Published in final edited form as:

Expert Rev Mol Diagn. 2011 November ; 11(8): 855–868. doi:10.1586/erm.11.70.

Adopting orphans: comprehensive genetic testing of Mendelian diseases of childhood by next-generation sequencing

Stephen F Kingsmore^{†,1}, Darrell L Dinwiddie¹, Neil A Miller¹, Sarah E Soden¹, Carol J Saunders¹, and The Children's Mercy Genomic Medicine Team^{*}

¹Children's Mercy Hospital & Clinics, 2401 Gillham Road, Kansas City, MO 64108, USA

Abstract

Orphan diseases are individually uncommon but collectively contribute significantly to pediatric morbidity, mortality and healthcare costs. Current molecular testing for rare genetic disorders is often a lengthy and costly endeavor, and in many cases a molecular diagnosis is never achieved despite extensive testing. Diseases with locus heterogeneity or overlapping signs and symptoms are especially challenging owing to the number of potential targets. Consequently, there is immense need for scalable, economical, rapid, multiplexed diagnostic testing for rare Mendelian diseases. Recent advances in next-generation sequencing and bioinformatic technologies have the potential to change the standard of care for the diagnosis of rare genetic disorders. These advances will be reviewed in the setting of a recently developed test for 592 autosomal recessive and X-linked diseases.

Keywords

bioinformatics; carrier screening; Mendelian diseases; molecular diagnostics; next-generation sequencing; orphan diseases

The burden of Mendelian disease

Of 6699 known disorders with suspected Mendelian inheritance, 1154 are recessive or X-linked and have an established molecular basis [101]. The number with a known molecular basis is rapidly increasing through next-generation sequencing (NGS) of whole genomes and exomes of affected individuals [1]. Recent reports indicate that Mendelian disorders account for approximately 17% of pediatric hospitalizations and an even greater proportion of healthcare costs [2–5]. Furthermore, it is reported that Mendelian diseases collectively affect approximately 25 million people in the USA [6]. By 2015, it is likely that over 7000 Mendelian disorders will have an established molecular cause.

© 2011 Expert Reviews Ltd

[†]Author for correspondence: Tel.: +1 816 234 3038 sfkingsmore@cmh.edu.

^{*}Members of the Children's Mercy Genomic Medicine Team are listed in the Acknowledgements section on page 866.

Financial & competing interests disclosure The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Conventional univariate genetic testing

There are hundreds of childhood recessive and X-linked illnesses for which molecular diagnosis is technically feasible but not currently available. Many are too rare to attract commercial interest for clinical testing. Such conditions are trapped in a state of rudimentary knowledge of mutation spectrum, allele frequencies and genotype–phenotype relationships (Figure 1). For the potentially treatable disorders, there is usually little opportunity for clinical trials of investigational new drugs (INDs; Figure 1). Many diseases for which testing is available exhibit locus and/or clinical heterogeneity, engendering lengthy and costly differential diagnostic odysseys. Thus, at our institution, many patients undergoing conventional genetic testing may never receive a molecular diagnosis despite, not uncommonly, having undergone US\$10,000 in molecular testing. Furthermore, serial univariate testing can be prolonged, delaying timely intervention or counseling. Scalable, economical, timely, multiplexed testing is needed to effectively translate molecular knowledge from the bench to the bedside.

Conventional molecular testing is most often performed with Sanger sequencing, a method that is not scalable, cost-effective or timely. However, this three-decade gold standard has generated valuable information regarding the clinical validity and utility of mutation analysis. In addition, it has prompted guidelines for laboratory procedures, variant interpretation and the ethical issues concerning the potential benefits and harms of testing minors that are extensible to NGS-based testing [7–10].

Benefits of comprehensive genetic testing

Comprehensive genetic testing is anticipated to extend the benefits demonstrated for conventional genetic testing to a broader range of dominant and recessive diseases. Timely diagnosis of affected individuals has several potential benefits [7,11–14]. While we acknowledge that there are also potential harms of broad genetic testing [7,15], particularly in children, these are beyond the remit of this article. Potential benefits are listed in the following paragraphs.

Potential for prevention of death & diminished disease severity

Effective therapy exists for a number of rare Mendelian diseases if diagnosis and intervention occur early in life. In the USA, newborns undergo screening for 30 core disorders and 25 secondary diseases [13,14,16,102]. Currently, newborn screens result in approximately 6500 diagnoses in the USA per annum, allowing for timely intervention. Examples include phenylketonuria and congenital hypothyroidism, where neonatal diagnosis and treatment prevent severe physical and intellectual disabilities. Likewise, with early diagnosis and treatment, death is prevented in certain forms of congenital adrenal hyperplasia, fatty acid oxidation disorders and galactosemia. Comprehensive diagnostic testing of affected children may provide earlier diagnosis of additional rare childhood diseases not detected by newborn screening, allowing for appropriate early intervention, which in turn may prevent death or diminish disease progression.

Genetic counseling

Identifying the molecular causes of Mendelian disorders will potentially have a significant impact on genetic counseling. Having a molecular diagnosis provides the family access to disease-specific support groups and accurate assessment of recurrence risk. For some disorders, a better understanding of prognosis and interventions may be available. Genetic counseling provides an opportunity to educate families about the natural history and treatment of disease, to explore the many psychosocial issues associated with genetic disease and to coordinate care among multiple subspecialists, as appropriate.

Allows the use of INDs

For some diseases, early diagnosis allows for therapeutic intervention before disease progression is no longer reversible or organ dysfunction occurs. For rare disorders, recruiting sufficient numbers of patients to participate in clinical trials is problematic; it is important that potential participants be identified early to allow for early referral, recruitment and entry into clinical trials of new therapies. Currently, utilization of INDs is constrained by late diagnosis, identification of patients after organ damage has already occurred and by low rates of ascertainment requiring years to complete recruitment and assess impact on outcomes. Finally, stratifying patients by genotype may allow for earlier detection of efficacy of INDs.

Improved quality of life

While many recessive disorders may never have curative treatments, timely diagnosis can nevertheless allow interventions that substantially improve quality of life. Such interventions may slow disease progression, lessen symptoms, prevent complications or improve function in affected organ systems. A specific diagnosis may increase the likelihood that these therapies will be covered by insurance. Knowing a diagnosis may also help parents advocate for school-based services, such as physical, occupational or speech therapy and special education programs. Without a genetic diagnosis, children may not qualify for services until significant delays manifest. For example, the First Steps Missouri program requires an undiagnosed child to display a >50% delay in more than one developmental domain. Similarly, school-aged children with a diagnosis can qualify for an individualized education program under the 'Other Health Impaired' category. Children without a specific diagnosis must meet specific criteria for academic performance, leaving some ineligible, particularly in districts that use a discrepancy model for learning disabilities.

