ARTICLE

Preconception Carrier Screening by Genome Sequencing: Results from the Clinical Laboratory

Sumit Punj,[1] Yassmine Akkari,[1,9,10] Jennifer Huang,[1,10] Fei Yang,[1] Allison Creason,[1] Christine Pak,[1] Amiee Potter,[1] Michael O. Dorschner,[2,3] Deborah A. Nickerson,[3] Peggy D. Robertson,[3] Gail P. Jarvik,[3,4] Laura M. Amendola,[4] Jennifer Schleit,[2] Dana Kostiner Simpson,[5] Alan F. Rope,[5] Jacob Reiss,[6] Tia Kauffman,[6] Marian J. Gilmore,[5] Patricia Himes,[5] Benjamin Wilfond,[7,8] Katrina A.B. Goddard,[6] and C. Sue Richards[1,*]

Advances in sequencing technologies permit the analysis of a larger selection of genes for preconception carrier screening. The study was designed as a sequential carrier screen using genome sequencing to analyze 728 gene-disorder pairs for carrier and medically actionable conditions in 131 women and their partners (n = 71) who were planning a pregnancy. We report here on the clinical laboratory results from this expanded carrier screening program. Variants were filtered and classified using the latest American College of Medical Genetics and Genomics (ACMG) guideline; only pathogenic and likely pathogenic variants were confirmed by orthologous methods before being reported. Novel missense variants were classified as variants of uncertain significance. We reported 304 variants in 202 participants. Twelve carrier couples (12/71 couples tested) were identified for common conditions; eight were carriers for hereditary hemochromatosis. Although both known and novel variants were reported, 48% of all reported variants were missense. For novel splice-site variants, RNA-splicing assays were performed to aid in classification. We reported ten copy-number variants and five variants in non-coding regions. One novel variant was reported in *F8*, associated with hemophilia A; prenatal testing showed that the male fetus harbored this variant and the neonate suffered a life-threatening hemorrhage which was anticipated and appropriately managed. Moreover, 3% of participants had variants that were medically actionable. Compared with targeted mutation screening, genome sequencing improves the sensitivity of detecting clinically significant variants. While certain novel variant interpretation remains challenging, the ACMG guidelines are useful to classify variants in a healthy population.

Introduction

Traditionally, carrier screening has focused on specific disorders that are known to have a higher prevalence in certain ethnic populations. More recently, lower sequencing costs coupled with higher accuracy of next generation sequencing-based methodologies have made it affordable for clinical laboratories to offer screening for substantially more conditions.[1–4] Both autosomal-recessive and X-linked conditions, which comprise a typical carrier-screening panel, are often observed in individuals with no family history of the condition. Therefore, for a healthy couple, offering pan-ethnic, expanded carrier screening is appropriate, particularly in a culturally and genetically heterogeneous population such as the United States. Many professional societies have developed their own practice guidelines on expanded carrier screening, in recognition of its increasing popularity. Furthermore, the American College of Medical Genetics and Genomics, American Congress of Obstetricians and Gynecologists, National Society of Genetic Counselors, Society of Maternal-Fetal Medicine, and the Perinatal Quality Foundation have

collaborated to issue a joint statement for healthcare providers and clinical laboratory personnel to educate and guide them on the use of this screening approach.[5]

Massively parallel sequencing or next-generation sequencing (NGS) has provided the technical means to not only screen the full gene, but also analyze multiple genes and multiple individuals simultaneously, as compared to the targeted mutation panel approach of traditional carrier screening. However, given the rapid pace of its application, there is a paucity of information on the downstream impact of NGS in the healthcare system and in routine medical care.

To this end, the NextGen study (Figure S1A), a part of the National Human Genome Research Institute's Clinical Sequencing Exploratory Research consortium (CSER), was focused on exploring the possibility of using genome sequencing as part of a preconception expanded carrier screening program from a variety of contexts. The multidisciplinary team generated evidence on a variety of goals to achieve this overarching objective including evaluating the clinical utility of genome sequencing (GS) in this clinical scenario, exploring critical interactions between

[1]Department of Molecular and Medical Genetics, Knight Diagnostic Laboratories, Oregon Health & Science University, Portland, OR 97239, USA; [2]Pathology, Division of Medical Genetics, University of Washington, Seattle, WA 98195, USA; [3]Genome Sciences, Division of Medical Genetics, University of Washington, Seattle, WA 98195, USA; [4]Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA 98195, USA; [5]Department of Medical Genetics, Kaiser Permanente Northwest, Portland, OR 97227, USA; [6]Center for Health Research, Kaiser Permanente Northwest, Portland, OR 97227, USA; [7]Department of Pediatrics, Division of Bioethics, University of Washington, Seattle, WA 98195, USA; [8]Treuman Katz Center for Pediatric Bioethics, Seattle Children's Hospital, Seattle, WA 98101, USA
[9]Present address: Legacy Laboratory Services, Legacy Health, Portland, OR 97232, USA
[10]These authors contributed equally to this work
*Correspondence: richarsu@ohsu.edu
https://doi.org/10.1016/j.ajhg.2018.04.004

individuals, providers, and laboratories that influence the implementation of clinical sequencing programs, and identifying and addressing barriers to integration of genomic and health data for clinical decision making. The study was designed as a randomized controlled trial with GS and analysis of a pre-selected list of 728 gene-disorder pairs (genes known to be associated with human disorders) for autosomal-recessive and X-linked conditions[6] as well as 148 genes[7,8] for conditions that are considered medically actionable. We reported known as well as novel variants that for this study were defined as those not previously reported in affected individuals. Here, we describe the analytic pipeline and the clinical laboratory results for subjects who received GS as part of the NextGen study.

Subjects and Methods

Selection of Participants and Study Design

All female participants were members of the Kaiser Permanente Northwest (KPNW) integrated healthcare delivery health management system. Our study is based on the sequential model and not the couple-based model of carrier screening. Females were first sequenced and if at least one positive carrier result was disclosed to the participant, her male partner was invited to join the GS arm of the study following consent (Figure S1B). To be eligible to participate in this experimental randomized controlled trial,[9] the female participants must have satisfied three criteria: (1) planning a pregnancy in the near future, (2) had a carrier screening test, usually cystic fibrosis (MIM: 219700), ordered by a clinician that was resulted and completed, and (3) not pregnant at the time of consent. All women who consented to participate in the study filled out a baseline survey including demographic information prior to being randomly assigned into the GS arm or the usual care arm of the study (Figure S1A). At KPNW, all participants (i.e., females and males) in the GS arm had a pre-test consent visit with a genetic counselor before their blood draw. Blood samples were sent to the CLIA laboratories at Illumina Clinical Services Laboratory and Oregon Health & Science University's (OHSU) Knight Diagnostic Laboratories for GS and variant confirmation, respectively, while secondary analysis was performed at the University of Washington. Positive carrier results were discussed with a genetic counselor for all participants during post-test counseling (Figure S1B). We sequenced a total of 202 participants: 131 females and 71 male partners (i.e., 71 couples). This study was reviewed and approved by the Institutional Review Board (IRB) at Kaiser Permanente Northwest, the University of Washington, and OHSU ceded IRB authority to KPNW. All participants received full written and IRB-approved consent and could withdraw at any time during the study without consequences.

