

Applying Genomic Analysis to Newborn Screening

B.D. Solomon^a D.E. Pineda-Alvarez^a K.A. Bear^{a, b}
NISC Comparative Sequencing Program^c J.C. Mullikin^c J.P. Evans^d

^aMedical Genetics Branch, National Human Genome Research Institute, National Institutes of Health,

^bDepartment of Neonatology, Walter Reed National Military Medical Center-Bethesda, Bethesda, Md.,

^cNIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Rockville, Md., and ^dDepartment of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, N.C., USA

Key Words

Exome sequencing · Genomic sequencing · Newborn screening · Whole-exome sequencing

Abstract

Large-scale genomic analysis such as whole-exome and whole-genome sequencing is becoming increasingly prevalent in the research arena. Clinically, many potential uses of this technology have been proposed. One such application is the extension or augmentation of newborn screening. In order to explore this application, we examined data from 3 children with normal newborn screens who underwent whole-exome sequencing as part of research participation. We analyzed sequence information for 151 selected genes associated with conditions ascertained by newborn screening. We compared findings with publicly available databases and results from over 500 individuals who underwent whole-exome sequencing at the same facility. Novel variants were confirmed through bidirectional dideoxynucleotide sequencing. High-density microarrays (Illumina Omni1-Quad) were also performed to detect potential copy number variations affecting these genes. We detected an average of 87 genetic variants per individual. After excluding artifacts,

96% of the variants were found to be reported in public databases and have no evidence of pathogenicity. No variants were identified that would predict disease in the tested individuals, which is in accordance with their normal newborn screens. However, we identified 6 previously reported variants and 2 novel variants that, according to published literature, could result in affected offspring if the reproductive partner were also a mutation carrier; other specific molecular findings highlight additional means by which genomic testing could augment newborn screening.

Copyright © 2012 S. Karger AG, Basel

Due to the availability of new high-throughput sequencing techniques, large-scale genomic analysis is becoming increasingly prevalent, and many potential clinical uses have been proposed. One such application is the extension or augmentation of newborn screening [Alexander and van Dyck, 2006]. The goal of newborn screening is primarily to identify, in an efficient and cost-effective manner, diseases in which early treatment is necessary to improve outcome. Relying on the American College of Medical Genetics recommendations, most United States newborn screening programs perform as-

says for 29 core conditions as well as 25 secondary targets that are part of the differential diagnosis for these core conditions [American College of Medical Genetics' Newborn Screening Expert Group, 2006; Burke et al., 2011].

In theory, gene-based screening has several advantages, such as the ability to bypass the need for substrate accumulation in affected patients and the potential to capture affected individuals missed by current newborn screening techniques [Schimmenti et al., 2011]. Additionally, genetic information could be used to complement the interpretation of currently available newborn screening results, potentially reducing the number of false-positive and non-clinically significant results generated. For example, certain genetic variants can lead to enzymatic differences that are ascertained by conventional newborn screening, falsely suggesting the presence of a disorder but not actually causing disease. Sequence-based information could avoid misidentifying these individuals as having positive newborn screens, thus avoiding attendant psychological stress, additional costs, and increased workload of those involved in newborn screening. Further, sequence-based data could enable rapid movement through the current algorithms for follow-up of abnormal results as many of these algorithms involve DNA-based testing (see the American College of Medical Genetics website for specific algorithms; www.acmg.net) [Tarini and Goldenberg, 2012].

There are, however, numerous challenges to the use of genomic sequencing to augment newborn screening. Major issues revolve around the difficulties inherent in interpretation of variants, the need to perform testing efficiently, achieving acceptable sensitivity and specificity, and the reality that only small amounts of DNA are typically available [Tarini and Goldenberg, 2012].

In order to begin to address such questions objectively, we analyzed data from 3 children with normal newborn screens who participated in a National Institutes of Health (NIH)/National Human Genome Research Institute (NHGRI) protocol on VACTERL association and who underwent whole-exome sequencing. Studies into the genetic causes of VACTERL association in these individuals are in progress; here, we analyze sequence data from genes known to be associated with conditions routinely ascertained by newborn screening in order to describe the types of findings that may arise when using high-throughput sequencing in conjunction with newborn screening.

