

# EX-HOM (EXome HOMozygosity): A Proof of Principle

Tommaso Pippucci<sup>a</sup> Matteo Benelli<sup>c, d, e</sup> Alberto Magi<sup>d, e, f</sup> Pier Luigi Martelli<sup>b</sup>  
Pamela Magini<sup>a</sup> Francesca Torricelli<sup>d</sup> Rita Casadio<sup>b</sup> Marco Seri<sup>a</sup>  
Giovanni Romeo<sup>a</sup>

<sup>a</sup>Medical Genetics Unit, Department of Gynecological, Obstetric and Pediatric Sciences, and <sup>b</sup>Laboratory of Biocomputing, CIRB/Department of Biology, University of Bologna, Bologna, <sup>c</sup>INFN, National Institute for Nuclear Physics, Section of Florence, <sup>d</sup>Diagnostic Genetic Unit, Careggi Hospital, AOUC, <sup>e</sup>Center for the Study of Complex Dynamics, and <sup>f</sup>Department of Medical and Surgical Critical Care, University of Florence, Florence, Italy

## Key Words

Exome sequencing · Homozygosity mapping ·  
Consanguinity · Autosomal recessive disorders

## Abstract

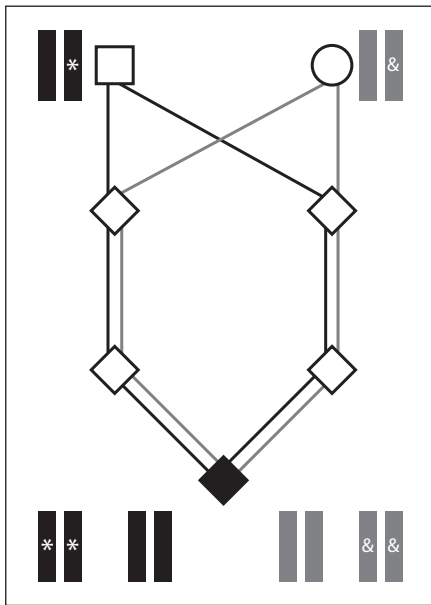
**Objective:** We provide the proof of principle that exome sequencing of only two affected siblings born to first-cousin parents is capable of directly identifying a single candidate gene for an autosomal recessive disorder. This strategy, which we call EX-HOM (EXome HOMozygosity), combines in a single step the capacity of exome sequencing to identify all the coding variants present in a genome with the property of homozygosity mapping to limit the search for candidate genes to specific chromosomal regions. **Methods:** We sequenced the exomes of two siblings born to first-cousin parents affected with dysmyelinating leukodystrophy and spastic paraparesis caused by a mutation in *FA2H*. We used exome sequencing data to identify homozygous regions shared by the two affected siblings (EX-HOM regions), compared them with the regions of maximum LOD score obtained with SNP genotyping, and selected the candidate variants within. **Results:** We identified regions of shared homozygosity (>1 Mb) accounting for about 290 Mb, containing only 3 candidate variants. Among these, the *FA2H* muta-

tion remained the only plausible one. **Conclusion:** In single consanguineous pedigrees with a few affected sibs, EX-HOM can be a one-step approach to identify the candidate genetic defect, bypassing obstacles such as genetic heterogeneity and the need for large pedigrees.

Copyright © 2011 S. Karger AG, Basel

## Introduction

In Mediterranean populations alone, homozygosity mapping using large consanguineous pedigrees led to the identification of almost 100 genes responsible for rare autosomal recessive disorders [1]. Today, advancements in sequencing technology make it feasible to map and identify the causative mutation in much smaller pedigrees or even in a single patient. Recent reports show how exome sequencing of a few affected descendants born to consanguineous parents can be in principle as powerful as the traditional strategy based on homozygosity mapping and mutation screening of positional candidate genes. SNP genotyping of two children of first cousins from the same pedigree showing pachygyria with mental retardation and seizures (MIM #600176), followed by exome sequencing in just one of them, discovered causative mutations in



**Fig. 1.** Variant transmission in autozygosity. Rare, possibly deleterious variants are assumed to be *at most* one per gene in the common great-grandparents (white square and circle at the top of the figure). Only half of the four great-grandparental gene copies (bars at the bottom of the figure) will carry a candidate rare pathogenic variant (\* and &) which will be transmitted through generations (white rhombs) along two of the four patterns of autozygosity (lines) to the affected offspring (black rhomb). Since the autozygous genome is expected to be 1/16, the autozygous genes harboring candidate rare, possibly deleterious, variants are expected to be 1/2 of the genes in the autozygous genome (1/32 of the whole genome).

