Published in final edited form as:

Mol Genet Metab. 2010 June ; 100(2): 184–192. doi:10.1016/j.ymgme.2010.03.004.

Next Generation Sequencing in Research and Diagnostics of Ocular Birth Defects

Gordana Raca^{1,2}, Craig Jackson¹, Berta Warman⁴, Tom Bair³, and Lisa A. Schimmenti⁴

¹UW Cytogenetic Services, Wisconsin State Laboratory of Hygiene

²Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison

³The University of Iowa DNA Facility, Iowa City, IA

⁴Department of Pediatrics, Division of Genetics and Metabolism, University of Minnesota

Abstract

Sequence capture enrichment (SCE) strategies and massively parallel next generation sequencing (NGS) are expected to increase the rate of gene discovery for genetically heterogeneous hereditary diseases, but at present, there are very few examples of successful application of these technologic advances in translational research and clinical testing. Our study assessed whether array based target enrichment followed by re-sequencing on the Roche Genome Sequencer FLX (GS FLX) system could be used for novel mutation identification in more than 1000 exons representing 100 candidate genes for ocular birth defects, and as a control, whether these methods could detect two known mutations in the *PAX2* gene. We assayed two samples with heterozygous sequence changes in *PAX2* that were previously identified by conventional Sanger sequencing. These changes were a c.527G>C (S176T) substitution and a single basepair deletion c.77delG. The nucleotide substitution c.527G>C was easily identified by NGS. A deletion of one base in a long polyG stretch (c.77delG) was not registered initially by the GS Reference Mapper, but was detected in repeated analysis using two different software packages. Different approaches were evaluated for distinguishing false positives (sequencing errors) and benign polymorphisms from potentially pathogenic sequence changes that require further follow-up. Although improvements will be necessary in accuracy, speed, ease of data analysis and cost, our study confirms that NGS can be used in research and diagnostic settings to screen for mutations in hundreds of loci in genetically heterogeneous human diseases.

Keywords

next generation sequencing; sequence capture; GS FLX; anophthalmia; microphthalmia; coloboma

INTRODUCTION

The goal of our study was to evaluate the capacity of array based sequence capture target enrichment (SCE) and massively parallel, next generation sequencing (NGS) to successfully

© 2009 Elsevier Inc. All rights reserved.

Corresponding Author: Gordana Raca, MD, PhD, FACMG, UW Cytogenetics Service, Wisconsin State Laboratory of Hygiene, 465 Henry Mall, Madison, WI 53706, Tel: 608-890-1113; FAX: 608-265-7818, racago@mail.slh.wisc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

identify mutations in candidate genes for the developmental ocular birth defects anophthalmia, microphthalmia and coloboma.

The advent of NGS technologies is expected to transform the practice of medical genetics [1–3]. With the high throughput and decreased sequencing costs achieved by NGS, it is no longer impossible to sequence hundreds or even thousands of exons and other genomic sequences in an individual with a suspected genetic disease. It is predicted that in the near future NGS might replace array based techniques and Sanger sequencing in their current clinical applications for the detection of mutations [1,3]. Additionally, NGS provides entirely new research and diagnostic capabilities, including whole genome screening for novel mutations and sequencing biological specimens for the genomic signature of novel infectious agents [4,5].

NGS could be particularly advantageous in research and testing for genetically heterogeneous hereditary conditions. Common disorders evaluated by clinical geneticists are caused by heterogeneous Mendelian loci and lend themselves to enrichment strategies followed by NGS. Examples include intellectual disability [6], deafness [7], familial cardiomyopathy [8] and retinitis pigmentosa [9]. In these conditions there are often very subtle phenotypic differences between affected patients to guide molecular diagnostics by indicating which gene is likely to be mutated in a particular individual. Current diagnostic evaluation proceeds by sequencing a series of genes, individually or in small sets, based on the relative frequency of the mutations and the sensitivity of available assays. If there is no predominant mutation(s) causing the disease, the pathogenic change often remains unknown even after very extensive and expensive molecular testing. With enrichment strategies followed by NGS, sequencing of all genes implicated in a particular genetic disorder could be performed simultaneously, efficiently and at low cost.

While clearly superior to traditional Sanger sequencing, NGS has had little impact on clinical testing to date. There are very few examples of successful application of NGS in translational research and diagnostics. Clinical testing using NGS is currently offered for Hypertrophic Cardiomyopathy (HCM), Dilated Cardiomyopathy (DCM) and Long QT syndrome (<http://www.genedx.com/>). NGS has also been explored as a method to perform rapid human leukocyte antigen (HLA) typing for high resolution allele identification [10,11], and to develop assays for Neurofibromatosis Type 1 [12], autosomal recessive ataxia [13] and mitochondrial disorders caused by mutations in the mitochondrial genome and 362 nuclear genes controlling mitochondrial function [14]. Ng et al. applied targeted sequencing of all coding regions (“exome”) to show the presence of causative mutations in four unrelated individuals with a rare dominantly inherited disorder, Freeman-Sheldon Syndrome (FSS) [15] and to discover the gene for a rare recessive disorder of previously unknown cause, Miller Disease [16]. Exciting applications have also been described in cancer research, where NGS has been applied in discovering new candidate genes for acute myeloid leukemia [17,18], glioblastoma multiforme [19] and other malignancies [20].

