Year : 2016

NEXT GENERATION SEQUENCING IN HUMAN DISEASES

Royer-Bertrand Béryl

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Département de Biologie Computationnelle

NEXT GENERATION SEQUENCING IN HUMAN DISEASES

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Béryl ROYER-BERTRAND

Ingénieur en Biosciences de l'INSA, Lyon, France

Jury

Prof. Barbara Wildhaber, Président
Dr. Carlo Rivolta, Directeur de thèse
Dr. Patrick Descombe, Expert
Dr. Pascal Escher, Expert

Lausanne 2016

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président · e	Madame Prof. Barbara **Wildhaber**
Directeur · rice de thèse	Monsieur Dr Carlo **Rivolta**
Experts · es	Monsieur Dr Pascal **Escher**
	Monsieur Dr Patrick **Descombes**

le Conseil de Faculté autorise l'impression de la thèse de

Madame Béryl Royer-Bertrand

Ingénieur en Biosciences de l' INSA, Lyon, France

intitulée

Next generation Sequencing in human diseases

Lausanne, le 9 décembre 2016

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Barbara Wildhaber

Table of Content

Acknowledgements

Foremost, I would like to thank my thesis supervisor, Dr. Carlo Rivolta, for giving me the possibility to join his group, for his useful comments and suggestions throughout the Ph.D. I also would like to acknowledge Dr. Alexandre Moulin and Prof. Andrea Superti-Furga for their guidance.

I would like to thank all the DGM/DBC members for their advice during the whole doctorate and the creation of this thesis. A special thank goes to Nicola Bedoni, who has always been present since the beginning of the thesis, as well as Katarina Cisarova and Mathieu Quinodoz for their patience, their fun and support. I am grateful to all my friends, colleagues and members of ADAS but especially to Nada and Anaïs for always being close no matter how far we were.

Finally, the biggest and warmest thank you goes to my parents and my family, who always believed in me and supported me throughout my entire studies and all my decisions. In particular, I want to thank Aurélien, who always helped me through every crisis and has been the most patient listener, great adviser and best friend (and much more).

Summary

Over the past years, Next-Generation Sequencing (NGS) has become an effective and accurate tool for the detection of potential causes and for the better understanding of human genetic diseases. Here, I applied bioinformatics analysis of NGS data to shed new light on the biology of two very different types of diseases, ocular melanomas and Mendelian diseases. For each patient, we sequenced either the coding regions of the genes, by means of Whole-Exome Sequencing (WES), or the full genome, by means of Whole-Genome Sequencing (WGS), this latter giving more complete and more diverse genetics alterations, compared to WES.

In the first part, I exploit the WGS data to understand the genetic causes of two subtypes of ocular melanomas, Uveal Melanoma (UM) and Conjunctival Melanoma (CM), both characterized by an overall high rate of metastatic transformation and mortality. Via the full genome sequencing of 2 patients with CM, we highlighted the presence of a high number of somatic mutations with a majority of cytosine to tyrosine (C>T) substitutions, specifically found at the 3' end of pyrimidine dinucleotide. They are classical features of UV light mutational signature. From a cohort of 34 patients with UM, we determined the genomic landscape of UM, which is very distinct from CM. More specifically, we detailed four classes of UM, depending on specific genetic alterations, corresponding to large copy number variants and mutations in driver genes (*BAP1*, *EIF1AX*, and *SF3B1*), which could be useful to infer a potential metastatic development.

In the second part, I was dealing with genomic analysis of rare inherited Mendelian diseases, more specifically with retinal degenerations and skeletal or metabolic disorders. I developed bioinformatics procedures aimed at detecting the variations present in the WES or the WGS data of patients of interest. Then, I set up *in silico* analyses to identify the disease-causing candidate DNA variants and to reduce the overall noise, which is rather abundant in all NGS-based approaches. These tools allowed us to discover two new developmental syndromes (caused by *HSPA9* and *NANS*), to characterize four genes causing retinal degeneration (*CEP78* and *TTLL5*), a skeletal phenotype (*ASAH1*) and a multisystem disorder (*NBAS*).

Résumé

Ces dernières années, le séquençage à haut débit (en anglais « Next-Generation Sequencing » - NGS) est devenu un outil efficace et précis pour la détection des causes potentielles des maladies génétiques humaines. Durant ma thèse, j'ai appliqué des analyses bioinformatiques pour étudier deux types de maladies bien différentes, les mélanomes oculaires et les maladies mendéliennes. Pour chaque patient, nous avons séquencé soit les régions codantes des gènes via la technique de « Whole Exome Sequencing » (WES), soit le génome en entier via du « Whole Genome Sequencing » (WGS). En comparaison au WES, le WGS donne des informations plus précises et plus complexes sur toutes les altérations génétiques.

D'une part, j'ai cherché à comprendre les causes génétiques de deux mélanomes oculaires, le Mélanome de l'Uvée (UM) et le Mélanome de la Conjonctive (CM), caractérisés par un taux élevé de métastases et un faible pronostic vital. En séquençant le génome de deux patients atteints de CM, j'ai mis en avant la présence d'un très grand nombre de mutations somatiques ; avec une majorité de substitutions de cytosine en tyrosine (C>T), trouvées spécifiquement du côté 3' des couplets de pyrimidine. Ce sont des modifications caractéristiques de l'action des rayons ultra-violets. A partir d'une cohorte de 34 patients atteints d'UM, j'ai déterminé les bases génomiques de ce cancer, très différentes de celles de CM. J'ai notamment détaillé quatre classes d'UM, chacune ayant des altérations génétiques spécifiques, consistant en de larges variations du nombre de copies et des mutations dans des gènes « drivers » (*BAP1*, *EIF1AX*, et *SF3B1*), ce qui pourrait être utile pour prédire des potentiels développements métastatiques.

D'une autre part, j'ai étudié des maladies mendéliennes rares, plus particulièrement les dégénérations rétiniennes ainsi que les maladies du squelette et métaboliques. J'ai développé des pipelines bioinformatiques détectant les variations génétiques présentes dans le WES ou le WGS de nos patients. J'ai ensuite mis en place des analyses *in-silico* pour identifier les variantes pouvant potentiellement expliquer la maladie, tout en réduisant les faux positifs, qui sont très fréquents en NGS. Ces outils m'ont permis de découvrir, avec mes collaborateurs, deux nouveaux syndromes développementaux (causés par les gènes *HSPA9* et *NANS*) et de caractériser quatre autres gènes causant une dégénérescence rétinienne (*CEP78* et *TTLL5*), une maladie squelettique (*ASAH1*) et une maladie multi-systémique (*NBAS*).

1 General introduction

Genetics is the study of heredity, conceptually acquiring its to-date from the pioneering work of Gregor Mendel in the mid-19th century. Developments in all fields of genetics and genetic technology in the first half of the 20th century provided the basis for the later developments and in particular for the implementation of Next-Generation Sequencing (NGS). The aim of this thesis focuses on the use of NGS technologies, their adaptation and improvement from a bioinformatics standpoint, in order to identify and better understand the causes underlying a large spectrum of human genetic disorders. In the first part of the Introduction, basic concepts of human genetics are presented, together with a description of the two main types of diseases studied in this thesis, i.e. ocular melanomas and Mendelian disorders. The second part describes the different NGS technologies. The third part focuses on the bioinformatics workflows or "pipelines" that were developed, optimized and used in order to process raw NGS-derived data for the identification of disease-causing mutations.

1.1 Human diseases with a genetic origin

DNA, genes and genetic variations

The instructions for the normal development and functioning of all living organisms are contained in the DeoxyriboNucleic Acid (DNA). Composed of more than three billions of bases in humans, this molecule is made of a repetition of four different nucleotides: Adenine (A), Cytosine (C), Guanine (G) and Thymine (T). The human DNA is compacted in the cell nucleus into 23 pairs of chromosomes, 22 of which are autosomal and one is a pair of sex chromosomes, either two X-chromosomes for females or one X-chromosome and one Y-chromosome for males. Less than 2% of the total DNA codes for proteins, and the molecular unit of heredity consists of the so-called genes. Genes are composed of coding parts (the exons), of non-coding parts (the introns) surrounded by

untranslated but transcribed regions (the UTRs) (Figure 1). In exons, a sequence of three nucleotides corresponds to a codon, translated into one of 20 different amino acids. The boundaries between exons and introns are defined by splicing sites, which provide relevant signals to the cell splicing machinery.

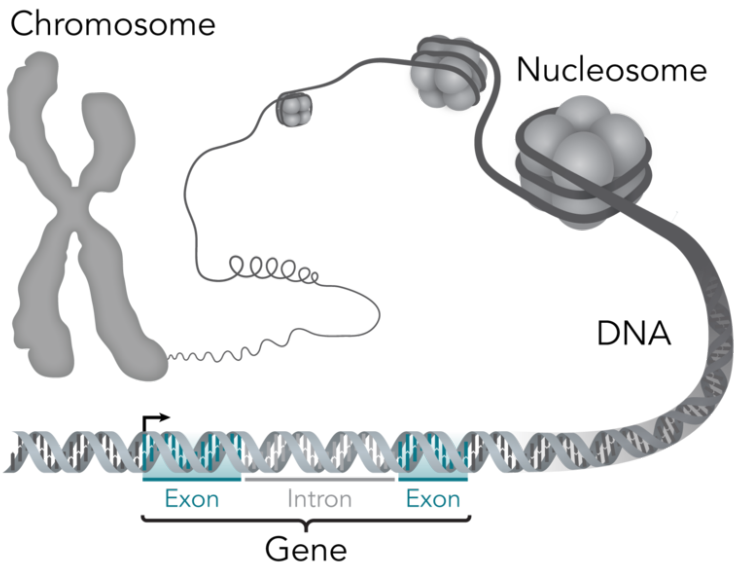


Figure 1: DNA molecule, from the chromosome to the genes (Create Commons®)

Genes are small segments of the DNA molecule, composed of exons (the coding and expressed regions), and introns (the non-coding regions). The DNA sequence is compacted into the 46 chromosomes of the human genome.

The genetic differences between two individuals are explained by less than 2% of variations in their DNA sequences. These genetic variations can be population-specific or person-specific. A variation is considered to be a polymorphism if its allelic frequency is more than 1% within a population, and rare otherwise. They can have different sizes, such as a single base, but also a part of chromosome to an entire one.

Single Nucleotide Variants (SNVs) are different combinations of single base substitutions, deletions or insertions. “Indels” represent the insertion or deletion of one or of a small stretch of nucleotides (Figure 2.A). If a SNV impacts the coding region of a gene, it can modify a codon and therefore the encoded amino acid residue, and the variant is considered to be non-synonymous. However, due to the degeneration of the

genetic code (different codons encode the same amino acid), the nucleotide variation may give rise to a triplet encoding the same amino acid as the original one. This type of variant is called synonymous. Larger regions of the genome (generally larger than 1000 bases) that vary between individuals, leading to an insertion, duplication, inversion or deletion of thousands of nucleotides, are called structural variants (SVs) [1]. Copy Number Variations (CNVs) are subtypes of SVs resulting from gain or loss of a copy of an entire DNA region by deletion or duplication (Figure 2 B) [2].

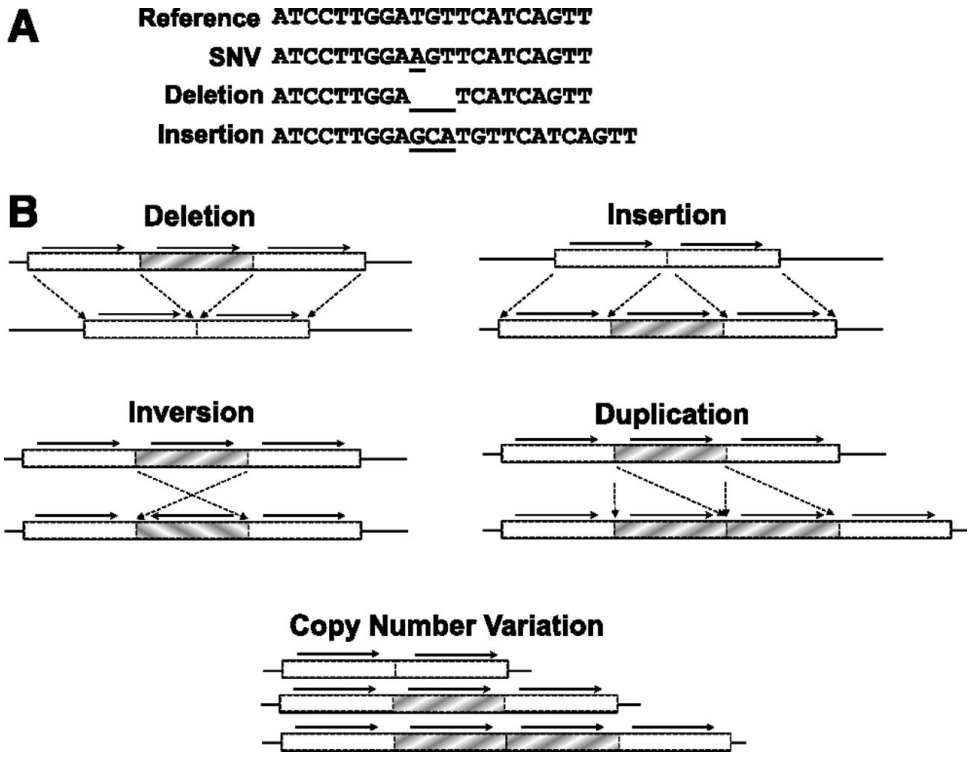


Figure 2: Types of genetic alterations in the genome, (A) small modifications (B) larger rearrangements [3]

Both SNVs and SVs are frequently present in the genome and they constitute the basis of evolution and adaptation. However, some of these variations may have a negative effect on the protein level, its activity, its function and its specificity, which may eventually lead to a disease.

Human genetic diseases

The spectrum of human genetic diseases is very large. It refers to abnormalities in the genome resulting in the complete gain, loss or disruption of the integrity of one or more genes. These DNA alterations can either be inherited from the parents (germline mutations) or they can be acquired throughout life (somatic mutations) because of aging and/or environmental conditions (Figure 3). Pathologies resulting from germline mutations, which can be transmitted to the following generations, are commonly referred to as hereditary diseases, while somatic DNA insults are usually not transmittable to the offspring and usually lead to malignancies.

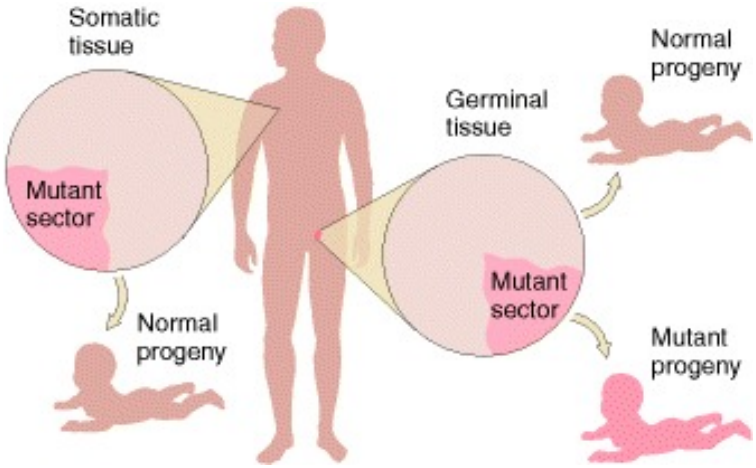


Figure 3: Somatic versus germline genetic variations [4]

SNVs or SVs can potentially modify the normal production of proteins; by directly affecting the genes encoding them, through deleterious non-synonymous variants, splice site variations or large rearrangements impacting one or more exons. Similarly, they can influence the normal function of the genes also by affecting non-genic regulatory regions.

The identification of the genetic basis of inherited diseases and cancer is essential for the understanding of the pathophysiological mechanisms underlying them.

Furthermore, it potentially provides hints of prognosis and eventual treatment as well as advice for the family members.

1.1.1 Introduction to ocular melanomas

Melanoma is a type of cancer originating from melanocytes, the cells producing melanin. The most common form of melanoma is cutaneous melanoma, originating from the melanocytes of the epidermis. The tumors often have similar appearances to moles, some may also develop from moles. Intense exposure to Ultra-Violet (UV) light is a key element for the development of this cancer. Cutaneous melanoma has a prevalence of more than 21 new cases per year per 100'000 individuals [5], counting for less than 5% of skin cancers but responsible for about 75% of all death from skin cancers [6]. Cutaneous melanoma is well curable with early recognition and treatment, but it is more challenging to treat it in advance stages, because of dissemination in other parts of the body.

Melanomas occurring in the different layers of the eye are referred to as ocular melanomas.

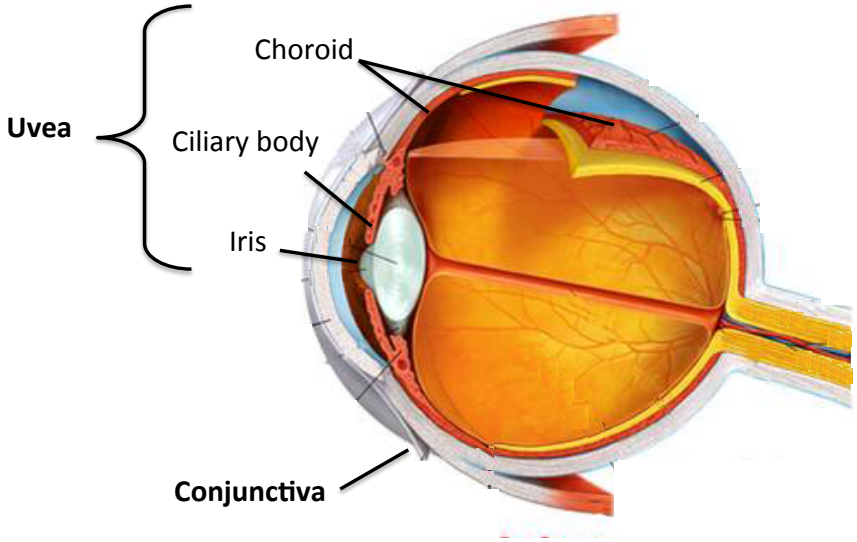


Figure 4: Structure of the eye, the conjunctiva and the uvea are highlighted

Conjunctival Melanoma

Conjunctival melanoma (CM) represents 5% of all ocular melanomas and it affects about 0.2 individuals per million worldwide. The primary site of the tumor is the conjunctiva, a thin membrane that covers part of the sclera, which is directly exposed to light, as well as the inner surface of the eyelids (Figure 4). It is a highly recurrent tumor leading to metastatic disease and death within 10 years in 25–30% of patients [7]. The most frequent sites of metastasis are the lungs, followed by the liver, the brain and the bones. Interestingly, CM incidence and morbidity have increased over the past few years, in particular in regions where sunlight is more abundant [8]. CM usually initiates as a pigmented nodular lesion in the conjunctiva during the sixth or seventh decade of life. Recent molecular investigations have indicated that conjunctival and cutaneous melanomas may share common features [9]. For instance, epidemiological studies have shown an association between decreasing latitude and increasing incidence of conjunctival melanoma [8], suggesting that exposure to sunlight has a role in its etiology [9]. However, strong molecular proof of such correlation has not been established yet. Currently, the treatment includes the surgical removal of the pigmented node and the surrounding area followed by cryotherapy, a cold-induced procedure decreasing cell growth and reproduction [10].

Uveal Melanoma

Despite having a very rare incidence of 5-8 new cases per million per year [11, 12], uveal melanoma (UM) is the most common primary intraocular tumor of the adult. It develops from melanocytes in the choroid, the ciliary body, or the iris (collectively called the "uvea", one of the inner layers of the eye) (Figure 4) and usually metastasizes through the

blood stream to the liver [13, 14]. Symptoms include variable and painless visual disturbances, often presenting when the tumor has already reached a considerable mass. Survival and potential therapeutic options depend, among other things, on the presence of specific genetic alterations [15]. While population studies suggest ethnic predisposition [16, 17], environmental factors that are directly involved in the transformation process have not been clearly delineated. For instance, a possible association with UV light exposure has been suggested [18-21], but it has recently been questioned based on molecular data [22].

UM shows two distinct classes of progression: the first class is characterized by a high survival rate (95% after 8 years) and without metastases, while the second class shows a lower survival rate (31% after 8 years) and patients usually present metastases, mainly in liver, with high resistance to chemotherapy [23]. Despite recent breakthroughs in understanding the molecular biology of this malignancy, e.g. mutations in driver genes (*BAP1*, *SF3B1*, *EIF1AX* [24]) and specific CNVs (chr3 monosomy [25]), UM prognosis has not increased over the last 40 years. After hepatic metastatic dissemination, the median overall survival is less than one year, and the standard treatments do not improve this outcome. Thus, it is necessary to develop new therapies based on our understanding of the genetic alterations of this melanoma.

1.1.2 Introduction to Mendelian diseases

Mendelian Diseases

Based on Mendel's rules, a recessive trait is the consequence of the action of two alleles on the two chromosomal copies of a gene, while a dominant trait only requires an allele on a single chromosome. In X-linked trait, both recessive and dominant rules may occur. In Mendelian diseases, pathogenicity is driven by mutations in a single gene. A

single heterozygous mutation can lead to a dominant disease due to haploinsufficiency or gain-of-function. Recessive disease requires disruption of both copies of a gene, either via the same mutation (homozygous variant), or two heterozygous mutations, one on each copy (compound heterozygous variants) (Figure 5).

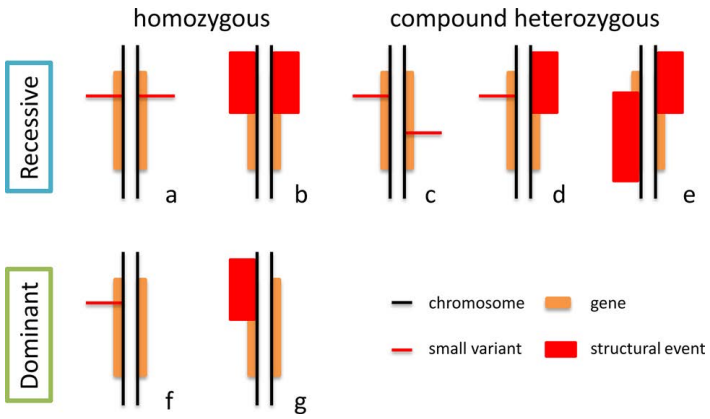


Figure 5: Possible configurations of pathogenic mutations for autosomal recessive and autosomal dominant inheritance [26].

Mutations can either be inherited from one parent or newly acquired in the patient. This latter is called a *de novo* mutation. Every individual usually carries 45 to 60 *do novo* mutations [27] [28]. It is important to note that, regardless of the type of mutation within single pedigrees, pathogenic Mendelian variants always co-segregate with the disease in affected individuals. Therefore, all patients within a family should necessarily share the same mutation(s) but not necessarily the same innocuous DNA variants.

Prevalence of Mendelian disorders

Mendelian diseases generally tend to have relatively low prevalence in the population; nevertheless they collectively affect millions of individuals [29]. Within this category, rare genetic disorders are estimated to range between 1'000 to 30'000 in

humans, and only half of them have been elucidated at the genetic level [30]. This is particularly true for rare Mendelian disorders leading to retinal degenerations such as retinitis pigmentosa, cone-rod dystrophy, Leber congenital amaurosis and other subtypes. These diseases are genetically heterogeneous, meaning that the disruption of any of 250 genes known so far [31] gives rise to a similar clinical phenotype.

Detection of disease-causing variants

Following the concept that rare diseases are caused by rare mutations, we focused on DNA variations that have an impact on protein function and occur with a very low frequency in the general healthy population. Indeed, since heterozygous recessive mutations are not disease causing, they can be present in the general population [32, 33], even at non-negligible frequencies. This simple genetic concept has tremendous consequences in NGS-based searching for mutations, as only about a dozen genes will harbor rare, non-synonymous variants in homozygous or compound heterozygous state genome-wide [26]. For this reason, recessive diseases are more easily diagnosed with NGS techniques but it is more complicated for dominant diseases. Different strategies have been developed to efficiently identify the variants of interest. Some of these strategies are described further in the bioinformatics section of this work.

1.2 Next-generation sequencing

1.2.1 Current usage of NGS

Next-generation sequencing (NGS), also called high-throughput sequencing, consists of modern technological methods used to query megabases of DNA and RNA at

once, through computer-based alignment of millions of short reads. Compared to the Sanger sequencing [34], one of the first sequencing techniques, NGS allows faster and cheaper sequencing results.

Since its beginning in 2008, NGS is widely used, both in diagnostic and research laboratories. This is especially true for the sequencing of all the coding regions of the genome, by means of Whole Exome Sequencing (WES); and the sequencing of the full genome, by means of Whole Genome Sequencing (WGS). The main reasons for this are the speed of processing, and the lowering price of NGS, which is now affordable for most laboratories and clinical diagnostic, with a exponential drop in the cost of human genome sequencing, starting from billions of dollars to about $ 1,000 per sample currently [35]). However, due to the generation of enormous amounts of raw data (from dozens to hundreds of Gigabytes), the analysis of any NGS data nowadays requires the usage of an adapted computational framework to generate readable information that is relevant for the phenotype studied.

Many different NGS technologies have been developed over the past decade, such as the ones from Illumina, Life Sciences, Oxford Nanopore, Pacific Biosciences or Complete Genomics. They allow us to sequence the exome, the genome, the transcriptome and the epigenome of any individuals and from many different tissues. All these sequencing platforms have their own advantages and disadvantages [36] [37] [38] based on different parameters, e.g. throughput (time of sequencing of a given number of samples), read length, accuracy, robustness, price and applications available after the sequencing, etc. [39].

For the most recent sequencing platforms, the time of sequencing is significantly reduced. This characteristic is of high importance for treatment of patients, especially in the field of neonatal care or aggressive cancers [40-42].

Short-reads sequencing is a useful technology, which has proven its performances over the past decade. However, with the recent development of long-reads sequencing platforms (e.g. from Pacific Biosciences), repetitive and complex regions of the genome can now be accessible, and essential for detecting more precisely indels of more than 50 base pairs, structural variants and genome phasing [39].

Single cell sequencing aims at detecting the cellular heterogeneity present in tissues, and has become particularly relevant for the screening, diagnostics and treatment of heterogeneous cancer types. It is also widely used in non invasive prenatal diagnostics [43].

1.2.2 NGS technology used during the thesis

Main workflow of NGS

DNA is extracted from a given tissue (e.g. blood) or biopsy after the signature of an informed consent by the patients and their geneticist, in accordance with regulations for studies on human subjects. The genome is then randomly fragmented into small pieces, followed by ligation of adapter sequences. After an amplification procedure, the short DNA fragments are then purified and sequenced into short reads (Figure 6, left panel).

Compared to single read sequencing, where only one end of the DNA fragment is sequenced, paired-end sequencing enables both ends of the DNA fragment to be sequenced. The reverse and forward reads are constructed with a known interval, which allows generating high-quality and well alignable sequence data. Moreover, paired-end sequencing facilitates the detection of genomic rearrangements, repetitive sequences, as well as novel transcripts.

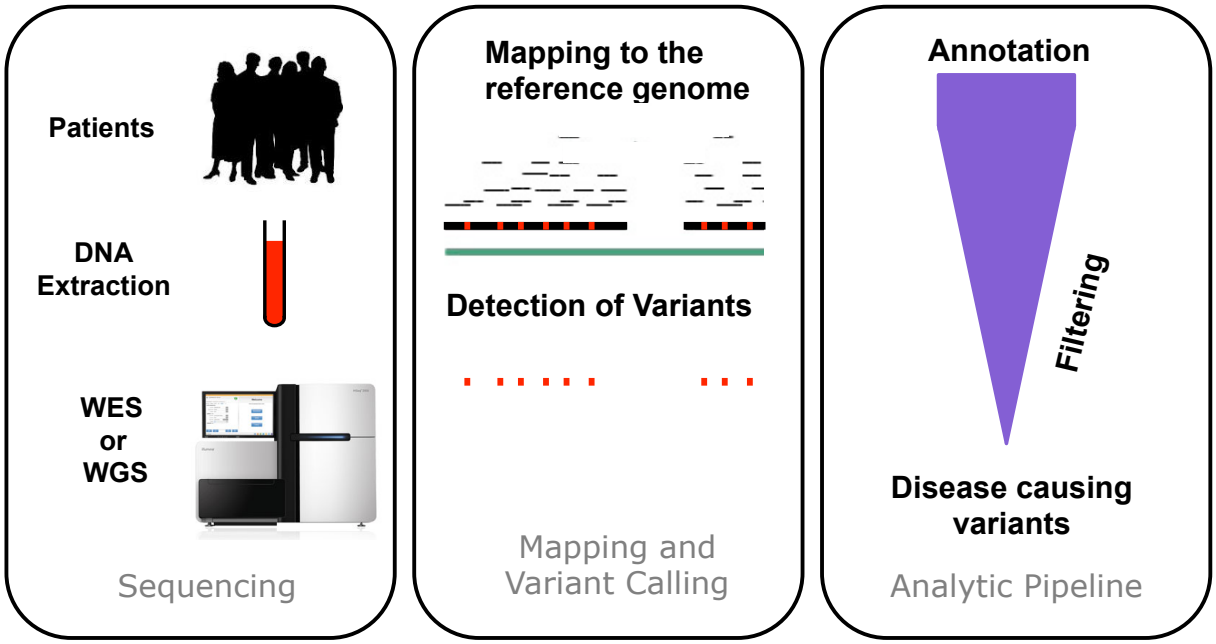


Figure 6: NGS methodology, starting with the sequencing of the DNA and followed by the generation all the variants present in a patient, and the detection of the disease-causing variants.

The coverage of sequencing refers to the average number of reads that align to known reference bases. Based on the different coverage level, we can determine the degree of confidence of a reference nucleotide or of a variation in the DNA sequence of a patient. The higher the coverage, the more confidence we have in the base quality. However, the price of the sequencing also increases with the depth of coverage.

WES versus WGS

All samples processed by WES and mentioned in this work were sequenced on an Illumina HiSeq 2000 machine, generating 100 bp paired-end reads with coverage of 50x to 99x, depending on the project. All the WGS samples were sequenced by the Complete Genomics technology [44], with paired-end reads and sequencing depth between >50x to >80x of coverage. The technology used by Complete Genomics is slightly different from Illumina sequencing. It uses high-density DNA nanoarrays that are populated with DNA nanoballs [44].

By definition, WES focuses exclusively on the expressed regions of the genome. In order to sequence only these exonic regions, an extra capturing step is needed in the

NGS procedure, using specific libraries of probes, which can selectively hybridize to the genomic regions of interest. Only these DNA fragments will then be amplified and sequenced. On the contrary, WGS queries every single base of the genome, with a better sensitivity and uniform coverage compared to WES results, especially in coding regions [26]. Indeed, the absence of the capturing step, characteristic of WES, permits a less artificial representation of the genome, which can be used to assess with better quality intronic and intergenic variants, but also large insertions and deletions which might be in heterozygous state (Figure 7) [26]. However, for highly heterogeneous and contaminated tumors, the usage of WES at very deep coverage (up to 10'000 times) has proven to be more efficient in detecting SNVs than WGS [45].

Feature	WGS	WES
Exonic variants	Yes	Yes
Intronic variants	Yes	No
Intergenic variants	Yes	No
Indels	Yes	Yes
Copy number variations	Yes	Not directly/imprecise
Large insertions and deletions	Yes	Not directly/imprecise
Transposable elements	Yes	Not directly/imprecise

Detection of copy number variations, large insertion and deletions, as well as of transposable elements are imprecise in WES since data are available for coding regions only, and these events can originate elsewhere

Figure 7: Features of WGS versus WES [26]

In Mendelian diseases, most of the disease-causing variants are located within the coding region of genes [32, 46, 47]. Therefore, WES, targeting only the coding part of our genome is a valid investigation tool, despite that the exome represents only 2% of the entire genome and does not provide accurate information for "atypical" mutations (e.g. structural variations, copy number variations). Conversely, WGS is more expensive and

requires more complex analysis, but it virtually provides information on all possible types of pathogenic genetic events, such as gene fusions, duplications, and deletions, and non-coding regions. In standard protocols, WES is usually the first sequencing method performed on a patient, and if negative, WGS can then be attempted.

1.3 Development of bioinformatics pipelines for genetic diseases

Mapping and variant calling step

The process of obtaining the full exome or genome sequence of an individual consists of a two-step computer-based procedure. First, the short NGS reads are mapped to the reference genome by assigning them specific genomic coordinates, corresponding to the amplified regions of the libraries used (Figure 6, middle panel, top). For this step, we used the Novoalign software (Novocraft, 2010), one of the most popular reads aligner with BWA [48], and Bowtie-2 [49]. While the two last aligners are both based on backward search with Burrows-Wheeler Transform, Novoalign uses a hash-based index to search the genome and gives very good performance results [50]. After a step removing the low-quality sequences data, mismatches between the reference genome and the individual genome are assessed via a bioinformatic process referred to as "variant calling" (Figure 6, middle panel, bottom). All the steps follow the guidelines provided by the Genome Analysis ToolKit (GATK) best practices [51], a bioinformatics suite developed by the Broad Institute, which offers a wide variety of tools with a primary focus on the discovery of variants and genotyping. The same protocol was applied to all our WES generated data, having the advantage of achieving data homogeneity, with the same technical errors. This simple practice enables the reduction of the sequencing "noise" (e.g. low quality calls, false positive variants). For WGS,

Complete Genomics directly performed these bioinformatics steps and provided us the raw files and the processed data for all the SNVs, SVs and CNVs.

Annotation and filtering steps

After the generation of all the variants present in an individual, two analytic procedures still remain and are highly important to identify the disease-causing variants (Figure 6, right panel). The first step is the annotation of all the genes and variants, which is then used for the second step that is the filtering of the variants based on specific criteria, which are relevant for the disease of interest.

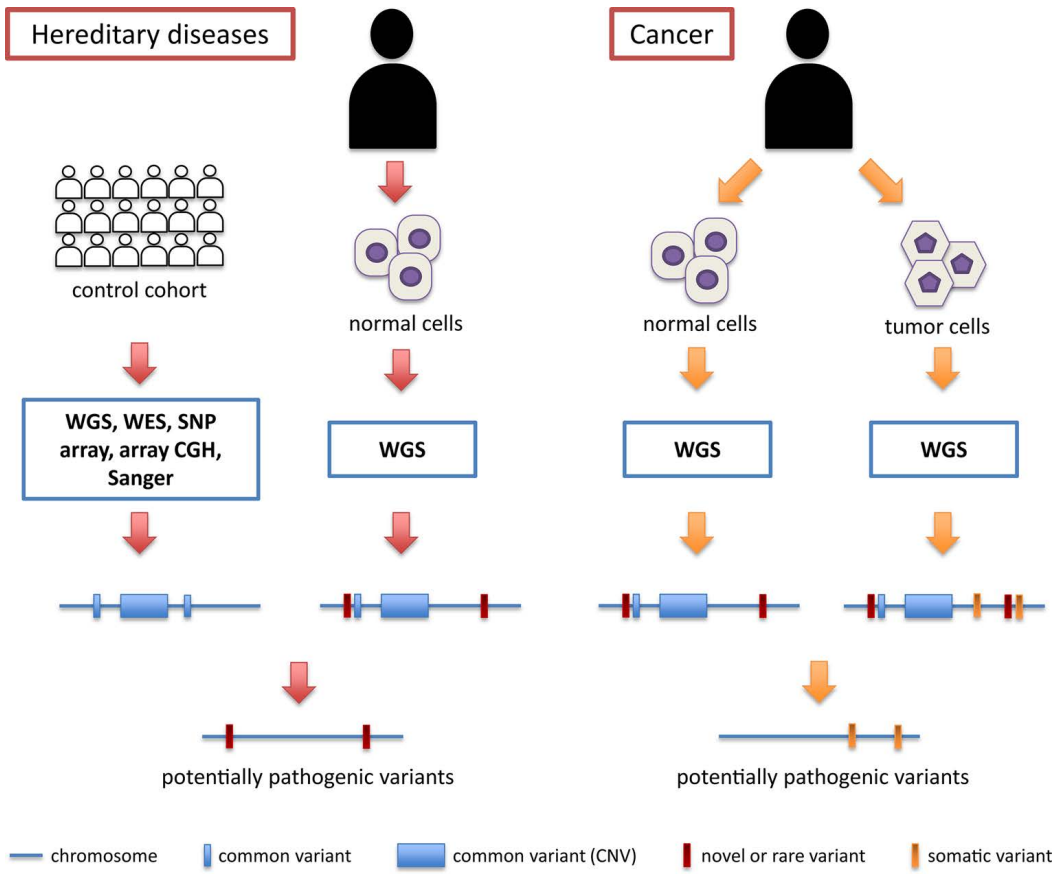


Figure 8: NGS methodology for hereditary diseases and cancer, extracted from [26]

More specifically, diverse bioinformatics analytical procedures or "pipelines" have been developed for application either in the context of cancer or Mendelian cases (Figure 8).

1.3.1 Bioinformatics for Cancer analysis

In cancer, especially for tumors caused by mutations acquired during an individual's lifetime, it is highly important to detect all the somatic mutations that are specific of the tumor. In term of experimental set-up approach, both tumor DNA and normal DNA of the same patient are sequenced. The cancer genome is extracted from tissue biopsy of the tumor, while its normal counterpart usually comes from blood or biopsy of a normal tissue. The simple comparison between tumor and normal genomes allows the detection of somatic mutations present only in the tumor (Figure 8, right part). In this situation, any somatic change is potentially a good candidate; no matter its frequency and type of mutation. However, a strong and deleterious mutation will be more interesting to study, as well as if the gene impacted has already been associated with a cancer or if it is involved in a relevant pathway for cancer progression or treatment. Chapter 5.1 provides a more detailed procedure of the analytic workflow for cancer. In case of heterogeneous tumors, one important criterion is the correct choice of the depth of sequencing, since high sequencing coverage may be able to detect variants present in very low percentage in the DNA. Single-cell sequencing is also a valuable tool to properly assess the cellular heterogeneity of the tumor sample.

1.3.2 Bioinformatics in Mendelian diseases

Filtering procedures

In Mendelian diseases, only one or two mutations in a single gene are responsible for the pathological phenotype. However, all individuals carry more or less frequent variations, which make them unique but are usually not disease-causing. The most critical part of the bioinformatics procedures applied to Mendelian disorders is: being

able to identify disease-causing candidate DNA variants while reducing the overall noise, both technical (false calls) and biological (innocuous variants). In order to efficiently identify these candidates, different steps are have been applied (Figure 9).

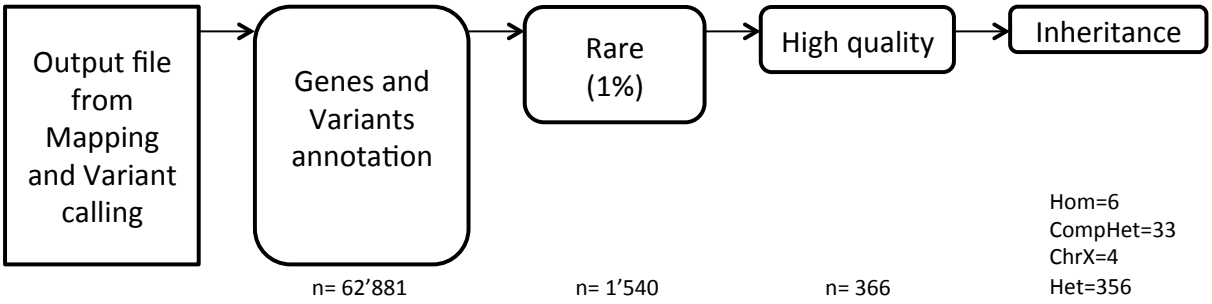


Figure 9: Annotation and filtering steps for rare Mendelian disorders

(with the average number of variants for 110 WES samples)

The input file corresponds to all the variants detected in a patient's WES or WGS data. It contains both rare and frequent variants; but also variants of different sequencing quality and different zygosity.

Genes annotation allows describing the impacted gene, if this gene is often mutated in control individuals (and so less likely to be disease-causing) (Table 1, A-B) and if it is a known disease gene (Table 1, C).

Annotation type	Description	Value	Data
A- "Polymorphic" gene	Number of mutation per gene in control population	Low, medium-low, medium, medium-high, high	Calculated from 1KG by Complete Genomics (CG)
	Number of nonsense and frameshift homozygous changes (in more that 2 individuals) per gene in a control population	NA, 0, 1, …	ExAC.r0.03 [52]
B- "False Positive" gene	Annotation of the genes which are often mutated in WES data	NA or FPGenes	[53]
C- Link with the phenotype	List of disease genes	Mendelian disorders, Retinal Degeneration, Skeletal, Metabolic, Mental Retardation, …	OMIM [54], RetNet [31], CeGaT,

Table 1: Description of the different annotations for the genes

Variants annotation: In this step, we add information on the variant's frequency from public databases of healthy controls (ExAC [52]), or of in-house control genetic databases (Table 2, D). The effect on the amino acid residue is also considered (Table 2, A), as well as its degree of deleteriousness (Table 2, B), and the conservation of the nucleotide impacted (Table 2, C). The information about known variants of clinical significance at the same position or in the surrounding 25 nucleotides is described (Table 2, E). Splicing events are calculated for every variant (Table 2, F) as well as their distance from the closest exon boundaries (Table 2, G).

Annotation type	Description	Value	Data
A- Protein impact	Description of the impact of the variant on the protein sequence	Nonsense, Frameshift, Splicing Site, Missense, Synonymous, intronic	annovar [55]
B- Amino Acid impact	Score reflecting the impact of the change based on known pathogenic mutations	Between 0 (poor impact) to 1 (high impact)	In house data
	VEST3 predictive score for the deleteriousness of the variant	Between 0 (poor impact) to 1 (high impact)	[56, 57]
C- Conservation of the nucleotide	PhyloP score between 46 mammals	Conserved if > 1.5	PhyloP [58]
D- Frequency of the variant	Frequency of the variant in public database or its absence	NA or frequency	Wellderly, from CG, 1KG from CG, ExAC [52]
	Frequency of the variant in in-house database (WES/WGS)	NA or number of individuals with the same variant	In house data
E- Clinical Significance	Presence of a variant with reported clinical significance	Pathogenic, likely pathogenic, likely benign, benign or NA	ClinVar [59]
	Number of pathogenic variants in 25 nucleotides around the variant	Number between 0 to 50	
F- Splice Site prediction	Prediction of creation or disruption of a splice site	Annotation if it creates or disrupts a splice site	MaxEntScan [60]
G- Exon boundaries	Calculation of the closest exon boundary from the variant	Number of nucleotides from the exon boundary	RefGene

Table 2: Description of the different annotations for the variants

After this intensive annotation step, it is now possible to filter the candidate variants based on the most relevant and biological information we can gather from the phenotype.

Rare (1%): By definition, we are focusing on rare genetic alterations, which could cause a rare disease. We usually set up the threshold of rarity of at most 1% of the allelic frequency, which takes into account healthy carrier of single heterozygous pathogenic mutations. We also filter the variants present in our in-house databases of hundreds of WES data of affected and unaffected individuals, for various diseases. Moreover, we remove all the rare variants that can be found in homozygous state in healthy carriers.

High quality: We only consider variants with a high sequencing quality, based on a minimal number of 15 reads, a percentage of alternative reads of at least 15% compared to the reference reads and a genotype quality of 50 or more (out of 99, being the best quality). Below these values, it is difficult to assess the real presence of the variant and its quality. During this step, most of the non-exonic variants are discarded, as they are outside of the range of WES. However, intronic variants of good quality in the surrounding exon boundaries may be kept, especially if they create new splicing events.

Inheritance pattern: Depending on the transmission mode of the disease, the variants are separated in 4 files: homozygous variants for autosomal recessive diseases as well as compound heterozygous variants, heterozygous variants for autosomal dominant diseases and variants on the X-chromosome for X-linked diseases.

Strategies to identify disease genes

Each disease and each patient is different. However, common strategies have been developed based on i) the inheritance pattern of the disease, ii) a clear clinical description of the phenotype, iii) the availability of family members and information, iv) the availability of genetic information from other patients with the same disorder. In the end, depending on the context, up to 6 different strategies can be applied (Figure 10) [61].

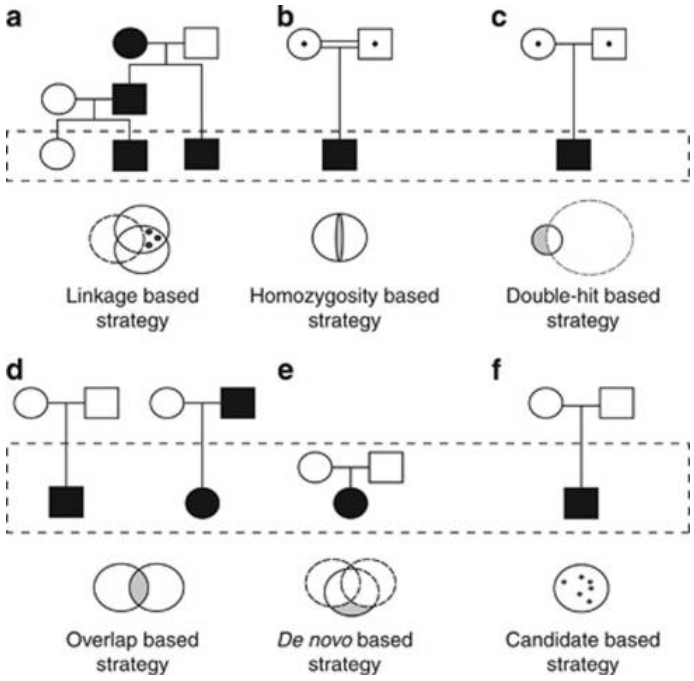


Figure 10: Strategies to identify disease-causing genes in NGS studies, extracted from [61]. Affected individuals have filled symbols; carriers are represented with a dot. Dashed rectangles contain the individuals that were sequenced. Circles below each pedigree represent the strategy used to identify the disease genes.

Linkage-based strategy, which compares the rare variants found in all the affected individuals versus healthy members of the same family.

Homozygous-base strategy, which focuses on homozygous variants inherited from both parents, in case of rare inherited recessive disease and suspected consanguinity.

Double-hit-based strategy, which analyses rare homozygous or compound heterozygous variants present in a patient without other family member available and when the disorder is suspected to be inherited in a recessive manner.

Overlap-based strategy, which searches for a single gene with different rare variants in multiple patients.

De novo-based strategy, which analyses both the patient and their parents in order to find a variant that has newly arisen in the patient only. It is also called a Trio analysis.

Candidate-based strategy, which corresponds to the strategy for single affected patients with a suspected dominant inherited disorder. In this case, a prioritization of the mutations is made based on different features, such as the impact of the variant on the protein function, the conservation of the modified amino acid and other.

During my Ph.D. work, all of these strategies have been applied on different patients. The overlap-based strategy was used in a large cohort of patients with recessive form of Cone-Rod Dystrophy and results are presented in Appendix 6.4 However, in the majority of cases, the inheritance pattern was unclear and additional patients or family members were not sequenced, for cost effective reasons. Thus, the candidate-base strategy was intensively used. In order to confirm the pathogenicity of a variant, co-segregation of a variant within the affected family members (when available) was checked. Note that Chapter 5.1 provides a more detailed procedure of the analytic workflow for Mendelian disorders.

1.4 Aim of this thesis

The aim of this Ph.D. work was to develop novel bioinformatics techniques specific to large scale sequencing sets in order to efficiently identify the genetic insults with high phenotypic effects, such as germline mutations in Mendelian hereditary disorders and somatic mutations in rare ocular cancers.

2 Results

This chapter gives a summary of the main results of the projects in which I took part during my doctoral work. Moreover, it contains the experimental follow-up carried out in the lab. A more detailed explanation of my personal contribution for each study is described at the beginning of each subchapter.

2.1 Overview of NGS analysis of human diseases

In the first part of my thesis, the applications of NGS technologies are reviewed to better understand the genetic alterations leading to human diseases. More specifically, we discuss the different existing methods for Mendelian disorders gene discovery on one side, and for the identification of causative somatic mutations in cancer on the other side. An emphasis is given on the benefits and disadvantages of using WGS compared to WES. Moreover, we detail the general pipelines used to study both Mendelian diseases and cancer. In this work, I reviewed the literature and wrote the manuscript. (See chapter 5.1, [26])

2.2 Ocular Melanomas

In this chapter, a focus is provided on the study of two different types of ocular melanomas: the conjunctival melanoma (CM) and the uveal melanoma (UM) by means of WGS. Our findings on CM highlight the presence of a high number of somatic mutations (more than 90,000 single nucleotide substitutions) and a majority of cytosine to thymine (C>T) substitutions, specifically found at the 3' end of pyrimidine dinucleotides. They are classical features of the presence of typical UV light mutation signature, a characteristic of cutaneous melanomas. In this project, I did all the bioinformatics

analysis and statistics, and I participated in the writing of the manuscript. (See chapter 5.2.1, [62])

Uveal Melanoma is the most common form of ocular melanoma and, unlike CM, it is genetically different from cutaneous melanoma. Whole-genome sequencing of 34 primary and metastatic UM patients allowed us to further characterize the genetic landscape proper to this type of cancer. In particular, we were able to classify the patients according to particular genetic events, such as Copy-Number Variations and mutations in cancer driver genes. We identified and designated four distinct classes of UM, with large genomic alterations and mutations in *BAP1* associated to classes A and B, usually linked with a bad prognosis and metastatic capabilities. On the other hand, C and D classes harbor minor alterations and single mutations in *SF3B1* or *EIF1AX*. Interestingly, when looking at the CNV levels of the long arm of chromosome 8, we were able to correlate high amplification level (ploidy of 5 or more) with metastatic UM (p-value=$8.6x10^{-4}$), while single-copy gain (ploidy of 3) was more likely to be associated with class D and *EIF1AX* mutations (p-value=$4.7x10^{-3}$), previously characterized in the literature as patients with late metastatic potential. For this project, I took care of all the bioinformatics set-up, analysis and creation of figures and tables and wrote the manuscript. (See chapter 5.2.2)

2.3 Rare inherited Mendelian disorders

The bioinformatic procedures developed in the course of my Ph.D. were applied to many types of rare inherited Mendelian disorders, ranging from retinal degenerations to skeletal and metabolic disorders. In this chapter, I detailed my major novel findings for a specific Mendelian disease. The appendix (Chapter 6) contained other projects for which the bioinformatics procedures were applied and allowed us to make new discoveries.

We were able to detect the mitochondrial chaperone *HSPA9* as the cause of a novel disorder, called EVEN-PLUS syndrome [MIM #616854] (Epiphyseal, Vertebral, Ear, Nose, PLUS associated findings). Interestingly, HSPA9 protein interacts with LONP1, another mitochondrial protein leading to a similar syndrome (CODAS [MIM #600373]) [63, 64]. This interaction delineates a family of "mitochondrial chaperonopathies", bringing first evidence of an unexplored role of mitochondrial chaperones in human embryonic morphogenesis. For this study, I analyzed the WES results and found the disease-causing gene in the three patients. I worked on the figures of the paper. (See chapter 5.3.1, [65])

Nine patients with infantile-onset severe developmental delay and skeletal dysplasia were sequenced and we identified biallelic mutations in *NANS*, the gene encoding the synthase for N-acetylneuraminic acid (sialic acid). Further functional experiments performed with collaborators showed that NANS-mediated synthesis of sialic acid is required for early brain development and skeletal growth. In this project, I analyzed the WES results of the patients, and generated the figures related to these *in-silico* findings. (See Appendix 6.1, [66])

In a family presenting with severe peripheral osteolysis, (a disease showing adult-onset progressive shortening of fingers and toes with redundancy of the overlying skin), we found, in all affected members, compound heterozygous mutations in *ASAH1*, the gene coding for acid ceramidase. This allowed us to confirm the diagnosis of Farber disease [MIM #228000] in what appear to be the oldest affected individuals known so far for this disease. Here, I generated the WES results and analyzed the data, which allowed us to identify the disease-causing gene. I produced the figures related to these *in-silico* findings. (See Appendix 6.2, [67])

We reported two unrelated patients with a very similar multisystemic disease involving the liver, vision, immune system, connective tissues, and bones, caused by recessive variants in the neuroblastoma amplified sequence (*NBAS*) gene. Although this gene has already been associated with human diseases of acute liver failure [MIM #616483] [68] and short stature and optic atrophy [MIM #614800] [69], we observed a wider phenotypic spectrum of NBAS deficiency than previously describes, including skeletal, hepatic, metabolic, and immunologic aspects. In this project, I generated the WES results and contributed to the identification of the *NBAS* gene. I made the figures related to these *in-silico* findings. (See Appendix 6.3, [70])

With the help of the bioinformatics pipeline, we discovered three individuals with cone-rod degeneration, a specific form of retinal degeneration affecting central vision, with bi-allelic variants in the centrosomal gene *CEP78*. The patients also developed hearing loss or hearing deficit with age. Our data revealed a previously undescribed clinical syndrome of a ciliary nature characterized by blindness and deafness but clearly distinct from Usher syndrome [MIM #276900], a condition for which visual impairment is due to retinitis pigmentosa, another type of retinal degeneration. For this project, I applied the WES pipeline I developed and helped in merging the WES results for the different patients of the cohort, which eventually led to the finding of this new gene in two different families. (See Appendix 6.4, [71])

TTLL5, a gene recently associated with Cone-Rod dystrophy [72], has been further characterized in our study of 8 patients suffering from a recessive form of cone-dominated retinopathy. We were able to link loss-of-function mutations in male with reduced sperm motility and infertility, delineating a novel, allele-specific syndrome causing defects in two unrelated functions, reproduction and vision. For this project, I contributed to the development of the bioinformatics pipeline, which allowed the identification of the disease-causing gene. (See Appendix 6.5, [73])

3 Discussion

Optimization of bioinformatics procedures

In this work, NGS technology has been used as the basis to decipher the genetic causes of various human diseases. Several bioinformatical procedures and filtering steps were developed and adapted to analyze different data sets. These procedures allowed us to identify the disease causing variants from thousand of benign genetic alterations and have been, of course, gradually optimized over the years. To date, five major versions of these steps have been developed and used, with a specific focus given on the genes, the variants and the different strategies to efficiently analyze the data. One of the most important features of these procedures has been the integration of annotated variants from thousand of individuals, such as the ones belonging to ExAC [52] or from our in-house database. Access to the frequency of variants has greatly modified our decision on selecting rare changes from frequent ones. Interestingly, some variants, reported to be disease causing in studies before the NGS revolution, have proven to be wrongly annotated [52]. These improvements of the different pipelines allowed me to provide more comprehensive lists of candidates' variants, including novel disease genes and novel mutations presented in this thesis. Moreover, five additional new candidate disease genes responsible for dominant and recessive skeletal disorders have been discovered and are currently being tested and/or are in the process of being submitted for publication. The pipelines described in this work will continuously be optimized to efficiently integrate all new information from NGS processes, which are being improved every day. Originally made for retinal dystrophies investigation, the pipelines were then adjusted for the diagnostic of skeletal and metabolic disorders. They are currently being applied for the analysis of intellectual disabilities and epilepsies, with promising results.

Differential approaches to cancer and Mendelian disorders

Experimental strategies and NGS-based approaches applied on ocular melanomas are very different that those used for rare Mendelian disorders. First of all, detection of cancer variants requires the acquisition of both normal and tumor DNAs, but this latter could be difficult to extract. Indeed, for ocular melanomas, only patients with eye enucleation and without any chemotherapy were considered. Since they are very rare cancers, only few samples were available for the study. Moreover, this double sequencing considerably increases the price of any study. Secondly, sequencing data are treated differently for these two types of diseases. While it is essential for Mendelian disorders to have a large list of variants present in healthy individuals, all somatic variants are potentially interesting in cancer and do not require a list of variations in the general population. However, the process of obtaining high quality somatic variants might be challenging, especially in highly heterogeneous cancers. This is an advantage of ocular melanomas, as they are usually very homogeneous. One important point in some types of cancer is the presence of predisposition factors, which are present in the germline genome (i.e. *BRCA1* in breast and ovarian cancers). This means that cancers with such hereditable conditions should also be studied similarly to Mendelian disorders, otherwise important information may be missed. Finally, the pathogenicity of a variant in a gene for a given disease can be assessed by various experiments, the functional testing being the most informative one. However, a good hint of probable pathogenicity is given in Mendelian diseases when the candidate variant segregates perfectly within the family. Such confirmation could not be undertaken for cancer.