*WDR62* [2]. The targeted sequencing of regions mapped by ancestral autozygosity in a family with diaphanospondylodysostosis led to the identification of a causative mutation in *BMPEP* in a single patient [3]. Kalay et al. [4] performed exome sequencing in a single individual born to first-cousin parents and, independently, linkage analysis in several small consanguineous families all affected with Seckel syndrome. Linkage analysis pointed to a founder homozygous haplotype on chromosome 15q21.1–q21.2 in which *CEP152* was found to be the gene carrying the same homozygous mutation in all the families [4]. In the single patient, a mapping approach based on exome sequencing data was used to select homozygous candidate regions and allowed the identification of 7 genes harboring candidate variants, among which *CEP152* was the only one embedded in a long homozygous stretch. Similarly, Becker et al. [5] used the homozygosity information obtained from exome sequencing to prioritize candidate

genes for autosomal recessive osteogenesis imperfecta in an affected subject born to second-cousin parents, identifying 7 homozygous regions containing 3 candidate variants.

All these studies support the idea that exome sequencing, even without the need for preliminary array SNP genotyping, can efficiently perform homozygosity mapping and restrict the pool of candidate genes to a very few.

Exome sequencing identifies thousands of single nucleotide variants (SNVs) per individual, which can be used to directly infer mapping information and at the same time select the candidate rare variants in regions of homozygosity. Candidate variants are all those variants not annotated in public databases such as dbSNP and the 1000 Genomes project, with possible deleterious effects on proteins. In published exome data, the frequency of genes harboring novel heterozygous rare variants is usually reported to be <500 per individual [6–11]. If downstream filtering is applied, such as comparisons against control exome databases, prioritization according to functional computational predictions or conservation during evolution, this number can drastically decrease to an average of about 250 [6, 8, 9, 11].

The rarer the autosomal recessive disorder, the higher the probability that affected children of consanguineous parents carry a mutation identical by descent in an autozygous genomic region. In an affected sibship of two, born to first-cousin parents, the causative mutation is likely to reside in one of the autozygous genes that they co-inherited from their great-grandparents. Affected children of first-cousin parents are on average autozygous in 1/16 of their genes. Of these, half (1/32) are expected to contain a possibly pathogenic variant in the autozygous state because each of the two unaffected great-grandparents was likely to carry one rare variant per gene at the most (fig. 1). The two affected siblings share 1/4 of this proportion ( $1/32 \times 1/4 = 1/128$ ). Assuming that a pool of 500 candidate variants were present in the great-grandparents, the two affected siblings would on average share 3–4 ( $500/128$ ) candidate variants. Thus, an exome sequencing strategy based on selecting families with two children born to first-cousin parents (which represents the most common consanguineous marriage) can provide the theoretical power to drastically restrict candidate variants. Furthermore, it can feature additional advantages compared to sequencing a single offspring. Sequencing two individuals, instead of one only, can improve the capacity to call variants in scarcely captured regions adding up the reads from the two independent experiments. Also, it is easier to exclude that genetic mechanisms other than autozygos-

ity might be causative for the disorder. For instance, it is virtually impossible that de novo mutations have occurred to cause the disease in both siblings.

The capacity to identify candidate autosomal recessive genes through a next-generation sequencing-based approach using a single consanguineous family with few affected individuals is a promising perspective in medical genetics. Extremely rare conditions often appear only in one family, and have not yet been studied or even reported. Furthermore, the gap in identifying genes responsible for rare Mendelian diseases becomes apparent considering that the molecular basis of at least 3,800 known or suspected Mendelian diseases is unknown (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>) [1].

Here, we provide a proof of principle of the validity of this strategy, which we call EX-HOM (Exome HOMozygosity) by sequencing the exomes of two brothers born to first-cousin parents affected with dysmyelinating leukodystrophy and spastic paraparesis (MIM #612443), who had been previously found to carry a homozygous mutation in *FA2H* [12].

## Materials and Methods

### Patients

A specific consent form was filled by the patients' parents, and the study was performed according to the Declaration of Helsinki principles. The patients' family was from the Southern Italian region of Calabria and consisted also of two unaffected siblings, a brother and a sister. We previously found that the two affected brothers carried a *FA2H* homozygous c.270+3A>T mutation (NM\_024306.2) leading to an aberrant mRNA product. Both patients showed childhood onset of progressive spastic paraparesis, mild pyramidal and cerebellar upper limb signs, severe cognitive impairment, white-matter disease, and cerebellar, brain stem, and spinal cord atrophy [12].

### Exome Capture and Sequencing

Genomic DNA samples were processed by in-solution hybridization using the human SureSelect All Exon Kit (Agilent Technologies, Inc., Santa Clara, Calif., USA) coupled with the Illumina DNA-seq pair-end sample prep kit (Illumina, Inc., Calif., USA).