We describe the first study which showed the feasibility of using genomic enrichment by sequence capture followed by NGS to investigate genetic causes of the ocular birth defects anophthalmia, microphthalmia and coloboma. These eye anomalies are among the most prevalent causes of childhood blindness, affecting annually ~2 per 10,000 newborns worldwide [21]. Although they can be of different origins, the majority are caused by defects in genes which regulate normal eye development [22–25]. There is increased evidence that mutations in large numbers (possibly hundreds) of different genes can cause congenital eye malformations, but no single gene is responsible for a high percentage of cases [23–25]. Anophthalmia, microphthalmia and coloboma therefore represent disorders where

simultaneous sequencing of large numbers of candidate genes by NGS is an ideal approach to study genetic causes and demonstrate the feasibility of NGS for clinical diagnostics.

We showed that the combination of array based SCE with re-sequencing on the GS FLX instrument using Titanium chemistry allows concurrent sequencing of more than 100 candidate genes for anophthalmia, microphthalmia and coloboma. However, improvements will be necessary in several areas including accuracy, speed, ease of data analysis and cost to allow successful diagnostic implementation of NGS for simultaneous mutation testing in hundreds of genes in genetically heterogeneous human diseases.

MATERIALS AND METHODS

We tested whether two known sequence changes in the renal-coloboma syndrome (a.k.a. Papillorenal syndrome, OMIM #120330) associated gene *PAX2*, which were previously characterized by Sanger sequencing, can reliably be detected by NGS. The first variant was a missense change in exon 5, c.527G>C, which resulted in serine to threonine amino-acid change S176T. This base substitution was identified in one of our previous studies in a father of a proband with ocular birth defects, but not in his affected child. Since detailed clinical information for the parent was not available, the change was described as a variant of unknown clinical significance.

Since short stretches of mono-, di-, and trinucleotide repeats represent hotspots for disease causing frameshift mutations in genomic DNA, we wanted to determine if this mutation type is detectable by NGS. Therefore, the second sequence change selected for the study was a deletion of one base in a polyG stretch in exon 2 of the *PAX2* gene (c.77delG), which has previously been described by Schimmenti et al [26].

We designed a custom 385,000 probe SCE array with more than 100 candidate genes for eye malformations. The list of selected candidate genes is provided in Table 1. The genes were chosen based on reports of mutations identified in patients with coloboma, microphthalmia and anophthalmia [22,23,25]. Additionally, a comprehensive literature search was performed for published mutations associated with ocular phenotypes in animal models [27,28]. Some genes were included based on their role in signaling and developmental pathways which are known as important for eye formation and function [24].

Appropriate SCE probes for the regions of interest were chosen in collaboration with Roche-NimbleGen design team (Roche-NimbleGen, Madison, WI). Only protein coding regions (coding exons) of the 112 candidate genes were targeted on the array. We selected 385,000 long oligonucleotide probes (>60bp) to tile the exons of the genes of interest with a very high density. All the probes had the uniqueness score of one (defined as having no match in the genome other than itself longer than 38 bp, allowing up to 5 insertions/deletions/mismatches in that match), to exclude repetitive regions from probe selection and avoid capturing pseudogene sequences [29]. Upon completion of the design and manufacturing of the array, the SCE on samples of genomic DNA from two patients with known mutations in the *PAX2* gene was performed at the NimbleGen service laboratory (Roche-NimbleGen, Madison, WI), following previously described protocols [30–32]. Briefly, ~20 micrograms of each patient's genomic DNA were randomly fragmented by nebulization to an average size of 500 bp. Linkers were ligated to the DNA fragments to provide a priming site for post-enrichment amplification of the eluted fragment pool. The fragments were denatured and hybridized to the custom SCE array. After a 72-hour hybridization, unbound material was removed by stringent washing. The arrays were transferred to the NimbleGen elution system, and the enriched fragment pool was eluted and recovered from the array. The enriched fragments were amplified with 22mer linkers to generate enough DNA template for downstream applications. After amplification, the

amount of captured DNA was measured by spectrophotometry and the product was tested for enrichment level by quantitative PCR with four proprietary QC control loci. These QC loci are conserved in both human and mouse genomes and have been empirically determined to accurately predict enrichment with several different array designs.

Sequencing of the two SCE prepared samples was performed on the GS FLX instrument using Titanium chemistry, at the University of Iowa DNA Facility following standard protocols. Two samples, separated by gaskets, were sequenced independently on two regions of the picotiter plate. Briefly, amplified fragments from SCE were end-repaired and ligated to adapter oligonucleotides. The library was diluted based on the results of a previously performed titration, so that upon denaturation single DNA fragments hybridized to individual beads containing sequences complementary to adapter oligonucleotides. The beads were compartmentalized into water-in-oil microvesicles to allow clonal expansion of separate DNA molecules bound to the beads by emulsion PCR. After amplification, the emulsion was disrupted, and the beads containing clonally amplified template DNA were enriched. The beads were again separated by limiting dilution, deposited into individual picotiter-plate wells, and combined with sequencing enzymes. Iterative pyrosequencing was performed on the picotiter plate by successive flow addition of the 4 dNTPs. A nucleotide-incorporation event in a well containing clonally amplified template produced pyrophosphate release and picotiter-plate well-localized luminescence, which was transmitted through the fiber-optic plate and recorded on a charge-coupled device camera. With the flow of each dNTP reagent, wells were imaged, analyzed for their signal-to-noise ratio, filtered according to quality criteria, and subsequently algorithmically translated into a linear sequence output [33–35].