Genes Analyzed

Within the GS arm of the study, only variants in pre-selected genes that were determined by the NextGen Return of Results Committee (RORC)[6] were chosen for analysis (Figure S1). These genes included those for carrier screening and medically actionable findings. For carrier screening, the 728 gene-disorder pairs, which comprised autosomal-recessive and X-linked conditions, were categorized into lifespan limiting (177 genes), serious (406 genes),

mild (93 genes), unpredictable (41 genes), and adult onset (11 genes). For medically actionable (also called secondary or additional) findings,[10] we used an expanded list compared with the most recent ACMG list.[11] This list was comprised of 121 genes for autosomal-dominant conditions, 23 genes for autosomal-recessive conditions, and 4 genes for X-linked conditions. The selection process for these genes has been previously published[7] and was based on their clinical validity and the clinical utility of medically actionable genes. The analytical validity of these genes was one of the metrics that was assessed in this study. The genes are listed in the Supplemental Note. All participants who consented to the study received results for at least the 177 conditions that were categorized as lifespan limiting.[6] The remaining categories of carrier conditions were optional and were returned only if requested. The National Center for Biotechnology Information (NCBI)-curated reference sequences ("NM and NP categories") were used for variant analysis. In addition, variants in the promoter region of CFTR (MIM: 602421) were also analyzed.

Sequencing

Genome sequencing was performed at the Illumina Clinical Services Laboratory. Briefly, genomic DNA was extracted from the participant's blood and processed for sequencing using the Illumina TruSeq DNA LT kit. The DNA sample was sequenced on a HiSeq 2000 or 2500 (Illumina, version 3 chemistry) with 100 base pair, paired-end reads. The sequenced fragments were assessed for quality and aligned to the NCBI reference genome (GRCh37/hg19) to generate BAM files. The BAM files were subsequently sent to the University of Washington for secondary analysis.

Bioinformatics Pipeline (Secondary Analysis)

For compatibility with the University of Washington's data analysis pipeline, FASTQ files, with the original read sequences, were generated from the BAM files received from the Illumina Clinical Services Laboratory. The reads were re-aligned to the NCBI GRCh37/hg19 reference sequence with the Burrows-Wheeler Aligner (v. 0.7.6a).[12] The aligned read data were subject to further analysis using tools from the Genome Analysis Tool Kit (GATK)[13] by removal of duplicate reads (Picard MarkDuplicates v.1.96), indel realignment (GATK RealignerTargetCreater and IndelRealigner v.2.6), and base-quality recalibration (GATK BaseRecalibrator v.2.6).

Single-Nucleotide Variants (SNV) and Small Insertion and Deletions (Indel)

SNVs and indel variants were called by the GATK UnifiedGenotyper v.2.6, followed by the GATK VariantAnnotation and VariantFiltration (to flag low-quality calls). Variant quality terms, QUAL and QD, were assigned by UnifiedGenotyper and the VariantFiltration tool assigned a "PASS" to all variants with a QUAL score > 100 and a QD (QUAL score normalized by allele depth) score > 5. Further annotation was performed by SeattleSeqAnnotation138,[14] using a local cache database that also served the website. These annotations that are included on the website are dbSNP[15] and clinical association data as well as scores from PolyPhen,[16] GERP,[17] CADD,[18] Grantham,[19] protein-protein interactions,[20] microRNAs from miRbase,[21] and population frequencies from NHLBI GO Exome Sequencing Project (ESP)[22] and Exome Aggregation Consortium (ExAC).[23] Also present were University of California Santa Cruz (UCSC) Browser[24] annotations: repeats, chimp alleles, CpG islands, and KEGG pathways. Additional annotations that are not part of the standard
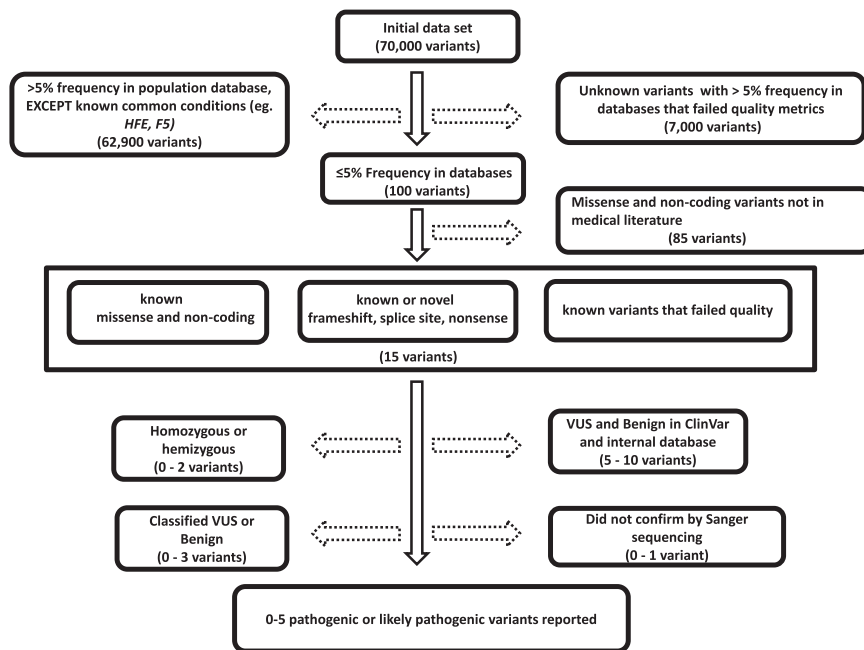
Figure 1. Schematic Representative of Variant Filtering.
The solid arrows are variants that were prioritized and stippled arrows indicate the variants that were discarded in the analysis. The numbers of variants depicted are averages per person.