Briefly, 5 µg of each genomic DNA was firstly sheared into fragments of 150–350 bp by nebulization. After end-repair and A base tailing, the Illumina pair-end adapters were ligated. After ligation, purification from not ligated adapters and size selection in the range of 150–350 bp, samples were amplified for 14 cycles to selectively enrich those DNA fragments having adapter molecules on both ends. The libraries were then hybridized and the targets recovered following the standard SureSelect Target Enrichment System protocol according to the manufacturer's recommendations. Libraries were sequenced as 36 bp pair-end reads following the standard Illumina Genome Analyzer Iix (GAIIx, Illumina, Inc.) workflow.

**Table 1.** SNV calls in the two siblings

SNVs	Sib 1	Sib 2
Validated	20,967	22,651
Transition to transversion ratio	2.91	2.89
In dbSNP	19,727 (94%)	21,237 (94%)
Heterozygous in dbSNP	10,372	11,159
Homozygous in dbSNP	9,355	10,078
Not in dbSNP	1,240 (6%)	1,414 (6%)
Heterozygous not in dbSNP	1,086	1,234
Homozygous not in dbSNP130	154	180

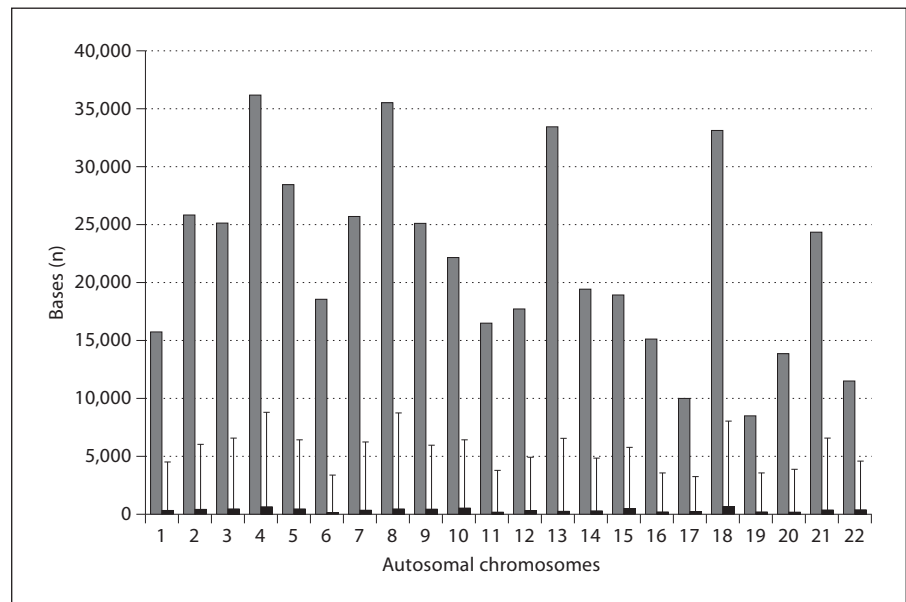
### Analysis of Exome Sequencing Data

The 36-bp paired end reads were aligned to the human reference genome hg18 with BWA version 0.5.8 [13]. Alignment and coverage statistics were generated with GATK version 1.0.4013, Depth of Coverage version 3.0 [14] (in a single and multiple sample approach), SAMTOOLS version 0.1.8 [15], and custom R scripts. Detection of SNVs and insertions/deletions (indels) was performed with the GATK whole-exome version 1 pipeline on the target defined by the Agilent SureSelect design. Standard parameters for validation of SNVs and indels were selected. Detected variants were annotated for novelty compared with both dbSNP build 130 (dbSNP130) and the 1000 Genomes project, and for position relative to RefSeq version 33. The GATK single-sample approach was used to generate general calls (table 1). The GATK multiple-sample approach, which combines information from different individuals to call variation at a site, was used instead to improve calling of novel homozygous variants. We can assume that in the EX-HOM regions, siblings born to consanguineous parents are genetically identical. Therefore, with a multiple sample approach we had a better chance of not missing true variants in scarcely covered sites, adding up genotype information from the two affected siblings' samples.

