Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants

Aziz Belkadi[a,b,1], Alexandre Bolze[c,f,1], Yuval Itan[c], Quentin B. Vincent[a,b], Alexander Antipenko[c], Bertrand Boisson[c], Jean-Laurent Casanova[a,b,c,d,e,2] and Laurent Abel[a,b,c,2]

[a] Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Paris, France, EU

[b] Paris Descartes University, Imagine Institute, Paris, France, EU

[c] St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, the Rockefeller University, New York, NY, USA

[d] Howard Hughes Medical Institute, New York, NY, USA

[e] Pediatric Hematology-Immunology Unit, Necker Hospital for Sick Children, Paris, France, EU

[f] Present address: Department of Cellular and Molecular Pharmacology, California Institute for Quantitative Biomedical Research, University of California, San Francisco, CA, USA

[1,2] Equal contributions

Corresponding authors: Jean-Laurent Casanova (casanova@rockefeller.edu) or Laurent Abel (laurent.abel@inserm.fr)

Key words :

Next generation sequencing, exome, genome, genetic variants, Mendelian disorders

Abstract

We compared whole-exome sequencing (WES) and whole-genome sequencing (WGS) for the detection of single-nucleotide variants (SNVs) in the exomes of six unrelated individuals. In the regions targeted by exome capture, the mean number of SNVs detected was 84,192 for WES and 84,968 for WGS. Only 96% of the variants were detected by both methods, with the same genotype identified for 99.2% of them. The distributions of coverage depth (CD), genotype quality (GQ), and minor read ratio (MRR) were much more homogeneous for WGS than for WES data. Most variants with discordant genotypes were filtered out when we used thresholds of CD≥8X, GQ≥20, and MRR≥0.2. However, a substantial number of coding variants were identified exclusively by WES (105 on average) or WGS (692). We Sanger sequenced a random selection of 170 of these exclusive variants, and estimated the mean number of false-positive coding variants per sample at 79 for WES and 36 for WGS. Importantly, the mean number of real coding variants identified by WGS and missed by WES (656) was much larger than the number of real coding variants identified by WES and missed by WGS (26). A substantial proportion of these exclusive variants (32%) were predicted to be damaging. In addition, about 380 genes were poorly covered (~27% of base pairs with CD<8X) by WES for all samples, including 49 genes underlying Mendelian disorders. We conclude that WGS is more powerful and reliable than WES for detecting potential disease-causing mutations in the exome.

Introduction

Whole-exome sequencing (WES) is now routinely used for detecting rare and common genetic variants in humans (1–7). Whole-genome sequencing (WGS) is becoming an attractive alternative approach, due to its decreasing cost (8, 9). However, it remains difficult to interpret variants lying outside the coding regions of the genome. Diagnostic and research laboratories, whether public or private, therefore tend to search for coding variants, which can be detected by WES, first. Such variants can also be detected by WGS, but few studies have compared the efficiencies of WES and WGS for this specific purpose (10–12). Here, we compared WES and WGS for the detection and quality of single-nucleotide variants (SNVs) located within the regions of the human genome covered by WES, using the most recent next-generation sequencing (NGS) technologies. Our goals were to identify the method most efficient and reliable for identifying SNVs in coding regions of the genome, to define the optimal analytical filters for decreasing the frequency of false-positive variants, and to characterize the genes that were hard to sequence by either technique.

Results

To compare the two NGS techniques, we performed WES with the Agilent Sure Select Human All Exon kit 71Mb (v4 + UTR), and WGS with the Illumina TruSeq DNA PCR-Free sample preparation kit on blood samples from six unrelated Caucasian patients with isolated congenital asplenia (OMIM #271400). We used the genome analysis toolkit (GATK) best-practice pipeline for the analysis of our data (13). We used the GATK Unified Genotyper (14) to call variants, and we restricted the calling process to the regions covered by the Sure Select Human All Exon kit 71Mb plus 50 bp of flanking sequences on either side of the each of the captured regions, for both WES and WGS samples. These regions, referred to as the WES71+50 region, included 180,830 full-length and 129,946 partial protein-coding exons from 20,229 genes (**Table S1**). There were 65 million reads per sample, on average, mapping to this region in WES, corresponding to a mean coverage of 73X (**Table S2**), consistent with the standards set by recent large-scale genomic projects aiming to decipher disease-causing variants by WES (11, 14, 15). On average, 35 million reads per sample mapped to this region by WGS, corresponding to a mean coverage of 39X (**Table S2**). The mean (range) number of SNVs detected was 84,192 (82,940-87,304) per exome and 84,968 (83,340-88,059) per genome. The mean number of SNVs per sample called by both methods was 81,192 (~96% of all variants)

(**Fig. S1A)**. For 99.2% of these SNVs, WES and WGS yielded the same genotype, and 62.4% of these concordant SNVs were identified as heterozygous (**Fig. S1B**). These results are similar to those obtained in previous WES studies (1, 5, 16). Most of the remaining SNVs (329 of 415) with discordant genotypes for these two techniques, were identified as homozygous variants by WES and as heterozygous variants by WGS. A smaller number of variants (86, on average), were identified as heterozygous by WES and homozygous by WGS (**Fig. S1B**).

We then investigated in WES and WGS data the distribution of the two main parameters assessing SNV quality generated by the GATK variant calling process (14): coverage depth (CD), corresponding to the number of aligned reads covering a single position; and genotype quality (GQ), which ranges from 0 to 100 (higher values reflect more accurate genotype calls). We also assessed the minor read ratio (MRR), which was defined as the ratio of reads for the less covered allele (reference or variant allele) over the total number of reads covering the position at which the variant was called. Overall, we noted reproducible differences in the distribution of these three parameters between WES and WGS. The distribution of CD was skewed to the right in the WES data, with a median at 50X but a mode at 18X, indicating low levels of coverage for a substantial proportion of variants (**Fig. 1A**). By contrast, the distribution of CD was normal-like for the WGS data, with the mode and median coinciding at 38X (**Fig. 1A**). We found that 4.3% of the WES variants had a CD < 8X, versus only 0.4% of the WGS variants. The vast majority of variants called by WES or WGS had a GQ close to 100. However, the proportion of variants called by WES with a GQ < 20 (3.1%) was, on average, twice that for WGS (1.3%) (**Fig. 1B**). MRR followed a similar overall distribution for WES and WGS heterozygous variants, but peaks corresponding to values of MRR of 1/7, 1/6, 1/5 and 1/4 were detected only for the WES variants (**Fig. 1C**). These peaks probably corresponded mostly to variants called at a position covered by only 7, 6, 5 and 4 reads, respectively. The overall distributions of these parameters indicated that the variants detected by WGS were of higher and more uniform quality than those detected by WES.

Next, we looked specifically at the distribution of these parameters for the variants with genotypes discordant between WES and WGS, denoted as discordant variants. The distribution of CD for WES variants showed that most discordant variants had low coverage, at about 2X, with a CD distribution very different from that of concordant variants (**Fig. S2A**). Moreover, most discordant variants had a GQ < 20 and a MRR < 0.2 for WES (**Fig. S2B**). By contrast, the distributions of CD, GQ, and MRR were very similar between WGS variants discordant with

WES results and WGS variants concordant with WES results (**Fig. S2**). All these results indicate that the discordance between the genotypes obtained by WES and WGS was largely due to the low quality of WES calls for the discordant variants. We therefore conducted subsequent analyses by filtering out low-quality variants. We retained SNVs with a CD ≥ 8X and a GQ ≥ 20, as previously suggested (17), and with a MRR ≥ 0.2. Overall, 93.8% of WES variants and 97.8% of WGS variants satisfied the filtering criterion (**Fig. 2A**). We recommend the use of these filters for projects requiring high-quality variants for analyses of WES data. More than half (57.7%) of the WES variants filtered out were present in the flanking 50 bp regions, whereas fewer (37.6%) of the WGS variants filtered out were present in these regions. In addition, 141 filtered WES variants and 70 filtered WGS variants per sample concerned the two base pairs adjacent to the exons, which are key positions for splicing. However, complete removal of the 50 bp flanking regions from the initial calling would result in a large decrease (~90,000) in the number of fully included protein coding exons (**Table S1**). After filtering, the two platforms called an average of 76,195 total SNVs per sample, and the mean proportion of variants for which the same genotype was obtained with both techniques was 99.92% (range: 99.91%-99.93%).

We then studied the high-quality (HQ) variants satisfying the filtering criterion but called by only one platform. On average, 2,734 variants (range: 2,344-2,915) were called by WES but not by WGS (**Fig. 2A**), and 6,841 variants (range: 5,623-7,231) were called by WGS but not WES (**Fig. 2A**). We used Annovar software (18) to annotate these HQ variants as coding variants, i.e., variants overlapping a coding exon, that refers only to coding exonic portion, but not UTR portion. Overall, 651 of the 2,734 WES-exclusive HQ variants and 1,113 of the 6,841 WGS-exclusive HQ variants were coding variants (**Fig. 2A)**. Using the Integrative Genomics Viewer (IGV) tool (19), we noticed that most WES-exclusive HQ variants were also present on the WGS tracks with quality criteria that were above our defined thresholds. We were unable to determine why they were not called by the Unified Genotyper. We therefore used the GATK Haplotype Caller to repeat the calling of SNVs for the WES and WGS experiments. With the same filters, 282 HQ coding variants were called exclusively by WES and 1,014 HQ coding variants were called exclusively by WGS. We combined the results obtained with Unified Genotyper and Haplotype Caller and limited subsequent analyses to the variants called by both callers. The mean number (range) of HQ coding SNVs called exclusively by WES fell to 105 (51-140) per sample, whereas the number called exclusively by WGS was 692 (506-802) (**Fig. 2B**) indicating that calling issues may account for ~80% of initial WES exclusive coding

variants and ~40% of initial WGS exclusive coding variants. The use of a combination of Unified Genotyper and Haplotype Caller therefore appeared to increase the reliability and accuracy of calls. With this combination, we obtained an average of 74,398 HQ SNVs (range: 72,867-77,373) called by both WES and WGS of which 19,222 (18,823-20,024) were coding variants; an average of 1,687 SNVs (range: 1,644-1,749) called by WES only; and 1,915 SNVs (range: 1,687-2,038) called by WGS only (**Fig. 2B**). The quality and distribution of CD, GQ and MRR obtained with this combined calling process were similar to those previously reported for Unified Genotyper (**Fig. S3**).

We further investigated the HQ coding variants called exclusively by one method when a combination of the two callers was used. We were able to separate the variants identified by only one technique into two categories: 1) those called by a single method and not at all by the other, which we refer to as fully exclusive variants, and 2) those called by both methods but filtered out by one method, which we refer to as partly exclusive variants. Of the HQ coding variants identified by WES only (105, on average, per sample), 61% were fully exclusive and 39% were partly exclusive. Of those identified by WGS only (692, on average) 21% were fully exclusive and 79% were partly exclusive. We performed Sanger sequencing on a random selection of 170 fully and partly exclusive WES/WGS variants. Out of 44 fully exclusive WES variants successfully Sanger sequenced, 40 (91%) were absent from the true sequence, indicating that most fully exclusive WES variants were false positives (**Table 1 and Table S3)**. In contrast, 39 (75%) of the 52 Sanger-sequenced fully exclusive WGS variants were found in the sequence, with the same genotype as predicted by WGS (including 2 homozygous), and 13 (25%) were false positives (**Table 1 and Table S3)**. These results are consistent with the observation that only 27.2% of the fully exclusive WES variants were reported in the 1000 genomes database (20), whereas most of the fully exclusive WGS variants (84.7%) were present in this database, with a broad distribution of minor allele frequencies (MAF) (**Fig. S4A**). Similar results were obtained for the partly exclusive variants. Only 10 (48%) of the 21 partly exclusive WES variants (including 3 homozygous) were real, whereas all (100%) of the 24 partly exclusive WGS variants (including 8 homozygous) were real. Using these findings, we estimated the overall numbers of false-positive and false-negative variants detected by these two techniques. WES identified a mean of 26 real coding variants per sample (including 5 homozygous) that were missed by WGS, and a mean of 79 false-positive variants. WGS identified a mean of 656 real coding variants per sample (including 104 homozygous) that were missed by WES, and a mean of 36 false-positive variants.

We noted that most of the false-positive fully exclusive WGS variants were located in the three genes (*ZNF717, OR8U1*, and *SLC25A5*) providing the largest number of exclusive variants on WGS (**Table S4**). Further investigations of the reads corresponding to these variants on the basis of blast experiments strongly suggested that these reads had not been correctly mapped (**Table 2**). Overall, we found that the majority of false positive WGS fully exclusive variants (11/13) and only a minority of false positive WES fully exclusive variants (4/40) could be explained by alignment and mapping mismatches (**Table 2**). We then determined whether the exclusive WES/WGS variants were likely to be deleterious and affect the search for disease-causing lesions. The distribution of combined annotation-dependent depletion (CADD) scores (21) for these variants is shown in **Fig S4B.** About 38.6% of the partly exclusive WES variants and 29.9% of the partly and fully exclusive WGS variants, which were mostly true positives, had a phred CADD score > 10 (i.e. they were among the 10% most deleterious substitutions possible in the human genome), and might include a potential disease-causing lesion. We found that 54.6% of fully exclusive WES variants, most of which were false positives, had a phred CADD score > 10, and could lead to useless investigations. Finally, we investigated whether some genes were particularly poorly covered by WES despite being targeted by the kit we used, by determining, for each sample, the 1,000 genes (approximately 5% of the full set of genes) with the lowest WES coverage (**Fig. S5**). Interestingly, 75.1% of these genes were common to at least four samples (of 6), and 38.4% were present in all six individuals. The percentage of exonic base pairs (bp) with more than 8X coverage for these 384 genes was, on average, 73.2% for WES (range: 0%-86.6%) and 99.5% for WGS (range: 63.6%-100%) (**Table S5**). These genes with low WES coverage in all patients comprised 47 genes underlying Mendelian diseases, including *EWSR1,* the causal gene of Ewing sarcoma, three genes (*IMPDH1, RDH12, NMNAT1*) responsible for Leber congenital amaurosis, and two genes (*IFNGR2, IL12B*) responsible for Mendelian susceptibility to mycobacterial diseases (**Table S5**).

Discussion

These results demonstrate that WGS can detect hundreds of potentially damaging coding variants per sample of which ~16% are homozygous, including some in genes known to be involved in Mendelian diseases, that would have been missed by WES in the regions targeted by the exome kit. In addition to the variants missed by WES in the targeted regions, a large number of genes, protein-coding exons, and non-coding RNA genes were not investigated by

WES despite being fully sequenced by WGS (**Fig. 3**). Finally, mutations outside protein-coding exons, or not in exons at all, might also affect the exome covered by WES, as mutations in the middle of long introns might impair the normal splicing of the exons (22). These mutations would be missed by WES, but would be picked up by WGS (and selected as candidate mutations if the mRNAs were studied in parallel, for example by RNAseq). The principal factors underlying the heterogeneous coverage of WES are probably related to the hybridization/capture and PCR amplification steps required for the preparation of sequencing libraries for WES (23). Here, we clearly confirmed that WGS provides much more uniform distribution of sequencing quality parameters (CD, GQ, MRR) than WES, as recently reported (12). In addition, we performed Sanger sequencing on a large number of variants to obtain a high-resolution estimate of the number of false positives and false negatives in both WES and WGS (**Fig. 3**). We further showed that a number of false-positive results, particularly for the WGS data, probably resulted from mapping problems. We also carried out a detailed characterization of the variants and genes for which the two methods yielded the most different results, providing a useful resource for investigators trying to identify the most appropriate sequencing method for their research projects. Further studies will explore whether similar results are also obtained for other types of variants (e.g. indels, CNVs). We provide open access to all the scripts used to perform this analysis at the software website GITHUB (https://github.com/HGID/WES_vs_WGS). We hope that researchers will find these tools helpful for analyses of data obtained by WES and WGS, two techniques that will continue to revolutionize human genetics and medicine.

Material and Methods

Study subjects:

The six subjects for this study (four females, two males) were collected in the context of a project on Isolated Congenital Asplenia (24). They were all of Caucasian origin (two from USA, and one from Spain, Poland, Croatia, and France), and unrelated. This study was conducted under the oversight of the Rockefeller University IRB. Written consent was obtained from all patients included in this study.

High-throughput Sequencing:

DNA was extracted from the ficoll pellet of 10mL of blood in heparin tubes. Four to six µg of unamplified, high molecular weight, RNase treated genomic DNA was used for WES and WGS. WES and WGS were done at the New York Genome Center (NYGC) using an Illumina HiSeq 2000. WES was performed using the Agilent 71Mb (V4 + UTR) single sample capture. Sequencing was done with 2x100 base-pairs (bps) paired-end reads, and 5 samples per lane were pooled. WGS was performed using the TruSeq DNA prep kit. Sequencing was done with the aim of 30X coverage from 2x100bp paired-end reads.

Analysis of high-throughput sequencing data:

We used the Genome Analysis Software Kit (GATK) best practice pipeline to analyse our WES and WGS data (13). Reads were aligned to the human reference genome (hg19) using the Maximum Exact Matches algorithm in Burrows-Wheeler Aligner (BWA) (25). Local realignment around indels was performed by the GATK (14). PCR duplicates were removed using Picard tools (http://picard.sourceforge.net). The GATK base quality score recalibrator was applied to correct sequencing artefacts. We called our 6 WES simultaneously together with 24 other WES using Unified Genotyper (UG) (14) as recommended by the software to increase the chance that the UG calls variants that are not well supported in individual samples rather than dismiss them as errors. All variants with a Phred-scaled SNP quality ≤ 30 were filtered out. The UG calling process in WGS was similar to that used for WES; we called our 6 WGS together with 20 other WGS. In both WES and WGS, the calling process targeted only regions covered by the WES 71 Mb kit + 50bp flanking each exon (12). When we expanded the WES regions with 100 and 200 bp flanking each exon as performed in some previous studies (26–30), we observed a higher genotype mismatch in variants called by WES and WGS, with a much lower quality of the WES variants located in those additional regions.

Matched and mismatched genotype statistics, analyses of variant coverage depth (CD), i.e. the number of reads passing quality control used to calculate the genotype at a specific site in a specific sample, genotype quality (GQ), i.e. a phred-scaled value representing the confidence that the called genotype is the true genotype, and minor read ratio (MRR), i.e. the ratio of reads for the less covered allele (reference or variant allele) over the total number of reads covering the position where the variant was called, were performed using a homemade R software script (31).