Data analysis was performed using the Roche proprietary software package for the GS FLX system. Image acquisition, image processing and signal processing were performed during the run. Post run analysis was conducted using the GS Reference Mapper. Sequence runs were mapped against the human reference genome (hg18), using the default software settings (minimum base overlap of 40bp and minimum overlap identity 90%). Sequence variations were detected automatically during mapping, and were annotated with known gene (refSeq genes from <http://genome.ucsc.edu/>) and SNP information (dbSNP129 from <http://genome.ucsc.edu/>). Variants were determined as high quality differences (HCDiffs) if the change was present in at least three non duplicate reads which included at least one read from each direction (forward and reverse). Additional analysis was performed with the CLC Genomic Workbench software (CLCbio, Aarhus, Denmark) and the NextGENe™ software (SoftGenetics, State College, PA).

RESULTS

We performed target enrichment by array based SCE and sequencing using GS FLX instrument on two DNA samples with known mutations in the *PAX2* gene. The sequencing was performed simultaneously for 112 candidate genes for ocular birth defects which were selected based on extensive literature search. A custom SCE array designed for target enrichment contained probes for all coding regions (total of 1017 exons) of the selected 112 genes. The size of the entire target region was 373,083 base pairs. Since 385,000 probe SCE arrays are able to capture up to 5MB of target sequence [32], our target region of ~0.37MB only used a portion of the capacity of the 385K array. This allowed the use of a large number of probes per each targeted region and helped in obtaining good enrichment. DNA yield after elution from the SCE arrays and PCR amplification was approximately 10 micrograms for each sample; control loci showed 721-fold enrichment for the sample with the missense change (Sample 1), and 697 fold enrichment for the sample with the frameshift mutation (Sample 2).

The run on the GS FLX instrument using Titanium chemistry yielded 214Mb of sequence for the first sample and 191Mb for the second sample. Both met the Roche standard of at least 150Mb per sample. 76.1Mb of sequence mapped to the targeted regions for Sample 1, and 89.6Mb for Sample 2 (Table 2); this proportion of bases on the target is comparable to what has previously been observed with array based SCE [12,15]. The average read lengths were 360bp for Sample 1 and 318bp for Sample 2. This is lower than the average read length for genomic samples but typical for samples prepared by SCE (The Roche standard is at least 300bp). Details about the sequencing coverage obtained in the target regions are shown in Table 2 and in Figures 1A and 1B.

For both samples, 94.7% of the captured regions showed $\geq 15\times$ coverage for all targeted bases (100%), with the average depth of coverage being ~ 206 for Sample 1 and ~ 175 for Sample 2. Only three regions showed complete lack of coverage. For 51 regions (out of a total of 1017) $15\times$ coverage was not obtained for all the bases within a region. Most regions with low coverage (44/51) were identical between the two samples. The regions which completely lacked coverage were in both cases the first coding exons of the genes *WNT4*, *WNT9A* and *LRP5*, which were noted to have a very high GC-content (83%, 81% and 83% respectively). Additionally, low complexity GC rich repetitive elements were present in the vicinity (100 bp upstream or downstream) of all the three exons without coverage. Unusual sequence characteristics with high GC content and a presence of repetitive elements are well known causes of poor enrichment by SCE and decreased sequencing efficiency [12].

Studies have shown that the coverage in the 10- to 15- fold range may be sufficient for resequencing applications, but higher coverage depths (50–60 fold) provide better alignment, assembly and accuracy [36]. Therefore, most of our target exons have higher coverage than needed for a robust re-sequencing assay, allowing to expand our target region by incorporating additional genes and to sequence our samples as a pool.

The missense change in Sample 1 (c.527G>C, S176T) was easily identified by NGS. A total of 151 reads covered the variant site, with 52% of the reads showing the wild type base and 48% showing the variant base, as would have been expected for a heterozygous allele. The c. 527G>C change identified by Sanger sequencing is shown in Figure 2A, while the same change detected by NGS is presented in Figure 2B.

A deletion of one base in a long polyG stretch was observed. A total of 76 reads covered the region, with $\sim 38\%$ showing the wild type allele (7 Gs), $\sim 54\%$ showing the mutation (6Gs) and the remaining $\sim 8\%$ showing different numbers of Gs which did not correspond to either allele present in the sample (Figure 3). Surprisingly, the c.77delG mutation was not registered by the GS Reference Mapper as a high quality difference. We therefore performed additional analysis with the CLC Genomic Workbench and the NextGENe™ analysis programs, focusing only on sequence changes in the coding region of the *PAX2* gene. The c.77delG was clearly identified both by the CLC Genomic Workbench and the NextGENe™ analysis. Reasons for the discrepancy between different analysis tools are unclear, however, it is well known that pyrosequencing based methods (like GS FLX) tend to erroneously interpret long stretches (>6) of the same nucleotide [33,37]. To compensate for the resulting “noise” at all homonucleotide sites the GS Reference Mapper may be less likely to register base deletions and insertions at such regions. We performed detailed analysis of sequence changes detected in targeted genes other than *PAX2*, to determine how well can SCE-NGS be applied to analysis of unknown samples from patients with eye anomalies. Since NGS applications usually identify a large amount of genetic variation, we wanted to determine if false positive findings (sequencing errors) and benign polymorphisms can readily be distinguished from possibly pathogenic changes which require further follow-up. The summary of this analysis is provided in Table 2. To consider clinically important variants, both in the SCE array design and in the analysis,