Disease gene identification by NGS: positive effects and limits

NGS technologies have revolutionized the field of medical genetics. They allow us to find disease-causing variants in a much quicker, cheaper, and more efficient way than ever. Every month, new genes are being identified as the genetic causes of existing or novel syndromes, or new therapeutic approaches in cancer. NGS is accelerating at an unprecedented scale the understanding of genes, pathways and family of diseases. In Mendelian diseases, it is very efficient for recessive families, as rare bi-allelic mutations are in the end only about a dozen in the human genome. It is even more powerful in consanguineous families, as regions of interest can be inferred easily from many NGS methods. However, it is more difficult to solve dominant cases because patients harbor much more rare variants in heterozygous state and many of them are harmless but difficult to discard.

Likelihood of a gene to be associated with a dominant condition

In order to reduce the number of candidate heterozygous variants to consider in patients with a dominant disease, we are developing a predicting tool estimating the likelihood of a gene to be associated with a dominant disease. Our method is based on the hypothesis that genes responsible for rare autosomal inherited diseases should have specific characteristics such as, i) low number of rare variants in the general healthy population ii) strong selection and therefore conservation across species, iii) function and expression pattern leading to possible haploinsufficiency or gain-of-function mutations. We selected many gene features and characteristics (OMIM [54], BioMart [74], ExAC [52]) and used linear discriminant analysis on a training set of 493 known dominant and recessive disease genes in order to generate the best model, which could separate efficiently the two categories of genes. In the end, from more than 200 features

selected, we retained 8 features, corresponding to the missenses load and donor site load per gene, the embryonic gene expression, the gene's conservation, the number of paralogs and Gene Ontology terms. Our predictive tool is being further tested to validate the results on known genes and it will then be applied to unsolved dominant cases.

Benefits of NGS results

NGS allows a better characterization of human diseases and treatable actions may be undertaken according to the patient's specific genetic alterations. One very good example is gene therapy for patients with retinal degenerations. The idea is to replace the non-functioning gene with a fully functional gene by the injection of a harmless virus in the retina, which delivers a correct version of the gene into the retinal cells. Although these treatments are limited for the moment to clinical trials and are only available for few genes, new therapies are continuously being developed for more and more patients and types of diseases. In cancer, via the use of NGS, personalized treatments are being undertaken according to the genetic landscape of the patients

NGS bias

We are often biased in NGS analyses as we mainly focus on the exome, a small region of the genome, which is easier to understand and which has been deeply investigated over the past years. WES allows us to focus only on these coding regions and also has limits caused by the particular sequencing protocols involving the use of defined libraries and amplification of DNA fragments. WGS gives the possibility to access every single base of the genome, but there is still no valid strategy to detect and assess the pathogenicity of singular mutations in the 98% of the genome constituted by non-coding sequences. The impact of specific mutations could also be determined via the use of other NGS data, such as the transcriptomics. In cancer research, epigenetic changes

detected with RNA-seq have shown to be important in the development of some cancers [75, 76], and our WGS study on Uveal Melanoma suggested that epigenetic changes might also have a role in the development of this cancer, especially for the metastatic forms. Further experiments including RNA-seq analysis on some of our WGS patients will be undertaken to understand better the progression of this melanoma.

Ethics in NGS

Although NGS is now routinely used in many laboratories, the privacy of the patients, their rights, and the safety of their genomic data are continuously questioned. Before sequencing someone's DNA, an informed consent needs to be signed according to the regulations of the country where it will be done. This document explains the different NGS methods, which one will be used, and the expected analyses the patient's DNA will be subject to as well as the related outcome. More specifically, secondary findings, i.e. discovery of variants leading to a specific disease, which is not the one prescribed by the geneticist, are under very particular regulation [77]. The patient has the right to refuse knowing such findings, even if the pathogenic but actionable mutations (e.g. known pathogenic mutations in *BRCA1*) should be communicated to the clinical doctor. In order to secure as much as possible the genetic data, all patient's personal information is directly anonymized before being sent for sequencing. One easy way to reduce the incidental findings in NGS clinical diagnostic is the usage of *in-silico* panels, which focused only on a list of genes already associated with the phenotype of the patients. Finally, all genetic data of any patient has to be securely stored, under strong firewall protection, and usually in an internal server with limited access to outside.

Conclusion

Human genetic diseases are numerous and many of them are not yet elucidated at their genetic level. Knowledge of the molecular defects leading to these disorders could offer many benefits to the patients, their family, but also to the society. Indeed, patients who are accurately diagnosed could receive appropriate treatment, if available. Furthermore, the progression of their disease could be predicted and monitored, and assessment of risk of inheritance to the offspring may help them in central decision-making processes (e.g. prenatal diagnostics, preventive screening).

NGS has been central to the rapid evolution of current human molecular genetics and bioinformatics analyses have certainly lead the way of this important and hopefully continuous process. In less than 10 years, the usage of NGS technology has become a routine in many laboratories and clinical facilities, allowing the identification of hundreds of new genes for novel or existing syndromes and a better understanding of the genetic landscape of cancers.

Nowadays, genomic medicine can be considered as personalized, as each patient can be examined individually with the sequencing of its genome or exome. Appropriate and sometimes unique treatments may be given to patients with very rare disorders, or specific subtypes of cancers. The personalized genetic information can also be used for more common conditions such as asthma or hypertension.

The price of sequencing is continuously decreasing while the accuracy is increasing, and we can imagine that, in the next future, NGS will be undertaken to many more patients and will replace current procedures, which might be too costly and with similar or less accurate results (e.g. CGH array, Sanger sequencing). The number of biobanks of genetic information is exponentially increasing. As a result, not only Mendelian diseases may be easily solved, but also complex genetic disorders may be better characterized,

leading to a better care for a larger number of subjects. In order to achieve so, constant technological development of NGS technology, and in particular of *in silico* analytical tools would play a fundamental role for tomorrow's medicine.

4 Bibliography

1. Feuk, L., A.R. Carson, and S.W. Scherer, *Structural variation in the human genome.* Nat Rev Genet, 2006. **7**(2): p. 85-97.
2. Freeman, J.L., et al., *Copy number variation: new insights in genome diversity.* Genome Res, 2006. **16**(8): p. 949-61.
3. Johnsen, J.M., D.A. Nickerson, and A.P. Reiner, *Massively parallel sequencing: the new frontier of hematologic genomics.* Blood, 2013. **122**(19): p. 3268-75.
4. Anthony JF Griffiths, J.H.M., David T Suzuki, Richard C Lewontin, and William M Gelbart, *An Introduction to Genetic Analysis, 7th edition*. 2000: W. H. Freeman.
5. Howlader N, N.A., Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA *SEER Cancer Statistics Review, 1975-2013.* 2016.
6. Shenenberger, D.W., *Cutaneous malignant melanoma: a primary care perspective.* Am Fam Physician, 2012. **85**(2): p. 161-8.
7. Missotten, G.S., et al., *Conjunctival melanoma in the Netherlands: a nationwide study.* Invest Ophthalmol Vis Sci, 2005. **46**(1): p. 75-82.
8. Yu, G.P., D.N. Hu, and S.A. McCormick, *Latitude and incidence of ocular melanoma.* Photochem Photobiol, 2006. **82**(6): p. 1621-6.
9. Griewank, K.G., et al., *Conjunctival melanomas harbor BRAF and NRAS mutations and copy number changes similar to cutaneous and mucosal melanomas.* Clin Cancer Res, 2013. **19**(12): p. 3143-52.
10. Shields, J.A., C.L. Shields, and P. De Potter, *Surgical management of conjunctival tumors. The 1994 Lynn B. McMahan Lecture.* Arch Ophthalmol, 1997. **115**(6): p. 808-15.
11. Singh, A.D., M.E. Turell, and A.K. Topham, *Uveal melanoma: trends in incidence, treatment, and survival.* Ophthalmology, 2011. **118**(9): p. 1881-5.
12. Bergman, L., et al., *Incidence of uveal melanoma in Sweden from 1960 to 1998.* Invest Ophthalmol Vis Sci, 2002. **43**(8): p. 2579-83.
13. Luke, J.J., et al., *Biology of advanced uveal melanoma and next steps for clinical therapeutics.* Pigment Cell Melanoma Res, 2015. **28**(2): p. 135-47.
14. Woodman, S.E., *Metastatic uveal melanoma: biology and emerging treatments.* Cancer J, 2012. **18**(2): p. 148-52.
15. Shields, J.A. and C.L. Shields, *Management of posterior uveal melanoma: past, present, and future: the 2014 Charles L. Schepens lecture.* Ophthalmology, 2015. **122**(2): p. 414-28.
16. Hu, D.N., et al., *Population-based incidence of uveal melanoma in various races and ethnic groups.* Am J Ophthalmol, 2005. **140**(4): p. 612-7.
17. Park, S.J., et al., *Nationwide Incidence of Ocular Melanoma in South Korea by Using the National Cancer Registry Database (1999-2011).* Invest Ophthalmol Vis Sci, 2015. **56**(8): p. 4719-24.
18. Holly, E.A., et al., *Intraocular melanoma linked to occupations and chemical exposures.* Epidemiology, 1996. **7**(1): p. 55-61.
19. Schmidt-Pokrzywniak, A., et al., *Positive interaction between light iris color and ultraviolet radiation in relation to the risk of uveal melanoma: a case-control study.* Ophthalmology, 2009. **116**(2): p. 340-8.
20. Seddon, J.M., et al., *Host factors, UV radiation, and risk of uveal melanoma. A case-control study.* Arch Ophthalmol, 1990. **108**(9): p. 1274-80.
21. Vajdic, C.M., et al., *Sun exposure predicts risk of ocular melanoma in Australia.* Int J Cancer, 2002. **101**(2): p. 175-82.
22. Furney, S.J., et al., *SF3B1 mutations are associated with alternative splicing in uveal melanoma.* Cancer Discov, 2013. **3**(10): p. 1122-9.

23. Onken, M.D., et al., *Gene expression profiling in uveal melanoma reveals two molecular classes and predicts metastatic death.* Cancer Res, 2004. **64**(20): p. 7205-9.
24. Decatur, C.L., et al., *Driver Mutations in Uveal Melanoma: Associations With Gene Expression Profile and Patient Outcomes.* JAMA Ophthalmol, 2016. **134**(7): p. 728-33.
25. Singh, N., A.D. Singh, and W. Hide, *Inferring an Evolutionary Tree of Uveal Melanoma From Genomic Copy Number Aberrations.* Invest Ophthalmol Vis Sci, 2015. **56**(11): p. 6801-9.
26. Royer-Bertrand, B. and C. Rivolta, *Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer.* Cell Mol Life Sci, 2015. **72**(8): p. 1463-71.
27. Campbell, C.D. and E.E. Eichler, *Properties and rates of germline mutations in humans.* Trends Genet, 2013. **29**(10): p. 575-84.
28. Kong, A., et al., *Rate of de novo mutations and the importance of father's age to disease risk.* Nature, 2012. **488**(7412): p. 471-5.
29. Shyr, C., et al., *FLAGS, frequently mutated genes in public exomes.* BMC Med Genomics, 2014. **7**: p. 64.
30. den Hollander, A.I., et al., *Lighting a candle in the dark: advances in genetics and gene therapy of recessive retinal dystrophies.* J Clin Invest, 2010. **120**(9): p. 3042-53.
31. *The Retinal Information Network*. 2009, Stephen P. Daiger, PhD, Administrator.
32. Nishiguchi, K.M., et al., *Whole genome sequencing in patients with retinitis pigmentosa reveals pathogenic DNA structural changes and NEK2 as a new disease gene.* Proc Natl Acad Sci U S A, 2013. **110**(40): p. 16139-44.
33. Stenson, P.D., et al., *The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine.* Hum Genet, 2014. **133**(1): p. 1-9.
34. Sanger, F., et al., *Nucleotide sequence of bacteriophage phi X174 DNA.* Nature, 1977. **265**(5596): p. 687-95.
35. Wetterstrand, K., *Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*
36. Quail, M.A., et al., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC Genomics, 2012. **13**: p. 341.
37. Loman, N.J., et al., *Performance comparison of benchtop high-throughput sequencing platforms.* Nat Biotechnol, 2012. **30**(5): p. 434-9.
38. Laver, T.W., et al., *Pitfalls of haplotype phasing from amplicon-based long-read sequencing.* Sci Rep, 2016. **6**: p. 21746.
39. Goodwin, S., J.D. McPherson, and W.R. McCombie, *Coming of age: ten years of next-generation sequencing technologies.* Nat Rev Genet, 2016. **17**(6): p. 333-51.
40. Sabour, L., M. Sabour, and S. Ghorbian, *Clinical Applications of Next-Generation Sequencing in Cancer Diagnosis.* Pathol Oncol Res, 2016.
41. Thomas, F., et al., *Impact of tumor sequencing on the use of anticancer drugs.* Curr Opin Oncol, 2014. **26**(3): p. 347-56.
42. Yu, H., et al., *Rapid molecular diagnostics of severe primary immunodeficiency determined by using targeted next-generation sequencing.* J Allergy Clin Immunol, 2016. **138**(4): p. 1142-1151 e2.
43. Zhu, W., et al., *Next-generation molecular diagnosis: single-cell sequencing from bench to bedside.* Cell Mol Life Sci, 2016.

44. Drmanac, R., et al., *Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays.* Science, 2010. **327**(5961): p. 78-81.
45. Griffith, M., et al., *Optimizing cancer genome sequencing and analysis.* Cell Syst, 2015. **1**(3): p. 210-223.
46. Corton, M., et al., *Exome sequencing of index patients with retinal dystrophies as a tool for molecular diagnosis.* PLoS One, 2013. **8**(6): p. e65574.
47. Gilissen, C., et al., *Genome sequencing identifies major causes of severe intellectual disability.* Nature, 2014. **511**(7509): p. 344-7.
48. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
49. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
50. Hwang, S., et al., *Systematic comparison of variant calling pipelines using gold standard personal exome variants.* Sci Rep, 2015. **5**: p. 17875.
51. Van der Auwera, G.A., et al., *From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.* Curr Protoc Bioinformatics, 2013. **11**(1110): p. 11.10.1-11.10.33.
52. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.
53. Fuentes Fajardo, K.V., et al., *Detecting false-positive signals in exome sequencing.* Hum Mutat, 2012. **33**(4): p. 609-13.
54. Amberger, J., et al., *McKusick's Online Mendelian Inheritance in Man (OMIM).* Nucleic Acids Res, 2009. **37**(Database issue): p. D793-6.
55. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.* Nucleic Acids Res, 2010. **38**(16): p. e164.
56. Liu, X., et al., *dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs.* Hum Mutat, 2016. **37**(3): p. 235-41.
57. Carter, H., et al., *Identifying Mendelian disease genes with the variant effect scoring tool.* BMC Genomics, 2013. **14 Suppl 3**: p. S3.
58. Pollard, K.S., et al., *Detection of nonneutral substitution rates on mammalian phylogenies.* Genome Res, 2010. **20**(1): p. 110-21.
59. Landrum, M.J., et al., *ClinVar: public archive of relationships among sequence variation and human phenotype.* Nucleic Acids Res, 2014. **42**(Database issue): p. D980-5.
60. Yeo, G. and C.B. Burge, *Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals.* J Comput Biol, 2004. **11**(2-3): p. 377-94.
61. Gilissen, C., et al., *Disease gene identification strategies for exome sequencing.* Eur J Hum Genet, 2012. **20**(5): p. 490-7.
62. Rivolta, C., et al., *UV light signature in conjunctival melanoma; not only skin should be protected from solar radiation.* J Hum Genet, 2016. **61**(4): p. 361-2.
63. Dikoglu, E., et al., *Mutations in LONP1, a mitochondrial matrix protease, cause CODAS syndrome.* Am J Med Genet A, 2015. **167**(7): p. 1501-9.
64. Strauss, K.A., et al., *CODAS syndrome is associated with mutations of LONP1, encoding mitochondrial AAA+ Lon protease.* Am J Hum Genet, 2015. **96**(1): p. 121-35.
65. Royer-Bertrand, B., et al., *Mutations in the heat-shock protein A9 (HSPA9) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia.* Sci Rep, 2015. **5**: p. 17154.

66. van Karnebeek, C.D., et al., *NANS-mediated synthesis of sialic acid is required for brain and skeletal development.* Nat Genet, 2016. **48**(7): p. 777-84.
67. Bonafe, L., et al., *Brief Report: Peripheral Osteolysis in Adults Linked to ASAH1 (Acid Ceramidase) Mutations: A New Presentation of Farber's Disease.* Arthritis Rheumatol, 2016. **68**(9): p. 2323-7.
68. Haack, T.B., et al., *Biallelic Mutations in NBAS Cause Recurrent Acute Liver Failure with Onset in Infancy.* Am J Hum Genet, 2015. **97**(1): p. 163-9.
69. Maksimova, N., et al., *Neuroblastoma amplified sequence gene is associated with a novel short stature syndrome characterised by optic nerve atrophy and Pelger-Huet anomaly.* J Med Genet, 2010. **47**(8): p. 538-48.
70. Segarra, N.G., et al., *NBAS mutations cause a multisystem disorder involving bone, connective tissue, liver, immune system, and retina.* Am J Med Genet A, 2015. **167A**(12): p. 2902-12.
71. Nikopoulos, K., et al., *Mutations in CEP78 Cause Cone-Rod Dystrophy and Hearing Loss Associated with Primary-Cilia Defects.* Am J Hum Genet, 2016. **99**(3): p. 770-6.
72. Sergouniotis, P.I., et al., *Biallelic variants in TTLL5, encoding a tubulin glutamylase, cause retinal dystrophy.* Am J Hum Genet, 2014. **94**(5): p. 760-9.
73. Bedoni, N., et al., *Mutations in the polyglutamylase gene TTLL5, expressed in photoreceptor cells and spermatozoa, are associated with cone-rod degeneration and reduced male fertility.* Hum Mol Genet, 2016.
74. Smedley, D., et al., *The BioMart community portal: an innovative alternative to large, centralized data repositories.* Nucleic Acids Res, 2015. **43**(W1): p. W589-98.
75. Cironi, L., et al., *Epigenetic features of human mesenchymal stem cells determine their permissiveness for induction of relevant transcriptional changes by SYT-SSX1.* PLoS One, 2009. **4**(11): p. e7904.
76. Sze, C.C. and A. Shilatifard, *MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer.* Cold Spring Harb Perspect Med, 2016.
77. Green, R.C., et al., *ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing.* Genet Med, 2013. **15**(7): p. 565-74.

5 Articles (main contributor)

5.1 Overview of NGS analysis of human diseases

Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer.

Own contribution: Reviewed of the literature concerning the usage of WGS in medical genetics and cancer.

Cell. Mol. Life Sci. (2015) 72:1463–1471
DOI 10.1007/s00018-014-1807-9

Cellular and Molecular Life Sciences

REVIEW

Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer

Beryl Royer-Bertrand · Carlo Rivolta

Received: 29 June 2014 / Revised: 12 December 2014 / Accepted: 15 December 2014 / Published online: 30 December 2014

Abstract The past decade has seen the emergence of next-generation sequencing (NGS) technologies, which have revolutionized the field of human molecular genetics. With NGS, significant portions of the human genome can now be assessed by direct sequence analysis, highlighting normal and pathological variants of our DNA. Recent advances have also allowed the sequencing of complete genomes, by a method referred to as whole genome sequencing (WGS). In this work, we review the use of WGS in medical genetics, with specific emphasis on the benefits and the disadvantages of this technique for detecting genomic alterations leading to Mendelian human diseases and to cancer.

Keywords Exome · Hereditary disease · Mutation · Polymorphism

Introduction

Identification of pathogenic DNA variants causing diseases is one of the main aims of medical genetic investigations. In the past, when direct DNA sequencing possibilities were limited, this goal was achieved only in cases for which the region of the genome harboring a given mutation could be reduced to a manageable size by other procedures, such as family-based linkage or haplotype analyses. In the absence of large pedigrees or of other favorable factors that could help this localization process, disease-causing variants could remain undetected for years. The recent commercialization of next-generation sequencing (NGS) platforms has introduced a substantial methodological shift in mutation detection procedures. Specifically, it has allowed the querying of megabases of DNA at once, through computer-based alignment of millions of short sequence reads [1]. Parallel sequencing of panels of candidate disease genes and whole exome sequencing (WES), investigating all of our protein-coding DNA ($\sim$2 % of the human genome), have now become routine procedures in most laboratories.

As the NGS technique develops, the price per sequenced base decreases, to the point that sequencing entire individual genomes is not a prohibitive effort any more. Compared to WES, the use of whole genome sequencing (WGS) in human genetics, and especially in medical genetics, is still in its infancy. The reasons for this delay are mainly two: WGS involves higher costs compared to WES and requires more complex analyses at the computational level. Unlike WES, however, WGS allows the identification of complex DNA variants that are not limited to the coding sequences of the genome and the detection of non-conventional events involving large stretches of DNA (Table 1). Moreover, WGS displays an increased sensitivity with respect to WES in relationship to coding sequences as well, as it analyzes contiguous DNA and allows better sequencing and mapping approaches. More specifically, since it is not limited by constraints originating from discontinuous DNA templates (captured exons), WGS can take advantage of information deriving from a "regional" context. For instance, WGS can identify gene fusions, duplications of exons, and other genetic defects that would likely be missed in the absence of information from surrounding, non-coding DNA, which is seldom targeted by pre-WES purification procedures. Coverage (number of times a given nucleotide is sequenced) in WGS

B. Royer-Bertrand · C. Rivolta (✉)
Department of Medical Genetics, University of Lausanne, Rue Du Bugnon 27, 1005 Lausanne, Switzerland
e-mail: carlo.rivolta@unil.ch

Table 1 Features of whole genome sequencing (WGS) vs. whole exome sequencing (WES)

Feature	WGS	WES
Exonic variants	Yes	Yes
Intronic variants	Yes	No
Intergenic variants	Yes	No
Indels	Yes	Yes
Copy number variations	Yes	Not directly/imprecise
Large insertions and deletions	Yes	Not directly/imprecise
Transposable elements	Yes	Not directly/imprecise

Detection of copy number variations, large insertion and deletions, as well as of transposable elements are imprecise in WES since data are available for coding regions only, and these events can originate elsewhere

is also in general more uniform, since genomic DNA is provided to the sequencer "as is", without undergoing selection procedures that may artificially create an uneven representation of the template material to be sequenced.

Unfortunately, the wealth of information produced by WGS, despite being preferable from a theoretical standpoint, may as well represent a burden for the identification of DNA variants meaningful to medical genetics. Such variants typically consist of one or a few mutations that have to be distinguished from thousands of benign DNA changes, and their identification has often been compared to the detection of a needle in a haystack. To follow the same analogy, WGS provides better chances of identifying pathological targets than WES, but at the same time it increases the size of the haystack, to the point that innocuous DNA changes may no longer be recognized as such. The advantages of WGS procedures can therefore be fully achieved only when analytical approaches can efficiently differentiate abnormal DNA changes from the multitude of benign variants that determine normal human heterogeneity.

To better illustrate all of these concepts, this review will focus on the use of WGS as a tool to detect rare DNA variants with a high phenotypic effect, such as germline mutations in Mendelian hereditary disorders and somatic mutations in cancer.

The medical genome: generalities and common procedures

The human reference genome

Because of the complexity of the human genome, NGS reads from WGS projects cannot be efficiently assembled via de novo procedures, but have to be mapped to a standard template sequence, the human "reference sequence". This human reference genome is a pooled sequence data of 13 healthy individuals with European ancestry [2], and has gradually evolved with the improvement of sequencing methods. It provides a common and unambiguous system of relationships between genomic coordinates and corresponding DNA bases.

Mapping of sequence reads and identification of variants

Following the generation of the raw DNA sequence reads by an NGS platform, the process of obtaining the full genome sequence of an individual (or, better, a reliable approximation of it), consists of a two-step, computer-based procedure. First, the short NGS reads are mapped to the reference genome by assigning to them specific genomic coordinates. This procedure is in general computer intensive and is achieved by the use of various algorithms (e.g., BWA [3], AGILE [4], NovoAlign [novocraft.com], or FastHASH [5]). Then, mismatches between the reference genome and the individual genome are assessed by a bioinformatic process referred to as "variant calling" (e.g., via software such as GATK [6] or VCMM [7]).

Both mapping and variant calling procedures can be highly parameterized and are susceptible to producing different outputs as a function of such parameters. Therefore, although for a given individual there is only one physical genome, made of DNA, at the present time we can only obtain one or more imperfect representations of it, made of bits and bytes. As a general rule, each step of any genome analysis produces both false positives, i.e., variants that are called but are not physically present in the genome, and false negatives, i.e., variants that are not called but are present in the physical genome. It is therefore important to minimize errors at these initial mapping and variant calling steps, since all of downstream analyses will be made on the assumption that these data are a faithful representation of the physical genome.

General filtering procedures

Since every WGS project produces on average ~4,000,000 called variants [8, 9], identification of mutations relies on a series of filtering procedures that have as goal to recognize rare DNA changes with a pathogenic effect and discard the multitude of variants that are unrelated to the disease studied. Comparison with databases reporting information from the unaffected population such as dbSNP [10], the ESP database (evs.gs.washington.edu), the Exome Aggregation Consortium (ExAC) (exac.broadinstitute.org), etc. represents the most consistent filtering step, under the assumptions that such public databases report (a) reliable information and (b) include polymorphic variants having

no direct relationship with genetic diseases. However, these databases have limitations such as the presence of very rare and pathogenic mutations [11] and artifacts [12].

The frequency of the detected variants in the general population could be taken into consideration during filtering procedures, since alleles from some (mostly recessive) diseases may very well be present in the general, unaffected population [13, 14]. Furthermore, most of these entries contain information about genotype and allele frequency in different human populations, allowing as well other important analyses. In addition to comparisons with data providing information on biological variability, filtering from technical errors should also be put in place. NGS platforms as well as mapping and variant calling pipelines tend to produce technical noise (false positives) that is luckily rather constant and sequence specific. Comparison with a small set of control samples sequenced by the same NGS platform and processed by the same informatics pipeline would help to remove errors from the genomes.

Since a considerable amount of variants still survive general filtering, it may be useful to incorporate in the analysis a predictive tool that scores the impact of coding DNA changes on the corresponding protein sequence and, possibly, function. There are currently many software packages that can perform these tasks and compute whether a given variant potentially affects protein formation, expression, and/or interaction with other proteins. Among those that are used most often, we can cite SIFT [15], PROVEAN [16], PolyPhen-2 [17], and GERP++ [18]. Since prediction tools are not always concordant and their output based on different parameters, most studies use a combination of two or more tools to infer the putative pathogenicity of the variants [19–21]. However, it is important to stress that all these packages provide information of predictive nature, and that filtering procedures based on them will have in the end only a relative value.

Databases of disease-associated variants

Many public databases reporting the direct relationship between DNA changes and specific traits exist and are publicly available. Some of them contain information on variants that underlie or are associated with diseases, such as the Human Gene Mutation Database [13] or the Online Mendelian Inheritance in Man database (OMIM) [22]. For structural variations, the Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER) [23] lists copy number variations present in the control and affected populations. For cancer studies, the Catalogue of Somatic Mutations in Cancer (COSMIC) [24] stores bona fide somatic mutations related to human cancers. Some other databases collect the results from pharmacogenetic studies to contribute to the development of individualized treatments (PharmGKB [25], Pharmaco-miR [26]). All these databases have increased substantially in size in recent years, due to NGS and larger and larger genetic studies. If integrated in WGS endeavors, they can be of great help in highlighting genetic variants associated with pathological traits.

Germline mutations

WGS in hereditary diseases

Pathogenic mutations with a high phenotypic effect can either be inherited from a person's parents (germline mutations) or be acquired throughout life (somatic mutations). Pathologies resulting from germline mutations, which can be transmitted to the following generations, are commonly referred to as hereditary diseases, while somatic DNA injuries are usually not transmittable to the offspring and lead in general to tumors. Both germline and somatic mutations can be efficiently identified by WGS; however, technical and analytical approaches to detect these pathogenic variants are rather different (Fig. 1). A review of the recent literature shows that hereditary complex disorders, for which a combination of common variants in different genes and environmental factors contribute to the pathology, are still mostly investigated via non-NGS techniques. Conversely, WGS is beginning to be systematically used as a tool to understand the causes of Mendelian inherited diseases, resulting from germline mutations in one gene (e.g., [20, 27, 28]).

The initial approach for the detection of Mendelian mutations by WGS is virtually the same as that used for WES-based studies. It consists of focusing first on the coding region of genes, more specifically on variants leading to a change in the amino acid sequence of future proteins. However, the real power of WGS emerges when events involving non-coding regions are investigated. Compared to other techniques, WGS allows us to specifically extract information from parts of the genome that are usually neglected, and at a base-pair resolution. Recent WGS studies have indeed shown that a number of unsolved cases from Mendelian disease can be explained by mutations in non-coding regions and, at various degrees, involving coding parts of disease genes (e.g., [8, 29]). Similar examples include the direct identification of gene disruption by the insertion of mobile elements, which are already known to play a significant role in the molecular etiology of hereditary diseases [30], but that are difficult to identify by other NGS techniques than WGS (own unpublished results).

It is important to note that, regardless of the type of mutation, in all Mendelian disorders and within single

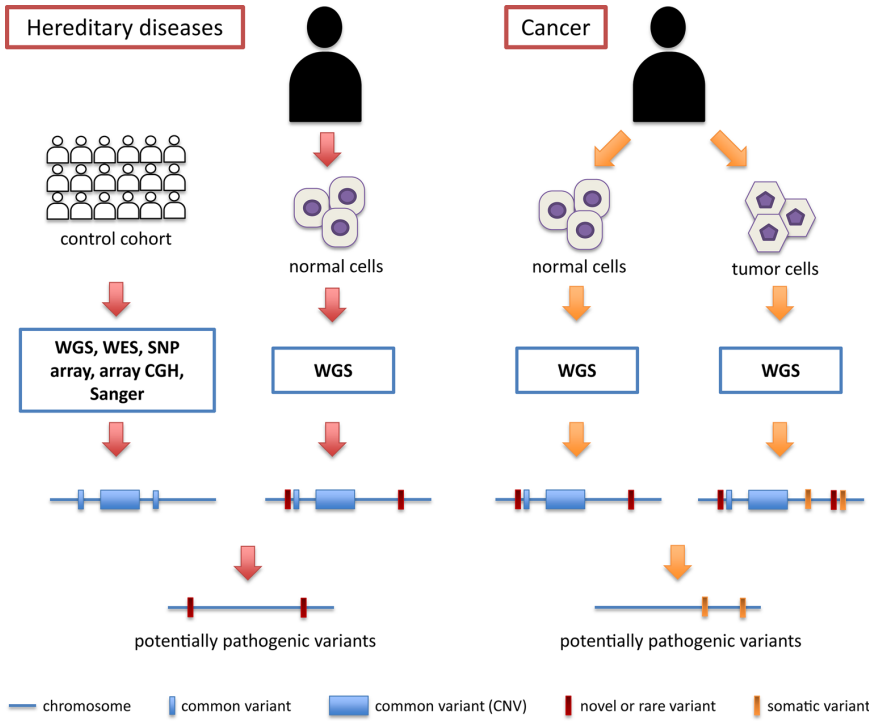


Fig. 1 Schematic workflow for the detection of potentially pathogenic DNA variants in hereditary diseases and in cancer. In hereditary diseases, the information from several genomes from a control cohort (white individuals) is assembled to produce a "metagenome" that includes all possible variants (both small events and copy number variations, or CNVs) that are allegedly not causing disease in the general population (*blue bars* and *boxes*). Potentially pathogenic variants are then deduced by comparing the WGS information from a patient (black individual) with that of the metagenome. In cancer, there is no need to query a control cohort, since the control information is provided by the genome of normal cells from the same patient. Regardless of their frequency in the general population, all these variants are then subtracted from the pool of DNA changes obtained from the tumor genome, making the detection process of pathogenic variants a more efficient and straightforward procedure

pedigrees, pathogenic variants always co-segregate with the disease in affected individuals. Therefore, all patients within a family should necessarily share the same mutation(s) but not necessarily the same innocuous DNA variants. This elementary concept of human genetics is one of the most powerful elements of investigation in NGS studies, including WGS, since it allows us to discard benign variants that cannot be immediately recognized as such. One of the first WGS projects that exploited this paradigm is the one performed by Roach et al., who, following the comparison of individual WGS output from two healthy parents and two affected children, could reduce the number of candidates genes, genomewide, from thousands to only four [31].

For monogenic disorders with no genetic heterogeneity, a similar strategy could be extended from a single pedigree to a group of unrelated patients. In these cases, merging genomic data from different patients and different pedigrees represents a much more powerful approach, because unrelated affected individuals would all tend to have rare variants (mutations) only in the disease gene [27]. Conversely, in Mendelian disorders displaying genetic heterogeneity, this approach may lead to false positive results, highlighting as pathogenic benign variants that may be coincidentally shared by a group of patients, and therefore it should not be used.

Identification of recessive, dominant, or X-linked mutations

Since heterozygous recessive mutations do not cause disease, they can be present, even at non-negligible frequencies, in the general population [13, 14]. Patients would conversely be either homozygotes for a mutation or compound heterozygotes for two different mutations in the same gene (Fig. 2). This simple genetic concept has tremendous consequences in WGS-based searching for mutations, as only about a dozen genes will harbor rare,

non-synonymous variants in the homozygous or compound heterozygous state genomewide. Other methods could be used to identify pathogenic variants, such as the elevated granularity of WGS data, which allows precise haplotype phasing in trios or quartets, to the point that meiotic recombination events in the parents of an index case could be detected. In other words, it is possible to identify all the regions of the genome identical by descent in affected individuals of a kindred, which by definition should harbor both the mutation transmitted from the father and from the mother [31, 32]. An extension of the same concept is autozygosity mapping, which reaches its highest possible precision when WGS information is used. This technique scores stretches of homozygous alleles (usually SNPs) in consanguineous families segregating a recessive disease to detect the single homozygous recessive mutation originating from a heterogeneous mutation in a common ancestor. Since alleles are transmitted from one generation to the other in large genomic "blocks" by meiotic recombination, the genomic region surrounding this homozygous mutation would also be completely homozygous for benign variants, which would act as a beacon for the presence of pathogenic recessive mutations [33].

In contrast, in autosomal dominant conditions, only one variant in a specific disease gene gives rise to the pathological phenotype. Compared to recessive cases, it is more difficult to infer pathogenicity of a given DNA variant since, in absence of other information, in principle all of the rare DNA changes detected genomewide can be the mutation causing disease (Fig. 2). Filtering steps as well as the careful use of clinical and public databases and pedigree-based co-segregation analyses become therefore essential. In case the condition is known to display no genetic heterogeneity, then the most powerful tool to infer pathogenicity becomes data merging across

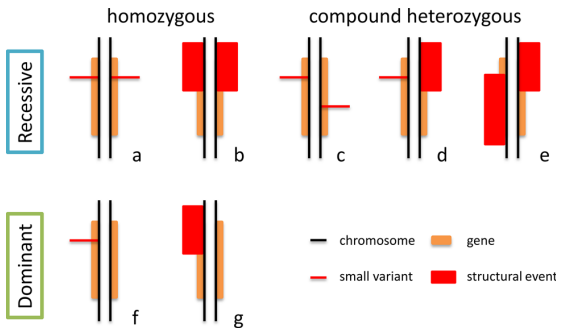


Fig. 2 Possible configurations of pathogenic mutations for autosomal recessive and autosomal dominant conditions. Structural events are usually better or exclusively detected via WGS procedures and therefore genotypes *b*, *d*, *e*, and *g* may be easily missed by other sequencing techniques

different unrelated patients, for the reasons described above.

Recent literature has shown that a substantial proportion of seemingly dominant cases may also result from the presence of de novo mutations [34–37]. In such cases, trio analyses would be the best strategy to choose, since appearance of de novo mutations would be easily scored by subtracting the list of genomic variants of the patient from those of their parents, without in principle the need to filter data from common variation databases.

Finally, procedures for X-linked cases would be substantially the same as those for dominant ones, with the exception that the genomic region to be considered would be limited only to the X chromosome.

Somatic mutations

WGS in cancer

As mentioned, DNA errors can also be acquired somatically through life. Because of age, environment, diet, etc., these mutations are usually not transmitted to the offspring but can accumulate and lead to disease. This is the case of most cancers, where somatic defects lead to a dysregulated cell growth and eventually to tumor and metastasis.

Detection of pathogenic somatic variants via WGS procedures is a much simpler effort, compared to that involving germline mutations in hereditary diseases. Indeed, the cancer genome of a given patient can be directly compared with that from tumor-free tissues from the same individual (usually blood leukocytes). This process eliminates the need for constructing an imprecise reference "metagenome" resulting from cohorts of unrelated patients. In this context, the fact that a given individual's germline genome carries polymorphic variants, rare DNA changes, or even large structural variations with respect to control genomes is completely irrelevant, since the mutations that count are those present in the cancer genome only (Fig. 1). In other words, the germline genome represents a baseline dataset used as a subtracting factor to obtain an unbiased count of all the acquired somatic mutations. Ley et al. were among the first to apply this method on an acute myeloid leukemia, identifying in the end two known mutations for cancer progression and eight novel mutations that could be used for possible targeted therapy [38].

Cancer appearance, progression, relapse

Since cancer is an evolving disorder, WGS can be used to score tumor progression, relapse, and remission by analyzing its genomic content at different time points.

Concerning tumor progression, a study by Ding et al. [39] investigated basal-like breast cancer via four parallel WGS procedures: on the peripheral blood of the patient to obtain a baseline genome, on the primary tumor to detect the somatic mutations, on a brain metastasis to understand metastatic transformation, and on the genome of a human-to-mouse primary tumor xenograph to understand the mechanisms of tumor changes following transplantation.

Tumor evolution in the context of therapeutic treatments can also be studied by WGS, as shown by a report on clonal evolution in acute myeloid leukemia cells, a cancer characterized by frequent relapses following chemotherapy treatment [40]. In this work, the authors noted two distinct patterns of tumor genome evolution: in the first one, the primary tumor clone gained mutations that made it to evolve into the relapse clone and therefore survive treatment; in the second one, chemotherapy applied a selective pressure enabling a specific sub-clone of the initial tumor to expand, and again survive treatment.

Tumorigenic pathways

Although every cancer has a unique landscape of somatic events, in some instances mutations tend to affect common genes, highlighting dysregulation of shared, important pathways for tumor progression. Analyses aimed at identifying such pathways can be done by considering multiple cases of the same tumor, to increase the signal represented by driver mutations (DNA changes providing selective advantage to a cancer cell clone) and minimize the noise deriving from passenger mutations (DNA changes that do not contribute to cancer etiology but accumulate in rapidly expanding clones). In a way, such analyses are very similar to those outlined above for hereditary diseases, for which multiplication of the patients' or controls' genomes to be analyzed helps to eliminate DNA changes which are not relevant to the disease. This approach has been applied to a relatively large number of different tumors, for a total of $\sim$150 genomes analyzed [41–49]. In addition to providing new insights into mutation-based differential prognosis, tumor molecular classification, progression mechanisms, etc., comparative WGS on multiple tumor samples helps identifying tumor signatures and mutational spectra across different types of cancer [50] or within the same tumor type, such as smoker and non-smoker lung cancer genomes [43].

Conclusions

From a genetic standpoint, there is nothing more exhaustive than the full sequence of a genome. It is therefore easy to predict that, when costs associated with WGS substantially decrease and better analytical tools are available, this procedure will become the technique of choice for most medical genetics investigations. WGS can in fact detect features of the human genome, such as copy number variations and intronic mutations, that other techniques cannot or struggle to identify, and that are becoming increasingly relevant to human genetic pathology. Furthermore, it is conceivable that many different genetic tests, which are currently performed as individual analyses (array CGH, sequence-specific mutation detection, gene panel screening, etc.) could be soon replaced by a single WGS run, which in fact can provide all of this information at once.

However, for WGS to become a popular tool in research and a routine test for DNA diagnosis, a few improvements still have to be made. From a clinical standpoint, diagnosis of the disease has to be very accurate, especially in terms of inheritance, because all downstream analyses would depend on it. Also, since a person's whole genome is unveiled, the risk of incidental findings is very high, revealing the need for integrating ethical policies adapted to this specific test. On the technical side, sequencing errors and noise have to be better estimated and eliminated, since false positive findings or long processing times are not compatible with diagnostic needs. This could be done by optimizing the reference genome, databases of common variants, prediction software, and also pre-WGS experimental design (e.g., by including specific information of a patient's family). In a more distant future, complex diseases will probably also be approached by WGS, to fully exploit the wealth of information that this technique produces in the context of variants that are not pathogenic *per se*, but that can cause disease via additive or multiplicative effects.

Acknowledgments We would like to acknowledge the Swiss National Science Foundation (Grant 310030_138346) and Fond'action contre le cancer Lausanne for their financial support. We are also indebted to Dr. Aaron F. McDaid for his critical revision of this text.

References

1. Pareek CS, Smoczynski R, Tretyn A (2011) Sequencing technologies and genome sequencing. J Appl Genet 52:413–435
2. (2010) E pluribus unum. Nat Methods 7:331
3. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754–1760
4. Misra S, Agrawal A, Liao WK, Choudhary A (2011) Anatomy of a hash-based long read sequence mapping algorithm for next generation DNA sequencing. Bioinformatics 27:189–195
5. Xin H, Lee D, Hormozdiari F, Yedkar S, Mutlu O, Alkan C (2013) Accelerating read mapping with FastHASH. BMC Genom 14(Suppl 1):S13
6. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

7. Shigemizu D, Fujimoto A, Akiyama S, Abe T, Nakano K, Boroevich KA, Yamamoto Y, Furuta M, Kubo M, Nakagawa H, Tsunoda T (2013) A practical method to detect SNVs and indels from whole genome and exome sequencing data. Sci Rep 3:2161
8. Nishiguchi KM, Tearle RG, Liu YP, Oh EC, Miyake N, Benaglio P, Harper S, Koskiniemi-Kuendig H, Venturini G, Sharon D, Koenekoop RK, Nakamura M, Kondo M, Ueno S, Yasuma TR, Beckmann JS, Ikegawa S, Matsumoto N, Terasaki H, Berson EL, Katsanis N, Rivolta C (2013) Whole genome sequencing in patients with retinitis pigmentosa reveals pathogenic DNA structural changes and NEK2 as a new disease gene. Proc Natl Acad Sci USA 110:16139–16144
9. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G, Dahl F, Fernandez A, Staker B, Pant KP, Baccash J, Borcherding AP, Brownley A, Cedeno R, Chen L, Chernikoff D, Cheung A, Chirita R, Curson B, Ebert JC, Hacker CR, Hartlage R, Hauser B, Huang S, Jiang Y, Karpinchyk V, Koenig M, Kong C, Landers T, Le C, Liu J, McBride CE, Morenzoni M, Morey RE, Mutch K, Perazich H, Perry K, Peters BA, Peterson J, Pethiyagoda CL, Pothuraju K, Richter C, Rosenbaum AM, Roy S, Shafto J, Sharanhovich U, Shannon KW, Sheppy CG, Sun M, Thakuria JV, Tran A, Vu D, Zaranek AW, Wu X, Drmanac S, Oliphant AR, Banyai WC, Martin B, Ballinger DG, Church GM, Reid CA (2010) Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327:78–81
10. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29:308–311
11. Kenna KP, McLaughlin RL, Hardiman O, Bradley DG (2013) Using reference databases of genetic variation to evaluate the potential pathogenicity of candidate disease variants. Hum Mutat 34:836–841
12. Yu X, Sun S (2013) Comparing a few SNP calling algorithms using low-coverage sequencing data. BMC Bioinformatics 14:274
13. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN (2014) The human gene mutation database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet 133:1–9
14. Nishiguchi KM, Rivolta C (2012) Genes associated with retinitis pigmentosa and allied diseases are frequently mutated in the general population. PLoS One 7:e41902
15. 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
16. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and indels. PLoS ONE 7:e46688
17. Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet Chapter 7(Unit7):20
18. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol 6:e1001025
19. Li MX, Kwan JS, Bao SY, Yang W, Ho SL, Song YQ, Sham PC (2013) Predicting mendelian disease-causing non-synonymous single nucleotide variants in exome sequencing studies. PLoS Genet 9:e1003143
20. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, Bainbridge M, Dinh H, Jing C, Wheeler DA, McGuire AL, Zhang F, Stankiewicz P, Halperin JJ, Yang C, Gehman C, Guo D, Irikat RK, Tom W, Fantin NJ, Muzny DM, Gibbs RA (2010) Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N Engl J Med 362:1181–1191
21. Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA, Pushkarev D, Neff NF, Hudgins L, Gong L, Hodges LM, Berlin DS, Thorn CF, Sangkuhl K, Hebert JM, Woon M, Sagreiya H, Whaley R, Knowles JW, Chou MF, Thakuria JV, Rosenbaum AM, Zaranek AW, Church GM, Greely HT, Quake SR, Altman RB (2010) Clinical assessment incorporating a personal genome. Lancet 375:1525–1535
22. Amberger J, Bocchini CA, Scott AF, Hamosh A (2009) McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res 37:D793–D796
23. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP (2009) DECIPHER: database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 84:524–533
24. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 39:D945–D950
25. Hewett M, Oliver DE, Rubin DL, Easton KL, Stuart JM, Altman RB, Klein TE (2002) PharmGKB: the Pharmacogenetics Knowledge Base. Nucleic Acids Res 30:163–165
26. Rukov JL, Wilentzik R, Jaffe I, Vinther J, Shomron N (2013) Pharmaco-miR: linking microRNAs and drug effects. Brief Bioinform
27. Gonzaga-Jauregui C, Lotze T, Jamal L, Penney S, Campbell IM, Pehlivan D, Hunter JV, Woodbury SL, Raymond G, Adesina AM, Jhangiani SN, Reid JG, Muzny DM, Boerwinkle E, Lupski JR, Gibbs RA, Wiszniewski W (2013) Mutations in VRK1 associated with complex motor and sensory axonal neuropathy plus microcephaly. JAMA Neurol 70:1491–1498
28. Bainbridge MN, Wiszniewski W, Murdock, Friedman J, Gonzaga-Jauregui C, Newsham I, Reid JG, Fink JK, Morgan MB, Gingras MC, Muzny DM, Hoang LD, Yousaf S, Lupski JR, Gibbs RA (2011) Whole-genome sequencing for optimized patient management. Sci Trans Med 3:87re83
29. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, Kwint M, Janssen IM, Hoischen A, Schenck A, Leach R, Klein R, Tearle R, Bo T, Pfundt R, Yntema HG, de Vries BB, Kleefstra T, Brunner HG, Vissers LE, Veltman JA (2014) Genome sequencing identifies major causes of severe intellectual disability. Nature
30. Deininger PL, Batzer MA (1999) Alu repeats and human disease. Mol Genet Metab 67:183–193
31. Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328:636–639
32. Dewey FE, Chen R, Cordero SP, Ormond KE, Caleshu C, Karczewski KJ, Whirl-Carrillo M, Wheeler MT, Dudley JT, Byrnes JK, Cornejo OE, Knowles JW, Woon M, Sangkuhl K, Gong L, Thorn CF, Hebert JM, Capriotti E, David SP, Pavlovic A, West A, Thakuria JV, Ball MP, Zaranek AW, Rehm HL, Church GM, West JS, Bustamante CD, Snyder M, Altman RB, Klein TE, Butte AJ, Ashley EA (2011) Phased whole-genome genetic risk in a family quartet using a major allele reference sequence. PLoS Genet 7:e1002280
33. Alkuraya FS (2010) Homozygosity mapping: one more tool in the clinical geneticist's toolbox. Genet Med 12:236–239
34. Kariminejad A, Barzegar M, Abdollahimajd F, Pramanik R, McGrath JA (2014) Olmsted syndrome in an Iranian boy with a new de novo mutation in TRPV3. Clin Exp Dermatol 39:492–495

35. Grozeva D, Carss K, Spasic-Boskovic O, Parker MJ, Archer H, Firth HV, Park SM, Canham N, Holder SE, Wilson M, Hackett A, Field M, Floyd JA, Hurles M, Raymond FL (2014) De novo loss-of-function mutations in SETD5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. Am J Hum Genet 94:618–624
36. 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 (2010) A de novo paradigm for mental retardation. Nat Genet 42:1109–1112
37. Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, Albrecht B, Bartholdi D, Beygo J, Di Donato N, Dufke A, Cremer K, Hempel M, Horn D, Hoyer J, Joset P, Ropke A, Moog U, Riess A, Thiel CT, Tzschach A, Wiesener A, Wohlleber E, Zweier C, Ekici AB, Zink AM, Rump A, Meisinger C, Grallert H, Sticht H, Schenck A, Engels H, Rappold G, Schrock E, Wieacker P, Riess O, Meitinger T, Reis A, Strom TM (2012) Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380:1674–1682
38. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, DiPersio JF, Wilson RK (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 456:66–72
39. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, Schmidt H, Kalicki J, Zhang Q, Chen L, Lin L, Wendl MC, McMichael JF, Magrini VJ, Cook L, McGrath SD, Vickery TL, Appelbaum E, Deschryver K, Davies S, Guintoli T, Lin L, Crowder R, Tao Y, Snider JE, Smith SM, Dukes AF, Sanderson GE, Pohl CS, Delehaunty KD, Fronick CC, Pape KA, Reed JS, Robinson JS, Hodges JS, Schierding W, Dees ND, Shen D, Locke DP, Wiechert ME, Eldred JM, Peck JB, Oberkfell BJ, Lolofie JT, Du F, Hawkins AE, O'Laughlin MD, Bernard KE, Cunningham M, Elliott G, Mason MD, Thompson DM Jr, Ivanovich JL, Goodfellow PJ, Perou CM, Weinstock GM, Aft R, Watson M, Ley TJ, Wilson RK, Mardis ER (2010) Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 464:999–1005
40. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, Wallis JW, Lu C, Shen D, Harris CC, Dooling DJ, Fulton RS, Fulton LL, Chen K, Schmidt H, Kalicki-Veizer J, Magrini VJ, Cook L, McGrath SD, Vickery TL, Wendl MC, Heath S, Watson MA, Link DC, Tomasson MH, Shannon WD, Payton JE, Kulkarni S, Westervelt P, Walter MJ, Graubert TA, Mardis ER, Wilson RK, DiPersio JF (2012) Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature 481:506–510
41. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, Escaramis G, Jares P, Bea S, Gonzalez-Diaz M, Bassaganyas L, Baumann T, Juan M, Lopez-Guerra M, Colomer D, Tubio JM, Lopez C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernandez JM, Puente DA, Freije JM, Velasco G, Gutierrez-Fernandez A, Costa D, Carrio A, Guijarro S, Enjuanes A, Hernandez L, Yague J, Nicolas P, Romeo-Casabona CM, Himmelbauer H, Castillo E, Dohm JC, de Sanjose S, Piris MA, de Alava E, San Miguel J, Royo R, Gelpi JL, Torrents D, Orozco M, Pisano DG, Valencia A, Guigo R, Bayes M, Heath S, Gut M, Klatt P, Marshall J, Raine K, Stebbings LA, Futreal PA, Stratton MR, Campbell PJ, Gut I, Lopez-Guillermo A, Estivill X, Montserrat E, Lopez-Otin C, Campo E (2011) Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature 475:101–105
42. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, Harview CL, Brunet JP, Ahmann GJ, Adli M, Anderson KC, Ardlie KG, Auclair D, Baker A, Bergsagel PL, Bernstein BE, Drier Y, Fonseca R, Gabriel SB, Hofmeister CC, Jagannath S, Jakubowiak AJ, Krishnan A, Levy J, Liefeld T, Lonial S, Mahan S, Mfuko B, Monti S, Perkins LM, Onofrio R, Pugh TJ, Rajkumar SV, Ramos AH, Siegel DS, Sivachenko A, Stewart AK, Trudel S, Vij R, Voet D, Winckler W, Zimmerman T, Carpten J, Trent J, Hahn WC, Garraway LA, Meyerson M, Lander ES, Getz G, Golub TR (2011) Initial genome sequencing and analysis of multiple myeloma. Nature 471:467–472
43. Liu J, Lee W, Jiang Z, Chen Z, Jhunjhunwala S, Haverty PM, Gnad F, Guan Y, Gilbert HN, Stinson J, Klijn C, Guillory J, Bhatt D, Vartanian S, Walter K, Chan J, Holcomb T, Dijkgraaf P, Johnson S, Koeman J, Minna JD, Gazdar AF, Stern HM, Hoeflich KP, Wu TD, Settleman J, de Sauvage FJ, Gentleman RC, Neve RM, Stokoe D, Modrusan Z, Seshagiri S, Shames DS, Zhang Z (2012) Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. Genome Res 22:2315–2327
44. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, Aoki M, Hosono N, Kubo M, Miya F, Arai Y, Takahashi H, Shirakihara T, Nagasaki M, Shibuya T, Nakano K, Watanabe-Makino K, Tanaka H, Nakamura H, Kusuda J, Ojima H, Shimada K, Okusaka T, Ueno M, Shigekawa Y, Kawakami Y, Arihiro K, Ohdan H, Gotoh K, Ishikawa O, Ariizumi S, Yamamoto M, Yamada T, Chayama K, Kosuge T, Yamaue H, Kamatani N, Miyano S, Nakagama H, Nakamura Y, Tsunoda T, Shibata T, Nakagawa H (2012) Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. Nat Genet 44:760–764
45. Kiel MJ, Velusamy T, Betz BL, Zhao L, Weigelin HG, Chiang MY, Huebner-Chan DR, Bailey NG, Yang DT, Bhagat G, Miranda RN, Bahler DW, Medeiros LJ, Lim MS, Elenitoba-Johnson KS (2012) Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. J Exp Med 209:1553–1565
46. Furney SJ, Pedersen M, Gentien D, Dumont AG, Rapinat A, Desjardins L, Turajlic S, Piperno-Neumann S, de la Grange P, Roman-Roman S, Stern MH, Marais R (2013) SF3B1 mutations are associated with alternative splicing in uveal melanoma. Cancer Discov 3:1122–1129
47. Weiss GJ, Liang WS, Demeure MJ, Kiefer JA, Hostetter G, Izatt T, Sinari S, Christoforides A, Aldrich J, Kurdoglu A, Phillips L, Benson H, Reiman R, Baker A, Marsh V, Von Hoff DD, Carpten JD, Craig DW (2013) A pilot study using next-generation sequencing in advanced cancers: feasibility and challenges. PLoS ONE 8:e76438
48. Craig DW, O'Shaughnessy JA, Kiefer JA, Aldrich J, Sinari S, Moses TM, Wong S, Dinh J, Christoforides A, Blum JL, Aitelli CL, Osborne CR, Izatt T, Kurdoglu A, Baker A, Koeman J, Barbacioru C, Sakarya O, De La Vega FM, Siddiqui A, Hoang L, Billings PR, Salhia B, Tolcher AW, Trent JM, Mousses S, Von Hoff D, Carpten JD (2013) Genome and transcriptome sequencing in prospective metastatic triple-negative breast cancer uncovers therapeutic vulnerabilities. Mol Cancer Ther 12:104–116
49. Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, Pugh TJ, Cherniack AD, Ambrogio L, Cibulskis K, Bertelsen B, Romero-Cordoba S, Trevino V, Vazquez-

Santillan K, Guadarrama AS, Wright AA, Rosenberg MW, Duke F, Kaplan B, Wang R, Nickerson E, Walline HM, Lawrence MS, Stewart C, Carter SL, McKenna A, Rodriguez-Sanchez IP, Espinosa-Castilla M, Woie K, Bjorge L, Wik E, Halle MK, Hoivik EA, Krakstad C, Gabino NB, Gomez-Macias GS, Valdez-Chapa LD, Garza-Rodriguez ML, Maytorena G, Vazquez J, Rodea C, Cravioto A, Cortes ML, Greulich H, Crum CP, Neuberg DS, Hidalgo-Miranda A, Escareno CR, Akslen LA, Carey TE, Vintermyr OK, Gabriel SB, Barrera-Saldana HA, Melendez-Zajgla J, Getz G, Salvesen HB, Meyerson M (2014) Landscape of genomic alterations in cervical carcinomas. Nature 506:371–375

50. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Jager N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, Lopez-Otin C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdes-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR (2013) Signatures of mutational processes in human cancer. Nature 500:415–421

5.2 Ocular Melanomas

5.2.1 Conjunctival Melanoma

UV light signature in conjunctival melanoma; not only skin should be protected from solar radiation

Own contribution: Analyzed the WGS data of CM patients and generated all the *in-silico* results.