We then filtered out variants with a CD < 8 or GQ < 20 or MRR < 20% a suggested in (17) using a homemade script .We used the Annovar tool (18) to annotate high quality (HQ) variants that were detected exclusively by one method. We checked manually some HQ coding variants detected exclusively by WES or WGS using the Integrative Genomics Viewer (IGV) (19), and we observed that some HQ coding WES exclusive variants, were also present in WGS but miscalled by the UG tool. To recall the UG miscalled SNVs, we used the GATK haplotype caller tool (HC) (14). Indels and SNVs were called simultaneously on 6 WES and 6 WGS, and SNV calls were extracted. The same DP, GQ and MRR filters were applied, and we used Annovar to annotate the HQ resulting variants. All scripts are available on https://github.com/HGID/WES_vs_WGS.

Sanger sequencing:

We randomly selected variants detected exclusively by WES or WGS to test them by Sanger sequencing. We chose more variants in the two categories of WES fully-exclusive and WGS fully-exclusive as we first hypothesized (wrongly) that most, if not all, partly-exclusive variants would be real. We chose less variants in sample S1, as we had few gDNA available for this sample, and we could not test any of the variants in S2 because of absence of remaining gDNA. No other criteria (position, gene, CADD score, frequency) was used for deciding which variants to Sanger sequence. The design of the primers and the sequencing technique are described in **Table S3**.

Analysis of the Sanger sequences was done using the DNASTAR SeqMan Pro software (v11.2.1) using the default settings. To facilitate the localization of the potential variants, we assembled the sequences obtained by Sanger with a 20bp fasta sequence centered on each variant. This sequence was obtained by creating a bed file of the region in the same way as described for the primer design (**Table S3**). Variants where either the forward or reverse sequence did not work were excluded from the analysis and assigned a NA on the Sanger

sequencing results **Table S3**. Sanger sequencing was only attempted once for each variant using the conditions described above.

Acknowledgements

We would like to thank Vincent Barlogis, Carlos Rodriguez Gallego, Jadranka Pac, and Malgorzata Pac for the recruitment of patients, Fabienne Jabot-Hanin, Maya Chrabieh, and Yelena Nemirovskaya for their invaluable help, and the New York Genome Center for conducting WES and WGS. The Laboratory of Human Genetics of Infectious Diseases is supported by grants from the March of Dimes (1-F12-440), National Center for Research Resources and the National Center for Advancing Sciences (NCATS) of the National Institutes of Health (8UL1TR000043), the St. Giles Foundation, the Rockefeller University, INSERM, and Paris Descartes University.

References

1. Ng SB, et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461(7261):272–276.

2. Byun M, et al. (2010) Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *J Exp Med* 207(11):2307–2312.

3. Bolze A, et al. (2010) Whole-exome-sequencing-based discovery of human FADD deficiency. *Am J Hum Genet* 87(6):873–881.

4. Bamshad MJ, et al. (2011) Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 12(11):745–755.

5. Tennessen JA, et al. (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337(6090):64–69.

6. Bolze A, et al. (2013) Ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. *Science* 340(6135):976–978.

7. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. *Cell* 155(1):27–38.

8. Genome of the Netherlands Consortium, Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 46(8):818–825.

9. Weaver JMJ, et al. (2014) Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nat Genet* 46(8):837–843.

10. Clark MJ, et al. (2011) Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol* 29(10):908–914.

11. Saunders CJ, et al. (2012) Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Transl Med* 4(154):154ra135.

12. Meynert AM, Ansari M, FitzPatrick DR, Taylor MS (2014) Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics* 15:247.

13. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.

14. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.

15. Wang JL, et al. (2010) TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain J Neurol* 133(Pt 12):3510–3518.

16. Choi M, et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 106(45):19096–19101.

17. Carson AR, et al. (2014) Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC Bioinformatics* 15:125.

18. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38(16):e164.

19. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14(2):178–192.

20. 1000 Genomes Project Consortium, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65.

21. Kircher M, et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46(3):310–315.

22. Spier I, et al. (2012) Deep intronic APC mutations explain a substantial proportion of patients with familial or early-onset adenomatous polyposis. *Hum Mutat* 33(7):1045–1050.

23. Kebschull JM, Zador AM (2014) Sources of PCR-induced distortions in high-throughput sequencing datasets. *bioRxiv*:008375.

24. Mahlaoui N, et al. (2011) Isolated congenital asplenia: a French nationwide retrospective survey of 20 cases. *J Pediatr* 158(1):142–148, 148.e1.

25. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl* 26(5):589–595.

26. Linderman MD, et al. (2014) Analytical validation of whole exome and whole genome sequencing for clinical applications. *BMC Med Genomics* 7:20.

27. Asan null, et al. (2011) Comprehensive comparison of three commercial human whole-exome capture platforms. *Genome Biol* 12(9):R95.

28. Sulonen A-M, et al. (2011) Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol* 12(9):R94.

29. Wang K, et al. (2011) Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* 43(12):1219–1223.

30. Szpiech ZA, et al. (2013) Long runs of homozygosity are enriched for deleterious variation. *Am J Hum Genet* 93(1):90–102.

31. R Development Core Team R Development Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

32. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.

Table 1: Results of Sanger sequencing for 170 WES and WGS fully and partly exclusive variants

Type of variant	Average # per sample (% homozygous)	# successfully sequenced / Total # sequenced	# (%) of real variants	# (%) of homozygous real variants	Estimated # of real variants*	Estimated # of false positives*
WES						
Fully exclusive	64 (0.5%)	44 / 56	4/44 (9%)	0/4 (0%)	6	58
Partly exclusive	41 (20%)	21 / 27	10/21 (48%)	3/10 (30%)	20	21
Total	105 (8%)	65 / 83	14/65 (22%)		26	79
WGS						
Fully exclusive	145 (6%)	52 / 60	39/52 (75%)†	2/39 (5%)	109	36
Partly exclusive	547 (44%)	24 / 27	24/24 (100%)	8/24 (33%)	547	0
Total	692 (36%)	76 / 87	63/76 (83%)		656	36

* : Estimated numbers of real variants and false positives were computed on the basis of real and false positives proportions applied on the average number of variants per sample
† : 1 real WGS fully exclusive variant was homozygous in Sanger and called heterozygous by WGS

Table 2: Blast results of WES and WGS fully-exclusive false-positive reads.

Origin of false positives *	Variant with reads mapping to a single region †	Variant with reads mapping to more than one region ‡
WES	36 (90%)	4 (10%)
WGS	3 (23.1%)	10 (76.9%)

* : All 40 WES and 13 WGS fully exclusive false-positive variants, according to the Sanger result across the 6 samples (**Table 1** and **Table S3**), were aligned using Blast (32) to the reference genome (hg19).
† : Number of variants with all reads mapping to a single region using Blast with default parameters (the threshold for identifying a mapped region is 80% of identities with the blasted sequence).
‡ : Number of variants with all reads mapping to 1) the initial region assigned by the WES or WGS analysis, and 2) at least another region with a higher alignment score (comprised between 95 and 100% of identities).

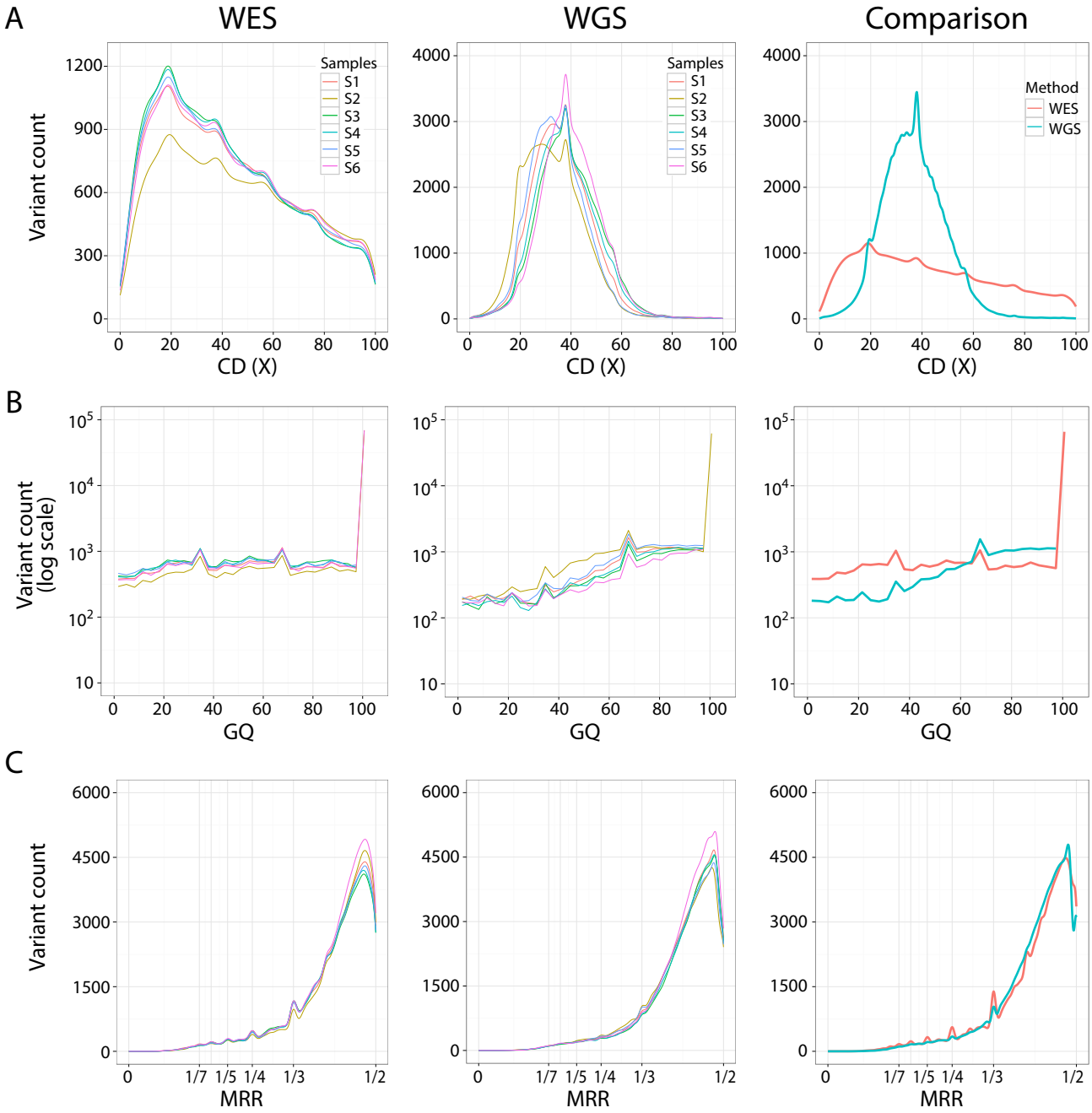

Figure 1: Distribution of the three main quality parameters for the variants detected by WES or WGS: (A) Coverage depth (CD), **(B)** genotype quality (GQ) score, and **(C)** minor read ratio (MRR). For each of the three parameters, we show: the 6 WES samples (left panel), the 6 WGS samples (middle panel), as well as the average over the 6 WES (red) and the 6 WGS (turquoise) samples (right panel).

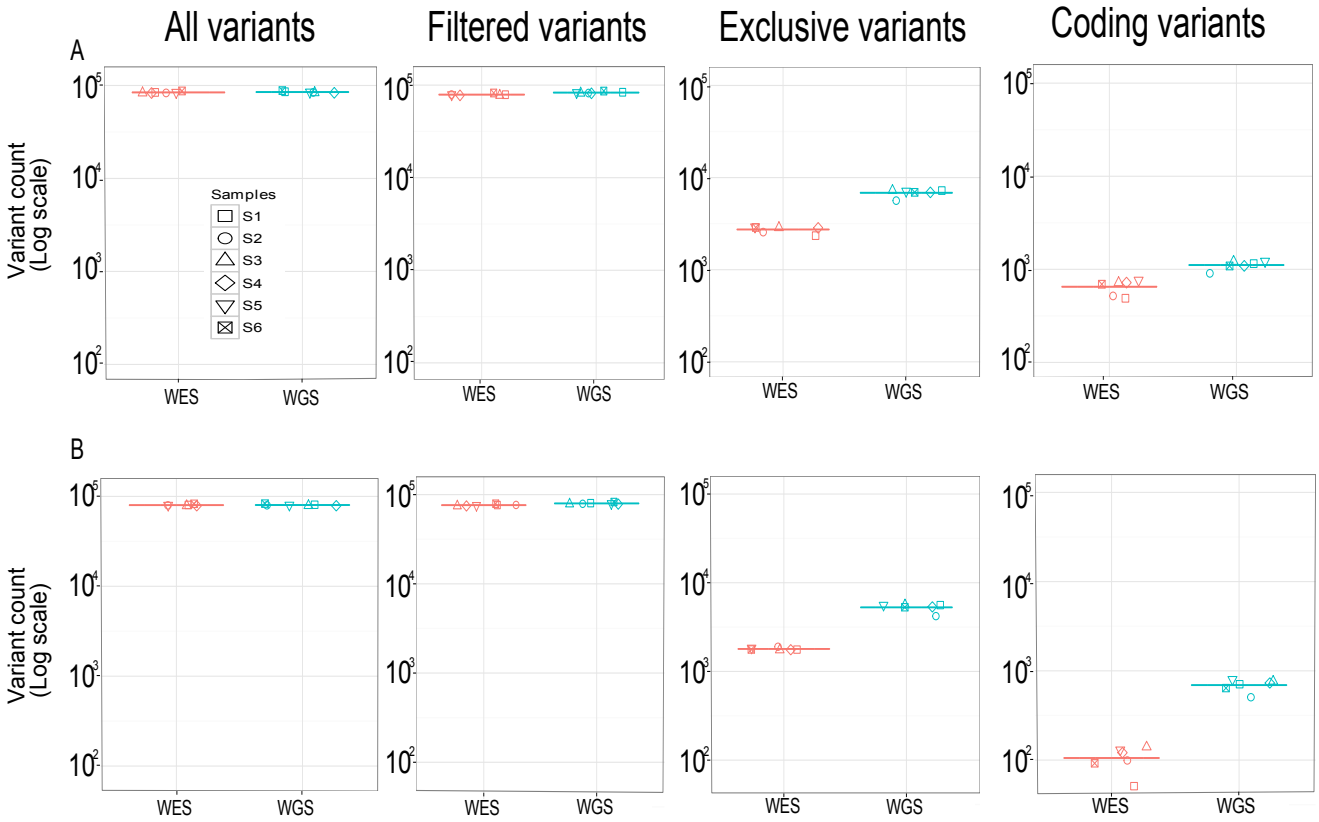

Figure 2: Numbers of SNVs in each WES or WGS sample following the application of various filters called with: (A) Unified Genotyper, and **(B)** the combination of Unified Genotyper and Haplotype Caller (bottom panel). For each of the two calling procedures, we show from left to right: Total number of SNVs called by WES (red) or WGS (turquoise) for each sample; Total number of high-quality SNVs satisfying the filtering criteria: CD ≥ 8X, GQ ≥ 20 and MRR ≥ 0.2 called by WES (red) or WGS (turquoise) for each sample; Number of high-quality SNVs called by only one method, after filtering: high-quality exclusive WES SNVs (red) and high-quality exclusive WGS SNVs (turquoise); Number of exclusive WES (red) and exclusive WGS (turquoise) high-quality coding SNVs.

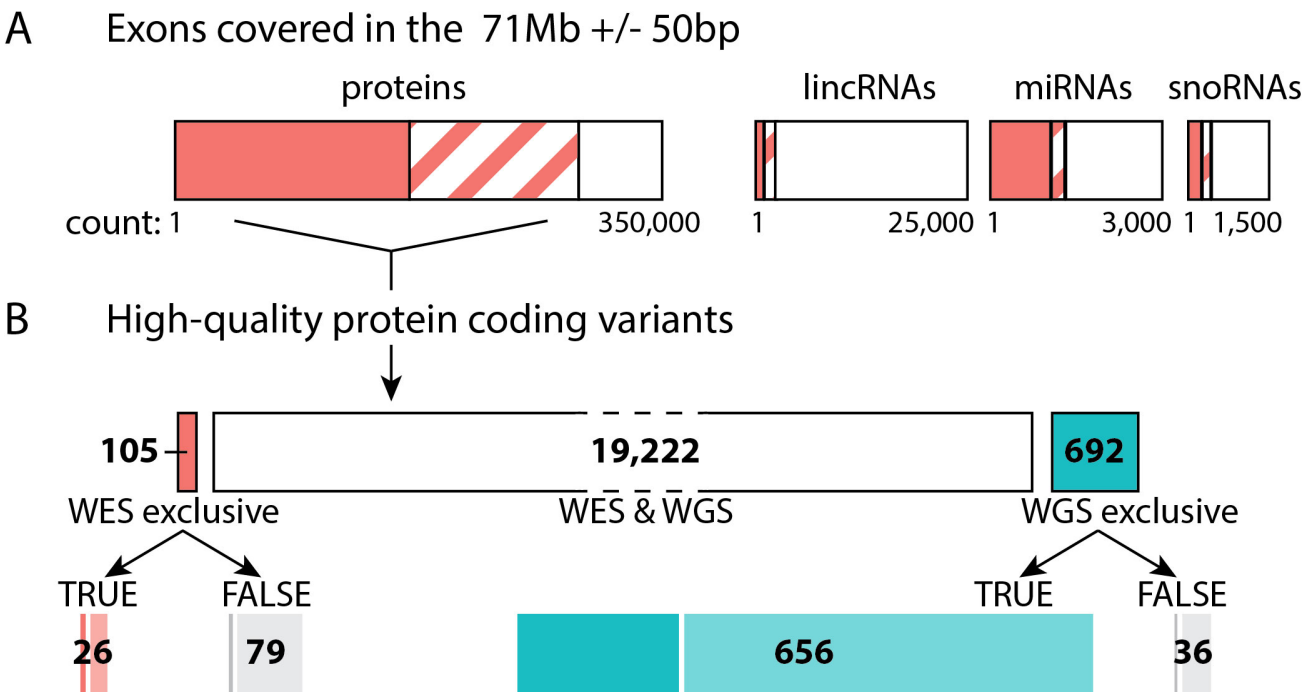

Figure 3: Diagram of the losses at various levels associated with the use of WES. (A) Exons that were covered by the Agilent Sure Select Human All Exon kit 71Mb (V4 + UTR) with the 50bps flanking regions. Exons fully covered are represented by boxes filled entirely in red; exons partly covered by boxes filled with red stripes; and exons not covered at all by white boxes. Numbers are shown in **Table S1**. **(B)** Number of high-quality coding variants called by WES and WGS (white box), by WES exclusively (red box), or by WGS exclusively (turquoise box). Details for the variants called exclusively by one method are provided underneath. TRUE: estimate based on variants detected by Sanger sequencing. FALSE: estimate based on variants that were not detected by Sanger sequencing (**Table 1**). Darker boxes (red, gray, or turquoise) represent homozygous variants. Lighter boxes (red, gray, or turquoise) represent heterozygous variants.