Psychosocial benefits

Some psychosocial issues associated with genetic disease may be alleviated by molecular testing through the resolution of uncertainty, allowing families to confront the issues directly. A specific diagnosis may give the family information about prognosis, providing the opportunity for psychological adjustment and to make realistic plans for the future, including plans for the patient's education, employment, insurance and personal relationships [7]. It may allow families to generate written care plans for severely affected individuals with a shortened lifespan or to plan for eventual guardianship as affected family members enter adulthood. Importantly, other family members can be notified of their potential risk so they can obtain appropriate genetic counseling. A diagnosis also provides the family an opportunity to connect with disease-specific support groups so that they might meet others affected by the same rare disorder and exchange information about useful therapies and educational strategies.

Narrow the differential diagnosis

Multiplexed testing allows many diseases to be ruled out simultaneously, avoiding costly and piecemeal diagnostic searches as well as unnecessary expensive or invasive treatments and procedures. For many diseases, a muscle or liver biopsy is necessary or helpful to narrow down the list of potential candidate genes by immunohistochemistry or other pathological findings. For example, patients with muscle weakness may undergo a muscle biopsy to guide the medical evaluation. If it were economically feasible to perform molecular testing for the dozens of genes on the differential, a faster diagnosis could be made using less invasive procedures. The availability of comprehensive molecular testing for mutations in the *DMD* gene has led to a marked change in the workup for the

dystrophinopathies, moving from biopsy-driven diagnosis to almost exclusive molecular testing.

Increased understanding of disease mechanisms

Molecular diagnosis has immense cumulative potential to lead to an understanding of disease mechanisms. For many Mendelian disorders, genotype–phenotype relationships exist that are relevant to age of onset, disease severity, rates of progression, distribution of affected organs, complications, pleiotropy and outcomes [17–20]. Such knowledge can only be accumulated for diseases in which molecular diagnosis is routinely performed. A broad understanding of genotype–phenotype relationships can lead to personalized or individualized care of patients with genetic diseases. This could include, for example, individualization of treatment intensity to minimize adverse effects, staging of predicted disease progression and prediction of severity and likely complications. For example, *MEN2A* mutations are classified based on their risk for aggressive thyroid cancer. The classification is used to guide recommendations regarding the age at which prophylactic thyroidectomy is performed, in addition to when biochemical screening for pheochromocytoma and hyperparathyroidism should be started [21,22]. The possibility of as yet undefined epistatic modifier genes for Mendelian disorders may change our current view from one of single-gene disorders to a more nuanced view of gene–gene interactions. Comprehensive testing can reveal the presence of variants in many genes concomitantly, allowing research into epistatic effects and thereby potentially further refining personalized therapy.

Improved variant databases

Although decades of molecular testing have led to the accumulation of numerous locus-specific and general research databases (e.g., the Cystic Fibrosis Database, the Tay–Sachs Database, the Single Nucleotide Polymorphism database, the Human Genome Mutation Database and the 1000 Genomes Project), a general clinical-grade mutation database does not currently exist. The development of such a tool is critically important for the interpretation of variant pathogenicity and represents a key bottleneck to the implementation of genomic medicine. This is exemplified by our verification study [23], which revealed 27% (122 out of 460) of literature-annotated disease mutations are either errors (incorrect mutation reported) or variants requiring reclassification as benign polymorphisms based on a frequency of >5% in samples tested and/or homozygosity in samples from people unaffected with the corresponding disease. As a first step towards addressing these problems, we plan to sequence several thousand unaffected individuals to reclassify common benign variants that have been curated as mutations based on the literature. Variants that are confirmed to be homozygous in multiple, unrelated unaffected individuals will be reclassified as benign. In addition, there is currently an active dialog among reference laboratories, the National Center for Biotechnology Information at the National Institutes of Health, the Human Variome Project and the Human Genome Variation Society to develop a clinical-grade mutation database. A tangential benefit of broader molecular testing for rare Mendelian disease will be the cumulative improvement of the quality of public mutation information.

Development of a universal diagnostic test

Children's Mercy Hospital (MO, USA), in collaboration with the Beyond Batten Disease Foundation and the National Center for Genome Resources, has developed a multiplexed NGS-based test for 592 recessive and X-linked childhood diseases [23]. The test was originally designed for comprehensive, economical preconception carrier testing of prospective parents (Table 1) [23–26]; however, it has been modified and expanded to include recessive and X-linked genes useful for the diagnosis of affected children at

Children's Mercy Hospital. Dominant diseases are currently excluded but merit addition in the future for diagnostic testing but not for preconception carrier testing.

While Sanger sequencing has high accuracy and subsequently high sensitivity and specificity for molecular diagnosis, it is not scalable or economical for testing large panels of genes. By contrast, NGS methods coupled with advanced bioinformatic processing allow economical testing of large panels of genes with accuracy levels that rival Sanger sequencing. Highly multiplexed molecular testing of millions of nucleotides in hundreds of genes is now possible, given multiple recent technological advances [23]. As shown in Figure 2, NGS technologies have had a remarkable effect on the cost of DNA sequencing. This trend is anticipated to continue as second-generation sequencing technologies mature and third-generation technologies become available. Commensurate advances in adjunctive technologies, such as target enrichment, molecular barcoding (for sample multiplexing), data storage and alignment algorithms have occurred. Finally, the governmental ruling deeming disease gene sequences found in nature 'unpatentable' is an important step toward the breakup of the monopolization and noncompetitive pricing of genetic testing [27].

The 592-disease test is envisaged to substantially lower the barrier to ordering molecular diagnostic and carrier testing, increasing the potential number of clinical presentations for which testing is ordered and changing the differential diagnostic paradigm from serial additive testing to inclusive, parallel and highly multiplexed testing.

Recessive disease inclusion for genetic testing

The 592 diseases were selected by three main assumptions:

- Cost-effectiveness. Models suggest that the incremental cost associated with adding diseases decreases toward an asymptote. Thus, the broadest coverage of diseases offered optimal analytic cost-benefit (as opposed to stratified panels of diseases or clinical presentations);
- Physician adoption and clinical utility were assumed to be optimal for diseases where diagnosis was likely to guide a medical, counseling or other potentially beneficial intervention for the family;
- Physician adoption and patient satisfaction were assumed to be optimal for diseases that are 'diagnostic dilemmas' (diseases with very similar or overlapping phenotypes with locus heterogeneity).