### Homozygosity Mapping with Exome SNVs

All the autosomal dbSNP130 sites and all the novel SNVs falling in the exome target regions served to create a genetic map consisting of 135,035 markers. Criteria for selection were that the site was at least 10-fold covered for heterozygous calls and 5-fold covered for homozygous calls. Statistics on the distribution of map markers along the genome were obtained with custom Perl and R scripts. To select individual and shared homozygous segments the runs of homozygosity (ROH) utility of PLINK version 1.06 [16] was used with parameters adapted to the map in use. We used a window of 50 SNVs (or 500 kb) allowing for 1 heterozygous and 5 missing sites per individual. We selected individual ROHs as stretches of 200 homozygous SNVs (or 2.5 Mb long homozygous regions) with at least 1 SNV every 50 kb and a maximal allowed gap between SNVs of 5 Mb. Identity-by-descent chromosomal segments of recent origin are expected to be quite long. Only homozygous segments longer than 5 Mb have been estimated to originate from a common paternal and maternal ancestor in the preceding six generations [17], and the length of the homozygous tract across the disease-linked locus in the affected offspring of first-cousin parents was found to be between 5 and 70 cM [18].

**Fig. 2.** Mean and median distances between EX-HOM map markers per chromosome. Grey histograms represent the mean, black histograms the median. Bars are 25th and 75th percentiles.



We considered the overlapping, allelically matching segments across the two siblings' individual ROHs at a threshold length of 1, 2.5, 5, and 10 Mb as EX-HOM candidate regions.

#### Comparison with Regions Detected with SNP Genotyping

We compared these EX-HOM regions with those detected with the low-density (6,090 SNPs) HumanLinkage V Illumina Panel Set (Illumina, Inc.), including in the analysis the parents and the two affected siblings. We performed parametric linkage analysis with Merlin [19], setting the disease-gene frequency at 0.0001 and penetrance for the aa, aA, and AA genotypes at 0.0, 0.0, and 1.0, respectively. We considered all regions reaching the maximum theoretical LOD score for this family (which is 1.8) as candidate regions.

#### Variant Filtering

All validated novel homozygous SNVs/indels detected in the EX-HOM regions with the multiple sample approach were evaluated for their probability to be disease related. We filtered variants retaining only small insertions or deletions, stop gain or stop loss changes, or SNVs to which a significant score by at least 2 out of 6 predictors of pathogenicity for amino acidic substitutions or splice site alterations was assigned. The 5 different widely used state-of-the-art predictors for SNVs we interrogated, of which the performances were recently compared [20], were MutPred [21], Polyphen [22], Polyphen2 [23], SIFT [24] and SNP&GO [25], while the one for splice-site alterations was NetGene2 [26].

Consultation of GENATLAS [27] and Human Protein Atlas [28] databases provided information on the expression and localization of the genes and the corresponding proteins.

#### Sanger Sequencing of Candidate Variants

Rare candidate SNVs and indels were validated by Sanger sequencing of DNA from the affected brothers and their parents. The primer oligonucleotides are listed in the online supplement

tary material ([www.karger.com/doi/10.1159/000330164](http://www.karger.com/doi/10.1159/000330164)). If Sanger sequencing confirmed the presence of the variant and its correct segregation in the family, 190 Italian control chromosomes were checked to verify if it was present in the Italian population.

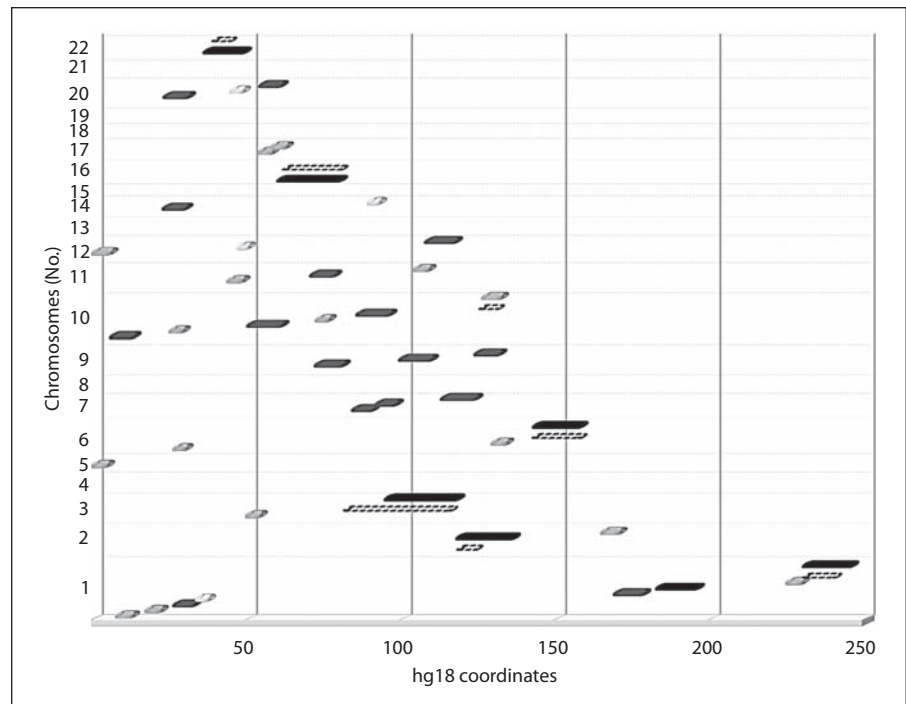
## Results

After removing duplicates (which represented 13% of the bases mapped to the reference genome in each of the two brothers), the target bases represented 48 and 55% of all those mapped to the reference genome, providing a mean coverage of 26- and 30-fold, respectively. The 86 and 91% (94% with the multiple sample approach) of all targeted bases with a base quality >30 were read more than 5 times (online suppl. material) and were therefore used for variant calling. Calls generated with the multiple sample approach were used to improve genotype calling of novel autosomal homozygous variants located in the EX-HOM regions.