Supporting information

Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants

Aziz Belkadi[a,b,1], Alexandre Bolze[c,f,1], Yuval Itan[c], Quentin B. Vincent[a,b], Alexander Antipenko[c], Bertrand Boisson[c], Jean-Laurent Casanova[a,b,c,d,e,2] and Laurent Abel[a,b,c,2]

[a] Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Paris, France, EU

[b] Paris Descartes University, Imagine Institute, Paris, France, EU

[c] St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, the Rockefeller University, New York, NY, USA

[d] Howard Hughes Medical Institute, New York, NY, USA

[e] Pediatric Hematology-Immunology Unit, Necker Hospital for Sick Children, Paris, France, EU

[f] Present address: Department of Cellular and Molecular Pharmacology, California Institute for Quantitative Biomedical Research, University of California, San Francisco, CA, USA

[1,2] Equal contributions

Corresponding authors: Jean-Laurent Casanova (casanova@rockefeller.edu) or Laurent Abel (laurent.abel@inserm.fr)

Supplementary text:

Sanger sequencing methods

Design of the primers: The first step was to create a bed file with each row representing a region of 400bp centered on the variants chosen for Sanger sequencing. The bed file was then uploaded in the UCSC genome browser using the 'add custom tracks' tab. The reference genome assembly used was GRCh37/hg19 (https://genome.ucsc.edu/cgi-bin/hgGateway). Fasta files with the sequence for each region were then downloaded from the UCSC website, and uploaded to BatchPrimer3 v1.0 (http://batchprimer3.bioinformatics.ucdavis.edu/cgi-bin/batchprimer3/batchprimer3.cgi) (1). We noticed that BatchPrimer3 worked better if the fasta files were copied and pasted rather uploaded using a link. We then requested for Sequencing primers using the following parameters: nb of return = 1 (1 towards 3', and 1 towards 5'); sequencing start = -1; primer size: Min = 18, Opt = 22, Max = 25; primer Tm: Min = 55, Opt = 58, Max = 62; Max self complementarity = 8; Max 3' self complementarity = 3. Lastly, variants for which one of the two primers was closer to 60bp to the variant were excluded from further sequencing and analysis. M13F or M13R sequences were added at the 5'-end of the forward or reverse primers. The full list of primers ordered is available at Table S3.

Sequencing of the variants: Amplification of the variants was performed using per reaction: H2O=11.5uL, 40% glycerol=4.5uL, 10X buffer (Denville without MgCl2)=2.25uL, MgCL2 (25mM)=0.9uL, dNTP (10mM)=0.225uL, primers (10uM)=0.5uL each, Taq Polymerase (Denville, #CB4050-2)=0.5uL, DNA=50-100ng. DNA was substituted by H2O in negative controls. 38 cycles of 94C (30''), 60C (30''), 72C (1') were performed on a Veriti Thermal Cycler (Life Technologies). Sequencing PCR was done using the Big Dye 1.1 (Life Technologies) protocol with 1 uL of amplification PCR product and either the M13F or the M13R primer on a Veriti Thermal Cycler (Life Technologies). Lastly the samples were sequenced on a ABI 3730 XL sequencer (Life Technologies).

Supplementary material:

5 supplementary figures

5 supplementary tables

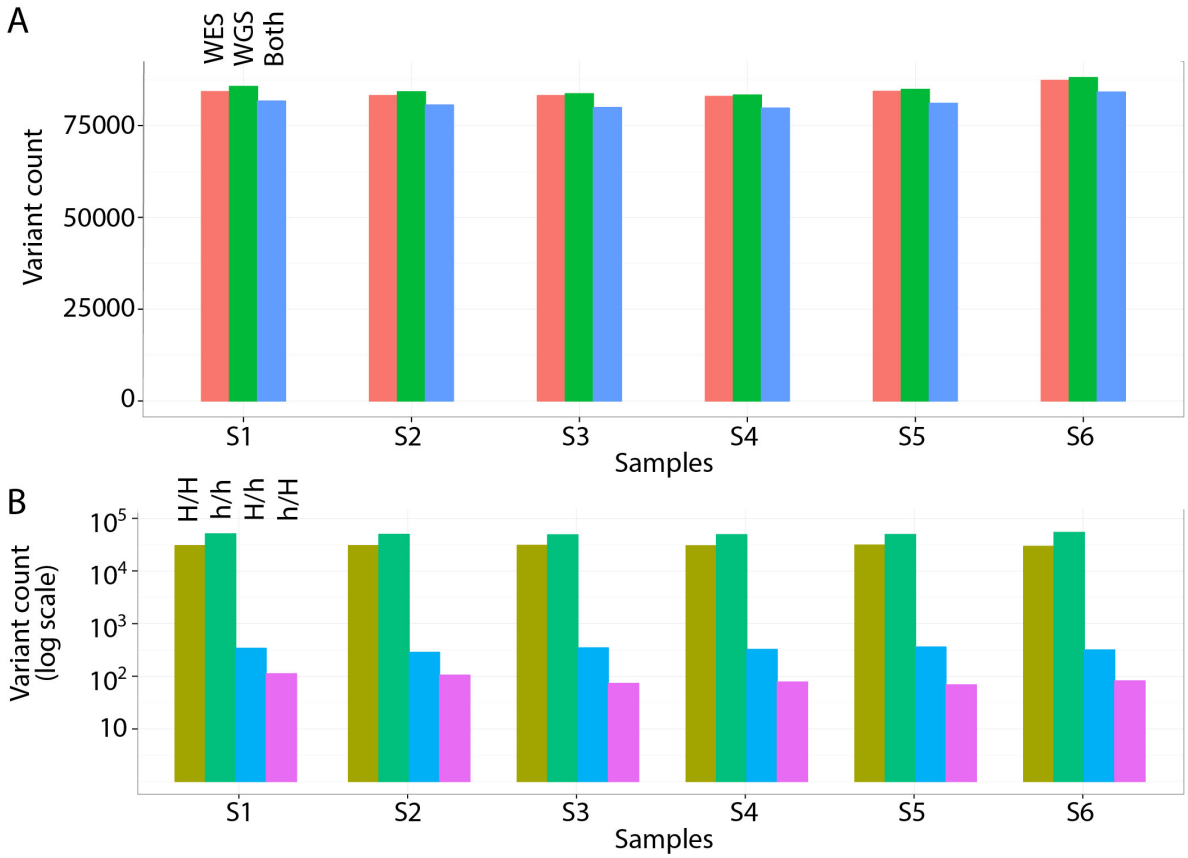

Figure S1: Number and general characteristics of single-nucleotide variants (SNVs) called by WES and WGS. (A) Total number of SNVs called by WES alone, WGS alone, and both platforms. **(B)** Characteristics of the SNVs called by both WES and WGS for each sample with four columns indicating the number of SNVs called homozygous by both methods (H/H, light green), called heterozygous by both methods (h/h, dark green), called homozygous by WES and heterozygous by WGS (H/h, blue), called heterozygous by WES and homozygous by WGS (h/H, purple)

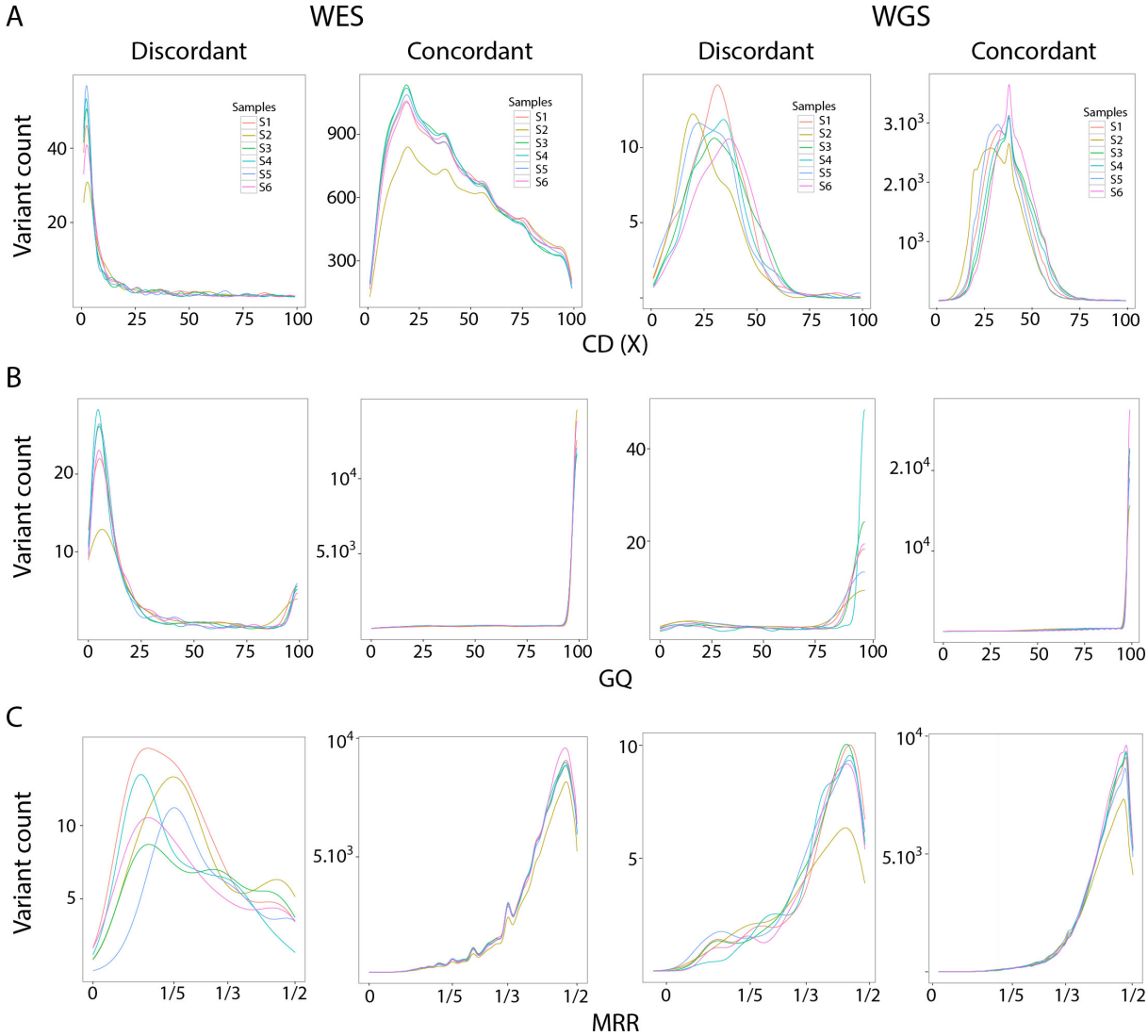

Figure S2: Distribution of the three main quality parameters for the variants with genotypes discordant between WES and WGS. (A) Coverage depth (CD), **(B)** genotype quality (GQ) score, and **(C)** minor read ratio (MRR). For each of the three parameters, four panels are shown: the two panels on the left show the characteristics of discordant and concordant SNVs in WES samples; the two panels on the right shown the characteristics of discordant and concordant SNVs in WGS samples.

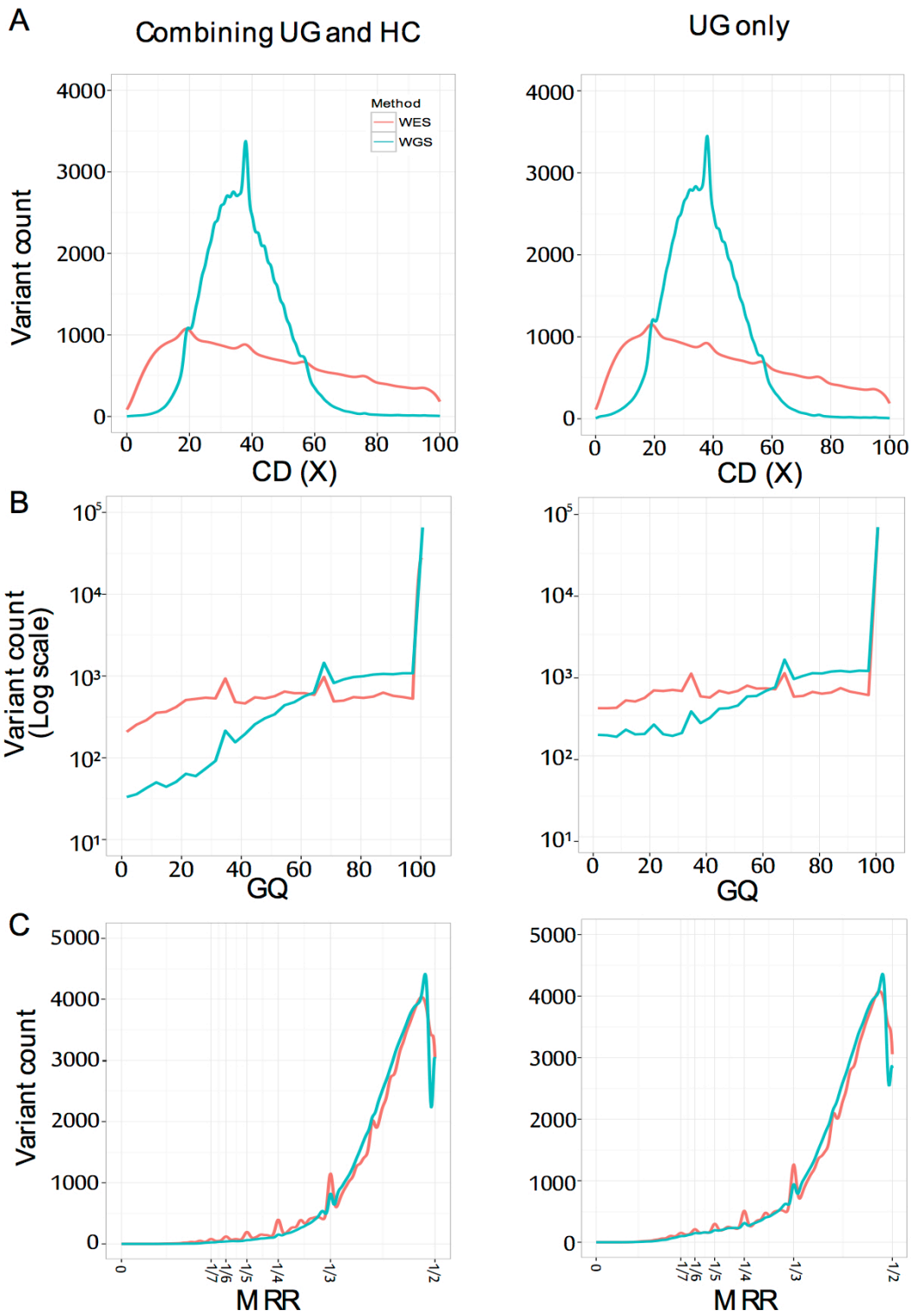

Figure S3: Comparison of the distribution of the three main quality parameters for the variants detected by WES or WGS, with either the combination of Unified Genotyper and Haplotype Caller, or with Unified Genotyper alone. (A) Coverage depth (CD), **(B)** genotype quality (GQ) score, and **(C)** minor read ratio (MRR). For each of the three parameters we show: the average over the 6 WES (red) and the 6 WGS (turquoise) samples for the combination of callers (left panel), and for Unified Genotyper alone (right panel).

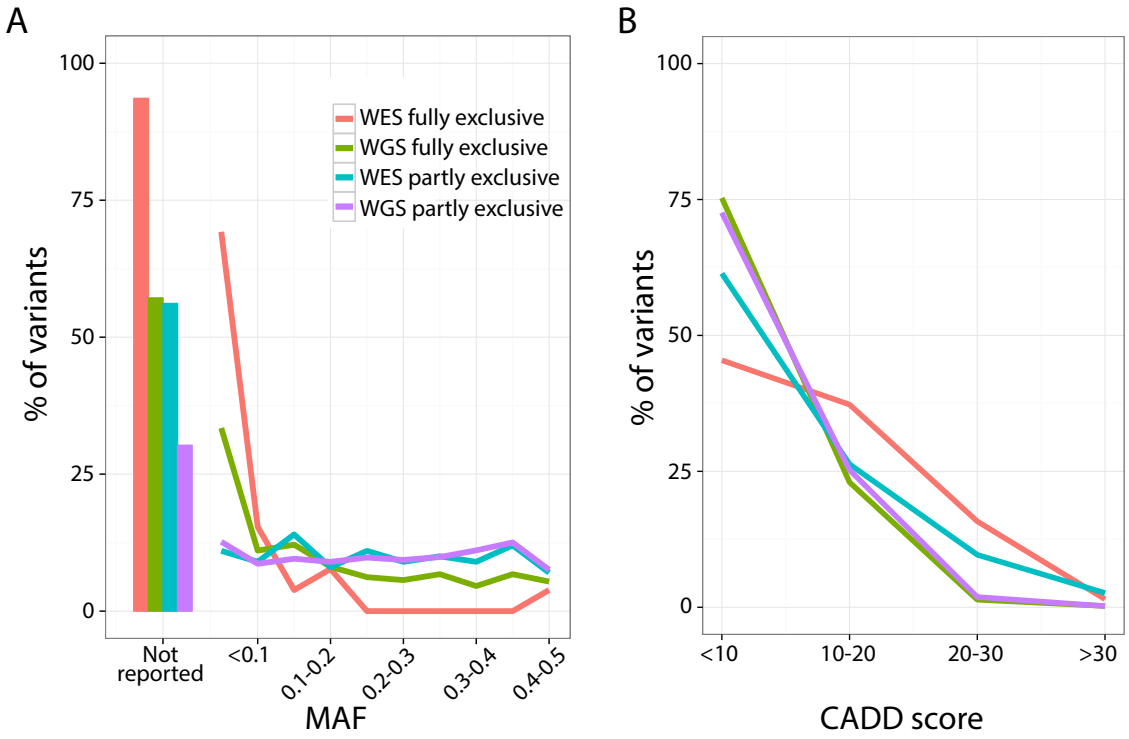

Figure S4: Distribution of high-quality coding SNVs identified exclusively by one technique according to: (A) their presence in the 1000 Genomes database, and their reported minor allele frequency (MAF) for those present in this database, **(B)** their CADD (combined annotation-dependent depletion) scores. Red: Fully exclusive high-quality WES coding SNVs, never identified by WGS. Turquoise: Partly exclusive high-quality WES coding SNVs, identified by WGS but filtered out due to their poor quality. Green: Fully exclusive high-quality WES coding SNVs, never called by WES. Purple: Partly exclusive high-quality WGS coding SNVs, identified by WES but filtered out due to their poor quality.

