KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Winand, R., Hens, K., Dondorp, W., de Wert, G., Moreau, Y., Vermeesch, J., Liebaers, I., Aerts, J. (2014). **In vitro screening of embryos by whole-genome sequencing: now, in the future or never?** Human Reproduction, 29 (4), 842-851.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://dx.doi.org/10.1093/humrep/deu005
Journal homepage	http://humrep.oxfordjournals.org/
Author contact	raf.winand@esat.kuleuven.be *+ 32 (0)16 32 86 43*
IR	url in Lirias https://lirias.kuleuven.be/handle/123456789/430436

(article begins on next page)

In vitro screening of embryos by whole-genome sequencing: now, in the future or never?

Raf Winand[1,2,†] & Kristien Hens[3,†], Wybo Dondorp[4], Guido de Wert[4], Yves Moreau[1,2], Joris Robert Vermeesch[5], Inge Liebaers[6], Jan Aerts[1,2]

[1]KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics

[2]iMinds Future Health Department

[3]Maastricht University, Health, Ethics & Society, Research Institute GROW, Maastricht University

[4]Department of Health, Ethics and Society, Maastricht University, Research Institutes CAPHRI and GROW, Maastricht University

[5]KU Leuven, Laboratory of Cytogenetics and Genome Research, Centre for Human Genetics (CME)

[6]University Hospital Brussels, Centre for Medical Genetics

[†]The authors consider that the first two authors should be regarded as joint First Authors

Abstract: Study question: What is the analytical and clinical validity and the clinical utility of in vitro screening of embryos by whole-genome sequencing?

Summary answer: At present there are still many limitations in terms of analytical and clinical validity and utility. In addition many ethical questions remain.

What is known already: Whole genome sequencing of in vitro embryos is technically possible. Many loss of function mutations exist in the general population without serious effects on the phenotype of the individual. Also, annotations of genes and the reference

genome are still not 100% correct.

Study design, size, duration: We used publicly available samples from the 1000 Genomes project and Complete Genomics, together with 42 samples from in-house research samples of parents from trios to investigate the presence of loss of function mutations in healthy individuals.

Participants/materials, setting, methods: In the samples we looked for mutations in genes that are associated with a selection of severe Mendelian disorders with a known molecular basis. We looked for mutations predicted to be damaging by PolyPhen and SIFT and for mutations annotated as disease causing in HGMD.

Main results and the role of chance: More than 40% of individuals who can be considered healthy have mutations that are predicted to be damaging in genes associated with severe Mendelian disorders or are annotated as disease causing.

Limitations, reasons for caution: The analysis relies on current knowledge and databases are continuously updated to reflect our increasing knowledge about the genome. In the process of our analysis several updates were already made.

Wider implications of the findings: At this moment it is not advisable to use whole-genome sequencing as a tool to set up health profiles to select embryos for transfer. We also raise some ethical questions that have to be addressed before this technology can be used for embryo selection.

Study funding/competing interest(s): None

Trial registration number: N/A

Introduction

Genetic testing of preimplantation embryos is a generally accepted approach in the context of preimplantation genetic diagnosis (PGD): patients with a known risk to transmit a specific genetic condition, or with a known chromosomal rearrangement, can opt for PGD to select embryos without the relevant disease-causing mutation. A second application of genetic testing of embryos is preimplantation genetic screening (PGS) for aneuploidy. Although not yet sufficiently proven by randomized controlled trials, this is offered by some centers to subfertile patients undergoing in vitro fertilization (IVF) as a treatment of infertility or as part of PGD, with the aim of improving their chances of a successful pregnancy. Whole-genome sequencing (WGS) and analysis, which determines and analyses the entire DNA sequence of an individual in one procedure, has been performed in single cells and single blastomeres (Navin, *et al.*, 2011, Xu, *et al.*, 2012, Voet, *et al.*, 2013). WGS (followed by a targeted analysis) could be a generic approach for PGD, avoiding time-consuming and labor-intensive customized PGD workups. WGS might also be an elegant alternative method for PGS, since full chromosomal aneuploidies and expected segmental imbalances associated with chromosome rearrangements can be easily identified (Harper and Sengupta, 2012). As WGS costs are now approaching the costs for array-based single-cell PGD or PGS, WGS may become a useful auxiliary technique for embryo testing as already performed in such context (Baslan, *et al.*, 2012, Hens, *et al.*, 2013, Simpson, *et al.*, 2013).

In addition to these applications, WGS could theoretically be used to considerably extend the scope of embryo testing. This would entail a widening of the aims of the procedure. In addition to either helping people to have a child without a specific disorder (as in PGD), or to attempt to increase the chances of a successful IVF pregnancy (as in PGS for aneuploidy), the aim of WGS in embryo testing

would be to ensure that children born after IVF or IVF/PGD are free from major disorders. Unlike classical PGD, embryo testing with these aims would be a form of medical screening, as it is a form of indiscriminate genetic testing without clinical data. One of the accepted criteria for responsible screening is that there should be a suitable test (Wilson and Jungner, 1968). This means that both the *analytical and clinical validity* of the test must have been demonstrated. The analytical validity of a genetic test is its ability to accurately determine the genotype of interest. Clinical validity is the accuracy with which the test can then predict a phenotype. If the test performs poor in these regards, this will adversely affect the *clinical utility* of the screening. This last concept refers to the balance of aim-related advantages and unavoidable disadvantages (drawbacks and costs), and is as such directly linked with the ethical acceptability of screening programs (Sanderson, *et al.*, 2005, Dondorp, *et al.*, 2010). In this paper, we assess the analytical and clinical validity of WGS-based testing as a necessary (though not a sufficient) condition for the clinical utility and ethical acceptability of extended or comprehensive embryo screening using this technology.

Materials and methods

We ascertained whether state-of-the-art technology and know-how could adequately distinguish benign polymorphisms from disease causing mutations and help deciding which embryo to transfer. We investigated how many apparently healthy adults carry mutations predicted to be damaging or annotated as disease causing - under the rationale that if such mutations were sufficient to cause the severe early-onset phenotype (and therefore predictive in a screening setting), they should be absent from the exome of apparently healthy adults.

Disease selection

We obtained the Online Mendelian Inheritance in Man (OMIM; McKusik-Nathns Institute of Genetic Medicine, 2012) list of diseases with a known molecular basis, consisting of more than 3000 diseases, from the OMIM website. This list was cross-referenced with the Human Phenotype Ontology (HPO)

database (Robinson and Mundlos, 2010) to provide an overview of diseases and their associated phenotypes. The resulting set contained 2,172 diseases from which we selected diseases characterized by dysmorphology and early onset symptoms. By selecting early onset disorders we ensure that individuals affected by one of these disorders would already show symptoms at the time of sequencing. In addition, only diseases that had the inheritance annotated in OMIM as 'autosomal dominant', 'autosomal dominant type; high penetrance' and 'autosomal recessive' were selected. As a result 132 autosomal dominant and 215 autosomal recessive diseases were retained. The complete list of these diseases can be found in supplementary Table 1 and Table 2.

Samples

For our analysis we used both private and publicly available samples from adult individuals who can be considered healthy at the time of sequencing (*i.e.* who did not exhibit clear signs of a congenital disorder at the time of sampling). We had access to 42 exome sequences from in-house research samples. These are trio samples of which the parents are considered healthy as they are symptom free. The sequences of affected children were also available but were not used in the initial analysis. In addition to our high quality exome sequences we downloaded two freely available sets from the 1000 Genomes project (McVean *et al.*, 2012) (1kG) and Complete Genomics (Drmanac *et al.*, 2010) (CG). The 1kG data came from the phase 1 integrated release version 3 from April 30, 2012, containing genotype calls for more than 1000 individuals. The data from CG was from 69 individuals, the sample names of which can be found in Table 3. All these sequences are from people who do not express symptoms of the selected disorders and therefore considered 'healthy'. Because some of the samples of these publicly available sets are from trios, we removed the samples from related individuals from the data sets leaving 1004 samples from the 1kG data and 50 from the CG data.

Transcripts

To localize the mutations in the genes and the different transcripts of those genes, we used the Ruby Ensembl API (Strozzi and Aerts, 2011) to connect to the Ensembl core database version 70. In

addition, we also looked at transcripts present in the Consensus CDS Project (CCDS) database Hs37.3 (Pruitt *et al.*, 2009).

Mutations predicted to be damaging

When analyzing genomes in the search of causative mutations, prediction algorithms are often used to predict the effect of a mutation on the protein. We used an in-house database called Annotate-it (Sifrim *et al.*, 2012) that holds detailed information on the selected genes, to retrieve prediction scores from SIFT (Ng, 2003) and PolyPhen (Adzhubei *et al.*, 2010) for the identified mutations. More information about these algorithms can be found in the supplements. Because PolyPhen and SIFT only give scores to missense mutations we considered nonsense and splice site mutations also to be damaging in the analysis. Indels were not included in the analysis.

Mutations described to be damaging in literature

In addition to the prediction algorithms, we also looked for mutations that are described as disease-causing in literature. For information on these mutations, we used the Human Gene Mutation Database (HGMD) containing more than 70,000 disease-causing mutations (Stenson *et al.*, 2003). For our analysis, we only selected those mutations with associations not considered tenuous by the curators of the HGMD and in the disease-causing category 'DM'.

Results

The results from this analysis show that many healthy individuals have mutations predicted to be damaging or annotated as disease causing in HGMD in genes associated with severe developmental disorders. For the 1kG samples all mutations are included, *i.e.* mutations that were found in both the low coverage and exome sequences. For the in-house data sets we only retained mutations with a read depth of at least 30x and phred score of 30. Relaxing these constraints leads to a higher number of mutations, which are described in supplementary Table 4 and Table 5.