reporting (Figure S1B). Only those SNVs and CNVs that were classified as either pathogenic or likely pathogenic were confirmed by an orthologous methodology (Sanger sequencing for SNVs; gene-focused array or multiplex PCR for CNVs) before reporting the results to the clinician and participant. Based on the frequency of the variant in the population databases (1000 Genomes Project, ESP, and ExAC), a 5% threshold for variant frequency was used for initially filtering variants (Figure 1). The 2015-ACMG guidelines[37] were used for variant interpretation and classification for SNVs and small indels. Confirmation for SNVs and small indels was performed, in both directions, by capillary electrophoresis-based Sanger sequencing (CE).[38] If a SNV was detected in a gene that was also known to have a pseudogene, long-range PCR followed by nested PCR was performed before CE. Structural variant confirmation for *HBA2* (MIM: 141850) deletion variants was performed by a multiplex polymerase chain reaction method.[39] Other structural variant confirmations were performed on a clinically validated gene-focused array, CytoSure Medical Research Exome Array (Oxford Gene Technology), according to the manufacturer's instructions. Finally, for splice-site variants, splicing analysis was performed by extracting RNA from whole blood, converting the mRNA to cDNA (SuperScript II Reverse Transcriptase, ThermoFisher Scientific), and sequencing the product by CE. Disorders associated with trinucleotide repeats were not analyzed. All reported variants were submitted to ClinVar.[27]

SeattleSeqAnnotation138 software suite were included to further support variant interpretation. These additional tools were SIFT[25] and SPIDEX for splicing,[26] data from ClinVar[27] and the professional version of Human Gene Mutation Database,[28] and population frequencies in 1000 Genomes Project,[29] ICR1000 UK exomes,[30] and 500 local exomes. The Human Genome Variation Society (HGVS)[31] notations were used for the nomenclature of SNVs and indels in exons.

Structural Variant Analysis of the Sequencing Data

Structural variants were called by LUMPY[32] and augmented by CNVnator[33] for the entire genome. For the RORC-selected genes, regions were defined that included 2,000 nucleotides upstream and downstream of the first and last exon in the longest transcript. Overlapping reads that included one or more exons of RORC-selected genes were then analyzed for structural variants. Quality metrics were applied by CNVnator and LUMPY to filter for signal type paired-end and/or split-read. Finally, structural variants in the BAM files were manually curated using the Integrated Genome Viewer (IGV).[34] Pathogenicity of copy-number variants (CNVs) were based on a laboratory-developed guideline that included consideration of the pathogenic mechanism of the variants, classifications in the population databases such as the Database of Genomic Variants (DGV)[35] and the human disease databases such as ClinVar and DECIPHER,[36] medical literature review, the variant frequency, and the consequence of the CNV; i.e., a deletion of several exons resulting in the remaining flanking exons being out of frame, or loss of the translation start site. Novel duplications were generally not reported due to lack of functional evidence of their effect on gene product or location, unless located in a well-studied gene (e.g., *DMD* [MIM: 300377]).

Tertiary Analysis and Variant Confirmation for SNVs and CNVs

The variants that were called and annotated at the University of Washington were sent via a secure site to the OHSU CLIA-laboratory for variant filtering, interpretation, confirmation, and

Clinical Report

The clinical reporting was performed in two phases to enable post-test genetic counseling and survey completion. The first phase included only the carrier results. For participants who chose to receive medically actionable (secondary) findings, an additional report containing positive results, if any, was provided at a later date. A sample clinical report is available in the Supplemental Note.

Results

Participant Choices for Results

The average participant age was 32 years for females (range: 21–46 years) and 34 years for males (range: 24–50 years). White/non-Hispanic participants comprised 78% of study participants. While most (93%) participants opted for receiving all categories of carrier results, some participants (~7%) did not want to know their carrier status for unpredictable or adult-onset onset conditions.

Table 1. Carrier Findings for Well-Known Disease-Associated Variants within Introns

Gene	Variant	Disorder	Condition Category	Variant Category	Classification
PTS (MIM: 612719)	NM_000317.2 (c.84_291A>G)	BH4-deficient hyperphenylalaninemia (MIM: 261640)	serious	known	likely pathogenic
GAA (MIM: 606800)	NM_001079804.2 (c.−32−13T>G)	glycogen storage disease II (MIM: 232300)	serious	known	pathogenic
GJB2 (MIM: 121011)	NM_004004.5 (c.−22−2A>C)	deafness and hearing loss (MIM: 220290)	mild	known	pathogenic
	NM_004004.5 (c.−23+1G>A)			known	pathogenic
PYGM (MIM: 608455)	NM_005609.3 (c.425−26A>G)	McArdle disease (MIM: 232600)	unpredictable	known	likely pathogenic

These variants are not in the canonical splice-site within introns. An *a priori* knowledge of the mutation spectrum in a gene is not required for using GS-based carrier testing. The variants are annotated according to the Human Genome Variation Society (HGVS)[28] recommended nomenclature. The variants in *GAA* and *GJB2* are located in the introns that are upstream (5′ in the coding strand) of the translational start codon (where the adenine position in ATG start codon is +1).

Almost every participant (99%) requested the return of medically actionable findings.

Sequencing Performance

For GS, the average depth of sequencing was 38.5× with an average of 79.5% of reads covered at a depth of at least 30×. *SMN1* ([MIM: 600354]; spinal muscular atrophy [MIM: 253300, 253550, 253400, 271150]) and *IKBKG* ([MIM: 300248]; X-linked hypohidrotic ectodermal dysplasia with immune deficiency [MIM: 300291] and anhidrotic ectodermal dysplasia with immune deficiency, osteopetrosis, and lymphedema [MIM: 300301]) were consistently sequenced at less than 10× coverage. This was due to the close proximity of these genes to their respective pseudogenes, thereby resulting in poor mapping quality of the sequence reads. Furthermore, specific bioinformatic strategies would be required for variant detection of these regions by NGS.[40] Therefore, variant interpretation for those gene-disorder pairs that were sequenced at less than 10× depth were not returned due to insufficient coverage.

Variant Filtering and Confirmation

On average, approximately 70,000 variants were detected in the 728 gene-disorder pairs for each participant. Based on the ACMG guidelines' stand-alone criterion for a benign classification, those variants with an average frequency of >5% in the population databases were classified benign, with the exception of the known clinically significant variants, GenBank: NM_000410.3 (c.845G>A [p.Cys282Tyr]) and GenBank: NM_000410.3 (c.187C>G [p.His63Asp]), variants in *HFE* (MIM: 613609), and the factor V (*F5*) (MIM: 612309) Leiden variant (GenBank: NM_000130.4; c.1601G>A [p.Arg534Gln]). The GenBank: NM_000410.3 (c.187C>G [p.His63Asp]) variant that has a frequency of 10.6% in the ExAC database was reported only if that participant's partner had consented to be tested and carried a heterozygous, GenBank: NM_000410.3 (c.845G>A [p.Cys282Tyr]) variant in *HFE*. Additionally, well-described, common variants (such as GenBank: NM_004004.5;

c.109G>A [p.Val37Ile] in *GJB2* [MIM: 121011] that has a frequency of 7.2% in the East Asian population) that met the criteria for pathogenicity albeit with high prevalence (>5%) in a specific ethnic population were retained. Based on the 5% threshold for filtering variants, approximately 98% of all variants were classified as benign.