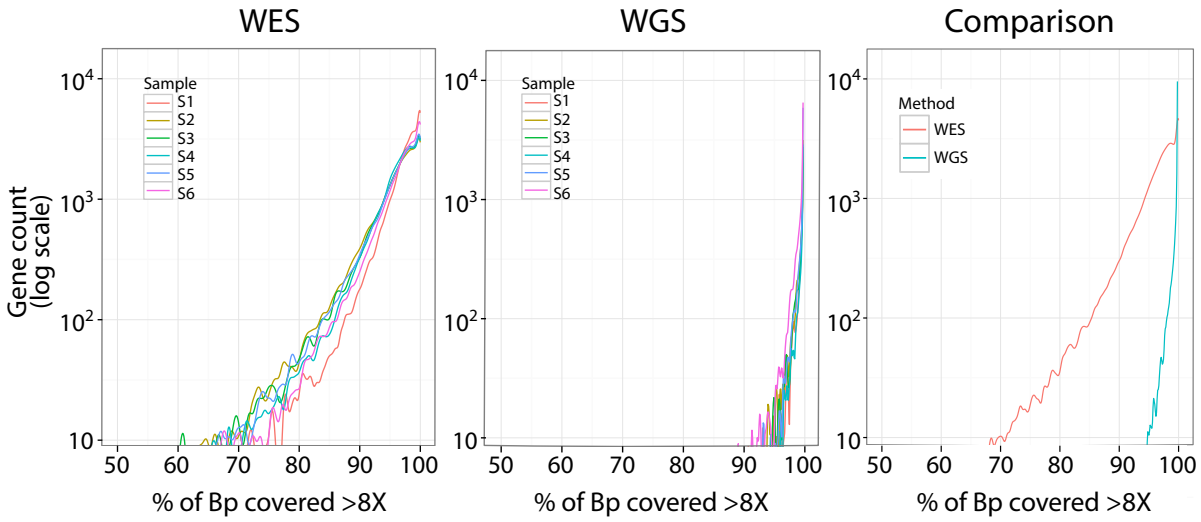

Figure S5: Distribution of the percentage of base pairs per gene with less than 8X coverage, for all genes in WES and WGS. *Y*-axis: number of genes (log-scale). *X*-axis: Percentage of base pairs for a given gene with at least 8X coverage. The figure shows data for the 6 WES samples (left panel), the 6 WGS samples (middle panel), and the average over the 6 WES (red), and the 6 WGS (turquoise) samples (right panel).

Table S1: Specific regions of the genome covered by WES using the 71Mb kit.

	Protein coding exons		lincRNA		MiRNA		snoRNA	
	71Mb	71Mb +/- 50Bps	71Mb	71Mb +/- 50Bps	71Mb	71Mb +/- 50Bps	71Mb	71Mb +/- 50Bps
Fully included	88,722	180,830	387	554	713	1,171	169	252
Partially included	219,328	129,946	965	855	508	94	130	93
Fully excluded	67,647	64,921	25,446	25,389	1,826	1,782	1,157	1,111
Total	375,697	375,697	26,798	26,798	3,047	3,047	1,456	1,456

Four types of genomic units were analyzed: protein-coding exons, miRNA exons, snoRNA exons, and lincRNA exons as defined in Ensembl Biomart (2) . We determined the number of these units using the R Biomart package (3) on the GRCh37/hg19 reference. For the counts, we excluded one of the duplicated units of the same type, or units entirely included in other units of the same type (only the longest unit would be counted in this case). We then determined the number of the remaining units that were fully or partly covered when considering the genomic regions defined by the Agilent Sure Select Human All Exon kit 71Mb (v4 + UTR) with or without the 50 bps flanking regions.

Table S2: Reads and coverage statistics for each WES and each WGS.

Sample	Total number of WES reads	Total number of WGS reads	Number of WES reads aligned in WES regions +/- 50 bps	Number of WGS reads aligned in WES regions +/- 50 bps	WES mean coverage in WES regions +/- 50 bps	WSG mean coverage in WES regions +/- 50 bps
S1	98,792,738	1,370,493,918	64,696,895	34,737,193	72.1	38.7
S2	124,483,242	1,303,868,290	80,970,674	31,743,245	90.3	35.3
S3	86,822,862	1,477,715,120	57,970,027	37,322,280	64.5	41.5
S4	89,521,104	1,438,287,290	59,084,117	36,600,011	65.9	40.7
S5	98,002,162	1,301,586,284	62,673,065	33,102,614	69.9	36.8
S6	100,056,600	1,445,702,068	68,002,983	37,619,386	75.8	41.9
Mean	99,613,118	1,389,608,828	65,566,294	35,187,455	73.1	39.2

Table S3: Sanger sequencing results.

Gene	Chr	Start	Ref	Obs	Genotype	Method	Sample	Sanger result	forward primer	reverse primer
CYP26B1	2	72359518	A	G	het	WES fully exclusive	S1	HET	TGTAAAACGACGGCCAGTGTG GGTCTTGGGTTAGACTGT	CAGGAAACAGCTATGACCGTAT AGCATCCGGGACACC
RSPH10B	7	5997562	G	A	het	WES fully exclusive	S1	NA	TGTAAAACGACGGCCAGTGCA GTGAGCCAAGATTGC	CAGGAAACAGCTATGACCATTT CTTCAAAGGAGCTCAAGG
TAS2R19	12	11174277	T	C	het	WES fully exclusive	S1	HET	TGTAAAACGACGGCCAGTTGC ACACATATACACCCATAAA	CAGGAAACAGCTATGACCCTTC CTCATGTTATTTGCCATT
ADAMTS18	16	77334230	T	G	het	WES fully exclusive	S1	WT	TGTAAAACGACGGCCAGTTCT CATAAAAGACAGTTCTTGGG	CAGGAAACAGCTATGACCCCAA TGTTAAGGTCAAAATGTCA
FAM209A	20	55100005	T	C	het	WES fully exclusive	S1	HET	TGTAAAACGACGGCCAGTAAA CCCGTCATGAGCAACT	CAGGAAACAGCTATGACCACTC ACTAGAACATCCGTTTCC
SRMS	20	62173927	G	A	het	WES fully exclusive	S1	WT	TGTAAAACGACGGCCAGTCTT GAGGGTTGGACAGCA	CAGGAAACAGCTATGACCCAGA GCAATGAGCTCCCA
RRP7A	22	42910165	C	T	het	WES fully exclusive	S1	NA	TGTAAAACGACGGCCAGTCCT CCTCGACCACCAAGT	CAGGAAACAGCTATGACCGATG GGATCACCTTCCTTG
CXCR7	2	237489904	C	T	het	WES partly exclusive	S1	HET	TGTAAAACGACGGCCAGTACA GCATCAAGGAGTGGCT	CAGGAAACAGCTATGACCCATC AGCTCGTACCTGTAGTTG
MYRIP	3	40251392	T	C	het	WES partly exclusive	S1	NA	TGTAAAACGACGGCCAGTGAA GAGAAAGCAGACCAGGTAA	CAGGAAACAGCTATGACCTTAC CTTCTTCAGCTCTTCCTG
SYNE1	6	152529260	G	A	het	WES partly exclusive	S1	NA	TGTAAAACGACGGCCAGTGCC TAAGAGGTGTGAGAACACT	CAGGAAACAGCTATGACCGATC ACTTCTCAGGGCTTAGG
EN2	7	155251433	C	T	het	WES partly exclusive	S1	HET	TGTAAAACGACGGCCAGTAAC TTCTTCATCGACAACATCC	CAGGAAACAGCTATGACCAGCG AGAGCGTCTTGGAG
MFSD3	8	145735026	T	G	het	WES partly exclusive	S1	WT	TGTAAAACGACGGCCAGTCCA AGGTTCTGTACGCTCC	CAGGAAACAGCTATGACCAGG AGCAGAAAGAGTTGCG
CES1	16	55862717	T	C	het	WES partly exclusive	S1	WT	TGTAAAACGACGGCCAGTACT CCAGAATGCTGTGAGAGTT	CAGGAAACAGCTATGACCATTT ATTCTCCATGTCCAGCAG
CXorf40A	X	148628490	A	T	hom	WES partly exclusive	S1	HOM	TGTAAAACGACGGCCAGTCAA TGCCCCGAAGACTTAAC	CAGGAAACAGCTATGACCCTGA GCAAAGGAACCTGTTTAC
AIM1L	1	26664968	C	T	het	WGS partly exclusive	S1	HET	TGTAAAACGACGGCCAGTACC AGCTACTTGGGACCAG	CAGGAAACAGCTATGACCCAGC TGCTGTGTGAAATTAGAG
ADCY2	5	7802363	C	T	het	WGS partly exclusive	S1	HET	TGTAAAACGACGGCCAGTGGC AAGTGGAGTAGGCATTT	CAGGAAACAGCTATGACCAGGC CACTATCCTGAAGTAAC
SOHLH1	9	138590928	C	T	hom	WGS partly exclusive	S1	HOM	TGTAAAACGACGGCCAGTCAG CCCCGAACATAATCTC	CAGGAAACAGCTATGACCTCCC TACGTGACCCAGTCT
OR8U1	11	56143716	T	C	het	WGS partly exclusive	S1	NA	TGTAAAACGACGGCCAGTCAT TCAACTTGTAGCAGTTCCTTA	CAGGAAACAGCTATGACCCTTG TCTGTGTCCAGGGC
SPATA5L1	15	45695382	G	A	het	WGS partly exclusive	S1	HET	TGTAAAACGACGGCCAGTCTG GGAGGTCTTTCGGAG	CAGGAAACAGCTATGACCGACA CAAGGCGTCCATCTC
GPX4	19	1106615	T	C	het	WGS partly exclusive	S1	HET	TGTAAAACGACGGCCAGTTAC GGACCCATGGAGGAG	CAGGAAACAGCTATGACCCAGA AAGATCCAGCAGGCTA
CDKL5	X	18638082	A	C	het	WGS partly exclusive	S1	HET	TGTAAAACGACGGCCAGTGGA ACCTAGTGTCATGCATTTT	CAGGAAACAGCTATGACCTAGA AAAGGCTCTGTTGAGAGG
C1orf94	1	34667784	A	C	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTATC CCTAAGGAAGTTGCGAT	CAGGAAACAGCTATGACCGGA AGGGATTCAGAGGAGTCTA
HMCN1	1	186052030	T	G	het	WES fully exclusive	S3	NA	TGTAAAACGACGGCCAGTCCT AATAAAAGCTAGCATCAGCA	CAGGAAACAGCTATGACCGGG GATTGAATGAGTATAGGCT
DYSF	2	71791292	T	G	het	WES fully exclusive	S3	NA	TGTAAAACGACGGCCAGTCTG GTGTGTCACCATCCC	CAGGAAACAGCTATGACCAGAC CTCTTCTCCTTCCAAGAC
ZSWIM2	2	187692949	A	T	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTTGA GACACAGGCTGTCTTGATA	CAGGAAACAGCTATGACCACAT TTTCCCAGGTATCTTCAA
USP49	6	41774685	C	G	het	WES fully exclusive	S3	NA	TGTAAAACGACGGCCAGTAGG TAACAGAACACGTAGAGATCC	CAGGAAACAGCTATGACCGGA GTTGAAAATGAATGAATCTA
MACC1	7	20198700	G	T	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTCCA CTTGAACACAAAAATCAAA	CAGGAAACAGCTATGACCTTGG GATTATATCCACAAAACC
ADCY8	8	131964235	C	G	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTGAG AGCACCCAAACACACAT	CAGGAAACAGCTATGACCAGA GTGCCTGGCAAATAATAAG
OR52B2	11	6190994	C	G	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTAAC ATAAGGATGACACAGAGGTG	CAGGAAACAGCTATGACCTTTG TGCCCCACTGAGATATAC
CAPN5	11	76796027	T	C	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTTTC ACAGGGCACATCAGG	CAGGAAACAGCTATGACCCACC CTCACTTTCTCAGCAG

RASAL1	12	113543517	A	C	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTGTGCCTGTCCATGTCCTG	CAGGAAACAGCTATGACCCTCTCTTCTCCCATCTCCTAGA
FMN1	15	33192236	G	T	het	WES fully exclusive	S3	NA	TGTAAAACGACGGCCAGTATATAAATGTTGTTAAGGGGAGGA	CAGGAAACAGCTATGACCTCCCGACAGCCTATTGAGTA
CES1	16	55862762	C	G	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTTCTTAAGGAGTCCAGAGCAAAG	CAGGAAACAGCTATGACCAAACTCCACCTGGAATCTGG
AKAP1	17	55184422	A	C	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTCAGTGAAGAGTTGCCGGA	CAGGAAACAGCTATGACCGACTGGCAGCCTTTCTCC
MUC16	19	9067022	T	G	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTGGTGTCTTCATCTGTTGTCAGT	CAGGAAACAGCTATGACCTCTACATCACAGGGCACATTTA
FKRP	19	47259734	G	C	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTGCTGCAACAAGGAGACCA	CAGGAAACAGCTATGACCGTACTGCACGCGGAAAAA
SGK2	20	42204913	A	C	het	WES fully exclusive	S3	WT	TGTAAAACGACGGCCAGTCTGTCTCTTTCCAGTCTGCC	CAGGAAACAGCTATGACCGTGTTAATGTGCTTCTGAGCTG
PLOD1	1	12010469	G	T	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTTCCATTTCCCAGATGGTG	CAGGAAACAGCTATGACCAGATTGCACGTCAAACAAGG
PLA2R1	2	160889514	G	A	het	WGS fully exclusive	S3	HOM	TGTAAAACGACGGCCAGTGTGAGAGTTTTGGGCCATATTA	CAGGAAACAGCTATGACCAAGACCTGGTTGTTTTTAATGG
ZNF717	3	75786202	T	C	het	WGS fully exclusive	S3	WT	TGTAAAACGACGGCCAGTGAGGTGTAGGTTGTGTGTTCAA	CAGGAAACAGCTATGACCTTTACGATAAGACAGTTCTCACCA
ZNF717	3	75786516	G	T	het	WGS fully exclusive	S3	NA	TGTAAAACGACGGCCAGTGCTTCTCACCTGTGTGAGTTCT	CAGGAAACAGCTATGACCATACATCAGAGAACTCACACCG
ATP13A4	3	193183940	T	C	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTGCCAACATGCACAGTACAAA	CAGGAAACAGCTATGACCCCGTTCCAGCATTTATGTATTT
ADCY2	5	7802363	C	T	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTGGCAAGTGGAGTAGGCATTT	CAGGAAACAGCTATGACCAGGCCACTATCCTGAAGTAAC
POMZP3	7	76240888	A	G	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTATGACAGCAGGTACCCTCAA	CAGGAAACAGCTATGACCCCAGATGAACTCAACAAGGC
TRAPPC9	8	140743340	G	T	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTAAAGATGCTACAGGAGGAACAG	CAGGAAACAGCTATGACCGATTCCTGGTGGCTTTGG
PTPLA	10	17659265	G	C	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTCGATGTCGTAGAAGGTGAGC	CAGGAAACAGCTATGACCGGTCGGTAGAGCTGGCTG
TMEM80	11	695842	G	A	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTACGGACTAATCGGGCCTC	CAGGAAACAGCTATGACCGCTTCTCGATGGGGTGAC
OR8U1	11	56143803	A	G	het	WGS fully exclusive	S3	WT	TGTAAAACGACGGCCAGTCCAACATTGTCAACCATTTCTA	CAGGAAACAGCTATGACCTCTTTCACCTCCTTATTCTGGA
CCDC88B	11	64124515	T	C	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTGGACATACCTGAGAACAGCATT	CAGGAAACAGCTATGACCACCGTGGAGGATCTCAGG
HECTD4	12	112601517	C	T	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTGATGTCTACCTTGAGGAACTCG	CAGGAAACAGCTATGACCGAAAGGACTGGGATGACCA
PLEKHH1	14	68024134	A	T	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTTGTGAGTGATGGGAAGACACTA	CAGGAAACAGCTATGACCCTGGCTTCTAATGAGCAGATGT
IRX3	16	54317628	G	A	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTAGGAGGACTGGTTTTATTTCTTTT	CAGGAAACAGCTATGACCTACAGTTAAACCCCAACACACA
MRC2	17	60769803	A	G	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTCTGGTGGTGGTGCTGATG	CAGGAAACAGCTATGACCAAGGGCACCCTTCCATAG
BPIFB4	20	31671663	T	C	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTGGAGAAATCCCACCTGGA	CAGGAAACAGCTATGACCCAAGACCCAAACCATGTAACTT
NEFH	22	29876587	C	A	het	WGS fully exclusive	S3	HET	TGTAAAACGACGGCCAGTCTGGACACGCTGAGCAAC	CAGGAAACAGCTATGACCCTCCAGGCGTAGCTGACC
ARSD	X	2833631	A	G	het	WGS fully exclusive	S3	NA	TGTAAAACGACGGCCAGTTCCCAAAGTGCTGGGATTA	CAGGAAACAGCTATGACCTGTGAATAGTGCTGGAGTGAAC
SLC25A5	X	118603929	C	T	het	WGS fully exclusive	S3	WT	TGTAAAACGACGGCCAGTATGTCATCAGATACTTCCCCAC	CAGGAAACAGCTATGACCAACTTACCCTTTGCAGTGTCAT
SPATA21	1	16730309	T	G	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTCTTTCACTTGTGACTAAAAGTCGT	CAGGAAACAGCTATGACCCTGTGATGACAGACACCAGG
KPRP	1	152732950	A	C	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTAGACCCAGGGCTCCTATG	CAGGAAACAGCTATGACCGGAGGAATCTCAACAGGACAC
CTNNB1	3	41278119	C	A	het	WES fully exclusive	S4	NA	TGTAAAACGACGGCCAGTAAGCTATTGAAGCTGAGGGAG	CAGGAAACAGCTATGACCGGAAACATCAATGCAAATGAA
FRYL	4	48559517	G	T	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTGAAAGATATTTGTTTGGTTATCACA	CAGGAAACAGCTATGACCATCCAGACAGCTCACCCTG
PGM3	6	83892687	C	A	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTCAAGATAATTTGTTCAGTAGACCA	CAGGAAACAGCTATGACCTAATGATTGGTTTTTTGGCTTC
ATP6V1C1	8	104078558	G	T	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTTTGAACTTGTAAAGGTAAAGGGAG	CAGGAAACAGCTATGACCTTCTTTCAATCATTTTTTTTCTGA
ATRNL1	10	117075090	T	G	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTTGCATTAACTATAGATGACCTTTCA	CAGGAAACAGCTATGACCCCTTAAGCAGAAACTGAAATTGTT