Autosomal dominant disorders – Mutations predicted to be damaging

When looking for mutations predicted to be damaging, we found that depending on the data set, 98-100% of healthy individuals had damaging mutations in genes associated with the selected autosomal dominant disorders. In the 1kG data set we found a median of 8 mutations (min. 3, max. 14), in the Complete Genomics data set a median of 9 mutations (min. 4, max. 14) and in the in-house data set a median of 2 mutations (min. 0, max. 5) per individual. In only one sample from the in-house data set no damaging mutations were found. The distribution of the number of mutations can be seen in Figure 1. Because variants that are frequently found in the population are unlikely to cause these severe Mendelian disorders, we only retained the variants with a minor allele frequency (MAF) in the 1000 genomes of < 1%. Applying this filter leads to a large decrease in identified variants with now 40-94% of individuals carrying damaging mutations. In the 1kG data set we found on a median of 0 (min. 0, max. 5) mutations, in the Complete Genomics data set a median of 2 mutations (min. 0, max. 6) and in the in-house data set a median of 0 mutations (min. 0, max. 4) per individuals. The distribution of the number of mutations can be seen in Figure 2.

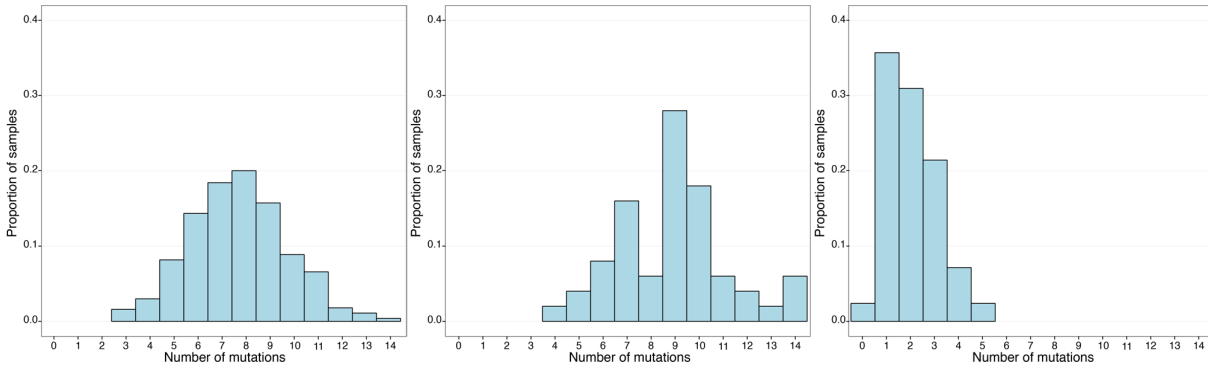


Figure 1

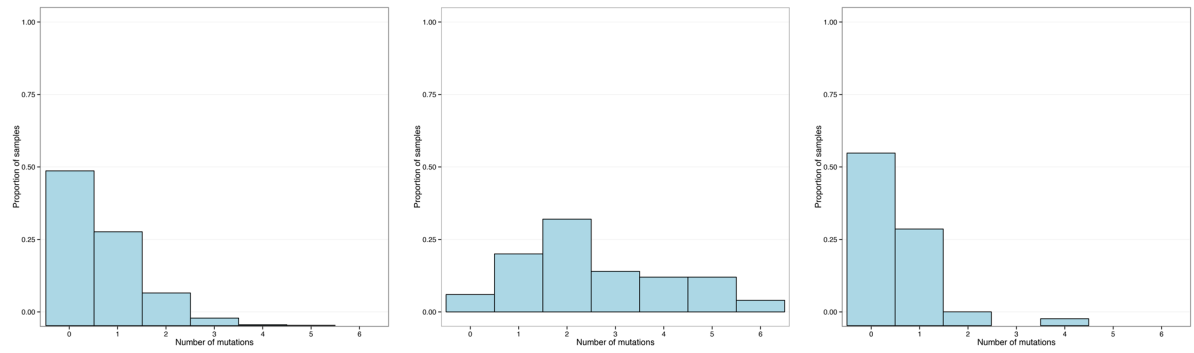


Figure 2

As explained by MacArthur *et al.* (2012) faulty gene annotation is a likely cause for these genes containing a large number of deleterious mutations. An example of this can be found in EDARADD (Ectodysplasin-A receptor-associated death domain), which is the gene with the highest number of unfiltered damaging mutations in all data sets. Mutations in EDARADD cause the autosomal dominant form of ectodermal dysplasia, characterized by sparse hair, missing or abnormal teeth and the inability to sweat (Cluzeau *et al.*, 2011). For this gene we found mutations in 83-98% of the samples. These mutations however are not annotated as disease causing in HGMD. In the 1kG data the most occurring mutation in EDARADD is at chromosome 1, position 236557771 G>A (dbSNP id: rs966365). So even though it is predicted to be damaging by both SIFT and PolyPhen, it is frequently found in that population (91%) suggesting an annotation error in the reference genome. Another observation we made was that most of the genes with the most common mutations have multiple transcripts and most have at least one transcript that is not affected by these mutations.

After filtering out the variants with a MAF in the 1000 genomes of > 1%, we find differences in the percentage of samples showing mutations between the data sets. For instance, the most occurring variant in the in-house data set was found in NOTCH2 which is associated with Hajdu-Cheney syndrome a disease characterized by coarse face, short neck, hirsutism, joint laxity, and bone dysplasias (Ramos, *et al.*, 1998). In this case mutations were found in 14% of the in-house samples, in 1% of the 1kG samples and 6% of the CG samples. The fact that a predicted damaging variant is frequently found in the local population but not in the 1000 genomes data set and dbSNP can indicate that the variant is a benign polymorphism in this population. An example of this is a variant that we identified at chromosome 3, position 98300354 (A>C) in CPOX, which is associated with hereditary coproporphyria. This mutation is found in almost 10% of the in-house samples while it doesn't occur in the other data sets and is not found in dbSNP.

In total we identified 323, 120, and 21 distinct mutations with a MAF of < 1% in the 1kG data in 69, 59, and 18 genes in respectively the 1kG, CG, and in-house data sets. An overview of the genes with

the largest number of predicted damaging mutations and some phenotypic information about the disorder is shown in supplementary Table 6 and Table 7

Autosomal dominant disorders – Mutations present in HGMD

Taking only the mutations into account that are annotated as disease causing in HGMD for the severe disorders from our list, we identified mutations in 20% of the 1kG samples, 22% of the Complete Genomics samples and 12% of the in-house samples. The number of affected samples with curated disease causing variants is thus much lower compared to those identified by the prediction programs. When looking at the in-house samples, we found that the mutations were considered to be damaging by either PolyPhen or SIFT but not by both. A list of the diseases of which causative mutations were identified can be found in supplementary Table 8.

In the in-house data we identified at total of six heterozygous mutations in five individuals. One of these mutations (CM023740) had a MAF >1% in the 1000 genomes indicating a possibly suspicious annotation or a variant with reduced penetrance. At the time of the initial analysis this mutation was annotated as causing extrahepatic biliary atresia, a feature of Alagille syndrome, but it was removed from HGMD at a later time. For Greig cephalopolysyndactyly syndrome two distinct mutations were found in GLI3, *i.e.* P707S (CM970684) (Wild *et al.*, 1997) and I808M (CM990707) (Kalff-Suske *et al.*, 1999). In a functional analysis, both mutations were found to cause misregulation of GLI3-localization by Krauss et al. (Krauss *et al.*, 2009). The mutation causing Rubinstein-Taybi syndrome - A981T (CM021081) in CREBBP - was identified by Coupry *et al.* (2002) in a set of 60 patients. Because these samples are part of trios we also had access to the samples of the children. Out of a total of 5 children, 4 children were heterozygous for the same mutation as their parent(s) but also did not express the disease. The fact that these variants are present in apparently healthy individuals may hint towards either (1) sequencing errors, (2) false positive entries in HGMD or (3) incomplete penetrance of certain variants which would make them of low predictive value in a PGS context.

Autosomal recessive disorders – Mutations predicted to be damaging

For autosomal recessive disorders there are two categories of affected individuals. Either they are homozygous for a mutation or they are compound heterozygous. We found almost the same number of samples that were homozygous for damaging mutations in genes associated with autosomal recessive disorders as we found mutations for the autosomal dominant disorders. Approximately 98-100% of samples were homozygous for at least one mutation predicted to be damaging. The number of damaging mutations per individual was lower however. In the 1kG data set we found a median of 4 mutations (min. 0, max. 9), in the Complete Genomics data set a median of 4 mutations (min. 1, max. 7) and also in the in-house data set a median of 4 mutations (min. 0, max. 7) mutations per individual. The distribution of the mutation counts can be seen in Figure 3. Limiting the variants to those with a MAF of < 1% in the 1000 genomes data show a large decrease especially in the 1kG data set. Less than 2% of the samples are homozygous for a damaging mutation in the 1kG data set and 56-69% in the CG and in-house data set. The median in the 1kG data sets is 0 mutations (min. 0, max. 1), in the Complete Genomics data set 1 mutation (min. 0, max. 2) and in the in-house data set 1 mutations (min. 0, max. 1). The distribution of the mutation counts can be seen in Figure 4. In total, we identified 17, 4, and 1 distinct homozygous mutations with MAF in 1kG of less than 1% in 16, 4, and 1 genes in respectively the 1kG, CG, and in-house data set. As for the autosomal dominant disorders, supplementary Table 9 and Table 10 show an overview of the genes with the largest number of mutations and the corresponding diseases.