We calculated the map density and distribution in terms of mean and median distance (fig. 2) across the 135,035 autosomal SNVs selected as map markers (on average 6,138 per chromosome). The means and medians were markedly different for all autosomes, as expected for an exomic target. The majority of markers laid close to each other (mean of the medians = 404 bp), but regions of the genome in which markers were more sparse (mean of the means = 21.9 kbp) were also present.



**Fig. 3.** Distribution and size of EX-HOM regions and their overlaps with maximum LOD score regions. Cylinders represent EX-HOM regions of different length (white: >1 Mb; light grey: >2.5 Mb; dark grey: >5 Mb; black: >10 Mb) displayed according to their physical position on the chromosome. Dotted line-bordered cylinders represent regions of maximum LOD score.



To assess whether long regions were not covered by markers due to low coverage rather than being gene-poor regions, we calculated the percentage of target sequences in those portions of the genome in which the distance between markers was >1 Mb. We found that the proportion of targets in these regions was 1% at the most, which indicates a minimal risk of losing mapping information.

We compared these EX-HOM mapping regions with those detected as LOD score peaks obtained by linkage analysis of the nuclear family (two parents and the two affected siblings only). Since our EX-HOM map was built with markers contained in the exome target regions, markers were densely but not uniformly distributed along the genome. We therefore did not expect a perfect match between regions identified by our linkage (which uses instead a low-density but uniform map) and EX-HOM analyses. We found 43 EX-HOM regions accounting for about 290 Mb at a threshold level of 1 Mb. The number of EX-HOM regions did not drastically change when the threshold was set at 2.5 Mb (39), but it rapidly decreased when the threshold was set at 5 Mb (23) and 10 Mb (7). There were 7 maximum LOD score regions. EX-HOM identified all LOD score peak regions with substantial overlap (87% on average), showing that EX-HOM performed with Plink can correctly identify disease-related long homozygous regions in consanguineous families

(fig. 3). Within all the regions of shared homozygosity, the siblings showed a total of 33 homozygous variants not included in dbSNP130 (online suppl. material). Only 3 final variants could be considered as candidates according to our criteria (table 2): a non-synonymous p.L731F in *EFCAB6* (NM\_198856) on chromosome 22, predicted to be probably damaging by Polyphen and Polyphen2; the intronic c.270+3A>T variant in *FA2H* (NM\_024306.2), predicted to alter a splice site by NetGene2.4; and a c.1246\_1246delG causing frame shift of *KIAA1407* (NM\_020817) on chromosome 3. Notably, all these 3 variants were in EX-HOM regions longer than 5 Mb overlapping maximum LOD score regions (fig. 3). The *EFCAB6* variant was found in the parents in the heterozygous state and in none of 190 control chromosomes, while the *KIAA1407* deletion was excluded after Sanger sequencing, since it appeared to be homozygous in the mother although not found in 190 control chromosomes.

Since dysmyelinating processes leading to leukodystrophy occur in the central nervous system (CNS), a further criterion to restrict potential candidates was to verify their level of expression in this tissue. *EFCAB6* has not been found in the brain, while it has been shown to be specifically expressed in the testis [29]. Conversely, *FA2H* is highly expressed in CNS, and functional information supports its role in myelinogenesis [30, 31]. For all these

