon update the protocols for molecular diagnosis of NDM [3,13]. We propose that after discovery of severe hyperglycemia in a neonate who is negative for serological markers of type 1 diabetes, a preliminary assessment of abnormalities of chromosome 6q24 can be performed (as at this stage, it is too early to guess NDM will be permanent or transient) followed by the search of a mutation in both *KCNJ11* and *INS* using Sanger sequencing as these two genes can be easily and quickly

Table 2. Details on exomic sequencing depth in NDM genes, obtained through the WES of the PNDM patient.

NDM genes	Chromosome	Start	End	Number of coding exons	Exomic size (bp)	Number of mapped 76 bp exomic reads	Mean exomic coverage	% of sequenced exomic genes according to several depth thresholds			
								≥8×	≥20×	≥50×	≥100×
KCNJ11	chr11	17365042	17366214	1	1173	984	57.1×	100	88.7	50.5	13.6
ABCC8	chr11	17371114	17454899	39	4746	6235	90.5×	94.6	85.0	54.5	31.6
INS	chr11	2137658	2138777	2	333	97	24.4×	87.1	65.8	0	0
GCK	chr7	44164197	44159587	12	2012	1055	46.1×	70.8	44.1	19.0	14.4
GLIS3	chr9	3818272	4115864	9	2328	3672	130.4×	84.8	79.4	67.8	39.9
EIF2AK3	chr2	88638369	88707907	17	3351	9507	214.1×	92.2	92.0	87.1	81.4
PDX1	chr13	27392276	27396838	2	852	179	27.2×	40.3	24.5	3.9	0
PTF1A	chr10	23521466	23522840	2	986	405	59.9×	38.8	36.4	26.3	7.0
SLC2A2	chr3	172198386	172227153	11	1575	3669	170.9×	99.6	99.2	88.8	69.8
HNF1B	chr17	33121488	33178988	9	1674	2089	94.5×	93.3	83.4	53.5	34.5
FOXP3	chrX	48994739	49001906	11	1296	402	25.6×	69.7	51.9	13.8	0

doi:10.1371/journal.pone.0013630.t002

sequenced. If negative, we propose a WES analysis of the patient's DNA which is the most comprehensive way to fully explore the molecular causes of this NDM case.

Materials and Methods

Study participant and DNA samples

For WES, we selected a patient of European origin, diagnosed with PNDM who was referred to the French Network for the Study of Neonatal Diabetes Mellitus [9]. He was born from non consanguineous parents and had no intra-uterine growth retardation (birth weight 2,900 g/birth length 50 cm at 40 gestational weeks). He underwent a thorough clinical examination and his medical records were reviewed. Assessment of neurological outcome was also performed. Initial diagnostic testing for mutations in *KCNJ11*, *ABCC8* and *INS*, and for chromosome 6q24 abnormalities was negative. DNA samples from the patient's parents and his young brother were available for genetic testing.

Ethics Statement

The study was approved by the local ethics committees (Assistance Publique – Hôpitaux de Paris, ClinicalTrials.gov Identifier: NCT00610038), and both parents gave written informed consent for the genetic testing of their child.

Targeted capture and massive parallel sequencing

Approximately 187,000 coding exons from 3 µg of genomic DNA from the patient were captured using the Agilent SureSelect Human All Exon kit, following the manufacturer's protocols. Briefly, DNA was sheared by acoustic fragmentation (Covaris) and purified using the QIAquick PCR Purification Kit (Qiagen). The quality of the fragmentation and purification was assessed with the Agilent 2100 Bioanalyzer. The fragment ends were repaired and adaptors were ligated to the fragments (NEBNext DNA sample prep, New England Biolabs). The resulting DNA library was purified using the QIAquick PCR Purification Kit, amplified by PCR and captured by hybridization to the biotinylated RNA library "baits" (Agilent). Bound genomic DNA was purified with streptavidin coated magnetic Dynal beads (Invitrogen) and re-amplified. The whole-exome DNA library was sequenced on the Illumina Genome Analyzer IIx in 76 bp paired-end reads and using two channels.

Read mapping, variant analysis and quality test of the sequencing protocol

Sequence reads were mapped to the reference human genome (UCSC NCBI36/hg18) using the ELANDv2 software (Illumina). Variant detection was performed with the CASAVA software (version 1.6, Illumina) and filtered to fit a CASAVA quality threshold ≥10 and depth of ≥8× CASAVA filters duplicate reads and reads without matched pairs.

A genomic DNA sample from the patient was genotyped on the Illumina Human 1M-Duo DNA Analysis BeadChips (with a call rate of 99.3%) as previously described [14]. We assessed the rate of exomic single nucleotide polymorphisms (SNPs) present on the array that were identified by WES and we calculated the concordance between the two methods.

Mutation validation

The p.Q485H mutation identified via WES was confirmed using the Sanger method. Primer sequences and PCR conditions are available upon request to authors. The PCR fragment fitting the 9th exon of *ABCC8* was sequenced using a standard protocol and the automated 3730xl DNA Analyser (Applied Biosystems). Electrophoregram reads were assembled and analysed with the Variant Reporter software (Applied Biosystems).

Acknowledgments

We are grateful to the patient and his family for participation in the study. We thank very much Sylvie Pradines MD (Grenoble, France) who also took care of the patient. We thank Aurélie Dechaume, Julien Philippe and Sabrina Pereira for their technical work in the Sanger sequencing and the mutation validation.

Author Contributions

Conceived and designed the experiments: AB MV PF. Performed the experiments: ED SL. Analyzed the data: AB OS FDG SG. Contributed reagents/materials/analysis tools: KB AS CBC LL R. Scharfmann JD R. Sladek MP MV PF. Wrote the paper: AB R. Sladek MV PF.

References

1. Vaxillaire M, DP, Bonnefond A, Froguel P (2009) Breakthroughs in monogenic diabetes genetics: from pediatric forms to young adulthood diabetes. Pediatr Endocrinol Rev 6: 405–417.
2. Aguilar-Bryan L, Bryan J (2008) Neonatal diabetes mellitus. Endocr Rev 29: 265–291.
3. Murphy R, Ellard S, Hattersley AT (2008) Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. Nat Clin Pract Endocrinol Metab 4: 200–213.
4. Babenko AP, Polak M, Cave H, Busiah K, Czernichow P, et al. (2006) Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. N Engl J Med 355: 456–466.
5. Pearson ER, Flechtner I, Njolstad PR, Malecki MT, Flanagan SE, et al. (2006) Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. N Engl J Med 355: 467–477.
6. Gloyn AL, Pearson ER, Antcliff JF, Proks P, Bruining GJ, et al. (2004) Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N Engl J Med 350: 1838–1849.
7. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461: 272–276.
8. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. A method and server for predicting damaging missense mutations. Nat Methods 7: 248–249.
9. Polak M, Dechaume A, Cave H, Nimri R, Crosnier H, et al. (2008) Heterozygous missense mutations in the insulin gene are linked to permanent diabetes appearing in the neonatal period or in early infancy: a report from the French ND (Neonatal Diabetes) Study Group. Diabetes 57: 1115–1119.
10. Vaxillaire M, Dechaume A, Busiah K, Cave H, Pereira S, et al. (2007) New ABCC8 mutations in relapsing neonatal diabetes and clinical features. Diabetes 56: 1737–1741.
11. Edghill EL, Minton JA, Groves CJ, Flanagan SE, Patch AM, et al. Sequencing of candidate genes selected by beta cell experts in monogenic diabetes of unknown aetiology. Jop 11: 14–17.
12. Ellard S, Flanagan SE, Girard CA, Patch AM, Harries LW, et al. (2007) Permanent neonatal diabetes caused by dominant, recessive, or compound heterozygous SUR1 mutations with opposite functional effects. Am J Hum Genet 81: 375–382.
13. Hattersley A, Bruining J, Shield J, Njolstad P, Donaghue KC (2009) The diagnosis and management of monogenic diabetes in children and adolescents. Pediatr Diabetes 10(Suppl 12): 33–42.
14. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, et al. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445: 881–885.

nature genetics

Identity-by-descent filtering of exome sequence data identifies *PIGV* mutations in hyperphosphatasia mental retardation syndrome

Peter M Krawitz[1–3,11], Michal R Schweiger[1,2,11], Christian Rödelsperger[1–3], Carlo Marcelis[4], Uwe Kölsch[5], Christian Meisel[5], Friederike Stephani[4], Taroh Kinoshita[6], Yoshiko Murakami[6], Sebastian Bauer[2], Melanie Isau[1], Axel Fischer[1], Andreas Dahl[1], Martin Kerick[1], Jochen Hecht[1,3], Sebastian Köhler[2], Marten Jäger[2], Johannes Grünhagen[2], Birgit Jonske de Condor[2], Sandra Doelken[2], Han G Brunner[4], Peter Meinecke[7], Eberhard Passarge[8], Miles D Thompson[9], David E Cole[9], Denise Horn[2], Tony Roscioli[4,10], Stefan Mundlos[1–3] & Peter N Robinson[1–3]

Hyperphosphatasia mental retardation (HPMR) syndrome is an autosomal recessive form of mental retardation with distinct facial features and elevated serum alkaline phosphatase. We performed whole-exome sequencing in three siblings of a nonconsanguineous union with HPMR and performed computational inference of regions identical by descent in all siblings to establish *PIGV*, encoding a member of the GPI-anchor biosynthesis pathway, as the gene mutated in HPMR. We identified homozygous or compound heterozygous mutations in *PIGV* in three additional families.

Recessive mutations are relatively common in the human genome, but their identification remains challenging. Initial efforts at using exome sequencing for disease gene discovery[1] analyzed small numbers of unrelated individuals, removed variants that are common or not predicted to be deleterious and then searched for genes with such variants in all affected individuals. The analysis of the exome sequences of two siblings and two further unrelated individuals affected by the autosomal recessive Miller syndrome led to the identification of *DHODH* as the disease gene[2]. Subsequently, researchers analyzed whole genome sequences of the same two siblings and their parents to identify chromosomal regions in which both siblings had inherited identical haplotypes from both parents, which allowed the number of gene candidates for Miller syndrome to be reduced from 34 to 4, showing that linkage information represents a useful filter for genome sequence data[3]. These studies illustrate the utility of sophisticated algorithmic analysis in reducing the candidate gene set beyond what can be achieved by a simple intersection filter.

HPMR, also known as Mabry syndrome (MIM%239300), was initially described as an autosomal recessive syndrome characterized by mental retardation and greatly elevated alkaline phosphatase levels[4,5]. Within a group of individuals with this rare syndrome, a previous study[6] delineated a specific clinical entity characterized by a distinct facial gestalt including hypertelorism, long palpebral fissures, a broad nasal bridge and tip, and a mouth with downturned corners and a thin upper lip, as well as brachytelephalangy. More variable neurological features included seizures and muscular hypotonia[6].

Here, DNA from three siblings of nonconsanguineous parents with this subtype of HPMR was analyzed by exome sequencing (**Supplementary Figs. 1** and **2** and **Supplementary Table 1**). Whole-exome sequencing using the ABI SOLiD platform was performed following enrichment of exonic sequences using Agilent's SureSelect whole-exome enrichment. Called variants were filtered to exclude variants not found in all affected persons as well as common variants identified in the dbSNP130 or HapMap databases, which left 14 candidate genes on multiple chromosomes (**Table 1** and **Supplementary Tables 2–4**).

In this work, we developed a statistical model that allowed us to infer regions that are identical by descent (IBD) from the exome sequences of only the affected children of a family in which an autosomal recessive disorder segregates. In consanguineous families, affected siblings share two haplotypes that are inherited from a single common ancestor at the disease locus and are thus homozygous by descent. In nonconsanguineous families, the affected children inherit identical maternal and paternal haplotypes in a region surrounding the disease gene, meaning that both haplotypes originated from the same maternal and paternal haplotype but are not necessarily from an identical ancestor (IBD = 2).

We developed an algorithm based on a Hidden Markov Model (HMM), a type of Bayesian network that is used to infer a sequence of hidden (that is, unobservable) states. We used the HMM algorithm to identify chromosomal regions with IBD = 2 in the presence of noisy (that is, potentially erroneous) sequence data. It is not possible to measure the IBD = 2 state directly; it is only possible to determine whether the genotypes of the siblings are compatible with identity-by-state status, that is, whether each sibling has the same homozygous

[1]Max Planck Institute for Molecular Genetics, Berlin, Germany. [2]Institut für Medizinische Genetik, Charité Universitätsmedizin Berlin, Berlin, Germany. [3]Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité -Universitätsmedizin Berlin, Berlin, Germany. [4]Department of Human Genetics, University Medical Centre St. Radboud, Nijmegen, The Netherlands. [5]Institut für Medizinische Immunologie, Charité Universitätsmedizin Berlin, Berlin, Germany. [6]Department of Immunoregulation, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan. [7]Medizinische Genetik, Altonaer Kinderkrankenhaus, Hamburg, Germany. [8]Institut für Humangenetik, Universitätsklinikum Essen, Essen, Germany. [9]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. [10]Department of Molecular and Clinical Genetics, University of Sydney, Sydney, Australia. [11]These authors contributed equally to this work. Correspondence should be addressed to S.M. (stefan.mundlos@charite.de) or P.N.R. (peter.robinson@charite.de).