HECTD4	12	112605691	A	C	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTTTTCTGAAACGGTTTGGCT	CAGGAAACAGCTATGACCCTGGGGTGGCTTCTTTCTA
GEMIN2	14	39601190	G	T	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTTGAGGCTTTCTTGTCTATACCC	CAGGAAACAGCTATGACCCCAATAAAATATTCCATGTGTTTTC
CATSPER2	15	43924422	T	C	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTCAGAATGTGACATACCAAACCA	CAGGAAACAGCTATGACCACAAATCTAGACGTGCTTTTCTG
MYLK3	16	46744689	C	A	het	WES fully exclusive	S4	NA	TGTAAAACGACGGCCAGTCATGAGTGACAAGCAATGAAAG	CAGGAAACAGCTATGACCCTTCCTCCCTTTAATGAACACA
CDC37	19	10506724	G	C	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTGTCCTGGTTGCAGGCTCT	CAGGAAACAGCTATGACCACGACTCCCCAGAGTTGATAG
PLK1S1	20	21143043	G	A	het	WES fully exclusive	S4	WT	TGTAAAACGACGGCCAGTACTCATTGCTTGGAGATAGGAA	CAGGAAACAGCTATGACCCATAAGATCACTACCACCCAGAA
TMEM88B	1	1361530	C	T	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTGTGGCTCTGGGACAGACAT	CAGGAAACAGCTATGACCAGGAGCACCAGCAGGAAG
RNF19B	1	33430102	T	G	het	WGS fully exclusive	S4	NA	TGTAAAACGACGGCCAGTACAGCGGACACTCCACCT	CAGGAAACAGCTATGACCGGGCTCCGAGAAGGACTC
TMEM87B	2	112813190	G	C	het	WGS fully exclusive	S4	WT	TGTAAAACGACGGCCAGTGTTTCCCAGAACTGCACG	CAGGAAACAGCTATGACCGGTCCCGACACTCCACTTA
BOC	3	113004240	C	T	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTGTGTGGTACCTCTTGATGTTCA	CAGGAAACAGCTATGACCCTCCTGGAACCAACCTGAG
SRD5A1	5	6651970	A	G	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTAAATCAAAATCCACTTTTAGCTTAG	CAGGAAACAGCTATGACCAAAGCAATGATGTGAACAAGG
FBXW11	5	171295669	G	C	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTTGAACTCTGCAAAAGTTGACAC	CAGGAAACAGCTATGACCGTGAGATATCAGGGGCTGTAAA
KCTD7	7	66098384	G	A	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTGGATTGAAGATGGAGCAGC	CAGGAAACAGCTATGACCTTGATCTCTTTCAATAAACCCATT
SNTB1	8	121824063	C	A	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTCAGAACGAGCCATTGGTG	CAGGAAACAGCTATGACCGACGCACTCTCCTCGCTC
AQP7	9	33385712	G	A	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTCACCCCTCAACACACAGG	CAGGAAACAGCTATGACCCACAGCATCTGCTCCTCAG
OR8U1	11	56143795	G	A	het	WGS fully exclusive	S4	WT	TGTAAAACGACGGCCAGTCCAACATTGTCAACCATTTCTA	CAGGAAACAGCTATGACCACCTCCTTATTCTGGAGGCTAT
TREH	11	118529127	G	A	het	WGS fully exclusive	S4	NA	TGTAAAACGACGGCCAGTGAACTGGTGCAGAGGTTTAATG	CAGGAAACAGCTATGACCTCAGTGTGCTCACCTGCAT
LRRC16B	14	24534337	C	A	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTGTTGCTAACTTACCCCGATTC	CAGGAAACAGCTATGACCAGGAAAAGGGGAAGACACAG
GALK2	15	49620200	C	T	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTGTCCTAAAATGTTTGATGACACC	CAGGAAACAGCTATGACCAAGTGCCCTAAGTAGTTTCTCTCA
ADCY9	16	4165432	T	C	hom	WGS fully exclusive	S4	HOM	TGTAAAACGACGGCCAGTGCAGCTAGAGGAGATGCTGTAT	CAGGAAACAGCTATGACCAACCACAGGAACAGATGGTG
C17orf96	17	36830108	T	G	het	WGS fully exclusive	S4	WT	TGTAAAACGACGGCCAGTACTCGGAGTGTCCAAGGC	CAGGAAACAGCTATGACCAATCTACGACCAGCTTCGC
LRRC45	17	79983379	C	T	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTGTGCATTCTGTCTGGTGACTAC	CAGGAAACAGCTATGACCGACAGTGCCCATGTGTGG
MED16	19	875395	C	T	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTATCTTGGTGCAGATCTCGGT	CAGGAAACAGCTATGACCGTCAGAGTCGAACTGCTCTTCT
GIPC1	19	14590236	C	T	het	WGS fully exclusive	S4	NA	TGTAAAACGACGGCCAGTCCAGCTACTTGGGAGGCT	CAGGAAACAGCTATGACCAAAGCCAGGAAGGACAAGTT
CCDC61	19	46518651	A	G	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTGTCTGGCCAAGGAGGTGA	CAGGAAACAGCTATGACCCTTAGGCTCCGCCTCATC
HELZ2	20	62190641	G	A	het	WGS fully exclusive	S4	HET	TGTAAAACGACGGCCAGTCTCCAAGTCCACCCACTTC	CAGGAAACAGCTATGACCCACCTGACCCTGACTGACTC
IRF6	1	209961970	C	G	het	WES fully exclusive	S5	WT	TGTAAAACGACGGCCAGTCTCTCCTGGGTTTGAAGGAT	CAGGAAACAGCTATGACCCAGAAGGATGGTCCAGAGAGAT
SNRK	3	43389767	G	T	het	WES fully exclusive	S5	WT	TGTAAAACGACGGCCAGTCCCACCAATACATCGGGTA	CAGGAAACAGCTATGACCGTAGCTGCAGCACGTTATTTTT
PIM1	6	37139029	C	G	het	WES fully exclusive	S5	NA	TGTAAAACGACGGCCAGTATGAGTGGGTGGGGTGAG	CAGGAAACAGCTATGACCCCGAAGTCGATGAGCTTG
STK3	8	99719384	A	C	het	WES fully exclusive	S5	WT	TGTAAAACGACGGCCAGTCAAATTTGGCTCAATTATGGTT	CAGGAAACAGCTATGACCCGTGGCATTTTAATTATGGTTT
DERA	12	16109969	T	G	het	WES fully exclusive	S5	WT	TGTAAAACGACGGCCAGTCCTTTCAAGGACCATGTAAAAAT	CAGGAAACAGCTATGACCGGATAAATGTGTTATCTTTCTCCAA
ELMSAN1	14	74194213	T	G	het	WES fully exclusive	S5	WT	TGTAAAACGACGGCCAGTCCACATACAGAAGCTCAAGGA	CAGGAAACAGCTATGACCGTTTTCGTAGGTGACAGGCT
CES1	16	55862791	T	C	het	WES fully exclusive	S5	WT	TGTAAAACGACGGCCAGTGACTGCCTTGACTCCTTCCT	CAGGAAACAGCTATGACCAAGGTCACTCACTTAGAAAGCG
TBX21	17	45820022	A	C	het	WES fully exclusive	S5	WT	TGTAAAACGACGGCCAGTAAACTCCCTAAACACCTTCCAG	CAGGAAACAGCTATGACCTCTAGGAATTAGGGGTAGGGG
CATSPERG	19	38851455	A	C	het	WES fully	S5	WT	TGTAAAACGACGGCCAGTCCT	CAGGAAACAGCTATGACCCTCC

						exclusive			CTTCTACGAAGACAGCAAA	TCTGAGCTTCCATAAGTG
STARD8	X	67940201	G	C	het	WES fully exclusive	S5	WT	TGTAAAACGACGGCCAGTCACCCCACCTGATCCTCT	CAGGAAACAGCTATGACCGGAAGGCCAGAGCAGTTC
PDE4DIP	1	144921924	G	A	het	WES partly exclusive	S5	NA	TGTAAAACGACGGCCAGTATTATGCAACTGACTCAAGGGT	CAGGAAACAGCTATGACCTTAGTCTTTGTGGGAGCTCAGT
WDR6	3	49049501	T	G	het	WES partly exclusive	S5	WT	TGTAAAACGACGGCCAGTATGTCTGACTGGATTTGGGAT	CAGGAAACAGCTATGACCCCCACCTTCCAGATACGAA
TBCK	4	107168386	T	G	het	WES partly exclusive	S5	WT	TGTAAAACGACGGCCAGTTTGTTGATAAGTTCAAAACTGAAAG	CAGGAAACAGCTATGACCAGACTCTGCAAAAGAGAGCTGTA
HLA-DRB5	6	32489786	T	G	hom	WES partly exclusive	S5	NA	TGTAAAACGACGGCCAGTCACACACACTCAGATTCCCA	CAGGAAACAGCTATGACCGACCGGATCCTTCGTGTC
GPRIN2	10	46999863	C	G	het	WES partly exclusive	S5	HET	TGTAAAACGACGGCCAGTGCGTCAGTGAGCGAGTCT	CAGGAAACAGCTATGACCATGTCATGCCCTCAGCATC
MUC6	11	1016928	C	G	het	WES partly exclusive	S5	NA	TGTAAAACGACGGCCAGTTTGGAGTCACCAAGGAGGT	CAGGAAACAGCTATGACCAATGACACCGACCACCAGT
IL32	16	3119304	A	G	het	WES partly exclusive	S5	WT	TGTAAAACGACGGCCAGTCAAGGTCATGAGATGGTTCC	CAGGAAACAGCTATGACCACAGCACCAGGTCAGAGC
RAD51C	17	56774108	T	G	het	WES partly exclusive	S5	WT	TGTAAAACGACGGCCAGTTAGACATTTCTGTTGCCTTGG	CAGGAAACAGCTATGACCAATGGAGTGTTGCTGAGGTCT
SIRPA	20	1895796	T	C	het	WES partly exclusive	S5	WT	TGTAAAACGACGGCCAGTGGTCAAATGAGATGATACATGC	CAGGAAACAGCTATGACCTGGAAAAGTCCATGTTGTTTCT
SGSM1	22	25272644	G	C	het	WES partly exclusive	S5	WT	TGTAAAACGACGGCCAGTTTGCTCTAGGGTGAGATTTCTG	CAGGAAACAGCTATGACCATTTCATGGCCAGGATTTAAC
KANSL3	2	97271090	G	A	het	WGS fully exclusive	S5	HET	TGTAAAACGACGGCCAGTACTCATGCCAACTTTACCCA	CAGGAAACAGCTATGACCATTGTGGAGGATCTCAACTCAG
IL17RB	3	53892830	T	C	het	WGS fully exclusive	S5	HET	TGTAAAACGACGGCCAGTCCAGAAAGAAGGGAAGTTTTG	CAGGAAACAGCTATGACCTCAGATTCTAGGTTCTCTGGGA
ZNF717	3	75787221	C	T	het	WGS fully exclusive	S5	NA	TGTAAAACGACGGCCAGTCAGTGAAAGGATTTTCCACATT	CAGGAAACAGCTATGACCTGAGTGTGGAAAACCCTTTATC
ZNF717	3	75788130	C	T	het	WGS fully exclusive	S5	WT	TGTAAAACGACGGCCAGTTGTGTGTGTCTGCTGATGTTTA	CAGGAAACAGCTATGACCAACAGTTCAGGAATGAAGCCT
COL19A1	6	70851789	A	G	het	WGS fully exclusive	S5	HET	TGTAAAACGACGGCCAGTTCATGTTTTAGAATGAACTCTCCTT	CAGGAAACAGCTATGACCTATACCTTTAGTCCTGGGCTTC
C11orf16	11	8953721	T	C	het	WGS fully exclusive	S5	HET	TGTAAAACGACGGCCAGTGTGACAGACCCCACACAGATA	CAGGAAACAGCTATGACCCTCAGGTAATGGTGGTGCCTAT
C1QTNF9B	13	24468329	A	G	het	WGS fully exclusive	S5	NA	TGTAAAACGACGGCCAGTCCCATCTGGAGAGTAAGAACTG	CAGGAAACAGCTATGACCAGCTCAGCACCCCAGATG
NDUFA7	19	8376431	G	A	het	WGS fully exclusive	S5	HET	TGTAAAACGACGGCCAGTGGAAACATGGTGAGACTCTGT	CAGGAAACAGCTATGACCCTGGAACACCCTGCTGTCT
CST7	20	24939590	G	C	het	WGS fully exclusive	S5	HET	TGTAAAACGACGGCCAGTGAAGCATTGCCCCAAGAT	CAGGAAACAGCTATGACCGTTAGAGACGTGGTGACGGT
SLC25A5	X	118604428	T	C	het	WGS fully exclusive	S5	WT	TGTAAAACGACGGCCAGTCCTTGTGTACAGATGACGTGTT	CAGGAAACAGCTATGACCCAGTTGTGGAACAGACACAGAT
FBLIM1	1	16096934	C	T	hom	WGS partly exclusive	S5	HOM	TGTAAAACGACGGCCAGTGATTCCTTTTTAATGCTCCTCA	CAGGAAACAGCTATGACCTCTAAGTGCTCAGCTCACTGC
SYN2	3	12046215	G	C	hom	WGS partly exclusive	S5	NA	TGTAAAACGACGGCCAGTCAGATGATGAACTTCCTGCG	CAGGAAACAGCTATGACCCGTCTGCTTTACCGCTTG
CLDN24	4	184242959	C	G	hom	WGS partly exclusive	S5	HOM	TGTAAAACGACGGCCAGTGATTTTAGAGGGAAGTGGGTCT	CAGGAAACAGCTATGACCACAAGACGGTTCAGGAGTTCT
PDZD2	5	32087253	A	G	het	WGS partly exclusive	S5	HET	TGTAAAACGACGGCCAGTATTACAAGCATGCGCCAC	CAGGAAACAGCTATGACCGAGCCTGACTGGAGACCTG
HOXA4	7	27169934	A	G	hom	WGS partly exclusive	S5	NA	TGTAAAACGACGGCCAGTGCTGACATGGATCTTCTTCATC	CAGGAAACAGCTATGACCTACCCCTATGGCTACCGC
GRK5	10	121196335	G	A	het	WGS partly exclusive	S5	HET	TGTAAAACGACGGCCAGTATGGCACTGTTCTTGTGCTC	CAGGAAACAGCTATGACCAGTCTGTCTGACTCTGCATCCT
USP28	11	113670052	T	A	hom	WGS partly exclusive	S5	HOM	TGTAAAACGACGGCCAGTCTAATCCTTTTCCCAAGGTGA	CAGGAAACAGCTATGACCGACCTTTGAGGTTAGGTAAGGG
ITGA5	12	54799450	A	G	hom	WGS partly exclusive	S5	HOM	TGTAAAACGACGGCCAGTGATCATCAGCTCTCAGCTCTTT	CAGGAAACAGCTATGACCGATACCCCTCAACCCCAC
PRIMA1	14	94245649	A	G	het	WGS partly exclusive	S5	HET	TGTAAAACGACGGCCAGTGGCCTAGGAAAACACAAAGAG	CAGGAAACAGCTATGACCACAACATTGTCCCCTTTGAA
TTLL13	15	90794102	G	A	het	WGS partly exclusive	S5	HET	TGTAAAACGACGGCCAGTTGAGGAAAAGGAATCTGAGAAG	CAGGAAACAGCTATGACCTGGTTCTGAATTTTGTTTCTGTT
ATAD3A	1	1452566	G	A	het	WES fully exclusive	S6	HET	TGTAAAACGACGGCCAGTCGGTCCACTCAGCAGGAT	CAGGAAACAGCTATGACCGGTCTTCCTCCTCTCCTCAG
ACVR2A	2	148676144	A	C	het	WES fully exclusive	S6	WT	TGTAAAACGACGGCCAGTACATATGGCCTTTGTCAAGAAC	CAGGAAACAGCTATGACCAAAATACTTCCTGGCCAATCTC
OTUD4	4	146071820	G	T	het	WES fully	S6	NA	TGTAAAACGACGGCCAGTTTA	CAGGAAACAGCTATGACCAGTG