OPEN

Journal of Human Genetics (2016) 61, 361–362
Official journal of the Japan Society of Human Genetics 1434-5161/16
www.nature.com/jhg

CORRESPONDENCE

UV light signature in conjunctival melanoma; not only skin should be protected from solar radiation

Journal of Human Genetics (2016) **61,** 361–362; doi:10.1038/jhg.2015.152; published online 10 December 2015

Conjunctival melanoma represents 5% of ocular melanomas. It is a highly recurrent tumor leading to metastatic disease and death at 10 years in 25–30% of patients.[1] Recent molecular investigations have indicated that conjunctival and cutaneous melanomas may share common features.[2] For instance, epidemiological studies have shown an association between decreasing latitude and increasing incidence of conjunctival melanoma,[3] suggesting that exposure to sunlight has a role in its etiology.[4] However, strong molecular proof of such correlation has not been established. We hereby report the results of a genome-wide sequencing effort of two conjunctival melanomas documenting a strong ultraviolet (UV) mutation signature.

We performed high-coverage whole-genome sequencing (average read depth = 80x, with 97.1% of the genome covered at least 40x) of two untreated perilimbic conjunctival melanomas occurring in two Caucasian women, aged 57 years (patient A) and 65 years (patient B) (Figures 1a and b). Somatic mutations were ascertained by comparing the DNA sequence of each tumor with their respective germline matches (blood leukocyte DNA). We identified a very high somatic mutation load composed of more than 96 000 mutations in each tumor, resulting in an average genome-wide mutational rate of 30 somatic mutations per million DNA bases. On average, we detected 802 mutations in the coding regions of genes. The ratio of non-synonymous to synonymous variants was 1.59:1, showing a prominence of passenger mutations. Non-synonymous mutations were composed mainly of missense (90%), followed by nonsense changes (7%).

Moreover, detailed analysis of the mutation spectrum uncovered a majority of cytosine to thymine (C>T) transitions, occurring in 88 (patient A) and 83% (patient B) of the single-nucleotide substitutions in these tumors (Figure 1c). In addition, 80 and 100% of these C>T changes, respectively, occurred at the 3′ end of pyrimidine dinucleotides and CC>TT changes represented >70% of all tandem substitutions, showing the presence of a typical UV mutational signature.[5]

Altogether, our molecular findings support the occurrence of UV-induced DNA damage in conjunctival melanoma indicating a link between solar radiation and development of this tumor. Awareness should therefore be raised in the general population about the benefits of wearing UV-shielding eyeglasses, in addition to skin sunscreens, when exposed to sunlight.

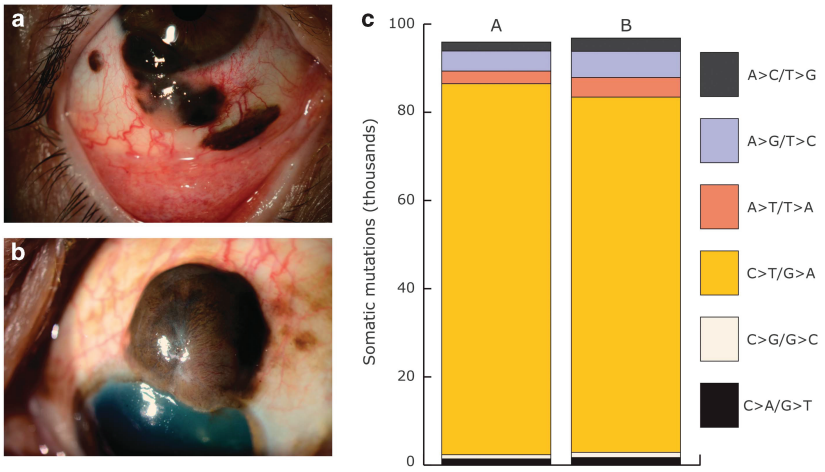


Figure 1 (**a**) A partially pigmented melanoma is seen in the inferior perilimbic area with extension onto the cornea. The isolated inferior and temporal pigmented areas were primary acquired melanosis (patient A). (**b**) A pigmented melanoma can be observed in the superior perilimbic area (patient B). (**c**) Mutation spectrum of the two conjunctival melanomas (A: patient A, B: patient B), showing a clear ultraviolet-induced DNA damage signature. No somatic mutations were detected in the melanoma-related genes *BRAF* and *NRAS*.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the 'Fond'action contre le cancer' foundation, Switzerland. We wish to thank Complete Genomics (Mountain View, CA, USA) for providing the whole-genome sequencing discussed in this report.

Carlo Rivolta[1], Beryl Royer-Bertrand[1], Donata Rimoldi[2], Ann Schalenbourg[3], Leonidas Zografos[3], Serge Leyvraz[4] and Alexandre Moulin[3]

[1]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland; [2]Ludwig Center for Cancer Research of the University of Lausanne, Épalinges, Switzerland; [3]Department of Ophthalmology, University of Lausanne, Jules Gonin Eye Hospital, Lausanne, Switzerland and [4]Department of Oncology, Lausanne University Hospital, Lausanne, Switzerland
E-mail: beryl.royer-bertrand@unil.ch

1 Missotten, G. S., Keijser, S., De Keizer, R. J. & De Wolff-Rouendaal, D. Conjunctival melanoma in the Netherlands: a nationwide study. *Invest. Ophthalmol. Vis. Sci.* **46**, 75–82 (2005).
2 Griewank, K. G., Westekemper, H., Murali, R., Mach, M., Schilling, B., Wiesner, T. *et al.* Conjunctival melanomas harbor BRAF and NRAS mutations and copy number changes similar to cutaneous and mucosal melanomas. *Clin. Cancer Res.* **19**, 3143–3152 (2013).
3 Yu, G. P., Hu, D. N. & McCormick, S. A. Latitude and incidence of ocular melanoma. *Photochem. Photobiol.* **82**, 1621–1626 (2006).
4 Griewank, K. G., Murali, R., Schilling, B., Scholz, S., Sucker, A., Song, M. *et al.* TERT promoter mutations in ocular melanoma distinguish between conjunctival and uveal tumours. *Br. J. Cancer* **109**, 497–501 (2013).
5 Ikehata, H. & Ono, T. The mechanisms of UV mutagenesis. *J. Radiat. Res.* **52**, 115–125 (2011).

5.2.2 Uveal Melanoma

Comprehensive genetic landscape of uveal melanoma by whole-genome sequencing

Own contribution: Analyzed the WGS data of UM patients and generated all the *in-silico* results.

Please cite this article in press as: Royer-Bertrand et al., Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing, The American Journal of Human Genetics (2016), http://dx.doi.org/10.1016/j.ajhg.2016.09.008

REPORT

Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing

Beryl Royer-Bertrand,[1,2] Matteo Torsello,[3] Donata Rimoldi,[4] Ikram El Zaoui,[1] Katarina Cisarova,[1] Rosanna Pescini-Gobert,[1] Franck Raynaud,[5,6] Leonidas Zografos,[7] Ann Schalenbourg,[7] Daniel Speiser,[4] Michael Nicolas,[7] Laureen Vallat,[7] Robert Klein,[8] Serge Leyvraz,[9] Giovanni Ciriello,[5,6] Nicolò Riggi,[3] Alexandre P. Moulin,[7,10] and Carlo Rivolta[1,10,*]

Uveal melanoma (UM) is a rare intraocular tumor that, similar to cutaneous melanoma, originates from melanocytes. To gain insights into its genetics, we performed whole-genome sequencing at very deep coverage of tumor-control pairs in 33 samples (24 primary and 9 metastases). Genome-wide, the number of coding mutations was rather low (only 17 variants per tumor on average; range 7–28), thus radically different from cutaneous melanoma, where hundreds of exonic DNA insults are usually detected. Furthermore, no UV light-induced mutational signature was identified. Recurrent coding mutations were found in the known UM drivers *GNAQ*, *GNA11*, *BAP1*, *EIF1AX*, and *SF3B1*. Other genes, i.e., *TP53BP1*, *CSMD1*, *TTC28*, *DLK2*, and *KTN1*, were also found to harbor somatic mutations in more than one individual, possibly indicating a previously undescribed association with UM pathogenesis. De novo assembly of unmatched reads from non-coding DNA revealed peculiar copy-number variations defining specific UM subtypes, which in turn could be associated with metastatic transformation. Mutational-driven comparison with other tumor types showed that UM is very similar to pediatric tumors, characterized by very few somatic insults and, possibly, important epigenetic changes. Through the analysis of whole-genome sequencing data, our findings shed new light on the molecular genetics of uveal melanoma, delineating it as an atypical tumor of the adult for which somatic events other than mutations in exonic DNA shape its genetic landscape and define its metastatic potential.

Despite having the very rare incidence of 5–8 new cases per million per year,[1,2] uveal melanoma (UM [MIM: 155720]) is the most common primary intraocular tumor of the adult. It develops from melanocytes in the choroid, the ciliary body, or the iris (collectively called the "uvea," one of the inner layers of the eye) and usually metastasizes through the blood stream to the liver.[3,4] Symptoms include variable and painless visual disturbances, often presenting when the tumor has already reached a considerable mass. Survival and potential therapeutic options depend, among other things, on the presence of specific genetic alterations.[5] Although population studies suggest ethnic predisposition,[6,7] environmental factors that are directly involved in the transformation process have not been clearly delineated. For instance, a possible association with UV light exposure has been suggested,[8–11] but questioned recently by molecular data.[12] Research on UM molecular genetics has been performed mostly by investigating coding mutations or copy-number variations detectable by direct sequencing of target genes, karyotype, array-CGH, MLPA, or SNP-array analyses.[13–17] As a result, mutations at codon 209 of the paralogous oncogenes *GNAQ* (MIM: 600998) and *GNA11* (MIM: 139313)[18,19] and in the tumor suppressor *BAP1* (MIM: 603089)[20] have been identified in the majority of UMs, whereas insults in *EIF1AX* (MIM: 300186) and *SF3B1* (MIM: 605590) or other genes seem to be less frequent, accounting for at most 20% of cases.[12,21–26] Moreover, copy gains and losses are common events in this tumor, typically involving chromosome 3 monosomy, 6p gain, and 8q gain.[14,17] After whole-genome sequencing of a series of tumor-control pairs, we present here an analytically unbiased and comprehensive assessment of the genetic landscape of UM.

We screened 33 UM samples (24 primary tumors and 9 unrelated metastases; Table S1) and corresponding normal tissue pairs by deep-coverage whole-genome sequencing (WGS), using the processing platform by Complete Genomics.[27] Written informed consent was obtained from all individuals enrolled in this study, and approval for human subject research was obtained from the Institutional Review Boards of all participating Institutions. Sequence reads were mapped to the human reference genome, assembly GRch37, and somatic variants in tumors were called by comparison with the matched normal genomes, as previously described.[28] All samples were surgically collected from eye enucleations or resected from liver metastases, allowing very clear post-surgery macroscopic

[1]Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, 1011 Lausanne Switzerland; [2]Center for Molecular Diseases, Lausanne University Hospital, 1011 Lausanne, Switzerland; [3]Experimental Pathology, Institute of Pathology, Lausanne University Hospital, 1011 Lausanne, Switzerland; [4]Ludwig Cancer Research, Department of Oncology, University of Lausanne, 1066 Epalinges, Switzerland; [5]Department of Computational Biology, Computational Systems Oncology, University of Lausanne, 1011 Lausanne, Switzerland; [6]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; [7]Jules-Gonin Eye Hospital, Department of Ophthalmology, Fondation Asile des Aveugles, University of Lausanne, 1004 Lausanne, Switzerland; [8]Formerly Complete Genomics, Mountain View, CA 94043, USA; [9]Department of Oncology, Lausanne University Hospital, 1011 Lausanne, Switzerland
[10]These authors contributed equally to this work
*Correspondence: carlo.rivolta@unil.ch
http://dx.doi.org/10.1016/j.ajhg.2016.09.008.

Please cite this article in press as: Royer-Bertrand et al., Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing, The American Journal of Human Genetics (2016), http://dx.doi.org/10.1016/j.ajhg.2016.09.008

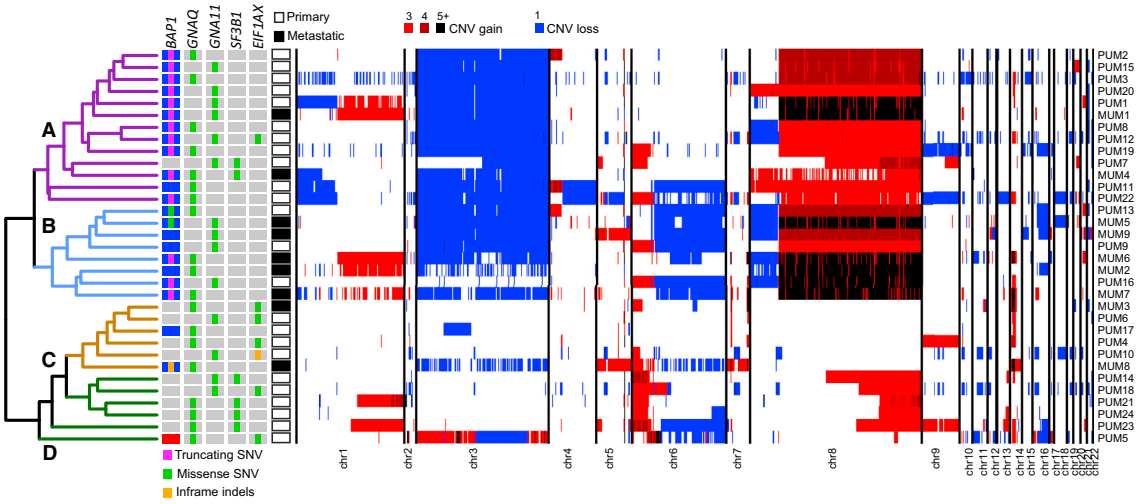


Figure 1. Unsupervised Hierarchical Clustering and Global Genetic Landscape of All Tumors Analyzed in This Study
Sample IDs are indicated on the right. CNV events are depicted in blue (copy losses) or in shades of red (copy gains) and ploidy is indicated in the legend provided at the top. SNVs in genes found to carry mutations in six or more individuals are shown on the left, with mutation classes provided within gray or blue boxes. Clustering identifies four classes—A, B, C, and D (see text)—indicated within the dendrogram.

isolation of tumor tissue from the surrounding environment. None of the 33 affected individuals received any treatment prior to primary tumor removal. Average coverage was 112× (range 102–118) for both tumor and control samples (>96% of the genome was covered 40× or more times), with minimal inter-individual variations (not shown). Mutation calls were performed genome-wide and included single-nucleotide variants (SNVs), copy-number variations (CNVs), as well as structural variations (SVs) such as chromosomal rearrangements. CNVs and SVs were assessed by comparing sequence coverage and especially de novo assembly of reads defining novel genetic junctions.[28] Data were extracted from MasterVar files and other relevant matrices by ad hoc Perl, bash, and R scripts, available upon request. Overall, we detected 37,321 SNVs (average per sample 1,166; range 576–2,131), 1,584 SVs (average per sample 50; range 13–182), and a number of CNVs corresponding to an average of 13.6% of the genome (range 0.03%–33.9%) (Table S2).

Unsupervised hierarchical clustering of our samples on CNVs revealed four major subgroups associated with mutational and metastatic status, branched two by two (Figure 1). Classes A and B (first branch) involved samples carrying chromosome 3 monosomy (by Fisher test, p value = 4.4×10^{-6}), chr 8q gain (p value = 2.8×10^{-9}), and in some instances chr 8p loss (p value = 3.0×10^{-2}). In addition, class B tumors also had loss of chr 6q (p value = 1.0×10^{-3}). Conversely, classes C and D represented more distinct subtypes with relatively few chromosomal rearrangements; class C tumor had no major aneuploidies, whereas class D reported gains of the distal part of chr 8q (p value = 2.0×10^{-3}). Seven samples presenting chr 1q gain were scattered across all classes, whereas loss of chr 1p was typical of class A tumors (p value = 5.0×10^{-3}).

Samples with monosomy of chr 3 were also associated (77% of cases) with somatic mutations in the tumor suppressor *BAP1*, lying on chr 3p21.1, in accordance with Knudson's two-hit model of tumorigenesis.[29] Indeed, *BAP1* SNVs included all kinds of somatic events, but mostly mutations leading to premature stop codons and therefore to functional protein knockout (Figure S1, Table S3). Hallmark driver mutations in the *GNAQ* and *GNA11* paralogs, encoding the components of the alpha subunit of the Gq protein heterotrimer, were present in 100% of the samples examined. They occurred in a perfectly mutually exclusive pattern and involved only four specific missenses—c.626A>T (p.Gln209Leu) (GenBank: NM_002072.3; 11 samples) and c.626A>C (p.Gln209Pro) (GenBank: NM_002072.3; 8 samples) in *GNAQ*, and c.626A>T (p.Gln209Leu) (GenBank: NM_002067.2; 13 samples) and c.626A>C (p.Gln209Pro) (GenBank: NM_002067.2; 1 sample) in *GNA11*—affecting the same functional amino acid residue and conferring oncogenic potential to this G protein.[18,19] Six tumors (18%) had missense mutations in *SF3B1* (Splicing Factor 3B, subunit 1), affecting codon 625 (5 cases) and codon 626 (1 case) (Figure S1, Table S3), a previously described hotspot region.[21] Finally, 7 other tumors had mutations impacting the first codons of *EIF1AX* (Eukaryotic Translation Initiation Factor 1A, X-Linked) (Figure S1, Table S3).[22] Mutations in *SF3B1* and *EIF1AX* seemed to occur in a mutually exclusive pattern and to be enriched with classes C and D (p value = 1.6×10^{-4}), with *SF3B1* preferentially mutated in class D. Except for one sample, *BAP1* mutations were never observed in

Please cite this article in press as: Royer-Bertrand et al., Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing, The American Journal of Human Genetics (2016), http://dx.doi.org/10.1016/j.ajhg.2016.09.008

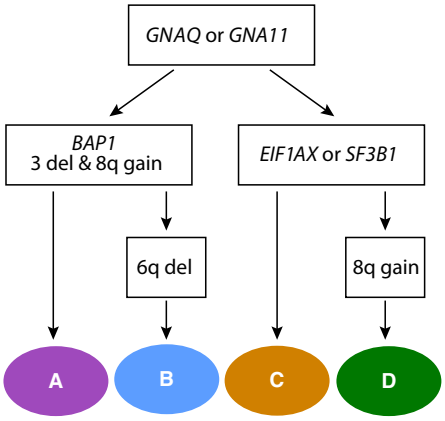


Figure 2. Inferred Somatic Events Defining Tumor Classes, as Identified by Clustering
Colors are the same as those shown in Figure 1. All steps determining branching are statistically significant.

case subjects carrying mutations in *SF3B1* or *EIF1AX* (p value = 1.4×10^{-5}), in agreement with findings from previous literature.[23,30] Also, consistent with the fact that all of the tumors analyzed harbored variants affecting either GNAQ or GNA11 Gln209, no somatic SNVs were observed in *PLCB4* or *CYSLTR2*, two genes that have been found to be mutated in a mutually exclusive pattern with respect to *GNAQ* or *GNA11* variants.[25,26]

Five genes (*TP53BP1* [MIM: 605230], *CSMD1* [MIM: 608397], *TTC28* [MIM: 615098], *DLK2*, and *KTN1* [MIM: 600381]) harbored somatic missense or truncating mutations in at least two samples, across all tumor classes (Table S3). TP53BP1 is a partner of the tumor suppressor protein p53, known to play a crucial role in maintaining genomic integrity as a mediator and effector of homologous recombination in response to double-strand breaks. This protein acts as a molecular scaffold that recruits responsive proteins, in order to repair damaged chromatin,[31] and its depletion has been associated with increased cell proliferation.[32] *CSMD1* (Cub and Sushi Multiple Domains-1) is a candidate tumor suppressor gene, the hyper-expression of which increases survival in mice with xenografted tumors.[33] Loss of *CSMD1* was detected in a large set of cancers, including head and neck, lung, breast, and skin primary tumors,[34] and associated with high tumor grade in invasive ductal breast carcinoma.[35] TTC28 (Tetratricopeptide Repeat Domain 28) is a ubiquitous protein, associated with diverse biological functions. Of note, TCC28 plays a critical role in the progress of mitosis and cytokinesis during mammalian cell cycle and its dysfunction was described as a potential component of tumorigenesis and tumor progression.[36,37] DLK2 (Delta-Like 2 homolog) is a transmembrane epidermal growth factor-like protein. It is highly homologous to DLK1, a protein that was found to be present at high levels in gliomas and involved in cell proliferation.[38] Similar to DLK1, DLK2 can bind to NOTCH1,[39] modulating the oncogenic potential of cultured melanoma cells.[40] Finally, KTN1 (Kinectin 1) is a protein of the endoplasmic reticulum membrane that interacts with kinesins.[41,42] Its role in cancer may be linked to dysregulation of cytoskeletal activity and mitosis. Two of these five genes were previously found to be mutated in UM: a missense in *TTC28* was detected in 1 out 35 samples from a WES screen,[22] and the cBioPortal repository reports a missense mutation in *KTN1* in one out of 80 tumors profiled by TCGA.[43,44]

Taken together, our clustering analysis indicates that initial events involve *GNAQ* or *GNA11* mutations, followed by a major branching determined by the functional loss of BAP1 and copy gains of chr 8q versus cases with a relatively normal chromosomal ploidy. These latter samples have conversely mutations in *EIF1AX* or in *SF3B1* (classes A and B versus C and D, respectively). In tumors with *BAP1* mutations, the long arm of chr 6 could eventually be lost, differentiating class B from class A. In tumors that are negative for *BAP1* mutations, part of chr 8q could undergo amplification, differentiating class D from class C (Figure 2).

Aggregate analyses on genetic data showed no significant differences between primary UM (PUM) and metastatic UM (MUM) samples, in terms of number of somatic coding SNVs, non-coding SNVs, CNVs, and SVs, indicating that the extent of genomic instability was here not associated with metastatic potential (Figures 3B–3E). Although singularly none of the main somatic drivers (chr 3, 6q, and 8q aneuploidies, as well as SNVs in *BAP1*, *GNAQ*, *GNA11*, *SF3B1*, *EIF1AX*) were computed as being statistically different, enrichment in PUMs versus MUMs showed very clear association trends (Figure 3A).

Remarkably, when considering specific levels of 8q amplification detectable by algorithms querying non-coding WGS data for CNVs and aneuploidies, we found a very clear association between metastatic potential and 8q ploidy of five copies or more (p value = 8.6×10^{-4}, Figure 3A). In addition, single-copy amplification of 8q (ploidy = 3) was indeed associated with primary tumors (p value = 4.7×10^{-3}, Figure 3A). Similarly, when mutational sets defining tumor sub-classes were considered, a significant association between sub-class B and metastases was identified (Figure 1, p value = 2.0×10^{-2}).

With respect to metastatic samples, the aneuploidies identified correlated well with those of 66 liver metastases from UM investigated previously, detecting chr 3 monosomy (73%), 8q gain (89%), 6q loss (64%), 1p loss (47%), 8p loss (45%), 1q gain (35%), and 16q loss (32%).[14] Similarly, the identified SNVs matched those on another study on five liver metastases.[45] Finally, the identification of a *SF3B1* mutation in one metastatic sample from our series is also in line with late-onset metastasis occurring in individuals with *SF3B1* mutations.[23]

Mutations targeting *BAP1* are one of the genetic landmarks of UM[20] and were found here to be associated with classes A and B (p value = 1.2×10^{-4}), classes that

Please cite this article in press as: Royer-Bertrand et al., Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing, The American Journal of Human Genetics (2016), http://dx.doi.org/10.1016/j.ajhg.2016.09.008

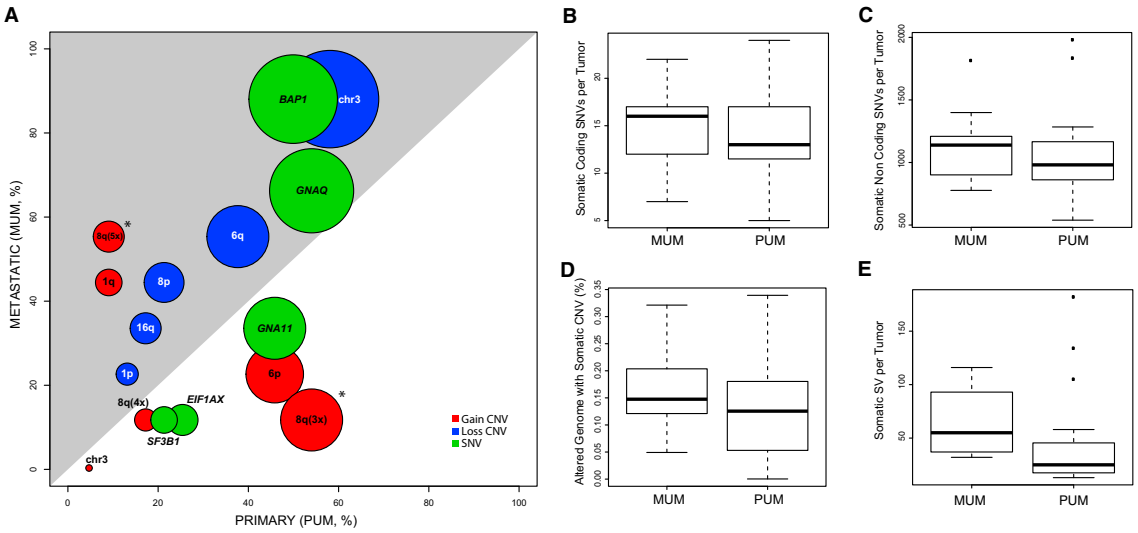


Figure 3. Genetic Features in Primary and Metastatic Tumors

(A) Overview of all major somatic events with respect to PUMs and MUMs. Each circle indicates a specific genetic event; its center corresponds to the percentage of samples carrying this feature in PUMs versus MUMs, whereas its diameter indicates the total number of such samples. Asterisks indicate statistical significance. The gray area depicts the surface of the plot for which there is an enrichment in MUMs.

(B–E) Boxplots of different types of genetic alterations, at the genome-wide scale.

are in fact defined mostly by the presence of chr 3 monosomy. To test for functional inactivation of the BAP1 protein, we assessed its nuclear staining in histological sections of all tumors (Figure 4). Of the 33 samples (60%), 20 displayed loss of nuclear localization (Figure 4). Of these, 17 (85%) presented chr 3 monosomy and a coding SNV (a truncating SNV in 14 cases and a missense or an in-frame deletion in 3 cases), accounting for loss of heterozygosity and protein delocalization.

The number of somatic SNVs involving coding and non-coding regions was strikingly low (Figure S2). Globally, the average load of coding mutations was 0.24/Mb (range 0.08–0.42/Mb, Table S4), one of the lowest detected so far in tumors. Comparison with other cancer types revealed that UM mutational load for coding regions was closer to that of pediatric tumors such as rhabdoid tumors, medulloblastoma (MIM: 155255), neuroblastoma, etc.,[46] rather than that of adult cancers (Figure S3). Pediatric tumors typically develop over a shorter period than most adult malignancies, frequently harbor few driver mutations, and may thereby have fewer sources of heterogeneity, facilitating the assessment of both the genetic and epigenetic determinants underlying their pathogenesis. Our data seem to suggest that, similarly to pediatric tumors, UM development may rely more on epigenetic drivers of transformation and tumor progression, rather than the classical accumulation of genetic events observed in the vast majority of adult malignancies.

The number of non-coding SNVs was also relatively low (736 per tumor on average, range 371–1,347) and mostly proportional to the number of coding SNVs (17 per tumor on average, range 7–28) (Figure S2), confirming that, compared to both cutaneous and conjunctival melanomas, which also originate from melanocytes, UM follows a different oncogenic pathway, characterized by significantly fewer mutations.[47,48] In addition, we failed to identify any statistically relevant non-coding SNVs for tumor-specific sites that were present in four samples or more, suggesting the absence of common regulatory variants in the landscape of these tumors, at least in our cohort.

Another difference between UM and cutaneous and conjunctival melanomas involved its mutational spectrum (Figure 5). Analysis of all coding and non-coding somatic single-nucleotide substitutions (SNSs) from our series showed the clear absence of an UV-induced mutation signature. This particular spectrum results from sunlight-driven formation of pyrimidine dimers on the DNA[49] and is found in both cutaneous and conjunctival melanomas (Figure 6A).[47,48] Direct analysis of genes known to be involved in cutaneous melanoma, such as *BRAF* (MIM: 164757), *NRAS* (MIM: 164790), and *NF1* (MIM: 162200) revealed no somatic mutations in UM, supporting again the notion that uveal and cutaneous melanomas have a different molecular etiology.[50,51] Conversely, the UM mutational spectrum was remarkably similar across all PUMs and MUMs and resembled that of apparently unrelated tumors, such as clear cell renal carcinoma, thyroid tumor, and glioblastoma (Figure 6). Notably, despite a different cellular origin, UM shares with these tumors recurrent genetic modifications; *BAP1* mutations and chr

Please cite this article in press as: Royer-Bertrand et al., Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing, The American Journal of Human Genetics (2016), http://dx.doi.org/10.1016/j.ajhg.2016.09.008

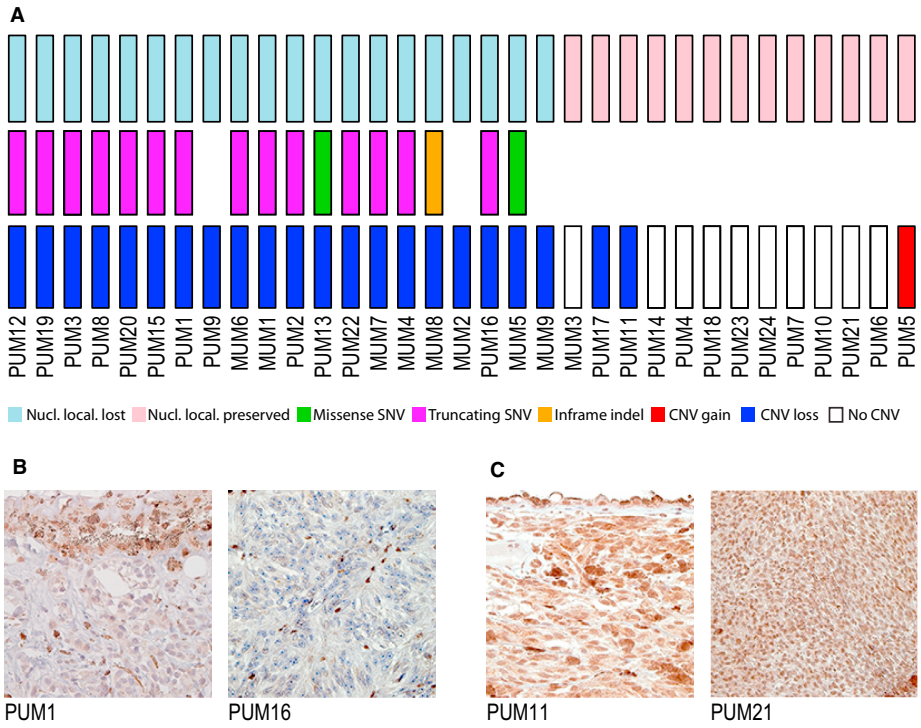


Figure 4. Landscape of Genetic Alterations Involving *BAP1* and Immunohistochemistry of BAP1 Protein
(A) Samples are ordered with respect to absence/presence of BAP1 nuclear localization, indicating loss or preservation of protein function, respectively. In general, loss of BAP1 function correlates with presence of a somatic SNV and loss of heterozygosity.
(B and C) Representative micrographs of paraffin-embedded UM samples showing absence (B) and presence (C) of BAP1 protein. Magnification: 252×.

3 monosomy are frequently seen in clear cell renal carcinoma,[52] while hotspot mutations in the first codons of *EIF1AX* are recurrent in papillary thyroid carcinoma.[53]

Analysis of all specific SNS types along with composition of the flanking bases allowed us to determine specific mutational signatures for UM, according to the classification of Alexandrov et al.[54] Our samples appeared to be enriched for signatures 12 or 16 (55% of the score), signature 1B (25%), and signature 6 (20%) (Figure 5B). Signature 1B corresponds to a rather ubiquitous pattern in cancer, resulting from the spontaneous deamination of 5-methylcytosine, which in turn is thought to correlate with the process of aging.[54,55] Conversely, signatures 12/16 and 6 are associated with defects in nucleotide excision repair and DNA mismatch repair, respectively.

A more global approach, considering the intersection between the SNVs detected in our series and the most frequently mutated genes in cancer (TCGA PANCAN list),[56] also revealed a minimal overlap, limited to *BAP1* and *SF3B1* (Figure S4).

A non-negligible number of structural variants (SVs) such as deletions, duplications, inversions, or inter- and intra-chromosomal rearrangements were also observed (Figure S5). Only a few of these events were recurrent, indicating the absence of major common drivers constituted by genetic events involving large parts of the genome. Among these, however, there were three inter-chromosomal events that were present in at least two individuals (Figure 7). Three samples (PUM20, PUM18, and PUM5) had a translocation involving chr 6 and chr 8 disrupting *UBE2W* (MIM: 614277) and *MYO6* (MIM: 600970) for the two first samples, respectively. The third event occurred in intergenic regions. Translocations between chr 13 and chr 17 (no genes involved) were also present in MUM9 and PUM5 and between chr 3 and chr 12 in PUM17 (no genes involved) and in PUM5 (disrupting *KDM2B* [MIM: 609078]) (Table S5). Although these translocations impacted roughly the same genomic areas, highlighting possible hotspot regions in UM genome, they neither targeted the same genes nor defined a specific tumor sub-category. Of note, one individual (PUM5) appeared to harbor a higher number of inter-chromosomal events and SVs than the average value of the other cases (Figure S5). Notably, this individual was also an outlier of our clustering analysis (Figure 1). However, neither the medical history nor tumor pathology revealed any uncommon feature, compared to the rest of the cohort.

Amplification of chr 8q is a well-known and important feature of UM.[17,57,58] Levels of chr 8q amplification seem

Please cite this article in press as: Royer-Bertrand et al., Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing, The American Journal of Human Genetics (2016), http://dx.doi.org/10.1016/j.ajhg.2016.09.008

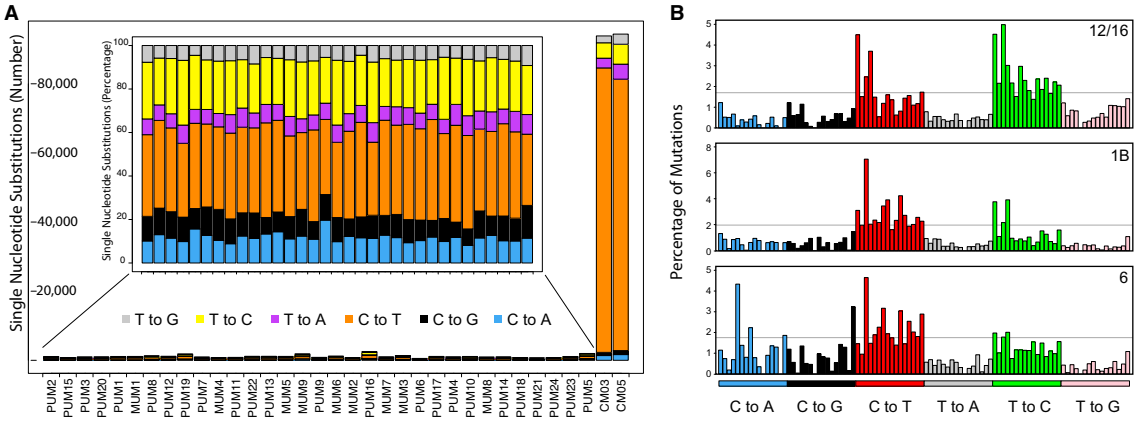


Figure 5. Mutational Signature of Our Samples, for SNSs, Genome-wide
(A) Main graph: comparison of mutational load of the UM samples studied with respect to two melanomas of the conjunctiva (CM), sequenced and processed according to the same methods.[48] The number of mutations is radically different in UM versus CM. Inset: mutational spectrum of each UM sample, in percentage, showing a relatively homogeneous pattern.
(B) Results of the analysis of mutational events according to the methods and the classification proposed by Alexandrov et al.[54,55] Three main signatures are detected in our samples, evocative of Alexandrov's signatures 12/16, 1B, and 6. The different peaks indicate specific genetic contexts of the altered nucleotide and are ordered according to the original article.[54]

to define prognostic status and metastatic potential in UM and differentiate class C from D (Figures 1 and 2). However, the molecular bases for this phenomenon are not known. One possible explanation is that the amplification is driven by the *MYC* oncogene (MIM: 190080), which lies in this region.[59] By comparing the pattern of chr 8q amplification in our samples, we determined the minimal region of overlap, involving a 2.3 Mb fragment toward its telomeric site (chr 8: 126,404,000–128,682,000).

Surprisingly, this region was very close to *MYC* (chr 8: 128,748,314–128,753,680) but did not include it. Conversely, it harbored six other genes (*POU5F1B*

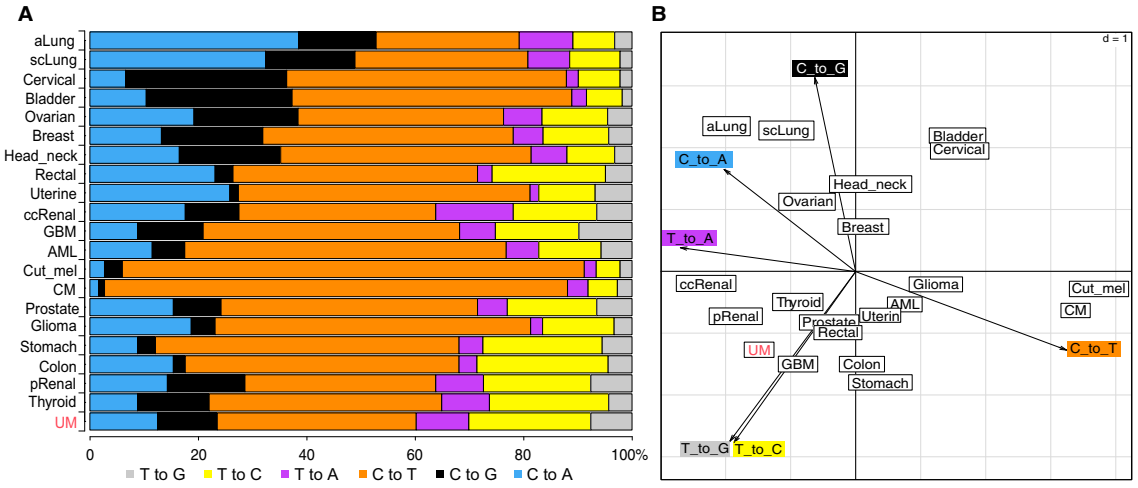


Figure 6. Analysis of the Mutational Spectrum in Our Samples versus Other Cancer Types, in Coding Regions Only
(A) The spectrum from UM is dissimilar from those from cutaneous and conjunctival melanomas, which are dominated by UV light-induced events (C to T transitions) and is conversely closer to that of thyroid and renal papillary cancer.
(B) Principal-component analysis (PCA) of the same data, showing the relatedness of UM with a few cancers but, again, not with other melanomas. Dimensions of the PCA are indicated by the arrowed axes. Primary data other than UM are from previously published sources.[48,62] Cancer types are indicated by the following abbreviations. aLung: lung, adenocarcinoma; scLung: lung, squamous cell; pRenal: renal, papillary; ccRenal: renal, clear cell; GBM: glioblastoma multiforme; AML: acute myeloid leukemia; Cut_mel: cutaneous melanoma; CM: conjunctival melanoma.

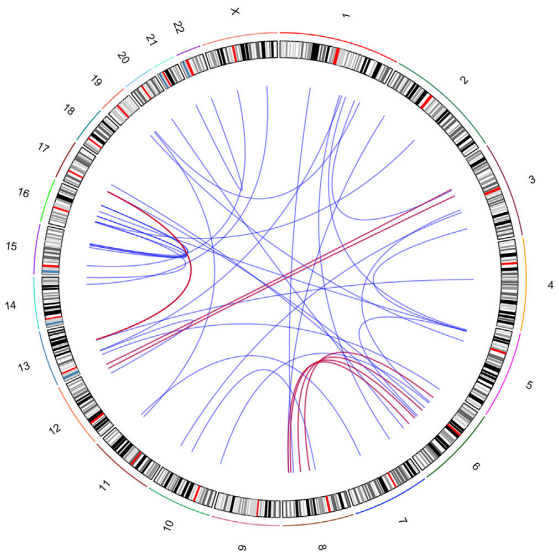


Figure 7. Circos Plot of All Somatic Interchromosomal Events, in All UM Samples
Red lines indicate events involving the same chromosomal regions in more than two individuals.

[MIM:615739], *FAM84B* [MIM: 609483], *TRIB1* [MIM: 609461], *LOC100130231/LINC00861*, *LOC100507056/CCAT1*, and *LOC727677/CASC8*). The most interesting of them was *POU5F1B* (POU Class 5 Homeobox 1B), a pseudogene of the POU5F1/OCT4 family, recently shown to be involved in prostatic and gastric cancer.[60,61] Real-time quantitative PCR experiments showed that only *POU5F1B*, *TRIB1*, *LINC00861*, and *CCAT1* were expressed in our UM samples, but no statistically significant correlation between their expression levels and 8q amplification or tumor class could be detected. The same held true for the *MYC* transcript, suggesting that none of these genes play a key role in UM pathogenesis (Figures S6 and S7).

By using a WGS-based, untargeted approach to investigate the genetic components of UM, we had the unique opportunity of assessing its genomic landscape as a whole, from single-nucleotide variants to interchromosomal rearrangements, providing the basis for future functional studies that go beyond the scope of our analysis. The global picture emerging from our work indicates that, genetically, UM is a relatively atypical tumor, mostly in virtue of the paucity of somatic events that characterize it. Driver mutations are very few and are confined to a relatively low number of genes, such as *BAP1*, *GNAQ*, and *GNA11*. Other genes, including those that were identified in this study, may have a role in tumorigenesis, but they are nonetheless present in a small fraction of the samples studied. Conversely, larger events such as extended copy-number and structural variations seem to shape UM's genome in a much more relevant way, possibly determining tumor progression and fate. Taken together, our results point to a critical role for non-canonical mechanisms of cellular transformation in UM development, where chromosomal rearrangements and non-coding SNVs potentially affecting distal regulatory elements may collaborate in the establishment of a permissive oncogenic landscape.

Supplemental Data

Supplemental Data include seven figures and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.09.008.

Acknowledgments

This work was supported by the "Fond'action contre le cancer" foundation, Switzerland, the Swiss National Science Foundation (Div III grant# 156260 to C.R. and SNF Professorship grant# PP00P3-157468/1 to N.R.), and the MEDIC Foundation. We wish to thank Complete Genomics (Mountain View, CA) for providing the whole-genome sequencing and Patricia Martin for her invaluable help. We thank as well the Lausanne Institutional Biobank and Danielle Minaidis for help in collecting biological material from affected individuals.

Received: July 22, 2016
Accepted: September 15, 2016
Published: October 13, 2016

Web Resources

GenBank, http://www.ncbi.nlm.nih.gov/genbank/
OMIM, http://www.omim.org/
Primary dataset, https://drive.switch.ch/index.php/s/7GjD2zQEoPOyaj9
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq

References

1. Singh, A.D., Turell, M.E., and Topham, A.K. (2011). Uveal melanoma: trends in incidence, treatment, and survival. Ophthalmology *118*, 1881–1885.
2. Bergman, L., Seregard, S., Nilsson, B., Ringborg, U., Lundell, G., and Ragnarsson-Olding, B. (2002). Incidence of uveal melanoma in Sweden from 1960 to 1998. Invest. Ophthalmol. Vis. Sci. *43*, 2579–2583.
3. Luke, J.J., Triozzi, P.L., McKenna, K.C., Van Meir, E.G., Gershenwald, J.E., Bastian, B.C., Gutkind, J.S., Bowcock, A.M., Streicher, H.Z., Patel, P.M., et al. (2015). Biology of advanced uveal melanoma and next steps for clinical therapeutics. Pigment Cell Melanoma Res. *28*, 135–147.
4. Woodman, S.E. (2012). Metastatic uveal melanoma: biology and emerging treatments. Cancer J. *18*, 148–152.
5. Shields, J.A., and Shields, C.L. (2015). Management of posterior uveal melanoma: past, present, and future: the 2014 Charles L. Schepens lecture. Ophthalmology *122*, 414–428.
6. Park, S.J., Oh, C.M., Kim, B.W., Woo, S.J., Cho, H., and Park, K.H. (2015). Nationwide incidence of ocular melanoma in South Korea by using the national cancer registry database (1999-2011). Invest. Ophthalmol. Vis. Sci. *56*, 4719–4724.
7. Hu, D.N., Yu, G.P., McCormick, S.A., Schneider, S., and Finger, P.T. (2005). Population-based incidence of uveal melanoma

Please cite this article in press as: Royer-Bertrand et al., Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing, The American Journal of Human Genetics (2016), http://dx.doi.org/10.1016/j.ajhg.2016.09.008

in various races and ethnic groups. Am. J. Ophthalmol. *140*, 612–617.

8. Seddon, J.M., Gragoudas, E.S., Glynn, R.J., Egan, K.M., Albert, D.M., and Blitzer, P.H. (1990). Host factors, UV radiation, and risk of uveal melanoma. A case-control study. Arch. Ophthalmol. *108*, 1274–1280.
9. Holly, E.A., Aston, D.A., Ahn, D.K., and Smith, A.H. (1996). Intraocular melanoma linked to occupations and chemical exposures. Epidemiology *7*, 55–61.
10. Vajdic, C.M., Kricker, A., Giblin, M., McKenzie, J., Aitken, J., Giles, G.G., and Armstrong, B.K. (2002). Sun exposure predicts risk of ocular melanoma in Australia. Int. J. Cancer *101*, 175–182.
11. Schmidt-Pokrzywniak, A., Jöckel, K.H., Bornfeld, N., Sauerwein, W., and Stang, A. (2009). Positive interaction between light iris color and ultraviolet radiation in relation to the risk of uveal melanoma: a case-control study. Ophthalmology *116*, 340–348.
12. Furney, S.J., Pedersen, M., Gentien, D., Dumont, A.G., Rapinat, A., Desjardins, L., Turajlic, S., Piperno-Neumann, S., de la Grange, P., Roman-Roman, S., et al. (2013). SF3B1 mutations are associated with alternative splicing in uveal melanoma. Cancer Discov. *3*, 1122–1129.
13. Prescher, G., Bornfeld, N., Hirche, H., Horsthemke, B., Jöckel, K.H., and Becher, R. (1996). Prognostic implications of monosomy 3 in uveal melanoma. Lancet *347*, 1222–1225.
14. Trolet, J., Hupé, P., Huon, I., Lebigot, I., Decraene, C., Delattre, O., Sastre-Garau, X., Saule, S., Thiéry, J.P., Plancher, C., et al. (2009). Genomic profiling and identification of high-risk uveal melanoma by array CGH analysis of primary tumors and liver metastases. Invest. Ophthalmol. Vis. Sci. *50*, 2572–2580.
15. Cassoux, N., Rodrigues, M.J., Plancher, C., Asselain, B., Levy-Gabriel, C., Lumbroso-Le Rouic, L., Piperno-Neumann, S., Dendale, R., Sastre, X., Desjardins, L., and Couturier, J. (2014). Genome-wide profiling is a clinically relevant and affordable prognostic test in posterior uveal melanoma. Br. J. Ophthalmol. *98*, 769–774.
16. van Essen, T.H., van Pelt, S.I., Versluis, M., Bronkhorst, I.H., van Duinen, S.G., Marinkovic, M., Kroes, W.G., Ruivenkamp, C.A., Shukla, S., de Klein, A., et al. (2014). Prognostic parameters in uveal melanoma and their association with BAP1 expression. Br. J. Ophthalmol. *98*, 1738–1743.
17. Damato, B., Dopierala, J.A., and Coupland, S.E. (2010). Genotypic profiling of 452 choroidal melanomas with multiplex ligation-dependent probe amplification. Clin. Cancer Res. *16*, 6083–6092.
18. Van Raamsdonk, C.D., Bezrookove, V., Green, G., Bauer, J., Gaugler, L., O'Brien, J.M., Simpson, E.M., Barsh, G.S., and Bastian, B.C. (2009). Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. Nature *457*, 599–602.
19. Van Raamsdonk, C.D., Griewank, K.G., Crosby, M.B., Garrido, M.C., Vemula, S., Wiesner, T., Obenauf, A.C., Wackernagel, W., Green, G., Bouvier, N., et al. (2010). Mutations in GNA11 in uveal melanoma. N. Engl. J. Med. *363*, 2191–2199.
20. Harbour, J.W., Onken, M.D., Roberson, E.D., Duan, S., Cao, L., Worley, L.A., Council, M.L., Matatall, K.A., Helms, C., and Bowcock, A.M. (2010). Frequent mutation of BAP1 in metastasizing uveal melanomas. Science *330*, 1410–1413.
21. Harbour, J.W., Roberson, E.D., Anbunathan, H., Onken, M.D., Worley, L.A., and Bowcock, A.M. (2013). Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. Nat. Genet. *45*, 133–135.
22. Martin, M., Maßhöfer, L., Temming, P., Rahmann, S., Metz, C., Bornfeld, N., van de Nes, J., Klein-Hitpass, L., Hinnebusch, A.G., Horsthemke, B., et al. (2013). Exome sequencing identifies recurrent somatic mutations in EIF1AX and SF3B1 in uveal melanoma with disomy 3. Nat. Genet. *45*, 933–936.
23. Yavuzyigitoglu, S., Koopmans, A.E., Verdijk, R.M., Vaarwater, J., Eussen, B., van Bodegom, A., Paridaens, D., Kiliç, E., and de Klein, A.; Rotterdam Ocular Melanoma Study Group (2016). Uveal melanomas with SF3B1 mutations: a distinct subclass associated with late-onset metastases. Ophthalmology *123*, 1118–1128.
24. Cimino, P.J., Kung, Y., Warrick, J.I., Chang, S.H., and Keene, C.D. (2016). Mutational status of IDH1 in uveal melanoma. Exp. Mol. Pathol. *100*, 476–481.
25. Johansson, P., Aoude, L.G., Wadt, K., Glasson, W.J., Warrier, S.K., Hewitt, A.W., Kiilgaard, J.F., Heegaard, S., Isaacs, T., Franchina, M., et al. (2016). Deep sequencing of uveal melanoma identifies a recurrent mutation in PLCB4. Oncotarget *7*, 4624–4631.
26. Moore, A.R., Ceraudo, E., Sher, J.J., Guan, Y., Shoushtari, A.N., Chang, M.T., Zhang, J.Q., Walczak, E.G., Kazmi, M.A., Taylor, B.S., et al. (2016). Recurrent activating mutations of G-protein-coupled receptor CYSLTR2 in uveal melanoma. Nat. Genet. *48*, 675–680.
27. Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., et al. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science *327*, 78–81.
28. Carnevali, P., Baccash, J., Halpern, A.L., Nazarenko, I., Nilsen, G.B., Pant, K.P., Ebert, J.C., Brownley, A., Morenzoni, M., Karpinchyk, V., et al. (2012). Computational techniques for human genome resequencing using mated gapped reads. J. Comput. Biol. *19*, 279–292.
29. Knudson, A.G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc. Natl. Acad. Sci. USA *68*, 820–823.
30. Decatur, C.L., Ong, E., Garg, N., Anbunathan, H., Bowcock, A.M., Field, M.G., and Harbour, J.W. (2016). Driver mutations in uveal melanoma: associations with gene expression profile and patient outcomes. JAMA Ophthalmol. *134*, 728–733.
31. Panier, S., and Boulton, S.J. (2014). Double-strand break repair: 53BP1 comes into focus. Nat. Rev. Mol. Cell Biol. *15*, 7–18.
32. Bi, J., Huang, A., Liu, T., Zhang, T., and Ma, H. (2015). Expression of DNA damage checkpoint 53BP1 is correlated with prognosis, cell proliferation and apoptosis in colorectal cancer. Int. J. Clin. Exp. Pathol. *8*, 6070–6082.
33. Tang, M.R., Wang, Y.X., Guo, S., Han, S.Y., and Wang, D. (2012). CSMD1 exhibits antitumor activity in A375 melanoma cells through activation of the Smad pathway. Apoptosis *17*, 927–937.
34. Ma, C., Quesnelle, K.M., Sparano, A., Rao, S., Park, M.S., Cohen, M.A., Wang, Y., Samanta, M., Kumar, M.S., Aziz, M.U., et al. (2009). Characterization CSMD1 in a large set of primary lung, head and neck, breast and skin cancer tissues. Cancer Biol. Ther. *8*, 907–916.
35. Kamal, M., Shaaban, A.M., Zhang, L., Walker, C., Gray, S., Thakker, N., Toomes, C., Speirs, V., and Bell, S.M. (2010). Loss of CSMD1 expression is associated with high tumour grade and poor survival in invasive ductal breast carcinoma. Breast Cancer Res. Treat. *121*, 555–563.
36. Izumiyama, T., Minoshima, S., Yoshida, T., and Shimizu, N. (2012). A novel big protein TPRBK possessing 25 units of

Please cite this article in press as: Royer-Bertrand et al., Comprehensive Genetic Landscape of Uveal Melanoma by Whole-Genome Sequencing, The American Journal of Human Genetics (2016), http://dx.doi.org/10.1016/j.ajhg.2016.09.008

TPR motif is essential for the progress of mitosis and cytokinesis. Gene *511*, 202–217.
37. Pitkänen, E., Cajuso, T., Katainen, R., Kaasinen, E., Välimäki, N., Palin, K., Taipale, J., Aaltonen, L.A., and Kilpivaara, O. (2014). Frequent L1 retrotranspositions originating from TTC28 in colorectal cancer. Oncotarget *5*, 853–859.
38. Yin, D., Xie, D., Sakajiri, S., Miller, C.W., Zhu, H., Popoviciu, M.L., Said, J.W., Black, K.L., and Koeffler, H.P. (2006). DLK1: increased expression in gliomas and associated with oncogenic activities. Oncogene *25*, 1852–1861.
39. Sánchez-Solana, B., Nueda, M.L., Ruvira, M.D., Ruiz-Hidalgo, M.J., Monsalve, E.M., Rivero, S., García-Ramírez, J.J., Díaz-Guerra, M.J., Baladrón, V., and Laborda, J. (2011). The EGF-like proteins DLK1 and DLK2 function as inhibitory non-canonical ligands of NOTCH1 receptor that modulate each other's activities. Biochim. Biophys. Acta *1813*, 1153–1164.
40. Nueda, M.L., Naranjo, A.I., Baladrón, V., and Laborda, J. (2014). The proteins DLK1 and DLK2 modulate NOTCH1-dependent proliferation and oncogenic potential of human SK-MEL-2 melanoma cells. Biochim. Biophys. Acta *1843*, 2674–2684.
41. Fütterer, A., Kruppa, G., Krämer, B., Lemke, H., and Krönke, M. (1995). Molecular cloning and characterization of human kinectin. Mol. Biol. Cell *6*, 161–170.
42. Machleidt, T., Geller, P., Schwandner, R., Scherer, G., and Krönke, M. (1998). Caspase 7-induced cleavage of kinectin in apoptotic cells. FEBS Lett. *436*, 51–54.
43. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404.
44. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1.
45. Luscan, A., Just, P.A., Briand, A., Burin des Roziers, C., Goussard, P., Nitschké, P., Vidaud, M., Avril, M.F., Terris, B., and Pasmant, E. (2015). Uveal melanoma hepatic metastases mutation spectrum analysis using targeted next-generation sequencing of 400 cancer genes. Br. J. Ophthalmol. *99*, 437–439.
46. Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature *499*, 214–218.
47. Cancer Genome Atlas, N.; Cancer Genome Atlas Network (2015). Genomic classification of cutaneous melanoma. Cell *161*, 1681–1696.
48. Rivolta, C., Royer-Bertrand, B., Rimoldi, D., Schalenbourg, A., Zografos, L., Leyvraz, S., and Moulin, A. (2016). UV light signature in conjunctival melanoma; not only skin should be protected from solar radiation. J. Hum. Genet. *61*, 361–362.
49. Ikehata, H., and Ono, T. (2011). The mechanisms of UV mutagenesis. J. Radiat. Res. (Tokyo) *52*, 115–125.
50. Foster, W.J., Fuller, C.E., Perry, A., and Harbour, J.W. (2003). Status of the NF1 tumor suppressor locus in uveal melanoma. Arch. Ophthalmol. *121*, 1311–1315.
51. Cruz, F., 3rd, Rubin, B.P., Wilson, D., Town, A., Schroeder, A., Haley, A., Bainbridge, T., Heinrich, M.C., and Corless, C.L. (2003). Absence of BRAF and NRAS mutations in uveal melanoma. Cancer Res. *63*, 5761–5766.
52. Cancer Genome Atlas Research Network (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature *499*, 43–49.
53. Cancer Genome Atlas Research Network (2014). Integrated genomic characterization of papillary thyroid carcinoma. Cell *159*, 676–690.
54. Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.
55. Kennedy, S.R., Salk, J.J., Schmitt, M.W., and Loeb, L.A. (2013). Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. PLoS Genet. *9*, e1003794.
56. Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Kandoth, C., Reimand, J., Lawrence, M.S., Getz, G., Bader, G.D., Ding, L., and Lopez-Bigas, N. (2013). Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci. Rep. *3*, 2650.
57. Ehlers, J.P., Worley, L., Onken, M.D., and Harbour, J.W. (2008). Integrative genomic analysis of aneuploidy in uveal melanoma. Clin. Cancer Res. *14*, 115–122.
58. Sisley, K., Rennie, I.G., Parsons, M.A., Jacques, R., Hammond, D.W., Bell, S.M., Potter, A.M., and Rees, R.C. (1997). Abnormalities of chromosomes 3 and 8 in posterior uveal melanoma correlate with prognosis. Genes Chromosomes Cancer *19*, 22–28.
59. Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.
60. Hayashi, H., Arao, T., Togashi, Y., Kato, H., Fujita, Y., De Velasco, M.A., Kimura, H., Matsumoto, K., Tanaka, K., Okamoto, I., et al. (2015). The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. Oncogene *34*, 199–208.
61. Breyer, J.P., Dorset, D.C., Clark, T.A., Bradley, K.M., Wahlfors, T.A., McReynolds, K.M., Maynard, W.H., Chang, S.S., Cookson, M.S., Smith, J.A., et al. (2014). An expressed retrogene of the master embryonic stem cell gene POU5F1 is associated with prostate cancer susceptibility. Am. J. Hum. Genet. *94*, 395–404.
62. Burns, M.B., Temiz, N.A., and Harris, R.S. (2013). Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat. Genet. *45*, 977–983.