Patients and Methods

We performed high-density microarrays (Illumina Omni1-Quad) and whole-exome sequencing for 3 children who participated in an established IRB-approved protocol on VACTERL association, a rare congenital disorder involving a combination of congenital anomalies. VACTERL association is not thought to have a classic biochemical basis such as would be ascertained by newborn screening. Full consent was obtained for all participants, and all participants and their families were seen in person at the NIH Clinical Center.

DNA Extraction

Blood was obtained via a peripheral venous sample, and DNA was initially extracted using a QIAamp DNA Blood Maxi Kit (Qiagen, Germantown, Md., USA). Phenol:chloroform purification was performed prior to whole-exome sequencing.

Microarray Performance and Analysis

Microarray analysis was performed using the Illumina Omni1-Quad SNP array per the Illumina 'infinium assay' protocol (Illumina Inc., San Diego, Calif., USA) [Gunderson et al., 2005]. In brief, extracted DNA was whole-genome amplified, fragmented, hybridized, fluorescently tagged, and scanned. The DNA samples were hybridized to the Illumina HumanOmni1-Quad Bead-Chips which contain >1 million SNP loci. We collected data using a BeadArray scanner and visualized data with the GenomeStudio (v2009.2, www.Illumina.com) genotyping module. The call rates for all the DNA samples were >99%. We used human genome build 36.1 (NCBI36/hg18) for analysis. Copy number variations (CNVs) were detected using PennCNV software filtered to annotate regions with at least 3 contiguous SNPs with the same imbalance [Wang et al., 2007]. Genomic imbalances were compared with known CNVs through the Database of Genomic Variants [Zhang et al., 2006].

Whole-Exome Sequencing

We performed solution hybridization exome capture with the SureSelect Human All Exon 38Mb and 50Mb Systems (Agilent Technologies, Santa Clara, Calif., USA) using biotinylated RNA baits to hybridize to sequences that correspond to exons [Gnirke et al., 2009]. We used the manufacturer's protocol version 1.0 compatible with Illumina paired-end sequencing except that the DNA fragment size and quality was measured using a 2% agarose gel stained with Sybr Gold rather than an Agilent Bioanalyzer. Manufacturer's specifications for the 38Mb kit state that the capture regions total approximately 38 Mb which accounts for 1.22% of the human genome, corresponding to the Consensus Conserved Domain Sequences database (CCDS) and >1,000 non-coding RNAs. The 50Mb kit also includes exons defined by the Gencode Project (<http://www.sanger.ac.uk/resources/databases/encode/>). Targeted regions included the exons of 18,113 CCDS genes, with a total of 37,640,396 bases in the human genome (All Exon 38Mb). The All Exon 50Mb kit includes all the regions in the All Exon 38Mb kit and adds exons of additional genes, miRNAs, and non-coding RNA genes, totaling 30,241 genomic features within a total of 51,646,629 targeted bases. Flowcell preparation and sequencing were carried out according to the protocol for the GAIIX sequencer (Illumina Inc.) [Bentley et al., 2008]. We used 76- or 101-bp paired-end lanes on a GAIIX flowcell in order

to generate sufficient reads to generate the aligned sequence. We performed image analysis and base calling on all data lanes using Illumina Genome Analyzer Pipeline software (GAPipeline versions 1.4.0 or greater) with default parameters.

Variant Analysis

Variants were analyzed using VarSifter software (<http://research.nhgri.nih.gov/software/VarSifter/>) [Teer et al., 2012]. In summary, we aligned reads to human genome build 36.1 (NCBI36/hg18) for analysis using 'efficient large-scale alignment of nucleotide databases' (ELAND, Illumina). For variants described here, although initial annotation was performed using NCBI36/hg18, variants are given here using NCBI37/hg19 coordinates. We grouped reads that aligned uniquely into genomic sequence intervals of approximately 100 kb; non-aligning reads were binned with their paired-end mates. Reads in each bin were subjected to a Smith-Waterman-based local alignment algorithm, *cross_match*, using the parameters *-minscore 21* and *-masklevel 0* to their respective 100-kb genomic sequence (<http://www.phrap.org>) [Smith and Waterman, 1981; Teer et al., 2012]. A total of 6 Gb of high-confidence mappable sequence data were generated in autosomal targeted regions per individual. Genotypes were called at all positions with high-quality sequence bases (Phred-like Q20 or greater) using a Bayesian algorithm (most probable genotype, MPG) [Teer et al., 2010, 2012]; goal read-depth is an average of at least 85% in targeted regions. Genotypes with an MPG score ≥ 10 (score/coverage ratio ≥ 0.5 , with a minimum of 10 reads) demonstrate $>99.89\%$ concordance with SNP Chip data. Targeted regions included the exons of 17,134 genes, with a total of 37,640,396 bases in the human genome (All Exon 38Mb: individual 3) or the exons of 30,241 genes and total 51,646,629 bases (All Exon 50Mb: individuals 1 and 2). The annotation of cSNVs (coding single nucleotide variants) was based on UCSC's 'known genes' dataset. We classified SNVs and short deletion-insertion variants with a custom suite of annotation scripts (PIANNO) as those in intronic, UTR, or within coding regions. The software categorized variants as belonging to one of the following subsets: 3'-UTR, 5'-UTR, downstream variants, frameshift (deletion, insertion, or substitution), intergenic, intronic, ncRNA (3'-UTR, 5'-UTR, exonic, intronic, or splicing), non-frameshift (deletion, insertion, or substitution), non-synonymous SNV, splicing, stop-gain SNV, stop-loss SNV, synonymous SNV, or upstream.