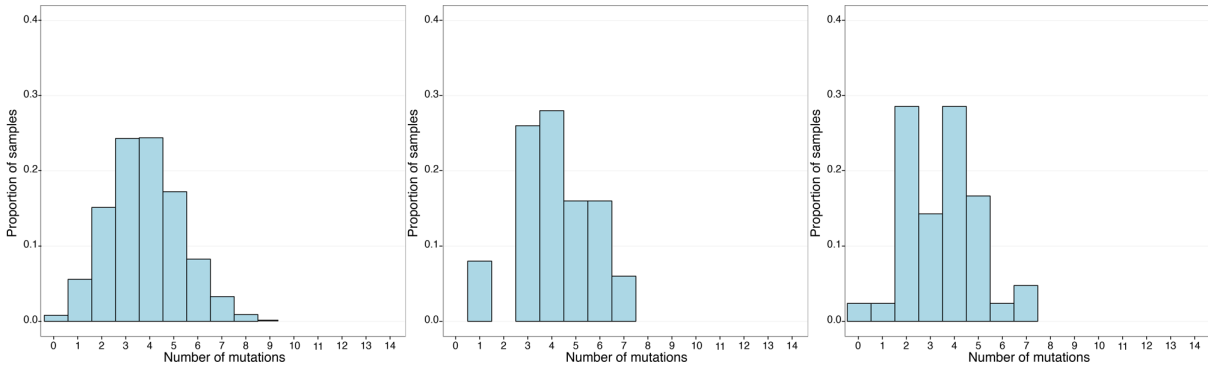


Figure 3

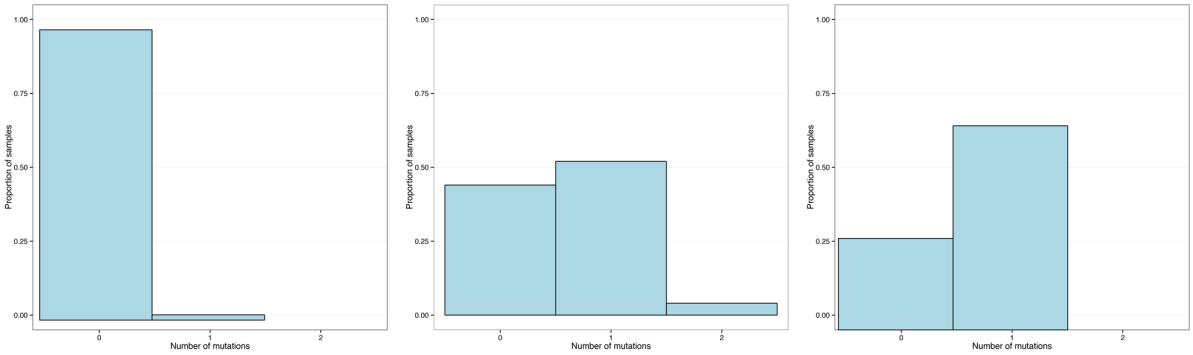


Figure 4

In addition to homozygosity, we also investigated the case of compound heterozygosity. Because no haplotype information was available for the in-house dataset, we identified the number of samples that have two different mutations predicted to be damaging in each gene. In this case the difference between the datasets is also large. While for some diseases a number of individuals are identified in one data set, there are no individuals with damaging mutations in another and vice versa. In this case the gene NEB, which is associated with nemaline myopathy, showed the highest number of individuals that had at least two distinct mutations with 2% of samples in the 1kG data set, 2% in the CG data set but no samples in the in-house data set affected. An overview of the number of affected samples and genes can be found in supplementary Table 11.

Autosomal recessive disorders – Mutations present in HGMD

For HGMD mutations, we found that the percentage of samples that are homozygous and therefore are expected to express the disease, is much lower than with autosomal dominant diseases. In the in-house dataset no homozygous HGMD mutations were found. An overview of all affected genes can be found in supplementary Table 12. The most frequently identified mutation was found in 16 individuals in the 1kG data (1.8%) on chromosome 17, position 7915912 C>T (dbSNP id: rs34598902). This mutation was thought to be causative for Leber Congenital Amaurosis (Zernant *et al.*, 2005), a disease characterized by congenital blindness in 10-20% of all blind children. However, it was later found in the normal population (Ito, 2004) with a MAF in the 1000 genomes of 8.26%. While the mutation was annotated as disease causing in HGMD at the time of our analysis, it has been

reclassified at a later time to the category 'DM?' indicating a tenuous association. Another mutation in the same gene at position 7912879 C>T (dbSNP id: rs28743021) was found in two individuals in the 1kG data who were homozygous for this mutation. This mutation is annotated as disease causing in HGMD based on a study by Koenekoop *et al*. (2002).

For compound heterozygous mutations, again counted as two distinct mutations, no positive samples were identified in the in-house data, only one in the CG dataset with a MAF >1% and a few in the 1kG dataset. An overview of the number of affected samples is shown in supplementary Table 13.

Discussion

Analytical validity

The analytical validity of WGS-based embryo testing is constituted by the quality of single-cell sequencing and the accuracy of interpretation. At present, single-cell genome-wide sequencing is not yet as good as sequencing based on multiple cells. On the assumption that this limitation will be overcome, we focus here on the accuracy of the interpretation of sequencing data.

We found that with increasing quality thresholds the number of individuals that carry mutations predicted to be damaging decreases in the samples from our in-house data set. Except for a common mutation found in the CG and in-house data set, we found fewer individuals with mutations predicted to be damaging in genes associated with autosomal recessive disorders. This may be due to the fact that homozygous mutations occur less frequently than heterozygous mutations. Or it may be linked to the clinical validity of the test: autosomal dominant disorders are more likely to show reduced penetrance or variable expressivity and therefore symptoms of the disorder might not be present in these individuals although the mutations are.

Clinical validity

More than 40% of genomes of healthy individuals in our study had a genetic mutation thought to be causative for severe autosomal dominant congenital disorders. Also, some healthy individuals were

homozygous or compound heterozygous for mutations associated with autosomal recessive disorders. These results are in line with the results of Xue *et al.* (2012) and MacArthur and Tyler-Smith (2010) who studied loss-of-function and disease-causing variants in healthy individuals. Our analysis relies on current prediction programs and databases. It is obvious from this analysis that current programs predicting protein damage based on exonic sequence information alone show a number of false positives and will need to improve to enable proper future phenotype prediction (Sifrim, *et al.*, 2013). In our results we see that filtering the identified mutations on the MAF in the 1000 genomes, significantly reduces the number of identified mutations per individual. In the CG and in-house data set however, a larger percentage of individuals still show mutations. This suggests that these mutations might be local polymorphisms that are not identified in the 1000 genomes project. An in-depth knowledge of the variants found in the local population can therefore lead to improved filters. Finally, the HGMD database is frequently updated and thus the results of each analysis might be different with each version of the database.

A specific challenge for embryo screening is that predictions have to be made in the absence of phenotypic information. This is different from the use of WGS-based testing in the postnatal context where this technology is being introduced as a means of finding a diagnosis for existing patients with a phenotype that could not be clarified with more traditional diagnostic approaches. However, in embryo selection, genetic information is all that is available, except for classical PGD where the phenotype of the parents can be taken into account. We considered to what extent adding WGS-based preconception testing of the prospective parents might help fill in this lacuna by providing additional context information. This approach may perhaps be helpful where findings related to dominant disorders are concerned, as information about genotype and phenotype of the parents will contribute to making better predictions about the health of children resulting from embryos with the relevant mutations found in WGS-based screening. However, preconception testing of the prospective parents seems less useful for the interpretation of findings related to recessive disorders, given our observation that healthy persons may be compound heterozygous or even homozygous for mutations in this category.

Advancing knowledge in genomics may enhance the clinical validity of WGS-based embryo testing, as more will be known about the influence of modifier and protective genes and possible epigenetic influences that may explain our findings (von Kanel, *et al.*, 2013). Also, other projects, such as the 'Deciphering Developmental Disorders' (Firth and Wright, 2011) study may contribute to identifying the genetic cause of disorders, making knowledge about the genome more complete. Although it can be expected that with increasing up-to-date knowledge about the relation between the genome and the phenotype it will become possible in the future to make better predictions about the health of children resulting from the transfer of an embryo with a specific genotype, we conclude that at present the clinical validity of WGS-based embryo screening is limited.

Clinical utility

As the analytical and clinical validity are still insufficient, a necessary condition for the introduction of extended or comprehensive embryo screening ('suitable test') is not met. Clearly, this would adversely affect the *clinical utility* of the screening, as it would lead to discarding embryos that may well develop into healthy children. Some may still defend the rapid introduction of WGS-based screening, arguing that mutations will be found that are thought to be causative at least in some cases, and that non-transfer of embryos with such mutations may still be warranted. However, as the number of available good-quality embryos in an IVF-cycle is limited, there may not be much scope for choosing a mutation-free embryo, and the option of a new cycle just to avoid a 'suspected' embryo may well be disproportional. Moreover, with the current state of the art, prospective parents would be faced with choices based on unreliable predictions about the health of the children they could have as a result of transferring this or that embryo. Overestimation of the predictive value of adverse findings of WGS-based embryo screening may lead to the couple remaining childless, or to undermining their confidence in the health of any children they still decide to have. In addition to this, they may also be confronted with equally unreliable information suggesting that they themselves are carriers of a potentially severe disease.