5.3 Rare inherited Mendelian disorders

5.3.1 HSPA9

Mutations in the heat-shock protein A9 (HSPA9) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia

Own contribution: Analyzed the WES of the three patients and identified the new disease-causing gene.

SCIENTIFIC REPORTS

OPEN

Received: 09 September 2015
Accepted: 26 October 2015
Published: 24 November 2015

Mutations in the heat-shock protein A9 (*HSPA9*) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia

Beryl Royer-Bertrand[1,2], Silvia Castillo-Taucher[3], Rodrigo Moreno-Salinas[4], Tae-Joon Cho[5], Jong-Hee Chae[6], Murim Choi[7], Ok-Hwa Kim[8], Esra Dikoglu[2], Belinda Campos-Xavier[2], Enrico Girardi[9], Giulio Superti-Furga[9], Luisa Bonafé[2], Carlo Rivolta[1], Sheila Unger[10] & Andrea Superti-Furga[2,11]

We and others have reported mutations in *LONP1*, a gene coding for a mitochondrial chaperone and protease, as the cause of the human CODAS (cerebral, ocular, dental, auricular and skeletal) syndrome (MIM 600373). Here, we delineate a similar but distinct condition that shares the epiphyseal, vertebral and ocular changes of CODAS but also included severe microtia, nasal hypoplasia, and other malformations, and for which we propose the name of EVEN-PLUS syndrome for epiphyseal, vertebral, ear, nose, plus associated findings. In three individuals from two families, no mutation in *LONP1* was found; instead, we found biallelic mutations in *HSPA9*, the gene that codes for mHSP70/mortalin, another highly conserved mitochondrial chaperone protein essential in mitochondrial protein import, folding, and degradation. The functional relationship between LONP1 and HSPA9 in mitochondrial protein chaperoning and the overlapping phenotypes of CODAS and EVEN-PLUS delineate a family of "mitochondrial chaperonopathies" and point to an unexplored role of mitochondrial chaperones in human embryonic morphogenesis.

Recently, we and others reported on the identification of mutations in *LONP1* as the cause of the human genetic disorder, CODAS (cerebral, ocular, dental, auricular and skeletal) syndrome (MIM 600373)[1,2]. *LONP1* codes for a phylogenetically conserved protein of the mitochondrial matrix that has both

[1]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. [2]Centre for Molecular Diseases, Department of Pediatrics, Lausanne University Hospital (CHUV), Lausanne, Switzerland. [3]Sección Genética, Hospital Clínico Universidad de Chile, and Sección Citogenética, Laboratorio, Clínica Alemana de Santiago, Santiago, Chile. [4]Unidad de Genética, Hospital Regional Rancagua, Rancagua, Chile; and ICBM, Facultad de Medicina, Universidad de Chile, Santiago, Chile. [5]Division of Pediatric Orthopaedics, Seoul National University Children's Hospital, Seoul, Republic of Korea. [6]Department of Pediatrics, Pediatric Clinical Neuroscience Center, Seoul National University Children's Hospital, Seoul, Republic of Korea. [7]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea. [8]Department of Radiology, Woorisoa Children's Hospital, Saemalro, Guro-gu, Seoul 08291, Republic of Korea. [9]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, 1090 Vienna, Austria. [10]Medical Genetics Service, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland. [11]Department of Pediatrics and Pediatric Surgery, University of Lausanne and Lausanne University Hospital (CHUV), Lausanne, Switzerland. Correspondence and requests for materials should be addressed to A.S.-F. (email: asuperti@unil.ch)

Figure 1. Photographs of patients 1 (A,B), 2 (C,D) and 3 (E,F). Common features include hypoplasia of the midface and of the nasal bones giving a flat nose, triangular nares, arched eyebrows with synophrys, and hypoplastic-dysplastic external ears. Panel G and H show the flat nose with triangular nares in patients 3 and 2, respectively, with a single central incisor in patient 3; and panel I shows the lesion of congenital aplasia cutis on the vertex of patient 3.

chaperone and protease activity. Subsequently, we have identified patients with a condition that shares the skeletal features of the CODAS syndrome but includes a distinct craniofacial dysmorphism and cardiac and intestinal malformations.

A review of the literature suggests that one sporadic patient reported as having CODAS[3], two sibs reported as having an EVE (epiphyseal-vertebral-ear) syndrome[4], and a further sporadic patient reported as having caudal regression with anal atresia and spondylo-epiphyseal dysplasia[5] may have had the disorder we describe here, for which we propose the name of EVEN-PLUS syndrome for epiphyseal, vertebral, ear, nose, plus associated findings.

We report here the identification of recessive mutations in the *HSPA9* gene in our three patients with this syndrome. The *HSPA9* gene codes for a mitochondrial chaperone that has been implicated in different physiologic processes, and thus has been known under several different names (heat-shock 70 kda protein 9 (HSPA9, HSPA9B), mortalin and mortalin 2 (MOT/MOT2), 75 KDa glucose-regulated protein (GRP75), among others)[6–8]. Notably, the HSPA9 protein participates in, and is necessary for, the proteolytic activity of LONP1[9–12].

Clinical Reports

Patient 1. This girl was the first child of a non-consanguineous couple of Korean origin. Upon specific questioning, there was no history of Parkinson disease in parents or grandparents. She was born at term (39 weeks) with a length of 38 cm and a weight of 2.2 kg. Her face showed midface hypoplasia with markedly hypoplastic nasal bones, giving her a flat nose with nares that were triangular in shape (Fig. 1). She had arched eyebrows and synophrys. Her external ears were markedly small and poorly formed (Fig. 1), while the ear canal was present. She was noted to have anal atresia that was subsequently surgically corrected. An atrioseptal defect (ASD) was present at birth but repeat ultrasound at age 20 months showed that it had closed spontaneously. At age 16 months, her developmental quotient was approximately 80; a cerebral MRI was normal, an abdominal ultrasound examination did not show abnormalities of kidneys or urinary tract, an ophthalmologic examination was normal (specifically, no cataracts were observed), and a CGH array study gave normal results. Skeletal radiographs obtained at age 16 months at 4 yrs showed lateral vertebral clefts, dysplasia of the proximal femurs and acetabula, "bifid" distal femurs and marked epiphyseal dysplasia at her knees (Fig. 2). These skeletal changes were considered typical of CODAS syndrome. However, no variants in the *LONP1* gene were found.

Patient 2. This girl was born to a consanguineous couple of Chilean origin; the parents are uncle once removed and niece. Family history was negative for Parkinson disease. Short long bones were noted prenatally. She was born at 38 weeks' gestation with a length of 39 cm (markedly below the 3rd percentile), a

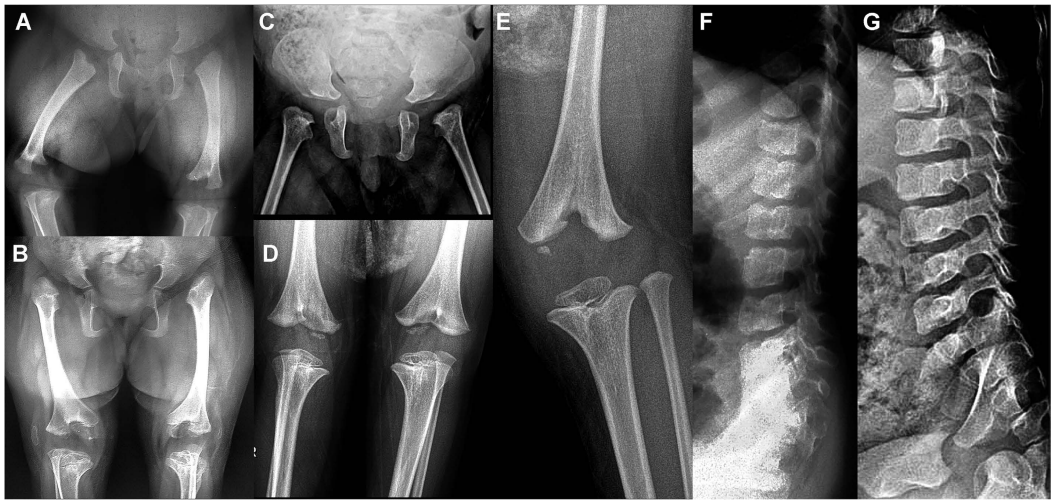


Figure 2. Skeletal-radiographic features. Panels A and B: patient 2 at birth (**A**) and at age 5 years (**B**) showing underossified pubic bones; bilateral dysplasia of the femoral heads at birth resulting in hip dislocation at age 5 yrs; and a "bifid" appearance of the distal femur with epiphyseal delay at birth, with dysplastic epiphyses that are "socketed" in the bifid femur at age 5 yrs; and small, laterally dislocated patellae. Panels C, D and E: Patient 1 at birth (**C**) with dysplastic femoral heads and at age 4 years (**D,E**) showing the bifid distal femur and the markedly dysplastic distal femoral epiphyses. The proximal tibial epiphyses are too small but less severely affected. Panels F and G: lateral lumbar spine of patient 3 at age 3 yrs (**F**) and of patient 1 at age 4 yrs (**G**) showing remnants of coronal clefts of the vertebral bodies. The clefts are prominent at birth and gradually disappear as ossification progresses.

weight of 2.8 kg, and a head circumference of 33,5 cm (slightly below the 10th percentile). Her karyotype was normal in blood cells. When seen in the genetics clinic at age 5 years, her weight was 12,5 kg, and she had severe short stature (markedly below the 3rd percentile for Chilean children) with a height of 80 cm; her head circumference 47,8 cm (below the 3rd percentile). She had severe bilateral microtia with apparently normal external ear duct, arched eyebrows with mild synophris, and a very flat nose with nares that were triangular in shape (Fig. 1). Her cranial fontanels were still open, she had two lateral hair whorls and a small area of aplasia cutis on her cranium. Her limbs looked short. An echocardiography showed a small ASD. Cerebral, abdominal and renal ultrasound scans gave normal results. Her language development was appropriate and she attended kindergarten. Radiographic findings at birth included dysplasia of the femoral heads and of the acetabulum as well as "bifid" distal femurs (Fig. 2). At age 5 yrs, the proximal femoral epiphyses were not ossified, the femoral heads appeared to be dislocated, and the epiphyses at the knee were dysplastic.

Patient 3. This girl is the younger sister of patient 2. Oligohydramnios and short long bones were noted prenatally. Born after 38 weeks of gestation, she was very short at birth, while weight and head circumference were at the lower limit of normal (weight, 2.7 kg; length, 39 cm; and head circumference, 34,5 cm); clinically, she appeared to have short limbs. She also had rectal atresia without fistula that required a surgical colostomy. She had a normal female karyotype in blood cells. At age 8 months, developmental delay was diagnosed. An echocardiography done at age 2 years 6 months showed persistent foramen ovale and aneurysmatic septum. When she was seen in the genetics clinics at age 3 years 6 months, her weight was 12 kg and her height 79 cm (both below the 3rd percentile). She had brachycephaly, severe bilateral microtia (Fig. 1) with apparently normal ear canal, aplasia cutis on the skull vertex, a very flat nose with triangular nares, arched eyebrows with mild synophrys, high palate, hypodontia (Fig. 1), short neck, and imperforate anus. She was developmentally delayed and attended a specialized Teleton institution. Imaging studies showed right vesico-ureteral reflux with right kidney nephropathy. A brain CT showed dysgenesis of the corpus callosum. Skeletal-radiographic findings are shown in Fig. 2. Radiographic findings were similar to those seen in her sister and included vertebral coronal clefts (Fig. 2) and agenesis of the coccyx.

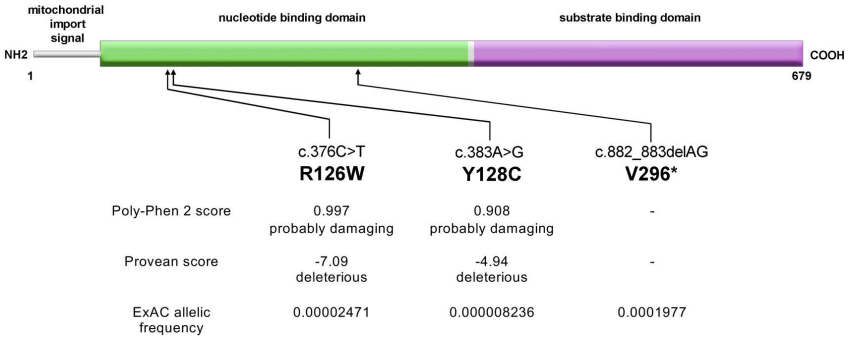


Figure 3. Scheme of the HSPA9 protein showing the localization of the mutations observed. The HSPA9 (mortalin) protein has a short mitochondrial import sequence and two main domains, the nucleotide (ATP/ADP) binding domain and the substrate binding domain (Dores-Silva *et al.*, 2015). The two amino acid substitution affect the nucleotide binding domain; the truncation mutation predicts the loss of part of the nucleotide binding domain and all of the substrate binding domain (unless the mRNA undergoes nonsense-mediated decay; see Results). The lower part shows a summary of the pathogenicity prediction software (PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/) and PROVEAN, http://provean.jcvi.org/index.php) as well as the allelic frequencies in the ExAC project (exac.broadinstitute.org/gene/ENSG00000113013).

Results

Exome sequencing identifies low-frequency mutations in the *HSPA9* gene in individuals with the EVEN-PLUS syndrome. The sequential analysis of the variants identified in the exomic sequence of the patients is presented in table S1. Variants were filtered for non-synonymy, for rarity, and for quality. Subsequently, genes were scored for the presence of either two variants at heterozygosity, or one or more variants at homozygosity, in the affected sibs as well as in the sporadic patient. There was only one gene that fit all criteria, namely, *HSPA9*. Patient 1 was found to be heterozygous for variants c.383A > G (p.Y128C) and c.882_883delAG (p.V296*). Patients 2 and 3 were found to be homozygous for variant c.376C > T (p.R126W). Both R126 and Y128 are extremely conserved (supplementary Figure S1). Results of prediction software PolyPhen-2[13] and Provean[14] suggested damaging results on protein structure (Fig. 3). The V296* truncation mutation abolishes more than half of the protein, including all of the substrate binding domain (Fig. 3); however, the premature termination codon is likely to promote nonsense-mediated decay. All mutations were confirmed by direct bidirectional Sanger sequencing of a second batch of genomic DNA; heterozygosity was confirmed in the unaffected parents. All three mutations were present at extremely low frequency in the ExAC browser (exac.broadinstitute.org/gene/ENSG00000113013) and were absent from the Exome Variant Server (http://evs.gs.washington.edu/EVS/) (Fig. 3). These frequencies are compatible with recessive inheritance of a rare disorder. Interestingly, the R126W mutation had been observed at heterozygosity in one individual out of three cohorts comprising over 1500 adult patients with Parkinson disease, and databases are since annotated with a possible role of this mutation in Parkinson disease[15,16] (but see discussion below).

3D molecular modeling of HSPA9 and mapping of the affected amino acid residues. Mapping of the mutated amino acids on the available HSPA9 nucleotide binding domain structure (NBD)[17] revealed that both R126W and Y128C are located next to each other on the surface of the protein, at some distance from the ATP/ADP binding site (Fig. 4A). Moreover, in our model the two mutations lie on a loop close to the predicted interface between the NBD and the substrate binding domain (SBD, Fig. 4B).

Discussion

Clinical delineation of the EVEN-PLUS syndrome. The identification of *LONP1* mutations in the CODAS syndrome allowed for the recognition of its wide phenotypic spectrum[1,2]. Subsequently, gene-based phenotypic sorting allowed us to identify a CODAS-related phenotype that is not associated with LONP1 mutations. This syndrome shares the skeletal features of the CODAS syndrome (vertebral and epiphyseal changes as shown in Fig. 2) but is further characterized by prenatal-onset short stature, a distinct craniofacial phenotype with microtia, a flat facial profile with flat nose and triangular nares, cardiac malformations, and other findings such as anal atresia, hypodontia, aplasia cutis, and others (see Fig. 1 and Table 1). Examples of this syndrome seem to have been previously reported: in 1990, Kozlowski *et al.* described a six-year-old boy, with « caudal regression and spondylo-epiphyseal dysplasia »[5]. The flat nose, dysplastic ears, and the combination of rectal and bladder incontinence with sacral agenesis observed in that boy are

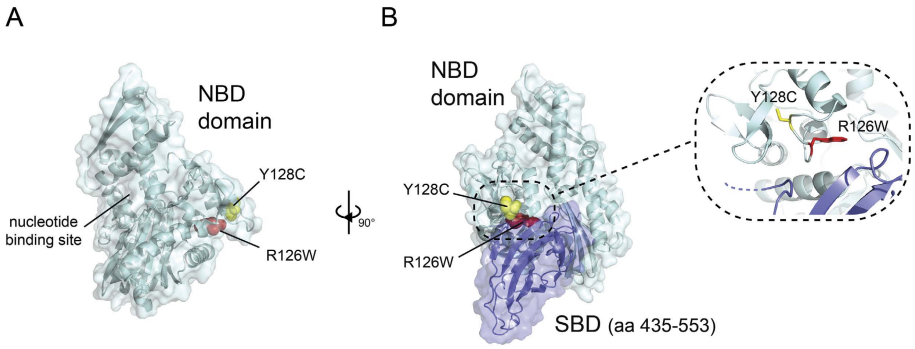


Figure 4. Localization of the mutations observed on the HSPA9 protein structure. (**A**) Mapping of the R126W and Y128C mutation of the crystal structure of the NBD (light blue). Both mutations lie on the surface of the protein, opposite to the nucleotide binding site. R126W is shown in red, Y128C in yellow. (**B**) Localization of the mutations in relation to the NBD/SBD interface (NBD and SBD shown in light green and purple respectively) in a model of the full length protein. Only a portion of the SBD is shown for clarity. Both mutations are located on a single loop near the interface, as shown in detail in the inset.

	Pat. 1	Pat. 2	Pat. 3
birth measurements	length 38 cm, weight 2.2 kg (at week 39); both values markedly below the normal range	length 39 cm, weight 2.8 kg (at week 38); length markedly below the normal range	length 39 cm, weight 2750 g (at week 38); length markedly below the normal range
nose	hypoplastic nose with vertical groove on tip (bifid tip) and triangular nares	hypoplastic nose with vertical groove on tip (bifid tip) and triangular nares	hypoplastic nose with vertical groove on tip (bifid tip) and triangular nares
ears	absent external ears (anotia), open ear duct	severe microtia with absent upper helix	absent external ears with open ear duct; possible hypoacusis
eyes	synophrys; no cataract	synophrys; no cataract	synophrys; no cataract
teeth			single upper central incisor, absence of some lateral incisors
skin	atopic dermatitis, sparse hair	two lateral hair whorls and area of aplasia cutis on the skull vertex	area of aplasia cutis on the skull vertex
heart	ASD (spontaneously closed at age 20 mos)	ASD (ostium secundum)	patent foramen ovale and aneurysmatic septum
gastrointestinal	anal atresia	normal abdominal ultrasonography	anal atresia
kidney/urogenital	No abnormalities on ultrasound	1 UTI at 1 year but normal renal ultrasonography	vesicoureteral reflux, hypoplastic right kidney
brain	normal MRI at age 5 mos	normal cerebral ultrasonography	agenesis of the corpus callosum with separated frontal horns
psychomotor development	Borderline-normal	Normal evaluation at kindergarten level, including language	Moderate developmental delay
HSPA9 mutations	*p.Y128C/p.V296**	*pR126W/p.R126W*	*pR126W/p.R126W*

Table 1. Synopsis of clinical features in the three individuals with EVEN-PLUS syndrome and *HSPA9* mutations.

reminiscent those in our patients; the radiographic changes, namely, vertebral clefts, severe epiphyseal dysplasia, "bifid" appearance of the distal femur, and coccygeal agenesis, are virtually identical (see Fig. 2). In 2009, Marlin *et al.* reported on a boy diagnosed as having CODAS[3]; however, while the facial features in the CODAS syndrome are non-specific, the craniofacial features and the skeletal findings of that patient closely match those observed in our patients. Finally, Amiel and coworkers had reported, in 1994, two sisters who had epiphyseal and vertebral dysplasia in combination with dysplastic external ears; they discussed CODAS as a possible diagnosis but concluded that those sisters probably represented a separate disorder for which they coined the name of EVE (epiphyses, vertebrae, ear) syndrome[4]. Although no DNA could be obtained from these individuals to test for the presence of HSPA9 mutations, the clinical resemblance and similarity of radiographic features strongly suggest that these four patients had the same disorder as the one we

describe here. Thus, we propose to retain the name suggested by Amiel and coworkers and to change it to "EVEN-PLUS syndrome" (for epiphyses, vertebrae, ears and nose, plus associated findings), reflecting the main clinical findings of the syndrome. While the skeletal findings in EVEN-PLUS are shared with CODAS, the facial features and the presence of associated malformations are distinct to the EVEN-PLUS syndrome.

Recessive mutations in the HSPA9 gene are the basis of the EVEN-PLUS syndrome. Several lines of evidence support the causative role of the *HSPA9* mutations in the pathogenesis of this complex malformation syndrome. The genetic evidence consists in the rarity or absence of the variants in our in-house population (to exclude systemic technical errors) and in publicly available databases (ExAC, EVS); their presence at compound heterozygosity in patient 1 (non-consanguineous parents) as well as in homozygosity in patients 2 and 3 (consanguineous parents); the phylogenetic conservation of the affected residues (supplementary Figure 1); as well as the consequences at the protein level both for the premature stop codon mutation predicting the loss of more than half of the protein, and for the amino acid substitutions predicted *in silico* to be damaging (Fig. 3). The R126W and Y128C mutations are located on the NDB but are unlikely to directly affect this activity, due to their distance from the active site. However, the proximity of the mutation sites to the predicted NBD/SBD interface could result in the disruption of this interaction, thereby affecting the function of the protein. Interestingly, our 3D model of HSPA9/mortalin is consistent with a recent in solution analysis of the HSPA9 structure[6], which places the NBD/SBD interface in similar orientation as the one predicted in our own molecular model. It seems therefore likely that the R126W and Y128C mutations result in similar consequences as the ones expected from the V296* mutation, which, if the mRNA escapes nonsense-mediated decay, abolishes the NBD/SBD interface and the whole substrate binding domain. Finally, albeit in another context, the R126W substitution has been shown to have adverse effects on HSPA9 protein function (see below).

HSPA9/mortalin, a protein with (too) many functions? The product of the HSPA9 gene, a 70 kDa heat-shock protein, has been extensively studied in the past 25 years and yielded precious insights on mitochondrial protein import and folding[6–8,10]. It is associated principally with the mitochondrial matrix but can be found in the cytoplasm as well as in the nucleus. Its main role is that of a chaperone that participates in the import of proteins from the cytosol to the mitochondrial matrix as well as their folding[18,19]. As a heat-shock protein with ATPase activity, it prevents the accumulation of unfolded, dysfunctional proteins[20]. Because of its different localizations and multiple binding partners, the HSPA9 protein has been known under several different names (mHSP70, mortalin, mot-2, GRP75) and has been associated with a large number of different processes ranging senescence to immortalization, oncogenesis, neurodegeneration, protection from oxidative stress, hematopoiesis, and even viral replication, among others[6,7]. Some of these studies, done in different contexts, have produced results that are relevant to our results. In 2005, Craven *et al.* studied a zebrafish mutant, *crimsonless*, ascertained for ineffective hematopoiesis. They identified a single amino acid substitution, G492E, in the *HSPA9b* gene (the zebrafish homologue of human *HSPA9*) as responsible for this phenotype. A morpholino knock-down model of *HSPA9b* effectively reproduced the *crimsonless* phenotype[21]. Although the study was focused on hematopoiesis, it was noted that the head and eyes appeared stunted as early as 38 hpf[21]; moreover, after 48 hpf all development appeared to halt, including further maturation of the musculature, fins, and internal organs, and the animals died around 72 hpf[21], indicating that HSPA9b knockdown had more widespread effects on embryonic development. Other investigators have investigated a possible role of HSPA9 in the pathogenesis of Parkinson disease, based on the physical and functional relationship of HSPA9 (often referred to as *mortalin* in that literature) with parkins. In two studies on the possible role of genetic variations in *HSPA9* in the pathogenesis of Parkinson disease, the R126W mutation has been identified at heterozygosity in 1 of 330 Spanish patients[16], and subsequently in 0 (none) of two cohorts of 286 and 1008 German patients[15]. In spite of the questionable statistical significance of this observation, expression studies using wild type and mutated HSPA9 were done, showing that the R126W mutation does impact the function of the HSPA9 protein measured by its effects on mitochondrial morphology, mitochondrial membrane potential, the production of reactive oxygen species[15,22] and the cell sensitivity to exogenous oxidative stress[22]. These expression studies confirm that the R126W mutation is not a neutral polymorphism but does impair HSPA9 function, in accordance with *in silico* predictions and 3D mapping data.

LONP1 mutations in the CODAS syndrome and HSPA9 mutations in the EVEN-PLUS syndrome suggest the existence of a family of "mitochondrial chaperonopathies". The concept of families of phenotypes coming from genes in the same pathway, or related by a common function (such as the cohesinopathies, or the cholesterol biosynthesis defects, or the mucopolysaccharidoses) is well established. While the identification of mutations in a mitochondrial matrix protease (*LONP1*) in patients with the CODAS syndrome was unexpected[1,2], the identification of mutations in a related gene (HSPA9) in EVEN-PLUS syndrome confirms the previous studies and suggests a common pathogenesis of the two syndromes. In fact, the functional relationship between HSPA9 and LONP1 in the mitochondrial chaperone-protease network[9,10,23] and the phenotypic overlap between CODAS and EVEN-PLUS syndromes delineate a new family of disorders – the "mitochondrial chaperonopathies".

Mechanistic pathogenesis of the EVEN-PLUS and CODAS syndrome remains unclear but is likely to be different than that of classic mitochondriopathies. In spite of the numerous studies on HSPA9 function, and of the studies of Strauss *et al.* on *LONP1* mutations found in the CODAS syndrome[1], the pathogenesis of the developmental defects associated with *LONP1* and *HSPA9* mutations remain unexplained. The phenotype of the CODAS syndrome, and even more so that of EVEN-PLUS, do not resemble those of other multisystem mitochondrial diseases; there is no indication of the "energy failure" process that is central to the pathogenesis of prototypical mitochondrial disorders such as the Pearson, Kerns-Sayre, MELAS, MERFF, or Alpers syndromes (reviewed by Andreux *et al.*, 2013[24]). Recessive mutations in a different mitochondrial chaperone-protease, CLPP, have been identified as one cause of the Perrault syndrome of ovarian failure and hearing loss, a condition that more closely resembles the classic mitochondriopathies[25]. Many of the clinical features of EVEN-PLUS syndrome (microtia, small nose, abnormal hair patterns, anal atresia, sacral agenesis) are the consequence of disturbed embryonic morphogenesis. But how can HSPA9 malfunction result in dysmorphogenesis? Heat-shock genes are widely expressed in vertebrate development even in absence of stress factors, suggesting a role in development[26]. Specifically, HSPA9 is known to interact with FGF1[27] as well as with Smad2, decreasing TGF-b signal transduction and affecting epithelial-mesenchymal transition[28]. Studies of the possible roles of HSPA9 in embryonic development may give an explanation for the defects observed in the EVEN-PLUS syndrome.

Awareness about the EVEN-PLUS syndrome and availability of molecular diagnosis will allow to recognize further cases, to increase our knowledge on the phenotypic spectrum associated with mutations in *HSPA9*, and to learn about the long-term outcome of affected individuals. The wealth on publications on the multiple roles of HSPA9/mortalin/mHSP70 notwithstanding, only these clinical observations may give us *in vivo* indications on the essential and non-essential roles of HSPA9 in prenatal development as well as over the entire human life-span.

Methods

The study was done in accordance with regulations for studies on human subjects of the hospitals in Lausanne, Seoul, and Rancagua. In addition, approvals from the IRB of the Seoul National Hospital, and of the Ethics Commission of the Lausanne University Hospital, were obtained. Peripheral blood was obtained with informed consent from the patients and their parents and genomic DNA was extracted by routine methods. Fragmented genomic DNA was purified with AMPure XP beads and the quality of the fragmented DNA was assessed with an Agilent Bioanalyzer. Preparation of the exome enriched, barcoded sequencing libraries was perfomed using Agilent SureSelect Human All Exon v4 kit. The final libraries were quantified with a Qubit Fluorometer (Life Technologies) and the correct size distribution was validated with an Agilent Bioanalyzer. Libraries were sequenced on Illumina HiSeq 2000, generating 100 bp paired-end reads. Raw reads were aligned onto the hg19 reference genome with Novoalign (http://www.novocraft.com) and the data cleanup and variant calling were performed according to GATK Best Practices recommendations[29]. Variant filtering was made with Annovar[30] and with own *perl* and *bash* scripts (available on request). Variants identified by this procedure were verified by direct PCR amplification of target exons from genomic DNA and bidirectional Sanger sequencing. A molecular model for the full-length human HSPA9 protein was generated with I-TASSER[31]. Figures were generated in the PyMOL Molecular Graphics System, Version 1.7.4 (Schrödinger, LLC).

References

1. Strauss, K. A. *et al.* CODAS syndrome is associated with mutations of LONP1, encoding mitochondrial AAA+ Lon protease. *Am J Hum Genet* **96(1),** 121–35 (2015).
2. Dikoglu, E. *et al.* Mutations in LONP1, a mitochondrial matrix protease, cause CODAS syndrome. *Am J Med Genet A* **167(7),** 1501–9 (2015).
3. Marlin, S. *et al.* Fourth case of cerebral, ocular, dental, auricular, skeletal syndrome (CODAS), description of new features and molecular analysis. *Am J Med Genet A* **152A(6),** 1510–4 (2010).
4. Amiel, J. *et al.* Epiphyseal, vertebral, and ear (EVE) dysplasia: a new syndrome? *J Med Genet* **36(7),** 561–4 (1999).
5. Kozlowski, K. *et al.* Caudal regression syndrome and spondyloepiphyseal dysplasia in a 6-year-old child. A new syndrome? *Pediatr Radiol* **21(1),** 75–7 (1990).
6. Dores-Silva, P. R. *et al.* Human mitochondrial Hsp70 (mortalin): shedding light on ATPase activity, interaction with adenosine nucleotides, solution structure and domain organization. *PLoS One* **10(1),** e0117170 (2015).
7. Flachbartova, Z. & Kovacech, B. Mortalin - a multipotent chaperone regulating cellular processes ranging from viral infection to neurodegeneration. *Acta Virol* **57(1),** 3–15 (2013).
8. Wadhwa, R., Taira, K. & Kaul, S. C. An Hsp70 family chaperone, mortalin/mthsp70/PBP74/Grp75: what, when, and where? *Cell Stress Chaperones* **7(3),** 309–16, 2002.
9. Savel'ev, A. S. *et al.* ATP-dependent proteolysis in mitochondria. m-AAA protease and PIM1 protease exert overlapping substrate specificities and cooperate with the mtHsp70 system. *J Biol Chem* **273(32),** 20596–602 (1998).
10. Voos, W. Chaperone-protease networks in mitochondrial protein homeostasis. *Biochim Biophys Acta* **1833(2),** 388–99 (2013).
11. Lim, J. H. *et al.* The mitochondrial Hsp70-dependent import system actively unfolds preproteins and shortens the lag phase of translocation. *EMBO J* **20(5),** 941–50 (2001).
12. Liu, Q. *et al.* Mitochondrial Hsp70 Ssc1: role in protein folding. *J Biol Chem* **276(9),** 6112–8 (2001).
13. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* Chapter 7, Unit 720 (2013).
14. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31(16),** 2745–7 (2015).

15. Burbulla, L. F. *et al.* Dissecting the role of the mitochondrial chaperone mortalin in Parkinson's disease: functional impact of disease-related variants on mitochondrial homeostasis. *Hum Mol Genet* **19(22),** 4437–52 (2010).
16. De Mena, L. *et al.* Mutational screening of the mortalin gene (HSPA9) in Parkinson's disease. *J Neural Transm* **116(10),** 1289–93 (2009).
17. Amick, J. *et al.* Crystal structure of the nucleotide-binding domain of mortalin, the mitochondrial Hsp70 chaperone. *Protein Sci* **23(6),** 833–42, 2014.
18. Kang, P. J. *et al.* Requirement for hsp70 in the mitochondrial matrix for translocation and folding of precursor proteins. *Nature* **348(6297),** 137–43 (1990).
19. Schneider, H. C. *et al.* Mitochondrial Hsp70/MIM44 complex facilitates protein import. *Nature* **371(6500),** 768–74 (1994).
20. Finka, A., Sharma, S. K. & Goloubinoff, P. Multi-layered molecular mechanisms of polypeptide holding, unfolding and disaggregation by HSP70/HSP110 chaperones. *Front Mol Biosci* **2(29),** 2015.
21. Craven, S. E. *et al.* Loss of Hspa9b in zebrafish recapitulates the ineffective hematopoiesis of the myelodysplastic syndrome. *Blood* **105(9),** 3528–34 (2005).
22. Wadhwa, R. *et al.* Functional significance of point mutations in stress chaperone mortalin and their relevance to Parkinson disease. *J Biol Chem* **290(13),** 8447–56 (2015).
23. Venkatesh, S. *et al.* Multitasking in the mitochondrion by the ATP-dependent Lon protease. *Biochim Biophys Acta* **1823(1),** 56–66 (2012).
24. Andreux, P. A., Houtkooper, R. H. & Auwerx, J. Pharmacological approaches to restore mitochondrial function. *Nat Rev Drug Discov* **12(6),** 465–83 (2013).
25. Jenkinson, E. M. *et al.* Perrault syndrome is caused by recessive mutations in CLPP, encoding a mitochondrial ATP-dependent chambered protease. *Am J Hum Genet* **92(4),** 605–13 (2013).
26. Rupik, W. *et al.* The expression patterns of heat shock genes and proteins and their role during vertebrate's development. *Comp Biochem Physiol A Mol Integr Physiol* **159(4),** 349–66 (2011).
27. Mizukoshi, E. *et al.* Fibroblast growth factor-1 interacts with the glucose-regulated protein GRP75/mortalin. *Biochem J* **343** Pt **2**, 461–6 (1999).
28. Li, Y., Kang, X. & Wang, Q. HSP70 decreases receptor-dependent phosphorylation of Smad2 and blocks TGF-beta-induced epithelial-mesenchymal transition. *J Genet Genomics* **38(3),** 111–6 (2011).
29. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **11(1110),** 1101–033 (2013).
30. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38(16),** e164 (2010).
31. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12(1),** 7–8 (2015).

Acknowledgements

We thank Fred De Sauvage (San Francisco) for discussions on the *crimsonless* fish model, and Wolfgang Voos (Bonn) and Pierre Goloubinoff (Lausanne) for discussions on HSPA9 and LONP1. We wish to acknowledge Keith Harshman and his team at the Functional Genomics Platform of the University of Lausanne for generating the raw exome sequencing data and that of Carole Chiesa for obtaining the Sanger sequences. This work was supported by the CoSMO-B research grant (to ASF, LB and SU) as well as by the Leenaards Foundation (www.leenaards.ch).

Author Contributions

A.S.F., S.U., L.B., B.R.B. and C.V. conceived the study; S.C.T., R.M., T.J.C. and J.H.C ascertained the patients; A.S.F., S.U., O.K.K. and L.B. reviewed the clinical and radiographic findings in all patients; B.R.B., M.C., S.U. and C.R. obtained and reviewed the exome sequence data; E.D. and B.C.X. performed all sequence confirmations and database analyses for the variants; E.G. and G.S.F. obtained and reviewed the protein model; A.S.F., S.U., C.R., L.B., E.G. and G.S.F. drafted and wrote the paper; all authors reviewed the paper.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Royer-Bertrand, B. *et al.* Mutations in the heat-shock protein A9 (*HSPA9*) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia. *Sci. Rep.* **5**, 17154; doi: 10.1038/srep17154 (2015).

6 Appendix

6.1 NANS

NANS-mediated synthesis of sialic acid is required for brain and skeletal development

Own contribution: Analyzed and described the pathogenic mutations identified in the WES of the patients.

nature genetics

NANS-mediated synthesis of sialic acid is required for brain and skeletal development

Clara D M van Karnebeek[1,2,28], Luisa Bonafé[3,28], Xiao-Yan Wen[4,5,28], Maja Tarailo-Graovac[2,6], Sara Balzano[7], Beryl Royer-Bertrand[3,7], Angel Ashikov[8], Livia Garavelli[9], Isabella Mammi[10], Licia Turolla[11], Catherine Breen[12], Dian Donnai[12], Valerie Cormier[13], Delphine Heron[13], Gen Nishimura[14], Shinichi Uchikawa[15], Belinda Campos-Xavier[3], Antonio Rossi[16], Thierry Hennet[17], Koroboshka Brand-Arzamendi[4,5], Jacob Rozmus[1], Keith Harshman[18], Brian J Stevenson[19], Enrico Girardi[20], Giulio Superti-Furga[20,21], Tammie Dewan[1], Alissa Collingridge[1], Jessie Halparin[1], Colin J Ross[1,2,6], Margot I Van Allen[6], Andrea Rossi[22], Udo F Engelke[23], Leo A J Kluijtmans[23], Ed van der Heeft[23], Herma Renkema[23], Arjan de Brouwer[24], Karin Huijben[23], Fokje Zijlstra[23], Thorben Heisse[25], Thomas Boltje[25], Wyeth W Wasserman[2,6], Carlo Rivolta[7], Sheila Unger[26], Dirk J Lefeber[8,23], Ron A Wevers[23,29] & Andrea Superti-Furga[3,27,29]

We identified biallelic mutations in *NANS*, the gene encoding the synthase for *N*-acetylneuraminic acid (NeuNAc; sialic acid), in nine individuals with infantile-onset severe developmental delay and skeletal dysplasia. Patient body fluids showed an elevation in *N*-acetyl-D-mannosamine levels, and patient-derived fibroblasts had reduced NANS activity and were unable to incorporate sialic acid precursors into sialylated glycoproteins. Knockdown of *nansa* in zebrafish embryos resulted in abnormal skeletal development, and exogenously added sialic acid partially rescued the skeletal phenotype. Thus, NANS-mediated synthesis of sialic acid is required for early brain development and skeletal growth. Normal sialylation of plasma proteins was observed in spite of NANS deficiency. Exploration of endogenous synthesis, nutritional absorption, and rescue pathways for sialic acid in different tissues and developmental phases is warranted to design therapeutic strategies to counteract NANS deficiency and to shed light on sialic acid metabolism and its implications for human nutrition.

Intellectual developmental disorders (IDDs) affect 2–2.5% of children and adults worldwide[1]. The developmental origin of IDDs is reflected in their definition as "substantial impairments of intellectual function and social or adaptive functioning present from early childhood" (ref. 2). Recent advances have shown that, in many cases, the etiology of these disorders is genetic, most frequently involving *de novo* mutations[3,4]. Along with a better understanding of the surrounding condition and prognosis, insights into the molecular basis of neurocognitive impairment allow for the development and application of targeted therapeutic strategies[5]. Although less frequent than IDDs, genetic disorders affecting skeletal development and growth (commonly called 'skeletal dysplasias') are a group of over 500 distinct disorders[6].

[1]Department of Pediatrics, University of British Columbia, Vancouver, British Columbia, Canada. [2]Centre for Molecular Medicine, Child and Family Research Institute, University of British Columbia, Vancouver, British Columbia, Canada. [3]Centre for Molecular Diseases, Lausanne University Hospital (CHUV), University of Lausanne, Lausanne, Switzerland. [4]Zebrafish Centre for Advanced Drug Discovery, Keenan Research Centre for Biomedical Science, St. Michael's Hospital, Toronto, Ontario, Canada. [5]Department of Medicine, University of Toronto, Toronto, Ontario, Canada. [6]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. [7]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. [8]Department of Neurology, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, the Netherlands. [9]Clinical Genetics Unit, IRCCS-S. Maria Nuova Hospital, Reggio Emilia, Italy. [10]Ambulatorio di Genetica Medica ULSS 13, U.O. Ginecologia e Ostetricia, Ospedale Dolo, Dolo, Italy. [11]Medical Genetics Unit, Local Health Authority (ULSS 9), Treviso, Italy. [12]Manchester Centre for Genomic Medicine, Institute of Human Development, Faculty of Medical and Human Sciences, University of Manchester, St. Mary's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK. [13]Institut IMAGINE, Hôpital Necker–Enfants Malades, Paris, France. [14]Department of Radiology, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan. [15]Department of Orthopedics, National Center for Child Health and Development, Tokyo, Japan. [16]Department of Molecular Medicine, Unit of Biochemistry, University of Pavia, Pavia, Italy. [17]Department of Physiology, University of Zürich, Zurich, Switzerland. [18]Genomic Technologies Facility, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland. [19]Vital-IT Group, Swiss Institute of Bioinformatics, University of Lausanne, Lausanne, Switzerland. [20]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. [21]Center for Physiology and Pharmacology, Medical University of Vienna, Vienna, Austria. [22]Neuroradiology Department, G. Gaslini Children's Hospital, Genoa, Italy. [23]Translational Metabolic Laboratory, Department of Laboratory Medicine, Radboud University Medical Center, Nijmegen, the Netherlands. [24]Department of Genetics, Radboud University Medical Center, Nijmegen, the Netherlands. [25]Department of Organic Chemistry, Radboud University, Nijmegen, the Netherlands. [26]Medical Genetics Service, Lausanne University Hospital, University of Lausanne, Lausanne, Switzerland. [27]Department of Pediatrics, Lausanne University Hospital, University of Lausanne, Lausanne, Switzerland. [28]These authors contributed equally to this work. [29]These authors jointly supervised this work. Correspondence should be addressed to A.S.-F. (asuperti@unil.ch), R.A.W. (ron.wevers@radboudumc.nl) or C.D.M.v.K. (cvankarnebeek@cw.bc.ca).

Received 2 March; accepted 29 April; published online 23 May 2016; doi:10.1038/ng.3578

Studying their molecular basis has provided valuable insights into the many factors necessary for skeletal development, ranging from minerals and structural molecules to enzymes and to signaling molecules and transcription factors[6,7]. We report here a genetic disorder presenting as a combination of severe IDD with skeletal dysplasia and short stature. Our data show that the pathogenic basis of this disorder is an inborn error of metabolism that affects the endogenous synthesis of NeuNAc. Exploration of the biochemical and molecular features of this disorder provides new information on the role of sialic acid in the development of brain and bone.

RESULTS

Clinical and radiographic phenotypes of NANS deficiency

Clinical patient reports are presented *in extenso* in the **Supplementary Note**. Nine patients from six families were studied; patients 1, 2, and 5 have previously been described[8,9]. The main clinical features of the disorder included a prenatal history that was unremarkable in all patients except for one (patient 8), in whom prenatal hydrocephalus was diagnosed. No specific signs and symptoms were present at birth, except for disproportionately short limbs that were observed in three patients (patients 3, 8, and 9). In the first months of life, all patients showed muscle hypotonia, and the achievement of early developmental milestones such as sitting and walking was delayed. Subsequently, global developmental delay, including cognitive impairment, was the major medical concern. All adult patients showed moderate-to-severe IDD; only one patient acquired speech, and none was living independently. Seizures were a prominent and early feature in one patient but were infrequent in some and absent in others. Social competences were relatively preserved. Body measurements at birth were normal or slightly lower than normal, but growth velocity decreased during the first or second year of life, and short stature with shortening of both the trunk and limbs was present in all adult patients. Facial features included a prominent forehead, mild synophrys, a sunken nasal bridge, a prominent bulbous nasal tip, and full lips (**Fig. 1**). No endocrine anomalies were noted, and adult patients had gone through pubertal development. Neuroimaging was available for six patients: patient 8 showed prenatal-onset hydrocephalus, whereas patient 9 showed perisylvian polymicrogyria, small basal ganglia, and reduced white matter mass (**Supplementary Fig. 1**); the other four patients showed moderate cerebral atrophy with nonspecific changes in white matter. Distinct features that permitted these individuals to be distinguished from among the large group of persons with severe IDD included the facial dysmorphisms and the skeletal dysplasia with short stature, premature carpal ossification, platyspondyly, longitudinal metaphyseal striations, and small epiphyses (see **Fig. 1** for details). For more radiographic details, see also the previous clinical reports for patients 1 and 2 (ref. 8) and for patient 5 (corresponding to patient 1 in ref. 9).

Identification of *NANS* mutations and functional analysis

Exome sequencing was performed on genomic DNA from patients 1–4. After passage through the filtering pipeline (Online Methods and **Supplementary Table 1**), only one gene, *NANS*, showed biallelic variants in all four patients (**Table 1**). A small indel just 5′ to exon 4, c.449–10_449–5delGATTACinsATGG, was seen in a heterozygous state in all four patients (reference transcript ENST00000210444, NCBI reference sequence NM_018946.3). In addition to this shared mutation, patients 1 and 2 (who are sisters) had a single-nucleotide insertion predicted to result in a frameshift with early truncation (c.389_390insT, p.Lys131Glnfs*8), whereas patients 3 and 4 (who are brother and sister) had a mutation of a canonical splice donor site (c.448+1G>A). Computational haplotype reconstruction using exome data identified a specific haplotype encompassing a region of 1.38 Mb on chromosome 9 (from 100,388,119 (rs10817858) to 101,767,385 (rs41305481) bp) in the four individuals carrying the intronic indel, indicative of a common origin for this variant. Sanger sequencing was used to confirm the mutations and to verify segregation with disease in all families. To test for the pathogenicity of the identified variants, *NANS* cDNA was retrotranscribed from fibroblast or lymphoblastoid cell RNA from patients 1, 3, and 4, as well as from the parents of patients 3 and 4, incubated with or without cycloheximide, PCR amplified, and analyzed by capillary electrophoresis (**Supplementary Fig. 2**). The c.449–10_449–5delGATTACinsATGG indel (patients 1–4) resulted in very low levels of mRNA, with the presence of wild-type mRNA as well as an isoform lacking both exons 3 and 4 (in frame); in the presence of cycloheximide, this allele seemed to produce additional abnormal isoforms, which apparently were subject to nonsense-mediated RNA decay (NMD). The c.448+1G>A variant (patients 3 and 4) produced two splicing isoforms, one lacking both exons 3 and 4 (similar to the previous mutation) and expressed at levels comparable to the wild-type allele and an out-of-frame isoform lacking exon 3 and part of exon 4 that was detected at very low levels. The exonic insertion c.389_390insT (patients 1 and 2) triggered NMD, as demonstrated by sequencing RT–PCR clones obtained from cells with or without cycloheximide treatment. Thus, all three mutations resulted either in unstable or nonfunctional *NANS* mRNA or in reduced levels of wild-type transcripts. We then screened patients 5–8 by selective PCR amplification of *NANS* exons from genomic DNA and direct bidirectional Sanger sequencing of the amplicons and found biallelic mutations in all the patients, including four missense mutations and one triplet insertion leading to the duplication of one amino acid (**Fig. 2** and **Table 1**).

Investigation of patient 9, the youngest patient in our series, followed a different course; this patient was enrolled into the TIDEX study, which combines genomics and metabolomics screening[10]. Metabolomic screening had identified an unusual metabolite, *N*-acetyl-D-mannosamine (ManNAc), in the plasma and urine of this child. Among the variants identified by exome sequencing in this patient, two missense mutations in *NANS* stood out as ManNAc-6-phosphate is the substrate of the NANS enzyme and malfunction of this enzyme could account for the accumulation of ManNAc in body fluids. The *NANS* variants identified in patient 9 were confirmed via Sanger sequencing and were shown to segregate correctly in the family. Of note, the exome results for patients 1–5 were examined for possible pathogenic mutations in the *GNE* gene, but none were observed. After this stage, the data from patients 1–8 and those from patient 9 were combined for all subsequent studies.

Mapping of mutations on the three-dimensional protein model of NANS

The NANS protein was modeled as a homodimer on the basis of the *Neisseria meningitidis* homolog, as well as recent results supporting dimer formation for the human protein[11]. Of the predicted amino acid changes (**Table 1**), four mapped in or near the active site: p.Lys131Gln, p.Gly133Val, p.Tyr188His, and p.Pro189Leu. It is likely that any of these would affect the catalytic activity of the enzyme, by changing the ability of the protein to bind substrate (Lys131 and Tyr188 are also predicted to make hydrogen bonds with substrates[11]; **Fig. 2**), by changing the pocket shape (p.Gly133Val), or by affecting functional residues in the active site (Pro189 is located right next to Tyr188). The p.His29Asn and p.Arg237Cys substitutions are localized at the dimer interface and likely affect protein folding, stability, and/or dimer formation. Although close to the dimer interface, the Ile327 residue that is duplicated is not involved in protein contacts in our model, the

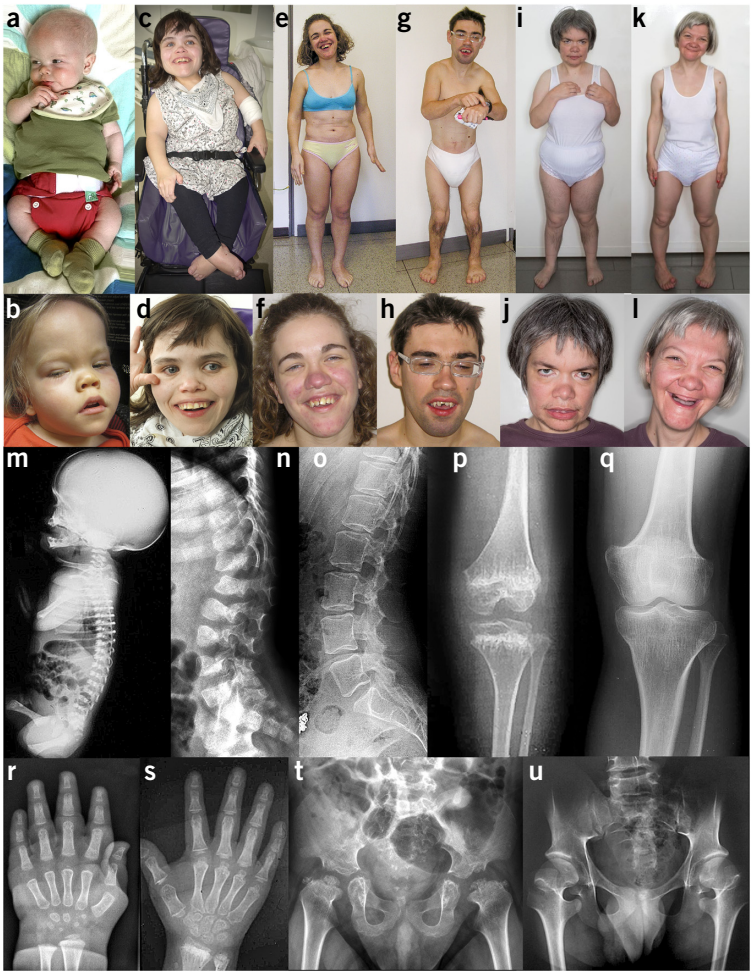


Figure 1 Morphological and skeletal features of NANS-deficient patients. (**a**,**b**) Patient 9 at age 6 months (**a**) and at age 3 years (**b**). (**c**,**d**) Patient 8 at age 28 years. (**e**,**f**) Patient 4 at age 24 years. (**g**,**h**) Patient 3 at age 28 years. (**i**,**j**) Patient 2 at age 38 years. (**k**,**l**) Patient 1 at age 40 years. For patient 9, only the phenotype of short limbs is apparent at age 6 months; facial dysmorphism with prominent forehead, saddle nose, full lips, and coarsening of traits is apparent at age 3 years. All adult patients have short stature (height between 130 and 150 cm). All photographs were obtained and published with consent. (**m**–**o**) Lateral spine films showing coronal clefts at birth (patient 8) (**m**) and severe vertebral body dysplasia at age 3 years (patient 9) (**n**); in adulthood, the vertebral bodies have a normal shape (patient 2) (**o**). (**p**,**q**) At the knees, there are metaphyseal striations and small epiphyses (epimetaphyseal dysplasia) at age 9 years (patient 8) (**p**); in adulthood, epiphyses remain small, whereas metaphyseal striations have disappeared (patient 2) (**q**). (**r**,**s**) Advanced carpal ossification at age 21 months (patient 9) (**r**) and metaphyseal striations (distal radius) with epiphyseal dysplasia at age 9 years (patient 8) (**s**). (**t**,**u**) Short femoral necks and small, irregular capital femoral epiphyses (epimetaphyseal dysplasia) at age 3 years, 7 months (patient 9) (**t**); in the adult (patient 3), short femoral neck and small epiphyses are present, but bone and cartilage structure appear normal (**u**). Compare the radiographic images of patients 1 and 2 and patient 5, as reported earlier[8,9].

duplication could be affecting folding instead. Finally, p.Arg151His is located on the surface of the protein and away from the dimer interface and substrate-binding site, possibly interfering with folding or with a critical protein interaction.