Received 29 March; accepted 3 August; published online 29 August 2010; doi:10.1038/ng.653

Table 1 Number of genes with nonsynonymous variants and acceptor or donor splice site mutations

	A1			A2			A3			A1 & A2 & A3		
Filter	Homozygous	Compound heterozygous	All	Homozygous	Compound heterozygous	All	Homozygous	Compound heterozygous	All	Homozygous	Compound heterozygous	All
NS/SS	2,752	934	3,385	2,900	1,090	3,640	2,806	1,070	3,625	1,728	273	1,928
Not in dbSNP130	182	35	216	218	47	262	200	38	235	12	2	14
IBD = 2	16	5	21	20	5	25	17	6	23	2	0	2
Sanger validated										2	0	2

Reducing the search space to the identical by descent (IBD = 2) regions and filtering out all common variants decreased the number of genes with nonsynonymous variants and acceptor or donor splice site mutation to two candidate genes. NS, nonsynonymous; SS, acceptor or donor splice site mutations.

or heterozygous genotype, a situation which we refer to as IBS*. In our model, every genetic locus was either IBD = 2 or IBD ≠ 2. The HMM was then used to predict the most likely sequence of IBD = 2 or IBD ≠ 2 chromosomal segments on the basis of the observed exome sequences of two or more affected siblings (**Supplementary Fig. 1** and **Supplementary Methods**).

HMM analysis decreased the search space to about 20% of the transcribed genome, reducing the number of candidate genes with mutations present in all three siblings from 14 to 2 (**Table 1**, **Supplementary Table 5** and **Supplementary Figs. 3–5**). The two mutations, c.[859G>A]+[859G>A] in *SLC9A1* and c.[1022C>A]+[1022C>A] in *PIGV*, were located within a 13-Mb homozygous block that was part of a larger 35-Mb IBD = 2 block. Runs of homozygosity of up to 4 Mb can occur in the European population even in individuals with no shared ancestors in the previous five to ten generations[7]. Both variants were confirmed with ABI Sanger sequencing and were not detected in 200 healthy, unrelated central European individuals. Further homozygous and compound heterozygous mutations were detected in *PIGV* in individuals from the families designated B[8], C[9] and D[10] (**Supplementary Note** and **Supplementary Tables 6** and **7**). All of these missense mutations affect evolutionarily highly conserved residues of PIGV (**Fig. 1a**).

PIGV, the second mannosyltransferase in the GPI anchor biosynthesis pathway[11], appeared to be of particular interest because alkaline phosphatase is a GPI-anchored protein. Over 100 mammalian proteins are modified by a glycosylphosphatidylinositol (GPI) anchor at their C terminus. The highly conserved backbone structure of the GPI anchor is synthesized in the endoplasmic reticulum through at least nine sequential reaction steps mediated by at least 18 proteins. GPI-anchored proteins comprise functionally divergent classes including hydrolytic enzymes, receptors, adhesion molecules and proteins with roles in the immune system[12]. Little is known to date about the phenotypic consequences of mutations of the GPI pathway in mammals. Abrogation of GPI biosynthesis in mice by knockdown of *Piga*, which encodes a protein that is involved in the first step of GPI-anchor biosynthesis, results in embryonic lethality[13]. Somatic loss-of-function mutations in *PIGA* in hematopoietic stem cells are associated with paroxysmal nocturnal hemoglobinuria[14], primarily because the progeny of affected stem cells are deficient in the GPI-anchored complement regulatory proteins CD55 and CD59, leading to the intravascular hemolysis characteristic of the disease. A promoter mutation in *PIGM*, encoding a subunit of the complex transferring the first mannose, reduces *PIGM* expression by over 90% and leads to an autosomal recessive syndrome characterized by hepatic venous thrombosis and absence seizures[15].

Defects in the GPI biosynthesis pathway can result in down-regulation of GPI-anchored proteins but not necessarily in a uniform reduction of all such proteins[12]. We therefore examined the surface expression of the GPI anchor itself on leukocytes of three individuals with HPMR using Alexa488-conjugated inactivated aerolysin (FLAER). All three subjects showed a substantial reduction of GPI-anchor expression. Correspondingly, expression of the GPI-anchored protein CD16 was markedly reduced (**Supplementary Fig. 6**). Wild-type *PIGV* cDNA and *PIGV* cDNA containing the p.Ala341Glu alteration were transiently transfected into PIGV-deficient Chinese hamster ovary (CHO) cells[11] to assess their effect on protein expression. Cells transfected with the mutant construct did not restore surface expression of GPI-anchored marker proteins (**Fig. 1b**), possibly because expressed PIGV protein levels were substantially reduced (**Fig. 1c**).

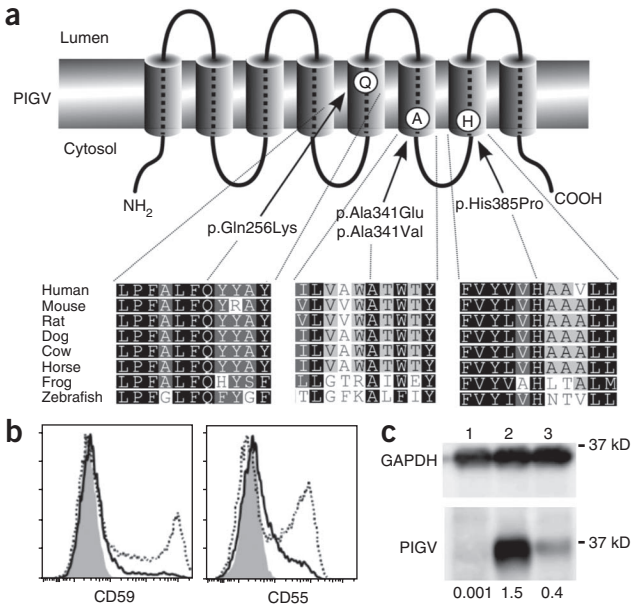


Figure 1 Identification of *PIGV* mutations in individuals with HPMR syndrome. (**a**) The homozygous *PIGV* mutation c.[1022C>A]+[1022C>A]; p.[Ala341Glu]+[Ala341Glu] was detected via whole exome sequencing in family A. Further homozygous and compound heterozygous mutations affecting evolutionarily highly conserved residues were found in three unrelated families: c.[1022C>A]+[1154A>C]; p.[Ala341Glu]+[His385Pro] in family B, c.[766C>A]+[766C>A]; p.[Gln256Lys]+[Gln256Lys] in family C, and c.[1022C>A]+[1022C>T]; p.[Ala341Glu]+[Ala341Val] in family D. (**b**) PIGV-deficient CHO cells were transiently transfected with wild-type (dashed lines) or p.Ala341Glu mutant (solid lines) *PIGV* cDNA in a weak expression vector or the empty vector (gray shadow). Wild-type *PIGV* efficiently restored the surface expression of CD59 (left) and CD55 (right), whereas p.Ala341Glu mutant PIGV induced only very low levels of CD59 and CD55. (**c**) PIGV protein levels were assessed 2 d after transfection of a control vector (lane 1), wild-type PIGV (lane 2) and PIGV with p.Ala341Glu (lane 3). The numbers beneath the gel indicate the relative intensity of PIGV to GAPDH expression.

In summary, we have identified *PIGV* mutations in HPMR using whole-exome capture and SOLiD sequencing in combination with an HMM algorithm to identify regions with IBD = 2 in siblings affected with autosomal recessive disorders. Our algorithm can be used in combination with other bioinformatic filters to streamline gene discovery in future exome sequencing projects.

Accession codes. The mutations in this work were numbered according to transcripts available in GenBank under the codes NM_003047.3 (*SLC9A1*) and NM_017837.2 (*PIGV*).

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

ACKNOWLEDGMENTS

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB 665) to S.M., by a grant from Bundesministerium für Bildung und Forschung (BMBF, project number 0313911) and an Australian National Health and Medical Research Council international research training fellowship to T.R., and by a grant of the Canadian Institutes of Health Research and Epilepsy Canada to M.D.T. We thank B. Fischer, U. Kornak, M. Ralser, E. van Beusekom, U. Marchfelder and D. Lefeber for their assistance in this project.

AUTHOR CONTRIBUTIONS

M.R.S., M.I. and A.D. performed targeted exome resequencing. P.M.K., C.R., A.F., M.K., S.B., S.K., M.J. and P.N.R. performed bioinformatic analysis. P.M.K., C. Marcelis, J.G., B.J.d.C., F.S. and T.R. performed mutation analysis and genotyping. D.H., C. Marcelis, M.D.T., D.E.C., S.D., P.M., E.P., T.R. and H.G.B. contributed to clinical evaluation of the affected individuals and delineation of the phenotype. P.M.K., U.K. and C. Meisel performed flow cytometric analysis. Y.M. and T.K. performed analysis of wild-type and A341E *PIGV* clones. P.M.K., M.R.S., D.H., J.H., H.G.B., P.N.R. and S.M. carried out the project planning and preparation of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

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

1. Ng, S.B. *et al. Nat. Genet.* **42**, 30–35 (2010).
2. Ng, S.B. *et al. Nature* **461**, 272–276 (2009).
3. Roach, J.C. *et al. Science* **328**, 636–639 (2010).
4. Mabry, C.C. *et al. J. Pediatr.* **77**, 74–85 (1970).
5. Kruse, K., Hanefeld, F., Kohlschutter, A., Rosskamp, R. & Gross-Selbeck, G. *J. Pediatr.* **112**, 436–439 (1988).
6. Horn, D., Schottmann, G. & Meinecke, P. *Eur. J. Med. Genet.* **53**, 85–88 (2010).
7. McQuillan, R. *et al. Am. J. Hum. Genet.* **83**, 359–372 (2008).
8. Rabe, P. *et al. Am. J. Med. Genet.* **41**, 350–354 (1991).
9. Marcelis, C.L., Rieu, P., Beemer, F. & Brunner, H.G. *Clin. Dysmorphol.* **16**, 73–76 (2007).
10. Thompson, M.D. *et al. Am. J. Med. Genet.* **152A**, 1661–1669 (2010).
11. Kang, J.Y. *et al. J. Biol. Chem.* **280**, 9489–9497 (2005).
12. Kinoshita, T., Fujita, M. & Maeda, Y. *J. Biochem.* **144**, 287–294 (2008).
13. Nozaki, M. *et al. Lab. Invest.* **79**, 293–299 (1999).
14. Takeda, J. *et al. Cell* **73**, 703–711 (1993).
15. Almeida, A.M. *et al. Nat. Med.* **12**, 846–851 (2006).

RAPID COMMUNICATION

Human Mutation

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

Unexpected Allelic Heterogeneity and Spectrum of Mutations in Fowler Syndrome Revealed by Next-Generation Exome Sequencing

Emilie Lalonde,[1,3†] Steffen Albrecht,[2†] Kevin C.H. Ha,[1,3] Karine Jacob,[3] Nathalie Bolduc,[3] Constantin Polychronakos,[3,4] Pierre Dechelotte,[5] Jacek Majewski,[1,3] and Nada Jabado[3,4*]

[1]McGill University and Genome Quebec Innovation Centre, Montreal, Canada; [2]Department of Pathology, Montreal Children's Hospital, McGill University Health Center, Montreal, Canada; [3]Departments of Human Genetics, McGill University Health Center, Montreal, Canada; [4]Departments of Pediatrics, Montreal Children's Hospital, McGill University Health Center, Montreal, Canada; [5]Department of Pathological Anatomy, CHU Clermont–Ferrand, Université d'Auvergne, France

Communicated by Graham R. Taylor

Received 16 April 2010; accepted revised manuscript 14 May 2010.
Published online 1 June 2010 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/humu.21293

ABSTRACT: Protein coding genes constitute approximately 1% of the human genome but harbor 85% of the mutations with large effects on disease-related traits. Therefore, efficient strategies for selectively sequencing complete coding regions (i.e., "whole exome") have the potential to contribute our understanding of human diseases. We used a method for whole-exome sequencing coupling Agilent whole-exome capture to the Illumina DNA-sequencing platform, and investigated two unrelated fetuses from nonconsanguineous families with Fowler Syndrome (FS), a stereotyped phenotype lethal disease. We report novel germline mutations in feline leukemia virus subgroup C cellular-receptor-family member 2, *FLVCR2*, which has recently been shown to cause FS. Using this technology, we identified three types of genetic abnormalities: point-mutations, insertions-deletions, and intronic splice-site changes (first pathogenic report using this technology), in the fetuses who both were compound heterozygotes for the disease. Although revealing a high level of allelic heterogeneity and mutational spectrum in FS, this study further illustrates the successful application of whole-exome sequencing to uncover genetic defects in rare Mendelian disorders. Of importance, we show that we can identify genes underlying rare, monogenic and recessive diseases using a limited number of patients ($n = 2$), in the absence of shared genetic heritage and in the presence of allelic heterogeneity.
Hum Mutat 31:918–923, 2010. © 2010 Wiley-Liss, Inc.