Conclusion and final remarks

At present the drawbacks of WGS-based embryo screening appear to outweigh the possible benefits for prospective parents, making the introduction of such screening in clinical practice unwarranted and at best premature. It may be that further scientific developments will lead to improving the predictive accuracy of WGS-based embryo. Although that would take away the drawback of decision-making on the basis of unreliable information, it does not automatically follow that WGS-based screening would then be unproblematic.

As screening aimed at simultaneously excluding a more than a limited number of genetic risk factors would very soon run into the problem of leaving no embryo for transfer (Hens, *et al.*, 2012), a possible approach would be to always select the embryo(s) with the best health profile (while maintaining a threshold of at least not transferring high-risk embryos). However, the clinical utility of that approach would depend on whether meaningful choices between embryos with various health profiles can indeed be made. It is at least not clear whether that is the case. And even if it were, the amount of relevant data and the fact that (where genetic susceptibilities are concerned) even reliable information would be about risks rather than certainties, the feasibility of well-considered decision making would not be obvious. Moreover, whose decisions should this be? As is it can be argued that both prospective parents (whose child it will be) and professionals (given their active involvement in the creation of the child) have a say in this matter, WGS-based screening may have the potential of leading to conflicts between those stakeholders. Last but not least a difficult problem is that choosing the embryo with the best profile will inevitably mean that children are born for whom some health prospects might already be known. The ethical question here is whether the future child should be allowed to decide for herself what she wants to know about her genome (Hens, *et al.*, 2013).

We conclude that even if current limitations in terms of analytical and clinical validity can be overcome, the notion of WGS-based embryo screening still raises some difficult questions. It is presently unclear whether these can be satisfactorily answered. A possible alternative approach with its own advantages and disadvantages (De Wert, 2009) would consist of preconception screening of the prospective parents followed by targeted PGD.

Acknowledgments

This research is supported by Research Council KU Leuven: GOA/10/09 MaNet, KUL PFV/10/016 SymBioSys ; IWT O&O ExaScience Life ; Hercules Stichting: Hercules III PacBio RS ; iMinds ; Art&D Instance ; VSC Tier 1: exome sequencing ; COST: Action BM1006: NGS Data analysis network. Kristien Hens was supported by the Centre for Society and Life Sciences (project number: 70.1.074)

Figure legends

Figure 1: Histograms showing the proportion of samples with a certain amount of mutations that are predicted to be damaging. Samples are from 1000 Genomes, Complete Genomics and in-house (left to right).

Figure 2: Histograms showing the proportion of samples with a certain amount of mutations that are predicted to be damaging and are found with a minor allele frequency of <1% in the 1000 Genomes data. Samples are from 1000 Genomes, Complete Genomics and in-house (left to right).

Figure 3: Histograms showing the proportion of samples with a certain amount of mutations that are predicted to be damaging. Samples are from 1000 Genomes, Complete Genomics and in-house (left to right). Only homozygous mutations are counted.

Figure 4: Histograms showing the proportion of samples with a certain amount of mutations that are predicted to be damaging and are found with a minor allele frequency of <1% in the 1000 Genomes data (left to right). Samples are from 1000 Genomes, Complete Genomics and in-house. Only homozygous mutations are counted.

References

1000 Genomes Project Consortium, Abecasis, GR, Auton, A, Brooks, LD, DePristo, MA, Durbin, RM, Handsaker, RE, Kang, HM, Marth, GT and McVean, GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; **491**:56-65.

Adzhubei I a, 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.

Baslan, T, Kendall, J, Rodgers, L, Cox, H, Riggs, M, Stepansky, A, Troge, J, Ravi, K, Esposito, D, Lakshmi, B *et al.* Genome-wide copy number analysis of single cells. *Nat Protoc*. 2012; **7**:1024-1041.

Cluzeau C, Hadj-Rabia S, Jambou M, Mansour S, Guigue P, Masmoudi S, Bal E, Chassaing N, Vincent M-C, Viot G, *et al.* Only four genes (EDA1, EDAR, EDARADD, and WNT10A) account for 90% of hypohidrotic/anhidrotic ectodermal dysplasia cases. *Hum Mutat* 2011;**32**:70–72.

Coupry, I, Roudaut, C, Stef, M, Delrue, MA, Marche, M, Burgelin, I, Taine, L, Cruaud, C, Lacombe, D and Arveiler, B. Molecular analysis of the CBP gene in 60 patients with Rubinstein-Taybi syndrome. *J Med Genet.* 2002; **39**:415-421.

De Wert, G. Preimplantation genetic testing: normative reflections. In Harper, J (ed), *Preimplantation Genetic Diagnosis.* 2009. Cambridge University Press, Cambridge, pp. 259-273.

Dondorp, W, Wert, G and Cornel, M. The Quality of Genetic Screening: An Integral Approach. In Kristoffersson, U, Schmidtke, J and Cassiman, JJ (eds), *Quality Issues in Clinical Genetic Services.* 2010. Springer Netherlands, pp. 165-172.

427 Drmanac, R, Sparks, AB, Callow, MJ, Halpern, AL, Burns, NL, Kermani, BG, Carnevali, P,
428 Nazarenko, I, Nilsen, GB, Yeung, G *et al.* Human genome sequencing using unchained base
429 reads on self-assembling DNA nanoarrays. *Science*. 2010; **327**:78-81.

430 Firth, HV and Wright, CF. The Deciphering Developmental Disorders (DDD) study. *Dev Med Child*
431 *Neurol*. 2011; **53**:702-703.

432 Harper, JC and Sengupta, SB. Preimplantation genetic diagnosis: state of the art 2011. *Hum Genet.*
433 2012; **131**:175-186.

434 Hens, K, Dondorp, W and de Wert, G. Embryos without secrets: An expert panel study on
435 comprehensive embryo testing and the responsibility of the clinician. *Eur J Med Genet.* 2012;
436 **56**:67-71.

437 Hens, K, Dondorp, W, Handyside, AH, Harper, J, Newson, AJ, Pennings, G, Rehmann-Sutter, C and
438 de Wert, G. Dynamics and ethics of comprehensive preimplantation genetic testing: a review of
439 the challenges. *Hum Reprod Update*. 2013; **19**:366-375.

440 Ito S. Novel Complex GUCY2D Mutation in Japanese Family with Cone-Rod Dystrophy.
441 *Investigative Ophthalmology & Visual Science* 2004;**45**:1480–1485.

442 Kalff-Suske, M, Wild, A, Topp, J, Wessling, M, Jacobsen, EM, Bornholdt, D, Engel, H, Heuer, H,
443 Aalfs, CM, Ausems, MG *et al.* Point mutations throughout the GLI3 gene cause Greig
444 cephalopolysyndactyly syndrome. *Hum Mol Genet*. 1999; **8**:1769-1777.

445 Koenekoop RK, Fishman G a, Iannaccone A, Ezzeldin H, Ciccarelli ML, Baldi A, Sunness JS, Lotery
446 AJ, Jablonski MM, Pittler SJ, *et al.* Electroretinographic abnormalities in parents of patients
447 with Leber congenital amaurosis who have heterozygous GUCY2D mutations. *Arch Ophthal*
448 2002;**120**:1325–1330.

449 Krauss S, So J, Hambrock M, Köhler A, Kunath M, Scharff C, Wessling M, Grzeschik K-H,
450 Schneider R, Schweiger S. Point mutations in GLI3 lead to misregulation of its subcellular
451 localization. *PLoS One* 2009;**4**:e7471.

452 MacArthur, DG and Tyler-Smith, C. Loss-of-function variants in the genomes of healthy humans.
453 *Hum Mol Genet.* 2010; **19**:R125-130.

454 MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger
455 L, Pickrell JK, Montgomery SB, *et al.* A systematic survey of loss-of-function variants in human
456 protein-coding genes. *Science* 2012;**335**:823–828.

457 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University. Online Mendelian
458 Inheritance in Man, OMIM. Accessed 2012-10-28, http://omim.org

459 Navin, N, Kendall, J, Troge, J, Andrews, P, Rodgers, L, McIndoo, J, Cook, K, Stepansky, A, Levy, D,
460 Esposito, D *et al.* Tumour evolution inferred by single-cell sequencing. *Nature*. 2011; **472**:90-
461 94.

462 Ng PC. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*
463 2003;**31**:3812–3814.

464 Pruitt KD, Harrow J, Harte R a, Wallin C, Diekhans M, Maglott DR, Searle S, Farrell CM, Loveland
465 JE, Ruef BJ, *et al.* The consensus coding sequence (CCDS) project: Identifying a common
466 protein-coding gene set for the human and mouse genomes. *Genome Res* 2009;**19**:1316–1323.

467 Ramos, FJ, Kaplan, BS, Bellah, RD, Zackai, EH and Kaplan, P. Further evidence that the Hajdu-
468 Cheney syndrome and the "serpentine fibula-polycystic kidney syndrome" are a single entity.
469 *Am J Med Genet.* 1998; **78**:474-481.

470 Robinson PN, Mundlos S. The human phenotype ontology. *Clin Genet* 2010;**77**:525–534.

471 Sanderson, S, Zimmern, R, Kroese, M, Higgins, J, Patch, C and Emery, J. How can the evaluation of
472 genetic tests be enhanced? Lessons learned from the ACCE framework and evaluating genetic
473 tests in the United Kingdom. *Genet Med.* 2005; **7**:495-500.