Database and literature searches as well as expert reviews were undertaken on 1123 diseases with recessive or X-linked inheritance and a known molecular basis [101,103], yielding 592 diseases with childhood onset and a potentially severe clinical course meeting criteria for clinical testing [8,10,11,26,28,102,103]. The latter are: strong evidence that mutations in the gene cause the disease (high clinical sensitivity and specificity, evidence from functional studies supporting pathogenicity and/or inclusion of multiple families or patients from more than one population); and provision of information relevant to the diagnosis, prognosis, treatment, medical surveillance or family planning that benefits the patient and/or the family undergoing testing.

Disease genes were included if mutations caused severe illness (lethal, life-threatening or greatly limiting quality of life) in a proportion of affected children, although we acknowledge the possibility of pleiotrophy and variable severity leading to the detection of mutations that may be less severe or later onset. The panel currently omits substantial sets of recessive disorders, such as most nonsyndromic deafness genes and adult-onset conditions. The expansion of our test menu will be physician-directed and, in the future, influenced by

the utility of the test for specific genes and diseases. Thus, we anticipate the near-term addition of comprehensive gene sets for conditions such as cardiomyopathies, deafness, epilepsy and monogenic diabetes mellitus. Indeed, NGS panels for some of these conditions have already been developed by other clinical laboratories [103]. Looking forward, Mendelian subtypes of complex disorders will be of mounting interest, as their contribution to common disease burden is increasingly recognized. As noted previously, a substantial increase in the number of disease genes exhibiting utility for clinical testing is anticipated in the next few years. For example, many new intellectual disability genes have been identified and will be reported in the literature in 2011 [ROPER H; PERS. COMM.]. Finally, there are certain populations in which the recessive disease burden is higher, either owing to founder effects or consanguinity, reflected in the limited mutation spectrum. The initial intent was to include disorders applicable to the general US population for disease inclusion; however, several diseases of concern to these specific populations were added following consultation with clinical geneticists and other healthcare advocates who indicated a keen awareness of particular diseases in these specific populations and, probably, a higher motivation for comprehensive testing.

Ordering considerations

Our clinical team mandated a pretest stratification of the patient's symptoms and differential diagnosis to limit the analysis of genes not relevant to the phenotype, since this information would not be of diagnostic use and would potentially result in revealing carrier status, which is undesirable in pediatric patients. This approach will also be helpful in narrowing the number of variants requiring interpretation. Capture of standardized data elements that reflect this stratification will be important in order entry to allow some individualization of the automated components of variant interpretation and reporting, including clinical correlation; however, the data entry process must not be too onerous for the ordering physician. We have identified two types of pretest clinical data that may assist in test interpretation and reporting:

- Controlled vocabularies of presenting symptoms and signs. A menu of clinical features from the Systematized Nomenclature of Medicine – Clinical Terms terminology has been assembled for the 592 recessive diseases in the initial panel that are selected by the ordering physician to create a subset of potential differential diagnoses;
- Pretest stratification will be based on suggested diagnostic group(s) of disorders, such as mitochondrial disorders, lysosomal storage diseases, congenital disorders of glycosylation, disordered sexual development, congenital hypotonia, intellectual disability or primary immunodeficiency. The ordering physician is anticipated to also select from such a menu of diagnostic groups that each define a set of diseases within the panel of 592 diseases.

Patients referred for the 592-disease test will meet with a genetic counselor, who will provide pretest counseling to the family and obtain informed consent. The counselor will complete an electronic order form recording patient name, standard demographic information, family history, symptom/diagnostic group(s), differential diagnosis and age of onset of symptoms. Pretest stratification of patients' phenotypes as well as identification of diagnostic groups based on presentation will create a subset of the 592 diseases for which variant information will be reported, alleviating the need for the ordering physician to be familiar with all of the listed diseases. This also assists in interpretation of the clinical relevance of variants and in assessment of clinical presentations, yielding the highest positive-predictive value from such testing.

Genomic sequencing strategy

We chose a targeted NGS approach to provide scalable comprehensive coverage of all genomic regions harboring mutations in the specific genes of interest. The number of mutations listed in the Human Gene Mutation Database (HGMD)[®] is steadily increasing, with a current tally of 102,433 [29]; a previous search indicated that 286 severe childhood autosomal recessive diseases involved 19,640 known mutations [23,29]. This makes a fixed-content method impractical. Array hybridization was initially favored because of its simplicity, cost, scalability and accuracy [30], but given the amount of allelic heterogeneity seen in most rare disorders and poor performance of arrays for detecting insertion-deletion (indel) mutations [29], a more sensitive method was chosen. The effectiveness and decline in cost of exome capture and NGS also influenced the strategy, although standard exome sequencing was not felt to be appropriate for routine molecular diagnosis in affected children. Exome sequencing is better suited for gene discovery rather than molecular diagnosis for the following reasons (Table 1) [1]:

- Exome sequencing will not detect nonexonic variants, some of which are very important for molecular diagnosis and readily detected by conventional genetic tests;
- Exome sequencing identifies carrier status as well as adult-onset disease variants, both of which are undesirable for pediatric testing. The American College of Medical Genetics (ACMG), American Society of Human Genetics, Alliance of Genetic Support Groups, Council of Regional Networks for Genetic Services, International Society of Nurses in Genetics, National Society of Genetic Counselors and American Academy of Pediatrics state: “If the medical or psychosocial benefits of a genetic test will not accrue until adulthood, as in the case of carrier status or adult-onset diseases, genetic testing generally should be deferred” [7]. In these respects, considerations for diagnostic testing in children differ materially from that in adults. The 592-disease test avoids the vast majority of such incidental findings in children by masking genes not relevant to the patient's symptoms;
- Exome sequencing uncovers many other genomic incidental findings, including variants of undetermined significance (VUS). ACMG guidelines mandate reporting of all significant variants and VUS [8,10]. Discovery and reporting of variants that are not relevant to the child's illness (the incidentalome, including nonhealthcare-related traits and risk factors for complex disorders) would be a barrier to adoption by pediatricians and genetic counselors [31]. Identification of incidental findings is minimized in the 592-disease test by masking regions that are irrelevant to the child's symptoms. This makes the pretest stratification step completed by the referring physician even more crucial;
- At approximately US\$2000 per sample, exome sequencing is currently significantly more expensive than the 592-disease test. The current fully loaded cost per sample for the 592 test, including testing and interpretation, is projected to be US\$618, of which US\$183 is consumables. The cost per test is predicted to fall to US\$200 per sample in 2015, of which US\$86 is predicted to be consumables, predicated principally on lower NGS cost-per-gigabase (GB) and economies of scale.