KEY WORDS: hydranencephaly–hydrocephaly; exome sequencing; deep sequencing; FLVCR2; Fowler syndrome

Introduction

Identification of rare monogenic diseases is of substantial interest to further our understanding of mechanisms of disease, biological pathways, and targeted therapies. To date, less than half of allelic variants responsible for monogenic disorders have been uncovered. This limited success is mainly due to the small numbers of available affected individuals for a given Mendelian disease (substantially reduced reproductive fitness leading to smaller pedigrees in affected families, etc.), or locus heterogeneity. All of these factors often lessen the power of traditional positional cloning strategies despite the advent of novel technologies, including single nucleotide polymorphism arrays (SNP microarrays). SNP arrays have been shown to help narrow areas of interest for further directed sequencing and are mainly powered (and thus limited) by families with shared genetic heritage (e.g., Mennonites, Amish, etc.) or with proven consanguinity [Laurier et al., 2006; Nishimura et al., 2005; Paisan-Ruiz et al., 2009; Strauss et al., 2005]. In contrast, deep resequencing of all human genes for discovery of allelic variants could potentially identify the gene underlying any given rare monogenic disease where a shared genetic heritage is not readily available [Shendure and Ji, 2008]. Massively parallel DNA sequencing technologies have rendered the whole-genome resequencing of individual humans increasingly practical, but cost remains a key consideration. An alternative approach involves the targeted resequencing of all protein-coding subsequences (that is, the exome), which requires $<2\%$ of the sequencing coverage required for the whole human genome [Choi et al., 2009; Ng et al., 2009] and has been very recently shown to be successful in identifying a Mendelian disorder [Ng et al., 2010].

The subject of our research, Fowler syndrome (FS), is a rare, prenatally lethal, disease that was first described in 1972 by Fowler et al. [1972]. There are less than 40 case reports [reviewed in Bessieres-Grattagliano et al., 2009; Williams et al., 2010], and all indicate a recurrent phenotype with hydrocephalus associated with progressive destruction of central nervous system tissue as a result of an unusual and characteristic proliferative vasculopathy. The occurrence of FS in consanguineous families and recurrence in both sexes is consistent with an autosomal recessive transmission. We investigated on the molecular level of two affected fetuses from two distinct French Canadian and French nonconsanguineous families. Using "next generation" exon sequencing (whole-exome sequencing coupling Agilent in-solution magnetic bead capture to the Illumina DNA sequencing platform), we identified

†The first two authors equally contributed to the manuscript.

*Correspondence to: Nada Jabado, Montreal Children's Hospital Research Institute, 4060 Ste. Catherine West, PT-239, Montreal, Qc, Canada H3Z 2Z3.
E-mail: nada.jabado@mcgill.ca

four separate germline alterations in the feline leukemia virus subgroup C cellular receptor family member 2 gene (*FLVCR2*, C14orf58 or FLJ20371 or MFSD7C; MIM♯ 610865) in the two patients who were compound heterozygotes for this disease. Mutations in *FLVCR2* have been very recently identified in FS [Meyer et al., 2010]. In three consanguineous families of Pakistani origin with FS, Meyer et al. [2010] performed a more traditional approach using autozygosity mapping (Affymetrix 250SNP arrays and microsatellite marker genotyping) in affected individuals. They demonstrated disease-linkage to chromosome 14q24.3, identified germline mutations in *FLVCR2* following direct sequencing of candidate genes within the target interval in the consanguineous cases, and further validated them in two additional patients with no documented consanguinity [Meyer et al., 2010]. Their approach illustrates a traditional gene mapping process, which even with the help of modern genotyping methods requires appropriate family structures and a sequential, usually lengthy process, involving identification of sizable candidate regions and systematic Sanger sequencing of a number of candidate genes. In our study, we use a much simpler and faster approach that makes full use of emerging technologies. We uncover two novel genetic alterations and a previously unsuspected high allelic heterogeneity in this very rare disorder. Importantly, we show that whole exome sequencing is a gene discovery approach that has major potential in identifying monogenic autosomal recessive disorders from a very small number affected individuals ($n = 2$) regardless of the need for shared genetic heritage status.

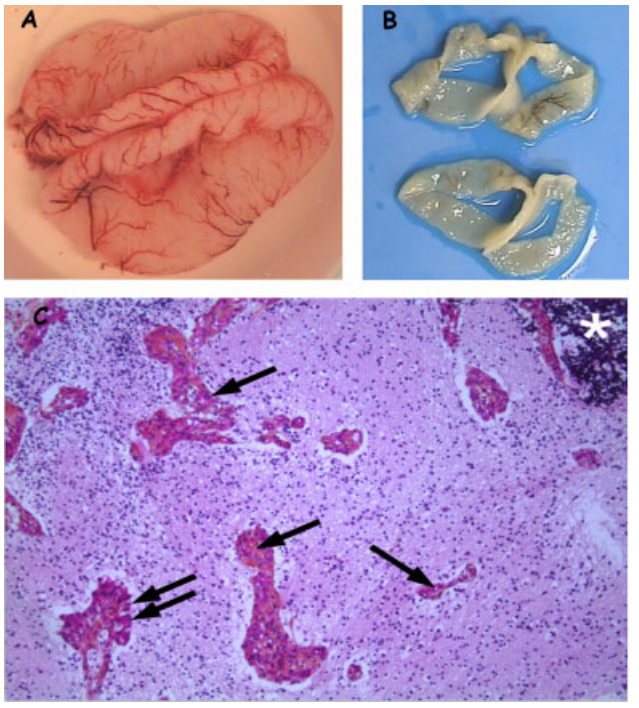


Figure 1. Neuropathological findings in case F1. **A:** Brain suspended in formalin solution. Notice the collapsed cerebral hemispheres. **B:** Coronal sections through the cerebral hemispheres. The cerebral mantle is extremely thin and there are no internal structures, such as basal ganglia and thalami. **C:** Histology. Medium power view shows hyperplastic microvessels (single arrows) and a glomeruloid structure (double arrow) along with an area of calcification (asterisk). (Hematoxylin-phloxine-saffron; original magnification 100 ×). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Materials and Methods

Patients Included in the Study

F1 was a female offspring of an unaffected nonconsanguineous French Canadian couple and was autopsied following pregnancy termination at the Montreal Children's Hospital. Pregnancy was terminated at 23 weeks' gestational age because of ventriculomegaly and limb deformities on fetal ultrasound examination. At autopsy, the fetus had features of fetal akinesia deformation sequence (arthrogryposis multiplex), with muscular atrophy, joint contractures, and cutaneous webbing. A 0.6-cm isolated atrial septal defect without other visceral malformations was also documented. Neuropathological findings showed bag-like cerebral hemispheres, with the cerebral mantle measuring only about 1–2 mm in thickness (Fig. 1A). No internal structures, such as basal ganglia or thalami, could be recognized (Fig. 1B). Brain stem and cerebellum were very small. Throughout the brain parenchyma, there were microcalcifications and hyperplastic microvessels forming glomeruloid structures (Fig. 1C). Some of the endothelial cells in the latter contained pale eosinophilic inclusions that stained strongly with PAS and were diastase resistant. Electron microscopy showed that these inclusions consisted of moderately electron-dense floculent material present in dilated rough endoplasmic reticulum (Fig. 1C). The residual brain parenchyma was highly disorganized. Spinal cord showed the same alterations and the eyes were normal. No vascular lesions were seen in any other organ. All of these abnormalities we report in F1 have been previously documented in the published FS cases [Al-Adnani et al., 2009; Castro-Gago et al., 2001; Fowler et al., 1972; Halder et al., 2003; Harper et al., 1983; Ibrahim et al., 2007; Usta et al., 2005; Witters et al., 2002]. Moreover, parents of F1 had lost a fetus with a phenotype of Fowler syndrome at 24 weeks of gestation 1 year prior to the termination of this pregnancy. Unfortunately, no frozen sample was available from the product of the first pregnancy and there were no other siblings in this family. DNA from F3, another fetus with FS sharing a similar phenotype (severely hypoplastic brain stem, cerebellum and spinal cord, and a glomeruloid vascular proliferation in all the brain including the cerebral cortex), was obtained through collaboration with a group in Clermont-Ferrand, France. This proband was from the kindred of family 4 described in Meyer et al. [2010]. All patient material was obtained following written consent from the families.

Exome Capture, Sequencing, and Short-Read Alignment

A total of 3 μg of DNA was subject to the exome capture procedure using the SureSelect Human All Exon Kit (solution magnetic bead capture), according to manufacturer's protocols. The captured DNA was then sequenced using Illumina GAIIx sequencing. We used three lanes of single-end, 76 base pair reads per sample. Reads were first processed using ELAND2 to extract reads that passed quality control in FASTQ format. Next, the reads for each individual were aligned to the human genome (UCSC hg18, NCBI build 36.1) using BWA (version 0.5.7) due to its proven ability to identify insertions and deletions [Li and Durbin, 2009]. We found that BWA consistently had a higher or equal proportion of mapped reads across each chromosome compared to ELAND2. Parameters were adjusted to allow for identification of insertions and deletions longer than four base pairs. Reads mapping nonuniquely or to unfinished sequences (chrN_random

and chrN_hap) were filtered while all remaining reads were used for downstream variant detection.

Mapping Statistics

We obtained approximately 78.5 million reads for F1 and 62.7 million reads for F3 that passed quality control and could be mapped to the reference genome. The exome was defined in a similar manner as in Ng et al. [2009] based on the NCBI Consensus Coding Sequence (CCDS) database (version 20090327), which includes the protein coding regions of well-supported genes. Although this exome definition may miss some poorly supported genes or exons, it ensures that the regions included are truly protein coding. This is important because the majority of deleterious mutations are in coding regions. By conservatively limiting ourselves to entries with the "public" status and merging exons with overlapping coordinates to account for transcript isoforms, we generated a set of 161,778 nonoverlapping regions covering 27.5 Mb. In total, approximately 27.2 Mb of the exome was mappable by all of our reads. The BWA alignment provided an average of 23.2 million and 20.1 million uniquely mapping reads per lane for F1 and F3, respectively. In addition, 2.1 million reads in F1 and 1.7 million reads in F3 aligned to multiple positions in the genome. These repetitive locations likely represent gene families, such as paralogues and pseudogenes. The remaining 13.9 million reads in F1 and 17.3 million reads in F3 were unmappable. Approximately 50% of all the mappable reads were targeted to the exome, whereas 63% targeted to regions within 200 bp of the exons. The capture of intronic sequences directly flanking the exons allows for detection of intronic variants that may affect pre-mRNA splicing. Overall, using three lanes of sequencing per sample we obtained an average coverage of 66.8 × of the CCDS exome. This depth of coverage allows confident calling of both homozygous and heterozygous mutations.

Variant Identification

SNP and indel information was extracted from the alignment data using the Samtools package (version 0.1.7) [Li and Durbin, 2009; Li et al., 2009]. Thresholds of Phred-like SNP quality of 20 and 50 were used for SNPs and indels, respectively. Additional custom filtering criteria were then imposed on SNPs to minimize the false positive rates. Specifically, to identify a variant base the following conditions had to be satisfied: the minimum coverage threshold was set to four reads per base; at least 30% of covering reads had to support the alternate base; only reads with a minimum base quality of Phred-like score of 35 for the SNP position were considered; and, for SNPs with sufficient coverage, the alternate base had to be supported by reads originating from both DNA strands.

Variant Annotation

After removal of potential artifacts, the variants were functionally annotated using an in-house script, created by Louis Letourneau, and SIFT 4.0.3 [Kumar et al., 2009]. Additionally, SeattleSeq Annotation 5.00 was used to identify potential splice-site SNPs. Using these annotations, variants were filtered first for those that are novel (not present in dbSNP or the 1,000 Genomes databases) and then for those that are likely deleterious. We predicted that damaging SNPs would be either missense, nonsense, or splice-site SNPs whereas damaging indels would be in coding regions. The variants that met these criteria were used for downstream analysis.