474 Sifrim A, Houdt JK Van, Tranchevent L-C, Nowakowska B, Sakai R, Pavlopoulos G a, Devriendt K,
475 Vermeesch JR, Moreau Y, Aerts J. Annotate-it: a Swiss-knife approach to annotation, analysis
476 and interpretation of single nucleotide variation in human disease. *Genome Med* 2012;**4**:73.

477 Sifrim, A, Popovic, D, Tranchevent, L-C, Ardeshirdavani, A, Sakai, R, Konings, P, Vermeesch, JR,
478 Aerts, J, De Moor, B and Moreau, Y. eXtasy: variant prioritization by genomic data fusion. *Nat*
479 *Methods.* 2013;**10**:1083-1084.

480 Simpson, JL, Rechitsky, S and Kuliev, A. Next-generation sequencing for preimplantation genetic
481 diagnosis. *Fertil Steril.* 2013; **99**:1203-1204.

482 Stenson PD, Ball E V, Mort M, Phillips AD, Shiel J a, Thomas NST, Abeysinghe S, Krawczak M,
483 Cooper DN. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 2003;**21**:577–
484 581.

485 Strozzi F, Aerts J. A Ruby API to query the Ensembl database for genomic features. *Bioinformatics*
486 2011;**27**:1013–1014.

487 Voet, T, Kumar, P, Van Loo, P, Cooke, SL, Marshall, J, Lin, ML, Zamani Esteki, M, Van der Aa, N,
488 Mateiu, L, McBride, DJ *et al.* Single-cell paired-end genome sequencing reveals structural
489 variation per cell cycle. *Nucleic Acids Res.* 2013; **41**:6119-6138.

490 von Kanel, T, Stanke, F, Weber, M, Schaller, A, Racine, J, Kraemer, R, Chanson, M, Tummler, B and
491 Gallati, S. Clinical and molecular characterization of the potential CF disease modifier syntaxin
492 1A. *Eur J Hum Genet* 2013; Epub ahead of print

493 Wild a, Kalff-Suske M, Vortkamp a, Bornholdt D, König R, Grzeschik KH. Point mutations in human
494 GLI3 cause Greig syndrome. *Hum Mol Genet* 1997;**6**:1979–1984.

495 Wilson, JM and Jungner, YG. Principles and practice of screening for disease. *WHO Chronicle:*
496 *Public Health Papers*. 1968; **22**:473.

497 Xu, X, Hou, Y, Yin, X, Bao, L, Tang, A, Song, L, Li, F, Tsang, S, Wu, K, Wu, H *et al.* Single-cell
498 exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell*.
499 2012; **148**:886-895.

500 Xue, Y, Chen, Y, Ayub, Q, Huang, N, Ball, EV, Mort, M, Phillips, AD, Shaw, K, Stenson, PD,
501 Cooper, DN *et al.* Deleterious- and disease-allele prevalence in healthy individuals: insights
502 from current predictions, mutation databases, and population-scale resequencing. *Am J Hum*
503 *Genet*. 2012; **91**:1022-1032.

504 Zernant, J, Kulm, M, Dharmaraj, S, den Hollander, AI, Perrault, I, Preising, MN, Lorenz, B, Kaplan,
505 J, Cremers, FP, Maumenee, I *et al.* Genotyping microarray (disease chip) for Leber congenital
506 amaurosis: detection of modifier alleles. *Invest Ophthalmol Vis Sci*. 2005; **46**:3052-3059.

507

200610	Achondrogenesis, type II
100800	Achondroplasia
101800	Acrodysostosis with hormone resistance
102370	Acromicric dysplasia
100300	Adams-oliver syndrome
103285	Adult syndrome
118450	Alagille syndrome1
170390	Andersen cardiodysrhythmic periodic paralysis
106260	Ankyloblepharon-ectodermal defects-cleft lip/palate
101200	Apert syndrome
108721	Atelosteogenesis type III
180500	Axenfeld-rieger syndrome, type I
153480	Bannayan-Riley-Ruvalcaba syndrome
123790	Beare-Stevenson cutis gyrata syndrome
130650	Beckwith-Wiedemann syndrome
110100	Blepharophimosis, ptosis, and epicanthus inversus
605039	Bohring-Opitz syndrome
112310	Boomerang dysplasia
113500	Brachyolmia type III
113620	Branchiooculofacial syndrome
613071	Bronchiectasis with or without elevated sweat chloride 3
114290	Campomelic dysplasia
131300	Camurati-Engelmann disease
115150	Cardiofaciocutaneous syndrome
169100	Char syndrome
214800	Charge syndrome
118400	Cherubism
119600	Cleidocranial dysplasia
262700	Combined pituitary hormone deficiency 4
255310	Congenital myopathy with fiber-type disproportion
122470	Cornelia de Lange syndrome 1
218040	Costello syndrome
123000	Craniometaphyseal dysplasia
123100	Craniosynostosis, type 1
123500	Crouzon syndrome
108120	Distal arthrogryposis type Ia
193700	Distal arthrogryposis type IIa
158300	Distal arthrogryposis type IX
121050	Distal arthrogryposis type VII
607323	Duane-Radial ray syndrome
604292	Ectrodactyly, ectodermal dysplasia, and cleft lip/palate syndrome
160120	Episodic ataxia, type I
164280	Feingold syndrome 1
614524	Fibrochondrogenesis 2
135100	Fibrodysplasia ossificans progressiva
175700	Greig cephalopolysyndactyly syndrome
102500	Hajdu-Cheney syndrome
121300	Hereditary coproporphyria
129500	Hidrotic ectodermal dysplasia
142946	Holoprosencephaly 4
609637	Holoprosencephaly 5
176670	Hutchinson-Gilford progeria syndrome
145900	Hypertrophic neuropathy of Dejerine-Sottas
146000	Hypochondroplasia
129490	Hypohidrotic ectodermal dysplasia
147750	Ivic syndrome
123150	Jackson-Weiss syndrome
147920	Kabuki syndrome 1
148050	KBG syndrome
610253	Kleefstra syndrome

118100	Klippel-Feil syndrome 1
156550	Kniest dysplasia
150250	Larsen syndrome
151100	Leopard syndrome 1
127300	Leri-Weill dyschondrosteosis
609192	Loeys-Dietz syndrome, type 1a
610168	Loeys-Dietz syndrome, type 1b
153400	Lymphedema-distichiasis syndrome
109150	Machado-Joseph disease
154700	Marfan syndrome
154780	Marshall syndrome
156400	Metaphyseal chondrodysplasia, jansen type
156500	Metaphyseal chondrodysplasia, schmid type
156530	Metatropic dysplasia
607932	Microphthalmia, syndromic 6
235730	Mowat-Wilson syndrome
602849	Muenke syndrome
186500	Multiple synostoses syndrome 1
139210	Myhre syndrome
161200	Nail-patella syndrome
601321	Neurofibromatosis-Noonan syndrome
163950	Noonan syndrome 1
164200	Oculodentodigital dysplasia
165800	Osteochondritis dissecans, short stature, and early-onset osteoarthritis
166200	Osteogenesis imperfecta, type I
166210	Osteogenesis imperfecta, type II
259420	Osteogenesis imperfecta, type III
166220	Osteogenesis imperfecta, type IV
166250	Osteoglophonic dysplasia
146510	Pallister-Hall syndrome
168400	Parastremmatic dwarfism
168500	Parietal foramina
168550	Parietal foramina with cleidocranial dysplasia
118800	Paroxysmal nonkinesigenic dyskinesia 1
169400	Pelger-Huet anomaly
101600	Pfeiffer syndrome
151210	Platyspondylic lethal skeletal dysplasia, Torrance type
119500	Popliteal pterygium syndrome
177170	Pseudoachondroplasia
129400	Rapp-Hodgkin syndrome
270450	Resistance to insulin-like growth factor I
180700	Robinow syndrome
180849	Rubinstein-Taybi syndrome 1
101400	Saethre-Chotzen syndrome
269150	Schinzel-Giedion midface retraction syndrome
269921	Sialuria
147891	Small patella syndrome
182290	Smith-Magenis syndrome
147250	Solitary median maxillary central incisor
603546	Spondyloepimetaphyseal dysplasia with joint laxity type 2
602111	Spondyloepimetaphyseal dysplasia, Missouri type
184250	Spondyloepimetaphyseal dysplasia, Strudwick type
183900	Spondyloepiphyseal dysplasia congenita
184095	Spondyloepiphyseal dysplasia, Maroteaux type
184252	Spondylometaphyseal dysplasia, Kozlowski type
271700	Spondyloperipheral dysplasia
108300	Stickler syndrome type I
604841	Stickler syndrome, type II
184840	Stickler syndrome, type III

184900	Stiff skin syndrome
206900	Syndromic microphthalmia 3
187600	Thanatophoric dysplasia, type I
187601	Thanatophoric dysplasia, type II
107480	Townes-Brocks syndrome
154500	Treacher Collins syndrome 1
190350	Trichorhinophalangeal syndrome, type I
190351	Trichorhinophalangeal syndrome, type III
119300	Van der Woude syndrome 1
148820	Waardenburg syndrome type III
277590	Weaver syndrome 1
608328	Weill-marchesani syndrome 2
277610	Weissenbacher-Zweymuller syndrome

Table I: List of OMIM IDs and corresponding disorders with autosomal dominant inheritance used in the analysis.