Specimen processing & tracking

Following informed consent, test requests will be submitted in an ordering system that includes standardized collection of clinical phenotypes to guide disease gene selection for

testing. Blood samples will initially be employed, but we plan to investigate the use of other noninvasive sample types, such as buccal smears or saliva, in the future. The workflow, which is depicted in Figure 3, begins with DNA extraction, library preparation and target enrichment which are conducted on dedicated, automated, robotic systems. Our current NGS strategy generates 100-nucleotide reads on an Illumina HiSeq™ 2000 to a depth of >3 GB per sample and approximately 96 samples per run (Table 2). Multiplexed NGS genomic DNA libraries are pooled into 12-plexes and enriched for all coding exons, splice site junctions and other regions known to contain common recurrent mutations in 527 genes corresponding to 592 childhood diseases by hybrid capture target enrichment. In addition, custom baits were designed to target boundaries of known indel mutations >25 nucleotides, as well as recurrent nonexonic mutations. Sequential rounds of bait design can decrease bias at under- and over-represented exons [23]. The platform could eventually be expanded to cover the coding regions of all genes known to cause Mendelian disease (currently 2993 diseases; see Table 2).

We previously reported that hybrid capture of 437 recessive disease genes required 29,891 120mer RNA baits [23]. An empiric goal for target enrichment is to accomplish >30% on-target nucleotides inexpensively, which corresponds to approximately 500-fold enrichment of an approximately 2 million nucleotide target. This was achieved in practice following one round of bait redesign for under-represented exons and decreased bait representation in over-represented exons [18]. An ideal enrichment protocol would give a narrow distribution of target coverage without tails or skewness (indicative of minimal enrichment-associated bias). In practice, the aligned sequence coverage distribution is unimodal but flat and right-skewed (Figure 4B). Thus hybrid capture requires over-sequencing of most targets to recruit a minority of poorly selected targets to adequate coverage levels. Median coverage increases linearly with sequence depth. The proportion of bases with >0 times coverage and >20 times coverage increased toward asymptotes at approximately 99 and 96%, respectively (Figure 4C). Targets with low (<3×) coverage tend to be highly reproducible and have high GC content. Most failing targets may be predicted and 'rescued' by individual PCR reactions or altered hybridization conditions.

Sequence analysis strategy

Sequence alignment and variant detection is performed using Alpheus™ version 3.0 [23,29,32–35], a scalable, modular analysis pipeline written using the Apache™ Hadoop™ software framework [104] and utilizing the map/reduce programming paradigm to enable fully distributed analysis on a compute cluster. Possible alternative software solutions include the Genome Analysis Toolkit from the Broad Institute and Genomics Workbench from CLC Bio. Alpheus uses and produces data in standard file formats such as FASTA for reference sequences, Gene Transfer Format for gene and transcript annotation, and Sequence/Alignment Map [36] for alignments. Variants are reported using a custom, tab-delimited format that will be updated to the standard variant call format in future releases. Alpheus has been developed to professional software engineering standards, including automated unit and functional testing of key components. Formal automated testing helps ensure software quality and allows performance of regression tests during development that provide a level of stability key for Clinical Laboratory Improvement Amendments (CLIA) compliance.

Unambiguous alignment of short sequences is often confounded by repetitive sequences. While the genomic regions targeted by the 592-disease test are overwhelmingly comprised of unique sequences, some loci have pseudogenes or paralogous sequences that can obfuscate unique alignment. Alignment scoring parameters were adjusted to improve alignments to sequences containing polynucleotide variants. These included rewarding

identities (+1) and penalizing mismatches (−1) and indels (−1−log[indel-length]); alignments are retained if the alignment covers at least 95% of the read and has a score of at least 80% of the theoretical highest score possible for that read.

Given the need for highly accurate genotype detection, we developed a decision tree for identifying and genotyping mutations based on previous studies of genotypes identified in NGS genome sequences (Figure 5) [32,33,37,38]. SNPs in 26 samples were genotyped by both Infinium arrays and NGS. The distribution of read count-based allele frequencies of 92,106 SNP calls was trimodal, with peaks corresponding to homozygous reference alleles, heterozygotes and homozygous variant alleles; the optimal genotyping cutoffs were 14 and 86% (Figures 4B & 6B). With these cutoffs and requirements for >20 times coverage and >ten reads of quality >20 to call a variant, the accuracy of sequence-based SNP genotyping was 98.8%, sensitivity was 94.9% and specificity was 99.99%. The positive-predictive value of sequence-based SNP genotypes was 99.96% and the negative-predictive value was 98.5%, as ascertained by array hybridization. At a sequence depth of 0.7–2.7 GB, sensitivity increased from 93.9 to 95.6% (Figure 6A). For known substitution, indel, splicing, gross deletion and regulatory mutations, sensitivity was 100% (113 out of 13 known alleles in 76 samples). Higher sensitivity for known mutations reflected manual curation. Reanalysis of this data indicates that modestly relaxed variant filters improved sensitivity without loss of specificity. Therefore, variants will be retained if present in >eight reads and >14% reads at nucleotides with >16-fold coverage. Heterozygous and homozygous variants are distinguished by a cutoff of 86%. These filters have a specificity >99.95% for single-nucleotide variants and indels <25 nucleotides in length. Fixed filters are more readily conforming to CLIA documentation and validation and will therefore be used rather than Bayesian inference-based algorithms.

While most mutations are single-nucleotide variants and small indels, comprehensive molecular diagnosis also requires sensitive and accurate detection of larger indels, gross deletions, gross insertions and complex rearrangements. Gross deletions are detected through a combination of methods including perfect alignment to mutant junction reference sequences (obviating low alignment scores when short reads containing complex polynucleotide variants are mapped to normal references) and by local changes in coverage (normalized to total sequence generated) [23]. Previous studies have identified copy number variations (CNVs) based on changes in regional coverage along a chromosome in an individual sample [32,33]. However, concomitant analysis of normalized coverage in batches of 192 samples circumvents the need for adjustment for GC content [39], allowing accurate detection of CNVs and deletions. It is anticipated that these approaches may be extensible to gross insertions and complex rearrangements but this requires additional validation.

An interpretation workflow has been designed (Figure 3) [23,34], which will be incorporated into a semi-automated pipeline featuring a user interface with a controlled vocabulary of clinical features for affected patients, as described earlier. In addition, the interface will include a variant table with fields for the following: gene symbol, gene name, disease symbol, Online Mendelian Inheritance in Man (OMIM) number, disease name, variant type, variant reads, average Q score, normal read number at that nucleotide, percentage of reads calling the variant, variant stratification (class and predicted functional consequence as described earlier), variant and chromosomal coordinate, variant and coding coordinate, amino acid change and protein coordinate, scores from *in silico* programs such as SIFT, PolyPhen-2, Splice Site Prediction by Neural Network, Exonic Splicing Enhancer finder, Single Nucleotide Polymorphism database accession, HGMD accession and classification, OMIM accession, Leiden Open Variation Database. or other locus-specific database accession and annotation, including prior annotation. Most of these fields can be

preloaded from relevant public databases or precalculated, providing an almost instant return of queries. While these fields can be dynamically changed [34], for clinical use they will probably be locked at optimized values. Each field and value can feature hyperlinks that open new windows at the relevant web resource and page [34]. Variant reporting is anticipated to be clustered by gene. Clicking on variants could also open a sequence view that shows all alignments of sequences that include that position [34]. This can be useful for checking polynucleotide indels that occasionally 'fool' alignment tools and are broken into two or more adjacent variants. A gene view and disease view page are also featured, which contain link outs to various gene, chromosome, protein and disease databases and browsers and references such as PubMed, OMIM and the London Medical Database [34]. Secondary clinical information, such as family history, race, ethnicity, ancestry, sex, age and age at onset of symptoms, may also be beneficial. Fields can readily be hyperlinked to definitions and gene ontology or other annotation.