NANS mutations lead to accumulation of *N*-acetylmannosamine *in vivo*

Next-generation metabolic screening (Online Methods) was first applied to the cerebrospinal fluid (CSF) of patient 9 (**Fig. 3** and **Supplementary Table 2**) and then to the plasma of patients 1–4 and 9 (data not shown), leading to the identification of an unusual compound, ManNAc, in all five patients. Quantitative nuclear magnetic resonance (NMR) spectroscopy was then used to determine the concentration of ManNAc in available urine samples from patients. In patients 1–4 and 8 (all adults at the time of study), the urine concentration of ManNAc ranged from 41 to 98 µmol/mmol creatinine (reference values, <10 µmol/mmol creatinine), whereas the excretion of ManNAc was highest (295 µmol/mmol creatinine) in patient 9 (age 3 years). NMR spectroscopy was applied to homogenates of cultured fibroblasts to explore the intracellular sialic acid synthesis pathway. In fibroblasts from two unrelated patients (patients 3 and 9), this analysis showed increased intracellular levels of ManNAc-6-phosphate rather than free ManNAc in comparison to control cells. NANS can act as a NeuNAc phosphate synthase (MIM 605202; Enzyme Commission (EC) database 2.5.1.57). This may account for the accumulation of ManNAc in body fluids and ManNAc-6-phosphate within cells in patients. The free NeuNAc concentration in patient-derived fibroblasts was normal (**Supplementary Table 2**). Finally, we evaluated whether deficient NANS activity would lead to systemic deficiency for sialic acid. The concentration of free NeuNAc was evaluated in the urine of five patients from four families (relative to creatinine excretion) as well as in the CSF of patient 9. In all patients, normal values were found (**Supplementary Table 3**), suggesting that there was no systemic depletion of free neuraminic acid. Analysis of plasma transferrin and apolipoprotein C-III isoforms, to evaluate the biosynthesis of N- and O-linked glycans, respectively, has been carried out repeatedly on several of the patients in the clinical setting, with these analyses giving normal results (**Supplementary Table 4**). This evidence further contributes to the notion that peripheral sialylation is not significantly affected, despite NANS deficiency.

NANS mutations impair enzyme activity

In two preliminary experiments, we explored the sialylation of proteins and lipids at the cell surface of patient-derived and control fibroblasts using FITC-labeled *Sambucus nigra* lectin (which specifically binds to terminal galactose-bound sialic acid residues) and FACS analysis to determine the cellular content of CMP-NeuNAc

Table 1 Overview of the *NANS* mutations observed in the nine patients

Genomic DNA position on chr. 9 (bp)	cDNA change[a]	Exon	Protein change	Present in databases[b]	PolyPhen score	Provean score	Observed in patient(s)
100,819,175	c.85C>A	1	p.His29Asn	No	Probably damaging	Deleterious	Pt. 8
100,839,236	c.389_390insT	3	p.Lys131GlnfsTer8	Wellderly (0.0008)	NA	NA	Pts. 1 and 2
100,839,249	c.398G>T	3	p.Gly133Val	No	Probably damaging	Deleterious	Pts. 6 and 7
100,839,300	c.448+1G>A	3	Aberrant splicing of exons 3 and 4 as seen in mRNA studies	ExAC (0.00002633)	NA	NA	Pts. 3 and 4
100,840,462	c.449–10_449–5delGATTACinsATGG[c]	4	Aberrant splicing of exons 3 and 4 as seen in mRNA studies	ExAC (0.00004947)	NA	NA	Pts. 1 and 2 Pts. 3 and 4
100,840,478	c.452G>A	4	p.Arg151His	Wellderly (0.0008)	Probably damaging	Neutral	Pt. 5 (homozygous)
100,840,588	c.562T>C	4	p.Tyr188His	No	Probably damaging	Deleterious	Pt. 9
100,840,592	c.566C>T	4	p.Pro189Leu	No	Probably damaging	Deleterious	Pt. 8
100,843,203	c.709C>T		p.Arg237Cys	No	Probably damaging	Deleterious	Pt. 9
100,845,238	c.981insATC	6	p.Ile327dup	No	Probably damaging	Deleterious	Pts. 6 and 7

NA, not applicable.[a]Based on transcript ENST00000210444 (CCDS6733), NCBI reference sequence NM_018946.3. [b]The databases considered included the ExAC database (accessed November 2015) and the Wellderly database (accessed January 2016) (see URLs). [c]This mutation was flagged as the following two distinct events, both in our pipeline and in the ExAC browser: (i) an intronic insertion, chr. 9: g.100840462C>CTGA (http://exac.broadinstitute.org/variant/9-100840462-C-CTGA), and (ii) an intronic deletion, chr. 9: g.100840465GATTAC>G (http://exac.broadinstitute.org/variant/9-100840465-GATTAC-G). This is due to the fact that both our pipeline and the ExAC pipeline used GATK variant calling software (see URLs). In ExAC data, the 'two' mutational events have been seen in the same number of individuals of European origin. Sanger sequencing in our samples confirms that the variant is a single mutational event.

and total NeuNAc before and after the addition of 10 mM ManNAc[12]. No differences between patient-derived and control fibroblasts were seen, possibly because the concentration of ManNAc used was substantially higher than what is found physiologically. We then developed a method to measure NANS enzyme activity in cell lysates by incubating lysates with ManNAc-6-phosphate and phosphoenolpyruvate (PEP) and quantifying newly formed NeuNAc by mass spectrometry (**Fig. 3d**). In comparison with fibroblasts from five healthy controls (613 ± 150 nmol NeuNAc/mg protein, mean ± s.d.), the three available patient-derived fibroblast lines showed reduced production of NeuNAc (163 ± 111 nmol NeuNAc/mg protein, mean ± s.d.) after 24 h of incubation. Fibroblasts from the heterozygous father of patients 3 and 4 showed intermediate NANS activity (403 nmol NeuNAc/mg protein). Although residual activity in this assay was high, these results are in agreement with an autosomal recessive defect in *NANS*.

Incorporation of *N*-acetyl-D-mannose into sialoglycoproteins is impaired

We then applied metabolic labeling of sialic acids using propargyloxycarbonyl (Poc)-derivatized analogs of ManNAc and NeuNAc (ManNPoc and NeuNPoc, respectively)[13], in a recently developed technique that had been useful in confirming deficient sialic acid incorporation in cells deficient for the Golgi transporter of CMP–sialic acid[14]. NeuNPoc, which enters the metabolic pathway downstream of the enzymatic step catalyzed by NANS, was incorporated efficiently into glycoproteins in all cell lines analyzed, whereas ManNPoc, which enters the pathway upstream of the NANS-catalyzed step (**Fig. 4**), was incorporated in fibroblasts from a control and a heterozygote for a *NANS* mutation but not in fibroblasts from patients 3, 8, and 9 (**Fig. 3e**). These data confirm the functional impairment of NANS activity in the metabolic pathway of sialic acid biosynthesis and protein sialylation. NANS deficiency should therefore be included on the list of congenital disorders of glycosylation (CDGs). Notably, these data suggest that exogenous NeuNAc might be used to bypass the enzymatic block.

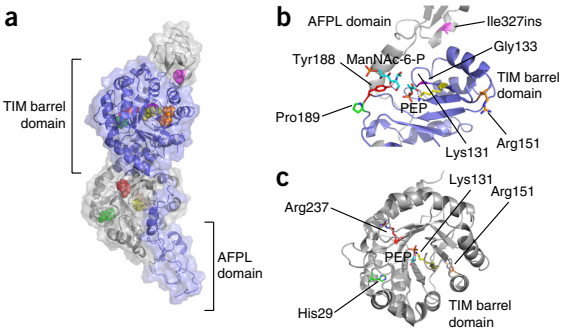


Figure 2 Three-dimensional model of the NANS protein and mapping of amino acid residues affected by mutations. (**a**) Model of the human NANS protein, showing one monomer in gray and one monomer in blue. Mutated residues are color-coded according to patient, with each allele mapped on a separate monomer. Lys131 and Arg151 are shown on both monomers. (**b**,**c**) Detailed views of the location of the alterations. PEP, phosphoenolpyruvate; TIM, triose phosphate isomerase.

nansa knockdown perturbs zebrafish skeletal development

There are two zebrafish orthologs for the human *NANS* gene, the *nansa* and *nansb* genes (see the Online Methods for details). *nansa* is expressed during early embryonic development, including during the stages with 50% epiboly in the axis, 1–13 somites at the notochord and polster, and 14–19 somites at the hatching gland. Thereafter, *nansa* is strongly expressed in the head, immature eye, myotome, optic tectum, and pharyngeal arch skeleton[15]. The expression pattern of *nansb* is unknown. We designed morpholino oligonucleotides (MOs) that block splicing to knock down both the *nansa* and *nansb* genes[16] (Online Methods). Microinjection of *nansa* MO into newly fertilized zebrafish eggs resulted in embryos with a small head, pericardial edema, and developmental anomalies of the skeleton, highlighted by Alcian blue staining, at 6 days post-fertilization (d.p.f.) (**Fig. 5a**,**b**). Interestingly, *nansa* morphants showed a complex phenotype in the area of the head, including hypoplastic or absent Meckel's cartilage, a lack of basihyal, shortened and abnormal ethmoid plate, trabecula, parachordal and palatoquadrate, and absence of the ceratobranchial structures (**Fig. 5b**)[17]. *nansb* morphants did not show an overt abnormal phenotype even at higher concentrations of the MOs (data not shown).

Partial rescue of skeletal development by exogenous NeuNAc

Given the position of Nansa (and NANS in human) in the synthetic pathway of NeuNAc, we tested whether the addition of sialic acid in zebrafish embryo water would rescue the head and skeleton

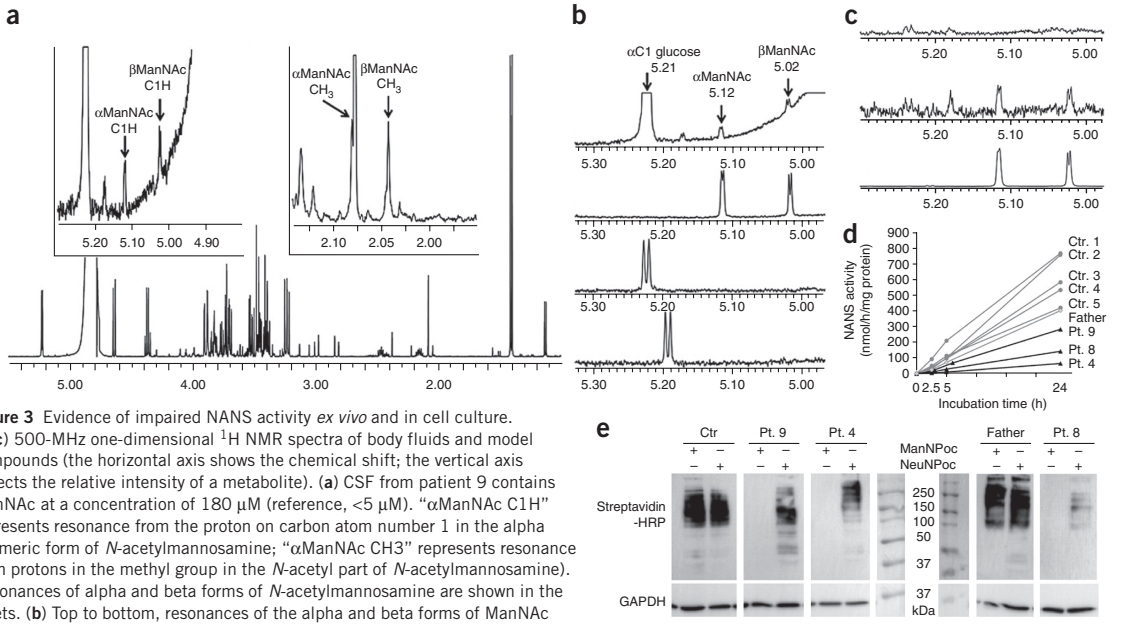


Figure 3 Evidence of impaired NANS activity *ex vivo* and in cell culture. (**a**–**c**) 500-MHz one-dimensional ^{1}H NMR spectra of body fluids and model compounds (the horizontal axis shows the chemical shift; the vertical axis reflects the relative intensity of a metabolite). (**a**) CSF from patient 9 contains ManNAc at a concentration of 180 µM (reference, <5 µM). "αManNAc C1H" represents resonance from the proton on carbon atom number 1 in the alpha anomeric form of *N*-acetylmannosamine; "αManNAc CH3" represents resonance from protons in the methyl group in the *N*-acetyl part of *N*-acetylmannosamine). Resonances of alpha and beta forms of *N*-acetylmannosamine are shown in the insets. (**b**) Top to bottom, resonances of the alpha and beta forms of ManNAc for the inset-relevant part of the CSF spectrum in **a** in comparison to the spectra for ManNAc, GalNAc (*N*-acetylgalactosamine), and GlcNAc (*N*-acetylglucosamine). (**c**) Top to bottom, spectra obtained from the fibroblasts of a control and patient 8 and the spectrum for a model compound, ManNAc-6-phosphate. (**d**) NANS enzyme activity in fibroblast lysates. Fibroblast lysates were from five controls (Ctr.), the heterozygous father of patients 3 and 4 (Father), and patients 4, 8, and 9. For details on the enzyme assay, see the Online Methods. (**e**) Metabolic labeling of sialylated proteins. Fibroblasts from a control, the heterozygous father of patients 3 and 4, and patients 4, 8, and 9 (**Table 1**) were incubated with peracetylated propargyloxycarbonyl analogs of ManNAc and NeuNAc—ManNPoc and NeuNPoc, respectively (see **Fig. 4** for the entry point of these molecules in the sialic acid biosynthetic pathway). Cell proteins were detected by immunoblotting using streptavidin-HRP (top) or antibody against GAPDH (bottom) as an internal control. In cells from the control and the heterozygous father, both ManNPoc and NeuNPoc result in strong labeling of proteins. In patient-derived cells, ManNPoc (which enters the synthetic pathway upstream of NANS) is unable to label proteins, whereas NeuNPoc (which enters the pathway downstream of the NANS enzyme) is incorporated efficiently. These findings confirm a metabolic block between ManNAc and NeuNAc, consistent with impaired NANS activity.

developmental phenotypes of *nansa* morphants. Addition of sialic acid at a concentration of 200 µM resulted in partial rescue of the skeletal phenotype, as measured by the reappearance and development of Meckel's cartilage structure; this structure was correctly formed in 9% of embryos under baseline *nansa* knockdown conditions but in 61% of embryos when sialic acid was added to the water (**Fig. 5c**,**d**).

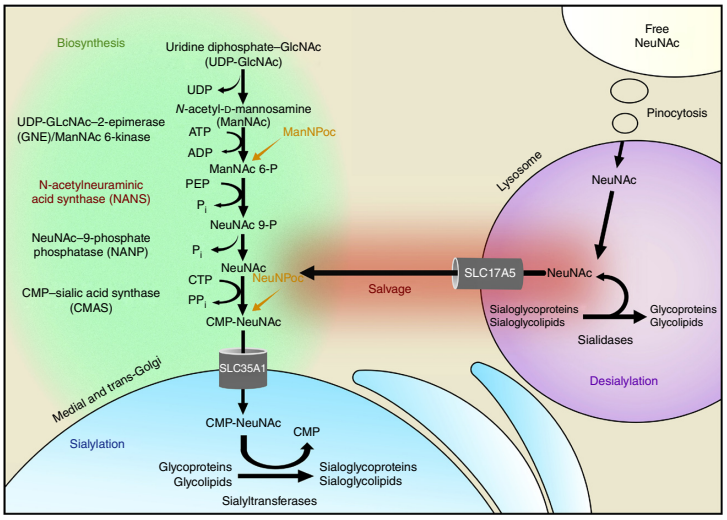


Figure 4 Simplified scheme of *N*-acetylneuraminic acid metabolism in human. Biosynthesis of NeuNAc is achieved largely in the cytoplasm, except for the CMP–sialic acid synthase (CMAS) reaction, which takes place in the nucleus (the nuclear envelope and its possible transporters are omitted here for simplicity). The synthesis of ManNAc is carried out in two steps by the bifunctional enzyme GNE/ManNAc 6-kinase. NeuNAc synthase (NANS) is highlighted in red. CMP–sialic acid is transferred by a specific transporter, SLC35A1, into the medial and trans-Golgi apparatus, where it is used as a substrate for the sialylation of proteins and lipids by various sialyltransferases. When sialylated glycoproteins and lipids are degraded in the lysosome, the free sialic acid released by sialidases can diffuse out of the lysosome through the lysosomal sialic acid transporter (SLC17A5). Free NeuNAc (as well as non-human sialic acid, NeuNGc) can be taken up by pinocytosis and released into the cytoplasm by the SLC17A5 transporter. Free sialic acid can then reenter the biosynthesis pathway (indicated as 'salvage' in the scheme). The relative importance of biosynthesis and salvage in different cell types and tissues is largely unknown. The metabolic scheme also shows the entry points of the two synthetic analogs used in this study, ManPoc and NeuNPoc, with the first being upstream and the second downstream of the step catalyzed by NANS.

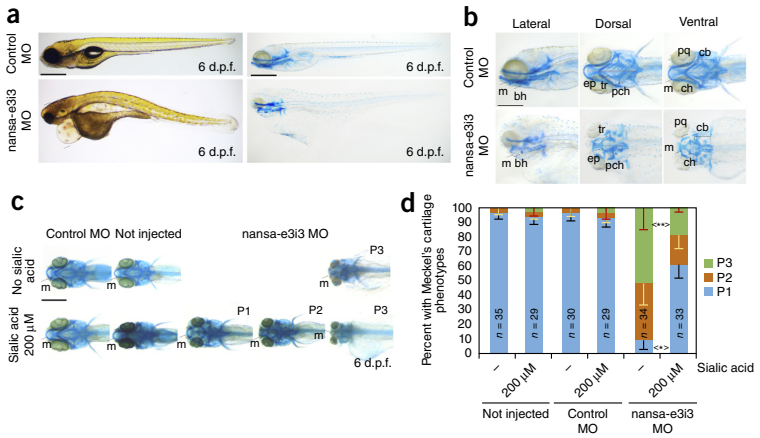


Figure 5 Abnormal skeletal development in zebrafish with morpholino-mediated knockdown of *nansa* is partially rescued by exogenous sialic acid. (**a**,**b**) nansa-e3i3 morphants (4 ng/nl) display a small head, pericardial edema, and abnormal cartilage and skeleton development, including hypoplastic or absent Meckel's cartilage (m) and lack of basihyal (bh) together with shortened and abnormal ethmoid plate (ep), trabecula (tr), parachordal (pch), palatoquadrate (pq), and ceratobranchial (cb) structures (dorsal and ventral views). ch, ceratohyal cartilage. Scale bars, 500 µm in **a** and 200 µm in **b**. (**c**) 200 µM sialic acid partially rescues the abnormal skeletal phenotype, as assessed by Meckel's cartilage measurement: P1, complete Meckel's cartilage; P2, incomplete Meckel's cartilage; P3, absent Meckel's cartilage. Scale bar, 200 µm. (**d**) The graph shows the proportion of fish with a Meckel's cartilage phenotype of P1, P2, or P3 with or without exogenously added sialic acid. For each experimental condition, approximately 10–12 embryos were analyzed (exact numbers are shown in the figure); experiments were carried out in triplicate. Error bars, s.d. from triplicate experiments. For *nansa* morphants, the difference between embryos with and without 200 µM sialic acid is statistically significant: $*P = 0.05$ for P1 embryos, $**P = 0.01$ for P3 embryos (two-tailed *t* test).

Interestingly, the rescue effect of sialic acid was dependent on the timing of its addition; rescue was observed when sialic acid was added to the embryo water right after MO injection but not when it was added 24 hours post-fertilization (h.p.f.), suggesting that sialic acid has a critical role in early embryonic development, potentially in cartilage and skeleton cell lineage specification or growth.

DISCUSSION

We present evidence that biallelic deleterious mutations in *NANS* are associated with severe IDD and skeletal dysplasia. First, we identified ten different *NANS* variants in nine patients from six unrelated families, segregating according to a recessive disease model. Second, the *NANS* mutations impaired the activity of the NeuNAc synthase enzyme, as evidenced by reduced enzyme activity and by the specific block of ManNAc analog incorporation in cultured cells. Third, dysfunction of NANS *in vivo* was confirmed by the accumulation of substrates of the missing enzyme, ManNAc in body fluids and ManNAc-6-phosphate in cultured cells. Finally, inactivation of the enzyme activity in zebrafish embryos resulted in a complex phenotype including abnormal development of skeletal structures. The conclusion that *NANS* mutations are the cause of the clinical phenotype is strengthened by the fact that it was reached by two independent approaches. Patients 1–8 were ascertained because of their phenotype of intellectual disability and specific skeletal dysplasia, and the genomic approach led to identification of *NANS* mutations and to their validation through RNA studies. Patient 9 was subjected to metabolomic screening first to elucidate the cause of severe IDD and dysmorphisms, leading to the identification of ManNAc in body fluids, and this biochemical phenotype allowed us to prioritize *NANS* mutations as the most likely to be pathogenic among the variants found subsequently through exome sequencing. Although most of our patients were ascertained retrospectively and the clinical assessment is heterogeneous, there are indications of different degrees of clinical severity. The study of additional patients is needed to determine the clinical spectrum of NANS deficiency and to establish possible genotype–phenotype correlations.

The brain contains the highest concentration of total sialic acid among human organs[18]. Sialic acid is present on glycoproteins and glycolipids such as the gangliosides, which are particularly abundant in nervous tissue, and NANS is highly expressed in the human brain (**Supplementary Fig. 3**). Genetic deficiency of either the sialyltransferase ST3GAL3 (MIM 604402) or ST3GAL5 (MIM 609056), two enzymes that use CMP-NeuNAc to add terminal sialic acid residues to the glycosidic antennae of glycoproteins and glycolipids, leads to infantile epilepsy and/or developmental arrest, suggesting that appropriately sialylated glycoproteins and/or glycolipids are necessary for higher brain functions[19–22]. Mutations affecting the CMP-NeuNAc transporter SLC35A1 (MIM 605634; see **Fig. 4**) result in developmental disability with ataxia and bleeding diathesis[23]. The observation of IDD in NANS-deficient patients and the brain dysplasia observed in two of them underline the relationship between sialylation and neurological functions and suggest that the requirements for sialic acid in the developing brain must be met at least partially by endogenous synthesis of sialic acid through the NANS pathway.

The short stature and skeletal dysplasia in NANS-deficient individuals also indicate that NANS-mediated sialic acid synthesis has a pivotal role in skeletal development, specifically in growth plate cartilage. The skeletal anomalies seen in the zebrafish knockdown model support this notion. The avascular nature of cartilage may make it dependent on endogenous synthesis. Several of the key players in cartilage and bone growth and development, such as chondroitin sulfate proteoglycans[24], bone sialoprotein[25], and osteopontin[26], are heavily sialylated and are candidates for further studies.

Tests detecting hyposialylated transferrin and apolipoprotein C-III yielded normal results in our NANS-deficient patients, and there was no clinical or laboratory evidence of hyposialylation of plasma proteins or clotting factors, suggesting that sialylation of plasma proteins is not greatly affected. How sialylation is achieved with impaired NANS activity is unclear. Several lines of explanation can be considered. First, the mutations in our patients may allow for some residual activity; none of our patients had a combination of two bona fide null alleles. Second, endogenous NeuNAc synthesis may be rate limiting in certain tissues and at certain times, where and when synthetic requirements are maximal (such as in the brain, during periods of rapid growth before birth and in the first 2 years of life[27–29], or in cartilage during infancy and childhood) but not in other tissues where

synthetic requirements for sialic acid may be lower. Such a mechanism has been put forward to explain the muscle-restricted phenotype of GNE myopathy (MIM 603824)[30], a disorder caused by recessive mutations in the *GNE* gene encoding UDP-*N*-acetylglucosamine 2-epimerase and ManNAc kinase activities in the sialic acid synthesis pathway (**Fig. 4**). Third, some tissues may be able to rescue and recycle sialic acid derived from the lysosomal breakdown of sialylated macromolecules (**Fig. 4**). A fourth possibility, not mutually exclusive with the others, is that nutrition-derived sialic acids may be entered into biosynthetic pathways and that this may occur in some tissues, such as the liver, but not or to a lesser extent in brain or cartilage. The observation of normal levels of free sialic acid in the urine of our NANS-deficient patients also indicates that there is no systemic depletion of sialic acid.

The role of nutrition-derived sialic acid raises the question of a possible treatment with oral sialic acid in NANS-deficient individuals, in analogy to other glycosylation and sialylation defects, such as CDG1B (MIM 602579) and GNE myopathy[31], which have been amenable to treatment by oral administration of specific sugars. Some of our data may point in this direction. When added to patient-derived cells in culture, the sialic acid analog NeuNPoc (used because of its detectability) but not the upstream metabolite ManNPoc was able to bypass the enzymatic block and be incorporated into macromolecules. In zebrafish embryos, exogenously added sialic acid was able to partially recue the developmental phenotype caused by *nansa* knockdown. It is possible that dietary supplementation with sialic acid could be beneficial for NANS-deficient patients. There is evidence that free sialic acid can be taken up and metabolized by cultured cells[32]; this could occur through pinocytosis and release to the cytoplasm by the lysosomal sialic acid exporter (SLC17A5; **Fig. 4**). In young mice and rats, sialic acid injected in the peritoneum is incorporated into macromolecules[33], and orally administered free sialic acid is found in plasma and, subsequently, in the liver and the brain[29,33]. *N*-glycolylneuraminic acid, a sialic acid analog that is widespread in mammals and apes but absent in man because of an evolutionary mutation, is found in human tissues as the result of dietary uptake from meat products[34,35]. Altogether, there is evidence suggesting that alimentary sialic acids could potentially be taken up and incorporated into biosynthetic pathways in human. The relative contributions of endogenously synthesized, nutritionally derived, and rescued sialic acid in different tissues and at different developmental stages in man remain to be investigated and may explain how the consequences of NANS deficiency are restricted to the developing brain and cartilage. In particular, if the brain pathology in NANS deficiency occurs in the first months of life (and perhaps even prenatally), this might be an obstacle to efficient treatment. Extensive studies in cell culture and *in vivo* are needed before envisaging any treatment possibility for NANS-deficient individuals. Such studies would also have the potential of clarifying the role of nutritional sialic acid in human.

Human milk contains a high concentration of free oligosaccharides, most of which are sialylated[27,28,36,37], as well as free sialic acid. These oligosaccharides, which are notably absent from cow's milk and infant formulas, have been attributed numerous functions, including stimulating brain development and cognition in infants[27,28,37,38]. Because oral administration of sialic acid in humans is considered safe and is well tolerated, nutritional supplementation with sialic acid in infancy, gestation, and advanced age has been proposed[27–29,37]. In view of the role of sialic acid and its potential use in nutrition, exploring the pathogenesis of brain dysfunction and skeletal dysplasia induced by NANS deficiency is worthwhile, not only because of the goal of elaborating therapeutic approaches for NANS-deficient individuals but also to shed light on sialic acid metabolism with its implications for human health.

URLs. Exome Aggregation Consortium (ExAC) database (accessed January 2016), http://exac.broadinstitute.org/; Wellderly database at the Scripps Wellderly Genome Resource, Scripps Wellderly Study (La Jolla, California) (accessed January 2016), https://genomics.scripps.edu/browser/files/wellderly/vcf/; PolyPhen-2 software for mutation assessment, http://genetics.bwh.harvard.edu/pph2/; Provean software for mutation assessment, http://provean.jcvi.org/index.php; Ensembl database, http://www.ensembl.org/; Novoalign software, http://www.novocraft.com/; Gaslini Cell Repository and Biobank, http://dppm.gaslini.org/biobank/; Leenaards Foundation in Lausanne, http://www.leenaards.ch/; Treatable Intellectual Disability Endeavor in British Columbia: First Collaborative Area of Innovation, http://www.tidebc.org/; Rare Diseases Models and Mechanisms Network, http://www.rare-diseases-catalyst-network.ca/index.php; Genome Analysis Toolkit (http://www.broadinstitute.org/gatk/); Enzyme Commission (EC) database, http://enzyme.expasy.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank M. Filocamo at the Gaslini Biobank (Genoa, Italy) for a fibroblast line for patient 1. We thank A. Reymond (CIG, FBM, Université de Lausanne) and his laboratory for lymphocyte immortalization. We thank C. Chiesa for Sanger sequencing and sample handling and shipment; S. de Boer for excellent technical assistance; B. Toh at the University of British Columbia for metabolic sample handling; X. Han for Sanger sequencing; B. Sayson for consenting and data management; M. Higginson for DNA extraction and sample handling; and A. Ghani for administrative assistance. We also thank R. Houben for skillfully preparing **Figure 4** and A. Bandi for all other figures. We are grateful to our clinical colleagues in Dolo, Genoa, Lausanne, Manchester, Paris, Reggio Emilia, Tokyo, Treviso, and Vancouver for patient management. A.S.-F. dedicates this paper to the memory of Paolo Durand who pointed out the relationship between sialic acid metabolism and IDD to him in 1980. Finally, we wish to thank the patients reported here as well as their parents for the enthusiasm they showed for our research efforts, for their patience, which was challenged by the studies lasting many years, and for their repeated donation of biological samples. They have been the source of continuous motivation for us.

This work has been supported by funding from the Leenaards Foundation in Lausanne, Switzerland; the Faculty of Biology and Medicine of the University of Lausanne; the BC Children's Hospital Foundation (Treatable Intellectual Disability Endeavour in British Columbia: First Collaborative Area of Innovation); Genome BC (grant SOF-195); the Rare Diseases Foundation; the Rare Diseases Models and Mechanisms Network; the Canadian Institutes of Health Research (grant 301221); and the Dutch Organization for Scientific Research, ZONMW (Medium Investment Grant 40-00506-98-9001 and VIDI Grant 91713359 to D.J.L.). The zebrafish studies were supported by funding to X.-Y.W. from the Canadian Rare Disease Models and Mechanisms Network, the Brain Canada Foundation, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canada Foundation for Innovation (CFI). The informatics infrastructures were supported by Genome BC and Genome Canada (ABC4DE Project) as well as by the Vital-IT Project of the Swiss Institute of Bioinformatics (SIB; Lausanne, Switzerland). C.J.R. is funded by a Canadian Institutes of Health Research New Investigator Award. C.D.M.v.K. is a recipient of the Michael Smith Foundation for Health Research Scholar Award (Vancouver, Canada). E.G. is supported by a Marie Skłodowska-Curie fellowship (MSCA-IF-661491).

AUTHOR CONTRIBUTIONS

C.D.M.v.K., L.B., S. Unger, R.A.W., and A.S.-F. conceived the study and coordinated and supervised the different teams. A.S.-F., L.B., C.D.M.v.K., T.D., A.C., M.I.V.A., C.J.R., J.H., L.G., L.T., V.C., D.H., D.D., C.B., I.M., and S. Uchikawa recruited the patients, reviewed the clinical and radiographic features, and obtained biological

materials from patients. A.S.-F., S. Unger, and G.N. reviewed the radiographic data. J.R. performed the bone marrow studies. Andrea Rossi reviewed the cerebral imaging. K. Harshman, B.J.S., B.C.-X., S.B., B.R.-B., H.R., C.R., M.T.-G., W.W.W., and A.d.B. were responsible for exome sequencing, haplotype reconstruction, Sanger sequencing, database studies, and mRNA–cDNA studies. R.A.W., L.A.J.K., E.v.d.H., and U.F.E. performed the metabolomics studies. Antonio Rossi studied ManNAc incorporation in fibroblasts. T. Hennet performed the lectin binding studies. A.A., K. Huijben, F.Z., and D.J.L. performed the NANS enzyme assays. D.J.L., A.A., T. Heisse, and T.B. studied the incorporation of sialic acid precursors in lymphocytes and fibroblasts. E.G. and G.S.-F. obtained the NANS three-dimensional model and mapped the affected amino acid residues. X.-Y.W. and K.B.-A. generated and phenotyped the zebrafish model. C.D.M.v.K. and A.S.-F. prepared the manuscript with contributions from all co-authors. All co-authors edited and reviewed the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Salvador-Carulla, L. *et al.* Intellectual developmental disorders: towards a new name, definition and framework for "mental retardation/intellectual disability" in ICD-11. *World Psychiatry* **10**, 175–180 (2011).
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 5th edn (American Psychiatric Publishing, 2013).
3. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
4. Gilissen, C. *et al.* Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**, 344–347 (2014).
5. van Karnebeek, C.D.M. & Stockler, S. Treatable inborn errors of metabolism causing intellectual disability: a systematic literature review. *Mol. Genet. Metab.* **105**, 368–381 (2012).
6. Bonafé, L. *et al.* Nosology and classification of genetic skeletal disorders: 2015 revision. *Am. J. Med. Genet.* **167**, 2869–2892 (2015).
7. Superti-Furga, A., Bonafé, L. & Rimoin, D.L. Molecular-pathogenetic classification of genetic disorders of the skeleton. *Am. J. Med. Genet.* **106**, 282–293 (2001).
8. Camera, G., Camera, A., Di Rocco, M. & Gatti, R. Sponastrime dysplasia: report on two siblings with metal retardation. *Pediatr. Radiol.* **23**, 611–614 (1993).
9. Geneviève, D. *et al.* Exclusion of the dymeclin and PAPSS2 genes in a novel form of spondyloepimetaphyseal dysplasia and mental retardation. *Eur. J. Hum. Genet.* **13**, 541–546 (2005).
10. Tarailo-Graovac, M. *et al.* Exome sequencing and the management of neurometabolic disorders. *N. Engl. J. Med.* (in the press).
11. Cotton, T.R., Joseph, D.D., Jiao, W. & Parker, E.J. Probing the determinants of phosphorylated sugar-substrate binding for human sialic acid synthase. *Biochim. Biophys. Acta* **1844**, 2257–2264 (2014).
12. Galuska, S.P. *et al.* Quantification of nucleotide-activated sialic acids by a combination of reduction and fluorescent labeling. *Anal. Chem.* **82**, 4591–4598 (2010).
13. Büll, C. *et al.* Sialic acid glycoengineering using an unnatural sialic acid for the detection of sialoglycan biosynthesis defects and on-cell synthesis of siglec ligands. *ACS Chem. Biol.* **10**, 2353–2363 (2015).
14. Riemersma, M. *et al.* Disease mutations in CMP–sialic acid transporter *SLC35A1* result in abnormal α-dystroglycan O-mannosylation, independent from sialic acid. *Hum. Mol. Genet.* **24**, 2241–2246 (2015).
15. Link, V., Shevchenko, A. & Heisenberg, C.P. Proteomics of early zebrafish embryos. *BMC Dev. Biol.* **6**, 1 (2006).
16. Eisen, J.S. & Smith, J.C. Controlling morpholino experiments: don't stop making antisense. *Development* **135**, 1735–1743 (2008).
17. Javidan, Y. & Schilling, T.F. Development of cartilage and bone. *Methods Cell Biol.* **76**, 415–436 (2004).
18. Wang, B. & Brand-Miller, J. The role and potential of sialic acid in human nutrition. *Eur. J. Clin. Nutr.* **57**, 1351–1369 (2003).
19. Simpson, M.A. *et al.* Infantile-onset symptomatic epilepsy syndrome caused by a homozygous loss-of-function mutation of GM3 synthase. *Nat. Genet.* **36**, 1225–1229 (2004).
20. Hu, H. *et al. ST3GAL3* mutations impair the development of higher cognitive functions. *Am. J. Hum. Genet.* **89**, 407–414 (2011).
21. Fragaki, K. *et al.* Refractory epilepsy and mitochondrial dysfunction due to GM3 synthase deficiency. *Eur. J. Hum. Genet.* **21**, 528–534 (2013).
22. Boccuto, L. *et al.* A mutation in a ganglioside biosynthetic enzyme, *ST3GAL5*, results in salt & pepper syndrome, a neurocutaneous disorder with altered glycolipid and glycoprotein glycosylation. *Hum. Mol. Genet.* **23**, 418–433 (2014).
23. Mohamed, M. *et al.* Intellectual disability and bleeding diathesis due to deficient CMP–sialic acid transport. *Neurology* **81**, 681–687 (2013).
24. Roughley, P.J., White, R.J. & Santer, V. Comparison of proteoglycans extracted from high and low weight–bearing human articular cartilage, with particular reference to sialic acid content. *J. Biol. Chem.* **256**, 12699–12704 (1981).
25. Vincent, K. & Durrant, M.C. A structural and functional model for human bone sialoprotein. *J. Mol. Graph. Model.* **39**, 108–117 (2013).
26. Sodek, J., Ganss, B. & McKee, M.D. Osteopontin. *Crit. Rev. Oral Biol. Med.* **11**, 279–303 (2000).
27. Wang, B. Sialic acid is an essential nutrient for brain development and cognition. *Annu. Rev. Nutr.* **29**, 177–222 (2009).
28. Wang, B. Molecular mechanism underlying sialic acid as an essential nutrient for brain development and cognition. *Adv. Nutr.* **3**, 465S–472S (2012).
29. Sprenger, N. & Duncan, P.I. Sialic acid utilization. *Adv. Nutr.* **3**, 392S–397S (2012).
30. Salama, I. *et al.* No overall hyposialylation in hereditary inclusion body myopathy myoblasts carrying the homozygous M712T *GNE* mutation. *Biochem. Biophys. Res. Commun.* **328**, 221–226 (2005).
31. Malicdan, M.C., Noguchi, S., Hayashi, Y.K., Nonaka, I. & Nishino, I. Prophylactic treatment with sialic acid metabolites precludes the development of the myopathic phenotype in the DMRV-hIBM mouse model. *Nat. Med.* **15**, 690–695 (2009).
32. Oetke, C. *et al.* Evidence for efficient uptake and incorporation of sialic acid by eukaryotic cells. *Eur. J. Biochem* **268**, 4553–4561 (2001).
33. Nöhle, U. & Schauer, R. Uptake, metabolism and excretion of orally and intravenously administered, ^{14}C- and ^{3}H-labeled *N*-acetylneuraminic acid mixture in the mouse and rat. *Hoppe-Seyler's Z. Physiol. Chem.* **362**, 1495–1506 (1981).
34. Tangvoranuntakul, P. *et al.* Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proc. Natl. Acad. Sci. USA* **100**, 12045–12050 (2003).
35. Samraj, A.N. *et al.* A red meat–derived glycan promotes inflammation and cancer progression. *Proc. Natl. Acad. Sci. USA* **112**, 542–547 (2015).
36. Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**, 1147–1162 (2012).
37. Fuhrer, A. *et al.* Milk sialyllactose influences colitis in mice through selective intestinal bacterial colonization. *J. Exp. Med.* **207**, 2843–2854 (2010).
38. Röhrig, C.H., Choi, S.S. & Baldwin, N. The nutritional role of free sialic acid, a human milk monosaccharide, and its application as a functional food ingredient. *Crit. Rev. Food Sci. Nutr.* doi:10.1080/10408398.2015.1040113 (2016).

ONLINE METHODS

Ethics. The studies were approved by the ethics boards of the following institutions: BC Children's and Women's Hospital, University of British Columbia (12-00067) and the Ethics Board of the Lausanne University Hospital. Research was performed according to the countries' ethics code of conduct. Parents or guardians provided written informed consent for the biochemical and genetic analysis and the publication of photographs and clinical data.

Identification of mutations in *NANS*. For patients 1–4, fragmented genomic DNA was purified with AMPure XP beads, and its quality was assessed with an Agilent Bioanalyzer. Preparation of the exome-enriched, barcoded sequencing libraries was performed using the SureSelect Human All Exon v4 kit (Agilent Technologies). The final libraries were quantified with a Qubit Fluorometer (Life Technologies), and the correct size distribution was validated on an Agilent Bioanalyzer. Libraries were then sequenced on a HiSeq 2000 machine (Illumina), generating 100-bp paired-end reads. Raw reads were aligned onto the hg19 reference genome using Novoalign software. Data cleanup and variant calling were performed according to GATK Best-Practices recommendations[39]. Variant filtering was performed with ANNOVAR[40] and with in-house Perl and Bash scripts, available upon request. For patient 9, who was genotyped in the Vancouver laboratory, similar procedures were used. To investigate the presence of a possible ancestral haplotype carrying the *NANS* indel in patients 1–4, we extracted all SNP alleles that were present in the region surrounding *NANS* and that were listed in the dbSNP database. Local genotypes were then scored, and likely haplotypes were constructed. In patients 5–8, all individual exons of *NANS* were amplified from genomic DNA (primer sequences available upon request) and sequenced directly in both directions using the Sanger method. For patient 9, exome sequencing was performed as part of the TIDEX gene discovery project using the Agilent SureSelect kit and an Illumina HiSeq 2000 instrument. The sequencing reads were aligned to human reference genome version hg19 (35× coverage), and rare variants were identified and assessed for their potential to disrupt protein function. In total, we identified 19 candidate genes affected by two rare heterozygous variants. Of these, *NANS* stood out as the most interesting functional candidate because of the preexisting next-generation metabolomics data. Sanger sequencing in the patient and his parents confirmed the mutations and their segregation with disease.

Construction of a three-dimensional protein model for NANS and mapping of the predicted mutations. A molecular model for dimeric full-length human NANS protein was generated with I-TASSER[41], using as templates the *N. meningitidis* homolog[42] and the human AFPL domain[43] structures. The model illustrations were generated with the PyMol Molecular Graphics System, version 1.7.4 (Schrödinger).

mRNA studies. Lymphoblastoid cell lines were cultured in suspension under 5% CO_2 in T25 flasks with RPMI-1640 medium supplemented with GlutaMAX-I (Gibco) containing 10% FBS (Gibco, 10270-106) and 1% penicillin-streptomycin (Gibco), whereas fibroblasts were cultured in DMEM (1×) with 1 g/L D-glucose L-glutamine (Gibco), supplemented with 10% FBS and 1% penicillin-streptomycin. To test for NMD, treatment with cycloheximide (Sigma-Aldrich) was performed in parallel with controls by incubating 10 million cells for 4 h in the presence of medium supplemented with 28 µg/ml of this chemical, according to published protocols[44]. Total RNA was isolated from both cycloheximide-treated and untreated fibroblast and lymphoblast cultures using the Direct-zol RNA MiniPrep kit (Zymo Research) according to the manufacturer's instructions. cDNA was prepared following the retrotranscription of 500 ng of RNA, using the PrimerScript RT–PCR kit (Clontech) and random hexamers; 10 ng of the produced cDNA was then used as template for downstream experiments. A specific primer pair spanning the exon–exon junctions for exons 1 and 2 and exons 5 and 6 (for sequences, see **Supplementary Table 5**) was designed to amplify the regions of the *NANS* cDNA containing all mutations studied. RT–PCR was performed in triplicate in a final reaction volume of 20 µl containing 5× Green GoTaq reaction buffer (Promega), 100 µM dNTP mix, 200 nM of each primer, and 0.1 U of GoTaq G2 DNA Polymerase (Promega). Reactions were incubated at 94 °C for 1 min followed by 35 cycles at 93 °C for 20 s, 64.1 °C for 30 s, and 72 °C for 1 min. The obtained products were resolved by capillary electrophoresis with the eGene HDA-GT12 Multi-Channel Genetic Analyzer, quantified, and finally ligated into the pCRII-TOPO TA vector (Invitrogen). Ligation mixes were used to transform chemically competent TOP-10 *Escherichia coli* (Invitrogen), and individual clones (at least 30 clones per electrophoresed sample) were sequenced by direct Sanger sequencing using BigDye terminator v1.1 (Applied Biosystems) with insert-specific primers. Sequencing data were analyzed using CLC Bio software (Qiagen) and compared with the corresponding human reference sequence (build hg19).

Next-generation metabolomics analysis. High-resolution untargeted metabolomics analysis of body fluids was performed using UHPLC–QTOF mass spectrometry. CSF and heparinized plasma samples were deproteinated in methanol:ethanol solution (50:50; 100 µl of each sample plus 400 µl of methanol:ethanol solution). Samples were analyzed in duplicate. A 2-µl sample was applied to an Acquity HSS T3 reverse-phase column (100 × 2.1 mm; 100 Å, 1.8 µm), and an Agilent 6540 UHD accurate mass UHPLC–QTOF mass spectrometer with acquisition in positive and negative modes was used. The buffers in positive mode consisted of buffer A (0.1 formic acid in water) and buffer B (0.1 formic acid in water:methanol solution (1:99)); in negative mode, the buffers consisted of buffer A (10 mM acetic acid) and buffer B (10 mM acetic acid in water:methanol solution (1:99)). After analysis and XCMS alignment, bioinformatics software developed in house showed the features (exact *m/z*, retention time, and intensity) deriving from metabolites that were significantly different in intensity in the patient sample as compared to age- and sex-matched controls. The Human Metabolome Database 3.0 was used to putatively annotate significantly different features[45]. This technique, also referred to as 'next-generation metabolic screening', was clinically validated using body fluids from patients for 25 known inborn errors of metabolism and introduced in the Nijmegen patient care research setting.

Proton (^{1}H) NMR spectroscopy of body fluids, fibroblast homogenates, and model compounds of N-acetylated sugars was performed on a 500-MHz NMR spectrometer with minor modifications[46]. The NMR spectrum for fibroblasts was recorded on a homogenate obtained after sonicating 7.5×10^6 cells in ^2H_2O (D_2O). The homogenate was deproteinated over a 10 kDa filter and trimethylsilyl-2,2,3,3-tetradeutero propionic (TSP) acid was used as a chemical shift reagent in the NMR spectrum. The model compounds ManNAc, *N*-acetylgalactosamine, and *N*-acetylglucosamine were purchased from Sigma Chemicals, and ManNAc-6-phosphate was purchased from Carbosynth (disodium salt).

Lectin staining on cultured fibroblasts. Fibroblasts were cultured for 48 h in DMEM supplemented with 10% FCS and 0, 1, or 10 mM ManNAc. After incubation, fibroblasts were trypsinized, washed twice in Hank's buffered saline solution (HBSS) containing 1% FCS, and resuspended at 1×10^6 cells/ml. Aliquots of 1×10^5 cells (in 100 µl) were incubated with 1 mM FITC-labeled *S. nigra* lectin (Vector Labs) in HBSS for 20 min on ice, washed twice in HBSS containing 1% FCS, and analyzed by flow cytometry using a FACSCanto II cytometer (BD Biosciences).

***N*-acetylmannosamine incorporation in fibroblast culture.** Skin fibroblasts were cultured and expanded in DMEM supplemented with 10% FCS at 37 °C in 5% CO_2 and were then cultured for 24 h in the absence of FCS. For the incorporation experiment, mycoplasma-negative cells were incubated in DMEM, without serum and antibiotics, with or without 10 mM ManNAc, at 37 °C in 5% CO_2 for 48 h. After incubation, the medium was removed, and cells were harvested in PBS and collected by low-speed centrifugation. The cell pellet was sonicated in buffer containing 20 mM Tris-HCl, pH 8.0, 5 mM EDTA, 150 mM NaCl, 1 µg aprotinin, and 1 mM PMSF, and the lysate was ultracentrifuged at 30,000*g* for 1 h at 4 °C to pellet the membrane fraction. In the clear supernatant, protein content was determined with the BCA protein assay (Pierce), and aliquots of the supernatant were ultrafiltered with an Amicon Ultra-0.5 Centrifugal Filter Unit with a cutoff of 3 kDa to separate sialic acid from soluble protein. The dried filtrate was used for sialic acid analysis. The CMP–sialic acid content was discriminated from free sialic acid by reduction: labeling with the fluorophore 1,2-diamino-4,5-methylenedioxybenzene (DMB) requires free keto as

well as carboxyl groups on the sialic acid molecule, and reduction of the keto group before the labeling process precludes the labeling of non-activated sialic acid[12]. Each lyophilized sample was split in two aliquots. To reduce free, non-activated monosaccharides, the dried samples were dissolved in 0.2 M sodium borate buffer, pH 8.0, containing 0.2 M sodium borohydride, incubated at 0 °C overnight, and dried in a SpeedVac concentrator. For fluorescent labeling of sialic acid, samples were hydrolyzed in 0.2 N trifluoroacetic acid (TFA) for 4 h at 80 °C, dried, redissolved twice in methanol, and dried again. To label sialic acids with DMB, hydrolysates were dissolved in 80 µl of 0.64 mg/ml DMB in 500 mM 2-mercaptoethanol, 9 mM sodium hydrosulfite, and 20 mM TFA and incubated for 2 h at 56 °C. Reactions were stopped by adding 10 µl of 0.2 M NaOH. The derivatized sialic acids were quantified by HPLC on a binary pump system (1525 Binary HPLC Pump, Waters) coupled to a fluorescence detector (2475 Multy λ Fluorescence Detector, Waters) set at an excitation wavelength of 372 nm and an emission wavelength of 456 nm. Chromatography was carried out at room temperature with a LichroCART 250-4 Superspher 100 RP18 (250 × 4 mm) column (Merck) and LiChrospher 100 RP18 (25 × 4 mm) (Merck) as a pre-column. Mobile phases were methanol:acetonitrile:water:TFA (4:4:92:0.1) and methanol:acetonitrile:water:TFA (45:45:10:0.1), and the flow rate was 0.3 ml/min[12].

Measurement of the *N*-acetylneuraminic acid content in fibroblasts. Fibroblasts (~2.5 million cells) in 250 µl of 50 mM Tris-HCl, pH 7.5, were sonicated on ice (3 × 8 s) and then centrifuged (10,000*g* for 10 min at 4 °C). To 100 µl of the supernatant (lysate) was added [3-^{13}C]NeuNAc (±23 µM) as an internal standard, and the solution was applied to a 30-kDa filter cup (Amicon Ultra) with 10 µl of 2% formic acid in the collection tube for deproteination. After centrifugation at 13,000*g* for 30 min at 4 °C, the flow-through was used for quantification of NeuNAc by mass spectrometry according to published methods[47,48]. Assays were performed in duplicate, and NeuNAc levels were normalized to protein levels in the lysates.

Determination on NANS activity in fibroblast cultures. Skin fibroblasts were obtained from affected individuals and healthy controls and cultured at 37 °C under 5% CO_2 in culture medium E199, supplemented with 10% FCS and 1% penicillin-streptomycin. All cultures were tested for mycoplasma infection before growth in culture. We developed an assay for NANS activity as follows: fibroblasts (~7.5 million cells) in 300 µl of 50 mM Tris-HCl, pH 7.5, were sonicated on ice (3 × 8 s) and then centrifuged (10,000*g* for 10 min at 4 °C). The supernatant (lysate) was used for the determination of protein concentration (BCA assay) and for enzyme activity assays. Incubation of ManNAc-6-phosphate disodium salt (Carbosynth; 2 mM) and 20 µl of lysate was carried out at 37 °C in duplicate in a total reaction volume of 100 µl. The reaction buffer consisted of 50 mM Tris-HCl, pH 7.5, 3 mM PEP (Roche), and 1 mM $MgCl_2$. Control incubations without ManNAc-6-phosphate were subtracted as a blank from the incubations. The reactions were allowed to proceed for 2.5, 5, and 24 h and then stopped by freezing (−20 °C). Samples (100 µl) were deproteinated by the addition of 50 µl of [3-^{13}C]NeuNAc (±23 µM) as an internal standard and centrifuged at 13,000*g* for 30 min at 4 °C on a 30-kDa filter cup (Amicon Ultra); the flow-through was collected in 10 µl of 2% formic acid. NeuNAc levels were quantified by mass spectrometry as described[47,48] and normalized for protein level in the lysates.

Metabolic labeling of sialic acids and glycoproteins in fibroblast cultures. Skin-derived fibroblasts were cultured in M199 medium (PAN biotech) supplemented with 10% FBS (PAA), non-essential amino acids (NEAA; Gibco), and 100 U/ml penicillin-streptomycin (Gibco) and tested for mycoplasma contamination. Eighty percent confluent cells were further grown for 5 d in medium containing 15 µM Ac5NeuNAc, Ac4ManNPoc, or Ac5NeuNPoc. Cells were collected in PBS by scraping and lysed in 150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 5 mM EDTA, 0.1% SDS, 1% Triton X-100, and 1× Complete protease inhibitor cocktail (Roche). To biotinylate propargyloxycarbonyl-containing glycoproteins, cell lysates were resolved on 10% SDS gels, blotted onto PVDF membrane. Unoccupied membrane sites were blocked with 5% dry milk in PBS, and the membrane was incubated with 500 µM $CuSO_4$, 250 µM l-histidine, 100 µM azide-PEG3-biotin, and 500 µM sodium ascorbate in PBS for 1 h at 37 °C. Biotinylated sialoglycoproteins were visualized by incubation with horseradish peroxidase (HRP)-conjugated streptavidin (GE Healthcare), followed by signal development with SuperSignal West Femto Maximum-Sensitivity Substrate (Thermo Fisher Scientific). As a loading control, immunostaining was performed with antibody to GAPDH (Abcam, ab8245) on the same membranes. On the basis of this protocol, we found that NeuNPoc at 15 µM concentration would readily label cellular sialoglycans after 5 d of incubation. For ManNPoc at the same concentration of 15 µM, longer incubation times were required to obtain strong labeling of glycoconjugates in fibroblasts from healthy donors (**Fig. 5**).

Determination of glycosylation of plasma proteins. Analysis of plasma transferrin N-glycosylation was carried out by isoelectric focusing[49] as well as by nanochip QTOF mass spectrometry[50]. The mucin-type O-linked glycosylation of plasma apolipoprotein C-III was analyzed by isoelectric focusing[49].

Zebrafish studies. Zebrafish were maintained at 28.5 °C on a 10-h dark/14-h light cycle. Protocols for experimental procedures were approved by the Research Ethics Board of St. Michael's Hospital (Toronto) (protocol ACC660). To knock down gene expression, we used splicing-blocking MOs for *nansa* and *nansb* knockdown plus a standard control MO (for sequences, see **Supplementary Table 5**). The zebrafish *nansa* gene has six exons, with the translation start codon in exon 1 (ENSDART00000067086; chromosome 1). We designed a splicing-blocking MO nansa-e3i3 that could block the splicing donor of exon 3, resulting in intron 3 retention and a truncated protein in translation (data not shown). *nansb* also has six exons, with the translation start codon in exon 1 (ENSDART00000169540; chromosome 25), and we designed an MO nansb-e1i1 to block the splicing donor site of exon 1 of *nansb*, resulting in intron 1 retention and a truncated Nansb protein (data not shown). MOs were synthesized by Gene Tools (for sequences, see **Supplementary Table 5**) and microinjected individually or in combination into zebrafish embryos at the one-cell stage. We injected embryos individually with nansa-e3i3 (4 ng/nl) and nansb-e1i1 (4 ng/nl) MOs, and also coinjected embryos with nansa-e3i3 (4 ng/nl) and nansb-e1i1 (4 ng/nl). Each injection was repeated at least three times. Knockdown of *nansa* and *nansb* was assessed by RT–PCR to confirm splicing defect and retention of the intron (data not shown). For control MO, embryos were injected at the one-cell stage with 4 ng/nl. At least 30 embryos were injected per condition and included in the analysis. At 24 h.p.f., embryos were manually dechorionated. Total RNA was extracted from embryos at 48 h.p.f. using TRIzol reagent (Invitrogen). The RNA concentration of each sample was determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). RNA integrity was verified by 1% agarose gel electrophoresis. The RNA template was converted into cDNA using Superscript II reverse transcriptase (Invitrogen) and used to amplify a *nansa*-specific cDNA (for primer sequences, see **Supplementary Table 5**).