246450	3-@hydroxy-3-methylglutaryl-CoA lyase deficiency
257920	3MC syndrome 1
265050	3MC syndrome 2
100100	Absence of abdominal muscles with urinary tract abnormality
276820	Absence of ulna and fibula with severe limb deficiency
231550	Achalasia-addisonianism-alacrima syndrome
200500	Acheiropody
200600	Achondrogenesis, type Ia
200990	Acrocallosal syndrome
607778	Acrocapitofemoral dysplasia
602875	Acromesomelic dysplasia, Maroteaux type
202370	Adrenoleukodystrophy
218000	Agenesis of the corpus callosum with peripheral neuropathy
608688	Aicar transformylase/imp cyclohydrolase deficiency
225750	Aicardi-Goutieres syndrome 1
203800	Alstrom syndrome
205100	Amyotrophic lateral sclerosis 2, juvenile
241200	Antenatal Bartter syndrome, type 2
201750	Antley-Bixler syndrome with genital anomalies and disordered steroidogenesis
207410	Antley-Bixler syndrome without genital anomalies or disordered steroidogenesis
208400	Aspartylglucosaminuria
256050	Atelosteogenesis, type II
251290	Band-like calcification with simplified gyration and polymicrogyria
209900	Bardet-Biedl syndrome
211530	Brown-Vialetto-van Laere syndrome
211750	C syndrome
271900	Canavan disease
608836	Carnitine palmitoyltransferase II deficiency, lethal neonatal
201000	Carpenter syndrome
224050	Cerebellar ataxia, mental retardation, and dysequilibrium syndrome
214150	Cerebrooculofacioskeletal syndrome 1
214500	Chediak-Higashi syndrome
215045	Chondrodysplasia, Blomstrand type
200700	Chondrodysplasia, Grebe type
225060	Cleft lip/palate-ectodermal dysplasia syndrome
216360	Coach syndrome
216400	Cockayne syndrome, type a
133540	Cockayne syndrome, type b
216550	Cohen syndrome
262600	Combined pituitary hormone deficiency 2
212065	Congenital disorder of glycosylation, type Ia
601110	Congenital disorder of glycosylation, type Id
608799	Congenital disorder of glycosylation, type Ie
212066	Congenital disorder of glycosylation, type IIa
608540	Congenital disorder of glycosylation, type Ik
610768	Congenital disorder of glycosylation, type Im
263700	Congenital erythropoietic porphyria
612540	Congenital myopathy Compton-North
275100	Congential nongoitrous hypothyroidism 4
260660	Cousin syndrome
218330	Cranioectodermal dysplasia 1
607812	Craniolenticulosutural dysplasia
601378	Crisponi syndrome
219700	Cystic fibrosis
261515	D-bifunctional protein deficiency
600060	Deafness 2
251450	Desbuquois dysplasia
602398	Desmosterolosis
608022	Diaphanospondylodysostosis
251850	Diarrhea 2, with microvillus atrophy

222600	Diastrophic dysplasia
604320	Distal spinal muscular atrophy 1
246200	Donohue syndrome
223800	Dyggve-Melchior-Clausen disease
224230	Dyskeratosis congenita 1
224410	Dyssegmental dysplasia, Silverman-Handmaker type
609304	Early infantile epileptic encephalopathy 3
225320	Ehlers-Danlos syndrome, cardiac valvular form
225400	Ehlers-Danlos syndrome, type VI
225410	Ehlers-Danlos syndrome, type VII
600002	Eiken skeletal dysplasia
225500	Ellis-van Creveld syndrome
612651	Endocrine-cerebroosteodysplasia
603034	Endplate acetylcholinesterase deficiency
226600	Epidermolysis bullosa dystrophica
226730	Epidermolysis bullosa junctionalis with pyloric atresia
226900	Epiphyseal dysplasia, multiple, 4
226980	Epiphyseal dysplasia, multiple, with early-onset diabetes mellitus
211500	Fazio-Londe disease
208150	Fetal akinesia deformation sequence
228520	Fibrochondrogenesis
228930	Fibular aplasia or hypoplasia, femoral bowing and poly-, syn-, and oligodactyly
219000	Fraser syndrome
256540	Galactosialidosis
230800	Gaucher disease, type I
230900	Gaucher disease, type II
231005	Gaucher disease, type IIIc
231050	Geleophysic dysplasia 1
231670	Glutaric acidemia I
232200	Glycogen storage disease Ia
232240	Glycogen storage disease Ic
232500	Glycogen storage disease IV
230500	GM1-gangliosidosis, type I
230600	GM1-gangliosidosis, type II
230650	GM1-gangliosidosis, type III
609460	Goldberg-Shprintzen megacolon syndrome
235510	Hennekam lymphangiectasia-lymphedema syndrome
241530	Hereditary hypophosphatemic rickets with hypercalciuria
608233	Hermansky-Pudlak syndrome 2
236490	Hyalinosis, infantile systemic
236680	Hydrolethalus syndrome 1
215140	Hydrops-ectopic calcification-moth-eaten skeletal dysplasia
243700	Hyper-IgE recurrent infection syndrome
239300	Hyperphosphatasia with mental retardation
241410	Hypoparathyroidism-retardation-dysmorphism syndrome
242900	Immunoosseous dysplasia, Schimke type
241500	Infantile hypophosphatasia
269920	Infantile sialic acid storage disorder
262400	Isolated growth hormone deficiency, type Ia
243800	Johanson-Blizzard syndrome
213300	Joubert syndrome
608091	Joubert syndrome 2
608629	Joubert syndrome 3
226700	Junctional epidermolysis bullosa, Herlitz type
262650	Kowarski syndrome
611722	Krabbe disease, atypical, due to saposin a deficiency
249700	Langer mesomelic dysplasia
262500	Laron syndrome
204000	Leber congenital amaurosis 1
204100	Leber congenital amaurosis 2

253310 Lethal congenital contracture syndrome 1
260600 Leukodystrophy, hypomyelinating, 3
278000 Lysosomal acid lipase deficiency
248370 Mandibuloacral dysplasia with type a lipodystrophy
608612 Mandibuloacral dysplasia with type b lipodystrophy
248500 Mannosidosis, alpha b, lysosomal
248800 Marinesco-Sjogren syndrome
212720 Martsolf syndrome
608728 Matrilin-3 related spondyloepimetaphyseal dysplasia
249000 Meckel syndrome, type 1
224690 Meier-Gorlin syndrome 1
613803 Meier-Gorlin syndrome 3
613804 Meier-Gorlin syndrome 4
613805 Meier-Gorlin syndrome 5
277380 Methylmalonic aciduria and homocystinuria, cblF type
610377 Mevalonic aciduria
210720 Microcephalic osteodysplastic primordial dwarfism, type II
206920 Microphthalmia with limb anomalies
255320 Minicore myopathy with external ophthalmoplegia
257300 Mosaic variegated aneuploidy syndrome 1
252500 Mucolipidosis II alpha/beta
252900 Mucopolysaccharidosis type IIIa
252920 Mucopolysaccharidosis type IIIb
252930 Mucopolysaccharidosis type IIIc
252940 Mucopolysaccharidosis type IIId
253000 Mucopolysaccharidosis type IVa
253010 Mucopolysaccharidosis type IVb
253200 Mucopolysaccharidosis type VI
253220 Mucopolysaccharidosis type VII
253250 Mulibrey nanism
245600 Multiple joint dislocations, short stature, craniofacial dysmorphism
265000 Multiple pterygium syndrome, Escobar variant
253290 Multiple pterygium syndrome, lethal type
236670 Muscular dystrophy-dystroglycanopathy
256030 Nemaline myopathy 2
256500 Netherton syndrome
610127 Neuronal ceroid lipofuscinosis 10
204500 Neuronal ceroid lipofuscinosis 2
204200 Neuronal ceroid lipofuscinosis 3
256731 Neuronal ceroid lipofuscinosis, 5
257200 Niemann-Pick disease, type a
257220 Niemann-Pick disease, type c1
251260 Nijmegen breakage syndrome
258315 Omodysplasia 1
610682 Osteogenesis imperfecta, type VII
610915 Osteogenesis imperfecta, type VIII
259700 Osteopetrosis 1
259730 Osteopetrosis 3
259720 Osteopetrosis 5
215150 Otospondylomegaepiphyseal dysplasia
239000 Paget disease, juvenile
608013 Perinatal lethal Gaucher disease
261540 Peters-Plus syndrome
262190 Pineal hyperplasia, insulin-resistant diabetes mellitus
263200 Polycystic kidney disease
225753 Pontocerebellar hypoplasia, type IV
263650 Popliteal pterygium syndrome, lethal type
263750 Postaxial acrofacial dysostosis
259900 Primary hyperoxaluria, type I
251200 Primary microcephaly 1

259775	Raine syndrome
275210	Restrictive dermopathy, lethal
215100	Rhizomelic chondrodysplasia punctata, type 1
222765	Rhizomelic chondrodysplasia punctata, type 2
600121	Rhizomelic chondrodysplasia punctata, type 3
268300	Roberts syndrome
268310	Robinow syndrome
269000	SC phocomelia syndrome
269250	Schneckenbecken dysplasia
255800	Schwartz-Jampel syndrome, type 1
269500	Sclerosteosis 1
210600	Seckel syndrome 1
612541	Severe congenital neutropenia 4
263520	Short rib-polydactyly syndrome, type II
260400	Shwachman-Diamond syndrome
270400	Smith-Lemli-Opitz syndrome
612936	Spastic quadriplegic cerebral palsy 3
272460	Spondylocarpotarsal synostosis syndrome
612350	Spondylocheirodysplasia, Ehlers-Danlos syndrome-like
277300	Spondylocostal dysostosis 1
143095	Spondyloepiphyseal dysplasia with congenital joint dislocations
271665	Spondylometaepiphyseal dysplasia, short limb-hand type
601559	Stuve-Wiedemann syndrome
272300	Sulfocysteinuria
272800	Tay-Sachs disease
273395	Tetraamelia
273750	Three M syndrome 1
259600	Torg-Winchester syndrome
609015	Trifunctional protein deficiency
264700	Vitamin D hydroxylation-deficient rickets, type 1a
600081	Vitamin D hydroxylation-deficient rickets, type 1b
277440	Vitamin D-dependent rickets, type 2a
600118	Warburg micro syndrome 1
277600	Weill-Marchesani syndrome 1
214100	Zellweger syndrome

Table II: List of OMIM IDs and associated disorders with autosomal recessive inheritance used in the analysis.