Analysis Related to Newborn Screening

From the exome and array-based data, 151 genes were selected and analyzed. Mutations in these genes (though often only in the homozygous/compound heterozygous state) would be predicted to result in disease that would be ascertainable by newborn screening (tables 1 and 2). For certain disorders, such as congenital deafness, every genetic disorder that could relate to detectable phenotypes would not be covered [Smith et al., 2012]. For our selected variant triage procedure, variants were first analyzed in multiple categories (see above) based solely on variant type. Second, for specific analysis related to newborn screening-associated genes, we focused on variants with the highest likelihood for a priori (i.e. not requiring in-depth functional analysis) pathogenicity: variants located in coding regions (e.g. excluding variants in the 3'- or 5'-UTR or captured intronic regions) and which were either in-frame or frameshift insertion-deletions, non-synony-

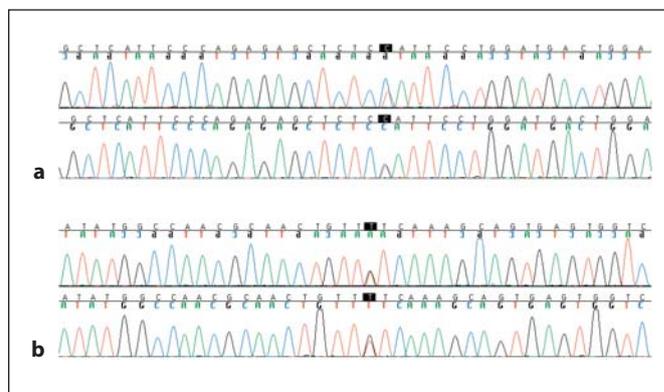


Fig. 1. a Novel heterozygous variant in *OTOA* (MIM 607039): c.674T>C, p.Tyr232His (individual 2). **b** Novel heterozygous variant in *SLC26A5* (MIM 613865): c.1777G>T, p.Val561Phe (individual 1).

mous, canonical splice-site, or other truncating variants (as annotated in tables 1 and 2). Third, variants found in public databases were included, and inclusion in these databases was not considered to be evidence of lack of pathogenicity, especially in recessive conditions; each such known variant was individually interrogated for possible reported health-related issues (accessed databases: dbSNP, build 131, Human Gene Mutation Database, last access December 2011) [Cooper et al., 1998; Smigielski et al., 2000]. Variants with only weak association with disease, such as those found via genome wide association studies, were not considered. Fourth, variants meeting the above criteria and thus still considered to be potentially deleterious (all of which were missense variants) were analyzed according to possible pathogenicity based upon predicted protein changes, including residue conservation, amino acid change type, and motif location [Teer et al., 2012].

Fifth, in order to detect likely artifacts, we performed further comparison of variants of interest versus results of whole-exome sequencing of 572 individuals (sequenced at the same facility as our patients) from the ClinSeq™ cohort which ascertains patients with a phenotypic continuum from unaffected to those who have had myocardial infarctions [Biesecker et al., 2009]. Annotated variants were considered to be highly likely to be artifacts when they were not noted to be previously known polymorphisms (appeared to be novel) and yet were seen in multiple comparison samples; as all variants thus determined to be artifacts involved repeat regions, this lent credence to the artifact assignment. Finally, other (non-artifact) novel variants were confirmed via bidirectional dideoxynucleotide sequencing (fig. 1).