For each participant, an average of four missense variants were identified that had a population frequency below the disease allele frequency for that respective gene, but with no published records in the medical literature at the time of analysis. These novel missense variants were classified as variants of uncertain significance (VUS) and were not reported.

We reported a total of 304 variants in this study. Approximately 92% (280/304) of variants passed the NGS variant quality filter (described in Subjects and Methods) and were confirmed by Sanger sequencing (CE), while 8.5% (26/304) of the variants did not pass the quality filter by NGS but met the criteria for pathogenicity and were confirmed by CE. In contrast, three variants were labeled as false positive calls from NGS-based analysis. These variants, which included two SNVs detected by NGS analysis pipeline and one CNV detected by the CNV analysis software, satisfied the criteria for pathogenicity but were not confirmed by CE and exon-centric aCGH, respectively, and therefore not reported.

Carrier Status and Types of Variants Reported

Among the 202 participants' samples analyzed, 78% received at least one positive carrier result. The average number of variants was 1.5 per individual, with a range of 0 to 5 variants and a mode of 1 variant per individual. We have reported all types of variants (Table S1) with known missense variants comprising the majority of a specific type of variant (48%). Table S2 lists all variants reported in this study. While most variants were reported within coding regions of the genes or affecting the canonical splice-site at the intron/exon junctions, we have reported five well-known pathogenic, non-coding variants that were not in the canonical splice site (Table 1).

Table 2. Classification of Putative Splice Variants before and after mRNA Analysis

Gene	Variant	Disorder	Condition Category	Variant Category	Initial Classification	Final Classification
CEP290 (MIM: 610142)	NM_025114.3 (c.6645+1G>A)	ciliopathies (MIM: 615991, 610188, 611755, 611134, 610189)	lifespan limiting	novel	likely pathogenic	likely pathogenic
CEP290 (MIM: 610142)	NM_025114.3 (c.6818+1_6818+2insGG)	ciliopathies (MIM: 615991, 610188, 611755, 611134, 610189)	lifespan limiting	novel	likely pathogenic	pathogenic
ERCC2 (MIM: 126340)	NM_000400.3 (c.594+2_594+5delTGAG)	trichothiodystrophy (MIM: 601675)	serious	novel	likely pathogenic	pathogenic
LRPPRC (MIM: 607544)	NM_133259.3 (c.469+1G>A)	Leigh syndrome, French-Canadian type (MIM: 220111)	lifespan limiting	novel	likely pathogenic	likely pathogenic
FAH (MIM: 613871)	NM_000137.2 (c.81+2T>A)	tyrosinemia type I (MIM: 276700)	serious	novel	likely pathogenic	VUS[a]

These genes are expressed in leukocytes. All individuals were heterozygotes.

[a]The assay results did not indicate a splicing defect in *FAH*; however, allele drop-out analysis was not performed, and further confirmation is necessary to determine the functional consequence of the GenBank: NM_000137.2 (c.81+2T>A) variant.

Approximately 64% of variants (195 of 304 total variants reported) were "distinct" (i.e., every reported variant counted only once). Within this distinct category, approximately 22% (44/195) were novel variants. The majority of the novel variants were classified as likely pathogenic, with the exception of some novel splice-site variants (Table 2) and CNVs (Table 3). For the genes that were expressed in blood (listed in Table 2), RNA-based splice-site analysis was performed on novel variants that were initially classified as likely pathogenic. RNA analysis was used as evidence to re-interpret the classification of the variant to determine whether the variant caused an in-frame exon skipping, out-of-frame exon skipping, intron retention, or no splicing defect. Thus, two variants were upgraded to pathogenic and one was downgraded to a VUS (Table 2).

Five individuals were identified as silent carriers for α-thalassemia (MIM: 604131); however, only four samples could be confirmed by multiplex PCR because there was no DNA available to confirm the fifth sample. Additionally, five other individuals harbored a CNV in a gene on either the carrier list or the medically actionable list (Table 3). Therefore, approximately 5% (10/202) of all participants were carriers of at least one clinically significant CNV.

Condition Categories Reported

The reported conditions comprised approximately 18% (134/728) of all the carrier gene-disorder pairs[6] analyzed in this study. Due to the small population size (n = 202) and ethnicity bias (78% were of European descent), there were variants and conditions that were observed multiple times (Table 4). For some disorders, the calculated carrier frequency in this study was higher than the estimated carrier frequency in the general population (Table 4). As expected, the common variants for *HFE*-associated hereditary hemochromatosis (MIM: 235200) comprised approximately 20% (40/202) of all heterozygotes in the study and 13% (40/304) of all variants reported.

Carrier Couples

Each couple's data were analyzed to determine whether they were carriers for the same condition as their respective partner. Not surprisingly, we identified carrier couples for a few common conditions: eight carrier couples for hereditary hemochromatosis (MIM: 235200), with each partner a carrier of either GenBank: NM_000410.3 (c.845G>A [p.Cys282Tyr]) or GenBank: NM_000410.3 (c.187C>G [p.His63Asp]), two couples for alpha-1 anti-trypsin deficiency (MIM: 613490), and one couple each for Factor V Leiden (MIM: 227400) and non-syndromic hearing loss ([MIM: 220290]; *GJB2*). We did not identify a carrier couple for any rare autosomal-recessive disorder. Three females were carriers for X-linked conditions, of which one was hemophilia A (MIM: 306700), a serious condition; the other two conditions were categorized as mild conditions (Table 5).