To visualize the cartilaginous structures, Alcian blue (Sigma) was dissolved in 70% ethanol and 1% hydrochloric acid. Zebrafish embryos (6 d.p.f.) were fixed in 4% paraformaldehyde overnight at 4 °C and maintained in 100% methanol at −20 °C until processing. The embryos were washed with PBS containing 0.1% Tween-20 (PBST). The embryos were bleached in 30% hydrogen peroxide for 2 h, washed with PBST, and transferred into Alcian blue solution. Embryos were stained overnight at room temperature. The embryos were rinsed four times with acidified ethanol (HCl-EtOH); 5% hydrochloric acid and 70% ethanol. Embryos were rinsed for 20 min in HCl-EtOH and rehydrated by washing 10 min in an HCl-EtOH/H_2Od series (75%, 25%, 50%, 50%, 25%, and 75%) and 100% H_2Od. Embryos were stored in 1 ml of glycerol-KOH.

39. Van der Auwera, G.A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 1–33 (2013).
40. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
41. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
42. Gunawan, J. *et al.* Structural and mechanistic analysis of sialic acid synthase NeuB from *Neisseria meningitidis* in complex with Mn^{2+}, phosphoenolpyruvate, and *N*-acetylmannosaminitol. *J. Biol. Chem.* **280**, 3555–3563 (2005).
43. Hamada, T. *et al.* Solution structure of the antifreeze-like domain of human sialic acid synthase. *Protein Sci.* **15**, 1010–1016 (2006).

44. Rajavel, K.S. & Neufeld, E.F. Nonsense-mediated decay of human *HEXA* mRNA. *Mol. Cell. Biol.* **21**, 5512–5519 (2001).
45. Wishart, D.S. *et al.* HMDB 3.0—the Human Metabolome Database in 2013. *Nucleic Acids Res.* **41**, D801–D807 (2013).
46. Engelke, U.F. *et al.* NMR spectroscopic studies on the late onset form of 3-methylglutaconic aciduria type I and other defects in leucine metabolism. *NMR Biomed.* **19**, 271–278 (2006).
47. Valianpour, F., Abeling, N.G., Duran, M., Huijmans, J.G. & Kulik, W. Quantification of free sialic acid in urine by HPLC–electrospray tandem mass spectrometry: a tool for the diagnosis of sialic acid storage disease. *Clin. Chem.* **50**, 403–409 (2004).
48. van der Ham, M. *et al.* Liquid chromatography–tandem mass spectrometry assay for the quantification of free and total sialic acid in human cerebrospinal fluid. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **878**, 1098–1102 (2010).
49. Jansen, J.C. *et al.* CCDC115 deficiency causes a disorder of Golgi homeostasis with abnormal protein glycosylation. *Am. J. Hum. Genet.* **98**, 310–321 (2016).
50. van Scherpenzeel, M., Steenbergen, G., Morava, E., Wevers, R.A. & Lefeber, D.J. High-resolution mass spectrometry glycoprofiling of intact transferrin for diagnosis and subtype identification in the congenital disorders of glycosylation. *Transl. Res.* **166**, 639–649 (2015).

6.2 ASAH1

Peripheral osteolysis in adults linked to ASAH1 (acid ceramidase) mutations: A new presentation of Farber disease

Own contribution: Analyzed the WES of the patients and helped in the identification of the new disease-causing gene.

ARTHRITIS & RHEUMATOLOGY
Vol. 68, No. 9, September 2016, pp 2323–2327
DOI 10.1002/art.39659

BRIEF REPORT

Peripheral Osteolysis in Adults Linked to *ASAH1* (Acid Ceramidase) Mutations: A New Presentation of Farber's Disease

Luisa Bonafé,[1] Ariana Kariminejad,[2] Jia Li,[3] Beryl Royer-Bertrand,[3] Virginie Garcia,[4] Shokouholsadat Mahdavi,[5] Bita Bozorgmehr,[2] Ralph L. Lachman,[6] Lauréane Mittaz-Crettol,[1] Belinda Campos-Xavier,[1] Sheela Nampoothiri,[7] Sheila Unger,[1] Carlo Rivolta,[3] Thierry Levade,[4] and Andrea Superti-Furga[1]

Objective. **To establish a diagnosis and provide counseling and treatment for 3 adult patients from one family presenting with peripheral osteolysis.**

Methods. **Following clinical and radiographic assessment, exome sequencing, targeted gene resequencing, and determination of enzyme activity in cultured fibroblasts were performed.**

Results. **The proband (age 40 years) had a history of episodic fever and pain in childhood that subsided around puberty. He and 2 of his older sisters (ages 58 and 60 years, respectively) showed adult-onset progressive shortening of fingers and toes with redundancy of the overlying skin. Radiographs showed severe osteolysis of the distal radius and ulna, carpal bones, metacarpal bones, and phalanges. Sequencing of the known genes for recessively inherited osteolysis, *MMP2* and *MMP14*, failed to show pathogenic mutations. Exome sequencing revealed compound heterozygosity for mutations c.505T>C (p.Trp169Arg) and c.760A>G (p.Arg254Gly) in *ASAH1*, the gene coding for acid ceramidase. Sanger sequencing confirmed correct segregation in the family, and enzyme activity in fibroblast cultures from the patients was reduced to ~8% of that in controls, confirming a diagnosis of Farber's disease.**

Conclusion. **Our findings indicate that hypomorphic mutations in *ASAH1* may result in an osteoarticular phenotype with a juvenile phase resembling rheumatoid arthritis that evolves to osteolysis as the final stage in the absence of neurologic signs. This observation delineates a novel type of recessively inherited peripheral osteolysis and illustrates the long-term skeletal manifestations of acid ceramidase deficiency (Farber's disease) in what appear to be the oldest affected individuals known so far.**

While resorption of bone and cartilage structures is a frequent finding in inflammatory conditions, there are at least 3 forms of osteolysis that are genetically determined (1). The more common, dominant form, multicentric carpotarsal osteolysis with or without nephropathy (MIM ID #166300), is caused by heterozygous mutations in the *MAFB* gene; its pathogenesis remains unclear. In addition, there are recessively inherited forms that have considerable phenotypic overlap and are known by different names, including Torg syndrome or Winchester syndrome (MIM ID #277950) and nodulosis arthropathy osteolysis syndrome or multicentric osteolysis nodulosis arthropathy syndrome (MONA; MIM ID #259600). Two genes, *MMP2* and *MMP14* (the genes encoding matrix metalloproteinase 2 (MMP-2) and MMP-14, respectively), are responsible for the recessive forms (1). The pathogenesis is unclear but may involve activation and presentation of transforming growth factor by matrix components.

PATIENTS AND METHODS

Characteristics of the patients. We observed a family in which 3 of 8 siblings were ascertained to have severe periph-

Supported by the Faculté de Biologie et Médecine de l'Université de Lausanne, the SNF (grant 156260), and the Leenaards Foundation, Lausanne.

[1]Luisa Bonafé, MD, Lauréane Mittaz-Crettol, MD, Belinda Campos-Xavier, PhD, Sheila Unger, MD, Andrea Superti-Furga, MD: Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland; [2]Ariana Kariminejad, MD, Bita Bozorgmehr, MD: Kariminejad-Najmabadi Pathology and Genetics Centre, Tehran, Iran; [3]Jia Li, PhD, Beryl Royer-Bertrand, PhD, Carlo Rivolta, PhD: University of Lausanne, Lausanne, Switzerland; [4]Virginie Garcia, PhD, Thierry Levade, MD, PhD: Centre Hospitalier Universitaire de Toulouse, INSERM UMR1037, and Université Toulouse III Paul-Sabatier, Toulouse, France; [5]Shokouholsadat Mahdavi, MD: Tehran Medical Genetics Laboratory, Tehran Welfare Organization, Tehran, Iran; [6]Ralph L. Lachman, MD: International Skeletal Dysplasia Registry and University of California, Los Angeles; [7]Sheela Nampoothiri, MD: Amrita Institute of Medical Sciences and Research Center, Cochin, India.

Drs. Bonafé and Kariminejad contributed equally to this work.

Address correspondence to Andrea Superti-Furga, MD, Centre Hospitalier Universitaire Vaudois, Bugnon 21, 1011 Lausanne, Switzerland. E-mail: asuperti@unil.ch.

Submitted for publication December 10, 2015; accepted in revised form February 23, 2016.

Figure 1. Pedigree and clinical features in the proband and his sisters. **A**, Pedigree showing the proband (**arrow**) and his 2 affected sisters (solid symbols). Other family members were clinically unaffected (open symbols). *ASAH1* genotypes are shown below the symbols. WT = wild-type. **B**, Photograph of the arms and hands of the older sister, showing shortened forearms, swollen wrists, and shortened fingers with redundant skin. **C** and **D**, Photographs of the proband, showing shortened forearms and fingers and redundant skin. **Arrows** indicate a subcutaneous nodule over the distal radius. **E** and **F**, Photographs of the arms and hands of the second sister, showing similar findings. Ulnar deviation at the wrist and phalangeal subluxations is evident. **G**, Photograph of the left foot of the proband, showing shortening of the midfoot and toes.

eral osteolysis as adults. Negative results of mutation analysis of the *MMP2* and *MMP14* genes prompted exome sequencing, leading to the discovery of compound heterozygous mutations in *ASAH1*, the gene coding for acid ceramidase. Deficient activity of acid ceramidase is associated with Farber's disease, a lysosomal storage disorder with accumulation of the neutral lipid ceramide. The findings in this family illustrate the presentation of Farber's disease as peripheral osteolysis in adults.

The proband in this Iranian family came to the attention of one of the authors (AK) at age 40 years. He was aware of a condition that affected him and 2 of his older sisters, and asked for genetic counseling since he planned to have children. His parents were not related. According to his mother's report, he had had episodes of fever and pain beginning at the age of 3 years. The episodes would last ~10–15 days, during which time he could not walk because of the pain. A small number of subcutaneous nodules developed, first on the ears, then over the wrist, elbows, and knee joints. Hoarseness developed at the same time and progressed over the years. Subsequently, the episodes subsided, and he remembers having joint pain from the age of ~12 years. At age 40 years, he has limitation of movement in the knee joint and limited flexion of all toes. His hands and toes have progressively shortened, and the skin over the hands is redundant (Figure 1). There is hyperextensibility in the fingers and thumbs. He has a few inconspicuous and indolent subcutaneous nodules on the ears, wrists (Figure 1), elbows, and ankles. His neurologic and developmental phenotype is normal, and he is professionally active as a telephone operator.

A radiographic study revealed severe osteolysis of the distal radius and ulna, metacarpal bones, and phalanges (Figure 2). The feet were similarly affected. The hips and knees showed signs of arthritis but no erosions, and the spine was unremarkable.

Two of the proband's older sisters were seen as adults (ages 58 years and 60 years, respectively). They had a similar clinical phenotype with short hands and redundant skin (Figure 1). Disease onset and progression were generally similar to those in the proband, with onset in childhood of episodes of pain and fever and appearance of nodules on the ears. The progression of the disease was similar to that in the proband, with shortening of the fingers, joint pain, and the appearance of discrete nodules on the elbows and knees. Hoarseness was present but milder than in the proband. There were no neurologic signs or history thereof. There was no history of frequent pulmonary infections in the proband or in his affected sisters.

Genetic analysis. Genomic DNA was obtained from the proband, from 2 affected siblings and 5 unaffected siblings, and from their mother. Their father was no longer alive. Informed consent was obtained from all family members. The molecular studies were performed with the approval of the ethics board of Lausanne University Hospital. Sanger sequencing of polymerase chain reaction amplicons of the *MMP2*, *MMP14*, and *ASAH1* exons was carried out using standard methods. Exome sequencing on genomic DNA from the 3 affected individuals in the pedigree was performed essentially as previously described (2). Genomic DNA was amplified using a SureSelect Human All Exon kit version 4 (Agilent) on a

Figure 2. Radiographic features in the acid ceramidase–deficient individuals as well as in other individuals with osteolysis. **A** and **B**, Radiographs of the hands of the proband (**A**) and of his sister (**B**), showing severe osteolysis of the distal radius and ulna, and of the carpal elements. The metacarpal bones have a diaphyseal constriction and their distal extremities are pointed. The proximal phalanges show a peculiar "cupping" around the distal end of the metacarpal bones. The changes at the proximal interphalangeal joints are more severe, but slightly reminiscent of those seen in longstanding psoriatic arthritis. **C**, Radiographs of the feet of the proband, showing severe osteolysis of the distal metatarsal bones and phalanges. **D**, Radiographs of the hand of an Indian woman with severe osteolysis. Her clinical and radiographic features were similar to those seen in the family described here. It seems plausible that she also had acid ceramidase deficiency, but this could not be determined since a blood sample was not available. **E** and **F**, Radiographs of the hands of adults with *MMP2* mutations. The radiographic pattern is different from that seen in the family described here. The metacarpal bones and phalanges resemble "sucked sugar candy," they do not show a diaphyseal constriction, and the proximal phalanges do not show cupping on the distal end of the metacarpal bones.

HiSeq 2000 machine (Illumina), and reads were aligned onto the *hg19* reference genome using NovoAlign (http://www.novocraft.com). Data quality control and variant calling were performed according to GATK Best Practices recommendations. Variant filtering was performed using Annovar and *perl* and *bash* scripts developed in-house (available from the corresponding author upon request) (2). Acid ceramidase activity in fibroblasts was determined as previously described (3).

RESULTS

No pathogenic variant could be identified in the exon sequences of the *MMP2* and *MMP14* genes of the proband. Whole-exome sequencing of the 3 affected individuals revealed the presence of 2 mutations in the *ASAH1* gene, c.505T>C (p.Trp169Arg) and c.760A>G (p.Arg254Gly). Details of the filtering process are presented in Supplementary Table 1, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.39659/abstract. Sanger sequencing of the other members of the family showed the expected segregation (Figure 1A). Acid ceramidase activity was determined in fibroblast cultures of the proband and one of his affected sisters; in both, activity was found to be ~7–8% of average control values.

DISCUSSION

The clinical presentation of our proband in adulthood with shortening of the hand bones, redundant skin, and subcutaneous nodules was highly suggestive of an osteolysis syndrome from the Winchester-Torg-MONA spectrum (see above), and radiographic studies confirmed that impression. Sequencing of the known genes for recessive osteolysis showed negative results. In the

absence of another obvious candidate gene, an exome study was done, and this revealed the presence of compound heterozygous mutations in the gene coding for acid ceramidase, *ASAH1*, that were found to segregate in the family according to a recessive model. Functional relevance of the mutations was confirmed by the demonstration of strongly reduced acid ceramidase activity in cultured cells. These mutations in *ASAH1* were unexpected; thus, the diagnosis of acid ceramidase deficiency came as a surprise. The 3 affected siblings illustrate a markedly protracted course of Farber's disease with an inflammatory phase in childhood followed by a quiescent phase resulting in progressive peripheral osteolysis, a few subcutaneous (and possibly laryngeal) nodules, and no neurologic signs.

The *ASAH1* mutations identified in this family have previously been observed in individuals with Farber's disease, but they have not been seen in combination with each other. The c.505T>C (p.Trp169Arg) mutation has been observed in 2 patients with "classic" Farber's disease (Levade T: unpublished observations). The c.760A>G (p.Arg254Gly) mutation has been reported at homozygosity in a young patient presenting in the first year of life with a clinical picture mimicking severe juvenile idiopathic arthritis (JIA) (4). Genotype–phenotype correlations are not easily recognized in Farber's disease, but the protracted clinical course associated with compound heterozygosity for these mutations is surprising and calls for further studies at the cellular and biochemical level.

While the classic "Farber's triad" consists of hoarseness, subcutaneous nodules, and joint pain, it is recognized today that the primary clinical manifestations of Farber's disease are neurologic (for review, see refs. 5 and 6); indeed, *ASAH1* mutations may manifest as a purely neurologic disorder, spinal muscular atrophy with progressive myoclonic epilepsy (MIM ID #159950) (7,8). Babies with the classic, infantile form of Farber's disease present with symptoms within the first months of life and rarely survive beyond age 2–3 years. In infants, Farber's disease may present as an acute inflammatory condition mimicking JIA (4,9). Hand radiographs of a boy described by Kostik et al (4) showed thinning of the diaphysis and osteopenia, and it is likely that skeletal involvement in that boy may have progressed to look like the osteolysis in our patients. Children with the so-called intermediate or mild forms may lack neurologic involvement but have severe articular and systemic involvement (particularly of the airways) and may have longer survival times. Some individuals with milder forms of Farber's disease have been known to reach early adulthood. Our proband was 40 years old at the time we first saw him; his affected sisters were 58 and 60 years old, respectively). A literature review and an examination of the database of the Toulouse laboratory (Centre Hospitalier Universitaire de Toulouse) suggest that these may be the oldest individuals with acid ceramidase deficiency known so far, illustrating the longest clinical course of Farber's disease.

From the perspective of the radiographic findings, this observation delineates a novel phenotype of peripheral osteolysis that is recessively inherited and caused by loss-of-function mutations in *ASAH1* and that should be considered in the differential diagnosis of peripheral osteolysis with nodules in adults. A skeletal presentation of acid ceramidase deficiency has not been described, and in contrast to many other lysosomal disorders, the condition has not been included in the nosology of genetic skeletal disorders so far (1). Comparison of the radiographs of the affected individuals in this family with those of individuals with *MMP2* and *MMP14* mutations revealed some radiographic signs (Figure 2) that may allow this condition to be distinguished from MMP-2 and MMP-14 osteolyses. The mechanisms by which acid ceramidase deficiency results in osteolysis remain unclear; in the patients in this family, the process seemed to be progressive in the absence of pain, swelling, or other clinical signs of inflammation.

Genetic counseling regarding recessive inheritance was given to our proband, a recommendation for laryngoscopy was formulated (to ascertain the presence of laryngeal nodules), and contacts are being made to explore the possibility of experimental acid ceramidase replacement therapy.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Superti-Furga had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Bonafé, Unger, Superti-Furga.

Acquisition of data. Bonafé, Kariminejad, Li, Royer-Bertrand, Garcia, Mahdavi, Bozorgmehr, Lachman, Mittaz-Crettol, Campos-Xavier, Nampoothiri, Unger, Rivolta, Levade, Superti-Furga.

Analysis and interpretation of data. Bonafé, Li, Royer-Bertrand, Garcia, Campos-Xavier, Unger, Rivolta, Levade, Superti-Furga.

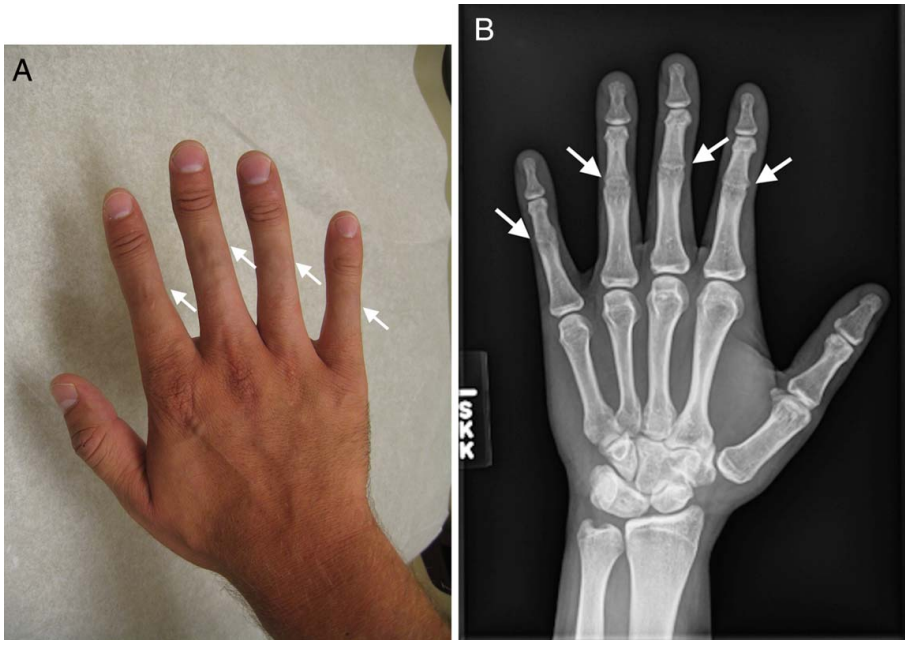
REFERENCES

1. Bonafe L, Cormier-Daire V, Hall C, Lachman R, Mortier G, Mundlos S, et al. Nosology and classification of genetic skeletal disorders: 2015 revision. Am J Med Genet A 2015;167:2869–92.
2. Royer-Bertrand B, Castillo-Taucher S, Moreno-Salinas R, Cho TJ, Chae JH, Choi M, et al. Mutations in the heat-shock protein A9 (HSPA9) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia. Sci Rep 2015;5:17154.
3. Bedia C, Camacho L, Abad JL, Fabriàs G, Levade T. A simple fluorogenic method for determination of acid ceramidase activity and diagnosis of Farber disease. J Lipid Res 2010;51:3542–7.

4. Kostik MM, Chikova IA, Avramenko VV, Vasyakina LI, Le Trionnaire E, Chasnyk VG, et al. Farber lipogranulomatosis with predominant joint involvement mimicking juvenile idiopathic arthritis. J Inherit Metab Dis 2013;36:1079–80.
5. Levade T, Sandhoff K, Schulze H, Medin JA. Acid ceramidase deficiency: Farber lipogranulomatosis. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW, Antonarakis SE, Ballabio A, Gibson KM, Mitchell G, editors. Online metabolic and molecular bases of inherited disease (OMMBID). New York: McGraw-Hill; 2009. URL: http://ommbid.mhmedical.com/content.aspx?sectionid=62643272&bookid=971&jumpsectionID=62643280&Resultclick=2.
6. Sands MS. Farber disease: understanding a fatal childhood disorder and dissecting ceramide biology. EMBO Mol Med 2013;5:799–801.
7. Zhou J, Tawk M, Tiziano FD, Veillet J, Bayes M, Nolent F, et al. Spinal muscular atrophy associated with progressive myoclonic epilepsy is caused by mutations in ASAH1. Am J Hum Genet 2012;91:5–14.
8. Gan JJ, Garcia V, Tian J, Tagliati M, Parisi JE, Chung JM, et al. Acid ceramidase deficiency associated with spinal muscular atrophy with progressive myoclonic epilepsy. Neuromuscul Disord 2015;25:959–63.
9. Torcoletti M, Petaccia A, Pinto RM, Hladnik U, Locatelli F, Agostoni C, et al. Farber disease in infancy resembling juvenile idiopathic arthritis: identification of two new mutations and a good early response to allogeneic haematopoietic stem cell transplantation. Rheumatology (Oxford) 2014;53:1533–4.

DOI: 10.1002/art.39752

Clinical Images: Symphalangism



The patient, a 34-year-old man with a history of noggin gene (*NOG*) mutation resulting in proximal symphalangism, presented to our primary care practice to establish care. He had chronic pain in the affected joints; the symptoms started during childhood and have progressively worsened since. His mother and brother have a similar condition. On physical examination it was noted that the typical skin folds over the proximal interphalangeal (PIP) joints of the fingers were absent bilaterally (**A**) (**arrows**) and there was minimal to no movement at the PIP joints. Plain radiography of the hands revealed bilateral osseous coalition/ankyloses of the PIP joints of the fingers (**B**) (**arrows**). Ankle range of motion was decreased bilaterally. Plain radiography of the feet confirmed osseous coalition/ankyloses at the talonavicular, lateral cuneiform-cuboid, and second tarsometatarsal joints. Symphalangism due to *NOG* mutation is a rare inherited condition with an autosomal-dominant pattern. The 2 commonly recognized types are proximal or Cushing's symphalangism (SYM1), which affects the PIP joints, wrists, and ankles as was the case in our patient, and distal symphalangism (SYM2), which mainly affects the distal interphalangeal (DIP) joints. Of note, patients with SYM1 may also have conductive deafness due to fusion of ossicles, which was absent in our patient. Currently, symptomatic care is the only treatment option in adults with symphalangism; however, surgical procedures have been attempted with some success in children (1).

1. Baek GH, Lee HJ. Classification and surgical treatment of symphalangism in interphalangeal joints of the hand. Clin Orthop Surg 2012;4:58–65.

Amirala S. Pasha, DO, MS
George Washington University
Washington, DC
Melissa Weimer, DO, MCR
Oregon Health & Science University
Portland, OR

6.3 NBAS

NBAS mutations cause a multisystem disorder involving bone, connective tissue, liver, immune system, and retina

Own contribution: Analyzed the WES of the patients and helped in the identification of the new disease-causing gene.

RAPID COMMUNICATION

AMERICAN JOURNAL OF medical genetics PART A

NBAS Mutations Cause a Multisystem Disorder Involving Bone, Connective Tissue, Liver, Immune System, and Retina

**Nuria Garcia Segarra,[1] Diana Ballhausen,[1] Heather Crawford,[2] Matthieu Perreau,[3] Belinda Campos-Xavier,[1] Karin van Spaendonck-Zwarts,[4] Cees Vermeer,[5] Michel Russo,[6] Pierre-Yves Zambelli,[7] Brian Stevenson,[8] Beryl Royer-Bertrand,[9] Carlo Rivolta,[9] Fabio Candotti,[3] Sheila Unger,[9] Francis L. Munier,[10] Andrea Superti-Furga,[11] and Luisa Bonafé[1]*

[1]Center for Molecular Diseases, Lausanne University Hospital, Lausanne, Switzerland

[2]Clinical Metabolic Genetics, Cook Children's Health Care System, Fort Worth, Texas, USA

[3]Division of Immunology and Allergy, Lausanne University Hospital, Lausanne, Switzerland

[4]Department of Clinical Genetics, Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands

[5]VitaK and Cardiovascular Research Institute Maastricht, Maastricht University, Maastricht, Netherlands

[6]Pediatric Service, Centre Hospitalier du Centre Valais, Sion, Switzerland

[7]Service of Pediatric Surgery, Lausanne University Hospital, Lausanne, Switzerland

[8]Vital-IT group, Swiss Institute of Bioinformatics, University of Lausanne, Switzerland

[9]Department of Medical Genetics, Lausanne University Hospital, Lausanne, Switzerland

[10]Jules-Gonin Eye Hospital, Lausanne, Switzerland

[11]Department of Pediatrics, Lausanne University Hospital, Lausanne, Switzerland

Manuscript Received: 25 July 2015; Manuscript Accepted: 10 August 2015

We report two unrelated patients with a multisystem disease involving liver, eye, immune system, connective tissue, and bone, caused by biallelic mutations in the neuroblastoma amplified sequence (*NBAS*) gene. Both presented as infants with recurrent episodes triggered by fever with vomiting, dehydration, and elevated transaminases. They had frequent infections, hypogammaglobulinemia, reduced natural killer cells, and the Pelger–Huët anomaly of their granulocytes. Their facial features were similar with a pointed chin and proptosis; loose skin and reduced subcutaneous fat gave them a progeroid appearance. Skeletal features included short stature, slender bones, epiphyseal dysplasia with multiple phalangeal pseudo-epiphyses, and small C1-C2 vertebrae causing cervical instability and myelopathy. Retinal dystrophy and optic atrophy were present in one patient. NBAS is a component of the synthaxin-18 complex and is involved in nonsense-mediated mRNA decay control. Putative loss-of-function mutations in *NBAS* are already known to cause disease in humans. A specific founder mutation has been associated with short stature, optic nerve atrophy and Pelger–Huët anomaly of granulocytes (SOPH) in the Siberian Yakut population. A more recent report associates *NBAS* mutations with recurrent acute liver failure in infancy in a group of patients of European descent. Our observations indicate that the phenotypic spectrum of NBAS deficiency is wider than previously known and includes skeletal, hepatic, metabolic, and immunologic aspects. Early recognition

How to Cite this Article:
Segarra NG, Ballhausen D, Crawford H, Perreau M, Campos-Xavier B, van Spaendonck-Zwarts K, Vermeer C, Russo M, Zambelli P-Y, Stevenson B, Royer-Bertrand B, Rivolta C, Candotti F, Unger S, Munier F, Superti-Furga A, Bonafé L. 2015. NBAS mutations cause a multisystem disorder involving bone, connective tissue, liver, immune system, and retina.
Am J Med Genet Part A 167A:2902–2912.

Grant sponsor: Swiss National Science Foundation; Grant number: 310030_132940.

*Corresponding to:
Prof. Luisa Bonafé, Center for Molecular Diseases, Lausanne University Hospital, CHUV-CI-02-31, Av. P. Decker 2, 1011 Lausanne, Switzerland.
E-mail: luisa.bonafe@chuv.ch

Article first published online in Wiley Online Library (wileyonlinelibrary.com): 19 August 2015
DOI 10.1002/ajmg.a.37338

of the skeletal phenotype is important for preventive management of cervical instability. © 2015 Wiley Periodicals, Inc.

Key words: NBAS; liver disease; transaminase; fatty acid oxidation; skeletal dysplasia; retinal dystrophy; optic atrophy; immunodeficiency; Pelger–Huët anomaly; cervical instability

INTRODUCTION

Mutations in *NBAS* (neuroblastoma amplified sequence; MIM 608025) were first observed by Maksimova et al. [2010] as the cause of short stature, optic nerve atrophy, and Pelger–Huët anomaly of granulocytes (SOPH syndrome, MIM 614800). This syndrome was observed in 31 Yakut families, an isolated population of east Russia, and associated with a biallelic recessive founder mutation in *NBAS* (c.5741G>A, which predicts p.Arg1914His). Clinical features included postnatal growth retardation, senile face, reduced skin turgor and elasticity, osteoporosis, optic atrophy with loss of visual acuity and color vision, underlobulated neutrophils, and normal intelligence. Recently, Haack et al. [2015] reported biallelic mutations in *NBAS* as the cause of recurrent acute infantile liver failure in a cohort of 11 individuals from ten German families. Here, we describe the clinical phenotype of two unrelated children with *NBAS* mutations that were ascertained for recurrent liver failure, short stature, loose skin, and abnormal skeletal radiographs.

CLINICAL REPORTS

Patient 1 was a 12-year-old boy born to healthy Swiss parents after uneventful pregnancy and delivery (Table I). Birth weight and length were 2,360 g (3rd–5th centile) and 46 cm (3rd–5th centile), respectively. He presented at 4.5 months of life with fever, vomiting, dehydration, and elevated transaminases (up to 2,500 U/L), which normalized within 48 hr following glucose infusion. He presented several similar episodes in the first year of life, always triggered by fever. He had hypoglycemia on two occasions (plasma glucose 2.1 and 2.7 mmol/L respectively, normal >3.0), but neither cholestasis nor significant coagulopathy were observed; factor V was normal (125%, normal: 70–140%) during crises. Serology tests for a large panel of viruses did not identify an infection. At 6 months of life, hepatomegaly was noted; a liver biopsy was performed at age 8 months, showing normal histology and ruling out an inflammatory or autoimmune hepatitis; there were no signs of metabolic derangement, such as steatosis or other storage material. His hepatomegaly normalized by the age of 3 years, despite recurrent acute liver disease episodes with elevated transaminases that persisted till the age of 8 years. Early suppression of fever by mefenamic acid seemed to help in preventing hepatic crises in some instances. Blood cell count was always normal but blood smears, reviewed retrospectively, showed the Pelger–Huët anomaly of granulocytes (Fig. S1). Metabolic evaluations showed mild hyperammonemia (70 μmol/L, normal <50), normal lactate and blood gases, and urinary ketosis (beta-hydroxybutyrate 118–258 mmol/mol creatinine, normal <91.8 mmol/mol creatinine in morning urine [Boulat et al., 2003]) but inappropriately low in respect to prolonged vomiting and fasting; dicarboxylic aciduria was noted at urinary organic acids: adipate 38–119 mmol/mol creatinine (normal <11), suberate 18–26 mmol/mol creatinine (normal <6.4), 3-hydroxyadipate 23–55 mmol/mol creatinine (normal <16), 3-hydroxysebacate 93–148 mmol/mol creatinine (normal <2.7), as well as unusual presence of 3-epoxy-acid excretion (not quantified). Organic acids normalized when he was not acutely ill. Amino acid and acylcarnitine profiles in plasma and total and free carnitine were normal during and between crises. Isoelectric focusing of serum transferrin was normal. Causes of exogenous intoxication were extensively searched for and excluded. An intravenous fructose challenge excluded fructose intolerance. Very long-chain fatty acids in plasma were normal. A skin biopsy was performed and fatty acid oxidation (FAO) enzymes quantified: 3-hydroxyacyl-CoA dehydrogenase (SCHAD) activity was significantly reduced (25.2 nmol/min/mg protein; reference range 40–200) and no mutation was found in the *HADH* gene (MIM 601609) (data not shown). Similarly, no mutation was found in any of the FAO genes (*ACSL3*, MIM 602371; *ACSL6*, MIM 604443; *ACSL5*, MIM 605677; *ACSL4*, MIM 300157; *ACSBG1*, MIM 614362; *ACSBG2*, MIM 614363; *SLC27A2*, MIM 603247; *ACSL1*, MIM 152425; *ECHS1*, MIM 602292; *EHHADH*, MIM 607037; *HSD17B10*, MIM 300256; *HADHB*, MIM 143450; *SCP2*, MIM 184755; *ACAA2*, MIM 604770; *ACAA1*, MIM 604054). Because of the recurrent hypoketosis with dicarboxylic and 3-hydroxydicarboxylic aciduria during crises, and the rapid reversal of liver disease with glucose infusion, a secondary disturbance of fatty acid oxidation was suspected; fasting was limited to a maximum of 10 hr, frequent meals with slow carbohydrates were recommended and supplements of maltodextrins administered regularly, with improvement of physical endurance. With this management, the child could participate in all physical activities in social settings and at school. His learning abilities were apparently normal.

From infancy, his growth curve was slightly below the 3rd centile for weight and length; at 4.5 months his weight was 5 kg (<3rd centile), length 60.5 cm (<3rd centile), OFC 42 cm (10th–25th centile). He had a "progeroid" appearance due to redundant abdominal skin and pseudo-exophthalmos caused by reduced peri-orbital fat. During childhood, the skeletal and connective tissue phenotype became more evident, with persistent proportionate short stature (stature at 6 years was 105 cm, −2 SD), short neck, dorsal kyphosis, slender habitus with reduced subcutaneous fat at the trunk, and redundant skin (Fig. 1). Facial features (Fig. 1) included triangular face, pointed chin, proptosis, and narrow forehead. Neurologic examination was normal and psychomotor development was apparently normal. Skeletal survey (Figs. 2–4) showed long bones with mild diaphyseal stenosis but flared metaphyses, and mild epiphyseal dysplasia. At 8.5 years, there was delayed epiphyseal ossification and discordance between hand bone age (7 years) and cubital bone age (5 years) and the presence of distal pseudoepiphyses at the proximal and middle phalanges. The bones were gracile and osteopenic, despite no history of fracture. Parathyroid hormone, osteocalcin, procollagen-1-N-propeptides (P1NP), thyroid function, growth hormone and insulin secretion tests were normal. At the age of 9 years, after a fall, he presented with acute paresthesia and transient paralysis of all four

TABLE I. Clinical Features of Patients With NBAS Deficiency

Clinical features and laboratory abnormalities	Patient 1	Patient 2
Age	12	5
Nationality	Swiss	American
Consanguinity	−	−
Stature (P)	−3 SD	−3.5 SD
Redundant skin, reduced subcutaneous fat	+	+
Skeletal dysplasia	+	+
Cervical instability (small C1-C2)	+	+
Fractures	−	+[a]
Max. level of AST / ALT	15,000/9,300	>7,500/9,700
Normal (U/L): <45/38 (12 years); <46/52 (5 years)		
Hepatomegaly in infancy	+	+
Acute liver failure (low PT, PTT, Factor V)	−	+
Pelger-Huët anomaly of granulocytes	+	+
Hypogammaglobulinemia: normal values [Jolliff et al., 1982]	+	−
IgG (g/L): 6.39–13.49 (12 years); 4.63–12.36 (4–5years)	IgG 4.9–5.2	IgG 7.39
IgA (g/L): 0.7–3.12 (12 years); 0.25–1.54 (4–5 y)	IgA 0.07–0.18	IgA 1.07
IgM (g/L): 0.56–352 (12 years); 0.43–1.96 (4–5 years)	IgM 0.43–0.56	IgM 1.61
Response to vaccinations	−	n.d.
Reduced natural killer (NK) cells	+	+
Absolute NK cell count	43	40
Normal (cells/μl): 70–480 (12 years); 130–720 (5 years)		
Frequent infections	+	+
Chronic/recurrent viral infections	+ (VZV)	+ (EBV)
Optic atrophy	+	−
Retinal dystrophy	+	−
Recurrent vomiting (crises)	+	+
Metabolic abnormalities during crises	+[b]	−
Neurological abnormalities	−	−
Cardiomyopathy[c]	−	−
Acute renal failure[c]	−	−
Erythema nodosum[c]	−	−
Inflammatory bowel disease[c]	−	−
Epilepsy[c]	−	−
Celiac disease[c]	−	−

n.d., not done/not tested.
VZV, Varicella Zoster Virus; EBV, Epstein Barr Virus.
[a]1 skull and 1 left femur fracture.
[b]Hypoketosis, dicarboxylic aciduria, and epoxy-acid excretion during crises.
[c]Additional clinical features reported by Haack et al. [2015].

limbs; X-rays and CT scan of the cervical spine showed small C1-C2 and cervical instability (Fig. 5) with compression of the cervical spinal cord, confirmed by pathologic brainstem evoked potentials. Stabilization surgery was performed with complete reversal of symptoms.

Eye examination at 5 years of age showed optic atrophy (Fig. S2) with retinal dystrophy at electroretinogram; he became symptomatic (photophobia, loss of visual acuity, and color vision) at the age of 8 years, therefore reading aids were provided.

At age 3 years, he was hospitalized for bacterial pneumonia, requiring intravenous antibiotics. He had five to eight episodes of middle ear infection per year from age 3 to 11 years, necessitating oral antibiotics. Since the age of 5 years he had recurrent labial herpes simplex virus (HSV) infection, with almost no free interval, associated with diffuse palpable lymphadenopathy. At 5 years of age, he was hospitalized for a chickenpox infection with prolonged fever, necessitating several weeks to recover; transaminases in plasma were as high as 15,000 U/L (reference values <45 U/L for AST and <38 U/L for ALT) during this infection. At age 7 years, he had a reactivation of varicella-zoster virus (VZV) infection in the thoracic region with ophthalmic involvement. A second recurrence of cutaneous and ophthalmic VZV occurred at age 11.5 years. Immunologic workup showed hypogammaglobulinemia (Table I), absent response to vaccinations, and reduced natural killer (NK) cell count (Table I). Subcutaneous weekly injections of immunoglobulins were started at age 11.5 years, restored normal IgG levels and substantially reduced both the upper airway infections and the labial HSV reactivations, with improvement of general well-being

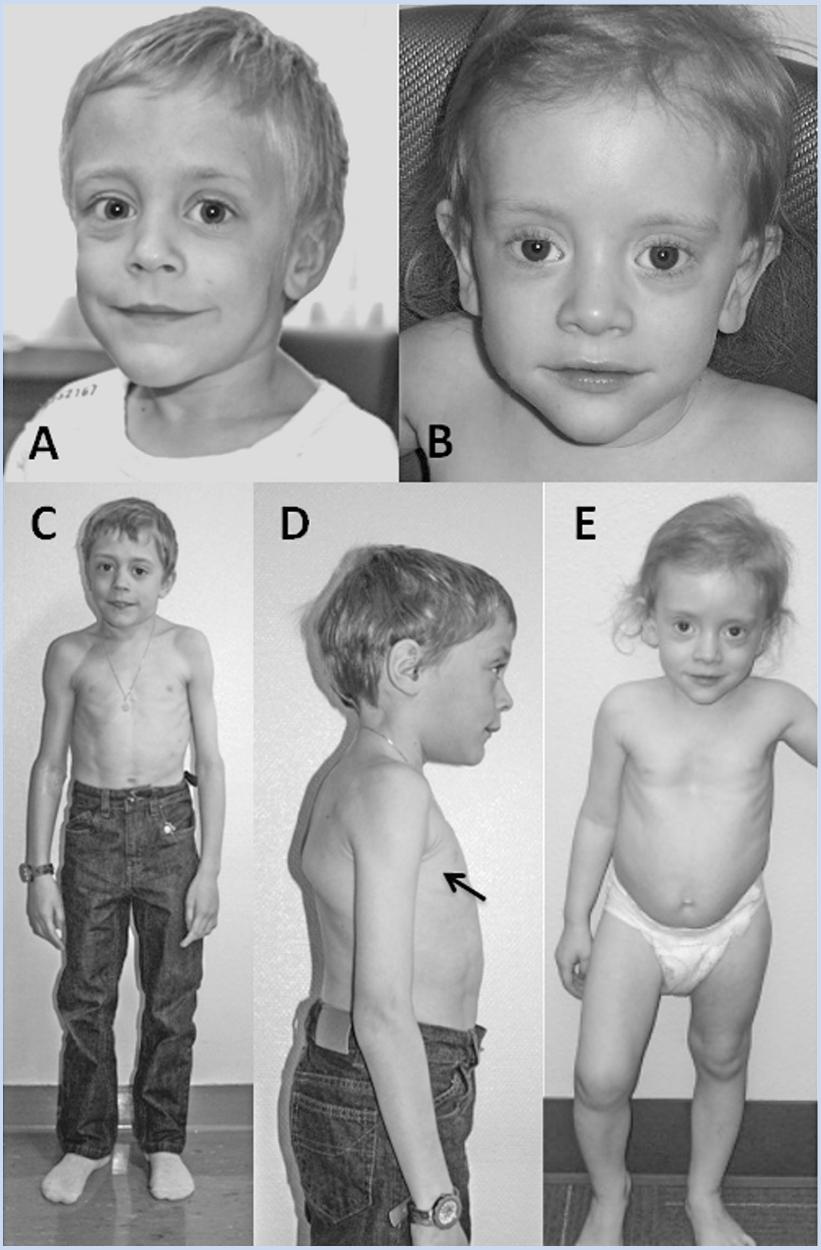


FIG. 1. Phenotype of patients with NBAS deficiency. A) and B) facial features of patient 1 (A) and 2 (B): triangular face, pointed chin, prominent eyes, narrow forehead. C) and D) Patient 1 at age 8 years: reduced subcutaneous fat with prominent musculature, skin folds (arrow), and prominent scapula. E) Patient 2 at age 3 years: pectus excavatum and prominent abdomen.

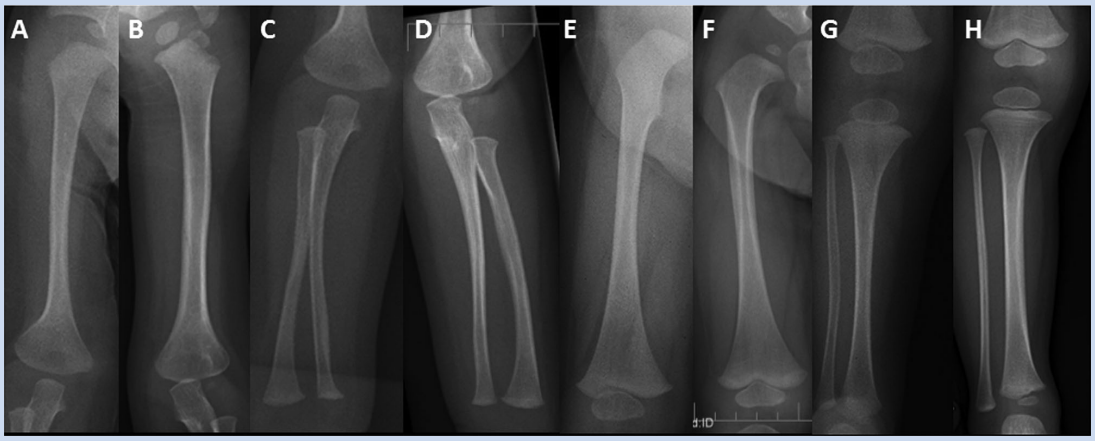


FIG. 2. Radiographic features of long bones in patients with NBAS deficiency. Panels A, C, E, G: patient 1 at age 1 year (humerus, radius, and ulna, femur, tibia, and fibula, respectively). Panels B, D, F, H: patient 2 at age 2 years (humerus, radius, and ulna, femur, tibia, and fibula, respectively). Note mild osteopenia, slender bones with stenotic diaphyseal shafts (particularly in panels B, E, and G), flared metaphyses. Knee epiphyses are mildly flattened.

of the patient. He has not had any further acute liver crisis after 8 years of age. At the most recent evaluation (12 years), his weight was 24.7 kg (3 kg <3rd centile), stature was 131 cm (5 cm <3rd centile; −3 SD) and OFC was 54 cm (50th centile).

Patient 2, a 5-year-old girl, was born to Caucasian parents at 37 weeks of gestation with a birth weight of 2.52 kg (50th centile) (Table I). She required one week of neonatal intensive care due to transient tachypnea of the newborn. At discharge, she was able to feed normally. However, she had feeding difficulties and delayed milestones by 6 months of age. She presented at 23 months of age with fever, recurrent episodes of non-bloody, non-bilious vomiting, and elevated transaminases. Transaminases were significantly

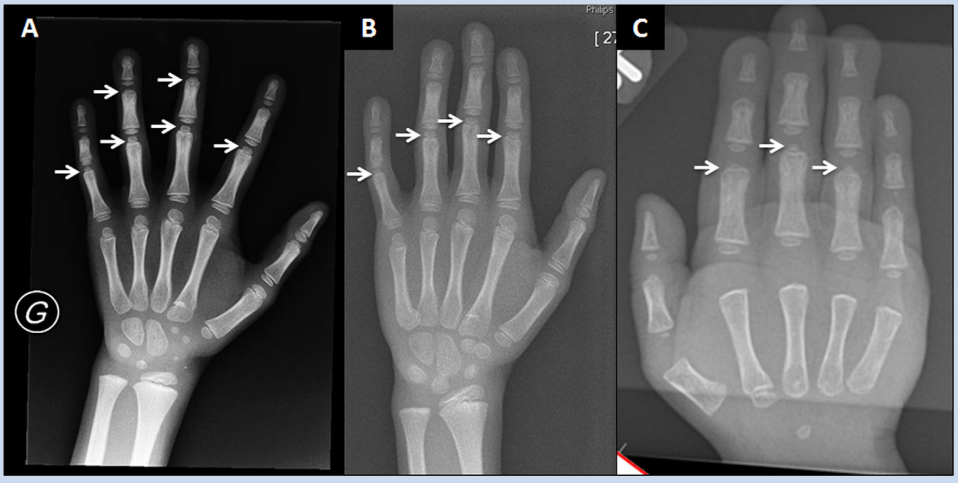


FIG. 3. Hand radiographs in patients with NBAS deficiency: A) Patient 1 at age 7 years; B) Patient 1 at age 8.5 years; C) Patient 2 at age 2 years. Note presence of pseudo-epiphyses at the distal phalanges (arrows). In panel B, note the dissociation of bone age: carpal bone and hand epiphyses are compatible with a bone age of 7 years; the delayed ossification of distal cubital epiphysis is compatible with a bone age of 5 years. In panel C, note delayed carpal bone age (<6 months) compared to epiphyseal bone age (1.5 years).

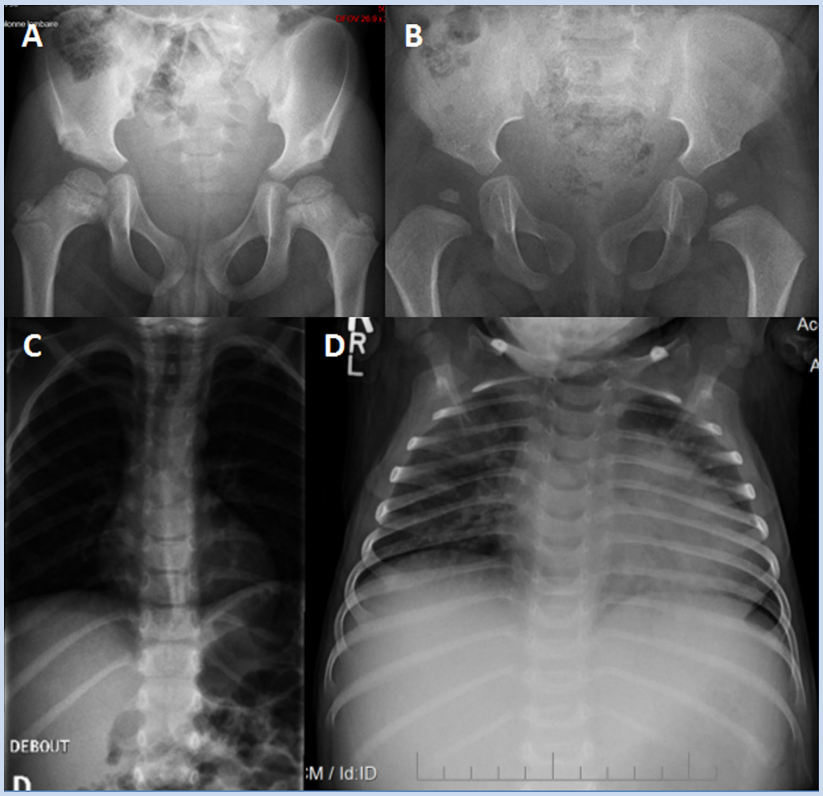


FIG. 4. Radiographs of the pelvis and chest in patients with NBAS deficiency. A) Pelvis of patient 1 at age 7 years and 10 months: the capital femoral epiphyses are mildly flattened; there are metaphyseal irregularities at the proximal femur, bilaterally. B) Pelvis of patient 2 at age 2 years; note small capital femoral epiphyses. C) Chest radiograph of patient 1 at age 7 years; note thin ribs and normal vertebral bodies. D) Chest radiograph of patient 2 at age 2 years; note short twelfth rib.

elevated (AST >7,500 U/L, normal: 16–46; ALT 9,070; normal: 33–52) which decreased during the initial crisis, but remained elevated upon discharge from the hospital. Coagulation studies were abnormal (PT 24.4 sec, normal: 11.4–15.1; PTT 39 sec, normal: 22.4–37.5, INR 2.3) and ferritin was increased (977 μg/L, normal: 10–95). She presented with hepatomegaly during the initial crisis. Liver biopsy was performed during the second crisis at 26 months of age and showed EBV nucleic acids and erythrophagocytosis suggestive of Epstein-Barr virus (EBV)-related hepatitis. Her hepatomegaly resolved after age 3 years. A metabolic evaluation during one of the crises including plasma lactate and ammonia, urine organic acids, plasma amino acids, acylcarnitine profile, succinylacetone, isoelectric focusing of transferrin, and alpha-1-antitrypsin was reported to be normal. The CSF cell counts, neurotransmitters, amino acids, and lactate were all normal. Sweat chloride testing was normal. Initial white blood cell count was low; however, all other blood counts were normal. Extensive serology for hepatotropic viruses was negative, except for chronic EBV infection. A blood smear was performed during the second acute crisis when she was 26 months and showed an absolute monocytosis. Repeat blood smear performed at 4 years showed Pelger–Huët anomaly in approximately 5% of neutrophils. An immunologic evaluation showed normal IgG and low NK cell count (Table I). In the first 2 years of life, she was hospitalized three times for acute liver failure. All crises began with fever and vomiting. Her transaminases normalized about 2 months after the prior acute crisis. Her facial phenotype included triangular face, pointed chin, down-turned corners of the mouth, and proptosis (Fig. 1). Her growth parameters were small for age; she had abundant skin and reduced subcutaneous fat (Fig. 1). During the hospitalization at 23 months of age, skeletal survey (Fig. 5) showed a J-shaped sella and small C1-C2. Cervical instability was diagnosed and fixation surgery performed. This study also showed a bone dysplasia similar to patient 1 (Figs 2–4). She has a history of two fractures (skull at 11 months and left femur at 4 years) that were the result of accidental falls. Her motor and speech development were delayed at age 3 years, but improved and motor development normalized. She still requires speech therapy for mild delays due to poor articulation. At last evaluation she was diag-

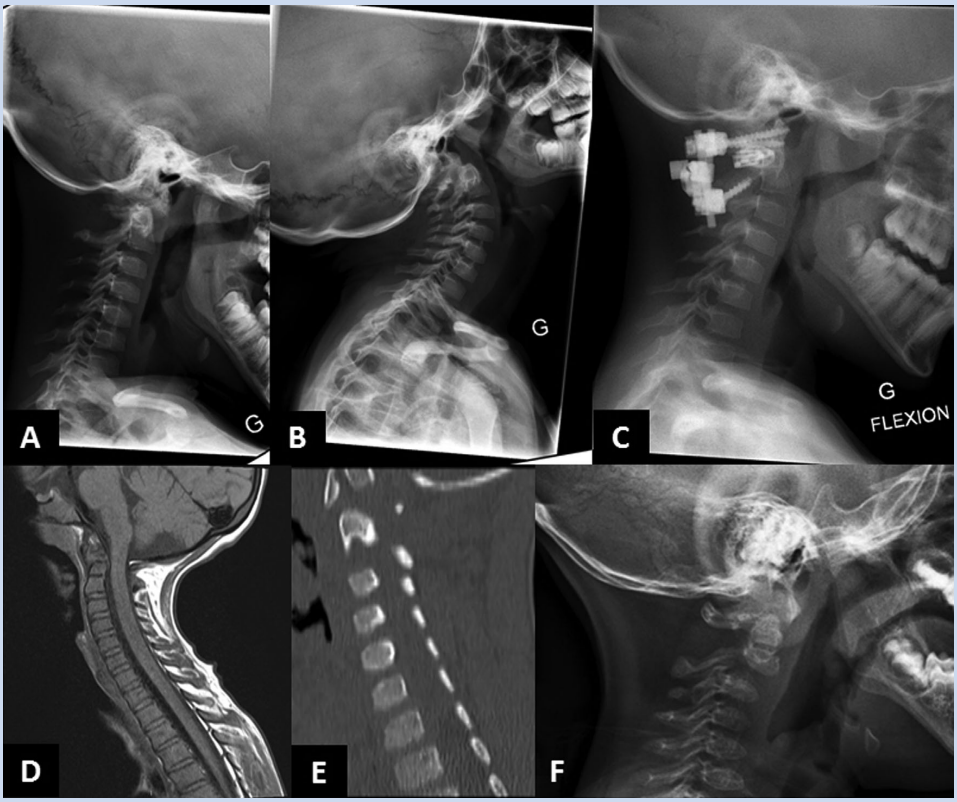


FIG. 5. Cervical instability in patients with NBAS deficiency. Panel A, B, and C: radiographs of the cervical spine of patient 1 at age 10 years (B in extension; C after fixation surgery). Panel D and E: MRI and CT scan of the cervical spine of patient 1 at 10 years. Panel F: radiograph of cervical spine of patient 2 at age 3 years. Note J-shaped sella in both patients (panels C and F) and the small C1 and C2 (all panels) with reduced space around the myelon (panel D).

nosed with systemic HSV infection and her growth parameters at that time (5 years) were as follows: weight 12.3 kg (1st centile), stature 91 cm (−3.5 SD) and OFC 48.5 cm (5th centile). Eye examination at 4 years did not identify optic atrophy or retinopathy, however she did have nearsightedness with astigmatism.

Exome Sequencing (ES)

The study was conducted within the frame of a Swiss National Science Foundation-funded research project on skeletal dysplasias for which ethical approval was obtained from the State Ethic Committee (Canton Vaud). DNA was extracted according to standard protocols from leucocytes from the two probands and their parents and siblings who consented to the study. We performed ES in the DNA of patients 1 and 2 as described [Shi et al., 2011]. The results were analyzed with an informatic pipeline excluding variants present at significant frequency in public databases, and variants frequently observed in the in-house database. Results were further prioritized in terms of data quality and assuming recessive inheritance in the two probands. Candidate mutations were screened using known annotations for gene function and associated phenotypes. Sanger sequencing was used to validate the presence and status of suspected pathogenic mutations in the two probands and to confirm segregation in the parents.