Of note, it is clear that there are different potential approaches to the management of the specific findings generated through this study. For the IRB-approved algorithm under which this study was conducted, as the identified variants were found in the heterozygous ('carrier') state in rare recessive disease-associated genes in the studied individuals, they would not meet criteria for return of information [Solomon et al., 2012].

Table 1. Summary of findings from analysis of 151 genes in which mutations (often in the homozygous or compound heterozygous state) could result in disorders potentially capturable by extended newborn screening

Individual	1	2	3
Overall variant summary			
total variants	108,003	102,166	63,298
non-synonymous SNV	10,693	10,533	8,473
INDEL (frameshift)	216	186	91
INDEL (in-frame)	132	139	91
stop-gain SNV	110	106	63
splicing	134	156	59
variants not in dbSNP (build 131)	40,226	36,675	20,147
Variants related to newborn screening-associated conditions			
total number of variants	93	94	74
number of known variants identified			
without reported evidence of pathogenicity	85	87	68
number of variants shown to be artifacts	4	4	4
variants (previously reported) with potential clinical relevance ^a	<p><i>ACADS</i> (MIM 201470): rs1799958: c.625G>A, p.Gly209Ser (possible association with SCAD)</p> <p><i>CBS</i> (MIM 236200): rs5742905: c.833A>G, p.Ile278Thr (homocystinuria)</p> <p><i>GALT</i> (MIM 230400): rs2070074: c.940A>G, p.Asn314Asp (Los Angeles/ D₁ allele, as no promoter deletion present)</p>	<p><i>ACADS</i> (MIM 201470): rs1799958: c.625G>A, p.Gly209Ser (possible association with SCAD)</p> <p><i>HPD</i> (MIM 140350; 276710): rs1154510: c.97A>G, p.Thr33Ala (hawkinsinuria, allelic with tyrosinemia type III)</p>	<p><i>DBT</i> (MIM 248600): rs12021720: c.1150A>G, p.Ser384Gly (MSUD)</p> <p><i>HPD</i> (MIM 140350; 276710): rs1154510: c.97A>G, p.Thr33Ala (hawkinsinuria, allelic with tyrosinemia type III)</p>
novel variants with potential clinical relevance ^b	<p><i>SLC26A5</i> (MIM 613865): c.1777G>T, p.Val561Phe (autosomal recessive deafness) (fig. 1b)</p>	<p><i>OTOA</i> (MIM 607039): c.674T>C, p.Tyr232His (autosomal recessive deafness) (fig. 1a)</p>	

In the upper part of the table, an overall summary of variants is presented; the lower portion of the table focuses on variants in genes associated with newborn screening-ascertained conditions. Some differences in variant results can be attributed to different methods related to exome sequencing.

dbSNP = Database of single nucleotide polymorphisms; INDEL = insertion/deletion variant; MSUD = maple syrup urine disease; SCAD = short chain acyl-CoA dehydrogenase deficiency; SNV = single nucleotide variant.

RefSeq identifiers: *ACADS*: RefSeq NM_000017; *CBS*: RefSeq NM_000071; *DBT*: RefSeq NM_001918 *GALT*: RefSeq NM_

000155; *HPD*: RefSeq NM_002150; *OTOA*: RefSeq NM_144672; *SLC26A5*: RefSeq NM_198999.

^a The presence of these variants would not be associated with likely clinical relevance in these individuals, but offspring could be affected with disease if a reproductive partner were also a heterozygous mutation carrier.

^b These variants are not included in publicly available databases and were not found in the 572 comparison samples. In each case, the presence of these variants was supported by their identification in other family members and was confirmed via bidirectional dideoxynucleotide sequencing.

Results

All 3 children had normal newborn screening results and exhibited no evidence later for any disorders that are ascertained by current newborn screening. By microarray analysis (Illumina Omni1-Quad), no patient had any CNVs affecting genes associated with disorders typically queried by current newborn screening (we did not use exome analysis to detect CNVs) [Sathirapongsasuti et al., 2011].