NA19240-L2-200-37ASM	NA12888-200-37-ASM	NA19700-200-37-ASM
NA19649-200-37-ASM	NA12890-200-37-ASM	NA19701-200-37-ASM
NA18505-200-37-ASM	NA12891-200-37-ASM	NA19703-200-37-ASM
NA19025-200-37-ASM	NA12893-200-37-ASM	NA19704-200-37-ASM
NA19238-L2-200-37-ASM	NA18501-200-37-ASM	NA19735-200-37-ASM
NA20846-200-37-ASM	NA18502-200-37-ASM	NA20502-200-37-ASM
NA19670-200-37-ASM	NA18504-200-37-ASM	NA20509-200-37-ASM
NA19834-200-37-ASM	NA18508-200-37-ASM	NA20510-200-37-ASM
NA18947-200-37-ASM	NA18517-200-37-ASM	NA20511-200-37-ASM
HG00731-200-37-ASM	NA18526-200-37-ASM	NA20845-200-37-ASM
HG00732-200-37-ASM	NA18537-200-37-ASM	NA20847-200-37-ASM
HG00733-200-37-ASM	NA18555-200-37-ASM	NA20850-200-37-ASM
NA06985-200-37-ASM	NA18558-200-37-ASM	NA21732-200-37-ASM
NA06994-200-37-ASM	NA18940-200-37-ASM	NA21733-200-37-ASM
NA07357-200-37-ASM	NA18942-200-37-ASM	NA21737-200-37-ASM
NA10851-200-37-ASM	NA18956-200-37-ASM	NA21767-200-37-ASM
NA12004-200-37-ASM	NA19017-200-37-ASM	NA12880-200-37-ASM
NA12877-200-37-ASM	NA19020-200-37-ASM	NA12881-200-37-ASM
NA12879-200-37-ASM	NA19026-200-37-ASM	NA12883-200-37-ASM
NA12882-200-37-ASM	NA19129-200-37-ASM	NA12886-L2-200-37-ASM
NA12884-200-37-ASM	NA19239-L2-200-37-ASM	NA12889-L2-200-37-ASM
NA12885-L2-200-37-ASM	NA19648-200-37-ASM	NA12878-200-37-ASM
NA12887-L2-200-37-ASM	NA19669-200-37-ASM	NA12892-L2-200-37-ASM

Table III: List of Coriel IDs used from the Complete Genomics dataset.

Phred score/read depth	**10**	**20**	**30**
Number of Individuals	22 (42)	19 (42)	17 (41)
Min	0 (1)	0 (1)	0 (0)
Max	4 (8)	4 (7)	4 (5)
Median	1 (4)	0 (3)	0 (2)

Table IV: Number of damaging mutations found in in-house samples in genes associated with autosomal dominant disorders. The minimum (Min), maximum (Max), and median number of mutations is given for the different threshold of read depth and Phred score. The number between brackets indicate the number found without filtering on a MAF of <1% in 1kG.

Phred/read depth	**10**	**20**	**30**
Individuals	34 (42)	33 (42)	29 (41)
Min	0 (1)	0 (1)	0 (0)
Max	1 (8)	1 (8)	1 (7)
Median	1 (4)	1 (4)	1 (4)

Table V: Number of damaging, homozygous mutations found in in-house samples in genes associated with autosomal recessive disorders. The minimum (Min), maximum (Max), and median number of mutations is given for the different thresholds of read depth and Phred score. The number between brackets indicate the number found without filtering on a MAF of <1% in 1kG.

Disease	Gene	1kG	% 1kG	CG	%CG	IH	%IH
Hajdu-Cheney syndrome	NOTCH2	14 (675)	1.39% (67.23%)	3 (7)	6.00% (14.00%)	6 (6)	14.29% (14.29%)
Coproporphyria, hereditary	CPOX	3 (26)	0.30% (2.59%)	1 (1)	2.00% (2.00%)	4 (4)	9.52% (9.52%)
Hypertrophic Neuropathy of Dejerine-Sotas	EGR2/MPZ/PMP22/PRX	10 (10)	1.00% (1.00%)	3 (3)	6.00% (6.00%)	2 (2)	4.76% (4.76%)
Arthrogryposis, distal, type 2a	MYH3	21 (21)	2.09% (2.09%)	2 (2)	4.00% (4.00%)	2 (2)	4.76% (4.76%)
Smith-Magenis syndrome	RAI1	39 (729)	3.88% (72.61%)	11 (37)	22.00% (74.00%)	0 (4)	0.00% (9.52%)
Ectodermal dysplasia	EDARADD EDAR	23 (988)	2.29% (98.41%)	1 (49)	2.00% (98.00%)	1 (35)	2.38% (83.33%)
KBG Syndrome	ANKRD11	23 (157)	2.29% (15.64%)	8 (13)	16.00% (26.00%)	0 (0)	0.00% (0.00%)
Osteochondritis dissecans, short stature and early-onset osteoarthritis	ACAN	70 (70)	6.97% (6.97%)	6 (6)	12.00% (12.00%)	0 (0)	0.00% (0.00%)
Trichorhinophalangeal syndrome, type I and II	TRPS1	17 (17)	1.69% (1.69%)	2 (2)	4.00% (4.00%)	1 (1)	2.38% (2.38%)
Marshall syndrome, Stickler syndrome, type II	COL11A1	15 (15)	1.49% (1.49%)	1 (1)	2.00% (2.00%)	1 (1)	2.38% (2.38%)

Table VI: List of genes with mutations predicted to be damaging, the corresponding diseases and the number and proportion of samples in which these mutations are found. The number between brackets indicate the number found without filtering on a MAF of <1% in 1kG.

Gene	Disease	Description
NOTCH2	Hajdu-Cheney Syndrome	Skeletal disorder characterized by short stature, coarse and dysmorphic facies, bowing of the long bones, and vertebral anomalies. Facial features include hypertelorism, bushy eyebrows, micrognathia, small mouth with dental anomalies, low-set ears, and short neck. There is progressive focal bone destruction, including acroosteolysis and generalized osteoporosis. Additional and variable features include hearing loss, renal cysts, and cardiovascular anomalies.
CPOX	Hereditary coproporphyria	Acute hepatic porphyria (see this term) characterized by the occurrence of neuro-visceral attacks and, more rarely, by the presence of cutaneous lesions. The prevalence in Europe is estimated at about 1/1,000,000. The disease manifests after puberty and preferentially affects women.
EGR2 MPZ PMP22 PRX	Hypertrophic Neuropathy of Dejerine-Sottas	Dejerine-Sottas neuropathy is a demyelinating peripheral neuropathy with onset in infancy. Affected individuals have delayed motor development due to severe distal motor and sensory impairment, resulting in difficulties in gait. Some patients have generalized hypotonia in infancy. Other features may include pes cavus, scoliosis, and sensory ataxia.
MYH3	Arthrogryposis, Distal, Type 2A	A rare congenital disorder characterised by dysmorphic status combining bone anomalies and joint contractures with typical facies. Affects less than 1 in 1,000,000.
RAI1	Smith-Magenis Syndrome	Variable intellectual deficit, sleep disturbance, craniofacial and skeletal anomalies, psychiatric disorders, and speech and motor delay. Affects 1 in 15,000 people.
EDARADD EDAR	Hypohidrotic Ectodermal Dysplasia	Hypotrichosis, hypodontia, and anhidrosis. Affects 1 in 17,000 people
ANKRD11	KBG Syndrome	KBG syndrome is a rare condition characterised by a typical facial dysmorphism, macrodontia of the upper central incisors, skeletal (mainly costovertebral) anomalies and developmental delay. To date it has been reported in 45 individuals.
ACAN	Osteochondritis dissecans, short stature and early-onset osteoarthritis	Familial osteochondritis dissecans is a rare genetic skeletal disorder characterized clinically by abnormal chondro-skeletal development, disproportionate short stature and skeletal deformation mainly affecting the knees, hips, ankles and elbows with onset generally in late childhood or adolescence.
TRPS1	Trichorhinophalangeal syndrome, type I	Trichorhinophalangeal syndrome type 1 is a malformation syndrome characterized by short stature, sparse hair, a bulbous nasal tip and cone-shaped epiphyses, as well as severe generalized shortening of all phalanges, metacarpals and metatarsal bones. Affects less than 1 in 1,000,000

	Trichorhinophalangeal syndrome, type II	Trichorhinophalangeal syndrome type 2 is characterized by the association of intellectual deficit and numerous other anomalies including redundant skin, multiple cartilaginous exostoses, characteristic facies and cone-shaped phalangeal epiphyses.
COL11A1	Marshall syndrome	Marshall syndrome is a multisystem disease characterized by ocular abnormalities, sensorineural hearing loss, craniofacial anomalies and anhidrotic ectodermal dysplasia. Affects less than 1 in 1,000,000
	Stickler syndrome, type III	Stickler syndrome is a clinically variable and genetically heterogeneous disorder characterized by ocular, auditory, skeletal, and orofacial abnormalities. Most forms of Stickler syndrome are characterized by the eye findings of high myopia, vitreoretinal degeneration, retinal detachment, and cataracts. Additional findings may include midline clefting (cleft palate or bifid uvula), Pierre Robin sequence, flat midface, sensorineural or conductive hearing loss, mild spondyloepiphyseal dysplasia, and early-onset osteoarthritis.