Medically Actionable (Secondary)

Based on our expanded medically actionable gene list (148 genes), additional findings were reported in 3.5% of participants (7 of 202); however, considering only the updated ACMG v2.0 list for secondary findings (59 genes),[11] the proportion of participants with such findings would be 2.9% (Table 6). The difference is due to the absence of *SERPINA1* (MIM: 107400) on the ACMG list. Although two participants were found to be compound heterozygotes in *HFE* (GenBank: NM_000410.3; c.845G>A [p.Cys282Tyr]; GenBank: NM_000410.3; c.187C>G [p.His63Asp]) for hereditary hemochromatosis, only homozygotes for the GenBank: NM_000410.3 (c.845G>A [p.Cys282Tyr]) variant in *HFE* would have been considered medically actionable. This study did not identify any individual who was homozygous for GenBank: NM_000410.3 (c.845G>A [p.Cys282Tyr]). In addition, we identified a male with a pathogenic variant, GenBank: NM_000059.3 (c.4965C>G [p.Tyr1655Ter]), in *BRCA2* (MIM: 600185). This variant was considered a carrier finding for Fanconi anemia (MIM: 605724) as well as a medically actionable

Table 3. CNVs Identified by NGS in Carriers and Confirmed by an Orthogonal Method

Gene	Deletion Variant	Disorder	Condition Category	Variant Category	Classification	No. of Heterozygotes
HBA2 (MIM: 141850)	whole gene (NM_000517.4)	alpha thalassemia (MIM: 604131)	carrier list: lifespan limiting	known	pathogenic	5[a]
FANCA (MIM: 607139)	exons 18–28 (NM_000135.3)	Fanconi anemia (MIM: 227650)	carrier list: serious	novel	likely pathogenic	1
TBCE (MIM: 604934)	exons 3–4 (NM_001079515.2)	hypoparathyroidism-retardation-dysmorphism-syndrome (MIM: 241410)	carrier list: serious	novel	pathogenic	1
INVS (MIM: 243305)	5′ UTR, exons 1 and 2 (NM_014425.4)	nephronophthisis 2 (MIM: 602088)	carrier list: lifespan limiting	novel	likely pathogenic	1
PROM1 (MIM: 604365)	5′ UTR and exon 1 (NM_006017.2)	retinitis pigmentosa 41 (MIM: 612095)	carrier list: mild	novel	likely pathogenic	1
BRCA1 (MIM: 113705)	5′ UTR exons 1–11 (NM_007300.3)	hereditary breast and ovarian cancer (MIM: 604370)	medically actionable	known	pathogenic	1

Variants were confirmed by high-density microarray, except the whole gene deletion of *HBA2*, which was confirmed by multiplex PCR (see Subjects and Methods).
[a]One variant was not confirmed because additional DNA was unavailable for confirmatory testing.

finding for an increased risk of male breast and prostate cancer (MIM: 114480 and 176807, respectively).

Moreover, we identified an incidental finding in an individual with mild hearing loss who harbored two variants, GenBank: NM_004004.5 (c.35delG [p.Gly12Valfs*2]; c.101T>C [p.Met34Thr]) in *GJB2* (MIM: 121011) for nonsyndromic hearing loss and deafness (DFNB1A [MIM: 220290]) (Table 6). This finding was not included in the medically actionable findings.

Discussion

One of the goals of this exploratory study was to learn more about the clinical utility of using GS for carrier screening in a clinical setting. To this end, a broad selection of gene/disorder pairs that would impact carrier status was analyzed. GS coupled with this large selection of gene/disorder pairs allowed us to increase the sensitivity of capturing most clinically significant variants. Recent studies have highlighted the advantages[41] and controversies[42] surrounding expanded carrier screening using an NGS-based approach. It is expected, as observed with this study, that clinically significant variants for nonserious and reduced penetrance, adult-onset conditions will be detected with high frequency. While the inclusion of mild or reduced penetrant conditions (hereditary hemochromatosis and factor V Leiden) may not be considered appropriate for clinical carrier screening, from a research perspective, it posits an unbiased approach to gather informative data on carrier status while offering autonomy of the participant's choices. As shown in our study, most participants requested results for all condition categories after appropriate genetic counseling.

In this study, the male partner of female carriers of pathogenic and likely pathogenic variants were invited to join the GS arm of the study. Again, only pathogenic and likely pathogenic variants in the genes reported for the female participant were reported to the male partner. This was designed to avoid prenatal diagnosis based on VUS results.

The NGS technology is advancing rapidly. Additionally, there is a simultaneous effort to improve and standardize the variant interpretation process.[37,43] To our knowledge, there are currently very few studies reported to use the 2015-ACMG guidelines for variant interpretation to analyze genomic data in individuals with no clinical phenotype. For novel variants that predicted a null effect in genes where loss-of-function is an established mechanism of disease, it was challenging to predict genotype-phenotype correlation. To add to the complexity of classifying novel variants in the absence of phenotype, it is also challenging to classify novel variants in a gene that is associated with clinical heterogeneity. For example, one participant was a carrier for a novel variant, GenBank: NM_020366.3 (c.1116delA [p.Lys372Asnfs*3]) in *RPGRIP1* (MIM: 605446). *RPGRIP1* pathogenic variants are associated with both Leber congenital amaurosis (MIM: 613826) and cone-rod dystrophy 3 (MIM: 608194), but it was not possible to predict for which condition the participant was a carrier. Thus, in our experience, the ability to accurately classify variants and predict outcomes is more challenging in a healthy population than in an affected individual and is less robust than in individuals presenting with an adverse phenotype.[44] While evidence based on phenotype is not a strong consideration in favor of pathogenicity according to the ACMG variant interpretation guidelines, sometimes highly specific phenotypic information does provide important evidence for interpreting variants associated with single-gene disorders.

Our finding of a variant in an X-linked condition, which was particularly impactful for a pregnancy outcome, illustrates the advantage of phenotypic information

Table 4. Genes with Variants Reported as Pathogenic or Likely Pathogenic More than Once

Gene	Variants Reported	Disorder	Condition Category	Variant Category	Classification	No. of heterozygotes (n = 202)	Frequency of Disorder in Study (n = 202)	Carrier Frequency (%)[a]
HFE (MIM: 613609)	NM_000410.3 (c.845G>A [p.Cys282Tyr])	hereditary hemochromatosis (MIM: 235200)	adult onset	Known	pathogenic	30	20%	6%–13%
	NM_000410.3 (c.187C>G [p.His63Asp])[b]			Known	pathogenic	10		
GJB2 (MIM: 121011)	NM_004004.5 (c.35delG [p.Gly12Valfs*2])	nonsyndromic, hearing loss (MIM: 220290)	mild	Known	pathogenic	3	9%	2.3%
	NM_004004.5 (c.35dupG [p.Val13Cysfs])			Known	pathogenic	1		
	NM_004004.5 (c.109G>A [p.Val37Ile])			Known	likely pathogenic	4		
	NM_004004.5 (c.−23+1G>A)			Known	pathogenic	1		
	NM_004004.5 (c. 101T>C [p.Met34Thr])			Known	likely pathogenic	4		
	NM_004004.5 (c.269T>C [p.Leu90Pro])			Known	pathogenic	3		
	NM_004004.5 (c.416G>A [p.Ser139Asn])			Known	likely pathogenic	1		
	NM_004004.5 (c.−22−2A>C)			Known	pathogenic	1		
F5 (MIM: 612309)	NM_000130.4 (c.1601G>A [p.Arg534Gln])	factor V Leiden thrombophilia (MIM: 227400)	unpredictable	known	pathogenic	17	8%	3%–8%
SERPINA1 (MIM: 107400)	NM_001127700.1 (c.1096G>A [p.Glu366Lys])	alpha-1 antitrypsin deficiency (MIM: 613490)	adult onset	known	pathogenic	6	6%	2.4%–4.8%
	NM_001127700.1 (c.863A>T [p.Glu288Val])			known	pathogenic	6		
ABCA4 (MIM: 601691)	NM_000350.2 (c.6089G>A [p.Arg2030Gln])	cone rod dystrophy 3 (MIM: 604116); Stargardt disease (MIM: 248200)	mild	known	pathogenic	1	4%	2%
	NM_000350.2 (c.1964T>G [p.Phe655Cys])			known	likely pathogenic	2		
	NM_000350.2 (c.4139C>T [p.Pro1380Leu])			known	pathogenic	1		
	NM_000350.2 (c.2588G>C [p.Gly863Ala])			known	likely pathogenic	2		
	NM_000350.2 (c.5882G>A [p.Gly1961Glu])			known	pathogenic	3		