**Table 2.** Candidate variants in the shared homozygous intervals

Chr	Gene	cDNA/protein level change	SIFT	SNP&GO	Polyphen	Polyphen2 (HumDiv)	Polyphen2 (HumVar)	MutPred	NetGene	Gene expression (GENAT-LAS)	Protein expression (Human Protein Atlas)
6	AKAP12	NM_005100 p.E367K	0.1	neutral, 5	benign	benign (0.020)	benign (0.030)	0.3	-	ear (highly), testis (highly), adipose tissue, endothelium	testis, kidney, ovary
16	DPEP2	NM_022355 p.D142N	0.31	neutral, 7	benign	benign (0.008)	benign (0.022)	0.37	-	uterus	most normal tissues, neuronal cells negative
22	EFCAB6	NM_022785 p.L883F	0.07	neutral, 6	benign	probably damaging (0.996)	probably damaging (0.956)	-	-	testis	exocrine pancreas, glandular cells of the gastrointestinal tract, surface epithelium of the bronchi and urinary bladder, paracortices of lymph nodes and tonsils
22	EFCAB6	NM_022785 p.R193Q	0.51	neutral, 9	benign	benign (0.000)	benign (0.000)	0.5	-	see above	see above
22	EFCAB6	NM_022785 p.I63T	0.02	neutral, 6	possibly damaging	possibly damaging (0.238)	benign (0.186)	0.61	-	see above	see above
22	PNPLA5	NM_001177675 p.R217G	0.33	neutral, 8	benign	possibly damaging (0.237)	benign (0.126)	0.39	-	brain, skin	-
16	FA2H	NM_024306.2 c.270+3A>T	-	-	-	-	-	-	0.95	meninges (highly), white matter (highly), uterus (highly), skin (highly)	-
3	KIAA1407	NM_020817 c.1246_1246delG	-	-	-	-	-	-	-	-	-

Chr = Chromosome; - = analysis did not provide any significant result.

reasons, *FA2H* could be considered the top candidate among all the genes harboring a rare homozygous mutation in the two affected brothers.

## Discussion

Consanguinity has always represented a potentially rich resource for the mapping of rare autosomal recessive diseases. Rare deleterious recessive mutations which usually go unnoticed in the heterozygous state can be ascer-

tained via homozygosity in highly inbred populations [1]. Homozygosity mapping exploits the high informativeness of meioses in consanguineous pedigrees but needs large families to narrow the linkage intervals. Next-generation sequencing-based approaches (either whole-exome or targeted resequencing) have performed more rapidly and efficiently than traditional Sanger sequencing of candidate genes following homozygosity mapping in single large families or several small, genetically homogeneous pedigrees [32, 33]. Moreover, the elevated power to drastically restrict candidate genes in autosomal

recessive disorders has been proven by the identification of causative mutations starting from the exome sequencing of only one to three affected siblings from consanguineous or even non-consanguineous families [5, 7, 34, 35]. However, especially when phenotypes are extremely rare and genetic heterogeneity is apparent, it becomes challenging to collect additional phenotypically matching cases bearing a mutation in the same gene. In their homozygosity analysis using exome sequencing variants from a single affected child of first-cousin parents, Kalay et al. [4] state that only a single candidate variant was associated with a region of extended homozygosity (35 Mb). This might lead to the conclusion that the use of only one affected descendant is sufficient to restrict candidate mutations to the causative one. However, it is unclear whether their method to measure homozygosity (percentage of homozygous calls within a sliding window of 100 variants) underestimates less extended stretches of homozygosity (i.e. 5 Mb) which can also contain the disease-causing mutation [18]. Becker et al. [5] used all dbSNP and variant sites, as we did in the present work, obtaining a much denser map which is supposed to be more sensitive to shorter stretches of homozygous variants. They found 3 candidate variants associated to long homozygous regions in a single offspring of second-cousin parents, which is in agreement with our findings. This is likely to be a coincidence, but it is intriguing since, on average, an offspring of second-cousin parents receives in autozygosity the same proportion of the genome which is expected to be shared by two siblings of first-cousin parents (1/64).

The present work proposes the EX-HOM as an approach to be applied when only a single, small consanguineous family is available to map and identify the genetic defect underlying a disease phenotype. Here, we provide the proof of principle that this method would have been able to identify the mutation responsible for the disease in the two affected sibs whose exomes we sequenced. Using only exome sequencing data, we succeeded in restricting the candidate genes to *FA2H* only. However, such a strategy is sensitive to the limitations which affect an exome sequencing approach involving only a single family. A first relevant issue is that specific variants might have been missed due to low coverage, even adding up the reads from two individuals. For instance, the *FA2H* mutation was covered by 7 reads (5 and 2 reads in the two sibs, respectively), providing a coverage of only 3 reads deeper than the threshold we set for variant calling ( $\geq 5$ ). Also, exome sequencing does not identify certain classes of mutations, such as large insertions or deletions, which would therefore be missed if they were the genetic

defect in the single family analyzed. Finally, efficiency of EX-HOM largely depends on the capacity to filter out all the variants which are not related to the disorder. We have shown how efficient EX-HOM can be in restricting candidate variants, taking advantage of the genetic analysis typical of consanguinity. Furthermore, with the increasing amount of variants we are discovering in population resequencing projects, more and more individual variants will be excluded because they are recognized as polymorphisms. We have to be very careful when filtering out variants found in one or two heterozygous carriers, since, in principle, they might be pathogenic in the homozygous state. Also, we have to remark that the candidate variant in *EFCAB6* could not be excluded due to the genomic and genetic information alone. If we had not been able to discard it on the basis of *EFCAB6* expression, it would have been confounding. This can be an issue when applying EX-HOM to different families.