we focused only on coding sequences. Both homozygous and heterozygous changes were included, since both recessive and dominant models of inheritance are considered for the studied disorders. We filtered out all known nonpathogenic single-nucleotide polymorphisms reported in the dbSNP database (<http://www.ncbi.nlm.nih.gov/projects/SNP/>). This left the total of 26 changes in coding regions not corresponding to known SNPs in both Sample 1 and Sample 2 (Table 2). We then applied three additional previously described parameters to further reduce the number of candidate mutations in tested samples. We focused only on changes detected in the regions with sufficient coverage ($>30\times$) with the rationale that those are less likely to represent false positives [36]. We also eliminated synonymous base changes, which did not result in amino-acid changes. Knowing that the samples were sequenced individually and that mosaicism for point mutations is very rare in patients with genetic disorders, we also required that the percentage of reads showing the variant allele exceeds 30% [12]. Using these filters we reduced the number of detected coding alterations which were likely to represent real sequence changes (rather than sequencing errors) to five (two frameshift and three missense mutations) in Sample 1 and three (two frameshift and one missense mutation) in Sample 2 (listed in Table 3). Only these eight alterations were selected for confirmation by Sanger sequencing, although the presence of other real sequence changes (which did not meet our selection criteria) could not be excluded. The base substitutions (missense mutations) detected by NGS were also identified by Sanger sequencing, but the frameshift mutations were not confirmed (Table 3). This is not surprising however, since all the frameshift mutations detected by NGS (deletion “CGA” in the *FGFR2* gene, deletion “T” in the *CHD2* gene, deletion “A” in the *APC* gene and deletion “A” in the *LRP2* gene) mapped to long homonucleotide stretches, where sequencing errors are known to occur with GS FLX technology. Although the frameshift changes listed in Table 3 meet stringent selection criteria, they still represent false positives, showing the difficulty in accurate detection of frameshift mutations by pyrosequencing based NGS technologies.

PolyPhen (<http://genetics.bwh.harvard.edu/pph/>) and SIFT (<http://sift.jcvi.org/>) analyses were performed to examine the likelihood that the observed missense changes alter the functions of the encoded proteins. Both analysis tools predicted one missense mutation (R942C in the *RAB3GAP1* protein) to be probably damaging, and one (I744V in *ZEB2*) to be likely benign (Table 3). However, for the remaining two amino acid changes (I3389V in *LRP2* and L1129S in *APC*) PolyPhen and SIFT gave different predictions, thus showing that an accurate determination of functional significance of amino acid changes in proteins cannot be achieved based solely on computer analysis. In summary, although confirmation by Sanger sequencing and the use of bioinformatics tools may have decreased the number of sequence variants which would have needed further studies in unknown samples, our examples illustrate that appropriate evaluation of detected changes in patient specimens will often require parental testing and functional analysis. This requirement may represent a significant challenge in the clinical settings.

Multiple known single nucleotide polymorphisms were detected, providing reassurance that both SCE and GS FLX sequencing performed as expected. For example, Sample 2 showed presence of the previously described polymorphism in the exon 8 of the *PAX2* gene 1521A>C, P326P (rs1800898) [38].

DISCUSSION

SCE and GS FLX sequencing allowed us to simultaneously sequence 112 candidate genes for ocular birth defects in two independent DNA samples, and to detect two known mutations in the *PAX2* gene.

Although whole genome sequencing represents the most comprehensive approach to identifying disease causing mutations in patients with hereditary disorders, such sequencing and analysis is not yet possible [39]. We therefore focused our studies on genes which are known to be associated with ocular birth defects, as well as numerous candidate genes which play a role in regulating early eye development.

To overcome the limitations of target enrichment by PCR, we used high-density SCE microarrays for selecting and capturing relevant exons [29–32]. The GS FLX instrument with Titanium chemistry was chosen based on its ability to produce long reads which are advantageous for sequencing complex genomes. Additionally, protocols for target enrichment by array based SCE were originally developed and optimized for the GS FLX system. However, our methods can be easily adapted for use with other sequencing platforms, like the Genome Analyzer IIx from Illumina (Illumina, Inc., San Diego, CA) and the SOLiD System from Applied Biosystems (Life Technologies, Carlsbad, CA). In our pilot study “hands on” and instrument run times for sequencing 112 genes in two independent specimens were measured in days, while the cost was less than \$8000 per sample (\$3000 for SCE and ~\$5000 for sequencing). Performing enrichment by traditional PCR and Sanger sequencing for the same number of genes would have been a daunting task, even for facilities with high levels of automation and high throughput. Although our costs for performing SCE and NGS sequencing were high compared to savings that can potentially be achieved with these technologies, they were still modest since the price for diagnostic Sanger sequencing for only one gene can be thousands of dollars. Furthermore, there are multiple ways to further optimize our approach.

Significant increase in sequencing quality and throughput and decrease in cost can be achieved by: 1) running multiple samples in the same run either by separating them physically (with gaskets) or by adding short oligonucleotide adapters as “barcodes”, and running the samples as a pool [40,41], 2) optimizing the custom SCE array to achieve more uniform coverage, by adding additional probes for poorly performing regions, and 3) using improved SCE protocols which require smaller amounts of DNA sample and allow adding GS FLX sequencing linkers during the SCE stage [42].