Measurement of Matrix Gamma-Carboxyglutamate (Gla) Protein (MGP) Activity

Circulating MGP levels were measured in plasma from patient 1. For this study, peripheral blood was drawn in EDTA, immediately centrifuged and frozen. The MGP was measured by ELISA methods as described [Cranenburg et al., 2011].

RESULTS

ES Data

Upon inspection of the filtered ES data, biallelic mutations were found in 14 genes in common between the two unrelated probands; among them, four mutations in *NBAS* met the criteria of correct

segregation and low frequency in control datasets: patient 1 was compound heterozygous for c.284C>T, p.Ala95Val (exon 4) and c.850A>T, p.Lys284Ter (exon 10) and patient 2 was compound heterozygous for c.409C>T, p.Arg137Trp (exon 7) and c.1186T>A, p.Trp396Arg (exon 14) (according to ENSEMBL Accession Number ENST00000281513). Three of these mutations were absent in variant databases but the missense mutation c.409C>T, p.Arg137Trp (exon 7) was listed twice in Exome Aggregation Consortium (ExAC) with a heterozygous frequency of 0.00001648 and once in Exome Variant Server (EVS). The four mutations were not found in our in-house database (>400 ES). Sanger sequencing showed the four variants were each heterozygous in one of the parents.

The PROVEAN software (Protein Variation Effect Analyzer) predicted a deleterious impact on the biological function of protein for p.Ala95Val (score −2.837; deleterious effect < −2.5), p. Arg137Trp (score −4.762) and p.Trp396Arg (score −12.671). The Polyphen-2 program v2.2.2r398 predicted the p.Ala95Val as probably damaging (score 0.982 with sensitivity 0.75 and specificity 0.96).

Figure 6 summarizes all *NBAS* mutations with software predictions, conservation, and population frequencies for mutations described herein.

Circulating MGP measurement in plasma of patient 1 showed normal concentration of uncarboxylated MGP: 120 pmol/L (reference values 71–756) [Cranenburg et al., 2011].

DISCUSSION

These patients presented with hepatic, metabolic, skeletal, retinal, and immunologic abnormalities that were undiagnosed until ES identified mutations in *NBAS*, in both. Upon review of their manifestations and other patients reported with *NBAS* mutations, we concluded that these variants were causative.

NBAS is a component of the syntaxin 18 complex involved in Golgi-to-endoplasmic reticulum (ER) retrograde transport of vesicles, allowing the distribution of proteins from the ER to the Golgi compartments [Aoki et al., 2009] necessary for balanced macromolecule intracellular trafficking. NBAS interacts with p31 of the syntaxin 18 complex in both HeLa cells [Aoki et al., 2009] and human fibroblasts [Haack et al., 2015]. A disturbance of retrograde Golgi-to-ER transport may result in ER stress, as suggested by expression profiling in NBAS-mutant fibroblasts [Haack et al., 2015].

NBAS is also important in nonsense-mediated mRNA decay (NMD) control. It forms an autoregulatory, conserved NMD circuit that regulates endogenous RNA targets [Longman et al., 2013]. Knockdown of *NBAS* causes up-regulation of genes involved in ER-stress, NMD factors, genes involved in response to wounding and nutrients, and in regulation of inflammatory response [Longman et al., 2013].

These functions suggest a pleiotropic role of this gene, potentially explaining the diversity of the phenotypic features of the patients.

Metabolic Aspects

The multisystem involvement observed in these patients resembles that of congenital disorders of glycosylation (CDG), in which defective glycosylation in the ER or Golgi affects the function of diverse proteins. However, isoelectric focusing of serum transferrin was normal in both children, thus excluding CDG type I or II and confirming the findings of Haack et al. [2015], who also found no glycosylation defect in their series.

The metabolic disturbances found in patient 1 were significant, although present exclusively during crises. The inappropriate hypoketosis and dicarboxylic aciduria during crises and the good response to glucose infusion suggested FAO dysfunction; the reduced activity of SCHAD enzyme in fibroblasts in the absence of any mutation in the *HADH* gene implicated a secondarily reduced *HADH* expression. Experiments to test this are underway. A generalized mitochondrial dysfunction as cause of the metabolic derangement in patient 1 seems unlikely because plasma lactate and urinary Krebs cycle intermediates were not elevated. Interestingly, 3-epoxy-acids were recurrently elevated in urine during crises, suggesting a peroxisomal disturbance. The NBAS protein contains a leucine zipper motif, three putative transmembrane helices, an ER signal sequence, and two peroxisomal targeting signals [Scott et al., 2003]. Our findings suggest a possible role of NBAS in regulating peroxisomal metabolism and/or protein trafficking in this organelle; further experiments are necessary to test this hypothesis.

Hepatic and Immunologic Aspects

Acute liver disease was triggered by fever in these patients and in those of Haack et al. [2015]. It has been proposed that the sensitivity to fever in these patients may be due to thermal susceptibility of the syntaxin complex 18 [Haack et al., 2015]. However, the evidence of immunodeficiency in the present patients and specific NBAS transcriptional targets identified by Longman et al. [2013] in HeLa cells suggest a link between immune dysregulation and recurrent liver inflammation. Knockdown of *NBAS* in HeLa cells caused upregulation of *CCL20* and *IL6R*. CCL20 is a chemokine known to be upregulated in alcoholic hepatitis; it is a biomarker and a mediator of lipopolysaccharide-induced liver inflammation [Affo et al., 2014]. Its upregulation in NBAS deficiency may therefore facilitate liver inflammation and cytolysis upon stress factors. IL6 is a multifunctional cytokine involved in induction of immunoglobulin production from activated B-cells, in T lymphocyte proliferation, survival and differentiation, and in regulation of acute phase proteins in the liver [Zohlnhofer et al., 1992]. *IL6R* is expressed predominantly by hepatocytes, neutrophils, monocytes/macrophages, and some lymphocytes [Wang et al., 2013]. IL6 down-regulates IL6R via endocytosis (by internalization) and lysosomal degradation. The reappearance of the receptor at the cell surface is cycloheximide-sensitive. Degradation is likely to be protective for the cell against overstimulation [Zohlnhofer et al., 1992]. NBAS deficiency is associated with reduced degradation of IL6R mRNA in HeLa cells [Longman et al., 2013]. This may explain the hypersensitivity of liver cells to inflammatory cytokines and may provide a link between the immunologic and hepatic abnormalities in NBAS deficiency. Soluble IL6R (sIL6R) plays a role in systemic inflammatory response and its expression is induced by viral infection via the cyclo-oxygenase 2 (COX-2) pathway and inhibits viral replication [Wang et al., 2013]. This model may

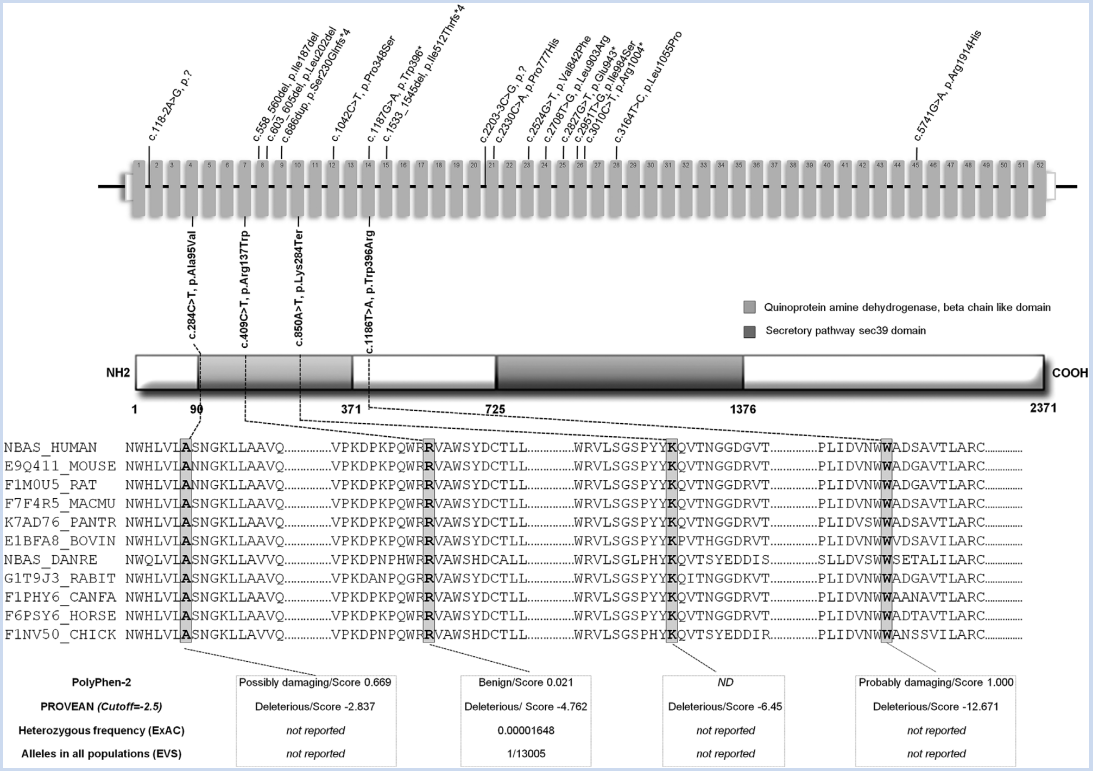


FIG. 6. The exon structure of the NBAS gene and a schematic representation of the NBAS protein with its domains are shown. At the top are the mutations described by Maksimova et al. [2010] and by Haack et al. [2015]. Between the gene cartoon and the protein cartoon are the four mutations identified in this paper. Below the protein cartoon are screenshots from the UniProt website showing that all four mutated amino acids are completely conserved in all species available. Finally, at the bottom are the results of pathogenicity predictions (PolyPhen-2 and PROVEAN) as well as allelic frequencies in the ExAC and EVS databases.

provide a rationale for the use of non-steroidal anti-inflammatory drugs at the onset of febrile illnesses to reduce liver inflammation in patients with NBAS deficiency.

The immunodeficiency in NBAS deficiency is subtle and possibly not consistent; patient 1 had a derangement of both humoral and cellular immune responses, and patient 2 only derangement of cellular immune response at last evaluation. Immunoglobulin replacement should be considered in affected patients to reduce frequency of infection and at the same time to avoid triggers of liver failure. Reduced NK cells may predispose to chronic viral infection and be a risk factor for tumorigenesis [Artis and Spits, 2015], for which patients should be monitored longitudinally.

Skeletal Aspects

Postnatal short stature was evident from early infancy and is a feature of the SOPH syndrome in the Yakuts [Maksimova et al., 2010]. No details on growth were provided by Haack et al. [2015], however those patients may not have been investigated because of the prominence of the hepatic involvement. Maksimova et al. [2010] suggested primary skeletal involvement in their patients, however those radiographic studies were done at adult age when dysplastic changes are usually less evident. They highlighted osteoporosis, delayed bone age, and late ossification of vertebral apophysis in one patient [Maksimova et al., 2010]. The skeletal dysplasia in the patients reported here showed thin, gracile bones, with stenotic diaphyses and flared metaphyses, and mild generalized epiphyseal dysplasia, suggesting a disturbance in bone mineralization rather than in skeletal patterning.

In the *NBAS* knockdown model of Longman et al. [2013], *MGP* was upregulated. MGP is a negative regulator of bone formation and deficiency causes Keutel syndrome, with abnormal calcification of cardiac valves, ear pinnae, and bronchial arches [Munroe et al., 1999]. Increased expression of *MGP* would be compatible with the bone phenotype observed in the present patients, with delayed maturation of epiphyses, thin bones, and underossified vertebrae. We measured circulating MGP in patient 1, which was in the reference range for adult controls, in contrast to upregulation

found in HeLa cells [Longman et al., 2013]. Our results, although limited to one patient, do not support a role for this protein in the skeletal phenotype of NBAS deficiency.

Independent from the pathogenetic mechanism, the under-ossification of the cervical spine observed in these patients indicates a serious risk of neurological complications and should be assessed in patients with NBAS deficiency.

Ophthalmologic Aspects

Finally, we confirm that eye involvement includes optic atrophy and retinal dystrophy, as reported by Maksimova et al. [2010] and is not limited to the Yakut population. While optic atrophy is a frequent feature in disturbances of energy metabolism, retinal dystrophy may be linked to disturbed Golgi-to-cilium vesicle transport [Evans et al., 2010]. Therefore, NBAS is likely to play an important role in retinal homeostasis and patients should be monitored accordingly. Patient 1 had signs of atypical cone-rod dystrophy from age 5, but became symptomatic at age 8. Patient 2 did not have evidence of retinal dystrophy but her last eye examination was done at age 4. No details are available about the visual function of patients reported by Haack et al. [2015].

Our observations highlight the heterogeneous and pleiotropic clinical manifestations of NBAS deficiency. This report should raise awareness on the syndrome; we have expanded the syndrome, highlighting some clinical issues, which could be important to recognize early, such as cervical instability and immunodeficiency. Our observations also suggest possible molecular and pathogenetic mechanisms to be investigated in future research to understand and treat this disorder.

WEB RESOURCES

Lausanne Genomic Technology Facility (LGTF), University of Lausanne: http://www.unil.ch/gtf/home.htlm

UniProt, http://www.uniprot.org/uniprot/P45452

The National Heart Lung and Blood Institute (NHLBI) Exome Sequencing Project, http://evs.gs.washington.edu/EVS/

Exome Aggregation Consortium (ExAC), http://exac.broadinstitute.org/

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/

PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/index.shtml

Protein Variation Effect Analyzer (PROVEAN), http://provean.jcvi.org/index.php

SIFT, http://sift.jcvi.org/

UniProt Knowledgebase (UniProtKB), http://www.uniprot.org/

ACKNOWLEDGMENTS

The authors have no conflict of interest to declare. The study was supported by the Swiss National Science Foundation (grant 310030_132940 to L.B.). We thank the Exome Aggregation Consortium (ExAC) server for the data on mutation frequency. We thank Michael J. Bennett, Children's hospital of Philadelphia, (PA), USA, for measuring SCHAD activity and Arnold W. Strauss, University of Cincinnati (OH), USA, for *HADH* mutation analysis. We thank Keith Harshman of the Lausanne Genomic Technologies Facility (LGTF) at the University of Lausanne, for performing ES experiments. We thank Olivier Boulat, Clothilde Roux and Olivier Braissant of the Biomedicine laboratory at the CHUV for their contribution in the metabolic workup. We also thank Carole Chiesa for technical laboratory assistance and the patients' families for their collaboration.

REFERENCES

Affo S, Morales-Ibanez O, Rodrigo-Torres D, Altamirano J, Blaya D, Dapito DH, Millan C, Coll M, Caviglia JM, Arroyo V, Caballeria J, Schwabe RF, Gines P, Bataller R, Sancho-Bru P. 2014. CCL20 mediates lipopolysaccharide induced liver injury and is a potential driver of inflammation and fibrosis in alcoholic hepatitis. Gut 63:1782–1792.

Aoki T, Ichimura S, Itoh A, Kuramoto M, Shinkawa T, Isobe T, Tagaya M. 2009. Identification of the neuroblastoma-amplified gene product as a component of the syntaxin 18 complex implicated in Golgi-to-endoplasmic reticulum retrograde transport. Mol Biol Cell 20:2639–2649.

Artis D, Spits H. 2015. The biology of innate lymphoid cells. Nature 517:293–301.

Boulat O, Gradwohl M, Matos V, Guignard JP, Bachmann C. 2003. Organic acids in the second morning urine in a healthy Swiss paediatric population. Clin Chem Lab Med 41:1642–1658.

Cranenburg EC, KY VANS-Z, Bonafe L, Mittaz Crettol L, Rodiger LA, Dikkers FG, AJ VANE, Superti-Furga A, Alexandrakis E, Vermeer C, Schurgers LJ, Laverman GD. 2011. Circulating matrix gamma-carboxyglutamate protein (MGP) species are refractory to vitamin K treatment in a new case of Keutel syndrome. J Thromb Haemost 9:1225–1235.

Evans RJ, Schwarz N, Nagel-Wolfrum K, Wolfrum U, Hardcastle AJ, Cheetham ME. 2010. The retinitis pigmentosa protein RP2 links pericentriolar vesicle transport between the Golgi and the primary cilium. Hum Mol Genet 19:1358–1367.

Haack TB, Staufner C, Kopke MG, Straub BK, Kolker S, Thiel C, Freisinger P, Baric I, McKiernan PJ, Dikow N, Harting I, Beisse F, Burgard P, Kotzaeridou U, Kuhr J, Himbert U, Taylor RW, Distelmaier F, Vockley J, Ghaloul-Gonzalez L, Zschocke J, Kremer LS, Graf E, Schwarzmayr T, Bader DM, Gagneur J, Wieland T, Terrile C, Strom TM, Meitinger T, Hoffmann GF, Prokisch H. 2015. Biallelic Mutations in NBAS Cause Recurrent Acute Liver Failure with Onset in Infancy. Am J Hum Genet 97:163–169.

Jolliff C, Cost K, Stivrins P, Grossman P, Nolte C, Franco S, Fijan K, Fletcher L, Shriner H. 1982. Reference intervals for serum IgG, IgA, IgM, C3, and C4 as determined by rate nephelometry. Clin Chem 28:126–128.

Longman D, Hug N, Keith M, Anastasaki C, Patton EE, Grimes G, Caceres JF. 2013. DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and Caenorhabditis elegans. Nucleic Acids Res 41:8319–8331.

Maksimova N, Hara K, Nikolaeva I, Chun-Feng T, Usui T, Takagi M, Nishihira Y, Miyashita A, Fujiwara H, Oyama T, Nogovicina A, Sukhomyasova A, Potapova S, Kuwano R, Takahashi H, Nishizawa M, Onodera O. 2010. Neuroblastoma amplified sequence gene is associated with a novel short stature syndrome characterised by optic nerve atrophy and Pelger–Huet anomaly. J Med Genet 47:538–548.

Munroe PB, Olgunturk RO, Fryns JP, Van Maldergem L, Ziereisen F, Yuksel B, Gardiner RM, Chung E. 1999. Mutations in the gene encoding the human matrix Gla protein cause Keutel syndrome. Nat Genet 21:142–144.

Scott DK, Board JR, Lu X, Pearson AD, Kenyon RM, Lunec J. 2003. The neuroblastoma amplified gene, NAG: Genomic structure and characterisation of the 7.3 kb transcript predominantly expressed in neuroblastoma. Gene 307:1–11.

Shi Y, Li Y, Zhang D, Zhang H, Li Y, Lu F, Liu X, He F, Gong B, Cai L, Li R, Liao S, Ma S, Lin H, Cheng J, Zheng H, Shan Y, Chen B, Hu J, Jin X, Zhao P, Chen Y, Zhang Y, Lin Y, Li X, Fan Y, Yang H, Wang J, Yang Z. 2011. Exome sequencing identifies ZNF644 mutations in high myopia. PLoS Genet 7:e1002084.

Wang Q, Chen X, Feng J, Cao Y, Song Y, Wang H, Zhu C, Liu S, Zhu Y. 2013. Soluble interleukin-6 receptor-mediated innate immune response to DNA and RNA viruses. J Virol 87:11244–11254.

Zohlnhofer D, Graeve L, Rose-John S, Schooltink H, Dittrich E, Heinrich PC. 1992. The hepatic interleukin-6 receptor. Down-regulation of the interleukin-6 binding subunit (gp80) by its ligand. FEBS Lett 306:219–222.

SUPPORTING INFORMATION

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

6.4 CEP78

Mutations in CEP78 Cause Cone-Rod Dystrophy and Hearing Loss Associated with Primary-Cilia Defects

Own contribution: Prepared the WES data for an overlap-based strategy on a cohort of patients with Cone-Rod Dystrophy.

REPORT

Mutations in *CEP78* Cause Cone-Rod Dystrophy and Hearing Loss Associated with Primary-Cilia Defects

Konstantinos Nikopoulos,[1,12] Pietro Farinelli,[1,12] Basilio Giangreco,[2] Chrysanthi Tsika,[3] Beryl Royer-Bertrand,[1,4] Martial K. Mbefo,[5] Nicola Bedoni,[1] Ulrika Kjellström,[6] Ikram El Zaoui,[1] Silvio Alessandro Di Gioia,[1] Sara Balzano,[1] Katarina Cisarova,[1] Andrea Messina,[7] Sarah Decembrini,[5] Sotiris Plainis,[3] Styliani V. Blazaki,[3] Muhammad Imran Khan,[8] Shazia Micheal,[8] Karsten Boldt,[9] Marius Ueffing,[9] Alexandre P. Moulin,[10] Frans P.M. Cremers,[8,11] Ronald Roepman,[8] Yvan Arsenijevic,[5] Miltiadis K. Tsilimbaris,[3] Sten Andréasson,[6] and Carlo Rivolta[1,*]

Cone-rod degeneration (CRD) belongs to the disease spectrum of retinal degenerations, a group of hereditary disorders characterized by an extreme clinical and genetic heterogeneity. It mainly differentiates from other retinal dystrophies, and in particular from the more frequent disease retinitis pigmentosa, because cone photoreceptors degenerate at a higher rate than rod photoreceptors, causing severe deficiency of central vision. After exome analysis of a cohort of individuals with CRD, we identified biallelic mutations in the orphan gene *CEP78* in three subjects from two families: one from Greece and another from Sweden. The Greek subject, from the island of Crete, was homozygous for the c.499+1G>T (IVS3+1G>T) mutation in intron 3. The Swedish subjects, two siblings, were compound heterozygotes for the nearby mutation c.499+5G>A (IVS3+5G>A) and for the frameshift-causing variant c.633delC (p.Trp212Glyfs*18). In addition to CRD, these three individuals had hearing loss or hearing deficit. Immunostaining highlighted the presence of CEP78 in the inner segments of retinal photoreceptors, predominantly of cones, and at the base of the primary cilium of fibroblasts. Interaction studies also showed that CEP78 binds to FAM161A, another ciliary protein associated with retinal degeneration. Finally, analysis of skin fibroblasts derived from affected individuals revealed abnormal ciliary morphology, as compared to that of control cells. Altogether, our data strongly suggest that mutations in *CEP78* cause a previously undescribed clinical entity of a ciliary nature characterized by blindness and deafness but clearly distinct from Usher syndrome, a condition for which visual impairment is due to retinitis pigmentosa.

Cone-rod degeneration (CRD [MIM: 120970]) represents an extremely rare class of hereditary diseases that affect the light-sensing neurons of the retina, the cone and rod photoreceptors.[1] Cones are involved in daytime vision, providing the brain with color information and central, precise visual input. Conversely, rods are active in very dim light conditions, are more abundant in the retinal periphery, and produce achromatic information, typical for instance of the visual stimulation provided by a landscape on a moonless night. Individuals with CRD experience initial loss of visual acuity (central vision) and aberrant color vision as a result of the prominent loss of cones, whereas rod functions remain relatively preserved.[2] As the disease progresses, both cone and rod functions deteriorate and central vision is severely impaired or lost, but peripheral islands of the retina might retain some residual activity.[3] On the basis of these clinical parameters, CRD can be distinguished from retinitis pigmentosa (RP [MIM: 26800], also called rod-cone degeneration), the most common form of hereditary retinal degeneration. In retinitis pigmentosa, rods are more severely affected than cones; initial symptoms include night blindness due to loss of rod function, and central vision (cone function) is often preserved until the very late stages of the disease.[4] CRD is almost invariantly inherited as a Mendelian trait, predominantly according to a recessive pattern of transmission, and is characterized by an elevated genetic and allelic heterogeneity.[5] Although as many as 33 CRD-associated genes have been identified to date (RetNet; see Web Resources), they are found to be mutated in only ~25% of clinical cases, implying that a substantial percentage of affected people might carry mutations in yet-to-be-identified genes.[6]

According to this rationale, we performed whole-exome sequencing (WES) in 34 unrelated individuals with CRD (29 from Greece and five from Sweden). Genomic DNA was extracted from peripheral blood leukocytes according to standard procedures, and then exomic libraries (SureSelect V5 kit, Agilent) were sequenced on an Illumina HiSeq 2000. Raw sequence files were assessed, trimmed, and finally mapped back to the human genome reference sequence (build hg19); DNA variants were called

[1]Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, 1011 Lausanne, Switzerland; [2]Center for Psychiatric Neuroscience, Department of Psychiatry, Lausanne University Hospital, 1008 Lausanne, Switzerland; [3]Department of Ophthalmology, Medical School, University of Crete, 71003 Heraklion, Greece; [4]Center for Molecular Diseases, Lausanne University Hospital, 1011 Lausanne, Switzerland; [5]Unit of Gene Therapy and Stem Cell Biology, Jules Gonin Ophthalmic Hospital, 1004 Lausanne, Switzerland; [6]Lund University, Skane University Hospital, Department of Ophthalmology, 20502 Lund, Sweden; [7]Service of Endocrinology, Diabetes and Metabolism, Lausanne University Hospital, 1011 Lausanne, Switzerland; [8]Department of Human Genetics and Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6525GA Nijmegen, The Netherlands; [9]Medical Proteome Center, Institute for Ophthalmic Research, University of Tuebingen, 72074 Tuebingen, Germany; [10]Department of Ophthalmology, Jules Gonin Ophthalmic Hospital, 1004 Lausanne, Switzerland; [11]Donders Center for Neuroscience, Radboud University, 6525EN Nijmegen, The Netherlands
[12]These authors contributed equally to this work
*Correspondence: carlo.rivolta@unil.ch
http://dx.doi.org/10.1016/j.ajhg.2016.07.009.

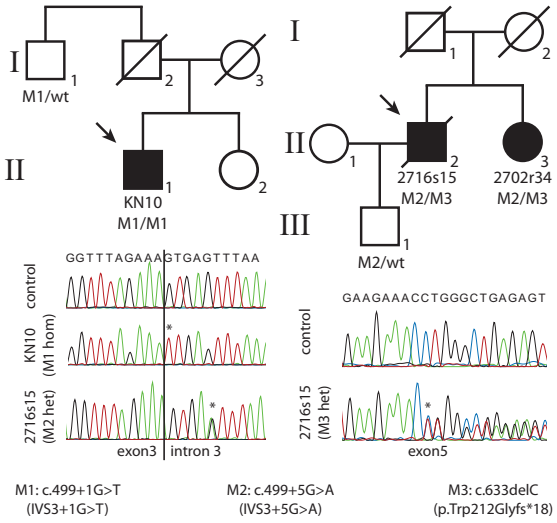


Figure 1. *CEP78* Mutations
Pedigrees and electropherograms of the identified DNA changes. Asterisks indicate the site of mutations. Abbreviations are as follows: Mx, mutant alleles; wt, wild-type allele.

and scored according to a specific in silico pipeline, described previously.[7] Aggregate data analysis and variant filtering procedures (Tables S1 and S2 in the Supplemental Data available online) revealed biallelic mutations in two probands, one from Greece and another from Sweden, in *CEP78* (centrosomal protein 78, composed of 16 exons for its longest coding isoform [GenBank: NM_001098802.1]). Both had classical signs and symptoms of CRD, clearly distinct from retinitis pigmentosa, as detailed below. Written informed consent was obtained from all individuals enrolled in this study, and approval for research on human subjects was obtained from the institutional review boards of all participating Institutions.

The Greek subject (KN10, individual II-1 on the left pedigree in Figure 1) was a 59-year-old male from the island of Crete and the eldest of two siblings of a non-consanguineous family. His sister was unaffected, and the family reported no history of retinal degeneration. Clinical history indicated hemarelopia since early adulthood (18–20 years of age); the condition progressed to severe central vision loss at the age of 35–40 years and evolved into severe visual impairment, nystagmus, and photophobia. Dyschromatopsia was also reported. Fundus examination at first visit showed normal color and normal vessels but a small atrophic foveal area with subjacent ring-like glistening in one eye and bull's-eye maculopathy in the other eye. A few atrophic lesions were present in the inferior periphery in one eye (Figure 2). The 30° static automated perimetry revealed a diffuse suppression of the visual field in both eyes and a relative conservation of the peripapillary and superior periphery. Full-field electroretinography (ERG) showed flat cone responses but still some residual rod-mediated signals in the left eye. This person also complained about minor hearing problems, and his audiogram exhibited relatively mild deficit; nonetheless, such deficit was clearly distinct and more severe than natural age-related hearing loss (presbycusis)[9] (Figure 2). KN10 carried a homozygous substitution in the first invariant base of intron 3 splice donor site c.499+1G>T (IVS3+1G>T) (Figure 1). This variant was located within a very small stretch of homozygosity that was not statistically significant for autozygosity, possibly indicating a mutational founder effect of geographic origin (not shown). The only relative who could be tested was his paternal uncle (individual I-1, left pedigree), who carried this DNA change heterozygously (Figure 1).

The Swedish proband (2716s15, individual II-2 on the right pedigree in Figure 1), now deceased, was last examined at 69 years of age. He was born from unaffected parents and was the first child of a kindred of two. His sister (2702r34, individual II-3, right pedigree), examined at age 65, also had retinal degeneration. Both had visual problems, including loss of color sensitivity and central vision, since childhood. Both also reported a hearing deficit since they were young, and both had hearing aids. Audiogram of the living Swedish subject at age 66 years revealed substantial sensorineural hearing loss, which did not seem to progress substantially over the following 11 years (Figure 2). Hospital records containing information on the hearing status of her deceased brother were destroyed upon his death. Fundus examination showed degenerative changes for both siblings in the macular region and some spicular pigment in the mid-periphery but fewer changes in the periphery (Figure 2). Progressive deterioration of the visual field was reported and documented as expanding from the center to the periphery. At last examinations, both siblings retained some residual vision at the periphery of the visual field, especially in dim-light conditions. Similar to the situation for the Greek subject, full-field ERG of both individuals highlighted almost no residual cone activity but still revealed some rod-mediated responses, even at these late ages. These siblings were compound heterozygotes for two *CEP78* mutations: a frameshift-causing single-nucleotide deletion (c.633delC; p.Trp212Glyfs*18) in exon 5 and an intronic base substitution (c.499+5G>A; IVS3+5G>A) in the vicinity of the donor site for intron 3, just four nucleotides away from the mutation identified in the Greek subject. Genetic examination of the proband's son (individual III-1, right pedigree) uncovered the presence of this latter mutation in heterozygosis, confirming the biallelic nature of the changes detected in his father and his aunt (Figure 1).

Sanger sequencing of the entire reading frame of *CEP78* in a cohort of 99 unrelated CRD-affected individuals of Swedish, Swiss, Dutch, and Pakistani ethnic background failed to identify any additional causative variants. The three mutations present in our two families were not detected in the genome of an internal control cohort of 350 unrelated individuals or in any other public

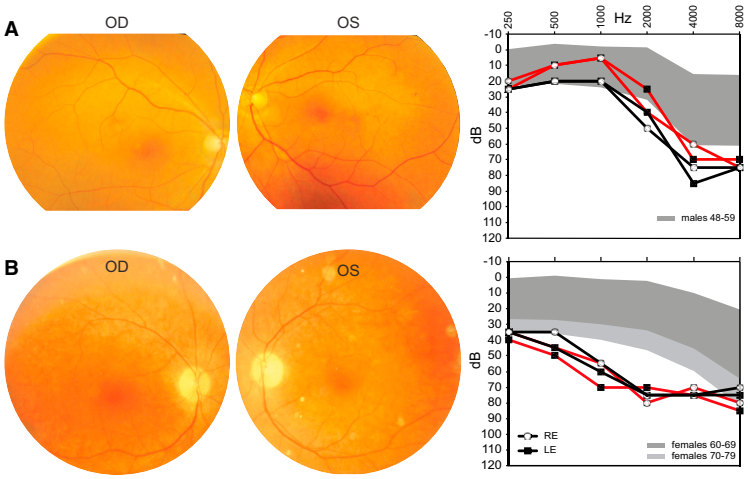


Figure 2. Clinical Features of the Analyzed Subjects
(A) Individual KN10. Fundus pictures (at age 53 years) reveal macular coalescent hypochromatic lesions in the right eye and minor macular changes in the left eye. Pure tone audiograms at age 57 years (black lines) and 59 years (red lines) show mild hearing impairment at higher frequencies compared to the normal range for gender and age, indicated by the shaded area.[8]
(B) Individual 2702r34. Fundus examination at age 65 years highlights attenuated vessels and degenerative changes in the posterior pole. Audiograms at age 66 years (black lines) and 77 years (red lines) show clear hearing loss at most frequencies. Abbreviations are as follows: OD, right eye; OS, left eye; RE, right ear; and LE, left ear.

database, including the 1000 Genomes Project, the Exome Variant Server (EVS), and the Exome Aggregation Consortium (ExAC) Browser, which reports sequencing data from more than 61,000 unrelated individuals. In addition, in silico assessment of the c.499+1G>T and c.499+5G>A mutations via two distinct web-based platforms, NNSPLICE 0.9[10] and Human Splicing Finder,[11] predicted for both variants the abolishment of the donor splicing site for intron 3.

To analyze the functional consequences of the three mutations, we obtained fresh blood samples and skin biopsies from the Greek proband and the living Swedish subject and performed the following experiments. We first retrotranscribed total RNA from immortalized lymphoblasts (GoScript Reverse Transcriptase, Promega). Then, after performing saturating RT-PCR of the region spanning all mutations (primers: 5′-TTTTGCAGAAGTCGTGTTCCT-3′ and 5′-TTCAAGGGCCTCTAGCAAAG-3′), we cloned the amplified products in *E. coli* (Zero Blunt PCR Cloning Kits, Invitrogen) and performed colony PCRs and capillary electrophoresis on 96 clones (48 clones per affected individual). Representative samples were Sanger sequenced, and relative numbers of splicing events were assessed and quantified. The c.499+1G>T mutation resulted invariantly in the skipping of exon 3, leading to the production of an aberrant isoform, never reported in genomic databases, for which exons 2 and 4 were joined together. This event ablated 24 codons and altered the reading frame of *CEP78*, leading to the formation of a premature termination codon at nucleotides 16–18 of exon 4. Therefore, this non-canonical transcript was predicted to trigger nonsense-mediated mRNA decay (NMD)[12] and result in no viable mRNA. The same exon-skipping occurrence was observed for the other, nearby mutation, c.499+5G>A, that was present in the Swedish subjects. Finally, the frameshift mutation c.633delC resulted in reduced mRNA amounts, as deduced by the low number of *E. coli* colonies carrying this cDNA clone (4 out of 48), again, probably as a result of the action of NMD. Immunoblot analysis in fibroblasts' extracts (antibodies: A301-799A, Bethyl Laboratories, epitope between residues 550 and 600 and A2066, Sigma, for beta actin) revealed the presence of CEP78 in very reduced amounts in the Greek subject and the absence of any detectable band in the Swedish subject, in agreement with the mRNA findings described above (Figure 3). More specifically, the homozygous c.499+1G>T mutation probably resulted in a few canonical mRNA forms not detected by our cloning experiments, in turn producing small amounts of CEP78. Concerning the Swedish subject, it is likely that both the c.633delC and the c.499+5G>A alleles produced mostly non-viable mRNA and minimal quantities of wild-type mRNA and protein. In fact, overexposed films showed a faint band corresponding to CEP78, indicating that the protein was present in trace amounts (not shown).

To gain insights into the relationship between vision and *CEP78*, we analyzed its expression in a panel of human tissues of cadaveric origin (Human Total RNA Master Panel II, Takara, primers 5′-GTTTCCCATTAATCAAAACACG-3′ and 5′-TCAACTTCAGAGGATGAAGGACT-3′ for *CEP78* and 5′-AGAGTGGTGCTGAGGATTGG-3′ and 5′-CCCTC ATGCTCTAGCGTGTC-3′ for the housekeeping gene *GUSB*). Although the number of *CEP78* transcripts in the retina was higher than in many other tissues and organs, retinal *CEP78* mRNA did not display the highest expression level (Figure 4). A time-course experiment on eyes of developing and postnatal mice (primers 5′-CTTCAGA AAGTGTCCAGGAAGC-3′ and 5′-GATCACTCTCTCCTCC TTCAGC-3′ for *Cep78* and 5′-CCTAAGATGAGCGCAAGTT GAA-3′ and 5′-CCACAGGACTAGAACACCTGCTAA-3′ for the housekeeping gene *Hprt1*) revealed high expression at embryonic stages, followed by a progressive decrease at perinatal stages and by a plateau at adulthood (Figure 4). This pattern is reminiscent of the expression of other genes involved in retinal degenerations and in early biogenesis and homeostasis of the centriole, notably

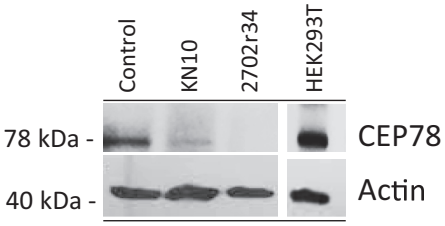


Figure 3. Immunoblot Analysis of Endogenous CEP78 in Fibroblasts from Affected Individuals
Numbers on the left refer to molecular-size markers. *Control* indicates human fibroblasts from a control individual; *HEK293T* indicates HEK293 cells containing SV40 large T antigen as a control for specificity of the anti-CEP78 antibody.

CEP76, *CEP110*, *CEP164* (MIM: 614848), *RAB8A* (MIM: 165040), *BBS4* (MIM: 600374), and *RPGR* (MIM: 312610).[13] Although we could not obtain primary data on *CEP78* expression in human cochlea, we assessed its RNaseq values from the only comprehensive transcriptome repository currently available for the inner ear.[14] *CEP78* displayed a FPKM (fragments per kilobase of exon per million reads mapped) value of 4.53 (average from three human cochleae), indicating moderate expression in this structure. Importantly, this FPKM value for *CEP78* appeared to be higher than that of most genes already known to be associated with hereditary deafness (40 out of 70, or 57%), as assessed in the same organs and in the same conditions (Table S3).

Little is known about the function of CEP78. Identified as a component of the centrosome by two independent proteomic screenings,[15,16] CEP78 is composed of five leucine-rich repeats located at the N-terminal half of the protein, as well as a coiled-coil domain at the C terminus. An important study using *Planaria* as the main experimental model revealed that miRNA-based knockdown of *CEP78* resulted in defective primary cilia assembly in flatworms and human RPE1 cells.[17] Intriguingly, *CEP78* was also found upregulated more than 5-fold by noise stress in rat cochlea.[18] The function and impact of CEP78 in human physiology, however, remain largely elusive.

The reported centrosomal localization of CEP78 prompted us to investigate a possible role in relationship to the photoreceptor primary cilium. Immunofluorescence of human retinal sections with anti-CEP78 antibody (IHC-00364, Bethyl Laboratories, epitope between residues 550 and 600) and anti-cone arrestin (SC-54355, Santa Cruz) showed that CEP78 is present in dot-shaped foci in the inner segments, probably at the base of the primary cilium in retinal photoreceptors, predominantly cones (Figure 5 and Figure S1). This observation was confirmed when CEP78 was labeled together with acetylated tubulin, staining the primary cilium of human skin fibroblasts (Figure 5). Interestingly, positive staining was observed in fibroblasts from KN10 and 2702r34 as well, confirming that *CEP78* was in fact expressed at the protein level in these individuals, as inferred (Figure 5). The mild differences between these experiments and the protein-expression observations might be due to different sensitivities of the two analytical tools and the use of distinct antibodies (A301-799A for immunoblot and IHC-00364 for immunofluorescence). No specific differences concerning CEP78 subcellular localization were observed in cells from affected individuals versus cells from controls.

Presence at the base of the connecting cilium is a characteristic that is shared by other proteins associated with retinal degeneration, and in particular by FAM161A, the deficiency of which causes the RP28 form of retinitis pigmentosa (MIM: 606068).[19–21] Indeed, tandem-affinity purification analysis performed with full-length FAM161A showed a positive interaction with CEP78,[22] and co-immunoprecipitation with an anti-CEP78 antibody (A301-800A, Bethyl Laboratories, epitope between residues 639 and

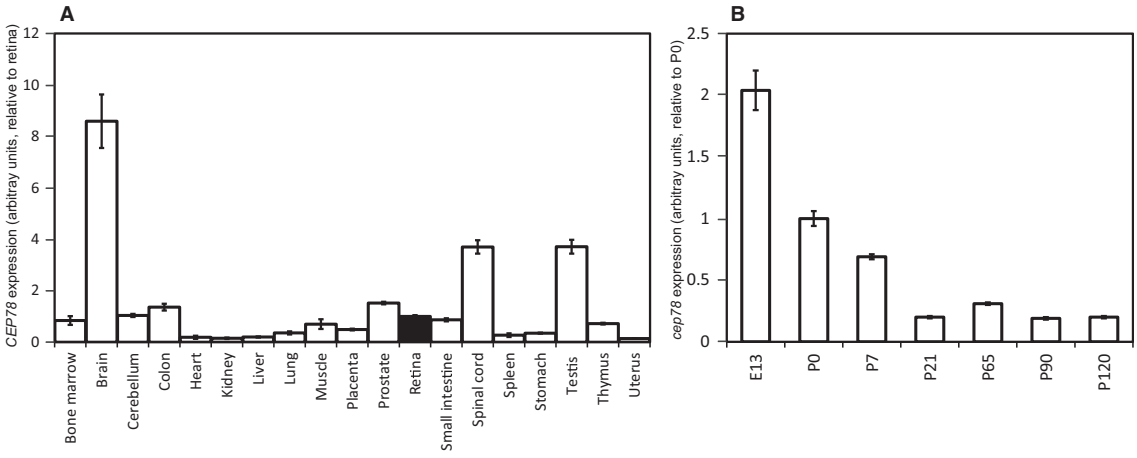


Figure 4. *CEP78* mRNA Expression in Various Human Tissues and Organs and in the Developing Murine Eye
Data are from real-time PCR relative expression analysis, for which *GUSB* and *Hprt1* were used as normalizing genes for (A) and (B), respectively.

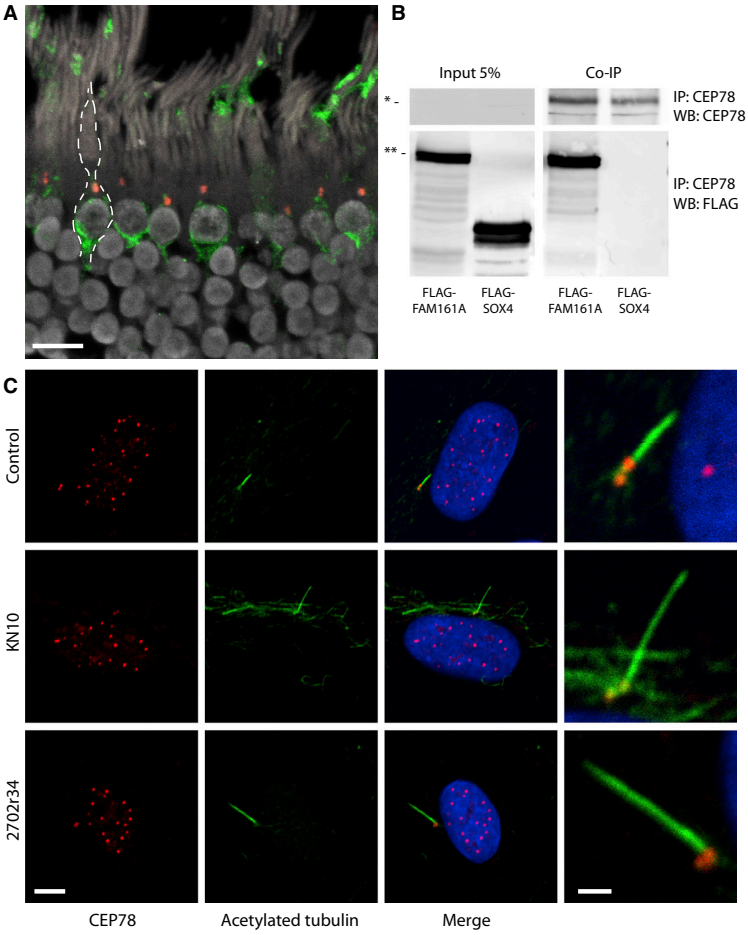


Figure 5. CEP78 in Human Cells and Its Interaction with FAM161A

(A) Immunostaining of CEP78 (red dots) and of cone arrestin (green) in a section of human retina. Margins of a cone photoreceptor are highlighted by a dotted line. The scale bar represents 10 μm.

(B). Co-immunoprecipitation of endogenous CEP78 and FLAG-tagged FAM161A in HEK293T cells. FLAG-SOX4 is a negative control. Abbreviations are as follows: IP:x, protein targeted by the antibody used in immunoprecipitation; and WB:x, protein or peptide targeted by the antibody used in immunoblots. The single and double asterisks indicate the expected molecular sizes for CEP78 (78 kDa) and the FLAG-FAM161A construct (90 kDa), respectively.

(C) Staining of CEP78 (red) and acetylated tubulin (green) in fibroblasts from a control individual, KN10, and 2702r34. CEP78 localizes at the centrioles and at the base of the induced primary cilium. The scale bars represent 5 and 1 μm for regular and magnification panels, respectively.

689) in HEK293T cells transfected with full-length FLAG-FAM161A revealed direct binding between these two ciliary proteins (Figure 5).

On the basis of these findings, we speculated that CRD due to mutations in *CEP78* could be a consequence of hindered ciliary function, similar to what occurs in many other retinal degenerations.[23] To test this hypothesis, we analyzed the morphology of primary cilia in fibroblasts derived from KN10 and 2702r34, with respect to four controls, after serum starvation. Unsupervised counting of at least 82 events per sample (207 events in affected individuals and 430 in controls) revealed that induced cilia in fibroblasts from KN10 and 2702r34 were significantly longer than those from control cells (Figure 6 and Figure S2), a phenomenon that has been previously associated with impaired function of this organelle. For instance, mutations in murine orthologs of *BBS4*, *ICK,* and *TSC1*, linked with ciliopathies such as Bardet-Biedl syndrome, endocrine-cerebro-osteodysplasia, and tuberous sclerosis, respectively, display kidney cells with elongated primary cilia.[24–26]

In recent years, a significant number of hereditary conditions have been recognized to be the consequence of abnormalities of the cellular cilium. These diseases, collectively called ciliopathies, form a genetically heterogeneous spectrum of disorders affecting various tissues and organs, for instance kidney, cochlea, brain, and retina.[23,27] Classical examples of ciliopathies involving retina and other tissues are Usher syndrome (MIM: 276900, blindness and deafness) and Bardet-Biedl syndrome (BBS [MIM: 209900], blindness and multi-organ defects) for both of which vision loss is due to retinitis pigmentosa.[28] In addition, syndromic ciliopathies such as Senior–Løken syndrome (SLS [MIM: 266900]), Joubert syndrome (JBTS [MIM: 213300]), and Jeune syndrome (JATD [MIM: 616300]) can occasionally be accompanied by retinal dystrophy and, in particular, retinitis pigmentosa and/or Leber congenital amaurosis (LCA [MIM: 204000]).[29–32] Another multi-organ ciliopathy is Alström syndrome (ALMS [MIM: 203800]),[33] caused by null mutations in the gene *ALMS1* (MIM: 606844). This disease is characterized by cone-rod degeneration, dilated cardiomyopathy, obesity, type 2 diabetes, and short stature, which can be accompanied by hepatosteatosis and defects in the lungs, kidney, and bladder.[33,34] Most cases also display progressive sensorineural hearing loss.[35] Of interest, mutations in *ALMS1* have recently been suggested as being causative of non-syndromic CRD.[36]

In this study, we show that mutations in *CEP78* result in cone-rod degeneration associated with hearing loss,

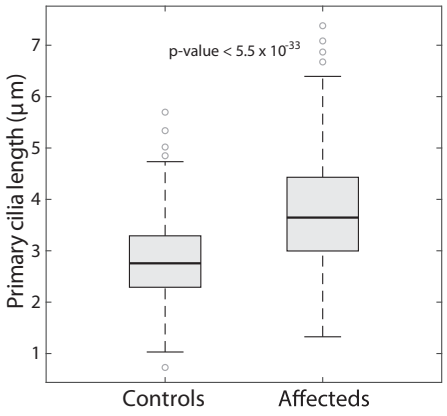


Figure 6. Analysis of Ciliary Lengths
Fibroblasts from affected subjects display significantly longer primary cilia than those from four unaffected controls, suggesting impaired functionality.

another hallmark of ciliopathy, but no other syndromic features. Interestingly, the two Swedish affected individuals had declared hearing loss, and the Greek subject had borderline hearing impairment. An intriguing possibility involves the presence of a genotype-phenotype correlation between *CEP78* alleles and hearing (but not vision), as it is the case for instance for mutations in *USH2A* and *ALMS1*.[36–38] Taken together, our data indicate that genetic defects in *CEP78* define a newly recognized ciliopathy, distinct from Usher and Alström syndromes, affecting both the visual and the hearing systems.

Accession Numbers

The mutations reported in this paper have been deposited in the Leiden Open Variation Database (LOVD) under variant ID numbers 0000119260, 0000119261, and 0000119262.

Supplemental Data

Supplemental Data include two figures and three tables and are available with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.07.009.

Acknowledgments

This work was supported by the Swiss National Science Foundation (grant #156260, to C.R.), by the European Community's Seventh Framework Programmes FP7/2009 under grant agreement #241955 (SYSCILIA) to R.R. and M.U., by the Netherlands Organization for Scientific Research (NWO Vici-865.12.005 to R.R), and by the Rotterdamse Stichting Blindenbelangen, the Stichting Blindenhulp, the Stichting A.F. Deutman Researchfonds Oogheelkunde, and the Stichting voor Ooglijders (to F.P.M.C. and M.I.K.). F.P.M.C. and M.I.K. were also supported by the following foundations: the Algemene Nederlandse Vereniging ter Voorkoming van Blindheid, the Landelijke Stichting voor Blinden en Slechtzienden, the Stichting Retina Nederland Fonds, and the Novartis fund, contributed through UitZicht. We would also like to acknowledge Drs. Andrea Superti-Furga and Luisa Bonafé from the Center for Molecular Diseases of the Lausanne University Hospital, Dr. Frauke Coppieters from the Ghent University Hospital, Dr. Guy Van Camp from the University of Antwerp, and Dr. Isabelle Schrauwen from the Translational Genomics Research Institute in Phoenix.

Received: March 17, 2016
Accepted: July 7, 2016
Published: September 1, 2016

Web Resources

1000 Genomes, http://browser.1000genomes.org
ExAC Browser, http://exac.broadinstitute.org/
GenBank, http://www.ncbi.nlm.nih.gov/genbank/
Leiden Open Variation Database (LOVD), http://www.lovd.nl/3.0/home
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
OMIM, http://www.omim.org/
RefSeq, http://www.ncbi.nlm.nih.gov/RefSeq
RetNet, https://sph.uth.edu/retnet/

References

1. Hamel, C.P. (2007). Cone rod dystrophies. Orphanet J. Rare Dis. *2*, 7.
2. Berson, E.L., Gouras, P., and Gunkel, R.D. (1968). Progressive cone-rod degeneration. Arch. Ophthalmol. *80*, 68–76.
3. Krauss, H.R., and Heckenlively, J.R. (1982). Visual field changes in cone-rod degenerations. Arch. Ophthalmol. *100*, 1784–1790.
4. Berson, E.L. (1993). Retinitis pigmentosa. The Friedenwald Lecture. Invest. Ophthalmol. Vis. Sci. *34*, 1659–1676.
5. Berger, W., Kloeckener-Gruissem, B., and Neidhardt, J. (2010). The molecular basis of human retinal and vitreoretinal diseases. Prog. Retin. Eye Res. *29*, 335–375.
6. Roosing, S., Thiadens, A.A., Hoyng, C.B., Klaver, C.C., den Hollander, A.I., and Cremers, F.P. (2014). Causes and consequences of inherited cone disorders. Prog. Retin. Eye Res. *42*, 1–26.
7. Royer-Bertrand, B., Castillo-Taucher, S., Moreno-Salinas, R., Cho, T.J., Chae, J.H., Choi, M., Kim, O.H., Dikoglu, E., Campos-Xavier, B., Girardi, E., et al. (2015). Mutations in the heat-shock protein A9 (HSPA9) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia. Sci. Rep. *5*, 17154.
8. Cruickshanks, K.J., Wiley, T.L., Tweed, T.S., Klein, B.E., Klein, R., Mares-Perlman, J.A., and Nondahl, D.M.; The Epidemiology of Hearing Loss Study (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. Am. J. Epidemiol. *148*, 879–886.
9. Jerger, J., Chmiel, R., Stach, B., and Spretnjak, M. (1993). Gender affects audiometric shape in presbyacusis. J. Am. Acad. Audiol. *4*, 42–49.
10. Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. J. Comput. Biol. *4*, 311–323.
11. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., and Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. *37*, e67.