Table 4. **Continued**

Gene	Variants Reported	Disorder	Condition Category	Variant Category	Classification	No. of heterozygotes (n = 202)	Frequency of Disorder in Study (n = 202)	Carrier Frequency (%)[a]
CYP21A2[c] (MIM: 613815)	NM_000500.8 (c.1360C>T [p.Pro454Ser])	congenital adrenal hyperplasia (MIM: 201910)	serious	known	pathogenic	3	4%[d]	1.6%–6%[e]
	NM_000500.8 (c.844G>T [p.Val282Leu])			known	pathogenic	3		
	NM_000500.8 (c.955C>T [p.Gln319Ter])			known	pathogenic	2		
CFTR (MIM: 602421)	NM_000492.3 (c.1521_1523delCTT [p.Phe508delPhe])	cystic fibrosis (MIM: 219700)	serious	known	pathogenic	6	3%	4%
SPG7 (MIM: 602783)	NM_003119.3 (c.1529C>T [(p.Ala510Val)])	spastic paraplegia 7 (MIM: 607259)	adult onset	known	pathogenic	3	2%	0.9%–1.5%
	NM_003119.3 (c.1045G>A [p.Gly349Ser])			known	pathogenic	2		

All variants use the HGVS nomenclature.

[a]Source: GeneReviews or Genetics Home Reference.

[b]The H63D variant in *HFE* was analyzed and reported only if the partner was a carrier of GenBank: NM_000410.3 (c.845G>A [p.Cys282Tyr]).

[c]The coding DNA nomenclature is used instead of the protein sequence nomenclature to avoid the discrepancy associated with the latter when using the hg19 genomic reference sequence; the coding DNA is also used for nomenclature for splice/intronic variants; one participant was homozygous for the *SERPINA1* Gen-Bank: NM_001127700.1 (c.1096G>A [p.Glu366Lys]) allele and was excluded from the table.

[d]Combined classic and non-classic with 3% carrying a variant for the non-classic form and 1% carrying a variant for the classic form.

[e]Combined frequency for classic and non-classic form; most of these variants were observed more than once.

for variant interpretation. In a female participant, we reported a heterozygous, nonsense variant in *F8* (MIM: 300841), which is associated with hemophilia A (MIM: 306700). This variant was novel and the limited evidence satisfied the criteria for only a likely pathogenic classification. The participant, who was already pregnant at the time of receiving this result, opted for prenatal testing. Her male fetus was found to harbor this variant, and subsequently, the newborn developed a complication, an acute subgaleal hemorrhage that is associated with the severe form hemophilia A. The prior knowledge that this infant was suspected to be affected with hemophilia A helped guide the immediate treatment plan (red cell and platelet transfusions with anti-hemophilic factor/von Willebrand factor complex) to avoid a fatal outcome. The additional evidence on phenotype prompted variant re-classification to pathogenic. Being novel, this variant would not have been detected on a targeted mutation panel. Overall, we conclude that the 2015-ACMG variant interpretation guideline is a powerful tool for systematic and organized classification for rare and novel variants that are detected by GS.

The variant classification process is continually evolving and this may explain the discrepancies in the carrier status results between our study and those from a previous study by Bell et al.[45] In the latter study, the authors focused on 448 recessive disorders in 104 individuals and reported a carrier burden of 2.8 per individual. In contrast, we analyzed more gene/disorder pairs (728) but reported 1.5 clinically significant variants per individual. The databases used for variant interpretation in the Bell et al. study[45] were limited compared with the currently available resources and the criteria for selecting pathogenicity of variants were also different. Of note, more than 70% (76/104) of individuals in the Bell et al. study[45] were either affected or known carriers of severe pediatric conditions, but in our study, only 3% of participants were previously known to be carriers for cystic fibrosis (MIM: 219700).

We also identified 3.5% of participants with a variant in a gene on our medically actionable list. While our list of genes (148 genes) was more extensive than that recommended by ACMG[11,46] (59 genes), all of the medically actionable findings detected in this study were in genes on the ACMG list, except *SERPINA1*. This result is consistent with that of a previous report[7] that used an expanded medically actionable findings list. Considering only the ACMG list (i.e., excluding *SERPINA1*), our rate of secondary findings also remains consistent with that of another study on 2,000 exomes.[47]

It is important to note that the lack of ethnic diversity and the small size of our individual cohort represented some limitations to our study. Although the study was designed to be offered to a pan-ethnic population, the data from this study were biased toward the fact that the majority of our participants were of European descent. Furthermore, data from 202 participants have limited

Table 5. Carrier Females for X-Linked Disorders

Gene	Variant	Disorder	Condition Category	Variant Category	Classification
F8 (MIM: 300841)	NM_000132.3 (c.3144G>A [p.Trp1048Ter])	hemophilia A (MIM: 306700)	serious	novel	pathogenic
TRAPPC2 (MIM: 300202)	NM_001128835.2 (c.12G>A [p.Trp4Ter])	X-linked spondyloepiphyseal dysplasia tarda (MIM: 313400)	mild	novel	likely pathogenic
G6PD (MIM: 305900)	NM_001042351.2 (c.376A>G [p.Asn126Asp]); NM_001042351.2 (c.202G>A[a] [Val68Met])	G6PD deficiency (MIM: 300908)	mild	known	pathogenic

[a]The A-haplotype in *G6PD* comprises two variants, GenBank: NM_001042351.2 (c.376A>G [p.Asn126Asp]) and GenBank: NM_001042351.2 (c.202G>A [p.Val68Met]) that are present in *cis*.

statistical power for analysis of carrier frequency. The small sample size may be a factor in the observed carrier frequency of certain conditions above the expected value. For example, in *SPG7* (MIM: 602783), which is associated with spastic paraplegia 7 (MIM: 607259) that has a reported prevalence of 2–6:100,000 (GeneReviews in Web Resources), the GenBank: NM_003119.3 (c.1529C>T [p.Ala510Val]) variant has an allele frequency of 0.0025 in the ExAC database; however, in our study, it was observed three times. While the frequency in our study suggested a variant of uncertain significance, we classified it as pathogenic based on other published evidence for its pathogenicity.[43]

Compared with mutation screening panels that were traditionally designed to target certain ethnic populations, NGS technologies are much better at detecting rare and novel pathogenic variants in a pan-ethnic population.