Since the capacity of the 385,000 probe array is higher than our current targeted region, it will be feasible to expand our project by adding new genes as they are implicated in ocular birth defects. In addition, we may perform further studies to compare the microarray capture method with solution phase based enrichment methods, such as Agilent’s SureSelect.

GS FLX sequencing clearly identified the known missense mutation in our Sample 1, but was less successful in detecting the one base deletion in a long homonucleotide stretch, that had previously been identified by Sanger sequencing. Homonucleotide stretches are particularly problematic to accurately analyze using the pyrosequencing based GS FLX system, as demonstrated by identification of alterations in the numbers of homonucleotides that could not be confirmed by Sanger sequencing. Since homonucleotide repeats represent hotspots for disease causing mutations in humans, accurate detection of insertions and deletions in these regions is very important for the use of NGS in the clinical setting, and will require further optimization. For the GS FLX system, the problem may be solvable by better analysis algorithms or manual review as well as Sanger sequencing confirmation for regions containing >6 copies of the same base. Alternatively, users may consider SOLiD or Illumina instruments, which are based on other chemistries and have a different error profile.

Current concerns in application of SCE-NGS methods for mutation analysis in research and clinical settings include 1) lack of definitive parameters for distinguishing false positives (sequencing errors) from real sequence changes, 2) insufficient information about false negative rate, 3) lack of guidelines for dealing with the overwhelming number of detected

sequence variants, 4) systematic technical artifacts due to duplicated sequences in the genome (large gene families, pseudogenes etc.), 5) decreased ability to detect gene interruptions due to insertions of Alu repeats and other repetitive sequences [12], 6) storage and management of massive amounts of data and 6) reporting and ethical considerations.

False positive and false negative rates are vital for clinical applications of NGS, and are known to significantly depend on the level of sequencing coverage. Smith et al. [36] have shown that ~15-fold redundant coverage allows accurate mutational profiling in haploid organisms, while deeper coverage should be applied for diploid genomes (we used 30× coverage as a threshold in our study). However, there are currently no recommendations regarding the minimum percentage of reads which should show a variant allele in order to confidently establish a heterozygous-allele call (we applied an arbitrary cut-off of 30%). Approaches that are typically used in NGS to filter huge numbers of detected (real) variants and distinguish clinically relevant changes from benign polymorphisms include 1) comparing detected changes with public databases of known sequence variants (dbSNP) 2) considering only non-synonymous sequence changes and 3) analyzing by PolyPhen or SIFT whether a change is likely to be damaging for the function of the encoded protein. Comparison with public databases is likely to become even more helpful in the future, when the 1000 Genomes Project [43] generates a catalogue of common variation that is more complete and more evenly ascertained than dbSNP. Interference from homologous sequences (pseudogenes, members of the same gene family) may result in false positives due to mismapped reads from duplicated regions with related sequences. This problem is ameliorated by removal of reads with non-unique placements during read alignment. However, this decreases sensitivity for detecting mutations, making it difficult to reliably test genes which have paralogs with highly identical sequences.

Data storage and analysis are large challenges for individual laboratories considering use of NGS. The issue of data storage is further complicated by the consideration of how much of the dataset to keep (image data, raw data, or just fasta files). The bioinformatics challenges are alleviated by core facilities offering data retention on their servers, access to analysis software and consultation with bioinformatics experts. Additionally, it has been proposed that cloud-computing solutions which provide internet access to large clusters of computers may allow investigators from small laboratories to access data analysis tools and significant hardware power at a reasonable cost [44].

Large sample sets will need to be studied and specific, objective, evidence based guidelines and quality controls developed to address main technical issues and enable use of SCE and NGS technologies to reliably detect pathogenic mutations in research and diagnostics laboratories. Routine Sanger sequencing will likely be required for a long time to confirm data generated by NGS. Even when technical limitations get resolved, serious test reporting issues and ethical considerations are likely to remain. For example, each person is likely to carry multiple clinically relevant mutations which can be detected by genome-wide sequencing. It is unclear should participants in NGS based studies receive all their sequencing information (even when it is not possible to implement effective medical treatments for clinically relevant variants) and how should this information be delivered and interpreted to patients [45].

In conclusion, our study showed that even with existing limitations, individual researchers working with core sequencing facilities can use NGS as a research tool with confidence and ease. In the near future, thanks to continuing improvements in throughput, accuracy, cost and ease of data analysis, it will also become feasible to apply NGS in the clinical setting.

Acknowledgments

We gratefully acknowledge Einat Snir and Jennifer Bair from The University of Iowa DNA Facility for technical assistance with GS FLX sequencing and manuscript review, and Eric Cabot from the University of Wisconsin-Madison Biotechnology Center for help with reviewing the sequencing data.