Due to the everyday decreasing costs of next-generation sequencing technologies, it will be feasible to annotate rare polymorphisms and their frequencies in population-specific mutational databases as already suggested for the Mediterranean region [1]. Exome sequencing of offspring of consanguineous parents will provide the advantage of eventually creating a homozygosity map of the human genome which will reveal whether these rare variants can be deleterious in the homozygous state or not.

### Web Resources

dbSNP: <http://www.ncbi.nlm.nih.gov/projects/SNP/>  
1000 Genomes project: <http://browser.1000genomes.org/index.html>

### Acknowledgements

This work was funded by the European Commission FP7 project 'CHERISH', grant No. 223692, the Italian Ministry of Health, Strategic Program, RFPS-4-631972, and the Comitato Telethon Fondazione Onlus, grant No. GGP0102. Targeted exome enrichment and Illumina sequencing experiments were performed at the Applied Genomics Institute (IGA), Udine, Italy.

## References

- Ozcelik T, Kanaan M, Avraham KB, Yanoukakos D, Mégarbané A, Tadmouri GO, Middleton L, Romeo G, King MC, Levy-Lahad E: Collaborative genomics for human health and cooperation in the Mediterranean region. *Nat Genet* 2010;42:641–645.
- Bilgüvar K, Oztürk AK, Louvi A, Kwan KY, Choi M, Tatli B, Yalnizoglu D, Tüysüz B, Çağlayan AO, Gökben S, Kaymakçalan H, Barak T, Bakircioğlu M, Yasuno K, Ho W, Sanders S, Zhu Y, Yilmaz S, Dinçer A, Johnson MH, Bronen RA, Koçer N, Per H, Mane S, Pamir MN, Yalçinkaya C, Kumandaş S, Topçu M, Ozmen M, Sestan N, Lifton RP, State MW, Günel M: Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 2010;9:207–210.
- Funari VA, Krakow D, Nevarez L, Chen Z, Funari TL, Vatanavicharn N, Wilcox WR, Rimoin DL, Nelson SF, Cohn DH: BMPER mutation in diaphanospondylodysostosis identified by ancestral autozygosity mapping and targeted high-throughput sequencing. *Am J Hum Genet* 2010;87:532–537.
- Kalay E, Yigit G, Aslan Y, Brown KE, Pohl E, Bicknell LS, Kayserili H, Li Y, Tüysüz B, Nürnberg G, Kiess W, Koegl M, Baessmann I, Buruk K, Toraman B, Kayipmaz S, Kul S, Ikbal M, Turner DJ, Taylor MS, Aerts J, Scott C, Milstein K, Dollfus H, Wieczorek D, Brunner HG, Hurler M, Jackson AP, Rauch A, Nürnberg P, Karagüzel A, Wollnik B: CEP152 is a genome maintenance protein disrupted in Seckel syndrome. *Nat Genet* 2011;43:23–26.
- Becker J, Semler O, Gilissen C, Li Y, Bolz HJ, Giunta C, Bergmann C, Rohrbach M, Koerber F, Zimmermann K, de Vries P, Wirth B, Schoenau E, Wollnik B, Veltman JA, Hoischen A, Netzer C: Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. *Am J Hum Genet* 2011;88:362–371.
- Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, Garimella KV, Fisher S, Abreu J, Barry AJ, Fennell T, Banks E, Ambrogio L, Cibulskis K, Kernysky A, Gonzalez E, Rudzicz N, Engert JC, DePristo MA, Daly MJ, Cohen JC, Hobbs HH, Altshuler D, Schonfeld G, Gabriel SB, Yue P, Kathiresan S: Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med* 2010;363:2220.
- Krawitz PM, Schweiger MR, Rödelsperger C, Marcelis C, Kölsch U, Meisel C, Stephani F, Kinoshita T, Murakami Y, Bauer S, Isau M, Fischer A, Dahl A, Kerick M, Hecht J, Köhler S, Jäger M, Grünhagen J, de Condor BJ, Doelken S, Brunner HG, Meinecke P, Passarge E, Thompson MD, Cole DE, Horn D, Roscioli T, Mundlos S, Robinson PN: Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nat Genet* 2010;42:827–829.
- Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BB, Veltman JA: De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* 2010;42:483–485.
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ: Exome sequencing identifies the cause of a Mendelian disorder. *Nat Genet* 2010;42:30–35.
- Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP: Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci* 2009;106:19096–19101.
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacherjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J: Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;460:272–276.
- Garone C, Pippucci T, Cordelli DM, Zuntini R, Castegnaro G, Marconi C, Graziano C, Marchiani V, Verrotti A, Seri M, Franzoni E: FA2H-related disorders: a novel c.270+3A>T splice-site mutation leads to a complex neurodegenerative phenotype. *Dev Med Child Neurol* 2011, E-pub ahead of print.
- Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–595.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: The Genome Analysis Toolkit: a Map-Reduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–1303.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: PLINK: a tool set for whole-genome association and population based linkage analyses. *Am J Hum Genet* 2007;81:559–575.
- McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, Macleod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson JF: Runs of homozygosity in European populations. *Am J Hum Genet* 2008;83:359–372.
- Woods CG, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, Stern R, Raymond FL, Sandford R, Malik Sharif S, Karbani G, Ahmed M, Bond J, Clayton D, Inglehearn CF: Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am J Hum Genet* 2006;78:889–896.
- Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin – rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002;30:97–101.
- Thusberg J, Olatubosun A, Vihinen M: Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat* 2011;32:358–368.
- Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P: Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 2009;25:2744–2750.
- Ramensky V, Bork P, Sunyaev S: Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002;30:3894–3900.
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–249.
- Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073–1081.
- Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R: Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat* 2009;30:1237–1244.
- Brunak S, Engelbrecht J, Knudsen S: Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol* 1991;220:49–65.
- Roux-Rouquie M, Chauvet ML, Munnich A, Frezal J: Human genes involved in chromatin remodeling in transcription initiation, and associated diseases: an overview using the GENATLAS database. *Mol Genet Metab* 1999;67:261–277.
- Berglund L, Björling E, Oksvold P, Fagerberg L, Asplund A, Szigvarto CA, Persson A, Ottosson J, Wernérus H, Nilsson P, Lundberg E, Sivertsson A, Navani S, Wester K, Kampf C, Hober S, Pontén F, Uhlén M: A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol Cell Proteomics* 2008;7:2019–2027.
- Niki T, Takahashi-Niki K, Taira T, Iguchi-Arigo SM, Ariga H: DJBP: a novel DJ-1-binding protein, negatively regulates the androgen receptor by recruiting histone deacetylase complex, and DJ-1 antagonizes this inhibition by abrogation of this complex. *Mol Cancer Res* 2003;1:247–261.