Identification of FS-Causative Gene

To find the disease-associated gene, we systematically searched for genes in the following order: a gene with a single homozygous mutation seen in both F1 and F3; a gene with two distinct homozygous mutations in F1 and F3; and finally a gene for which F1 and F3 are compound heterozygotes. Only mutations passing all filters previously mentioned were considered in this pipeline. Mutation numbering is based on the NM_017791.2 cDNA sequence, with position +1 referring to the A of the ATG initiation codon, in accordance with the journal guidelines (www.hgvs.org/mutnomen). For protein nomenclature, codon 1 refers the initiation codon from the reference NP_060261.

Results and Discussion

Exome Sequencing Identifies FLVCR2 as Causal in FS

We sequenced exomes in a total of two unrelated fetuses with FS (F1 and F3) from two independent kindreds using Illumina Genome Analyzer IIx. An average of 5.3 Gb of mappable sequence was generated per affected individual as single-end, 76-bp reads. To distinguish potentially pathogenic mutations from other variants, we focused on nonsynonymous (NS) variants, splice acceptor, and donor site mutations (SS), and short coding insertions or deletions (indels; I), anticipating that synonymous variants would be far less likely to be pathogenic. We also presumed that the variants responsible for FS would be rare and therefore likely to be previously unidentified. A novel variant was defined as one that did not exist in the databases used for comparison, namely, dbSNP and 1000 Genomes.

As FS is a recessive disease, each proband was required to have two inactivated copies of the same gene; that is, both fetuses are either homozygotes or compound heterozygotes for a NS/SS/I mutation in the disease-causing gene. To identify the causative gene, we systematically searched for genes with a unique homozygous mutation in both samples, genes with distinct homozygous mutations in either sample, and genes for which both individuals are compound heterozygotes. Each fetus was found to have at least a single homozygous NS/SS/I variant in ~4,300 genes and at least a single heterozygous NS/SS/I variant in ~8,200 genes. Filtering these variants against the dbSNP and 1,000 Genomes databases reduced the candidate gene pool 100-fold for homozygous mutations and 10-fold for heterozygous mutations (Table 1).

Table 1. The Number of Genes Found in the Two Probands with Potentially Pathogenic Variants in Various Categories

	F1	F3
Homozygous		
NS/SS/I	4,275	4,408
Novel	59	81
Common to both samples	0	0
Different mutations in same gene	0	0
Heterozygous		
NS/SS/I	8,370	8,040
Novel	868	998
Compound heterozygotes	83	92
Common to both samples	1	1

Genes were screened for both homozygous and heterozygous variants. Each subsequent row introduces an additional filter on the variants within the two classes of variants. Novel variants refer to those not present in the dbSNP and 1,000 Genomes databases.

We first searched, unsuccessfully, for a homozygous mutation present in both probands. Similarly, no evidence was found for a single gene harboring different homozygous mutations in the two samples. We then proceeded to search for potentially damaging compound heterozygous variants. Individually, 83 and 92 candidate genes were identified in F1 and F3, respectively. Taking the intersection of these two lists revealed only one candidate gene, *FLVCR2*, which encodes the feline leukemia virus subgroup C cellular receptor family member 2 gene. Thus, comparison of exome data from two unrelated affected individuals was sufficient to identify *FLVCR2* as the sole candidate gene for FS. This finding is further corroborated by the recent publication by Meyer et al. [2010] at the time of finalizing our data.

The protein FLVCR2 has 12 transmembrane domains joined by extracellular loops (Fig. 2) and is predicted to belong to the major facilitator superfamily [Finn et al., 2010; Hunter et al., 2009], a class of small solute transporter proteins responsive to chemiosmotic ion gradients. Although we can identify the location of our variants with respect to FLVCR2's domains, further work is required to fully understand the exact consequences of these variants.

F1 is a compound heterozygote for two novel mutations. The first, c.997C>T, is in exon 4 (coordinates 75169725–75169792) and introduces an Alanine to Valine change, p.Ala326Val, at the end of the sixth transmembrane domain. The second one is in the 5′ splice site downstream of exon 7 (coordinates 75177051–75177156): c.1341+2T>C. It changes the GT donor site to GC and is the first pathogenic splice-site mutation found by exome sequencing. This mutation most likely leads to the skipping of exon 7 or retention of the intron following exon 7, disrupting the tenth transmembrane domain and potentially resulting in nonsense mediated decay, although the exact effects of this splice site substitution cannot be determined from this data. Neither mutation is annotated as a polymorphism in dbSNP nor 1,000 Genomes.

F3 is also a compound heterozygote. We identified the c.1192C>G (encoding p.Leu398Val) substitution in exon 6 and the c.329_334del6 deletion, p.Asn110_Phe112delinsIle, in exon 1 described previously by the group of Meyer et al. [2010]. This deletion results in a frameshift and in the shortening of the first extracellular loop in by three amino acids while the substitution is located two amino acids upstream of the ninth transmembrane domain. The *FLVCR2* genotype of this patient is identical to case 6 in family 4 of this report.

To determine whether the wild-type amino acids that are affected by the mutations are phylogenetically conserved, we used phastCons scores based on a Multiz alignment of 28 vertebrates, taken from the UCSC Genome Browser [Blanchette et al., 2004]. Scores range 0 to 1, where 1 is the most conserved. For F1, c.977C>T has a score of 0.997375 and c.1341+2T>C has a score of 0.976378. The scores of the F3 mutations are 1 and 0.992126 for c329_334del6 and c.1192C>G, respectively. Each mutation found affects at least one phylogenetically conserved residue, located either within a transmembrane domain or an extracellular loop. Altering these amino acids could result in structural changes within these important domains potentially leading to functional protein changes. Further functional studies are warranted and ongoing to test the impact of these mutations on protein function.

Validation of the Novel FLVCR2 Mutations Identified in F1

We had no DNA available from the parents of F1 to document that they were heterozygotes for the genetic changes seen in *FLVCR2*. However, we tested 95 DNA samples from unaffected

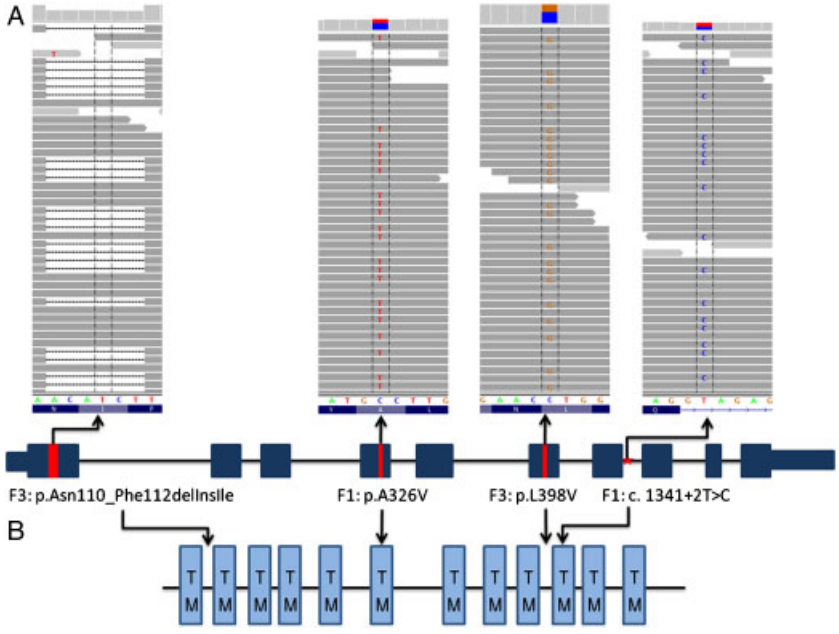


Figure 2. Schematic of the mutations found in F1 and F3 in both the gene and protein views. **A:** The four mutations in *FLVCR2* identified in F1 and F3 using whole exome sequencing as visualized using the Integrative Genomics Viewer from the Broad Institute (http://www.broadinstitute.org/igv). Mutations are also shown in their relative position along the gene (not to scale). **B:** Mutations in FLVCR2 are shown relative to the protein domains. Although the splice site mutation is not encoded, it likely results in retention of intron 7, disrupting the 10th transmembrane domain. (TM = transmembrane). Mutation numbering is based on the NM_017791.2 cDNA sequence, with position +1 referring to the A of the ATG initiation codon, in accordance with the journal guidelines (www.hgvs.org/mutnomen). For protein nomenclature, codon 1 refers the initiation codon from the reference NP_060261. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

individuals with similar geographic origin and did not identify the genetic changes seen in F1 on exons 1 and 4 of *FLVCR2*, further confirming that these are causative changes and not nonpathological variants. Along with the results of Meyer et al. [2010], these findings provide very strong evidence for the role of *FLVCR2* in FS, as well as the causative function the two new mutations identified by our group.

FLVCR2 Mutations and the Pathogenesis of FS

We identified four different mutations in a very rare monogenic disease: FS. In total, when combining our findings with those of Meyer et al. [2010], seven different mutations, including five missense variants, one nonsense mutation, and one deletion/insertion change, will have been found in *FLVCR2* and are potentially causative of FS. The features of FS have been reviewed in detail in two recent publications [Bessieres-Grattagliano et al., 2009; Williams et al., 2010]. Its pathological hallmark is the presence of hyperplastic microvessels forming glomeruloid structures throughout the central nervouse system (CNS). Some of these endothelial cells contain eosinophilic, PAS-positive, diastase-resistant inclusions that on electron microscopy are composed of moderately electron-dense floculent material present in dilated rough endoplasmic reticulum. The microvascular proliferation is associated with extensive necrosis and calcification of the CNS tissue. Destruction of brain tissue leads to hydranencephaly. The residual brain tissue is highly disorganized. The fetuses often but not invariably have features of fetal akinesia deformation sequence (arthrogryposis multiplex). There are no visceral malformations and the proliferative microangiopathy has not been described outside of the CNS. Importantly, it is unclear whether the microvascular proliferation is the result of the brain necrosis or its cause.

It is interesting to note that *FLVCR2* expression is not restricted to the CNS or to the vasculature (endothelium or pericytes). However, mutations in this gene induce a defect that is mainly, if not exclusively, restricted to the CNS and its vasculature, reenforcing the notion that brain angiogenesis is differentially regulated from angiogenesis in other body organs. FLVCR2 belongs to the major facilitator superfamily (MFS) of secondary carriers that transport small solutes in response to chemiosmotic ion gradients. It has been previously suggested that this transporter is specific for a calcium–chelate complex, and is involved in the regulation of growth and calcium metabolism, two processes involved in regulating angiogenesis, and the crosstalk between endothelial cells and pericytes. FLVCR1 (MIM♯ 609144), a paralog of FLVCR2 that shares 52% amino acid sequence identity, has been shown to be a human exporter of heme and to serve as a receptor for feline leukemia virus subgroup C (FeLV-C). FLVCR1 is critical for early erythropoiesis, and mutations in this receptor induce decreased to abrogated heme export, heme accumulation in erythroid progenitor, and their subsequent death [Keel et al., 2008]. *FLVCR1* null mice die in utero and have reduced myeloid and lymphoid cell growth and a disruption in early erythropoiesis with craniofacial and limb deformities [Keel et al., 2008]. A recent study shows that alternative splicing of *FLVCR1* transcripts lead to subsequent FLVCR1 insufficiency, and this in turn, acts as a contributing factor to the erythropoietic defect observed in Diamond-Blackfan anemia [Rey et al., 2008]. Wild-type FLVCR2 does not function as a receptor for FeLV-C or as a heme exporter; however, a single mutation of Asn463 to an acidic Asp residue in FLVCR2 extracellular domain 6 is sufficient to render it functional as an FeLV-C receptor [Brown et al., 2006]. This finding and data on FLVCR1 insufficiency secondary to a splicing defect in the gene reenforce our mutational findings as causative of abrogation/change of function in *FLVCR2*. If heme export is the major physiological function of FLVCR1, finding what FLVCR2 is exporting will prove crucial in understanding the pathogenesis of FS. Further studies are needed to determine what are the processes altered through *FLVCR2* mutations that are responsible of the exuberant vascular outgrowth and subsequent hydranencephaly.