In summary, as presented in tables 1 and 2, we detected a total of 261 variants related to newborn screening-associated genes for the 3 individuals, with an average of 87 variants per individual. Consistent with the normal results from standard newborn screening, no variants were identified that, in their specific allelic state, would predict disease in the tested individuals. However, each individual had multiple variants that could, according to published literature and publicly available mutation databases (dbSNP, HGMD), result in affected offspring if their reproductive partner were also a heterozygous mutation carrier. All such variants were missense substitutions. Individual 1 had 3 such variants; 2 have been previously reported (in *ACADS*, associated with short chain acyl-CoA dehydrogenase deficiency, MIM 201470; and *CBS*, associated with homocystinuria, MIM 236200), while 1 (in *SLC26A5*, associated with autosomal recessive deafness, MIM 613865) was novel [Hu et al., 1993; Corydon et al., 2001; Liu et al., 2003]. Additionally, individual 1 was found to have the Los Angeles/D₁ allele in *GALT*. Homozygous/compound heterozygous mutations in *GALT* are associated with galactosemia (MIM 230400), but the identified *GALT* allele is not known to cause pathogenicity. This is a clinically important distinction, as the finding of this allele (as opposed to a more deleterious allele) by molecular testing directs clinical decision-making in an infant with an abnormal conventional newborn screen for galactosemia [Tedesco, 1972; Langley et al., 1997; Elsas et al., 2001]. Individual 2 had 3 potentially relevant variants, including 2 that have been previously reported (in *ACADS*, associated with short chain acyl-CoA dehydrogenase deficiency, MIM 201470; and *HPD*, associated with hawkinsinuria, allelic with tyrosinemia type III, MIM 140350), and 1 novel variant (in *OTOA*, associated with autosomal recessive deafness, MIM 607039) [Tomoeda et al., 2000; Corydon et al., 2001; Zwaenepoel et al., 2002]. Individual 3 had two such variants, neither of which were novel (in *DBT*, associated with maple syrup urine disease, MIM 248600; and *HPD*, associated with hawkinsinuria, allelic with ty-

rosinemia type III, MIM 140350) [Tsuruta et al., 1998; Tomoeda et al., 2000].

On initial analysis, variants in 7 genes (*CPT1A* (MIM 255120), *CYP21A2* (MIM 201910), *DUOX2* (MIM 607200), *ETFB* (MIM 231680), *OTOA* (MIM 607039), *TAT* (MIM 276600), and *TRIOBP* (MIM 609823)) appeared to be novel according to publicly available databases, and these variants were not found in the 572 comparison individuals sequenced at the same facility. On reexamination of the same databases a short time later, these were found to be newly included in updated versions of the databases; none were reported as pathogenic or disease-associated.

Using a database of 572 individuals sequenced at the same facility, we were able to detect that variants affecting 5 genes (*CYP21A2* (MIM 201910), *GPSM2* (MIM 613557), *HADHB* (MIM 609015), *TMIE* (MIM 600971), and *TRIOBP* (MIM 609823)) were sequencing artifacts (table 2). All of these variants involve repeat regions.

Discussion

Despite the small sample size, our findings highlight several important elements that should serve to inform and inspire further study. First, this analysis demonstrates challenges related to the interpretation of variants of unknown significance. Some variants would clearly be deleterious in the homozygous/compound heterozygous state. Others are common variants with no evidence for direct involvement in Mendelian disorders such as many included in newborn screening. However, many variants fall into a 'gray zone', especially in polymorphic genes. Using publicly available (as well as private) databases can be helpful in terms of determining whether these variants have been identified previously, but the critical step in terms of determining pathogenicity remains daunting and fraught with potential error. This is especially true for certain variant types such as single amino acid substitutions [Berg et al., 2011]. In fact, the frequency in our small cohort of some of these purportedly clinically-relevant variants, in contrast to the overall prevalences of the associated recessive diseases, argues against their pathogenicity and points to the need for care when interpreting public databases as well as reported findings.

In order to avoid problems regarding variants of unknown significance, one possibility would be to use a custom-designed assay to test only for known deleterious mutations. This would also help address the problem associated with DNA quantity requirements (though it must be stated that technological improvements will like-

ly help with the DNA quantity issue in the near future). However, this approach would also preclude the identification of many mutations in genes in which there are a high proportion of family-specific novel variants. It will probably be far more expedient to 'sort' all variants informatically and simply ignore (for now) those which cannot be clearly defined as deleterious. Such an informatics approach has the added important advantage of allowing reanalysis of genomic data as more variants (and relevant genes) are identified and will also allow research analysis of novel variants in an ongoing manner [Berg et al., 2011]. Moreover, accumulation of rich genomic data in those undergoing concurrent traditional newborn screening with subsequent informatics-based analysis of the results will allow accrual of critical data ultimately necessary for accurately interpreting novel variants. For example, in a disease-free individual, when novel variants appear *in trans* with variants previously documented as disease-causing, the novel variant can be typically be assigned to non-pathogenic status.