Enrollment phenotypic data (symptom group, diagnostic group and suggested diagnosis) will automatically generate an inclusive subset of the 592-disease test panel that is potentially relevant to the patient's diagnosis. Non-relevant disease genes will be deselected (masked) to avoid inadvertent identification of carrier status in genes not affecting the clinical phenotype; however, carrier status discovered for relevant genes will be reported. Masking will considerably reduce the number of variants requiring interpretation.

Variant stratification

Pathologic interpretation and reporting of results will follow the ACMG guidelines for reporting sequencing variants [8,10], which stratify variants as follows:

- Category 1 (known mutations). Previously reported variants recognized as pathogenic;
- Category 2. Previously unreported variants that are of a type expected to cause the disorder (frame shifts, nonsense, disruption of ATG or normal stop codon, splice junction mutations or deletion of one or more exons to shift reading frame);
- Category 3 (VUS). Previously unreported variants that are of a type that may or may not be pathogenic (cryptic splice site mutations, in-frame amino acid insertion/deletion, in-frame exon deletion or any missense mutation);
- Category 4. Previously unreported variants that are unlikely to be disease causing. They include synonymous or intronic variants unlikely to affect splicing. Recognized neutral variants will not be reported.

Variant interpretation

Conventional molecular testing leans heavily on expert variant-by-variant interpretation of pathogenicity and frequently requires additional, unstructured communication between the ordering physician/genetic counselor and laboratory director for interpretation and reporting. For easy interpretation and reporting of variants detected in hundreds of disease genes in samples at high throughput, a new paradigm must be developed where, for example, approximately 90% of variant results can be interpreted automatically, approximately 9% are interpreted manually and the remaining 0.9% require a clinical case conference or similar interactions.

Finding that a patient is homozygous or compound heterozygous for two previously reported pathogenic mutations (category 1), or two novel mutations expected to be pathogenic (category 2) or a combination of the two, in a gene relevant to the phenotype will be sufficient for diagnosis (Figure 5). Clinical correlation will be performed as part of the

laboratory quality assurance and proficiency program. Follow-up testing may be warranted, such as enzyme or other biochemical analysis (Figure 5).

Reporting of category 3 variants is particularly complex, especially for genes for which there is limited knowledge of the mutation spectrum. For some genes, such as those associated with metabolic disorders, interpretation is anticipated to be difficult owing to a preponderance of missense mutations. Biochemical and other testing will be used, as appropriate. The extent to which a sequence variant is considered pathogenic is influenced by a number of factors, including (but not limited to) the level of clinical suspicion for the associated disorder, segregation of the variant with phenotype in the family, the position and evolutionary conservation of the affected amino acid, co-occurrence with a deleterious mutation, evaluation of chromosomes from an equivalent population, as well as frequency data based on cumulative experience. *In silico* tools such as SIFT, PolyPhen-2, Splice Site Prediction by Neural Network and Exonic Splicing Enhancer finder will be used as appropriate. Category 3 variants will be further grouped into 3a and 3b according to these data, with 3a being more likely to be pathogenic than 3b. Despite all the available data and tools at hand, there will still be VUS, which will be reported as such and classified as either category 3 or 4, using the aforementioned criteria. The supporting evidence used to make this determination will be provided in the report, along with language stating whether the variant is likely or unlikely to be pathogenic with a caveat that the tools used are not definitive. Variants classified as group 4 will be reported as VUS unlikely contributing to disease. The maturity of the reference database will increase with time by recording the dynamic cumulative frequency of each variant in each ethnic group, which will be helpful in reassigning frequent VUS as benign.

Reporting combinations of category 1 or 2 variants with VUS also requires manual evaluation (Figure 5). Many factors must be considered, including clinical correlation, prevalence of the disease in question, mutation spectrum and the other factors considered in the interpretation of VUS. In some cases, a single group 1 or 2 variant with a specific phenotype may be diagnostic. For example, a single frameshift mutation in the *MUNC13-4* gene in a patient with symptoms of familial hemophagocytic lymphohistiocytosis is considered diagnostic, given the clinical correlation and the observation that 50% of patients with *MUNC13-4*-related hemophagocytic lymphohistiocytosis test positive for just one mutation [40]. In such cases, it is assumed that the second allele was undetected and further supportive testing may be warranted. By contrast, a child with failure to thrive and one pathogenic mutation in the *CFTR* gene (carrier frequency 1/29 in the general population) would not necessarily be diagnosed with cystic fibrosis, particularly in the absence of positive sweat chloride testing. Results where any category 1 or 2 mutation combined with VUS considered unlikely to be pathogenic must also be evaluated on a case-by-case basis. Patients with ambiguous results will be discussed anonymously at a biweekly case conference. For novel variants, further studies such as functional assays or family studies may be helpful to determine pathogenicity.

Our reporting interface will incorporate pull-down menus, clickable boxes and write-in text fields for entry of interpretation of the automated content. Terms, definitions and menu options will need to be defined for these pages in order to minimize write-ins and provide maximal standardization of interpretation. Examples of key terms will be pathogenic, likely pathogenic and uncertain. The lines of evidence supporting the key terms will be required. Each variant will require an electronic 'interpretation'. Interpretation will have four phases: automated interpretation (by Alpheus); primary interpretation (by a variant specialist/technician); review/approval by the laboratory director; and secondary team review in patients with ambiguous results. Action boxes will also be available for ordering follow-up studies, such as repeat or confirmatory sequencing or functional testing.

Reporting of genetic variants to physicians

The 592-disease test interpretation and reporting tools have been in development for almost 5 years [23,29,32–35,41,42]. Originally designed for NGS data analysis by nonexperts, we anticipate somewhat straightforward adaptation for 592-disease test interpretation and reporting. A principal challenge in developing interpretation and reporting tools is to present data in a manner that is prioritized and staged so that it is not overwhelming to users. However, it must retain the functionality needed for interpretation and reporting of the vast majority of variants using minimal nonstandardized text. A succinct tabular electronic report is in development, which will contain the key findings, interpretation and supporting evidence with hyperlinks to disease information related to genetic counseling, inheritance, disease progression, treatment and resources.