						exclusive			CCTTATGATCTGTGAAGGTGTC	TCAGGGAAGAAGATGAAA
FOXK1	7	4801940	A	C	het	WES fully exclusive	S6	WT	TGTAAAACGACGGCCAGTTATAGGGGACTTGAAAAAAGCA	CAGGAAACAGCTATGACCCAGGTGACCTCACTCCCC
CRTAC1	10	99770893	A	C	het	WES fully exclusive	S6	WT	TGTAAAACGACGGCCAGTCTGCAGTAGCAAAAGACAAGGT	CAGGAAACAGCTATGACCAGGATGTTACCGTTCCTGCT
AQP2	12	50344816	A	C	het	WES fully exclusive	S6	WT	TGTAAAACGACGGCCAGTCTCCATAGCCTTCTCCAGG	CAGGAAACAGCTATGACCGATGGCAAAGTTGTGGCTACT
ERN2	16	23718102	T	G	het	WES fully exclusive	S6	WT	TGTAAAACGACGGCCAGTAGCTCTATTCCTGGCTCCTAGT	CAGGAAACAGCTATGACCAGCAGAGGCAGGGATCTAAG
RAD51C	17	56774108	T	G	het	WES fully exclusive	S6	NA	TGTAAAACGACGGCCAGTTAGACATTTCTGTTGCCTTGG	CAGGAAACAGCTATGACCAATGGAGTGTTGCTGAGGTCT
HIPK4	19	40895487	A	G	het	WES fully exclusive	S6	WT	TGTAAAACGACGGCCAGTGGGAAAAAGACAAGGAACTAGG	CAGGAAACAGCTATGACCCAAGAATGACGCCTACCG
LILRB2	19	54780769	G	C	het	WES fully exclusive	S6	NA	TGTAAAACGACGGCCAGTCCAGTGGTTTGGATTCTCTTT	CAGGAAACAGCTATGACCTCTGAGCGTCAGTTTTTCATC
FLG	1	152281007	A	G	het	WES partly exclusive	S6	NA	TGTAAAACGACGGCCAGTGGGAGGCATCAGACCTTC	CAGGAAACAGCTATGACCACACAGTCAGTGTCAGCACAG
FANCD2	3	10088404	C	T	het	WES partly exclusive	S6	WT	TGTAAAACGACGGCCAGTTTAACTGTTTTTCTGTTGTTGCAT	CAGGAAACAGCTATGACCTAAATAGGATACGGAAGGCCA
TRIP6	7	100468284	A	G	het	WES partly exclusive	S6	HET	TGTAAAACGACGGCCAGTGGAGGCTGGGAGACAGAG	CAGGAAACAGCTATGACCCTTTTAGCACCGTTCCTCCT
OR1L6	9	125512770	T	C	hom	WES partly exclusive	S6	HOM	TGTAAAACGACGGCCAGTCTCCCACCTACATTCCCTGT	CAGGAAACAGCTATGACCGTACATAACTGTGGCTACCCG
PPYR1	10	47086915	C	T	het	WES partly exclusive	S6	HET	TGTAAAACGACGGCCAGTCCCTCAAGTGTATCACTTAGTTCA	CAGGAAACAGCTATGACCAGTAGTCCATGATGGTGTAGACG
SLC22A12	11	64367862	T	C	het	WES partly exclusive	S6	HET	TGTAAAACGACGGCCAGTAGCAGATTGTGGGTGTGG	CAGGAAACAGCTATGACCATGCATGACATGAACATCTAGG
SKA3	13	21750538	G	A	het	WES partly exclusive	S6	WT	TGTAAAACGACGGCCAGTGTGGGACATACCGTCCACT	CAGGAAACAGCTATGACCCGAGATTCAAACTAGTGGCG
OR4N4	15	22383064	C	A	het	WES partly exclusive	S6	HET	TGTAAAACGACGGCCAGTTGTTCAACTGTCATGAACCCTA	CAGGAAACAGCTATGACCAAGGGCACATGTAGATGAAGAT
KCNJ12	17	21319079	C	A	het	WES partly exclusive	S6	WT	TGTAAAACGACGGCCAGTGGTACATGCTGCTCATCTTCTC	CAGGAAACAGCTATGACCACCAATCATGAAGGAGTCGAT
CXorf40A	X	148628490	A	T	hom	WES partly exclusive	S6	HOM	TGTAAAACGACGGCCAGTCAATGCCCCGAAGACTTAAC	CAGGAAACAGCTATGACCCTGAGCAAAGGAACCTGTTTAC
LOC440563	1	13183115	G	A	het	WGS fully exclusive	S6	WT	TGTAAAACGACGGCCAGTAAAATTTGTTGTTAGACAAGCTCC	CAGGAAACAGCTATGACCCCCAGATAAAACAGAAAGTGGA
SIPA1L2	1	232539219	C	T	het	WGS fully exclusive	S6	HET	TGTAAAACGACGGCCAGTAAGTAGTCCCACTCAGTCCCTT	CAGGAAACAGCTATGACCTTTAGCTATTGCATTTCCACAA
ZNF717	3	75786620	G	A	het	WGS fully exclusive	S6	WT	TGTAAAACGACGGCCAGTTTTCTCTCCTGAGTGAGTCCC	CAGGAAACAGCTATGACCGAAAAACCTTTCATCGCAAGT
ZNF717	3	75788192	T	C	het	WGS fully exclusive	S6	NA	TGTAAAACGACGGCCAGTCCTACCTGAGTTATCACTTGGAC	CAGGAAACAGCTATGACCTTTAATTTGAACTCAAACCATGT
GET4	7	930689	C	T	het	WGS fully exclusive	S6	HET	TGTAAAACGACGGCCAGTCCCCTTTCCTTTTCTGTGTTAT	CAGGAAACAGCTATGACCTTATGAAAAATCATGGGTCAGG
OR8U1	11	56143819	G	C	het	WGS fully exclusive	S6	WT	TGTAAAACGACGGCCAGTTCTATTGTGATGACATGCCTCT	CAGGAAACAGCTATGACCTCTTCAGAGCTTCTTTCACCTC
FRY	13	32776616	T	A	het	WGS fully exclusive	S6	HET	TGTAAAACGACGGCCAGTTGCTCATGAGATATCCAGCTAA	CAGGAAACAGCTATGACCCGTGCCTGGTCATAACTCTAA
TICRR	15	90168410	A	C	het	WGS fully exclusive	S6	WT	TGTAAAACGACGGCCAGTACCTATGAGGTTGAGCTGGAG	CAGGAAACAGCTATGACCCTGGGCCAGTCTTTAATTATGT
FSD1	19	4322990	G	A	het	WGS fully exclusive	S6	HET	TGTAAAACGACGGCCAGTATAGCTGGGAACCTGAGGAGTA	CAGGAAACAGCTATGACCCAGCACCTTGACCTTGTTG
SLC25A5	X	118604409	C	T	het	WGS fully exclusive	S6	WT	TGTAAAACGACGGCCAGTGAAGCCAAGATCATCCAATG	CAGGAAACAGCTATGACCAACAGACACAGATGCTATCAACC
DTX2	7	76121509	C	T	het	WGS partly exclusive	S6	HET	TGTAAAACGACGGCCAGTAGGAAAACAAAACCAAAGGC	CAGGAAACAGCTATGACCAAAGAGGCACTGCTCCCC
SOHLH1	9	138586966	G	A	hom	WGS partly exclusive	S6	HOM	TGTAAAACGACGGCCAGTCTTCCAGATGCCGAGAAAG	CAGGAAACAGCTATGACCCATCTGACTTCTCTCCCAGAAC
LRP4	11	46898771	T	C	het	WGS partly exclusive	S6	HET	TGTAAAACGACGGCCAGTTCTCACAACCAAAGAGAGAGTG	CAGGAAACAGCTATGACCATGAGTTTCAGTTTGCCTGATT
CHGA	14	93397655	C	T	het	WGS partly exclusive	S6	HET	TGTAAAACGACGGCCAGTTAACCCTAATCGTTGTCCTGG	CAGGAAACAGCTATGACCCTGTGGGCCTGGGTATTT
HYDIN	16	70883822	T	C	het	WGS partly exclusive	S6	HET	TGTAAAACGACGGCCAGTCCTTGATTATGAGTTCCAGGTC	CAGGAAACAGCTATGACCTCCTGCTAGAATATCTGACTCCA
MYOM1	18	3067278	A	G	het	WGS partly exclusive	S6	HET	TGTAAAACGACGGCCAGTAAAGTGTCATTAGTTGGTGCTTTT	CAGGAAACAGCTATGACCCTCAGACGACCACTGCAAC
TMEM86B	19	55739689	G	A	het	WGS partly	S6	HET	TGTAAAACGACGGCCAGTCTG	CAGGAAACAGCTATGACCAGAT

						exclusive			GGGCCTCTCTCACAC	CTGAGTCCCAAGAATGG
SOGA1	20	35491551	A	G	hom	WGS partly exclusive	S6	HOM	TGTAAAACGACGGCCAGTGAC ACCTCCGAGCTGCTAT	CAGGAAACAGCTATGACCCCGG AGAGGAAAAAGAGC
SLC16A8	22	38477930	G	A	het	WGS partly exclusive	S6	HET	TGTAAAACGACGGCCAGTACT TCGAAGACTGTCCCTCATA	CAGGAAACAGCTATGACCCGGA GGTGACCTTATTCCTTA
ATP11C	X	138897130	A	C	hom	WGS partly exclusive	S6	HOM	TGTAAAACGACGGCCAGTCAC TTTAAAATGGTGTATTTTTACC	CAGGAAACAGCTATGACCTGAA AGTGTGTCTCAGATTTGC

Table S4: Genes carrying at least two variants called exclusively by WES and at least 3 variants called exclusively by WGS.

WES			WGS		
Gene	Samples carrying variants	Number of variants	Gene	Samples carrying variants	Number of variants
HLA-DRB1	3	7	ZNF717	5	54
CES1	4	6	OR8U1	6	16
PDE4DIP	4	6	SLC25A5	4	15
SIRPB1	5	5	SYN2	6	11
ADAM21	2	5	MUC5B	5	11
MUC6	2	4	AQP7	6	9
CEP170	4	3	TAS2R43	3	9
APOBEC3H	3	3	CROCC	6	8
GPRIN2	3	3	HLA-DRB1	4	8
ZNF717	3	3	GRIN3B	6	7
HLA-DQA2	2	3	OR51A2	5	7
KCNJ12	2	3	TPSD1	3	7
PLEC	2	3	LONRF2	6	6
SIRPA	2	3	FLJ43860	5	6
HLA-A	1	3	HLA-C	4	6
MUC20	1	3	GRID2IP	6	5
OR9G1	1	3	IDUA	6	5
TAS2R43	1	3	LOC440563	6	5
FAT3	4	2	PRODH	5	5
HECTD4	4	2	SELO	5	5
MYLK3	4	2	HEG1	3	5
ACSM5	3	2	TAS2R19	3	5
CLIP1	3	2	ARSD	2	5
DZANK1	3	2	FBRSL1	6	4
IL31RA	3	2	KRT83	6	4
IL32	3	2	SAC3D1	6	4
PNKP	3	2	TREH	6	4
PPYR1	3	2	ANKRD24	5	4
SF3B3	3	2	CPAMD8	5	4
TNC	3	2	FAM131C	5	4
GBP7	2	2	ZNF598	5	4
KPRP	2	2	C2CD2	4	4
MLL3	2	2	PLCL2	4	4
PCMTD1	2	2	SEC22B	4	4
PKHD1L1	2	2	TMEM88B	4	4
SPANXD	2	2	CPZ	3	4
ZSWIM2	2	2	HLA-A	3	4
CAPN5	1	2	LRRN4	3	4

CATSPER2	1	2	MAP2K3	3	4
CFHR1	1	2	OBSCN	3	4
HLA-DRB5	1	2	TTLL1	3	4
SBSN	1	2	IER5	2	4
TMEM128	6	1	LAMA5	2	4
CXorf40A	5	1	PABPC3	2	4
ACVR2A	4	1	SYTL1	1	4
BBS4	4	1	BAIAP2L2	6	3
CCNA1	4	1	COL4A1	6	3
CCNE1	4	1	PI4K2B	6	3
CTNNB1	4	1	PKD1L2	6	3
LGALS3	4	1	SNX19	6	3
MFSD3	4	1	SPRN	6	3
NCF4	4	1	WTIP	6	3
NOTCH1	4	1	ALDH4A1	5	3
OR52B2	4	1	ANKS6	5	3
PGM3	4	1	CIT	5	3
RAD51C	4	1	COL22A1	5	3
RHPN2	4	1	GRIN2D	5	3
SGK2	4	1	KALRN	5	3
ATF7IP	3	1	NAV2	5	3
ATRNL1	3	1	NBPF3	5	3
BCOR	3	1	TG	5	3
C19orf44	3	1	ZAN	5	3
DDX18	3	1	DPP3	4	3
DYSF	3	1	FRY	4	3
FAM135A	3	1	HLA-DRB5	4	3
FMN1	3	1	HMMR	4	3
FRMD4A	3	1	KIAA1211	4	3
FRYL	3	1	MRS2	4	3
HIPK4	3	1	PCNT	4	3
MAP2K3	3	1	PKD1L1	4	3
MDGA2	3	1	SOHLH1	4	3
MITF	3	1	TMEM158	4	3
MUC16	3	1	BAHCC1	3	3
NKX2-8	3	1	C8orf73	3	3
OTUD4	3	1	CCDC57	3	3
OXCT2	3	1	CYP2A7	3	3
SELRC1	3	1	DRD4	3	3
SLC35E2	3	1	FHOD3	3	3
SLC5A12	3	1	GAB4	3	3
SPTA1	3	1	LILRB3	3	3
TICRR	3	1	LOC653486	3	3
UBE2D1	3	1	LRP8	3	3
USH2A	3	1	MED16	3	3

USP49	3	1	MUC12	3	3
WFDC1	3	1	PDE4DIP	3	3
ADCY8	2	1	PILRB	3	3
AKAP1	2	1	SLC16A8	3	3
ALPPL2	2	1	SORT1	3	3
AQP2	2	1	TMED8	3	3
ATP6V1A	2	1	TMEM44	3	3
BCL9	2	1	CCDC61	2	3
C1orf94	2	1	CD200R1	2	3
C2CD3	2	1	CD24	2	3
C5orf60	2	1	GPR31	2	3
CACNA1S	2	1	HLA-DQB1	2	3
CATSPERG	2	1	LGALS8	2	3
CCAR1	2	1	MMP20	2	3
CHRNA4	2	1	OR2T4	2	3
CNTD1	2	1	PEX6	2	3
CRTAC1	2	1	PIEZO1	2	3
CYB561	2	1	PPP1R37	2	3
DERA	2	1	PRR5	2	3
DOCK5	2	1	TCF3	2	3
FAM13A	2	1	TMEM86B	2	3
FAT2	2	1	TRIM50	2	3
FKRP	2	1	UHRF1	2	3
FOXK1	2	1	ZFPM1	2	3
GCGR	2	1	CACNA1B	1	3
GEMIN2	2	1	CHGA	1	3
GPATCH8	2	1	GABBR1	1	3
HMCN1	2	1	MUC20	1	3
HSF4	2	1	ZNF700	1	3
ISYNA1	2	1	AMH	6	2
KCNH6	2	1	ATP10B	6	2
LAT	2	1	BCLAF1	6	2
LCN12	2	1	CCT5	6	2
LDLRAD3	2	1	CERS1	6	2
LYZL2	2	1	DNAH17	6	2
MACC1	2	1	EMR1	6	2
MICAL2	2	1	LOC100507462	6	2
NGLY1	2	1	OTUD7A	6	2
OR1L4	2	1	WDR86	6	2
OR1L6	2	1	YBX2	6	2
PARD3B	2	1	ABCC6	5	2
PHACTR3	2	1	ANKLE1	5	2
PIM1	2	1	ANKRD36	5	2
PLA2R1	2	1	ARHGEF10L	5	2
PLXNA1	2	1	ATXN2	5	2

POU2F2	2	1
RASAL1	2	1
RRP7A	2	1
SF3B5	2	1
SKA3	2	1
SNRK	2	1
SOGA3	2	1
SPATA21	2	1
SVIL	2	1
TBX21	2	1
TFAM	2	1
TNRC6A	2	1
TPST2	2	1
TRAPPC12	2	1
TSHR	2	1
ZNF527	2	1
ZPLD1	2	1

C2orf72	5	2
CABP5	5	2
CCDC175	5	2
CCDC33	5	2
CTDP1	5	2
EXD3	5	2
FBP1	5	2
FPGS	5	2
HCN2	5	2
MFAP2	5	2
NUGGC	5	2
PPP1R3G	5	2
SPHK1	5	2
SYNM	5	2
TMEM221	5	2
TRIM22	5	2
WNK2	5	2
ADAM11	4	2
AGRN	4	2
AHNAK2	4	2
ATP11A	4	2
CARD14	4	2
ENPP7	4	2
GFRA4	4	2
HOXA4	4	2
HSPG2	4	2
HYDIN	4	2
INCENP	4	2
IRF2BP2	4	2
KIAA0284	4	2
LCN15	4	2
MEF2D	4	2
MIDN	4	2
MYO1C	4	2
PANX2	4	2
PLXND1	4	2
POM121C	4	2
PRIMA1	4	2
SYNE3	4	2
TBKBP1	4	2
THEM4	4	2
TMTC1	4	2
TSPAN11	4	2
UBR4	4	2
AGBL1	3	2

ALK	3	2
ARID3A	3	2
ASTL	3	2
ATG2A	3	2
C6orf10	3	2
C9orf96	3	2
CAPN14	3	2
CAPN9	3	2
CCDC90A	3	2
CNTN5	3	2
CTSF	3	2
FAM174B	3	2
FAM59B	3	2
FBXW8	3	2
FCGBP	3	2
FOXD1	3	2
GATA5	3	2
HEATR1	3	2
HID1	3	2
HIVEP3	3	2
HMHA1	3	2
IGDCC4	3	2
IL17RB	3	2
ITIH3	3	2
KIR3DL1	3	2
LPIN1	3	2
MEGF6	3	2
MGAM	3	2
NAF1	3	2
PALM	3	2
PLEKHG4B	3	2
RYK	3	2
RYR3	3	2
SCN9A	3	2
SRD5A1	3	2
TBC1D22B	3	2
TBC1D2B	3	2
TGM6	3	2
TNS1	3	2
TSPAN10	3	2
UQCRFS1	3	2
WDR27	3	2
XPO5	3	2
XPO7	3	2
ACAN	2	2

Table S5: List of 380 genes poorly covered in all 6 WES samples indicating those that are known to be involved in Mendelian diseases (source: OMIM).