Second, our findings demonstrate the strengths and weaknesses of relying heavily upon publicly available databases. For example, variants in 7 genes appeared to be novel on first analysis, but on re-examination a few months later, these variants were found to be newly included in the updated databases, highlighting the need for both timely curation of databases and iterative analysis of patient data. Conversely, using such databases to assign pathogenicity can be equally problematic, especially in the case of recessive or low-penetrance mutations and because such databases have frequently included seemingly pathogenic mutations that are in fact benign. In other words, it must be abundantly clear that the inclusion of variants in these databases is not a sign of clinical irrelevance.

Third, this study emphasizes pitfalls in high-throughput sequencing, both in terms of incomplete coverage of all relevant regions as well as the inevitable presence of artifacts. Using a database of 572 individuals sequenced at the same facility, 'variants' affecting 5 genes were found to be sequencing artifacts. Unsurprisingly, all of these variants involved repeat regions. Such concerns raise questions about both false-positive and false-negative data and the need to confirm clinically actionable findings before reporting them, especially until next-generation sequencing platforms achieve better accuracy.

Though this study highlights numerous impediments to the use of genomic data to augment newborn screening, it also illustrates several potential benefits. First, as described in the results section, we identified a variant in

GALT (p.Asn314Asp) that can be associated with reduced enzyme activity when linked with certain variants *in cis* [Langley et al., 1997; Elsas et al., 2001]. The lack of these linked variants (and the presence of variants linked to the allele conferring normal enzymatic activity) confirms that this is not a clinically concerning finding. Having information like this immediately available could be an effective way to help correlate results from current newborn screening techniques. Second, we identified a heterozygous, established disease-associated missense variant in *PHYH* (MIM 266500): rs28938169: c.85C>T, p.Pro29Ser; mutations in this gene are associated with Refsum disease [Jansen et al., 2000]. One of the manifestations of Refsum disease is deafness, but the onset is typically slightly older than would be ascertained by newborn screening. This illustrates how genomic screening could complement conventional screening by ascertaining clinically actionable disorders that would not be ascertainable by current newborn screening methods or disorders whose rarity precludes inclusion in conventional newborn screening panels.

Acknowledgements

This research was supported by the Division of Intramural Research, National Human Genome Research Institute (NHGRI), National Institutes of Health and Human Services, United States of America. The authors are extremely grateful to Dr. Leslie G. Biesecker (Chief and Senior Investigator, Genetic Disease Research Branch, NHGRI) for access to large-scale sequencing data for the use as comparison samples and to Dr. Max Muenke (Chief and Senior Investigator, Medical Genetics Branch, NHGRI) for his support and mentorship. Pertaining to Dr. Bear, the views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Army nor the US Government.

References

- Alexander D, van Dyck PC: A vision of the future of newborn screening. *Pediatrics* 117:S350–S354 (2006).
- American College of Medical Genetics' Newborn Screening Expert Group: Newborn screening: toward a uniform screening panel and system. *Genet Med* 8:1S-252S (2006).
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al: Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53–59 (2008).
- Berg JS, Khoury MJ, Evans JP: Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. *Genet Med* 13:499–504 (2011).