Supported by the Translational Research Award from the University of Wisconsin Institute for Clinical and Translational Research-ICTR (P. I. Raca)

REFERENCES

1. ten Bosch JR, Grody WW. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *J Mol Diagn* 2008;10:484–492. [PubMed: 18832462]
2. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 2009;55:641–658. [PubMed: 19246620]
3. Tucker T, Marra M, Friedman JM. Massively parallel sequencing: the next big thing in genetic medicine. *Am J Hum Genet* 2009;85:142–154. [PubMed: 19679224]
4. Schuster SC. Next-generation sequencing transforms today's biology. *Nat Methods* 2008;5:16–18. [PubMed: 18165802]
5. Marguerat S, Wilhelm BT, Bahler J. Next-generation sequencing: applications beyond genomes. *Biochem Soc Trans* 2008;36:1091–1096. [PubMed: 18793195]
6. Ropers HH. Genetics of intellectual disability. *Curr Opin Genet Dev* 2008;18:241–250. [PubMed: 18694825]
7. Nance WE. The genetics of deafness. *Ment Retard Dev Disabil Res Rev* 2003;9:109–119. [PubMed: 12784229]
8. Marian AJ. Genetic determinants of cardiac hypertrophy. *Curr Opin Cardiol* 2008;23:199–205. [PubMed: 18382207]
9. Goodwin P. Hereditary retinal disease. *Curr Opin Ophthalmol* 2008;19:255–262. [PubMed: 18408503]
10. Gabriel C, Danzer M, Hackl C, Kopal G, Hufnagl P, Hofer K, Polin H, Stabenheiner S, Proll J. Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification. *Hum Immunol* 2009;70:960–964. [PubMed: 19706315]
11. Bentley G, Higurashi R, Hoglund B, Goodridge D, Sayer D, Trachtenberg EA, Erlich HA. High-resolution, high-throughput HLA genotyping by next-generation sequencing. 2009;74:393–403.
12. Chou LS, Liu CS, Boese B, Zhang X, Mao R. DNA Sequence Capture and Enrichment by Microarray Followed by Next-Generation Sequencing for Targeted Resequencing: Neurofibromatosis Type 1 Gene as a Model. *Clin Chem* 2010;56:1–11. [PubMed: 19892841]
13. Hoischen A, Gilissen C, Arts P, Wieskamp N, van der Vliet W, Vermeer S, Steehouwer M, de Vries P, Meijer R, Seiquer J, Knoers NVAM, Buckley MF, Scheffer H, Veltman JA. Massively parallel sequencing of ataxia genes after array-based enrichment. *Hum Mutat*. 2010 Feb 11; Published online DOI 10.1002/humu.21221.
14. Vasta V, Ng SB, Turner EH, Shendure J, Hahn SH. Next generation sequence analysis for mitochondrial disorders. *Genome Med* 2009;1:100. [PubMed: 19852779]
15. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272–276. [PubMed: 19684571]
16. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;42:30–35. [PubMed: 19915526]
17. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R,

- Walter MJ, Link DC, Graubert TA, DiPersio JF, Wilson RK. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008;456:66–72. [PubMed: 18987736]
18. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, Fulton LA, Locke DP, Magrini VJ, Abbott RM, Vickery TL, Reed JS, Robinson JS, Wylie T, Smith SM, Carmichael L, Eldred JM, Harris CC, Walker J, Peck JB, Du F, Dukes AF, Sanderson GE, Brummett AM, Clark E, McMichael JF, Meyer RJ, Schindler JK, Pohl CS, Wallis JW, Shi X, Lin L, Schmidt H, Tang Y, Haipek C, Wiechert ME, Ivy JV, Kalicki J, Elliott G, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson MA, Baty J, Heath S, Shannon WD, Nagarajan R, Link DC, Walter MJ, Graubert TA, DiPersio JF, Wilson RK, Ley TJ. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 2009;361:1058–1066. [PubMed: 19657110]
 19. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA, Hartigan J Jr, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008;321:1807–1812. [PubMed: 18772396]
 20. Mardis ER, Wilson RK. Cancer genome sequencing: a review. *Hum Mol Genet* 2009;18:R163–R168. [PubMed: 19808792]
 21. Stoll C, Alembik Y, Dott B, Roth MP. Congenital eye malformations in 212,479 consecutive births. *Ann Genet* 1997;40:122–128. [PubMed: 9259960]
 22. Guercio JR, Martyn LJ. Congenital malformations of the eye and orbit. *Otolaryngol Clin North Am* 2007;40:113–140. vii. [PubMed: 17346564]
 23. Gregory-Evans CY, Williams MJ, Halford S, Gregory-Evans K. Ocular coloboma: a reassessment in the age of molecular neuroscience. *J Med Genet* 2004;41:881–891. [PubMed: 15591273]
 24. Fitzpatrick DR, van Heyningen V. Developmental eye disorders. *Curr Opin Genet Dev* 2005;15:348–353. [PubMed: 15917212]
 25. Chang L, Blain D, Bertuzzi S, Brooks BP. Uveal coloboma: clinical and basic science update. *Curr Opin Ophthalmol* 2006;447–470. [PubMed: 16932062]
 26. Schimmenti LA, Shim HH, Wirtschafter JD, Panzarino VA, Kashtan CE, Kirkpatrick SJ, Wargowski DS, France TD, Michel E, Dobyns WB. Homonucleotide expansion and contraction mutations of PAX2 and inclusion of Chiari 1 malformation as part of renal-coloboma syndrome. *Hum Mutat* 1999;14:369–376. [PubMed: 10533062]
 27. Gross JM, Perkins BD. Zebrafish mutants as models for congenital ocular disorders in humans. *Mol Reprod Dev* 2008;75:547–555. [PubMed: 18058918]
 28. Alescaron C, Ernst RT. Anterior eye development and ocular mesenchyme: new insights from mouse models and human diseases. *Bioessays* 2004;26:374–386. [PubMed: 15057935]
 29. Roche-NimbleGen Inc, Roche Nimble-Gen probe design fundamentals. Application Notes. [Accessed December 2009]. http://www.nimblegen.com/products/lit/probe_design_2008_06_04.pdf
 30. Roche-NimbleGen Inc, NimbleGen services user's guides: sequence capture service. Application Notes. [Accessed December 2009]. http://www.nimblegen.com/products/lit/SeqCap_UsersGuide_Service_v3p0.pdf
 31. Olson M. Enrichment of super-sized resequencing targets from the human genome. *Nat Methods* 2007;4:891–892. [PubMed: 17971778]
 32. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA. Direct selection of human genomic loci by microarray hybridization. *Nat Methods* 2007;4:903–905. [PubMed: 17934467]
 33. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y,

- Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437:376–380. [PubMed: 16056220]
34. Rothberg JM, Leamon JH. The development and impact of 454 sequencing. *Nat Biotechnol* 2008;26:1117–1124. [PubMed: 18846085]
35. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 2008;9:387–402. [PubMed: 18576944]
36. Smith DR, Quinlan AR, Peckham HE, Makowsky K, Tao W, Woolf B, Shen L, Donahue WF, Tusneem N, Stromberg MP, Stewart DA, Zhang L, Ranade SS, Warner JB, Lee CC, Coleman BE, Zhang Z, McLaughlin SF, Malek JA, Sorenson JM, Blanchard AP, Chapman J, Hillman D, Chen F, Rokhsar DS, McKernan KJ, Jeffries TW, Marth GT, Richardson PM. Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res* 2008;18:1638–1642. [PubMed: 18775913]
37. Wicker T, Schlagenhauf E, Graner A, Close TJ, Keller B, Stein N. 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* 2006;7:275. [PubMed: 17067373]
38. Shim HH, Nakamura BN, Cantor RM, Schimmenti LA. Identification of two single nucleotide polymorphisms in exon 8 of PAX2. *Mol Genet Metab* 1999;68:507–510. [PubMed: 10607481]
39. Summerer D. Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics* 2009;94:363–368. [PubMed: 19720138]
40. Meyer M, Stenzel U, Myles S, Prufer K, Hofreiter M. Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res* 2007;35:e97. [PubMed: 17670798]
41. Meyer M, Stenzel U, Hofreiter M. Parallel tagged sequencing on the 454 platform. *Nat Protoc* 2008;3:267–278. [PubMed: 18274529]
42. Roche-NimbleGen Inc, NimbleGen Titanium Optimized Sequence Capture 385K Arrays. Application Notes. [Accessed December 2009].
http://www.454.com/downloads/products-solutions/SeqCapTitan385K_Final.pdf
43. Siva N. 1000 Genomes project. *Nat Biotechnol* 2008;26:256. [PubMed: 18327223]
44. McPherson JD. Next-generation gap. *Nat Methods* 2009;6:S2–S5. [PubMed: 19844227]
45. Biesecker LG. Exome sequencing makes medical genomics a reality. *Nat Genet* 2010;42:13–14. [PubMed: 20037612]

Figure 1. Figure 1A and 1B. Average numbers of reads per base (depth of coverage) for each targeted region in Sample 1 (Fig 1A) and Sample 2 (Fig 1B). Although significant variation is noticeable

between regions, sufficient coverage for reliable SNP detection was obtained for the majority of targeted exons.

Figure 2.

Figure 2A. Sanger sequencing chromatograms showing the c.527G>C, S176T mutation in two different DNA strands. The mutated base is marked by the red arrow.

Figure 2B. A screenshot from the CLCbio analysis software showing the misense mutation c. 527G>C, S176T. Reads from different directions are shown in different colors (red and blue). The mutated base is shown between the two vertical red lines.

NT_030059 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

Consensus - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02HGSYZ.455-473 - TTCCTC-TCA-T-CCA

F05MJXW02FNU95.88-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02G4G00.439-477 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG

F05MJXW02GICJS.32-1 - TTCCT

F05MJXW02HI2RU.51-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C

F05MJXW02GSAK5.183-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02J5MSR.404-456 - TTCCTC-TCA-T-CCA-GTT-CCA-G-C-A-CGG-CCT-CCCC-T-CC

F05MJXW02G7QRJ.56-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT

F05MJXW02FD22F.137-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02JFKFL.83-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G

F05MJXW02FWQA.138-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02IBUUM.383-438 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02H2JY9.381-419 - TTCCTC-TCA-T-CCA-GTT-CCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T

F05MJXW02JDUUO.379-390 - T

F05MJXW02JP73W.112-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02F9R7N.105-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02HOSAK.365-478 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02JJGKC.105-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02IK4YU.364-489 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02IFRPT.363-468 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02GT7JZ.361-424 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CC

F05MJXW02PF40M.361-479 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02H9DMT.359-483 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02HAJSH.169-1 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T-GGGA-

F05MJXW02HDVX7.356-397 - TTCCTC-TCA-T-CCA-GTT-CCCA-G-C-A-CGG-CCT-CCCC-T-CC-T--G-TTT--CCA-G-CG-CCTCCAA-T--G-A-CCC-A-G-T

Figure 3.