Associated Gene Name	Chr	Mendelian diseases	WES % of BP coverage > 8X	WGS % of BP coverage > 8X	Description
WNT4	1	46,XX SEX REVERSAL WITH DYSGENESIS OF KIDNEYS, ADRENALS, AND LUNGS / MAYER-ROKITANSKY-KUSTER-HAUSER SYNDROM / MULLERIAN APLASIA AND HYPERANDROGENISM	85.0	99.5	wingless-type MMTV integration site family, member 4
HSD11B2	16	APPARENT MINERALOCORTICOID EXCESS; AME	82.5	99.9	hydroxysteroid (11-beta) dehydrogenase 2
IFNGR2	21	ATYPICAL MYCOBACTERIOSIS, FAMILIAL	83.8	100.0	interferon gamma receptor 2 (interferon gamma transducer 1)
IL12B	5	ATYPICAL MYCOBACTERIOSIS, FAMILIAL / PSORIASIS SUSCEPTIBILITY 11; PSORS11	81.3	100.0	interleukin 12B
SDHA	5	CARDIOMYOPATHY, DILATED, 1GG; CMD1GG / LEIGH SYNDROME; LS / MITOCHONDRIAL COMPLEX II DEFICIENCY / PARAGANGLIOMAS 5; PGL5	67.2	100.0	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)
LIM2	19	CATARACT 19; CTRCT19	75.8	100.0	lens intrinsic membrane protein 2, 19kDa
SLC6A8	X	CEREBRAL CREATINE DEFICIENCY SYNDROME 1; CCDS1	44.0	96.0	solute carrier family 6 (neurotransmitter transporter), member 8
DNAI2	17	CILIARY DYSKINESIA, PRIMARY, 9; CILD9	83.8	100.0	dynein, axonemal, intermediate chain 2
KRT18	12	CIRRHOSIS, FAMILIAL	77.6	100.0	keratin 18
CRLF1	19	COLD-INDUCED SWEATING SYNDROME 1; CISS1	83.6	100.0	cytokine receptor-like factor 1
GDF1	19	CONOTRUNCAL HEART MALFORMATIONS; CTHM / RIGHT ATRIAL ISOMERISM; RAI / TETRALOGY OF FALLOT; TOF / TRANSPOSITION OF THE GREAT ARTERIES, DEXTRO-LOOPED 3; DTGA3	77.5	98.8	growth differentiation factor 1
TUBB3	16	CORTICAL DYSPLASIA, COMPLEX, WITH OTHER BRAIN MALFORMATIONS 1; CDCBM1 / FIBROSIS OF EXTRAOCULAR MUSCLES, CONGENITAL, 3A, WITH OR WITHOUT EXTRAOCULAR	82.3	100.0	tubulin, beta 3 class III
TUBB4A	19	DYSTONIA 4, TORSION, AUTOSOMAL DOMINANT; DYT4 / LEUKODYSTROPHY, HYPOMYELINATING, 6; HLD6	77.0	100.0	tubulin, beta 4A class IVa
EWSR1	22	EWING SARCOMA; ES / HISTIOCYTOMA, ANGIOMATOID FIBROUS	84.9	100.0	EWS RNA-binding protein 1
GGT1	22	GLUTATHIONURIA	63.8	100.0	gamma-glutamyltransferase 1
GK	X	GLYCEROL KINASE DEFICIENCY	78.1	100.0	glycerol kinase
BLOC1S3	19	HERMANSKY-PUDLAK SYNDROME 8; HPS8	72.4	100.0	biogenesis of lysosomal organelles complex-1, subunit 3
ACVR2B	3	HETEROTAXY, VISCERAL, 4, AUTOSOMAL; HTX4	83.4	99.7	activin A receptor, type IIB
HS6ST1	2	HYPOGONADOTROPIC HYPOGONADISM 15 WITH OR WITHOUT ANOSMIA; HH15	82.1	100.0	heparan sulfate 6-O-sulfotransferase 1
FGF8	10	HYPOGONADOTROPIC HYPOGONADISM 6 WITH OR WITHOUT ANOSMIA; HH6	64.8	100.0	fibroblast growth factor 8 (androgen-induced)
SOX18	20	HYPOTRICHOSIS-LYMPHEDEMA-TELANGIECTASIA SYNDROME; HLTS	78.5	99.0	SRY (sex determining region Y)-box 18
MGP	12	KEUTEL SYNDROME	73.9	100.0	matrix Gla protein
IMPDH1	7	LEBER CONGENITAL AMAUROSIS 11; LCA11 / RETINITIS PIGMENTOSA 10; RP10	79.7	99.8	IMP (inosine 5'-monophosphate) dehydrogenase 1
RDH12	14	LEBER CONGENITAL AMAUROSIS 13; LCA13	77.1	100.0	retinol dehydrogenase 12 (all-trans/9-cis/11-cis)
NMNAT1	1	LEBER CONGENITAL AMAUROSIS 9; LCA9	85.0	100.0	nicotinamide nucleotide adenylyltransferase 1

SURF1	9	LEIGH SYNDROME; LS	80.5	99.7	surfeit 1
PIP5K1C	19	LETHAL CONGENITAL CONTRACTURE SYNDROME 3; LCCS3	82.7	100.0	phosphatidylinositol-4-phosphate 5-kinase, type I, gamma
SNTA1	20	LONG QT SYNDROME 12; LQT12	81.5	100.0	syntrophin, alpha 1
LHB	19	LUTEINIZING HORMONE, BETA POLYPEPTIDE; LHB	23.5	100.0	luteinizing hormone beta polypeptide
DHFR	5	MEGALOBLASTIC ANEMIA DUE TO DIHYDROFOLATE REDUCTASE DEFICIENCY	80.6	100.0	dihydrofolate reductase
TSPAN7	X	MENTAL RETARDATION, X-LINKED 58; MRX58	78.9	99.3	tetraspanin 7
SMS	X	MENTAL RETARDATION, X-LINKED, SYNDROMIC, SNYDER-ROBINSON TYPE; MRXSSR	69.5	100.0	spermine synthase
VSX2	14	MICROPHTHALMIA, ISOLATED 2; MCOP2 / MICROPHTHALMIA, ISOLATED, WITH COLOBOMA 3; MCOPCB3	78.4	100.0	visual system homeobox 2
KRT83	12	MONILETHRIX	82.2	100.0	keratin 83
POMT2	14	MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH BRAIN AND EYE / MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (CONGENITAL WITH MENTAL RETARDATION) / MUSCULAR DYSTROPHY-DYSTROGLYCANOPATHY (LIMB-GIRDLE), TYPE C, 2; MDDGC2	85.6	100.0	protein-O-mannosyltransferase 2
DOK7	4	MYASTHENIA, LIMB-GIRDLE, FAMILIAL	83.0	99.8	docking protein 7
BANF1	11	NESTOR-GUILLERMO PROGERIA SYNDROME; NGPS	70.6	99.8	barrier to autointegration factor 1
REEP1	2	NEURONOPATHY, DISTAL HEREDITARY MOTOR, TYPE VB; HMN5B / SPASTIC PARAPLEGIA 31, AUTOSOMAL DOMINANT; SPG31	84.3	100.0	receptor accessory protein 1
NAA10	X	OGDEN SYNDROME; OGDNS	83.1	98.7	N(alpha)-acetyltransferase 10, NatA catalytic subunit
PPIB	15	OSTEOGENESIS IMPERFECTA, TYPE IX; OI9	74.2	100.0	peptidylprolyl isomerase B (cyclophilin B)
SPINK1	5	PANCREATITIS, HEREDITARY; PCTT / TROPICAL CALCIFIC PANCREATITIS	80.7	100.0	serine peptidase inhibitor, Kazal type 1
AMH	19	PERSISTENT MULLERIAN DUCT SYNDROME, TYPES I AND II; PMDS	82.1	100.0	anti-Mullerian hormone
PSPH	7	PHOSPHOSERINE PHOSPHATASE DEFICIENCY; PSPHD	72.4	100.0	phosphoserine phosphatase
IGFBP7	4	RETINAL ARTERIAL MACROANEURYSM WITH SUPRAVALVULAR PULMONIC STENOSIS;	81.4	100.0	insulin-like growth factor binding protein 7
PRPF31	19	RETINITIS PIGMENTOSA 11; RP11	84.3	100.0	pre-mRNA processing factor 31
RP9	7	RETINITIS PIGMENTOSA 9; RP9	79.5	100.0	retinitis pigmentosa 9 (autosomal dominant)
KCNC3	19	SPINOCEREBELLAR ATAXIA 13; SCA13	77.2	97.4	potassium voltage-gated channel, Shaw-related subfamily, member 3
HES7	17	SPONDYLOCOSTAL DYSOSTOSIS 4, AUTOSOMAL RECESSIVE; SCDO4	66.0	100.0	hes family bHLH transcription factor 7
DDX11	12	WARSAW BREAKAGE SYNDROME; WABS	64.3	100.0	DEAD/H (Asp-Glu-Ala-Asp/His) box helicase 11
MXRA8	1	None	79.8	99.4	matrix-remodelling associated 8
ANKRD65	1	None	64.5	100.0	ankyrin repeat domain 65
TMEM88B	1	None	34.1	97.0	transmembrane protein 88B
C1orf233	1	None	78.1	99.7	chromosome 1 open reading frame 233
MMP23B	1	None	62.5	98.0	matrix metallopeptidase 23B
C1orf86	1	None	84.1	97.9	chromosome 1 open reading frame 86
GPR153	1	None	85.7	100.0	G protein-coupled receptor 153
APITD1	1	None	86.6	100.0	apoptosis-inducing, TAF9-like domain 1

PRAMEF1	1	None	13.5	100.0	PRAME family member 1
FAM131C	1	None	67.3	99.2	family with sequence similarity 131, member C
CROCC	1	None	76.4	100.0	ciliary rootlet coiled-coil, rootletin
IGSF21	1	None	84.1	99.2	immunoglobin superfamily, member 21
AKR7A3	1	None	78.8	100.0	aldo-keto reductase family 7, member A3 (aflatoxin aldehyde reductase)
AKR7A2	1	None	80.7	100.0	aldo-keto reductase family 7, member A2 (aflatoxin aldehyde reductase)
CAMK2N1	1	None	68.6	99.6	calcium/calmodulin-dependent protein kinase II inhibitor 1
TRNP1	1	None	73.6	100.0	TMF1-regulated nuclear protein 1
RAB42	1	None	69.8	99.8	RAB42, member RAS oncogene family
HDAC1	1	None	84.4	100.0	histone deacetylase 1
FAM229A	1	None	79.9	100.0	family with sequence similarity 229, member A
BMP8A	1	None	77.6	100.0	bone morphogenetic protein 8a
BMP8B	1	None	84.9	100.0	bone morphogenetic protein 8b
YBX1	1	None	85.1	98.9	Y box binding protein 1
LDLRAD1	1	None	84.0	100.0	low density lipoprotein receptor class A domain containing 1
SSBP3	1	None	81.1	99.9	single stranded DNA binding protein 3
FAM19A3	1	None	76.8	100.0	family with sequence similarity 19 (chemokine (C-C motif)-like), member A3
C1orf106	1	None	82.4	100.0	chromosome 1 open reading frame 106
NENF	1	None	76.0	100.0	neudesin neurotrophic factor
ABCB10	1	None	77.1	99.9	ATP-binding cassette, sub-family B (MDR/TAP), member 10
OPN3	1	None	82.4	98.4	opsin 3
C1orf229	1	None	70.1	100.0	chromosome 1 open reading frame 229
OR2L8	1	None	2.6	100.0	olfactory receptor, family 2, subfamily L, member 8 (gene/pseudogene)
OR2M3	1	None	78.3	100.0	olfactory receptor, family 2, subfamily M, member 3
CYS1	2	None	79.7	100.0	cystin 1
PQLC3	2	None	82.7	94.6	PQ loop repeat containing 3
CGREF1	2	None	83.8	100.0	cell growth regulator with EF-hand domain 1
MEMO1	2	None	80.7	100.0	mediator of cell motility 1
PKDCC	2	None	83.3	99.5	protein kinase domain containing, cytoplasmic
RPS27A	2	None	82.8	100.0	ribosomal protein S27a
C1D	2	None	78.3	100.0	C1D nuclear receptor corepressor
CD8B	2	None	65.9	100.0	CD8b molecule
FOXI3	2	None	79.1	100.0	forkhead box I3
TRIM43B	2	None	65.6	100.0	tripartite motif containing

					43B
ANKRD36	2	None	67.6	100.0	ankyrin repeat domain 36
PDCL3	2	None	76.6	100.0	phosducin-like 3
POU3F3	2	None	61.8	96.1	POU class 3 homeobox 3
TMEM37	2	None	85.2	99.9	transmembrane protein 37
HNRNPA3	2	None	80.3	100.0	heterogeneous nuclear ribonucleoprotein A3
NDUFB3	2	None	47.5	100.0	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa
CCNYL1	2	None	84.9	100.0	cyclin Y-like 1
WNT6	2	None	78.5	100.0	wingless-type MMTV integration site family, member 6
NPPC	2	None	70.4	100.0	natriuretic peptide C
ASB18	2	None	80.0	99.9	ankyrin repeat and SOCS box containing 18
HES6	2	None	74.4	100.0	hes family bHLH transcription factor 6
PRR21	2	None	42.3	100.0	proline rich 21
HIGD1A	3	None	78.8	100.0	HIG1 hypoxia inducible domain family, member 1A
TMEM42	3	None	82.4	100.0	transmembrane protein 42
MRP63	3	None	73.6	100.0	-
PODXL2	3	None	85.1	100.0	podocalyxin-like 2
EFCC1	3	None	83.7	100.0	EF-hand and coiled-coil domain containing 1
RAB43	3	None	70.9	94.8	RAB43, member RAS oncogene family
CDV3	3	None	85.2	100.0	CDV3 homolog (mouse)
CAMK2N2	3	None	77.4	99.3	calcium/calmodulin-dependent protein kinase II inhibitor 2
IGF2BP2	3	None	85.6	99.8	insulin-like growth factor 2 mRNA binding protein 2
RPL39L	3	None	79.8	100.0	ribosomal protein L39-like
MFI2	3	None	86.6	100.0	antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5
CTBP1	4	None	79.7	99.7	C-terminal binding protein 1
C4orf48	4	None	81.0	100.0	chromosome 4 open reading frame 48
ADRA2C	4	None	77.7	100.0	adrenoceptor alpha 2C
UGT2B28	4	None	10.1	100.0	UDP glucuronosyltransferase 2 family, polypeptide B28
OSTC	4	None	71.5	100.0	oligosaccharyltransferase complex subunit (non-catalytic)
RPS3A	4	None	83.2	100.0	ribosomal protein S3A
PRSS48	4	None	85.4	100.0	protease, serine, 48
ANKRD33B	5	None	79.1	99.9	ankyrin repeat domain 33B
FOXD1	5	None	57.3	100.0	forkhead box D1
VDAC1	5	None	77.7	100.0	voltage-dependent anion channel 1
CDKN2AIPNL	5	None	71.4	100.0	CDKN2A interacting protein N-terminal like
SAP30L	5	None	82.5	100.0	SAP30-like

FABP6	5	None	81.0	100.0	fatty acid binding protein 6, ileal
ATP6V0E1	5	None	86.3	100.0	ATPase, H+ transporting, lysosomal 9kDa, V0 subunit e1
C5orf47	5	None	77.6	100.0	chromosome 5 open reading frame 47
PRR7	5	None	83.2	100.0	proline rich 7 (synaptic)
C5orf60	5	None	79.8	100.0	chromosome 5 open reading frame 60
TUBB2A	6	None	71.3	100.0	tubulin, beta 2A class IIa
HIST1H2BK	6	None	49.9	100.0	histone cluster 1, H2bk
LSM2	6	None	71.3	100.0	LSM2 homolog, U6 small nuclear RNA associated (S. cerevisiae)
RPS10-NUDT3	6	None	73.6	100.0	RPS10-NUDT3 readthrough
RPL10A	6	None	73.7	100.0	ribosomal protein L10a
CLPSL2	6	None	73.7	100.0	colipase-like 2
SLC35B2	6	None	64.9	100.0	solute carrier family 35 (adenosine 3'-phospho 5'-phosphosulfate transporter), member B2
ANKRD66	6	None	71.3	100.0	ankyrin repeat domain 66
CD24	6	None	59.4	98.6	CD24 molecule
METTL24	6	None	79.4	100.0	methyltransferase like 24
FAM26F	6	None	80.7	100.0	family with sequence similarity 26, member F
NUS1	6	None	82.3	100.0	nuclear undecaprenyl pyrophosphate synthase 1 homolog (S. cerevisiae)
CENPW	6	None	81.8	100.0	centromere protein W
PHF10	6	None	76.8	100.0	PHD finger protein 10
UNCX	7	None	70.8	99.9	UNC homeobox
NUDT1	7	None	77.0	100.0	nudix (nucleoside diphosphate linked moiety X)-type motif 1
RSPH10B2	7	None	68.3	99.9	radial spoke head 10 homolog B2 (Chlamydomonas)
NFE2L3	7	None	81.9	100.0	nuclear factor, erythroid 2-like 3
SEPT7	7	None	67.3	99.9	septin 7
VOPP1	7	None	74.8	100.0	vesicular, overexpressed in cancer, prosurvival protein 1
CHCHD2	7	None	74.5	100.0	coiled-coil-helix-coiled-coil-helix domain containing 2
ATP5J2	7	None	36.4	100.0	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit F2
CLEC2L	7	None	80.6	100.0	C-type lectin domain family 2, member L
MKRN1	7	None	69.4	99.9	makorin ring finger protein 1
XRCC2	7	None	84.2	100.0	X-ray repair complementing defective repair in Chinese hamster cells 2
FBXO16	8	None	80.7	100.0	F-box protein 16
NKX6-3	8	None	67.1	99.8	NK6 homeobox 3
CEBPD	8	None	83.8	100.0	CCAAT/enhancer binding protein (C/EBP), delta