- Biesecker LG, Mullikin JC, Facio FM, Turner C, Cherukuri PF, et al: The ClinSeq Project: piloting large-scale genome sequencing for research in genomic medicine. *Genome Res* 19:1665–1674 (2009).
- Burke W, Tarini B, Press NA, Evans JP: Genetic screening. *Epidemiol Rev* 33:148–164 (2011).
- Cooper DN, Ball EV, Krawczak M: The human gene mutation database. *Nucleic Acids Res* 26:285–287 (1998).
- Corydon MJ, Vockley J, Rinaldo P, Rhead WJ, Kjeldsen M, et al: Role of common gene variations in the molecular pathogenesis of short-chain acyl CoA dehydrogenase deficiency. *Pediatr Res* 49:18–23 (2001).
- Elsas LJ, Lai K, Saunders CJ, Langley SD: Functional analysis of the human galactose-1-phosphate uridylyltransferase promoter in Duarte and LA variant galactosemia. *Mol Genet Metab* 72:297–305 (2001).
- Gnirke A, Melnikov A, Maguire J, Rogov P, Leproust EM, et al: Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27:182–189 (2009).
- Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS: A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* 37:549–554 (2005).
- Hu FL, Gu Z, Kozich V, Kraus JP, Ramesh V, Shih VE: Molecular basis of cystathionine beta-synthase deficiency in pyridoxine responsive and nonresponsive homocystinuria. *Hum Mol Genet* 2:1857–1860 (1993).
- Jansen GA, Hogenhout EM, Ferdinandusse S, Waterham HR, Ofman R, et al: Human phytanoyl-CoA hydroxylase: resolution of the gene structure and the molecular basis of Refsum's disease. *Hum Mol Genet* 9:1195–1200 (2000).
- Langley SD, Lai K, Dembure PP, Hjelm LN, Elsas LJ: Molecular basis for Duarte and Los Angeles variant galactosemia. *Am J Hum Genet* 60:366–372 (1997).
- Liu XZ, Ouyang XM, Xia XJ, Zheng J, Pandya A, et al: Prestin, a cochlear motor protein, is defective in non-syndromic hearing loss. *Hum Mol Genet* 12:1155–1162 (2003).
- Sathirapongsasuti JF, Lee H, Horst BA, Brunner G, Cochran AJ, et al: Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics* 27:2648–2654 (2011).
- Schimmenti LA, Warman B, Schleiss MR, Daly KA, Ross JA, et al: Evaluation of newborn screening bloodspot-based genetic testing as second tier screen for bedside newborn hearing screening. *Genet Med* 13:1006–1010 (2011).
- Smigielski EM, Sirotkin K, Ward M, Sherry ST: dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res* 28:352–355 (2000).
- Smith RJH, Hildebrand MS, Van Camp G: Deafness and hereditary hearing loss overview; in Pagon RA, Bird TD, Dolan CR, Stephens K (eds): *GeneReviews* [Internet] (University of Washington, Seattle 2012) available at <http://www.ncbi.nlm.nih.gov/books/NBK1434/>.
- Smith TF, Waterman MS: Identification of common molecular subsequences. *J Mol Biol* 147:195–197 (1981).
- Solomon BD, Hadley DW, Pineda-Alvarez DE, NISC Comparative Sequencing Program, Kamat A, et al: Incidental medical information in whole-exome sequencing. *Pediatrics* 129:e1605–1611 (2012).
- Tarini BA, Goldenberg AJ: Ethical issues with newborn screening in the genomics era. *Annu Rev Genomics Hum Genet* [Epub ahead of print] (2012).
- Tedesco TA: Human galactose 1-phosphate uridylyltransferase. Purification, antibody production, and comparison of the wild type, Duarte variant, and galactosemic gene products. *J Biol Chem* 247:6631–6636 (1972).
- Teer JK, Bonnycastle LL, Chines PS, Hansen NF, Aoyama N, et al: Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. *Genome Res* 20:1420–1431 (2010).
- Teer JK, Green ED, Mullikin JC, Biesecker LG: VarSifter: visualizing and analyzing exome-scale sequence variation data on a desktop computer. *Bioinformatics* 28:599–600 (2012).
- Tomoeda K, Awata H, Matsuura T, Matsuda I, Ploechl E, et al: Mutations in the 4-hydroxyphenylpyruvic acid dioxygenase gene are responsible for tyrosinemia type III and hawkinsinuria. *Mol Genet Metab* 71:506–510 (2000).
- Tsuruta M, Mitsubuchi H, Mardy S, Miura Y, Hayashida Y, et al: Molecular basis of intermittent maple syrup urine disease: novel mutations in the *E2* gene of the branched-chain alpha-keto acid dehydrogenase complex. *J Hum Genet* 43:91–100 (1998).
- Wang K, Li M, Hadley D, Liu R, Glessner J, et al: PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 17:1665–1674 (2007).
- Zhang J, Feuk L, Duggan GE, Khaja R, Scherer SW: Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. *Cytogenet Genome Res* 115:205–214 (2006).
- Zwaenepoel I, Mustapha M, Leibovici M, Verpy E, Goodyear R, et al: Otoancorin, an inner ear protein restricted to the interface between the apical surface of sensory epithelia and their overlying acellular gels, is defective in autosomal recessive deafness DFNB22. *Proc Natl Acad Sci USA* 99:6240–6245 (2002).