We identify herein the gene responsible for a rare autosomal recessive disorder using direct whole exome resequencing on the smallest sample size reported for the identification of a monogenic disorder ($n = 2$). To date, there is only one other report that used this technology for the identification of a monogenic disorder. Ng et al. [2010] performed this type of methodology on four affected individuals including two siblings, and identified the gene responsible for Miller disease in addition to identifying mutations in *DNAH5* causing primary ciliary dyskineasia in the affected sibling pair. We were able, using this technology, to identify missense mutations, in-deletions, and intronic splice-site changes (first report using this technology to identify a pathogenic splice-site mutation), in only two individuals who were unrelated and who, in addition, came from kindred with no shared family heritage. Although we help uncover high allelic heterogeneity and mutation spectrum in a rare genetic disorder, FS, we show that this genetic approach is a much simpler and faster approach than traditional homozygosity mapping that makes full use of emerging technologies and is capable of identifying rare or common monogenic, recessive disorders. For dominant conditions, additional individuals are required, preferably from a single family whose family structure is known. By performing a linkage analysis prior to whole exome sequencing or resequencing, the region of interest could be narrowed to a reasonable size. Furthermore, this report emphasizes that it is now possible to directly identify a gene in a recessive Mendelian disorder using as few as two unrelated individuals and regardless of their kindred shared genetic heritage (consanguineous families and/or specific communities or genetic backgrounds) and in the presence of allelic heterogeneity.

Acknowledgments

The Cole Foundation (to N.J.). N.J. is the recipient of a Chercheur Boursier award from Fonds de Recherche en Sante du Quebec. J.M. is a recipient of a Canada Research Chair. We would like to thank Louis Letourneau for the use of his bioinformatics tools for filtering and annotations of SNPs. We would also like to thank the staff of the Genome Quebec sequencing platform, particularly Joana Dias, Alexandre Montpetit, and Pierre Lepage, with active participation and help in optimizing the techniques used in this project.

References

Al-Adnani M, Kiho L, Scheimberg I. 2009. Fowler syndrome presenting as a Dandy-Walker malformation: a second case report. Pediatr Dev Pathol 12:68–72.

Bessieres-Grattagliano B, Foliguet B, Devisme L, Loeuillet L, Marcorelles P, Bonniere M, Laquerriere A, Fallet-Bianco C, Martinovic J, Zrelli S, Leticee N, Cayol V, Etchevers HC, Vekemans M, Attie-Bitach T, Encha-Razavi F. 2009. Refining the clinicopathological pattern of cerebral proliferative glomeruloid vasculopathy (Fowler syndrome): report of 16 fetal cases. Eur J Med Genet 52:386–392.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, Haussler D, Miller W. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res 14:708–715.

Brown JK, Fung C, Tailor CS. 2006. Comprehensive mapping of receptor-functioning domains in feline leukemia virus subgroup C receptor FLVCR1. J Virol 80:1742–1751.

Castro-Gago M, Pintos-Martinez E, Forteza-Vila J, Iglesias-Diz M, Ucieda-Somoza R, Silva-Villar I, Codesido-Lopez J, Viso-Lorenzo A, Campos Y, Arenas J, Eirís-Puñal J. 2001. Congenital hydranencephalic–hydrocephalic syndrome with proliferative vasculopathy: a possible relation with mitochondrial dysfunction. J Child Neurol 16:858–862.

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloglu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci USA 106:19096–19101.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A. 2010. The Pfam protein families database. Nucleic Acids Res 38:D211–D222.

Fowler M, Dow R, White TA, Greer CH. 1972. Congenital hydrocephalus–hydrencephaly in five siblings, with autopsy studies: a new disease. Dev Med Child Neurol 14:173–188.

Halder A, Panigrahi I, Pal L. 2003. Fowler-like syndrome with extreme oligohydramnios, growth restriction and without muscular hypoplasia. Indian Pediatr 40:418–423.

Harper CG, Thomas W, ApSimon HT. 1983. Hydrocephalus and hypothalamic dysfunction in a young woman. Med J Aust 2:394–398.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. 2009. InterPro: the integrative protein signature database. Nucleic Acids Res 37:D211–D215.

Ibrahim A, Murthy P, Arunkalaivanan AS. 2007. A case of recurrent first-trimester Fowler syndrome. J Obstet Gynaecol 27:201–202.

Keel SB, Doty RT, Yang Z, Quigley JG, Chen J, Knoblaugh S, Kingsley PD, De Domenico I, Vaughn MB, Kaplan J, Palis J, Abkowitz JL. 2008. A heme export protein is required for red blood cell differentiation and iron homeostasis. Science 319:825–828.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081.

Laurier V, Stoetzel C, Muller J, Thibault C, Corbani S, Jalkh N, Salem N, Chouery E, Poch O, Licaire S, Danse JM, Amati-Bonneau P, Bonneau D, Mégarbané A, Mandel JL, Dollfus H. 2006. Pitfalls of homozygosity mapping: an extended consanguineous Bardet-Biedl syndrome family with two mutant genes (BBS2, BBS10), three mutations, but no triallelism. Eur J Hum Genet 14:1195–1203.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Meyer E, Ricketts C, Morgan NV, Morris MR, Pasha S, Tee LJ, Rahman F, Bazin A, Bessieres B, Dechelotte P, Yacoubi MT, Al-Adnani M, Marton T, Tannahill D, Trembath RC, Fallet-Bianco C, Cox P, Williams D, Maher ER. 2010. Mutations in FLVCR2 are associated with proliferative vasculopathy and hydranencephaly–hydrocephaly syndrome (Fowler syndrome). Am J Hum Genet 86:471–478.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. 2010. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42:30–35.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461:272–276.

Nishimura DY, Swiderski RE, Searby CC, Berg EM, Ferguson AL, Hennekam R, Merin S, Weleber RG, Biesecker LG, Stone EM, Sheffield VC. 2005. Comparative genomics and gene expression analysis identifies BBS9, a new Bardet-Biedl syndrome gene. Am J Hum Genet 77:1021–1033.

Paisan-Ruiz C, Scopes G, Lee P, Houlden H. 2009. Homozygosity mapping through whole genome analysis identifies a COL18A1 mutation in an Indian family presenting with an autosomal recessive neurological disorder. Am J Med Genet B Neuropsychiatr Genet 150B:993–997.

Rey MA, Duffy SP, Brown JK, Kennedy JA, Dick JE, Dror Y, Tailor CS. 2008. Enhanced alternative splicing of the FLVCR1 gene in Diamond Blackfan anemia disrupts FLVCR1 expression and function that are critical for erythropoiesis. Haematologica 93:1617–1626.

Shendure J, Ji H. 2008. Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145.

Strauss KA, Puffenberger EG, Craig DW, Panganiban CB, Lee AM, Hu-Lince D, Stephan DA, Morton DH. 2005. Genome-wide SNP arrays as a diagnostic tool: clinical description, genetic mapping, and molecular characterization of Salla disease in an Old Order Mennonite population. Am J Med Genet A 138A:262–267.

Usta IM, AbuMusa AA, Khoury NG, Nassar AH. 2005. Early ultrasonographic changes in Fowler syndrome features and review of the literature. Prenat Diagn 25:1019–1023.

Williams D, Patel C, Fallet-Bianco C, Kalyanasundaram K, Yacoubi M, Dechelotte P, Scott R, Bazin A, Bessieres B, Marton T, Cox P. 2010. Fowler syndrome—a clinical, radiological, and pathological study of 14 cases. Am J Med Genet A 152A:153–160.

Witters I, Moerman P, Devriendt K, Braet P, Van Schoubroeck D, Van Assche FA, Fryns JP. 2002. Two siblings with early onset fetal akinesia deformation sequence and hydranencephaly: further evidence for autosomal recessive inheritance of hydranencephaly, fowler type. Am J Med Genet 108:41–44.

Reference List / Abstracts

Nat Genet. 2010 Dec; 42(12):1131-4.

Exome sequencing identifies ACAD9 mutations as a cause of complex I deficiency.

Haack TB, Danhauser K, Haberberger B, Hoser J, Strecker V, Boehm D, Uziel G, Lamantea E, Invernizzi F, Poulton J, Rolinski B, Iuso A, Biskup S, Schmidt T, Mewes HW, Wittig I, Meitinger T, Zeviani M, Prokisch H.

[1] Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [2] Institute of Human Genetics, Technische Universität München, Munich, Germany. [3] These authors contributed equally to this work.

Abstract

An isolated defect of respiratory chain complex I activity is a frequent biochemical abnormality in mitochondrial disorders. Despite intensive investigation in recent years, in most instances, the molecular basis underpinning complex I defects remains unknown. We report whole-exome sequencing of a single individual with severe, isolated complex I deficiency. This analysis, followed by filtering with a prioritization of mitochondrial proteins, led us to identify compound heterozygous mutations in ACAD9, which encodes a poorly understood member of the mitochondrial acyl-CoA dehydrogenase protein family. We demonstrated the pathogenic role of the ACAD9 variants by the correction of the complex I defect on expression of the wildtype ACAD9 protein in fibroblasts derived from affected individuals. ACAD9 screening of 120 additional complex I-defective index cases led us to identify two additional unrelated cases and a total of five pathogenic ACAD9 alleles.

Proc Natl Acad Sci U S A. 2010 Dec 7;107(49):21104-9.

Comprehensive genetic testing for hereditary hearing loss using massively parallel sequencing.

Shearer AE, Deluca AP, Hildebrand MS, Taylor KR, Gurrola J 2nd, Scherer S, Scheetz TE, Smith RJ.

Department of Otolaryngology, Head and Neck Surgery, University of Iowa, Iowa City, IA 52242.

Abstract

The extreme genetic heterogeneity of nonsyndromic hearing loss (NSHL) makes genetic diagnosis expensive and time consuming using available methods. To assess the feasibility of target-enrichment and massively parallel sequencing technologies to interrogate all exons of all genes implicated in NSHL, we tested nine patients diagnosed with hearing loss. Solid-phase (NimbleGen) or solution-based (SureSelect) sequence capture, followed by 454 or Illumina sequencing, respectively, were compared. Sequencing reads were mapped using GSMAPPER, BFAST, and BOWTIE, and pathogenic variants were identified using a custom-variant calling and annotation pipeline (ASAP) that incorporates publicly available in silico pathogenicity prediction tools (SIFT, BLOSUM, Polyphen2, and Align-GVGD). Samples included one negative control, three positive controls (one biological replicate), and six unknowns (10 samples total), in which we genotyped 605 single nucleotide polymorphisms (SNPs) by Sanger sequencing to measure sensitivity and specificity for SureSelect-Illumina and NimbleGen-454 methods at saturating sequence coverage. Causative mutations were identified in the positive controls but not in the negative control. In five of six idiopathic hearing loss patients we identified the pathogenic mutation. Massively parallel sequencing technologies provide sensitivity, specificity, and reproducibility at levels sufficient to perform genetic diagnosis of hearing loss.

N Engl J Med. 2010 Dec 2;363(23):2220-7.

Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia.

Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, Garimella KV, Fisher S, Abreu J, Barry AJ, Fennell T, Banks E, Ambrogio L, Cibulskis K, Kernytsky A, Gonzalez E, Rudzicz N, Engert JC, DePristo MA, Daly MJ, Cohen JC, Hobbs HH, Altshuler D, Schonfeld G, Gabriel SB, Yue P, Kathiresan S.

Cardiovascular Research Center, Massachusetts General Hospital, and Department of Medicine, Boston University School of Public Health, Boston, MA 02114, USA.

Abstract

We sequenced all protein-coding regions of the genome (the "exome") in two family members with combined hypolipidemia, marked by extremely low plasma levels of low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides. These two participants were compound heterozygotes for two distinct nonsense mutations in ANGPTL3 (encoding the angiopoietin-like 3 protein). ANGPTL3 has been reported to inhibit lipoprotein lipase and endothelial lipase, thereby increasing plasma triglyceride and HDL cholesterol levels in rodents. Our finding of ANGPTL3 mutations highlights a role for the gene in LDL cholesterol metabolism in humans and shows the usefulness of exome sequencing for identification of novel genetic causes of inherited disorders. (Funded by the National Human Genome Research Institute and others.).

Nature. 2010 Nov 24

Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation.

Nazarian R, Shi H, Wang Q, Kong X, Koya RC, Lee H, Chen Z, Lee MK, Attar N, Sazegar H, Chodon T, Nelson SF, McArthur G, Sosman JA, Ribas A, Lo RS.

[1] Division of Dermatology/Department of Medicine, UCLA's Jonsson Comprehensive Cancer Center, 52-121 CHS, Los Angeles, California 90095-1750, USA [2] David Geffen School of Medicine, University of California, Los Angeles, California 90095-1750, USA [3] These authors contributed equally to this work.