Need for confirmation of positive findings in a clinical laboratory

The 592-disease test will be offered as a laboratory developed test in our clinical laboratory, which has the requisite facilities and equipment for CLIA compliance, including unidirectional work-flow, positive pressure for preamplification rooms, containment hoods, physical separation of pre- and post-PCR steps, conditioned emergency power and environmental controls. Analytic validation of the 592-disease test is in progress and will include blinded testing of approximately 950 archived patient samples with known mutations previously identified through clinical testing. Previous verification of the NGS procedure for the 448-disease version of the test was performed in a research setting [23]. Based on this study, we do not anticipate a need to carry out confirmatory Sanger sequencing for nucleotide substitution, indel and exonic deletion results [23], although this will be defined by formal validation studies in our clinical laboratory. Confirmatory testing will be needed for CNVs in a clinical microarray or cytogenetics laboratory by array comparative genomic hybridization, quantitative PCR or fluorescence *in situ* hybridization, as appropriate.

Conclusion

The 592-disease test will be a strategically important tool to enable the practice of personalized genomic diagnosis for orphan pediatric disease. The lessons and results of these studies will be extensible to other applications, such as expanded newborn screening or comprehensive preconception carrier screening. Thus, early molecular diagnosis will potentially improve the treatment for childhood recessive diseases by increasing ascertainment, allowing stratification in INDs and delivering molecular diagnosis before organ involvement. In those diseases for which treatment is available, this will theoretically allow time for referral and early intervention before disease progression.

Expert commentary

Molecular diagnosis of Mendelian diseases in affected children will soon be dramatically impacted by the remarkable power of NGS allied with advanced computational analysis. This is one of the first areas of medicine that will be transformed by genomic medicine. For affected children and their families, this is likely to mean faster diagnosis and greater likelihood of definitive diagnosis. For healthcare systems, there will potentially be diagnostic cost savings. However, resources for genetic counseling and pathologic interpretation will be strained by the greater test volumes anticipated for some time.

Five-year view

Over the next 5 years, NGS allied with advanced computational analysis will become the 'gold standard' for the molecular diagnosis of Mendelian diseases in affected children. Testing will probably be centralized at regional reference laboratories, although point-of-care devices for rapid diagnosis will be available for a select subset of diseases and clinical scenarios. This will have had several impacts on healthcare provision. First, medical students, pediatricians, clinical pathologists and genetic counseling students will require considerably more instruction about the technology and its interpretation. Second, family physicians and generalist pediatricians will increasingly make initial diagnoses of uncommon Mendelian diseases using such technologies and will then refer those patients to specialists for further assessment and treatment. There will be an immense need for genetic counselors. Regional centers of excellence will emerge for the assessment and treatment of specific classes of Mendelian disorders.

In light of enrichment costs and the associated burden of over-sequencing, whole-genome sequencing will probably become a cost-effective alternative to target enrichment in the next 5 years. The promise of the US\$1000 genome will become a reality as the cost of sequencing continues to decrease and impending technological advances materialize. Whole-genome NGS has advantages of broadest coverage of disease genes and regions harboring mutations and better-developed tools for CNVs. Whole-genome sequencing for the molecular diagnosis of childhood diseases, newborn screening and preconception carrier testing will be technically and economically feasible. Thus, test formats, stratification of panels and indications for testing are likely to change substantially compared with what has been described in this article. Bioinformatic analysis and associated data handling will change dramatically and new standards will be established. A key development will be the establishment of a clinical-grade database of mutations that is broadly accepted as integral to variant interpretation. Again, the requirement of multimillion dollar computer facilities to analyze genome sequences that cost US\$100 will result in centralized regional testing for the majority of the population.

As the rate of gene discovery continues to increase through exome and genome sequencing of individual cases, we predict that the causal genes for approximately 15,000 Mendelian disorders will be identified in the next 5 years. Genotype-phenotype relationships will be understood for several thousand Mendelian disorders. Ongoing research in Mendelian diseases will also increase our understanding of the function of noncoding variants and their relationship to phenotype. Many epistatic modifier genes for recessive illnesses will be identified, leading us to change our perception of Mendelian disorders from being single-gene conditions to a more nuanced view. Some diseases that are today regarded as common and complex will become reclassified as hundreds of specific Mendelian or semi-Mendelian disorders. These will include conditions such as intellectual disability, autism spectrum disorders, mental illness, and other neurologic and psychiatric disorders.

Acknowledgments

The authors would like to acknowledge the Children's Mercy Genomic Medicine Team: Ahmed Abdelmoity, Michael Artman, Mara Becker, Michael Begleiter, Julia Bracken, Shannon Carpenter, Christie Ciacio, Mark Clements, Bryce Heese, John Lantos, Steven Leeder, Jean Baptiste LePichon, Ann Modrcin, Laurie Smith and Tarak Srivastava. Hans Hilger Ropers at the Max Planck Institute in Berlin, Germany, is also part of this team.

References

Papers of special note have been highlighted as:

- of interest
- of considerable interest

1. Kingsmore SF, Saunders CJ. Deep sequencing of patient genomes for disease diagnosis: when will it become routine? *Sci. Transl. Med.* 2011; 3(87):87ps23. •• Provides a good overview of the current status of next-generation sequencing (NGS) for clinical use, including barriers to routine adoption.
2. Costa T, Scriver CR, Childs B. The effect of Mendelian disease on human health: a measurement. *Am. J. Med. Genet.* 1985; 21(2):231–242. [PubMed: 4014310]
3. Lialiaris T, Mantadakis E, Kareli D, Mpountoukas P, Tsalkidis A, Chatzimichail A. Frequency of genetic diseases and health coverage of children requiring admission in a general pediatric clinic of northern Greece. *Ital. J. Pediatr.* 2010; 36:9. [PubMed: 20205810]
4. McCandless SE, Brunger JW, Cassidy SB. The burden of genetic disease on inpatient care in a children's hospital. *Am. J. Hum. Genet.* 2004; 74(1):121–127. [PubMed: 14681831]
5. Yoon PW, Olney RS, Khoury MJ, Sappenfield WM, Chavez GF, Taylor D. Contribution of birth defects and genetic diseases to pediatric hospitalizations. A population-based study. *Arch. Pediatr. Adolesc. Med.* 1997; 151(11):1096–1103. [PubMed: 9369870]
6. Berry RJ, Buehler JW, Strauss LT, Hogue CJ, Smith JC. Birth weight-specific infant mortality due to congenital anomalies, 1960 and 1980. *Public Health Rep.* 1987; 102(2):171–181. [PubMed: 3104974]
7. Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. American Society of Human Genetics Board of Directors, American College of Medical Genetics Board of Directors. *Am. J. Hum. Genet.* 1995; 57(5):1233–1241. No authors listed. [PubMed: 7485175]
8. Maddalena A, Bale S, Das S, Grody W, Richards S. Technical standards and guidelines: molecular genetic testing for ultra-rare disorders. *Genet. Med.* 2005; 7(8):571–583. [PubMed: 16247296] • Provides guidance for clinical laboratory testing and reporting of variants in ultra-rare disorders.
9. Zoccoli, MA.; Chan, M.; Erker, JC.; Ferreira-Gonzalez, A.; Lubin, IM. Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine; Approved Guideline. NCCLS; PA, USA: 2004. NCCLS document MM9-A
10. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet. Med.* 2008; 10(4):294–300. [PubMed: 18414213]
11. Grosse SDKL, Khoury MJ. Evaluation of the validity and utility of genetic testing for rare diseases. *Adv. Exp. Med. Biol.* 2010; 686:115–131. [PubMed: 20824443]
12. Heshka JT, Palleschi C, Howley H, Wilson B, Wells PS. A systematic review of perceived risks, psychological and behavioral impacts of genetic testing. *Genet. Med.* 2008; 10(1):19–32. [PubMed: 18197053]
13. Newborn screening: toward a uniform screening panel and system. *Genet. Med.* 2006; 8(Suppl. 1): 1S–252S. No authors listed. [PubMed: 16783161]
14. American College of Medical Genetics Newborn Screening Expert Group. Newborn screening: toward a uniform screening panel and system – executive summary. *Pediatrics.* 2006; 117(5 Pt 2):S296–S307. [PubMed: 16735256]
15. Dhondt JL. Expanded newborn screening: social and ethical issues. *J. Inherit. Metab. Dis.* 2010; 33(Suppl. 2):S211–S217. [PubMed: 20544288]
16. McCabe LL, Therrell BL Jr, McCabe ER. Newborn screening: rationale for a comprehensive, fully integrated public health system. *Mol. Genet. Metab.* 2002; 77(4):267–273. [PubMed: 12468271]
17. Aartsma-Rus A, Van Deutekom JC, Fokkema IF, Van Ommen GJ, Den Dunnen JT. Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. *Muscle Nerve.* 2006; 34(2):135–144. [PubMed: 16770791]
18. Braun AT, Farrell PM, Ferec C, et al. Cystic fibrosis mutations and genotype-pulmonary phenotype analysis. *J. Cyst. Fibros.* 2006; 5(1):33–41. [PubMed: 16275171]

19. Skordis N, Shammas C, Efstathiou E, Kaffe K, Neocleous V, Phylactou LA. Endocrine profile and phenotype-genotype correlation in unrelated patients with non-classical congenital adrenal hyperplasia. *Clin. Biochem.* 2011; 44(12):959–963. [PubMed: 21635882]
20. Wedell A. Molecular genetics of 21-hydroxylase deficiency. *Endocr. Dev.* 2011; 20:80–87. [PubMed: 21164261]
21. Moore FD, Dluhy RG. Prophylactic thyroidectomy in MEN-2A – a stitch in time? *N. Engl. J. Med.* 2005; 353(11):1162–1164. [PubMed: 16162888]
22. Kloos RT, Eng C, Evans DB, et al. Medullary thyroid cancer: management guidelines of the American Thyroid Association. *Thyroid.* 2009; 19(6):565–612. [PubMed: 19469690]
23. Bell CJ, Dinwiddie DL, Miller NA, et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci. Transl. Med.* 2011; 3(65):65ra64. •• Demonstrates proof of principle for a highly multiplexed NGS-based test currently being validated clinically.
24. Gross SJ, Pletcher BA, Monaghan KG. Carrier screening in individuals of Ashkenazi Jewish descent. *Genet. Med.* 2008; 10(1):54–56. [PubMed: 18197057]
25. Pletcher BA, Bocian M. Preconception and prenatal testing of biologic fathers for carrier status. American College of Medical Genetics. *Genet. Med.* 2006; 8(2):134–135. [PubMed: 16481898]
26. McCabe LL, McCabe ER. Genetic screening: carriers and affected individuals. *Annu. Rev. Genomics Hum. Genet.* 2004; 5:57–69. [PubMed: 15485343] •• This article, which advocates using newborn screening as a model system for all genetic testing, gives a brief history of newborn screening and outlines issues associated with genetic testing, particularly population screening.
27. Association for Molecular Pathology et al. v. United States Patent and Trademark Office et al. 09-cv-04515. NY, USA: 2010.
28. Khoury MJ, Coates RJ, Evans JP. Evidence-based classification of recommendations on use of genomic tests in clinical practice: dealing with insufficient evidence. *Genet. Med.* 2010; 12(11):680–683. [PubMed: 20975567]
29. Sugarbaker DJ, Richards WG, Gordon GJ, et al. Transcriptome sequencing of malignant pleural mesothelioma tumors. *Proc. Natl Acad. Sci. USA.* 2008; 105(9):3521–3526. [PubMed: 18303113]
30. Srinivasan BS, Evans EA, Flannick J, et al. A universal carrier test for the long tail of Mendelian disease. *Reprod. Biomed. Online.* 2010; 21(4):537–551. [PubMed: 20729146] •• Describes the introduction of a highly multiplexed carrier test using an array-based method.
31. Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. *JAMA.* 2006; 296(2):212–215. [PubMed: 16835427] •• Outlines the practical issues of broad sequence-based testing, especially for rare disorders, in the general population.
32. Baranzini SE, Mudge J, van Velkinburgh JC, et al. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature.* 2010; 464(7293):1351–1356. [PubMed: 20428171]
33. Kim JI, Ju YS, Park H, et al. A highly annotated whole-genome sequence of a Korean individual. *Nature.* 2009; 460(7258):1011–1015. [PubMed: 19587683]
34. Miller NA, Kingsmore SF, Farmer A, et al. Management of high-throughput DNA sequencing projects: Alpheus. *J. Comput. Sci. Syst. Biol.* 2008; 1:132. [PubMed: 20151039]
35. Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature.* 2008; 456(7221):470–476. [PubMed: 18978772]
36. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009; 25(16):2078–2079. [PubMed: 19505943]
37. Dahl F, Stenberg J, Fredriksson S, et al. Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc. Natl Acad. Sci. USA.* 2007; 104(22):9387–9392. [PubMed: 17517648]
38. Summerer D, Hevroni D, Jain A, et al. A flexible and fully integrated system for amplification, detection and genotyping of genomic DNA targets based on microfluidic oligonucleotide arrays. *Nat. Biotechnol.* 2010; 27(2):149–155.
39. Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. Analysis of the size distributions of fetal and maternal cell-free DNA by paired-end sequencing. *Clin. Chem.* 2010; 56(8):1279–1286. [PubMed: 20558635]