- 30 Eckhardt M, Yaghootfam A, Fewou SN, Zöller I, Gieselmann V: A mammalian fatty acid hydroxylase responsible for the formation of alpha-hydroxylated galactosylceramide in myelin. *Biochem J* 2003;388:245–254.
- 31 Zöller I, Meixner M, Hartmann D, Büsow H, Meyer R, Gieselmann V, Eckhardt M: Absence of 2-hydroxylated sphingolipids is compatible with normal neural development but causes late-onset axon and myelin sheath degeneration. *J Neurosci* 2008;24:9741–9754.
- 32 Walsh T, Shahin H, Elkan-Miller T, Lee MK, Thornton AM, Roeb W, Abu Rayyan A, Loulus S, Avraham KB, King MC, Kanaan M: Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPM2 as the cause of nonsyndromic hearing loss DFNB82. *Am J Hum Genet* 2010;87:90–94.
- 33 Sirmaci A, Walsh T, Akay H, Spiliopoulos M, Sakalar YB, Hasanefendioglu-Bayrak A, Duman D, Farooq A, King MC, Tekin M: MASP1 mutations in patients with facial, umbilical, coccygeal, and auditory findings of Carnevale, Malpuech, OSA, and Michels syndromes. *Am J Hum Genet* 2010;87:679–686.
- 34 Pierce SB, Walsh T, Chisholm KM, Lee MK, Thornton AM, Fiumara A, Opitz JM, Levy-Lahad E, Klevit RE, King MC: Mutations in the DBP-deficiency protein HSD17B4 cause ovarian dysgenesis, hearing loss, and ataxia of Perrault Syndrome. *Am J Hum Genet* 2010;87:282–288.
- 35 Gilissen C, Arts HH, Hoischen A, Spruijt L, Mans DA, Arts P, van Lier B, Stehouwer M, van Reeuwijk J, Kant SG, Roepman R, Knoppers NV, Veltman JA, Brunner HG: Exome sequencing identifies WDR35 variants involved in Sensenbrenner syndrome. *Am J Hum Genet* 2010;87:418–423.

## Erratum

---

In the article by Gordon et al., ‘A New Expectation-Maximization Statistical Test for Case-Control Association Studies Considering Rare Variants Obtained by High-Throughput Sequencing’ [*Hum Hered* 2011;71:113–125], an author name has been misspelled: Francisco De La Vega has to be Francisco M. De La Vega.