12. Hentze, M.W., and Kulozik, A.E. (1999). A perfect message: RNA surveillance and nonsense-mediated decay. Cell *96*, 307–310.
13. Zhang, S.S., Xu, X., Liu, M.G., Zhao, H., Soares, M.B., Barnstable, C.J., and Fu, X.Y. (2006). A biphasic pattern of gene expression during mouse retina development. BMC Dev. Biol. *6*, 48.
14. Schrauwen, I., Hasin-Brumshtein, Y., Corneveaux, J.J., Ohmen, J., White, C., Allen, A.N., Lusis, A.J., Van Camp, G., Huentelman, M.J., and Friedman, R.A. (2016). A comprehensive catalogue of the coding and non-coding transcripts of the human inner ear. Hear. Res. *333*, 266–274.
15. Andersen, J.S., Wilkinson, C.J., Mayor, T., Mortensen, P., Nigg, E.A., and Mann, M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. Nature *426*, 570–574.
16. Jakobsen, L., Vanselow, K., Skogs, M., Toyoda, Y., Lundberg, E., Poser, I., Falkenby, L.G., Bennetzen, M., Westendorf, J., Nigg, E.A., et al. (2011). Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. EMBO J. *30*, 1520–1535.
17. Azimzadeh, J., Wong, M.L., Downhour, D.M., Sánchez Alvarado, A., and Marshall, W.F. (2012). Centrosome loss in the evolution of planarians. Science *335*, 461–463.
18. Han, Y., Hong, L., Zhong, C., Chen, Y., Wang, Y., Mao, X., Zhao, D., and Qiu, J. (2012). Identification of new altered genes in rat cochleae with noise-induced hearing loss. Gene *499*, 318–322.
19. Bandah-Rozenfeld, D., Mizrahi-Meissonnier, L., Farhy, C., Obolensky, A., Chowers, I., Pe'er, J., Merin, S., Ben-Yosef, T., Ashery-Padan, R., Banin, E., and Sharon, D. (2010). Homozygosity mapping reveals null mutations in *FAM161A* as a cause of autosomal-recessive retinitis pigmentosa. Am. J. Hum. Genet. *87*, 382–391.
20. Langmann, T., Di Gioia, S.A., Rau, I., Stöhr, H., Maksimovic, N.S., Corbo, J.C., Renner, A.B., Zrenner, E., Kumaramanickavel, G., Karlstetter, M., et al. (2010). Nonsense mutations in *FAM161A* cause RP28-associated recessive retinitis pigmentosa. Am. J. Hum. Genet. *87*, 376–381.
21. Di Gioia, S.A., Letteboer, S.J., Kostic, C., Bandah-Rozenfeld, D., Hetterschijt, L., Sharon, D., Arsenijevic, Y., Roepman, R., and Rivolta, C. (2012). FAM161A, associated with retinitis pigmentosa, is a component of the cilia-basal body complex and interacts with proteins involved in ciliopathies. Hum. Mol. Genet. *21*, 5174–5184.
22. Di Gioia, S.A. (2013). Targeted sequence capture and ultra high throughput sequencing for gene discovery in inherited diseases. PhD Thesis (University of Lausanne, Lausanne, Switzerland).
23. Wright, A.F., Chakarova, C.F., Abd El-Aziz, M.M., and Bhattacharya, S.S. (2010). Photoreceptor degeneration: Genetic and mechanistic dissection of a complex trait. Nat. Rev. Genet. *11*, 273–284.
24. Armour, E.A., Carson, R.P., and Ess, K.C. (2012). Cystogenesis and elongated primary cilia in *Tsc1*-deficient distal convoluted tubules. Am. J. Physiol. Renal Physiol. *303*, F584–F592.
25. Moon, H., Song, J., Shin, J.O., Lee, H., Kim, H.K., Eggenschwiller, J.T., Bok, J., and Ko, H.W. (2014). Intestinal cell kinase, a protein associated with endocrine-cerebro-osteodysplasia syndrome, is a key regulator of cilia length and Hedgehog signaling. Proc. Natl. Acad. Sci. USA *111*, 8541–8546.
26. Mokrzan, E.M., Lewis, J.S., and Mykytyn, K. (2007). Differences in renal tubule primary cilia length in a mouse model of Bardet-Biedl syndrome. Nephron, Exp. Nephrol. *106*, e88–e96.
27. Waters, A.M., and Beales, P.L. (2011). Ciliopathies: An expanding disease spectrum. Pediatr. Nephrol. *26*, 1039–1056.
28. Koenig, R. (2003). Bardet-Biedl syndrome and Usher syndrome. Dev. Ophthalmol. *37*, 126–140.
29. Bachmann-Gagescu, R., Dempsey, J.C., Phelps, I.G., O'Roak, B.J., Knutzen, D.M., Rue, T.C., Ishak, G.E., Isabella, C.R., Gorden, N., Adkins, J., et al.; University of Washington Center for Mendelian Genomics (2015). Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity. J. Med. Genet. *52*, 514–522.
30. Bard, L.A., Bard, P.A., Owens, G.W., and Hall, B.D. (1978). Retinal involvement in thoracic-pelvic-phalangeal dystrophy. Arch. Ophthalmol. *96*, 278–281.
31. Wilson, D.J., Weleber, R.G., and Beals, R.K. (1987). Retinal dystrophy in Jeune's syndrome. Arch. Ophthalmol. *105*, 651–657.
32. Hildebrandt, F., and Zhou, W. (2007). Nephronophthisis-associated ciliopathies. J. Am. Soc. Nephrol. *18*, 1855–1871.
33. Marshall, J.D., Muller, J., Collin, G.B., Milan, G., Kingsmore, S.F., Dinwiddie, D., Farrow, E.G., Miller, N.A., Favaretto, F., Maffei, P., et al. (2015). Alström syndrome: Mutation spectrum of *ALMS1*. Hum. Mutat. *36*, 660–668.
34. Marshall, J.D., Maffei, P., Collin, G.B., and Naggert, J.K. (2011). Alström syndrome: genetics and clinical overview. Curr. Genomics *12*, 225–235.
35. Marshall, J.D., Bronson, R.T., Collin, G.B., Nordstrom, A.D., Maffei, P., Paisey, R.B., Carey, C., Macdermott, S., Russell-Eggitt, I., Shea, S.E., et al. (2005). New Alström syndrome phenotypes based on the evaluation of 182 cases. Arch. Intern. Med. *165*, 675–683.
36. Lazar, C.H., Kimchi, A., Namburi, P., Mutsuddi, M., Zelinger, L., Beryozkin, A., Ben-Simhon, S., Obolensky, A., Ben-Neriah, Z., Argov, Z., et al. (2015). Nonsyndromic early-onset cone-rod dystrophy and limb-girdle muscular dystrophy in a consanguineous Israeli family are caused by two independent yet linked mutations in *ALMS1* and *DYSF*. Hum. Mutat. *36*, 836–841.
37. Rivolta, C., Sweklo, E.A., Berson, E.L., and Dryja, T.P. (2000). Missense mutation in the *USH2A* gene: association with recessive retinitis pigmentosa without hearing loss. Am. J. Hum. Genet. *66*, 1975–1978.
38. Lenassi, E., Vincent, A., Li, Z., Saihan, Z., Coffey, A.J., Steele-Stallard, H.B., Moore, A.T., Steel, K.P., Luxon, L.M., Héon, E., et al. (2015). A detailed clinical and molecular survey of subjects with nonsyndromic USH2A retinopathy reveals an allelic hierarchy of disease-causing variants. Eur. J. Hum. Genet. *23*, 1318–1327.

6.5 TTLL5

Mutations in the polyglutamylase gene TTLL5, expressed in photoreceptor cells and spermatozoa, are associated with cone-rod degeneration and reduced male fertility

Own contribution: Helped in the analytic bioinformatics procedures, which identify the disease-causing gene.

HMG Advance Access published September 15, 2016

Human Molecular Genetics, 2016, Vol. 0, No. 0 1–10

doi: 10.1093/hmg/ddw282
Advance Access Publication Date: 22 August 2016
Original Article

ORIGINAL ARTICLE

Mutations in the polyglutamylase gene *TTLL5*, expressed in photoreceptor cells and spermatozoa, are associated with cone-rod degeneration and reduced male fertility

Nicola Bedoni[1,†], Lonneke Haer-Wigman[2,3,†], Veronika Vaclavik[4,5], Viet H. Tran[4], Pietro Farinelli[1], Sara Balzano[1], Beryl Royer-Bertrand[1,6], Mohammed E. El-Asrag[7], Olivier Bonny[8], Christos Ikonomidis[9], Yan Litzistorf[9], Konstantinos Nikopoulos[1], Georgia G. Yioti[10], Maria I. Stefaniotou[10], Martin McKibbin[11], Adam P. Booth[12], Jamie M. Ellingford[13], Graeme C. Black[13], Carmel Toomes[7], Chris F. Inglehearn[7], Carel B. Hoyng[14], Nathalie Bax[14], Caroline C.W. Klaver[14,15], Alberta A. Thiadens[15], Fabien Murisier[16], Daniel F. Schorderet[17], Manir Ali[7], Frans P.M. Cremers[2,3], Sten Andréasson[18], Francis L. Munier[4,†] and Carlo Rivolta[1,†,*]

[1]Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, Lausanne, Switzerland, [2]Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands, [3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands, [4]Jules Gonin Eye Hospital, Lausanne, Switzerland, [5]Department of Ophthalmology, Fribourg Hospital HFR, Fribourg, Switzerland, [6]Center for Molecular Diseases, Department of Pediatrics, Lausanne University Hospital (CHUV), Lausanne, Switzerland, [7]Section of Ophthalmology & Neuroscience, Leeds Institute of Biomedical & Clinical Sciences, University of Leeds, Leeds, UK, [8]Service of Nephrology, Lausanne University Hospital (CHUV), Lausanne, Switzerland, [9]Department of Otorhinolaryngology, Head and Neck Surgery, Lausanne University Hospital (CHUV), Lausanne, Switzerland, [10]Department of Ophthalmology, University of Ioannina School of Medicine, Ioannina, Greece, [11]The Eye Department, St. James's University Hospital, Leeds, UK, [12]Royal Eye Infirmary, Derriford Hospital, Plymouth, UK, [13]Centre for Genomic Medicine, St. Mary's Hospital, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK, [14]Department of Ophthalmology, Radboud University Medical Center, Nijmegen, the Netherlands, [15]Department of Ophthalmology, Erasmus Medical Center, Rotterdam, the Netherlands, [16]Fertas Andrology Laboratory, Lausanne, Switzerland, [17]Institute for Research in Ophtalmology, University of Lausanne and Ecole Polytechnique Federale de Lausanne, Switzerland and [18]Department of Ophthalmology, Lund University, Lund, Sweden

[†]The authors wish it to be known that, in their opinion, the first 2 authors and last 2 authors should be regarded as joint First and Last Authors, respectively.
Received: July 8, 2016. **Revised:** August 19, 2016. **Accepted:** August 20, 2016

*To whom correspondence should be addressed at: Carlo Rivolta, Department of Computational Biology, Unit of Medical Genetics, University of Lausanne, Rue du Bugnon 27, 1011 Lausanne, Switzerland. Tel: +41-21-6925451; Fax: +41-21-6925455; Email: carlo.rivolta@unil.ch

Abstract

Hereditary retinal degenerations encompass a group of genetic diseases characterized by extreme clinical variability. Following next-generation sequencing and autozygome-based screening of patients presenting with a peculiar, recessive form of cone-dominated retinopathy, we identified five homozygous variants [p.(Asp594fs), p.(Gln117*), p.(Met712fs), p.(Ile756Phe), and p.(Glu543Lys)] in the polyglutamylase-encoding gene *TTLL5*, in eight patients from six families. The two male patients carrying truncating *TTLL5* variants also displayed a substantial reduction in sperm motility and infertility, whereas those carrying missense changes were fertile. Defects in this polyglutamylase in humans have recently been associated with cone photoreceptor dystrophy, while mouse models carrying truncating mutations in the same gene also display reduced fertility in male animals. We examined the expression levels of *TTLL5* in various human tissues and determined that this gene has multiple viable isoforms, being highly expressed in testis and retina. In addition, antibodies against TTLL5 stained the basal body of photoreceptor cells in rat and the centrosome of the spermatozoon flagellum in humans, suggesting a common mechanism of action in these two cell types. Taken together, our data indicate that mutations in *TTLL5* delineate a novel, allele-specific syndrome causing defects in two as yet pathogenically unrelated functions, reproduction and vision.

Introduction

Cone dystrophies (CDs) and cone-rod dystrophies (CRDs) are rare heterogeneous retinal disorders with an estimated prevalence of ~1:40,000 (1). They lead to severe visual impairment, primarily or exclusively due to the degeneration of cone photoreceptors. Patients experience progressive loss of visual acuity, defective colour vision, photophobia, and have central scotomas. Only later, as the disease progresses, in some cases loss of peripheral vision may also occur (2,3).

The progressive degeneration of retinal photoreceptors in CDs and CRDs is mostly nonsyndromic and has been associated with multiple genetic causes, with at least 20 associated disease genes (RetNet; http://www.sph.uth.tmc.edu/RetNet/; date last accessed July 8, 2016). However, more than 75% of cases presenting with dominant or recessive forms of these conditions are genetically unsolved (4). Recent discoveries in CD molecular genetics include the identification of pathogenic variants in the tubulin polyglutamylase *TTLL5* (Tubulin Tyrosine Ligase-Like Protein 5) gene, found to cause retinal dystrophy in four British families (5). This gene, like the 12 other members of the TTLL superfamily, is involved in post-translational modifications of α- and β-tubulin, which are components of the axonemes of both cilia and flagella.

Interestingly, male mice with a defective TTLL5 display dramatically reduced fertility associated with defects in sperm motility (6). Most sperm tails of mutant mice were found to have disrupted axonemes with loss of tubulin doublets and a significantly decreased polyglutamylation in the upper and lower segments. No abnormal phenotype of retinal photoreceptors or of cochlear cells was initially observed, based on histologic examination (6). A second study with a more thorough characterization of the ocular phenotype in the same mouse model showed a decline of electroretinographic (ERG) amplitudes for both rods and cones in aged mice (20-22 mo). However, no microtubule defects were found after examination of electron micrographs (7).

Ciliopathies represent a class of hereditary disorders involving deficiencies in ciliary and cilia-associated proteins, often affecting a variety of tissues and organs (8). Due to the presence of an immotile cilium in both rods and cones photoreceptors, many ciliopathies display a retinal phenotype, either as part of a syndromic condition (associated with hearing defects, renal nephronophtisis, liver fibrosis, bone and/or brain anomalies) or as the sole pathological sign (9–11).

Following the investigation of a cohort of patients displaying CD or CRD, we identified mutations in *TTLL5* that are associated with both retinal degeneration and reduced sperm motility in humans, possibly defining a novel syndromic ciliopathy.

Results

Clinical and molecular findings

Our research started with the molecular characterization of a Swiss male patient (P1), aged 75 years, presenting with a late-onset cone dystrophy (CD). He was the eldest of three brothers and his parents were first cousins (Fig. 1, F1), without any history of ocular problems. The patient was first seen at the age of 33 years, when he first noticed blurred vision. His best corrected Snellen visual acuity (BCVA) at that time was 0.6 in the right eye and 0.8 in his left eye. Seven years later, his BCVA was still stable, but worsened when he was 53 years old, dropping dramatically to 0.05 in the right eye and 0.08 in the left eye. He was also complaining of a reduced dark adaptation. Twenty-two years later his vision remained stable, with the patient using more of his peripheral vision.

The clinical examination was typical of a cone dystrophy: the fundus examination showed central foveal atrophy with peripapillary hyperpigmentation and atrophy, while the peripheral retina was within normal limits (Fig. 2). The first ERGs (performed when the patient was 56 years old) showed normal scotopic responses, whereas photopic responses had severely reduced amplitude. The following ERG testing, when the patient was 70 years old, showed some rod involvement, with reduced rod-specific b-wave. The 30-Hz flicker was undetectable. Autofluorescent (AF) images at age 66 years (Fig. 2) showed central hypofluorescence corresponding to atrophy, surrounded by a large hyperfluorescent ring. AF imaging at age 72 indicated that the ring mildly increased in diameter, and so did the area of hypofluorescence (Fig. 2). Kinetic visual field tested at this later age showed mild constriction and central scotoma in both eyes.

The DNA of the patient was first screened for mutations in known disease genes. Following the negative output of a panel-

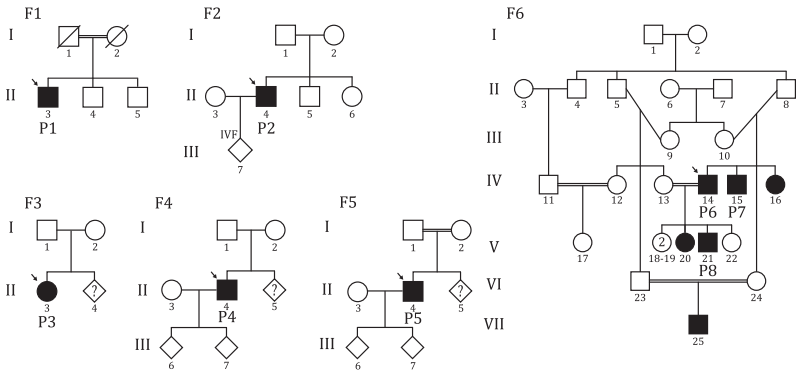


Figure 1. Pedigrees from families with pathogenic *TTLL5* variants. The probands are: subject P1 [II:3, family F1; c.1782del;p.(Asp594Glufs*29)], subject P2 [II:4, family F2; c.349C > T;p.(Gln117*)], subject P3 [II:3, family F3; c.2132_2133insGATA;p.(Met712IleAspfs*15)], subject P4 [II:4, family F4; c.1627G > A;p.(Glu543Lys)], subject P5 [II:4, family F5; c.2266A > T;p.(Ile756Phe)], and subject P6 [IV:14, family F6; c.1627G > A;p.(Glu543Lys)].

based Next-Generation Sequencing screening [the IROme (12)], we performed whole-genome sequencing (WGS) of the patient's DNA. This latter procedure resulted in more than 4 million DNA variants with respect to the human reference genome (Build hg19). These were evaluated by the use of an internal *in-silico* pipeline assessing their frequency in the general population, quality, etc. (Supplementary Material, Table S1), as well as their presence within autozygous regions (Supplementary Material, Fig. S1). At the end of this process, we were left with 19 variants (7 in autozygous regions), including c.1782delT in *TTLL5* (Fig. 3), a gene known to be involved in microtubule posttranslational modifications and associated with ciliary microtubule stabilization (13). The variant was in exon 20 and consisted of a 1-bp deletion causing a frameshift starting from codon 594 and terminating with the creation of a premature stop triplet 29 codons downstream [p.(Asp594Glufs*29)]. It was also not detected in the genome of 400 healthy controls from the same geographic region of Switzerland [data from the CoLaus study (14)], and from publicly available databases [ExAC (15), dbSNP (16)]. Importantly, this DNA change localized to a region of chromosome 14 (75,986,579-80,875,911, Build hg19) showing clear autozygosity (Supplementary Material, Fig. S1) in P1's genome.

Because of the possible involvement of TTLL5 in extraocular ciliary functions, the patient underwent more detailed clinical examinations. To assess the presence of additional subtle abnormalities in organs known to be affected by ciliopathies, he was evaluated for both renal and otorhinolaryngological functions. He did not display the classical clinical features of renal ciliopathy (no polyuria/polydipsia), but chronic renal insufficiency without proteinuria (stage G3aA1) (17) was identified, associated with low-grade chronic anemia. By ultrasound, the size of the kidneys appeared to be preserved, but a small asymmetry was noted (9.6 cm on the left vs. 10.7 cm on the right side), without significant renal artery stenosis. No cysts were visible and no biopsy was performed. Altogether, the renal features were consistent with age-related decreased renal function, but low-grade and late appearance nephronophtisis cannot be fully excluded. The observed minor sensorineural hearing loss, and loss of some high frequencies, was compatible with the natural course of hearing about a person of the age of the subject (Supplementary Material, Fig. S2). The clinical examination and patient's history did not reveal any upper airway pathology necessitating further investigations for impaired mucocillary function in the sinus or the bronchi. The patient reported a history of infertility due to reduced sperm motility, diagnosed when he was in his late 20's. He could not have offspring and adopted two children. Following our findings, a new semen analysis was performed at age 75, revealing azoospermia, a sign that nonetheless could simply be related to the difficulty in obtaining an ejaculate at his current age.

Based on these findings, we extended our analyses to a number of additional cohorts of 365 patients with CD and CRD from Switzerland, Sweden, Greece, The Netherlands, and Britain. In a Swedish male patient of Iraqi descent (P2), we identified the homozygous nonsense variant c.349C > T;p.(Gln117*) by targeted Sanger sequencing of *TTLL5* (Figs. 1, F2; 3). Again, this variant was absent from controls and publicly available databases. Although the patient did not report any history of consanguinity, the occurrence of this extremely rare variant in a homozygous state suggests the presence of residual consanguinity or of a geographical founder effect, which was not tested at the genome level. The patient, aged 46 years, reported no family history of similar visual impairment. Fundus examination revealed degenerative changes, especially in the posterior pole, but more normal features in the periphery. Visual field analysis by Goldmann perimetry showed residual fields in the periphery, but a large central scotoma. Full-field ERG demonstrated residual cone and rod response, consistent with a diagnosis of cone-rod degeneration. In addition to these signs and symptoms typical of CRD, the patient had high myopia (Table 1). This subject was also infertile, but had been able to have a child by *in vitro* fertilization. His semen analysis revealed a normal spermatozoa count but, similar to P1, reduced sperm motility. He also had morphologically normal spermatozoa for 5% of the count, over three independent assays.

Following autozygome-based analysis (Supplementary Material, Fig. S1), we identified homozygous mutations in three additional patients from the Netherlands (Fig. 1, F3–5). All of them were diagnosed with CD, with full-field ERGs showing reduction of cone function but preserved rod responses (Table 1). Similar to the cases described above, the female patient P3 had a *TTLL5* truncating mutation, c.2132_2133insGATA;p.(Met712IleAspfs*15), resulting in a premature termination codon. The other two patients (P4, and P5) were males and both carried homozygous missense mutations, namely c.1627G > A;p.(Glu543Lys) and c.2266A > T;p.(Ile756Phe),

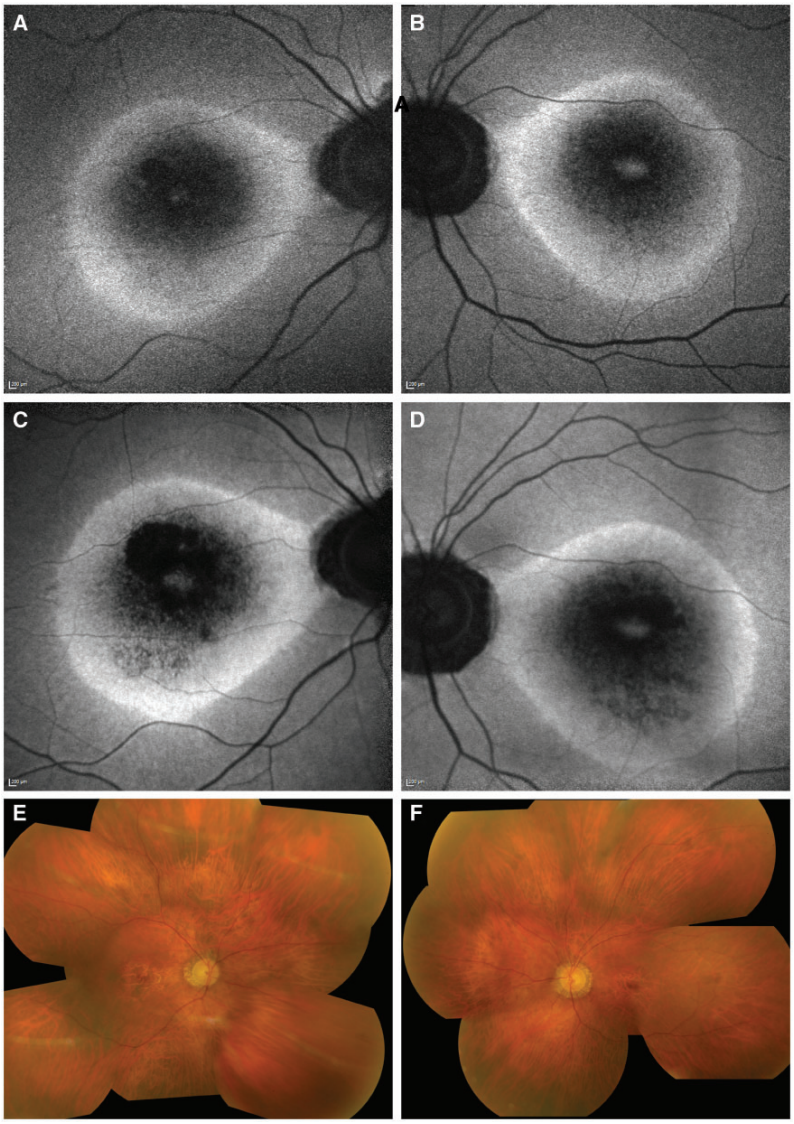


Figure 2. Fundi of index patient P1. Autofluorescence imaging of the right (**A**) and left (**B**) eye at age 66 years, showing a distinctive pattern of abnormalities, including a hyperfluorescent ring and hyperfluorescent area at the fovea surrounded by patchy hypofluorescence in both eyes. Hypofluorescence around the optic nerve was also present. The same images, obtained 6 years later (**C,D**), showed an increase of hypofluorescent areas within the ring and a mild enlargement of the hyperfluorescent ring in both eyes. Composite pictures of the fundi at age 72, showing atrophic areas around the fovea and around the optic nerve (**E,F**). Peripheral retina was within normal limits.

respectively (Fig. 3). Following colour vision testing, substantial mistakes were made in all three colour axes. All patients had a myopic refractive error, but none of them showed any additional extraocular symptoms. In particular, they did not report any fertility problems, and both male patients had offspring.

Finally, a male patient (P6) of Pakistani origin was screened by whole-exome sequencing (WES). He was a member of a consanguineous pedigree that included six additional individuals with high myopia and an acquired CD or CRD with loss of corrected visual acuity from the second decade onwards (Fig. 1, F6). In this case as well, a homozygous mutation in *TTLL5* was identified within an autozygous region in chromosome 14 (Supplementary Material, Fig. S1). It was the same missense detected above [c.1627G > A;p.(Glu543Lys), Fig. 3], which perfectly co-segregated with the disease in the three affected and two unaffected individuals for which DNA samples were available.

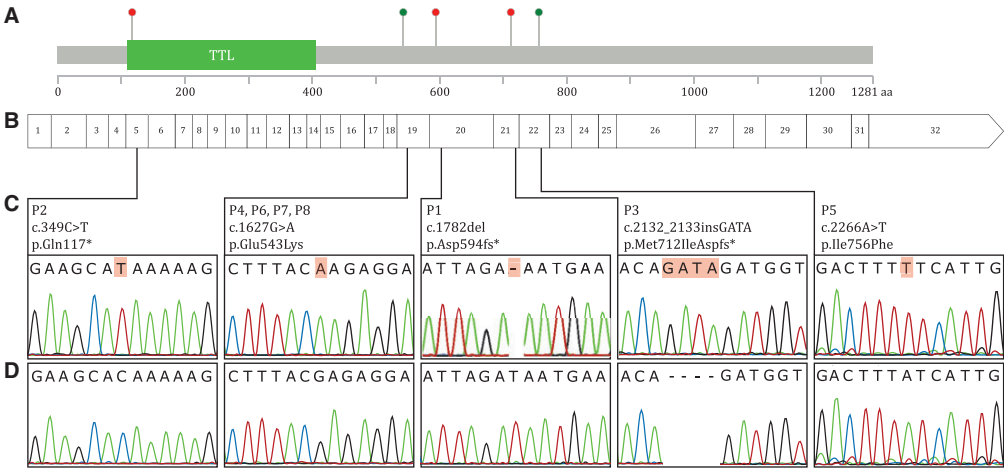


Figure 3. Mutation diagram of the TTLL5 protein (**A**), and corresponding cDNA. Red and green dots indicate truncating and missense mutations, respectively. Exons are numbered and drawn to scale (**B**) with respect to the protein sequence. The TTL domain responsible for polyglutamylation activity is indicated. Chromatograms of the mutations identified (**C**), compared to their relevant wild type sequences (**D**).

Of note, in addition to retinal dystrophy, all patients from this pedigree reported high to very-high myopia (-5 to -22 diopters) (Table 1). None of the patients reported fertility problems, and indeed patient P6 had five children.

TTLL5 RNA isoforms and expression in different tissues

According to publicly-available databases, *TTLL5* produces six protein-coding and alternatively spliced isoforms (transcripts 001, 002, 003, 016, 017, 018 of the Ensembl database GRCh37, release 84), presenting a rather widespread expression throughout different tissues and organs (UniGene). To gain insights into this topic, we investigated the composition and abundance of *TTLL5* transcripts in a panel of cadaveric organs and tissues (Fig. 4).

Our data confirmed that *TTLL5* has an extremely variable expression pattern, both in terms of isoforms and of presence in various tissues. However, qualitative and quantitative assessment of all transcripts revealed that expression of the canonical isoform 001 was overwhelmingly more abundant (more than 40-fold higher than the average of the remainder) and that expression in the retina and testis represented ∼64% of TTLL5 presence across all tissues and organs examined (31 and 33% in testis and retina, respectively; Fig. 4). Interestingly, all mutations identified in our cohort of patients affected isoform 001, and individuals with inactivating mutations showed reduced fertility (Table 1).

TTLL5 mRNA level in the index patient P1

Since mutations leading to premature termination codons trigger nonsense-mediated mRNA decay (NMD) and result in no or in short-lived transcripts (18), we analysed expression of *TTLL5* in skin fibroblasts from patient P1, displaying retinal degeneration and infertility. Quantitative real time PCR (q-PCR) resulted in a dramatically reduced detection of the transcript of interest as compared to a healthy control fibroblast mRNA (∼10%; Fig. 4).

TTLL5 protein in ciliated fibroblasts, retina, and spermatozoa

To better understand the role of TTLL5 with respect to the cellular cilium, we analysed control fibroblasts following serum starvation, a procedure that induces ciliogenesis. Immunofluorescence analysis revealed the clear localization of TTLL5 at both centrioles (Fig. 5, panels A–D).

Subsequently, we performed immunofluorescence analyses in the retina and sperm cells from rat and human, respectively (Fig. 5, panels E-G). In agreement with previous results in mouse and human (5), the anti-TTLL5 antibody decorated the inner segment of photoreceptors in proximity of the basal body and the connecting cilium, suggesting that TTLL5 may in fact play its functions at the base of the photoreceptor primary cilium.

Staining of mature human spermatozoa also indicated for the first time a clear centrosomal localization of TTLL5, with no overlap with the polyglutamylated α- and β-tubulin of the flagellum.

Discussion

Both primary (or immotile) and motile cilia play crucial roles in the normal function of most tissues of the human body. These tiny hair-like organelles participate in a wide range of cellular functions during development, tissue morphogenesis and homeostasis. It is therefore not surprising that mutations in ciliary genes are often associated with a broad range of conditions, classified as ciliopathies, either involving single organs or causing syndromic phenotypes (8). Some examples of diseases affecting primary cilia are polycystic kidney disease, Usher syndrome, retinitis pigmentosa, Bardet-Biedl and Joubert syndromes (19–22). On the other hand, motile cilia defects have been shown to be causative for Kartagener syndrome and allied diseases, collectively grouped under the disease spectrum of primary ciliary dyskinesias.

Cilia and flagella, highly conserved in their core structure, are ancestral organelles composed of more than 650 proteins (10,23). The building units of both ciliary and flagellar

Table 1. Patients with TTLL5 mutations and clinical features

Family	Patient	TTLL5 protein change	Zygosity	Sex	Age at last examination	Visual acuity	Correction	Full field ERG at last examination		Macula	Periphery	Other features
								cone response	rod response			
F1	P1	c.1782del; p.Asp594Glufs*29	hom	M	75	0.05	−3.6	absent	reduced	atrophy	normal	azoospermia
F2	P2	c.349C>T; p.Gln117*	hom	M	46	0.05	−10.00	residual	residual	atrophy	minor changes	reduced motility of sperm; normal anterior eye segment
F3	P3	c.2132_2133insGATA; p.Met712IleAspfs*15	hom	F	58	0.16	−3	reduced	normal		normal	
F4	P4	c.1627G>A; p.Glu543Lys	hom	M	61	0.03	−5	absent	normal	atrophy	normal	
F5	P5	c.2266A>T; p.Ile756Phe	hom	M	38	0.33	−8	absent	normal	pigmentary changes	normal	
F6	P6	c.1627G>A; p.Glu543Lys	hom	M	53	NA	−8.00	NA	NA	atrophy	pigmentary changes	phthisical right eye
	P7	c.1627G>A; p.Glu543Lys	hom	M	38	NA	−5.00	NA	NA	NA	NA	posterior subcapsular cataract
	P8	c.1627G>A; p.Glu543Lys	hom	M	18	NA	−22.00	NA	NA	NA	NA	fairly normal retina

microtubules, α- and β-tubulin, are subject to post-translational modifications, accomplished by enzymes catalyzing different reactions such as the generation of Δ2-tubulin, acetylation (24), tyrosination (25), polyglutamylation (26) and polyglycylation (27,28). Among members of the TTLL superfamily there are glutamylases and glycilases (29–31). TTLL5 initiates the formation of side chains within the C-terminal tail of α- and β-tubulin, with a preference for α-tubulin (32), and current models indicate that the role of polyglutamylation is to provide the necessary conditions for proper MT-MAPs (microtubule and microtubule-associated proteins) interactions. Studies have shown that polyglutamylation exerts differential regulation by selectively recruiting different MAPs: MAP1B, MAP2, tau, and neuronal kinesins have higher affinity for tubulins with 1-3 glutamyl units, whereas MAP1A has higher affinity for longer side chains (25,26,28). Moreover, it has been shown that masking polyglutamylated sites with a specific anti-polyglutamylated tubulin antibody (GT335) affects the amplitude of flagellar beating in sea urchin sperm axonemes, suggesting a key role of polyglutamylated sites for interaction with ciliary dyneins (33). Centriole stability was also shown to be influenced by the degree of polyglutamylation, and GT335 antibody-loaded HeLa cells showed a complete transient disappearance of the centriole pair (13). Finally, members of the TTLL family, Ttll3 and Ttll6, play a role in cilia structure and motility in zebrafish (34). All TTLL proteins have a preference for either α- or β-tubulin and participate to either initiation or elongation of the polyglutamyl side chain. TTLL5, together with TTLL4 and TTLL7, initiates polyglutamylation, while other members function in the elongation of the polyglutamyl side chain or in the initiation or the elongation of polyglycylation.

Specific patterns of modifications on microtubules might be responsible for various functions. In the case of polyglutamylation, the side chains are built within the carboxy-terminal tail of tubulin, where the binding sites of motor and MT-associated proteins (MAPs) are also found. Thus, it is plausible that the interaction of MTs with such proteins might depend on specific patterns of modifications (35). Additional studies highlighted the importance of polyglutamylation for the proper beating of airway cilia (36), as well as for providing a molecular traffic sign required by motor proteins in order to maintain continuous synaptic transmission (37). Major evidence of the implication of polyglutamylation in photoreceptor ciliary function was recently reported (7), showing that *Ttll5*$^{-/-}$ mice developed a similar retinal phenotype to *Rpgr*$^{-/-}$ mice, a known mouse model for retinitis pigmentosa. In addition, *Ttll5*$^{-/-}$ mice display strongly reduced glutamylation of RPGR^{ORF15}, a retina-specific variant of RPGR (38). Altogether, current evidence strongly supports the notion that the presence and length of polyglutamyl side chains, not only on tubulin but also on other substrates, is crucial for proper functioning of both motile and immotile cilia.

In our work we show that mutations in a gene involved in the polyglutamylation of α-tubulin is associated with defects in the retina and spermatozoa. Clinically, these molecular phenotypes translate into cone-first CRD and reduced sperm motility, likely due to the functional impairment of the primary cilium and the flagellum, respectively. Our assumption is supported by immunofluorescence data, demonstrating that TTLL5 localizes at the basal body of the cilia in photoreceptors, as well as at the base of the spermatozoal axoneme and in ciliated skin fibroblasts. Moreover, we reveal that the highest levels of expression of the major *TTLL5* protein-coding isoform is in the retina and testis. It is also very interesting to note that, in terms of fertility and *TTLL5* pathogenic variants, there is an apparent

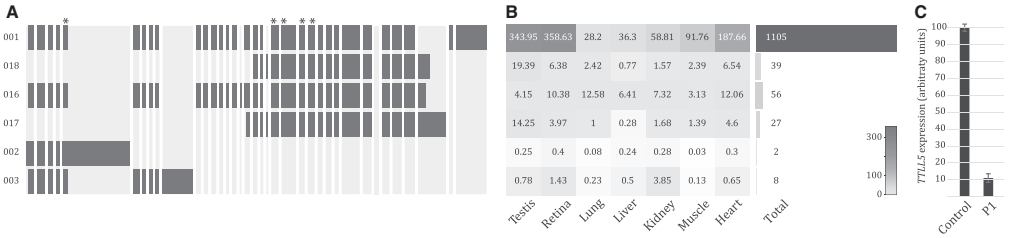


Figure 4. Relative expression of *TTLL5* isoforms and their expression. *TTLL5* has six known alternative transcripts (001, 002, 003, 016, 017, and 018), resulting from the splicing of the exons indicated here in dark grey (**A**). Their expression within seven different human tissues, measured by quantitative real time PCR, is indicated by both numerical values and shades of grey (**B**). Although all isoforms seem to be widely expressed, isoform 001 is the most prominent one, among all tissues considered ('Total' column). Asterisks show the position of the mutations identified in this work. *TTLL5* isoform 001 expression in fibroblasts of the index patient P1 vs. a control, by quantitative PCR (**C**). *TTLL5* mRNA amounts were normalized with respect to the housekeeping gene *HPRT1*.

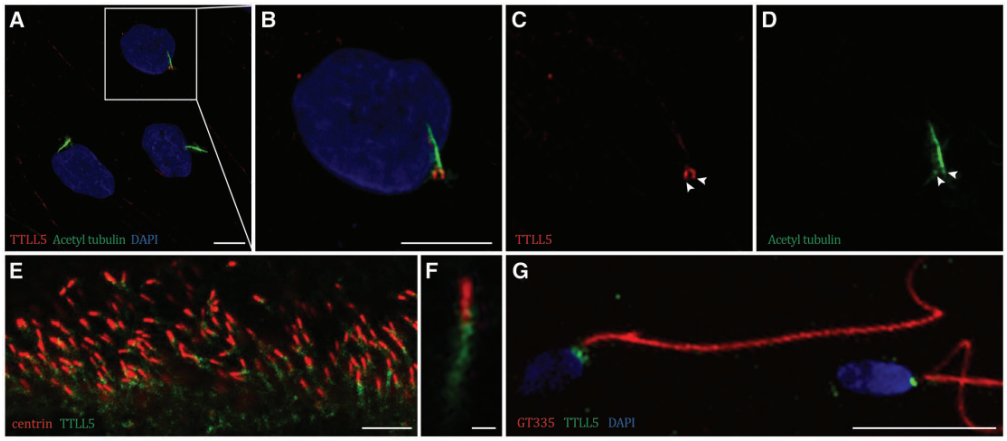


Figure 5. Immunofluorescence staining of ciliated control human skin fibroblasts (**A-D**). TTLL5 co-localizes with acetylated tubulin at the centrioles, as indicated by arrowheads (C,D). Immunofluorescence in rat retina sections (**E**, and magnified cilium in **F**) and in control human spermatozoon (**G**). TTLL5 decorates the basal body in photoreceptors and the centrioles in spermatozoa. Scale bars: A-E, 5 μm; F, 1 μm; G, 20 μm.

genotype/phenotype correlation, which seems to be irrelevant for retinal degeneration. In other words, the phenotype elicited by *TTLL5* pathogenic changes appears to depend on mutation classes. Missense variants are seemingly associated with a non-syndromic phenotype that is limited to the retina, whereas inactivating mutations appear to disrupt the functions of both photoreceptors and spermatozoa, thus defining a novel allele-specific syndrome. Yet, three male patients with truncating *TTLL5* mutations were previously reported having offspring (5), raising the possibility of variable expressivity or reduced penetrance of this class of mutations. In support of the latter hypothesis, *Ttll5* knockout mice display extremely reduced but not completely abolished fertility (6). Alternatively, this genotype/phenotype correlation in our cohort could also be coincidental. The association of *TTLL5* mutations with severe myopia is another intriguing hypothesis that warrants additional investigation in a larger cohort of patients, especially given the complex pattern of inheritance of nearsightedness in humans.

In conclusion, we show that mutations in *TTLL5* are associated with a newly-defined syndrome affecting vision and the male reproductive system. Despite the fact that cilia and flagella have different morphologies and functions, they may share similar physiological mechanisms, and the enzymatic reaction of polyglutamylation performed by TTLL5 may be one of these common elements.

Materials and Methods

Patients and controls

Patient P1 was recruited from the Jules Gonin Ophthalmic Hospital (Lausanne, Switzerland); and patient P2 from the Department of Ophthalmology of Lund University Hospital (Lund, Sweden). Patients P3, P4 and P5 were recruited from the Radboud University Medical Center (Nijmegen, The Netherlands) and the Erasmus University Medical Center (Rotterdam, The Netherlands). Patients P6-8 were sampled by author MMK, an ophthalmologist based at St James's University Hospital (Leeds, England), while on a field trip to Pakistan. DNA of all subjects was extracted from peripheral blood leukocytes. A control sperm sample was provided by a healthy donor. Our research has been conducted in accordance with the tenets of the Declaration of Helsinki and was approved by the Institutional Review Boards of our respective Organizations.

Clinical evaluation

For patients P1-5, ophthalmologic examination included assessment of BCVA, slit-lamp examination, funduscopy, fundus photography, and optical coherence tomography. Full-field ERGs were also recorded, as prescribed by the International Society for Clinical Electrophysiology of Vision (ISCEV).

For patients P6-8, ophthalmologic assessment was limited to slit-lamp examination and fundus inspection using direct and indirect ophthalmoscopy through dilated pupils. Owing to the non-hospital setting, electrodiagnostic and other testing was not available. Visual acuity was recorded together with a history of nyctalopia or photoaversion.

Semen analysis was carried out for patients P1 and P2 by standard procedures of andrology laboratories, and according to WHO guidelines (39). For P1 patient, two Leja chambers (Leja) filled with 6 µl of semen were entirely scanned under phase contrast microscopy to confirm the absence of spermatozoa. Additional clinical features were assessed only for patient P1, due to substantial problems in getting back to the other probands. These tests included a thorough otorhinolaryngologic examination assessing structure and function of the nasopharyngeal mucosa, the ear canal and the hearing (by pure-tone audiometry), as well as a full renal examination, including a complete checkup of renal function, urine and blood analysis, assessment of blood pressure, and ultrasonography. Analyses involving other tissues and organs known to be involved in other ciliopathies were not performed due to a negative clinical history: normal body mass index, no respiratory complains, no metabolic disturbance and no skeletal abnormalities.

Whole-genome and whole-exome sequencing

WGS in the Swiss index patient P1 was performed using 4 µg of DNA. Sequencing was performed by Complete Genomics Inc. (Mountain View, CA, USA), as described previously (40). Genetic variants were identified using v2.0 of the Complete Genomics pipeline (41). WES was performed for proband P6 using 3 µg of DNA. Protein-coding regions were captured using the SureSelect All Exon v4 kit (Agilent) and paired-end sequencing was performed using the Illumina HiSeq 2500 platform. Single nucleotide variants and small insertions or deletions were detected using the Genome Analysis Tool Kit (GATK v2.4.7) software package, using the Best Practice Guidelines identified by the developers (42). The pathogenicity of genetic variants detected through WGS and WES were assessed after functional annotation through ANNOVAR (43).

Homozygosity mapping

Genomic regions with high homozygosity were determined using the free web-based tool HomozygosityMapper (44).

Mutation screening

Primer pairs for *TTLL5* exons and flanking intron boundaries were designed using the CLCbio Genomics Workbench (Qiagen, Supplementary Material, Table S2). PCR amplification was performed in a 20 µl total volume containing 10 ng genomic DNA, 1x GoTaq buffer, 0.1 mM dNTPs, 10 µM of each primer, and 5 U/µl of GoTag polymerase (Promega). PCR products were purified (ExoSAP-IT, USB) and a sequencing reaction was performed in a total volume of 5 µl using 1 µl primer 3.3 µM, 0.5 µl BigDye Terminator v1.1, and 1 µl of the provided Buffer (Applied Biosystems).

Antibodies

Commercial goat polyclonal anti-TTLL5 antibody (Santa Cruz Biotechnology Inc), raised against a peptide mapping near the C-terminus of TTLL5 human origin, was used at a 1:100 dilution. Mouse monoclonal anti-centrin clone 20H5 antibody was purchased from Millipore and used at a dilution of 1:1000. Anti-polyglutamylated tubulin GT335 and monoclonal anti-acetylated tubulin antibodies were provided by Dr C. Janke (Institut Curie, Orsay, France) and purchased from Sigma-Aldrich, respectively. Secondary donkey anti-goat antibodies conjugated with Alexa Fluor 488 were purchased from Invitrogen and secondary goat anti-mouse antibodies (Life Technologies) were conjugated with Alexa Fluor 594 (1:1000).

Fibroblast immortalization and culture

Primary skin fibroblasts were immortalized with exogenous hTERT by the use of pLOX-TERT-iresTK (45) and grown in DMEM(1x) + 1g/L D-glucose L-Glutamine Pyruvate (Gibco), supplemented with 10% FBS, 1% penicillin-streptomycin, and 1% fungizone, adapted from previously published protocols (46).

Immunofluorescence

Immunofluorescence was performed for TTLL5 localization in human control spermatozoa and rat retinal sections. The sperm staining procedure used was adapted from a previously published protocol (47). After washing five times with PBS 1X in a 15 ml Falcon tube and centrifugation steps of 5' at 800g, all at room temperature (RT), the semen pellet was resuspended and fixed with 4% (v/w) PFA in PBS, and incubated for 30 min on ice. The sample was then washed three times with PBS and stored at 4 °C for future use. The immunostaining procedure used 20 µl aliquots of fixed sperm cells, transferred to a 1.5 µl Eppendorf tube. Blocking was done for 30 min at RT in 100 µl PBS containing 3% (w/v) BSA (PBSA). Primary antibody incubation was performed with specified dilutions in 100 µl PBSA, overnight at 4 °C. Samples were washed three times with PBS containing 0.1% (v/v) TX100 (PBST). Secondary antibody incubation for 30 min was carried out at RT in 100 µl PBSAT [PBST containing 3% (w/v) BSA and 0.1% (v/v) TX100]. After three washes with PBST and two washes with PBS the final pellet was resuspended in PBS and 5 µl were placed on a slide. 5 µl DAPI vectashield were added to the sample, which was then coverslipped and fixed with nail polish.

Unfixed Sprague-Dawley rat eyes and C57BL/6J mouse eyes were isolated and soaked for 3 h in PBS containing 30% sucrose. Eyes were embedded in Yazulla medium (30% egg albumen and 3% gelatin in water) and cryosectioned (12 µm) onto Superfrost Plus slides (Thermo Scientific). Sections were washed three times with PBS and a stepwise procedure was followed similar to that described for sperm immunostaining, the only difference being that this was carried out on the slide and with lower volumes.

Quantitative real time PCR

Primer pairs used for q-PCR are listed in Supplementary Material, Table S3. The q-PCR product was visualized on 1%

agarose gel to verify the primer's specificity. A standard curve using a control cDNA template prepared from human normal tissues total RNA (BioChain) was used to test the efficiency of each primer pair. *HPRT1* was used as a normalization control, as described (48). Amplification was performed using the SYBR Green PCR Master Mix (Applied Biosystems).

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We would like to thank the patients and their families who participated in this study.

Conflict of Interest statement. None declared.

Funding

This work was supported by grants from the Swiss National Science Foundation to CR [grant numbers 138346, 156260], the Foundation Fighting Blindness USA to FPMC [grant number C-GE-0811-0545-RAD01], and RP Fighting Blindness and Fight For Sight [RP Genome Project GR586]. MEE was funded by an Egyptian Government Scholarship. Family F6 was analysed by the UK Inherited Retinal Disease Consortium.

References

1. Hamel, C.P. (2007) Cone rod dystrophies. *Orphanet J. Rare Dis.*, **2**, 7.
2. Berson, E.L., Gouras, P. and Gunkel, R.D. (1968) Progressive cone-rod degeneration. *Arch. Ophthalmol.*, **80**, 68–76.
3. Berson, E.L., Gouras, P. and Gunkel, R.D. (1968) Progressive cone degeneration, dominantly inherited. *Arch. Ophthalmol.*, **80**, 77–83.
4. Roosing, S., Thiadens, A.A., Hoyng, C.B., Klaver, C.C., den Hollander, A.I. and Cremers, F.P. (2014) Causes and consequences of inherited cone disorders. *Prog. Retin. Eye Res.*, **42**, 1–26.
5. Sergouniotis, P.I., Chakarova, C., Murphy, C., Becker, M., Lenassi, E., Arno, G., Lek, M., MacArthur, D.G., Consortium, U.C.E., Bhattacharya, S.S., *et al.* (2014) Biallelic variants in TTLL5, encoding a tubulin glutamylase, cause retinal dystrophy. *Am. J. Hum. Genet.*, **94**, 760–769.
6. Lee, G.S., He, Y., Dougherty, E.J., Jimenez-Movilla, M., Avella, M., Grullon, S., Sharlin, D.S., Guo, C., Blackford, J.A., Jr., Awasthi, S., *et al.* (2013) Disruption of Ttll5/stamp gene (tubulin tyrosine ligase-like protein 5/SRC-1 and TIF2-associated modulatory protein gene) in male mice causes sperm malformation and infertility. *J. Biol. Chem.*, **288**, 15167–15180.
7. Sun, X., Park, J.H., Gumerson, J., Wu, Z., Swaroop, A., Qian, H., Roll-Mecak, A. and Li, T. (2016) Loss of RPGR glutamylation underlies the pathogenic mechanism of retinal dystrophy caused by TTLL5 mutations. *Proc. Natl Acad. Sci. U. S. A*, **113**, E2925–E2934.
8. Hildebrandt, F., Benzing, T. and Katsanis, N. (2011) Ciliopathies. *N. Engl. J. Med.*, **364**, 1533–1543.
9. Waters, A.M. and Beales, P.L. (2011) Ciliopathies: an expanding disease spectrum. *Pediatr. Nephrol.*, **26**, 1039–1056.
10. Fliegauf, M., Benzing, T. and Omran, H. (2007) When cilia go bad: cilia defects and ciliopathies. *Nat. Rev. Mol. Cell Biol.*, **8**, 880–893.
11. Hildebrandt, F., Attanasio, M. and Otto, E. (2009) Nephronophthisis: disease mechanisms of a ciliopathy. *J. Am. Soc. Nephrol.*, **20**, 23–35.
12. Schorderet, D.F., Bernasconi, M., Tiab, L., Favez, T. and Escher, P. (2014) IROme, a new high-throughput molecular tool for the diagnosis of inherited retinal dystrophies-a price comparison with Sanger sequencing. *Adv. Exp. Med. Biol.*, **801**, 171–176.
13. Bobinnec, Y., Khodjakov, A., Mir, L.M., Rieder, C.L., Edde, B. and Bornens, M. (1998) Centriole disassembly in vivo and its effect on centrosome structure and function in vertebrate cells. *J. Cell Biol.*, **143**, 1575–1589.
14. Firmann, M., Mayor, V., Vidal, P.M., Bochud, M., Pecoud, A., Hayoz, D., Paccaud, F., Preisig, M., Song, K.S., Yuan, X., *et al.* (2008) The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.*, **8**, 6.
15. Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., O'Donnell-Luria, A., Ware, J., Hill, A., Cummings, B., *et al.* (2015) Analysis of protein-coding genetic variation in 60,706 humans. *Nature.*, **536**, 285–291.
16. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
17. KDIGO. (2013) Clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl.*, **3**, 150.
18. Hentze, M.W. and Kulozik, A.E. (1999) A perfect message: RNA surveillance and nonsense-mediated decay. *Cell*, **96**, 307–310.
19. Badano, J.L., Mitsuma, N., Beales, P.L. and Katsanis, N. (2006) The ciliopathies: an emerging class of human genetic disorders. *Annu. Rev. Genomics Hum. Genet.*, **7**, 125–148.
20. Zaghloul, N.A. and Katsanis, N. (2009) Mechanistic insights into Bardet-Biedl syndrome, a model ciliopathy. *J. Clin. Invest.*, **119**, 428–437.
21. Forsythe, E. and Beales, P.L. (2013) Bardet-Biedl syndrome. *Eur. J. Hum. Genet.*, **21**, 8–13.
22. Falk, N., Losl, M., Schroder, N. and Giessl, A. (2015) Specialized Cilia in Mammalian Sensory Systems. *Cells*, **4**, 500–519.
23. Carvalho-Santos, Z., Azimzadeh, J., Pereira-Leal, J.B. and Bettencourt-Dias, M. (2011) Evolution: Tracing the origins of centrioles, cilia, and flagella. *J. Cell Biol.*, **194**, 165–175.
24. Piperno, G. and Fuller, M.T. (1985) Monoclonal antibodies specific for an acetylated form of alpha-tubulin recognize the antigen in cilia and flagella from a variety of organisms. *J. Cell Biol.*, **101**, 2085–2094.
25. Gundersen, G.G. and Bulinski, J.C. (1986) Distribution of tyrosinated and nontyrosinated alpha-tubulin during mitosis. *J. Cell Biol.*, **102**, 1118–1126.
26. Edde, B., Rossier, J., Le Caer, J.P., Desbruyeres, E., Gros, F. and Denoulet, P. (1990) Posttranslational glutamylation of alpha-tubulin. *Science*, **247**, 83–85.
27. Westermann, S. and Weber, K. (2003) Post-translational modifications regulate microtubule function. *Nat. Rev. Mol. Cell Biol.*, **4**, 938–947.
28. Janke, C. and Kneussel, M. (2010) Tubulin post-translational modifications: encoding functions on the neuronal microtubule cytoskeleton. *Trends Neurosci.*, **33**, 362–372.

29. Boucher, D., Larcher, J.C., Gros, F. and Denoulet, P. (1994) Polyglutamylation of tubulin as a progressive regulator of in vitro interactions between the microtubule-associated protein Tau and tubulin. *Biochemistry*, **33**, 12471–12477.
30. Larcher, J.C., Boucher, D., Lazereg, S., Gros, F. and Denoulet, P. (1996) Interaction of kinesin motor domains with alpha- and beta-tubulin subunits at a tau-independent binding site. Regulation by polyglutamylation. *J. Biol. Chem.*, **271**, 22117–22124.
31. Bonnet, C., Boucher, D., Lazereg, S., Pedrotti, B., Islam, K., Denoulet, P. and Larcher, J.C. (2001) Differential binding regulation of microtubule-associated proteins MAP1A, MAP1B, and MAP2 by tubulin polyglutamylation. *J. Biol. Chem.*, **276**, 12839–12848.
32. Janke, C., Rogowski, K., Wloga, D., Regnard, C., Kajava, A.V., Strub, J.M., Temurak, N., van Dijk, J., Boucher, D., van Dorsselaer, A., *et al.* (2005) Tubulin polyglutamylase enzymes are members of the TTL domain protein family. *Science*, **308**, 1758–1762.
33. Gagnon, C., White, D., Cosson, J., Huitorel, P., Edde, B., Desbruyeres, E., Paturle-Lafanechere, L., Multigner, L., Job, D. and Cibert, C. (1996) The polyglutamylated lateral chain of alpha-tubulin plays a key role in flagellar motility. *J. Cell Sci.*, **109**, 1545–1553.,
34. Pathak, N., Austin, C.A. and Drummond, I.A. (2011) Tubulin tyrosine ligase-like genes ttll3 and ttll6 maintain zebrafish cilia structure and motility. *J. Biol. Chem.*, **286**, 11685–11695.
35. Janke, C., Rogowski, K. and van Dijk, J. (2008) Polyglutamylation: a fine-regulator of protein function? 'Protein Modifications: beyond the usual suspects' review series. *EMBO Rep.*, **9**, 636–641.
36. Ikegami, K., Sato, S., Nakamura, K., Ostrowski, L.E. and Setou, M. (2010) Tubulin polyglutamylation is essential for airway ciliary function through the regulation of beating asymmetry. *Proc. Natl Acad. Sci. U. S. A*, **107**, 10490–10495.
37. Ikegami, K., Heier, R.L., Taruishi, M., Takagi, H., Mukai, M., Shimma, S., Taira, S., Hatanaka, K., Morone, N., Yao, I., *et al.* (2007) Loss of alpha-tubulin polyglutamylation in ROSA22 mice is associated with abnormal targeting of KIF1A and modulated synaptic function. *Proc. Natl Acad. Sci. U. S. A*, **104**, 3213–3218.
38. Vervoort, R., Lennon, A., Bird, A.C., Tulloch, B., Axton, R., Miano, M.G., Meindl, A., Meitinger, T., Ciccodicola, A. and Wright, A.F. (2000) Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nat. Genet.*, **25**, 462–466.
39. WHO. (2010) WHO laboratory manual for the examination and processing of human semen.
40. Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., *et al.* (2010) Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*, **327**, 78–81.
41. Carnevali, P., Baccash, J., Halpern, A.L., Nazarenko, I., Nilsen, G.B., Pant, K.P., Ebert, J.C., Brownley, A., Morenzoni, M., Karpinchyk, V., *et al.* (2012) Computational techniques for human genome resequencing using mated gapped reads. *J. Comput. Biol.*, **19**, 279–292.
42. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.*, **43**, 491–498.
43. Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.
44. Seelow, D., Schuelke, M., Hildebrandt, F. and Nurnberg, P. (2009) HomozygosityMapper–an interactive approach to homozygosity mapping. *Nucleic Acids Res.*, **37**, W593–W599.
45. Salmon, P., Oberholzer, J., Occhiodoro, T., Morel, P., Lou, J. and Trono, D. (2000) Reversible immortalization of human primary cells by lentivector-mediated transfer of specific genes. *Mol. Ther.*, **2**, 404–414.
46. van Karnebeek, C.D., Bonafe, L., Wen, X.Y., Tarailo-Graovac, M., Balzano, S., Royer-Bertrand, B., Ashikov, A., Garavelli, L., Mammi, I., Turolla, L., *et al.* (2016) NANS-mediated synthesis of sialic acid is required for brain and skeletal development. *Nat. Genet.*, **48**, 777–784.
47. Nishimura, H., Gupta, S., Myles, D.G. and Primakoff, P. (2011) Characterization of mouse sperm TMEM190, a small transmembrane protein with the trefoil domain: evidence for co-localization with IZUMO1 and complex formation with other sperm proteins. *Reproduction*, **141**, 437–451.
48. Hoggart, C.J., Venturini, G., Mangino, M., Gomez, F., Ascari, G., Zhao, J.H., Teumer, A., Winkler, T.W., Tsernikova, N., Luan, J., *et al.* (2014) Novel approach identifies SNPs in SLC2A10 and KCNK9 with evidence for parent-of-origin effect on body mass index. *PLoS Genet.*, **10**, e1004508.