40. zur Stadt U, Schmidt S, Kasper B, et al. Linkage of familial hemophagocytic lymphohistiocytosis (FHL) type-4 to chromosome 6q24 and identification of mutations in syntaxin 11. *Hum. Mol. Genet.* 2005; 14(6):827–834. [PubMed: 15703195]
41. Mudge J, Miller NA, Khrebtukova I, et al. Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. *PLoS ONE.* 2008; 3(11):e3625. [PubMed: 18985160]
42. Zhang G, Beck BB, Luo W, Wu F, Kingsmore SF, Dai D. Development of a phylogenetic tree model to investigate the role of genetic mutations in endometrial tumors. *Oncol. Rep.* 2011; 25(5): 1447–1454. [PubMed: 21327331]

Websites

101. [(Accessed 27 May 2011)] OMIM®: Online Mendelian Inheritance in Man. www.ncbi.nlm.nih.gov/omim
102. [(Accessed 27 May 2011)] National Newborn Screening & Genetics Resource Center. <http://nnsis.uthscsa.edu>
103. [(Accessed 2 June 2011)] GeneTests. www.genetests.org
104. [(Accessed 2 June 2011)] Apache™ Hadoop™. <http://hadoop.apache.org>

Key issues

- Although individually rare, Mendelian diseases collectively account for significant morbidity and mortality, and are responsible for approximately 20% of infant mortality and approximately 10% of pediatric hospitalizations. In total, 13 million people are estimated to be affected by Mendelian disease in the USA.
- Current traditional molecular univariate or small panel genetic testing is often prolonged, expensive and fails to provide a diagnosis for many patients.
- Comprehensive genetic testing offers numerous benefits over traditional testing, including improved testing yields, more effective genetic counseling, possible prevention of death or decreased disease severity in some cases, and rapid improvement in variant databases.
- Recent advances in next-generation sequencing and bioinformatic analysis have allowed for the development of an inexpensive universal diagnostic test that can simultaneously test for pathogenic mutations in 592 diseases.
- Pretest stratification and masking of genes not consistent with clinical presentation is possible by physician entry of standardized data elements at the time of ordering.
- The 592-disease test will utilize multiplexed, next-generation sequencing of 527 hybrid-enriched genes to a level of at least 3 gigabases of sequence per sample.
- We have previously achieved a genotyping accuracy of 98.8%, sensitivity of 94.9% and specificity of 99.99%, with an at least 20 times coverage for heterozygous base calls using Alpheus™ and a cutoff of 14–86%.
- Variant interpretation for a large number of poorly characterized genes is highly complex and will require the use of a clinical-grade variant database, which is currently under development, and multiple lines of evidence to support a disease-causing annotation.

Figure 1. Current circle of hopelessness of orphan diseases
R&D: Research and development.

Figure 2. Cost per gigabase of DNA sequence
Courtesy of the National Human Genome Research Institute.

Figure 3. 592-disease test workflow

LIMS: Laboratory information management system; QC: Quality control; qPCR: Quantitative PCR; UDT: Universal diagnostic test; UV: Ultraviolet.

Figure 4. Analytic metrics of the 592-disease test
 GB: Gigabase. Modified with permission from [23].

Figure 5. Decision tree to classify sequence variation and evaluate genotype
(A) Variant detection and genotyping, (B) variant categorization and (C) variant interpretation for recessive disorders. BLOSUM: Blocks of Amino Acid Substitution Matrix; bp: Base pair; Dx: Diagnosis; GSNAP: Genomic Short-read Nucleotide Alignment Program; Indel: Insertion-deletion; NCBI: National Center for Biotechnology Information; SD: Standard deviation.

Figure 6. Diagnostic metrics of the 592-disease test

GB: Gigabase; Het: Heterozygotes; Hom: Homozygotes; NPV: Negative-predictive value; PPV: Positive-predictive value; Ref: Homozygous reference alleles. Modified with permission from [23].

Table 1

Properties of single-gene, 592-disease test and exome molecular testing: comparison of potential approaches.

Property	Serial single MDx	UDT	Exome sequencing
Dx non-exonic mutations	Yes	Yes	No
Year 1 diseases covered/test	1	592	~3000 [†]
Year 4 diseases covered/test	1	>1000	>3000
Year 1 cost (US\$)	1000→10,000	618	~2000
Year 4 cost (US\$)	1000→10,000	<200 [‡]	~350 [‡]
Year 1 time-to-Dx	4 weeks→6 months	4 weeks	4 weeks
Year 4 time-to-Dx	4 weeks→6 months	2 weeks	2 weeks
Year 4 exonic haplotypes	No	Yes	Yes
Builds knowledge base	Very slowly	Rapidly	Very rapidly
Clinically irrelevant information	Little	Some	Lots
VUS	Some	More	Lots

Dx: Diagnosis; MDx: Molecular diagnosis; UDT: Universal diagnostic test; VUS: Variants of undetermined significance.

Data taken from [1].

[†]Covers more diseases but not all have sufficient evidence for pathogenicity/clinical utility.

[‡]Cost of sequence generation is a small component of year 4 test cost.

Table 2

Annual 592-disease test metrics.

Metric	Time point						
	February 2011		Year 1	Year 2	Year 3	Year 4	
CLIA-conforming	No	Yes	Yes	Yes	Yes	Yes	
Number of diseases	448	592	700	850	1000		
Turnaround time (weeks)	6	4	3	2	2		
Cost (US\$)	ND	618	ND	205	ND		
Batch size	96	192	ND	ND	ND		
Multiplexing	12	12	ND	ND	ND		
Fold enrichment	442×	500×	525×	550×	575×		
Assembled read length (nt)	149	200	250	Exonic haplotypes	Exonic haplotypes		
Sequence/sample (GB)	1.5	2	3	Optimized	Optimized		
O× coverage nucleotides (%)	2.30	1.80	1.50	1.20	<1		
16× coverage nucleotides (%)	94	95	97	98	>99		
SNP genotype sensitivity (%)	95.6	95.6	96	97.5	99		
SNP genotype specificity (%)	99.99	99.99	99.99	99.99	99.99		
% failure (all sources)	ND	10	8	6	<5		
Sample/year/module	ND	5184	ND	20,000	ND		

CLIA: Clinical Laboratory Improvement Amendments; GB: Gigabase; ND: Not determined; nt: Nucleotide.
Data taken from [23].