Abstract

Activating B-RAF(V600E) (also known as BRAF) kinase mutations occur in 7% of human malignancies and 60% of melanomas. Early clinical experience with a novel class I RAF-selective inhibitor, PLX4032, demonstrated an unprecedented 80% anti-tumour response rate among patients with B-RAF(V600E)-positive melanomas, but acquired drug resistance frequently develops after initial responses. Hypotheses for mechanisms of acquired resistance to B-RAF inhibition include secondary mutations in B-RAF(V600E), MAPK reactivation, and activation of alternative survival pathways. Here we show that acquired resistance to PLX4032 develops by mutually exclusive PDGFR (also known as PDGFRB) upregulation or N-RAS (also known as NRAS) mutations but not through secondary mutations in B-RAF(V600E). We used PLX4032-resistant sub-lines artificially derived from B-RAF(V600E)-positive melanoma cell lines and validated key findings in PLX4032-resistant tumours and tumour-matched, short-term cultures from clinical trial patients. Induction of PDGFR RNA, protein and tyrosine phosphorylation emerged as a dominant feature of acquired PLX4032 resistance in a subset of melanoma sub-lines, patient-derived biopsies and short-term cultures. PDGFR-upregulated tumour cells have low activated RAS levels and, when treated with PLX4032, do not reactivate the MAPK pathway significantly. In another subset, high levels of activated N-RAS resulting from mutations lead to significant MAPK pathway reactivation upon PLX4032 treatment. Knockdown of PDGFR or N-RAS reduced growth of the respective PLX4032-resistant subsets. Overexpression of PDGFR or N-RAS(Q61K) conferred PLX4032 resistance to PLX4032-sensitive parental cell lines. Importantly, MAPK reactivation predicts MEK inhibitor sensitivity. Thus, melanomas escape B-RAF(V600E) targeting not through secondary B-RAF(V600E) mutations but via receptor tyrosine kinase (RTK)-mediated activation of alternative survival pathway(s) or activated RAS-mediated reactivation of the MAPK pathway, suggesting additional therapeutic strategies.

Nucleic Acids Res. 2010 Nov 1;38(20):6985-96.

Targeted next-generation sequencing of DNA regions proximal to a conserved GXGXXG signaling motif enables systematic discovery of tyrosine kinase fusions in cancer.

Chmielecki J, Peifer M, Jia P, Socci ND, Hutchinson K, Viale A, Zhao Z, Thomas RK, Pao W.

Weill Graduate School of Medical Sciences, Cornell University, New York, NY 10021, USA.

Abstract

Tyrosine kinase (TK) fusions are attractive drug targets in cancers. However, rapid identification of these lesions has been hampered by experimental limitations. Our in silico analysis of known cancer-derived TK fusions revealed that most breakpoints occur within a defined region upstream of a conserved GXGXXG kinase motif. We therefore designed a novel DNA-based targeted sequencing approach to screen systematically for fusions within the 90 human TKs; it should detect 92% of known TK fusions. We deliberately paired 'in-solution' DNA capture with 454 sequencing to minimize starting material requirements, take advantage of long sequence reads, and facilitate mapping of fusions. To validate this platform, we analyzed genomic DNA from thyroid cancer cells (TPC-1) and leukemia cells (KG-1) with fusions known only at the mRNA level. We readily identified for the first time the genomic fusion sequences of CCDC6-RET in TPC-1 cells and FGFR1OP2-FGFR1 in KG-1 cells. These data demonstrate the feasibility of this approach to identify TK fusions across multiple human cancers in a high-throughput, unbiased manner. This method is distinct from other similar efforts, because it focuses specifically on targets with therapeutic potential, uses only 1.5 µg of DNA, and circumvents the need for complex computational sequence analysis.

Nat Genet. 2010 Dec;42(12):1109-12.

A de novo paradigm for mental retardation.

Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, Del Rosario M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA.

[1] Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [2] These authors contributed equally to this work.

Abstract

The per-generation mutation rate in humans is high. De novo mutations may compensate for allele loss due to severely reduced fecundity in common neurodevelopmental and psychiatric diseases, explaining a major paradox in evolutionary genetic theory. Here we used a family based exome sequencing approach to test this de novo mutation hypothesis in ten individuals with unexplained mental retardation. We identified and validated unique non-synonymous de novo mutations in nine genes. Six of these, identified in six different individuals, are likely to be pathogenic based on gene function, evolutionary conservation and mutation impact. Our findings provide strong experimental support for a de novo paradigm for mental retardation. Together with de novo copy number variation, de novo point mutations of large effect could explain the majority of all mental retardation cases in the population.

BMC Genomics. 2010 Nov 18;11(1):641.

Combining target enrichment with barcode multiplexing for high throughput SNP discovery.

Cummings N, King R, Rickers A, Kaspi A, Lunke S, Haviv I, Jowett JB.

Abstract

BACKGROUND: The primary goal of genetic linkage analysis is to identify genes affecting a phenotypic trait. After localisation of the linkage region, efficient genetic dissection of the disease linked loci requires that functional variants are identified across the loci. These functional variations are difficult to detect due to extent of genetic diversity and, to date, incomplete cataloguing of the large number of variants present both within and between populations. Massively parallel sequencing platforms offer unprecedented capacity for variant discovery, however the number of samples analysed are still limited by cost per sample. Some progress has been made in reducing the cost of resequencing using either multiplexing methodologies or through the utilisation of targeted enrichment technologies which provide the ability to resequence genomic areas of interest rather that full genome sequencing.

RESULTS: We developed a method that combines current multiplexing methodologies with a solution-based target enrichment method to further reduce the cost of resequencing where region-specific sequencing is required. Our multiplex/enrichment strategy produced high quality data with nominal reduction of sequencing depth. We undertook a genotyping study and were successful in the discovery of novel SNP alleles in all samples at uniplex, duplex and pentaplex levels.

CONCLUSION: Our work describes the successful combination of a targeted enrichment method and index barcode multiplexing to reduce costs, time and labour associated with processing large sample sets. Furthermore, we have shown that the sequencing depth obtained is adequate for credible SNP genotyping analysis at uniplex, duplex and pentaplex levels.

Science. 2010 Oct 8;330(6001):228-31.

Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma.

Jones S, Wang TL, Shih IeM, Mao TL, Nakayama K, Roden R, Glas R, Slamon D, Diaz LA Jr, Vogelstein B, Kinzler KW, Velculescu VE, Papadopoulos N.

Ludwig Center for Cancer Genetics and Therapeutics and Howard Hughes Medical Institute, Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA.

Abstract

Ovarian clear cell carcinoma (OCCC) is an aggressive human cancer that is generally resistant to therapy. To explore the genetic origin of OCCC, we determined the exomic sequences of eight tumors after immunoaffinity purification of cancer cells. Through comparative analyses of normal cells from the same patients, we identified four genes that were mutated in at least two tumors. PIK3CA, which encodes a subunit of phosphatidylinositol-3 kinase, and KRAS, which encodes a well-known oncoprotein, had previously been implicated in OCCC. The other two mutated genes were previously unknown to be involved in OCCC: PPP2R1A encodes a regulatory subunit of serine/threonine phosphatase 2, and ARID1A encodes adenine-thymine (AT)-rich interactive domain-containing protein 1A, which participates in chromatin remodeling. The nature and pattern of the mutations suggest that PPP2R1A functions as an oncogene and ARID1A as a tumor-suppressor gene. In a total of 42 OCCCs, 7% had mutations in PPP2R1A and 57% had mutations in ARID1A. These results suggest that aberrant chromatin remodeling contributes to the pathogenesis of OCCC.

PLoS One. 2010 Oct 26;5(10):e13630.

Molecular diagnosis of neonatal diabetes mellitus using next-generation sequencing of the whole exome.

Bonnefond A, Durand E, Sand O, De Graeve F, Gallina S, Busiah K, Lobbens S, Simon A, Bellanné-Chantelot C, Létourneau L, Scharfmann R, Delplanque J, Sladek R, Polak M, Vaxillaire M, Froguel P.

CNRS-UMR-8199, Univ Lille Nord de France, UDSL, Lille, France.

Abstract

BACKGROUND: Accurate molecular diagnosis of monogenic non-autoimmune neonatal diabetes mellitus (NDM) is critical for patient care, as patients carrying a mutation in KCNJ11 or ABCC8 can be treated by oral sulfonylurea drugs instead of insulin therapy. This diagnosis is currently based on Sanger sequencing of at least 42 PCR fragments from the KCNJ11, ABCC8, and INS genes. Here, we assessed the feasibility of using the next-generation whole exome sequencing (WES) for the NDM molecular diagnosis.

METHODOLOGY/PRINCIPAL FINDINGS: We carried out WES for a patient presenting with permanent NDM, for whom mutations in KCNJ11, ABCC8 and INS and abnormalities in chromosome 6q24 had been previously excluded. A solution hybridization selection was performed to generate WES in 76 bp paired-end reads, by using two channels of the sequencing instrument. WES quality was assessed using a high-resolution oligonucleotide whole-genome genotyping array. From our WES with high-quality reads, we identified a novel non-synonymous mutation in ABCC8 (c.1455G>C/p.Q485H), despite a previous negative sequencing of this gene. This mutation, confirmed by Sanger sequencing, was not present in 348 controls and in the patient's mother, father and young brother, all of whom are normoglycemic.

CONCLUSIONS/SIGNIFICANCE: WES identified a novel de novo ABCC8 mutation in a NDM patient. Compared to the current Sanger protocol, WES is a comprehensive, cost-efficient and rapid method to identify mutations in NDM patients. We suggest WES as a near future tool of choice for further molecular diagnosis of NDM cases, negative for chr6q24, KCNJ11 and INS abnormalities.

Nat Genet. 2010 Oct;42(10):827-9.

Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome.

Krawitz PM, Schweiger MR, Rödelsperger C, Marcelis C, Kölsch U, Meisel C, Stephani F, Kinoshita T, Murakami Y, Bauer S, Isau M, Fischer A, Dahl A, Kerick M, Hecht J, Köhler S, Jäger M, Grünhagen J, de Condor BJ, Doelken S, Brunner HG, Meinecke P, Passarge E, Thompson MD, Cole DE, Horn D, Roscioli T, Mundlos S, Robinson PN.

Max Planck Institute for Molecular Genetics, Berlin, Germany. peter.robinson@charite.de

Abstract

Hyperphosphatasia mental retardation (HPMR) syndrome is an autosomal recessive form of mental retardation with distinct facial features and elevated serum alkaline phosphatase. We performed whole-exome sequencing in three siblings of a nonconsanguineous union with HPMR and performed computational inference of regions identical by descent in all siblings to establish PIGV, encoding a member of the GPI-anchor biosynthesis pathway, as the gene mutated in HPMR. We identified homozygous or compound heterozygous mutations in PIGV in three additional families.

Am J Hum Genet. 2010 Oct 8;87(4):553-9.

Mutations in SCARF2 are responsible for Van Den Ende-Gupta syndrome.

Anastasio N, Ben-Omran T, Teebi A, Ha KC, Lalonde E, Ali R, Almureikhi M, Der Kaloustian VM, Liu J, Rosenblatt DS, Majewski J, Jerome-Majewska LA.

Department of Human Genetics, McGill University, Montreal, Quebec H3A 1B1, Canada.

Abstract

Van Den Ende-Gupta syndrome (VDEGS) is an extremely rare autosomal-recessive disorder characterized by distinctive craniofacial features, which include blepharophimosis, malar and/or maxillary hypoplasia, a narrow and beaked nose, and an everted lower lip. Other features are arachnodactyly, camptodactyly, peculiar skeletal abnormalities, and normal development and intelligence. We present molecular data on four VDEGS patients from three consanguineous Qatari families belonging to the same highly inbred Bedouin tribe. The patients were genotyped with SNP microarrays, and a 2.4 Mb homozygous region was found on chromosome 22q11 in an area overlapping the DiGeorge critical region. This region contained 44 genes, including SCARF2, a gene that is expressed during development in a number of mouse tissues relevant to the symptoms described above. Sanger sequencing identified a missense change, c.773G>A (p.C258Y), in exon 4 in the two closely related patients and a 2 bp deletion in exon 8, c.1328_1329delTG (p.V443DfsX83), in two unrelated individuals. In parallel with the candidate gene approach, complete exome sequencing was used to confirm that SCARF2 was the gene responsible for VDEGS. SCARF2 contains putative epidermal growth factor-like domains in its extracellular domain, along with a number of positively charged residues in its intracellular domain, indicating that it may be involved in intracellular signaling. However, the function of SCARF2 has not been characterized, and this study reports that phenotypic effects can be associated with defects in the scavenger receptor F family of genes.

Hum Mol Genet. 2010 Oct 15;19(20):4112-20.

Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study.

Bowden DW, An SS, Palmer ND, Brown WM, Norris JM, Haffner SM, Hawkins GA, Guo X, Rotter JI, Chen YD, Wagenknecht LE, Langefeld CD.

Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA. dbowden@wfubmc.edu

Abstract

Family-based linkage analysis has been a powerful tool for identification of genes contributing to traits with monogenic patterns of inheritance. These approaches have been of limited utility in identification of genes underlying complex traits. In contrast, searches for common genetic variants associated with complex traits have been highly successful. It is now widely recognized that common variations frequently explain only part of the inter-individual variation in populations. 'Rare' genetic variants have been hypothesized to contribute significantly to phenotypic variation in the population. We have developed a combination of family-based linkage, whole-exome sequencing, direct sequencing and association methods to efficiently identify rare variants of large effect. Key to the successful application of the method was the recognition that only a few families in a sample contribute significantly to a linkage signal. Thus, a search for mutations can be targeted to a small number of families in a chromosome interval restricted to the linkage peak. This approach has been used to identify a rare (1.1%) G45R mutation in the gene encoding adiponectin, ADIPOQ. This variant explains a strong linkage signal (LOD > 8.0) and accounts for 17% of the variance in plasma adiponectin levels in a sample of 1240 Hispanic Americans and 63% of the variance in families carrying the mutation. Individuals carrying the G45R mutation have mean adiponectin levels that are 19% of non-carriers. We propose that rare variants may be a common explanation for linkage peaks observed in complex trait genetics. This approach is applicable to a wide range of family studies and has potential to be a discovery tool for identification of novel genes influencing complex traits.

J Exp Med. 2010 Oct 25;207(11):2307-12.

Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma.

Byun M, Abhyankar A, Lelarge V, Plancoulaine S, Palanduz A, Telhan L, Boisson B, Picard C, Dewell S, Zhao C, Jouanguy E, Feske S, Abel L, Casanova JL.

St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 10065, USA. miby769@rockefeller.edu

Abstract

Classic Kaposi sarcoma (KS) is exceedingly rare in children from the Mediterranean Basin, despite the high prevalence of human herpesvirus-8 (HHV-8) infection in this region. We hypothesized that rare single-gene inborn errors of immunity to HHV-8 may underlie classic KS in childhood. We investigated a child with no other unusually severe infectious or tumoral phenotype who died from disseminated KS at two years of age. Whole-exome sequencing in the patient revealed a homozygous splice-site mutation in STIM1, the gene encoding stromal interaction molecule 1, which regulates store-operated Ca(2+) entry. STIM1 mRNA splicing, protein production, and Ca(2+) influx were completely abolished in EBV-transformed B cell lines from the patient, but were rescued by the expression of wild-type STIM1. Based on the previous discovery of STIM1 deficiency in a single family with a severe T cell immunodeficiency and the much higher risk of KS in individuals with acquired T cell deficiencies, we conclude that STIM1 T cell deficiency precipitated the development of lethal KS in this child upon infection with HHV-8. Our report provides the first evidence that isolated classic KS in childhood may result from single-gene defects and provides proof-of-principle that whole-exome sequencing in single patients can decipher the genetic basis of rare inborn errors.

Am J Hum Genet. 2010 Sep 10;87(3):418-23.

Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome.

Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Steehouwer M, van Reeuwijk J, Kant SG, Roepman R, Knoers NV, Veltman JA, Brunner HG.

Department of Human Genetics, Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, The Netherlands.

Abstract

Sensenbrenner syndrome/cranioectodermal dysplasia (CED) is an autosomal-recessive disease that is characterized by craniosynostosis and ectodermal and skeletal abnormalities. We sequenced the exomes of two unrelated CED patients and identified compound heterozygous mutations in WDR35 as the cause of the disease in each of the two patients independently, showing that it is possible to find the causative gene by sequencing the exome of a single sporadic patient. With RT-PCR, we demonstrate that a splice-site mutation in exon 2 of WDR35 alters splicing of RNA on the affected allele, introducing a premature stop codon. WDR35 is homologous to TULP4 (from the Tubby superfamily) and has previously been characterized as an intraflagellar transport component, confirming that Sensenbrenner syndrome is a ciliary disorder.

Nat Genet. 2010 Sep;42(9):790-3.

Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J.

Department of Genome Sciences, University of Washington, Seattle, Washington, USA.

Abstract

We demonstrate the successful application of exome sequencing to discover a gene for an autosomal dominant disorder, Kabuki syndrome (OMIM%147920). We subjected the exomes of ten unrelated probands to massively parallel sequencing. After filtering against existing SNP databases, there was no compelling candidate gene containing previously unknown variants in all affected individuals. Less stringent filtering criteria allowed for the presence of modest genetic heterogeneity or missing data but also identified multiple candidate genes. However, genotypic and phenotypic stratification highlighted MLL2, which encodes a Trithorax-group histone methyltransferase: seven probands had newly identified nonsense or frameshift mutations in this gene. Follow-up Sanger sequencing detected MLL2 mutations in two of the three remaining individuals with Kabuki syndrome (cases) and in 26 of 43 additional cases. In families where parental DNA was available, the mutation was confirmed to be de novo (n = 12) or transmitted (n = 2) in concordance with phenotype. Our results strongly suggest that mutations in MLL2 are a major cause of Kabuki syndrome.

Hum Mutat. 2010 Aug;31(8):918-23.

Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing.

Lalonde E, Albrecht S, Ha KC, Jacob K, Bolduc N, Polychronakos C, Dechelotte P, Majewski J, Jabado N.

McGill University and Genome Quebec Innovation Centre, Montreal, Canada.

Abstract

Protein coding genes constitute approximately 1% of the human genome but harbor 85% of the mutations with large effects on disease-related traits. Therefore, efficient strategies for selectively sequencing complete coding regions (i.e., "whole exome") have the potential to contribute our understanding of human diseases. We used a method for whole-exome sequencing coupling Agilent whole-exome capture to the Illumina DNA-sequencing platform, and investigated two unrelated fetuses from nonconsanguineous families with Fowler Syndrome (FS), a stereotyped phenotype lethal disease. We report novel germline mutations in feline leukemia virus subgroup C cellular-receptor-family member 2, FLVCR2, which has recently been shown to cause FS. Using this technology, we identified three types of genetic abnormalities: point-mutations, insertions-deletions, and intronic splice-site changes (first pathogenic report using this technology), in the fetuses who both were compound heterozygotes for the disease. Although revealing a high level of allelic heterogeneity and mutational spectrum in FS, this study further illustrates the successful application of whole-exome sequencing to uncover genetic defects in rare Mendelian disorders. Of importance, we show that we can identify genes underlying rare, monogenic and recessive diseases using a limited number of patients (n=2), in the absence of shared genetic heritage and in the presence of allelic heterogeneity.

Am J Hum Genet. 2010 Aug 13;87(2):282-8.

Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome.

Pierce SB, Walsh T, Chisholm KM, Lee MK, Thornton AM, Fiumara A, Opitz JM, Levy-Lahad E, Klevit RE, King MC.

Department of Medicine, University of Washington, Seattle, WA 98195, USA.

Abstract

Perrault syndrome is a recessive disorder characterized by ovarian dysgenesis in females, sensorineural deafness in both males and females, and in some patients, neurological manifestations. No genes for Perrault syndrome have heretofore been identified. A small family of mixed European ancestry includes two sisters with well-characterized Perrault syndrome. Whole-exome sequencing of genomic DNA from one of these sisters revealed exactly one gene with two rare functional variants: HSD17B4, which encodes 17beta-hydroxysteroid dehydrogenase type 4 (HSD17B4), also known as D-bifunctional protein (DBP). HSD17B4/DBP is a multifunctional peroxisomal enzyme involved in fatty acid beta-oxidation and steroid metabolism. Both sisters are compound heterozygotes for HSD17B4 c.650A>G (p.Y217C) (maternal allele) and HSB17B4 c.1704T>A (p.Y568X) (paternal allele). The missense mutation is predicted by structural analysis to destabilize the HSD17B4 dehydrogenase domain. The nonsense mutation leads to very low levels of HSD17B4 transcript. Expression of mutant HSD17B4 protein in a compound heterozygote was severely reduced. Mutations in HSD17B4 are known to cause DBP deficiency, an autosomal-recessive disorder of peroxisomal fatty acid beta-oxidation that is generally fatal within the first two years of life. No females with DBP deficiency surviving past puberty have been reported, and ovarian dysgenesis has not previously been associated with this illness. Six other families with Perrault syndrome have wild-type sequences of HSD17B4. These results indicate that Perrault syndrome and DBP deficiency overlap clinically; that Perrault syndrome is genetically heterogeneous; that DBP deficiency may be underdiagnosed; and that whole-exome sequencing can reveal critical genes in small, nonconsanguineous families.

Am J Hum Genet. 2010 Jul 9;87(1):146-53.

Terminal osseous dysplasia is caused by a single recurrent mutation in the FLNA gene.

Sun Y, Almomani R, Aten E, Celli J, van der Heijden J, Venselaar H, Robertson SP, Baroncini A, Franco B, Basel-Vanagaite L, Horii E, Drut R, Ariyurek Y, den Dunnen JT, Breuning MH.

Center for Human and Clinical Genetics, Leiden University Medical Center, 2300RC Leiden, The Netherlands.

Abstract

Terminal osseous dysplasia (TOD) is an X-linked dominant male-lethal disease characterized by skeletal dysplasia of the limbs, pigmentary defects of the skin, and recurrent digital fibroma with onset in female infancy. After performing X-exome capture and sequencing, we identified a mutation at the last nucleotide of exon 31 of the FLNA gene as the most likely cause of the disease. The variant c.5217G>A was found in six unrelated cases (three families and three sporadic cases) and was not found in 400 control X chromosomes, pilot data from the 1000 Genomes Project, or the FLNA gene variant database. In the families, the variant segregated with the disease, and it was transmitted four times from a mildly affected mother to a more seriously affected daughter. We show that, because of nonrandom X chromosome inactivation, the mutant allele was not expressed in patient fibroblasts. RNA expression of the mutant allele was detected only in cultured fibroma cells obtained from 15-year-old surgically removed material. The variant activates a cryptic splice site, removing the last 48 nucleotides from exon 31. At the protein level, this results in a loss of 16 amino acids (p.Val1724_Thr1739del), predicted to remove a sequence at the surface of filamin repeat 15. Our data show that TOD is caused by this single recurrent mutation in the FLNA gene.

Am J Hum Genet. 2010 Jul 9;87(1):90-4.

Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82.

Walsh T, Shahin H, Elkan-Miller T, Lee MK, Thornton AM, Roeb W, Abu Rayyan A, Loulus S, Avraham KB, King MC, Kanaan M.

Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA 98195, USA.

Abstract

Massively parallel sequencing of targeted regions, exomes, and complete genomes has begun to dramatically increase the pace of discovery of genes responsible for human disorders. Here we describe how exome sequencing in conjunction with homozygosity mapping led to rapid identification of the causative allele for nonsyndromic hearing loss DFNB82 in a consanguineous Palestinian family. After filtering out worldwide and population-specific polymorphisms from the whole exome sequence, only a single deleterious mutation remained in the homozygous region linked to DFNB82. The nonsense mutation leads to an early truncation of the G protein signaling modulator GPSM2, a protein that is essential for maintenance of cell polarity and spindle orientation. In the mouse inner ear, GPSM2 is localized to apical surfaces of hair cells and supporting cells and is most highly expressed during embryonic development. Identification of GPSM2 as essential to the development of normal hearing suggests dysregulation of cell polarity as a mechanism underlying hearing loss.

Proc Natl Acad Sci U S A. 2010 Jul 13;107(28):12629-33.

Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing.

Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, Nord AS, Mandell JB, Swisher EM, King MC.

Department of Medicine, University of Washington, Seattle, WA 98195, USA.

Abstract

Inherited loss-of-function mutations in the tumor suppressor genes BRCA1, BRCA2, and multiple other genes predispose to high risks of breast and/or ovarian cancer. Cancer-associated inherited mutations in these genes are collectively quite common, but individually rare or even private. Genetic testing for BRCA1 and BRCA2 mutations has become an integral part of clinical practice, but testing is generally limited to these two genes and to women with severe family histories of breast or ovarian cancer. To determine whether massively parallel, "next-generation" sequencing would enable accurate, thorough, and cost-effective identification of inherited mutations for breast and ovarian cancer, we developed a genomic assay to capture, sequence, and detect all mutations in 21 genes, including BRCA1 and BRCA2, with inherited mutations that predispose to breast or ovarian cancer. Constitutional genomic DNA from subjects with known inherited mutations, ranging in size from 1 to >100,000 bp, was hybridized to custom oligonucleotides and then sequenced using a genome analyzer. Analysis was carried out blind to the mutation in each sample. Average coverage was >1200 reads per base pair. After filtering sequences for quality and number of reads, all single-nucleotide substitutions, small insertion and deletion mutations, and large genomic duplications and deletions were detected. There were zero false-positive calls of nonsense mutations, frameshift mutations, or genomic rearrangements for any gene in any of the test samples. This approach enables widespread genetic testing and personalized risk assessment for breast and ovarian cancer.

Nat Genet. 2010 Jun;42(6):483-5.

De novo mutations of SETBP1 cause Schinzel-Giedion syndrome.

Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BB, Veltman JA.

Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Abstract

Schinzel-Giedion syndrome is characterized by severe mental retardation, distinctive facial features and multiple congenital malformations; most affected individuals die before the age of ten. We sequenced the exomes of four affected individuals (cases) and found heterozygous de novo variants in SETBP1 in all four. We also identified SETBP1 mutations in eight additional cases using Sanger sequencing. All mutations clustered to a highly conserved 11-bp exonic region, suggesting a dominant-negative or gain-of-function effect.

Am J Hum Genet. 2010 May 14;86(5):743-8.

Massively parallel sequencing of exons on the X chromosome identifies RBM10 as the gene that causes a syndromic form of cleft palate.

Johnston JJ, Teer JK, Cherukuri PF, Hansen NF, Loftus SK; NIH Intramural Sequencing Center, Chong K, Mullikin JC, Biesecker LG.

National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892-4472, USA.

Abstract

Micrognathia, glossoptosis, and cleft palate comprise one of the most common malformation sequences, Robin sequence. It is a component of the TARP syndrome, talipes equinovarus, atrial septal defect, Robin sequence, and persistent left superior vena cava. This disorder is X-linked and severe, with apparently 100% pre- or postnatal lethality in affected males. Here we characterize a second family with TARP syndrome, confirm linkage to Xp11.23-q13.3, perform massively parallel sequencing of X chromosome exons, filter the results via a number of criteria including the linkage region, use a unique algorithm to characterize sequence changes, and show that TARP syndrome is caused by mutations in the RBM10 gene, which encodes RNA binding motif 10. We further show that this previously uncharacterized gene is expressed in midgestation mouse embryos in the branchial arches and limbs, consistent with the human phenotype. We conclude that massively parallel sequencing is useful to characterize large candidate linkage intervals and that it can be used successfully to allow identification of disease-causing gene mutations.

Genome Biol. 2009;10(10):R116.

Enrichment of sequencing targets from the human genome by solution hybridization.

Tewhey R, Nakano M, Wang X, Pabón-Peña C, Novak B, Giuffre A, Lin E, Happe S, Roberts DN, LeProust EM, Topol EJ, Harismendy O, Frazer KA.

Scripps Genomic Medicine, Scripps Translational Science Institute, The Scripps Research Institute, 3344 N Torrey Pines Court, La Jolla, CA 92037, USA. rtewhey@ucsd.edu

Abstract

To exploit fully the potential of current sequencing technologies for population-based studies, one must enrich for loci from the human genome. Here we evaluate the hybridization-based approach by using oligonucleotide capture probes in solution to enrich for approximately 3.9 Mb of sequence target. We demonstrate that the tiling probe frequency is important for generating sequence data with high uniform coverage of targets. We obtained 93% sensitivity to detect SNPs, with a calling accuracy greater than 99%.

Nat Genet. 2010 Apr;42(4):338-42.

PHF6 mutations in T-cell acute lymphoblastic leukemia.

Van Vlierberghe P, Palomero T, Khiabanian H, Van der Meulen J, Castillo M, Van Roy N, De Moerloose B, Philippé J, González-García S, Toribio ML, Taghon T, Zuurbier L, Cauwelier B, Harrison CJ, Schwab C, Pisecker M, Strehl S, Langerak AW, Gecz J, Sonneveld E, Pieters R, Paietta E, Rowe JM, Wiernik PH, Benoit Y, Soulier J, Poppe B, Yao X, Cordon-Cardo C, Meijerink J, Rabadan R, Speleman F, Ferrando A.

Institute for Cancer Genetics, Columbia University Medical Center, New York, New York, USA.

Abstract

Tumor suppressor genes on the X chromosome may skew the gender distribution of specific types of cancer. T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematological malignancy with an increased incidence in males. In this study, we report the identification of inactivating mutations and deletions in the X-linked plant homeodomain finger 6 (PHF6) gene in 16% of pediatric and 38% of adult primary T-ALL samples. Notably, PHF6 mutations are almost exclusively found in T-ALL samples from male subjects. Mutational loss of PHF6 is importantly associated with leukemias driven by aberrant expression of the homeobox transcription factor oncogenes TLX1 and TLX3. Overall, these results identify PHF6 as a new X-linked tumor suppressor in T-ALL and point to a strong genetic interaction between PHF6 loss and aberrant expression of TLX transcription factors in the pathogenesis of this disease. reproducibility at levels sufficient to perform genetic diagnosis of hearing loss.

Nat Genet. 2010 Jan;42(1):30-5.

Exome sequencing identifies the cause of a mendelian disorder.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ.

Department of Genome Sciences, University of Washington, Seattle, Washington, USA.

Abstract

We demonstrate the first successful application of exome sequencing to discover the gene for a rare mendelian disorder of unknown cause, Miller syndrome (MIM%263750). For four affected individuals in three independent kindreds, we captured and sequenced coding regions to a mean coverage of 40x and sufficient depth to call variants at approximately 97% of each targeted exome. Filtering against public SNP databases and eight HapMap exomes for genes with two previously unknown variants in each of the four individuals identified a single candidate gene, DHODH, which encodes a key enzyme in the pyrimidine de novo biosynthesis pathway. Sanger sequencing confirmed the presence of DHODH mutations in three additional families with Miller syndrome. Exome sequencing of a small number of unrelated affected individuals is a powerful, efficient strategy for identifying the genes underlying rare mendelian disorders and will likely transform the genetic analysis of monogenic traits.

Methods. 2010 Apr;50(4):S15-8.

Rapid quantification of DNA libraries for next-generation sequencing.

Buehler B, Hogrefe HH, Scott G, Ravi H, Pabón-Peña C, O'Brien S, Formosa R, Happe S.

Agilent Technologies, 11011 N. Torrey Pines Road, La Jolla, CA 92037, USA.

Abstract

The next-generation DNA sequencing workflows require an accurate quantification of the DNA molecules to be sequenced which assures optimal performance of the instrument. Here, we demonstrate the use of qPCR for quantification of DNA libraries used in next-generation sequencing. In addition, we find that qPCR quantification may allow improvements to current NGS workflows, including reducing the amount of library DNA required, increasing the accuracy in quantifying amplifiable DNA, and avoiding amplification bias by reducing or eliminating the need to amplify DNA before sequencing.

Nature. 2009 Sep 10;461(7261):272-6. Epub 2009 Aug 16.

Targeted capture and massively parallel sequencing of 12 human exomes.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J.

Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. sarahng@u.washington.edu

Abstract

Genome-wide association studies suggest that common genetic variants explain only a modest fraction of heritable risk for common diseases, raising the question of whether rare variants account for a significant fraction of unexplained heritability. Although DNA sequencing costs have fallen markedly, they remain far from what is necessary for rare and novel variants to be routinely identified at a genome-wide scale in large cohorts. We have therefore sought to develop second-generation methods for targeted sequencing of all protein-coding regions ('exomes'), to reduce costs while enriching for discovery of highly penetrant variants. Here we report on the targeted capture and massively parallel sequencing of the exomes of 12 humans. These include eight HapMap individuals representing three populations, and four unrelated individuals with a rare dominantly inherited disorder, Freeman-Sheldon syndrome (FSS). We demonstrate the sensitive and specific identification of rare and common variants in over 300 megabases of coding sequence. Using FSS as a proof-of-concept, we show that candidate genes for Mendelian disorders can be identified by exome sequencing of a small number of unrelated, affected individuals. This strategy may be extendable to diseases with more complex genetics through larger sample sizes and appropriate weighting of non-synonymous variants by predicted functional impact.

Nat Biotechnol. 2009 Feb;27(2):182-9. Epub 2009 Feb 1.

Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C.

Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. gnirke@broad.mit.edu

Abstract

Targeting genomic loci by massively parallel sequencing requires new methods to enrich templates to be sequenced. We developed a capture method that uses biotinylated RNA 'baits' to fish targets out of a 'pond' of DNA fragments. The RNA is transcribed from PCR-amplified oligodeoxynucleotides originally synthesized on a microarray, generating sufficient bait for multiple captures at concentrations high enough to drive the hybridization. We tested this method with 170-mer baits that target >15,000 coding exons (2.5 Mb) and four regions (1.7 Mb total) using Illumina sequencing as read-out. About 90% of uniquely aligning bases fell on or near bait sequence; up to 50% lay on exons proper. The uniformity was such that approximately 60% of target bases in the exonic 'catch', and approximately 80% in the regional catch, had at least half the mean coverage. One lane of Illumina sequence was sufficient to call high-confidence genotypes for 89% of the targeted exon space.

Leukemia. 2010 Oct;24(10):1799-804.

Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies.

Makishima H, Jankowska AM, Tiu RV, Szpurka H, Sugimoto Y, Hu Z, Saunthararajah Y, Guinta K, Keddache MA, -Putnam P, Sekeres MA, Moliterno AR, List AF, McDevitt MA, Maciejewski JP.

Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH, USA

Abstract

Segmental somatic uniparental disomy (UPD) is a common defect in many cases of myelodysplastic syndrome, myeloproliferative neoplasms , and acute myeloid leukemia (AML). Previously, we reviewed single-nucleotide polymorphism arrays karyograms in a large screen of myeloid malignancies, and identified 15 patients with somatic UPD of the long arm of chromosome 7. In this study, we sought to identify the gene(s) that might be mutated and drive the generation of uniparental disomy UPD7 in myeloid malignancies. We generated exome chromosome 7 libraries that were enriched for the content of chromosome 7 coding sequences using the SureSelect capture synthetic biotinylated RNA probes from Agilent (Santa Clara, CA, USA), tiling all the coding regions from chromosome 7. Libraries were generated from two cases with UPD7q and subjected to high-throughput sequencing on an Illumina Genome Analyzer IIx (San Diego, CA, USA). Our study demonstrates for the first time detection of EZH2 mutations in patients with aggressive myeloid malignancies. Similar to previously identified mutations in UTX and ASXL1 in myeloid malignancies, EZH2 is a polycomb-associated gene. Our findings support an increasing possibility that mutations in these genes represent a new class of molecular lesions conveying a clonal epigenetic instability phenotype. Consequently, mutations involved in epigenetic regulation can constitute leukemogenic events.

Neuron. 2010 Dec 9;68(5):857-864.

Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS.

Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, Gibbs JR, Brunetti M, Gronka S, Wuu J, Ding J, McCluskey L, Martinez-Lage M, Falcone D, Hernandez DG, Arepalli S, Chong S, Schymick JC, Rothstein J, Landi F, Wang YD, Calvo A, Mora G, Sabatelli M, Monsurrò MR, Battistini S, Salvi F, Spataro R, Sola P, Borghero G; The ITALSGEN Consortium, Galassi G, Scholz SW, Taylor JP, Restagno G, Chiò A, Traynor BJ.

Neuromuscular Diseases Research Group, Laboratory of Neurogenetics, Porter Neuroscience Building, NIA, NIH, Bethesda, MD 20892, USA.

Abstract

Using exome sequencing, we identified a p.R191Q amino acid change in the valosin-containing protein (VCP) gene in an Italian family with autosomal dominantly inherited amyotrophic lateral sclerosis (ALS). Mutations in VCP have previously been identified in families with Inclusion Body Myopathy, Paget disease, and Frontotemporal Dementia (IBMPFD). Screening of VCP in a cohort of 210 familial ALS cases and 78 autopsy-proven ALS cases identified four additional mutations including a p.R155H mutation in a pathologically proven case of ALS. VCP protein is essential for maturation of ubiquitin-containing autophagosomes, and mutant VCP toxicity is partially mediated through its effect on TDP-43 protein, a major constituent of ubiquitin inclusions that neuropathologically characterize ALS. Our data broaden the phenotype of IBMPFD to include motor neuron degeneration, suggest that VCP mutations may account for 1%-2% of familial ALS, and provide evidence directly implicating defects in the ubiquitination/protein degradation pathway in motor neuron degeneration.

2100 Bioanalyzer Reference List

Genome Biol. 2010 Dec 8;11(12):R119. [Epub ahead of print]

Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition.

Adey A, Morrison HG, No Last Name A, Xun X, Kitzman JO, Turner EH, Stackhouse B, Mackenzie AP, Caruccio NC, Zhang X, Shendure J.

Agilent Technologies, Application Note, Publication Number: 5990-5008EN

Improving sample quality for SureSelect target enrichment and next-generation sequencing with the High Sensitivity DNA kit

Kirill Gromadski, Ruediger Salowsky, Susanne Glueck

Agilent Technologies, Technical Note, Publication Number: 5990-4417EN

Performance characteristics of the High Sensitivity DNA assay for the Agilent 2100 Bioanalyzer

www.agilent.com/genomics/sureselect

For Research Use Only. Not for use with diagnostics procedures.

Published in USA, January 13, 2011
Publication Number 5990-7233EN

Agilent Technologies