LYPLA1	8	None	82.9	100.0	lysophospholipase I
TCF24	8	None	78.1	100.0	transcription factor 24
FABP5	8	None	65.5	100.0	fatty acid binding protein 5 (psoriasis-associated)
YWHAZ	8	None	79.9	100.0	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta
KHDRBS3	8	None	79.1	99.9	KH domain containing, RNA binding, signal transduction associated 3
BOP1	8	None	72.2	82.8	block of proliferation 1
RPL8	8	None	84.8	100.0	ribosomal protein L8
AK3	9	None	75.8	100.0	adenylate kinase 3
IFNA4	9	None	13.5	100.0	interferon, alpha 4
ANKRD18B	9	None	67.2	100.0	ankyrin repeat domain 18B
ANKRD18A	9	None	66.1	100.0	ankyrin repeat domain 18A
SUSD3	9	None	79.6	100.0	sushi domain containing 3
TMEFF1	9	None	45.2	100.0	transmembrane protein with EGF-like and two follistatin-like domains 1
OR13C2	9	None	63.9	100.0	olfactory receptor, family 13, subfamily C, member 2
GNG10	9	None	61.2	100.0	guanine nucleotide binding protein (G protein), gamma 10
FPGS	9	None	85.0	99.6	folylpolyglutamate synthase
SET	9	None	78.9	99.7	SET nuclear oncogene
SH3GLB2	9	None	79.6	100.0	SH3-domain GRB2-like endophilin B2
IER5L	9	None	83.8	100.0	immediate early response 5-like
NCS1	9	None	84.9	98.9	neuronal calcium sensor 1
C9orf172	9	None	75.1	99.9	chromosome 9 open reading frame 172
C9orf37	9	None	0.0	100.0	chromosome 9 open reading frame 37
TMEM236	10	None	53.7	97.8	transmembrane protein 236
BMI1	10	None	6.0	92.8	BMI1 polycomb ring finger oncogene
MTRNR2L7	10	None	8.4	100.0	MT-RNR2-like 7
UTF1	10	None	49.8	100.0	undifferentiated embryonic cell transcription factor 1
SCT	11	None	65.3	99.6	secretin
DUSP8	11	None	79.9	100.0	dual specificity phosphatase 8
KRTAP5-3	11	None	58.0	100.0	keratin associated protein 5-3
C11orf91	11	None	83.1	97.4	chromosome 11 open reading frame 91
SYT7	11	None	82.6	100.0	synaptotagmin VII
C11orf83	11	None	79.8	100.0	chromosome 11 open reading frame 83
CNIH2	11	None	82.6	99.8	cornichon family AMPA receptor auxiliary protein 2
ANAPC15	11	None	67.5	100.0	anaphase promoting complex subunit 15
RAB6A	11	None	82.3	100.0	RAB6A, member RAS oncogene family
CLNS1A	11	None	82.2	100.0	chloride channel, nucleotide-sensitive, 1A

TMPRSS5	11	None	83.8	100.0	transmembrane protease, serine 5
NRGN	11	None	78.1	100.0	neurogranin (protein kinase C substrate, RC3)
PTMS	12	None	81.1	100.0	parathymosin
NANOG	12	None	70.5	100.0	Nanog homeobox
KLRC4-KLRK1	12	None	54.2	100.0	KLRC4-KLRK1 readthrough
PRR4	12	None	84.9	100.0	proline rich 4 (lacrimal)
LALBA	12	None	73.2	100.0	lactalbumin, alpha-
DNAJC22	12	None	75.1	100.0	DnaJ (Hsp40) homolog, subfamily C, member 22
POU6F1	12	None	78.4	100.0	POU class 6 homeobox 1
SMAGP	12	None	51.0	100.0	small cell adhesion glycoprotein
FIGNL2	12	None	84.2	100.0	fidgetin-like 2
EIF4B	12	None	77.9	100.0	eukaryotic translation initiation factor 4B
MARCH9	12	None	78.3	99.8	membrane-associated ring finger (C3HC4) 9
LLPH	12	None	14.8	100.0	LLP homolog, long-term synaptic facilitation (Aplysia)
C12orf73	12	None	67.8	100.0	chromosome 12 open reading frame 73
C12orf75	12	None	74.7	100.0	chromosome 12 open reading frame 75
CKAP4	12	None	80.5	100.0	cytoskeleton-associated protein 4
CCDC42B	12	None	80.5	100.0	coiled-coil domain containing 42B
SDS	12	None	79.7	100.0	serine dehydratase
C12orf49	12	None	82.0	100.0	chromosome 12 open reading frame 49
HRK	12	None	54.1	100.0	harakiri, BCL2 interacting protein
RPLP0	12	None	84.1	100.0	ribosomal protein, large, P0
SETD8	12	None	73.5	100.0	SET domain containing (lysine methyltransferase) 8
MMP17	12	None	86.2	100.0	matrix metallopeptidase 17 (membrane-inserted)
FBRSL1	12	None	85.8	99.8	fibrosin-like 1
PXMP2	12	None	81.4	100.0	peroxisomal membrane protein 2, 22kDa
IL17D	13	None	81.1	100.0	interleukin 17D
USP12	13	None	67.9	100.0	ubiquitin specific peptidase 12
OR4N2	14	None	66.5	100.0	olfactory receptor, family 4, subfamily N, member 2
CCNB1IP1	14	None	74.0	100.0	cyclin B1 interacting protein 1, E3 ubiquitin protein ligase
TPPP2	14	None	81.7	100.0	tubulin polymerization-promoting protein family member 2
RPS29	14	None	73.4	100.0	ribosomal protein S29
PLEK2	14	None	84.2	100.0	pleckstrin 2
ACOT4	14	None	83.2	100.0	acyl-CoA thioesterase 4
TMED10	14	None	83.3	99.3	transmembrane emp24-like trafficking protein 10 (yeast)
COX8C	14	None	69.6	100.0	cytochrome c oxidase subunit VIIIC
IFI27L1	14	None	70.9	100.0	interferon, alpha-inducible

					protein 27-like 1
HHIPL1	14	None	83.3	97.7	HHIP-like 1
NUDT14	14	None	80.4	99.6	nudix (nucleoside diphosphate linked moiety X)-type motif 14
TEX22	14	None	57.2	100.0	testis expressed 22
CRIP1	14	None	64.8	100.0	cysteine-rich protein 1 (intestinal)
AVEN	15	None	81.1	99.9	apoptosis, caspase activation inhibitor
GOLGA8B	15	None	24.8	100.0	golgin A8 family, member B
MAPK6	15	None	76.1	99.9	mitogen-activated protein kinase 6
TMEM202	15	None	76.2	100.0	transmembrane protein 202
COX5A	15	None	80.7	100.0	cytochrome c oxidase subunit Va
COMMD4	15	None	79.5	100.0	COMM domain containing 4
ADAMTS7	15	None	78.6	99.7	ADAM metallopeptidase with thrombospondin type 1 motif, 7
MORF4L1	15	None	75.6	100.0	mortality factor 4 like 1
WHAMM	15	None	83.5	100.0	WAS protein homolog associated with actin, golgi membranes and microtubules
FAM103A1	15	None	74.7	100.0	family with sequence similarity 103, member A1
HDGFRP3	15	None	80.3	100.0	Hepatoma-derived growth factor-related protein 3
HBZ	16	None	59.8	78.1	hemoglobin, zeta
NME4	16	None	83.2	99.9	NME/NM23 nucleoside diphosphate kinase 4
C16orf13	16	None	71.7	100.0	chromosome 16 open reading frame 13
METRN	16	None	66.3	99.3	meteorin, glial cell differentiation regulator
TPSD1	16	None	75.6	100.0	tryptase delta 1
HS3ST6	16	None	77.2	98.8	heparan sulfate (glucosamine) 3-O-sulfotransferase 6
SLC9A3R2	16	None	80.4	100.0	solute carrier family 9, subfamily A (NHE3, cation proton antiporter 3), member 3 regulator 2
TCEB2	16	None	5.3	99.6	transcription elongation factor B (SIII), polypeptide 2 (18kDa, elongin B)
HCFC1R1	16	None	82.1	100.0	host cell factor C1 regulator 1 (XPO1 dependent)
MTRNR2L4	16	None	58.1	100.0	MT-RNR2-like 4
DEXI	16	None	31.1	100.0	Dexi homolog (mouse)
SOCS1	16	None	77.2	100.0	suppressor of cytokine signaling 1
MPV17L	16	None	73.5	100.0	MPV17 mitochondrial membrane protein-like
C16orf52	16	None	65.6	98.7	chromosome 16 open reading frame 52
FAM57B	16	None	79.0	97.1	family with sequence similarity 57, member B
CTF1	16	None	75.3	100.0	cardiotrophin 1
COX6A2	16	None	75.0	94.3	cytochrome c oxidase subunit VIa polypeptide 2

BRD7	16	None	82.2	99.9	bromodomain containing 7
BCAR1	16	None	85.1	100.0	breast cancer anti-estrogen resistance 1
WFDC1	16	None	80.5	100.0	WAP four-disulfide core domain 1
ZFPM1	16	None	77.2	99.9	zinc finger protein, FOG family member 1
DBNDD1	16	None	84.8	100.0	dysbindin (dystrobrevin binding protein 1) domain containing 1
RILP	17	None	82.9	99.1	Rab interacting lysosomal protein
C1QBP	17	None	70.2	99.9	complement component 1, q subcomponent binding protein
MAP2K4	17	None	84.4	100.0	mitogen-activated protein kinase kinase 4
FAM18B2	17	None	77.9	99.8	trans-golgi network vesicle protein 23 homolog C (S. cerevisiae)
LGALS9	17	None	59.4	100.0	lectin, galactoside-binding, soluble, 9
C17orf50	17	None	78.2	100.0	chromosome 17 open reading frame 50
CCL3	17	None	71.8	100.0	chemokine (C-C motif) ligand 3
CCL4	17	None	60.6	100.0	chemokine (C-C motif) ligand 4
PTGES3L	17	None	80.1	99.9	prostaglandin E synthase 3 (cytosolic)-like
C17orf105	17	None	82.2	100.0	chromosome 17 open reading frame 105
FAM171A2	17	None	76.2	100.0	family with sequence similarity 171, member A2
TBKBP1	17	None	82.3	98.9	TBK1 binding protein 1
CBX1	17	None	66.9	100.0	chromobox homolog 1
SNX11	17	None	84.8	100.0	sorting nexin 11
ATP5G1	17	None	73.4	99.9	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit C1 (subunit 9)
C17orf77	17	None	44.3	100.0	chromosome 17 open reading frame 77
SUMO2	17	None	76.2	100.0	small ubiquitin-like modifier 2
SYNGR2	17	None	81.9	100.0	synaptogyrin 2
CHMP6	17	None	82.6	100.0	charged multivesicular body protein 6
C17orf89	17	None	65.9	100.0	chromosome 17 open reading frame 89
NOTUM	17	None	80.5	98.4	notum pectinacetylesterase homolog (Drosophila)
RAC3	17	None	80.0	98.0	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3)
FN3K	17	None	84.4	99.8	fructosamine 3 kinase
METRNL	17	None	80.9	99.7	meteorin, glial cell differentiation regulator-like
TUBB6	18	None	83.0	100.0	tubulin, beta 6 class V
SLMO1	18	None	78.4	100.0	slowmo homolog 1 (Drosophila)
SERPINB10	18	None	81.9	100.0	serpin peptidase inhibitor, clade B (ovalbumin), member 10

SHC2	19	None	76.9	99.5	SHC (Src homology 2 domain containing) transforming protein 2
ODF3L2	19	None	80.5	100.0	outer dense fiber of sperm tails 3-like 2
HCN2	19	None	58.8	95.5	hyperpolarization activated cyclic nucleotide-gated potassium channel 2
FGF22	19	None	72.6	100.0	fibroblast growth factor 22
RNF126	19	None	80.6	99.1	ring finger protein 126
PALM	19	None	77.9	98.7	paralemmin
R3HDM4	19	None	78.6	100.0	R3H domain containing 4
GRIN3B	19	None	79.7	100.0	glutamate receptor, ionotropic, N-methyl-D-aspartate 3B
C19orf26	19	None	79.4	99.9	chromosome 19 open reading frame 26
EFNA2	19	None	76.7	99.2	ephrin-A2
RPS15	19	None	62.2	100.0	ribosomal protein S15
MEX3D	19	None	77.1	97.2	mex-3 RNA binding family member D
TCF3	19	None	78.6	99.9	transcription factor 3
ONECUT3	19	None	53.9	98.5	one cut homeobox 3
KLF16	19	None	61.0	99.7	Kruppel-like factor 16
ABHD17A	19	None	80.0	100.0	abhydrolase domain containing 17A
CSNK1G2	19	None	83.5	98.9	casein kinase 1, gamma 2
BTBD2	19	None	85.2	98.7	BTB (POZ) domain containing 2
GNG7	19	None	78.7	100.0	guanine nucleotide binding protein (G protein), gamma 7
MPND	19	None	82.2	100.0	MPN domain containing
CHAF1A	19	None	85.2	100.0	chromatin assembly factor 1, subunit A (p150)
RPL36	19	None	80.5	100.0	ribosomal protein L36
C19orf70	19	None	71.4	100.0	chromosome 19 open reading frame 70
MLLT1	19	None	83.6	99.2	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 1
ALKBH7	19	None	72.3	100.0	alkB, alkylation repair homolog 7 (E. coli)
PET100	19	None	81.2	100.0	PET100 homolog (S. cerevisiae)
PIN1	19	None	80.7	100.0	peptidylprolyl cis/trans isomerase, NIMA-interacting 1
S1PR5	19	None	83.9	100.0	sphingosine-1-phosphate receptor 5
C19orf80	19	None	70.6	100.0	chromosome 19 open reading frame 80
TSPAN16	19	None	78.0	100.0	tetraspanin 16
ZNF69	19	None	73.0	100.0	zinc finger protein 69
SAMD1	19	None	79.1	98.6	sterile alpha motif domain containing 1
NDUFB7	19	None	70.5	99.7	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7, 18kDa
TMEM221	19	None	71.6	100.0	transmembrane protein 221

CCDC124	19	None	75.7	99.7	coiled-coil domain containing 124
PBX4	19	None	84.7	100.0	pre-B-cell leukemia homeobox 4
ZNF626	19	None	72.9	100.0	zinc finger protein 626
UQCRFS1	19	None	67.4	100.0	ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1
PDCD5	19	None	79.1	100.0	programmed cell death 5
RHPN2	19	None	79.3	99.1	rhophilin, Rho GTPase binding protein 2
SLC7A10	19	None	80.9	100.0	solute carrier family 7 (neutral amino acid transporter light chain, asc system), member 10
USF2	19	None	79.7	96.4	upstream transcription factor 2, c-fos interacting
LGALS7	19	None	17.8	100.0	lectin, galactoside-binding, soluble, 7
C19orf69	19	None	50.2	100.0	glutamate-rich 4
GRIK5	19	None	83.4	100.0	glutamate receptor, ionotropic, kainate 5
PSG2	19	None	44.4	100.0	pregnancy specific beta-1-glycoprotein 2
APOC1	19	None	69.7	100.0	apolipoprotein C-I
BBC3	19	None	74.4	100.0	BCL2 binding component 3
PRR24	19	None	81.7	88.5	proline rich 24
MEIS3	19	None	81.1	100.0	Meis homeobox 3
DBP	19	None	82.1	100.0	D site of albumin promoter (albumin D-box) binding protein
CGB7	19	None	11.0	100.0	chorionic gonadotropin, beta polypeptide 7
LIN7B	19	None	83.1	100.0	lin-7 homolog B (C. elegans)
ZNF578	19	None	74.3	100.0	zinc finger protein 578
TMEM86B	19	None	82.2	100.0	transmembrane protein 86B
TMEM238	19	None	46.1	100.0	transmembrane protein 238
UBE2S	19	None	75.1	100.0	ubiquitin-conjugating enzyme E2S
RFPL4A	19	None	52.9	100.0	ret finger protein-like 4A
ZBTB45	19	None	82.1	100.0	zinc finger and BTB domain containing 45
TCF15	20	None	74.9	100.0	transcription factor 15 (basic helix-loop-helix)
SIRPB1	20	None	66.5	63.6	signal-regulatory protein beta 1
EBF4	20	None	80.1	99.1	early B-cell factor 4
SNX5	20	None	80.9	100.0	sorting nexin 5
DEFB119	20	None	72.6	100.0	defensin, beta 119
CCM2L	20	None	83.0	99.9	cerebral cavernous malformation 2-like
GHRH	20	None	81.9	100.0	growth hormone releasing hormone
EMILIN3	20	None	75.9	100.0	elastin microfibril interfacer 3
WFDC8	20	None	80.1	100.0	WAP four-disulfide core domain 8
FAM210B	20	None	83.1	100.0	family with sequence similarity 210, member B
TAF4	20	None	85.9	97.5	TAF4 RNA polymerase II, TATA box binding protein

					(TBP)-associated factor, 135kDa
TCFL5	20	None	65.7	99.1	transcription factor-like 5 (basic helix-loop-helix)
LIME1	20	None	73.9	100.0	Lck interacting transmembrane adaptor 1
LKAAEAR1	20	None	73.3	100.0	LKAAEAR motif containing 1
MRPS6	21	None	85.1	97.5	mitochondrial ribosomal protein S6
HMGN1	21	None	74.5	97.9	high mobility group nucleosome binding domain 1
FAM207A	21	None	66.6	100.0	family with sequence similarity 207, member A
GSC2	22	None	76.9	100.0	goosecoid homeobox 2
RTN4R	22	None	83.5	100.0	reticulon 4 receptor
EIF4ENIF1	22	None	86.0	100.0	eukaryotic translation initiation factor 4E nuclear import factor 1
SLC16A8	22	None	75.9	100.0	solute carrier family 16 (monocarboxylate transporter), member 8
ST13	22	None	72.3	100.0	suppression of tumorigenicity 13 (colon carcinoma) (Hsp70 interacting protein)
PRR5	22	None	78.4	100.0	proline rich 5 (renal)
ARHGAP8	22	None	0.0	98.5	Rho GTPase activating protein 8
PRKX	X	None	70.8	100.0	protein kinase, X-linked
MTRNR2L10	X	None	34.5	100.0	MT-RNR2-like 10
EDA2R	X	None	83.9	100.0	ectodysplasin A2 receptor
NONO	X	None	80.8	100.0	non-POU domain containing, octamer-binding
FAM50A	X	None	82.4	99.9	family with sequence similarity 50, member A

References for supporting information:

1. You FM et al. (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9:253.

2. Flicek P et al. (2014) Ensembl 2014. *Nucleic Acids Res* 42:D749–D755.

3. Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 4:1184–1191.