Table VII: Overview of genes with the highest number of damaging mutations, the corresponding diseases and a short description of that disease. Compiled from information from OMIM, Orphanet (INSERM, 1997) and U.S. National Library of Medicine. (National Institutes of Health, 2013)

Disease	Gene	1kG	% 1kG	CG	%CG	IH	%IH
Alagille syndrome (extrahepatic biliary atresia)	JAG1	6 (85)	0.60% (8.47%)	0 (4)	0.00% (8.00%)	0 (1)	0.00% (2.38%)
Greig cephalopolysyndactyly syndrome	GLI3	2 (2)	0.20% (0.20%)	0 (0)	0.00% (0.00%)	3 (3)	7.14% (7.14%)
Kallmann syndrome 3	PROKR2	16 (16)	1.59% (1.59%)	3 (3)	6.00% (6.00%)	0 (0)	0.00% (0.00%)
Marfan syndrome	FBN1	8 (8)	0.80% (0.80%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Rubinstein-Taybi syndrome 1	CREBBP	18 (18)	1.79% (1.79%)	0 (0)	0.00% (0.00%)	2 (2)	4.76% (4.76%)

Table VIII: List of genes with mutations predicted to be disease causing (DM) in HGMD, the corresponding diseases and the number and proportion of samples in which these mutations are found. The number between brackets indicate the number found without filtering on a MAF of <1% in 1kG.

Disease	Gene	1kG	% 1kG	CG	%CG	IH	%IH
Joubert syndrome 2	TMEM216	0 (0)	0.00% (0.00%)	27 (27)	54.00% (54.00%)	29 (29)	69.05% (69.05%)
Alstrom syndrome	ALMS1	2 (4)	0.20% (0.40%)	1 (1)	2.00% (2.00%)	0 (0)	0.00% (0.00%)
Epidermolysis bullosa, junctional, Herlitz type	LAMA3 LAMB3 LAMC2	1 (12)	0.10% (1.20%)	1 (1)	2.00% (2.00%)	0 (0)	0.00% (0.00%)
Mannosidosis, Alpha B	MAN2B1	0 (1)	0.00% (0.10%)	1 (1)	2.00% (2.00%)	0 (0)	0.00% (0.00%)
Band-like calcification with simplified gyration	OCLN	2 (2)	0.20% (0.20%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Ellis-Van Creveld syndrome	EVC/EVC2	2 (11)	0.20% (1.10%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)

Table IX: List of genes with homozygous mutations predicted to be damaging, the corresponding diseases and the number and proportion of samples in which these mutations are found. The number between brackets indicate the number found without filtering on a MAF of <1% in 1kG.

TMEM216	Joubert Syndrome 2	Joubert syndrome is a genetically heterogeneous autosomal recessive disorder characterized by a specific hindbrain malformation, which is referred to as the 'molar tooth sign' (MTS) on brain MRI, hypotonia, developmental delay, oculomotor apraxia, and breathing abnormalities.
ALMS1	Alstrom syndrome	Alström syndrome is a multisystemic disorder characterized by cone-rod dystrophy, hearing loss, obesity, insulin resistance and hyperinsulinemia, type 2 diabetes mellitus, dilated cardiomyopathy (DCM), and progressive hepatic and renal dysfunction. Approximately 450 cases have been identified worldwide.
LAMA3 LAMB3 LAMC2	Epidermolysis bullosa, junctional, Herlitz type	Junctional epidermolysis bullosa, Herlitz-type is a severe subtype of junctional epidermolysis bullosa (JEB, see this term) characterized by blisters and extensive erosions, localized to the skin and mucous membranes. Prevalence is 1-9 in 1,000,000
MAN2B1	Alpha-mannosidosis B	Alpha-mannosidosis is an inherited lysosomal storage disorder characterized by immune deficiency, facial and skeletal abnormalities, hearing impairment, and intellectual deficit. It occurs in approximately 1 in 500,000 live births.
OCLN	Band-like calcification with simplified gyration and polymicrogyria	Congenital intrauterine infection-like syndrome is characterised by the presence of microcephaly and intracranial calcifications at birth accompanied by neurological delay, seizures and a clinical course similar to that seen in patients after intrauterine infection with Toxoplasma gondii, Rubella, Cytomegalovirus, Herpes simplex. Prevalence is less than 1 in 1,000,000
EVC/EVC2	Ellis-Van Creveld syndrome	Ellis-van Creveld syndrome (EVC) is a chondral and ectodermal dysplasia characterized by short ribs, polydactyly, growth retardation, and ectodermal and heart defects. It is a rare disease with approximately 150 cases reported worldwide.

Table X: Overview of genes associated with autosomal recessive disorders with the highest number of damaging mutations, the corresponding diseases and a short description of that disease. Compiled from information from OMIM, Orphanet, and U.S. National Library of Medicine.

Disease	Gene	1kG	% 1kG	CG	%CG	IH	%IH
Acrocallosal syndrome	KIF7	0 (20)	1.99% (1.99%)	0 (2)	0.00% (2.00%)	2 (2)	4.76% (4.76%)
Dyssegmental dysplasia, Silverman-Handmaker type	HSPG2	2 (10)	0.20% (1.00%)	0 (0)	0.00% (0.00%)	1 (3)	2.38% (7.14%)
Hyper-IgE recurrent infection syndrome	DOCK8	0 (2)	0.00% (0.20%)	0 (1)	0.00% (2.00%)	1 (2)	2.38% (4.76%)
Schwartz-Jampel syndrome, type I	HSPG2	2 (10)	0.20% (1.00%)	0 (0)	0.00% (0.00%)	1 (3)	2.38% (7.14%)
Nemaline myopathy 2	NEB	23 (68)	2.29% (6.77%)	1 (14)	2.00% (28.00%)	0 (0)	0.00% (0.00%)
Fraser syndrome	FRAS1/ FREM2	10 (16)	1.00% (1.59%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Cystic fibrosis	CFTR	8 (8)	0.80% (0.80%)	1 (1)	2.00% (2.00%)	0 (0)	0.00% (0.00%)
Pontocerebellar hypoplasia, type IV	TSEN54	0 (156)	0.00% (15.54%)	0 (3)	0.00% (6.00%)	0 (2)	0.00% (4.76%)

Table XI: List of genes that have at least two distinct mutations that are predicted to be damaging. The corresponding diseases and the number and proportion of samples in which these mutations are found is also given. The number between brackets indicate the number found without filtering on a MAF of <1% in 1kG.

Disease	Gene	1kG	% 1kG	CG	%CG	IH	%IH
Polycystic kidney disease	PKHD1	1 (2)	0.10% (0.20%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Leber congenital amaurosis I	GUCY2D	0 (18)	0.00% (1.79%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Ellis-van Creveld syndrome	EVC	0 (10)	0.00% (1.00%)	0 (1)	0.00% (2.00%)	0 (0)	0.00% (0.00%)
Trifunctional protein deficiency	HADHA	0 (5)	0.00% (0.50%)	0 (1)	0.00% (2.00%)	0 (0)	0.00% (0.00%)
Pituitary hormone deficiency, combined, 2	PROP1	0 (4)	0.00% (0.40%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Mucopolysaccharidosis type IIIc	HGSNAT	0 (3)	0.00% (0.30%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Aspartylglucosaminuria	AGA	0 (1)	0.00% (0.10%)	0 (1)	0.00% (2.00%)	0 (0)	0.00% (0.00%)
Mucopolysaccharidosis type VI	ARSB	0 (1)	0.00% (0.10%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Cystic fibrosis	CFTR	1 (1)	0.10% (0.10%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Carnitine palmitoyltransferase II deficiency	CPT2	1 (1)	0.10% (0.10%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Mucopolysaccharidosis type IVa	GALNS	1 (1)	0.10% (0.10%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)

Table XII: List of genes with mutations predicted to be disease causing (DM) in HGMD, the corresponding diseases and the number and proportion of samples in which these mutations are found. The number between brackets indicate the number found without filtering on a MAF of <1% in 1kG.

Disease	Gene	1kG	% 1kG	CG	%CG	IH	%IH
Polycystic kidney disease	PKHD1	4 (14)	0.40% (1.39%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Cystic fibrosis	CFTR	4 (4)	0.40% (0.40%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Leber congenital amaurosis 1	GUCY2D	0 (3)	0.00% (0.30%)	0 (1)	0.00% (2.00%)	0 (0)	0.00% (0.00%)
Aspartylglucosaminuria	AGA	2 (2)	0.20% (0.20%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Congenital disorder of glycosylation, type Ia	PMM2	1 (1)	0.10% (0.10%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)
Mucopolysaccharidosis type IIIc	HGSNAT	0 (1)	0.00 (0.10%)	0 (0)	0.00% (0.00%)	0 (0)	0.00% (0.00%)

Table XIII: List of genes that have at least two distinct mutations that are predicted to be disease causing (DM) in HGMD. The corresponding diseases and the number and proportion of samples in which these mutations are found is also given. The number between brackets indicate the number found without filtering on a MAF of <1% in 1kG.