A screenshot from the CLCbio analysis software showing the frameshift mutation c.del77G. The mutated base is shown between the two vertical grey lines. The reads showing the mutant allele (6Gs) are grouped at the top, while the reads generated from the wild-type allele (7Gs) are shown at the bottom.

Table 1

Candidate genes for anophthalmia, microphthalmia and coloboma represented on the SCE array

Implicated in syndromic anophthalmia, microphthalmia and coloboma	Implicated in non-syndromic anophthalmia, microphthalmia and coloboma	Code for proteins on the SHH signaling pathway	Code for proteins on the WNT signaling pathway	Other
<i>DPYD, SIX3, RAB3GAP, ZFHX1B, ALG3, PITX2, SHH, PTCH, POMT1, KIAA1279, RBP4, PAX2, PAX6, PTPN11, CREBBP, SALL1 NDP, MKS1, SALL4, PQBP1, BCOR, PORCN, IGBP1, FLNA, CHD7, CC2D2A, HMX1, CLDN19, LRP2, TFAP2A, IGBP1, IKBKG, B3GALTL, GDF6, SOX2, OTX2, JAG1, BMP4, MITF, HESX1,</i>	<i>CHX10, MAF, SIX6, RAX</i>	<i>GLI2, GLI3, SMO, VAX1, VAX2, SUFU, DRM, CER1</i>	<i>DVL1, WNT1, WNT16, WNT5A, WNT5B, WNT7B, WNT7A, WNT2, WNT8A, WNT10A, WNT6, WNT8B, WNT3, WNT3A, WNT9A, WNT9B, WNT11, WNT4, WNT2B, LRP5, LRP6, FZD6, FZD9, FZD2, FZD1, FZD7, FZD5, FZD10, AXIN2, GSK3A, APC, CTNNB1</i>	<i>CRYAA, CRYAB, CRYBA1, CRYBA2, CRYBA4, CRYBB1, CRYBB2, CRYBB3, CRYGA, CRYGB, CRYGC, CRYGD, CRYGS, CRYZ, SOX10, ZNF703, ZNF503, CRX, FOXG1, BMP7, CHD2, DLX1, DLX2, TBX2, TBX5, FGF8, HES1, LHX1</i>
40 loci	4 loci	8 loci	32 loci	28 loci

Table 2

GS FLX Sequencing Summary

	Sample 1	Sample 2
Sequence yield	214Mb	191Mb
¹ Bases on target	76.1Mb	89.6Mb
² Average depth	206	175.17
³ Average coverage	99.24%	99.28%
%Covered $\geq 15\times$	97.35	97.35
# regions with $<100\%$ $15\times$ coverage	54	54
# regions without coverage	3	3
Total # of detected sequence changes	376	372
# changes in coding regions	110	114
# changes not corresponding to known SNPs	116	108
# changes in coding regions not corresponding to known SNPs	26	26
⁴ # changes likely to be real sequence alterations**	5	3

¹ Bases on target were determined based on rigorous criteria, so that bases that are a single nucleotide outside of the captured region are not included in the count, even if they map to regions flanking the targets

² Average depth is defined as average number of reads covering individual targeted nucleotides.

³ Average coverage is the average proportion of bases within targeted exons which were covered by reads.

⁴ These sequence variants met the following selection criteria: 1. they do not represent synonymous substitutions, 2. there is at least $30\times$ coverage at the variant site and 3. the percentage of reads showing the variant allele exceeds 30%.

Table 3

Non-synonymous coding sequence changes detected in Samples 1 and 2, which do not correspond to known SNPs.

Sample 1												
Chromosome	Reference Position	Gene Name	Coverage Depth	Reference Nucleotide	Variant Nucleotide	Variant Percentage	Reference AA	Variant AA	Homo-nucleotide	Sanger Sequencing Result	PolyPhen prediction	SIFT prediction
chr10	123,235,026	<i>FGFR2</i>	99	CGA	deletion	32%	FG	PX (frameshift)	YES (four Cs)	NOT confirmed		
chr2	169,746,208	<i>LRP2</i>	168	T	C	52%	I	V	-	Confirmed	possibly damaging	tolerated
chr2	144,872,994	<i>ZEB2</i>	150	T	C	46%	I	V	-	Confirmed	benign	tolerated
chr2	135,642,699	<i>RAB3GAP1</i>	158	C	T	49%	R	C	-	Confirmed	probably damaging	affects protein function
chr15	91,356,594	<i>CHD2</i>	146	T	deletion	37%	F	X (frameshift)	YES (five Ts)	NOT confirmed		

Sample 2												
Chromosome	Reference Position	Gene Name	Coverage Depth	Reference Nucleotide	Variant Nucleotide	Variant Percentage	Reference AA	Variant AA	Homo-polymer	NOT confirmed	PolyPhen prediction	SIFT prediction
									-	Confirmed	benign	* affects protein function (low confidence prediction)
chr5	112,202,576	<i>APC</i>	188	T	C	49%	L	S				
chr5	112,203,856	<i>APC</i>	148	A	deletion	37%	T	*X (frameshift)	YES (six As)	NOT confirmed		
chr2	169,750,453	<i>LRP2</i>	155	A	deletion	36%	Y	X (frameshift)	YES (four As)	NOT confirmed		

* this SIFT prediction was described as "low confidence". The amino acid change may have been predicted to affect the protein function just because other protein sequences used for comparison were not diverse enough.