Some of the current drawbacks of GS include the inability to detect mosaicism because of low read depth, high data storage, and generating a large number of VUSs. However, within existing NGS platforms, GS can address some of the limitations of a capture or amplicon-based sequencing approach. GS can bypass the disadvantages of PCR-based library preparation, which can be a source of introducing variant artifacts and PCR biases resulting in a non-uniform representation of the DNA library. It is a superior method for determining structural variation in the genome because the exact breakpoint in the DNA sequence can be identified. Finally, from a cost perspective, it can overcome the continuous need to re-design and validate clinical targeted gene panels when new pathogenic variants are identified in non-coding regions.

Current clinical NGS applications still do not have the sensitivity and specificity to detect all types of variants,

Table 6. Secondary and Incidental Findings in Participants

Gene	Variant	Disorder	Condition Category	Variant Category	Classification
BRCA1 (MIM: 113705)	NM_007300.3 (c.2071delA [p.Arg691Aspfs*10])	hereditary breast and ovarian cancer (MIM: 604370, 612555)	medically actionable	known	pathogenic
	NM_007300.3 (c.3485delA [p.Asp1162Valfs*48])			known	pathogenic
	NM_007300.3 (c.(?_-30)_ (4185+1_4186-1)del)			known	pathogenic
BRCA2[a] (MIM: 600185)	NM_000059.3 (c.4965C>G [p.Tyr1655Ter])		medically actionable and serious (carrier)	known	pathogenic
SERPINA1[b] (MIM: 107400)	NM_001127700.1 (c.1096G>A [p.Glu366Lys])	alpha-1 antitrypsin deficiency (MIM: 613490)	medically actionable	known	pathogenic
APC (MIM: 611731)	NM_000038.5 (c.1042C>T [p.Arg348Ter])	*APC*-associated polyposis conditions (MIM: 175100)	medically actionable	known	pathogenic
GJB2[c] (MIM: 121011)	NM_004004.5 (c.35delG [p.Gly12Valfs*2]); NM_004004.5 (c.101T>C [Met34Thr])	nonsyndromic hereditary hearing loss (DFNB1) (MIM: 220290)	carrier list: mild	known	pathogenic
SMAD3 (MIM: 603109)	NM_005902.3 (c.484G>T [p.Glu162Ter])	Loeys-Dietz syndrome type 3 1C (MIM: 613795)	medically actionable	novel	likely pathogenic

These results are based on screening 130 female and 69 male participants; 3 individuals did not opt-in to receive these results.
[a]Carrier for Fanconi anemia.
[b]Homozygous GenBank: NM_001127700.1 (c.1096G>A [p.Glu366Lys]) (PI*ZZ).
[c]Incidental finding for the GenBank: NM_004004.5 (c.35delG [p.Gly12Valfs*2]); NM_004004.5 (c.101T>C [p.Met34Thr]) variants in one individual presumed to be in *trans*.

such as those causing triplet repeat disorders (e.g., fragile X [MIM: 300624]) and regions of the genome with high homology (pseudogenes). Due to the inherent limitations of the short-read approach in NGS platforms that are currently utilized in most clinical laboratories, multiple methodologies would need to be used to support detection of the full range of variant classes. However, these challenges will be short-lived because the implementation of long-read DNA sequencing (third generation) technologies[48] coupled with advances in bioinformatic pipelines for detecting short tandem repeats[49] and copy number variation[50] is imminent in the clinical laboratory. NGS is a paradigm shift in the rate at which carrier status is determined for several hundred disorders simultaneously. It may soon replace other methodologies and become a unifying platform for performing most molecular genetic tests.

Accession Numbers

The accession number for the sequence data reported in this paper is dbGap: phs00927.

Supplemental Data

Supplemental Data include one figure, two tables, and Supplemental Note (clinical report) and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.04.004.

Acknowledgments

Web Resources

ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/
dbGaP, https://www.ncbi.nlm.nih.gov/gap
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
GeneReviews, Casari, G., and Marconi, R. (1993). Spastic Paraplegia 7. https://www.ncbi.nlm.nih.gov/books/NBK1107/
OMIM, http://www.omim.org/
SeattleSeq Annotation 138, http://snp.gs.washington.edu/SeattleSeqAnnotation138/

References

1. Prior, T.W. (2014). Next-generation carrier screening: are we ready? Genome Med. 6, 62.
2. Azimi, M., Schmaus, K., Greger, V., Neitzel, D., Rochelle, R., and Dinh, T. (2016). Carrier screening by next-generation sequencing: health benefits and cost effectiveness. Mol. Genet. Genomic Med. 4, 292–302.
3. Nazareth, S.B., Lazarin, G.A., and Goldberg, J.D. (2015). Changing trends in carrier screening for genetic disease in the United States. Prenat. Diagn. 35, 931–935.
4. Arjunan, A., Litwack, K., Collins, N., and Charrow, J. (2016). Carrier screening in the era of expanding genetic technology. Genet. Med. 18, 1214–1217.
5. Edwards, J.G., Feldman, G., Goldberg, J., Gregg, A.R., Norton, M.E., Rose, N.C., Schneider, A., Stoll, K., Wapner, R., and Watson, M.S. (2015). Expanded carrier screening in reproductive medicine-points to consider: a joint statement of the American College of Medical Genetics and Genomics, American College of Obstetricians and Gynecologists, National Society of Genetic Counselors, Perinatal Quality Foundation, and Society for Maternal-Fetal Medicine. Obstet. Gynecol. 125, 653–662.
6. Himes, P., Kauffman, T.L., Muessig, K.R., Amendola, L.M., Berg, J.S., Dorschner, M.O., Gilmore, M., Nickerson, D.A., Reiss, J.A., Richards, C.S., et al. (2017). Genome sequencing and carrier testing: decisions on categorization and whether to disclose results of carrier testing. Genet. Med. 19, 803–808.
7. Dorschner, M.O., Amendola, L.M., Turner, E.H., Robertson, P.D., Shirts, B.H., Gallego, C.J., Bennett, R.L., Jones, K.L., Tokita, M.J., Bennett, J.T., et al.; National Heart, Lung, and Blood Institute Grand Opportunity Exome Sequencing Project (2013). Actionable, pathogenic incidental findings in 1,000 participants' exomes. Am. J. Hum. Genet. 93, 631–640.
8. Amendola, L.M., Dorschner, M.O., Robertson, P.D., Salama, J.S., Hart, R., Shirts, B.H., Murray, M.L., Tokita, M.J., Gallego, C.J., Kim, D.S., et al. (2015). Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res. 25, 305–315.
9. Kauffman, T.L., Wilfond, B.S., Jarvik, G.P., Leo, M.C., Lynch, F.L., Reiss, J.A., Richards, C.S., McMullen, C., Nickerson, D., Dorschner, M.O., and Goddard, K.A. (2017). Design of a randomized controlled trial for genomic carrier screening in healthy patients seeking preconception genetic testing. Contemp. Clin. Trials 53, 100–105.
10. Tan, N., Amendola, L.M., O'Daniel, J.M., Burt, A., Horike-Pyne, M.J., Boshe, L., Henderson, G.E., Rini, C., Roche, M.I., Hisama, F.M., et al. (2017). Is "incidental finding" the best term?: a study of patients' preferences. Genet. Med. 19, 176–181.
11. Kalia, S.S., Adelman, K., Bale, S.J., Chung, W.K., Eng, C., Evans, J.P., Herman, G.E., Hufnagel, S.B., Klein, T.E., Korf, B.R., et al. (2017). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet. Med. 19, 249–255.
12. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595.
13. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.
14. Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461, 272–276.

15. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311.

16. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.

17. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput. Biol. *6*, e1001025.

18. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.

19. Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. Science *185*, 862–864.

20. Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447–D452.

21. Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*, D68–D73.

22. Lam, D.W., and LeRoith, D. (2012). The worldwide diabetes epidemic. Curr. Opin. Endocrinol. Diabetes Obes. *19*, 93–96.

23. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

24. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

25. Ng, P.C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. Genome Res. *11*, 863–874.

26. Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K., Hua, Y., Gueroussov, S., Najafabadi, H.S., Hughes, T.R., et al. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. Science *347*, 1254806.

27. Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. *44* (D1), D862–D868.

28. Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., Shiel, J.A., Thomas, N.S., Abeysinghe, S., Krawczak, M., and Cooper, D.N. (2003). Human Gene Mutation Database (HGMD): 2003 update. Hum. Mutat. *21*, 577–581.

29. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

30. Ruark, E., Münz, M., Renwick, A., Clarke, M., Ramsay, E., Hanks, S., Mahamdallie, S., Elliott, A., Seal, S., Strydom, A., et al. (2015). The ICR1000 UK exome series: a resource of gene variation in an outbred population. F1000Res. *4*, 883.

31. den Dunnen, J.T., Dalgleish, R., Maglott, D.R., Hart, R.K., Greenblatt, M.S., McGowan-Jordan, J., Roux, A.F., Smith, T., Antonarakis, S.E., and Taschner, P.E. (2016). HGVS recommendations for the description of sequence variants: 2016 Update. Hum. Mutat. *37*, 564–569.

32. Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. *15*, R84.

33. Abyzov, A., Urban, A.E., Snyder, M., and Gerstein, M. (2011). CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. *21*, 974–984.

34. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

35. MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. *42*, D986–D992.

36. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am. J. Hum. Genet. *84*, 524–533.

37. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al.; ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424.

38. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

39. Liu, Y.T., Old, J.M., Miles, K., Fisher, C.A., Weatherall, D.J., and Clegg, J.B. (2000). Rapid detection of alpha-thalassaemia deletions and alpha-globin gene triplication by multiplex polymerase chain reactions. Br. J. Haematol. *108*, 295–299.

40. Feng, Y., Ge, X., Meng, L., Scull, J., Li, J., Tian, X., Zhang, T., Jin, W., Cheng, H., Wang, X., et al. (2017). The next generation of population-based spinal muscular atrophy carrier screening: comprehensive pan-ethnic SMN1 copy-number and sequence variant analysis by massively parallel sequencing. Genet. Med. *19*, 936–944.

41. van der Hout, S., Holtkamp, K.C., Henneman, L., de Wert, G., and Dondorp, W.J. (2016). Advantages of expanded universal carrier screening: what is at stake? Eur. J. Hum. Genet. *25*, 17–21.

42. Lazarin, G.A., and Goldberg, J.D. (2016). Current controversies in traditional and expanded carrier screening. Curr. Opin. Obstet. Gynecol. *28*, 136–141.

43. Amendola, L.M., Jarvik, G.P., Leo, M.C., McLaughlin, H.M., Akkari, Y., Amaral, M.D., Berg, J.S., Biswas, S., Bowling, K.M., Conlin, L.K., et al. (2016). Performance of ACMG-AMP variant-interpretation guidelines among nine laboratories in the clinical sequencing exploratory research consortium. Am. J. Hum. Genet. *98*, 1067–1076.

44. Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. *2012*, 251364.

45. Bell, C.J., Dinwiddie, D.L., Miller, N.A., Hateley, S.L., Ganusova, E.E., Mudge, J., Langley, R.J., Zhang, L., Lee, C.C., Schilkey, F.D., et al. (2011). Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci. Transl. Med. *3*, 65ra4.

46. Green, R.C., Berg, J.S., Grody, W.W., Kalia, S.S., Korf, B.R., Martin, C.L., McGuire, A.L., Nussbaum, R.L., O'Daniel, J.M., Ormond, K.E., et al.; American College of Medical Genetics and Genomics (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet. Med. *15*, 565–574.

47. Yang, Y., Muzny, D.M., Xia, F., Niu, Z., Person, R., Ding, Y., Ward, P., Braxton, A., Wang, M., Buhay, C., et al. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. JAMA *312*, 1870–1879.

48. Nakano, K., Shiroma, A., Shimoji, M., Tamotsu, H., Ashimine, N., Ohki, S., Shinzato, M., Minami, M., Nakanishi, T., Teruya, K., et al. (2017). Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. Hum. Cell *30*, 149–161.

49. Fungtammasan, A., Ananda, G., Hile, S.E., Su, M.S., Sun, C., Harris, R., Medvedev, P., Eckert, K., and Makova, K.D. (2015). Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. Genome Res. *25*, 736–749.

50. Hehir-Kwa, J.Y., Pfundt, R., and Veltman, J.A. (2015). Exome sequencing and whole genome sequencing for the detection of copy number variation. Expert Rev. Mol. Diagn. *15*, 1023–1032.