Sha Tang[1], Kelly Farwell[1], Layla Shahmirzadi[1], Dima El-Khechen[1], Zoe Powis[1], Stephanie Gandomi[1], Ira Lu[1], Cameron Mroske[1], Ruth Baxter[1], Xiang Li[1], Brigette Tippin[1], Wenqi Zeng[1], Elizabeth C. Chao[1,2]
1) Ambry Genetics Corporation, Aliso Viejo, CA 92656; 2) Department of Pediatrics, University of California, Irvine, Irvine, CA 92697

Mutations in Newly Discovered Mendelian Disease Genes Represent a Significant Portion of Positive Findings in Diagnostic Exome Sequencing (DES)

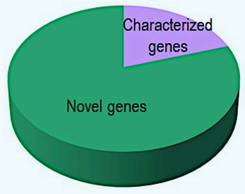
BACKGROUND

- Based on the research studies for the recent 5 years and the clinical experience for over 2 years, whole exome sequencing has been remarkably successful as both a diagnostic and novel gene discovery tool.
- OMIM phenotypes for which the molecular basis is known almost doubled in the recent 6 years, mostly attributing to the utilization of whole exome sequencing. The pace for novel disease-causing gene discovery has been accelerating and new genes are discovered weekly.
- Translating the most up-to-date research discoveries into diagnostic exome sequencing (DES) analysis is extremely important for improving diagnostic sensitivity.
- Instead of entirely relying on the most recent version of HGMD and/or OMIM Morbid databases, we employ an internal database for known disease-causing genes and related phenotypes. Our database features a dynamic design for shuffling genes between "characterized" and "novel" in real time and is curated on a weekly basis to incorporate the latest discoveries.
- The majority of DES tests performed in our laboratory involve whole exome sequencing of the proband plus two first-degree relatives. The resultant well-defined gene/alteration list after trio filtering enables the possibility for thorough literature search of each gene prioritized, minimizing the chance of missing a positive finding in a newly discovered locus.

METHODS

- **Patients:** Genomic deoxyribonucleic acid (gDNA) was isolated from whole blood from the proband and relatives referred to Ambry Genetics (Aliso Viejo, CA) for diagnostic exome sequencing (DES). Informed consent was obtained from all family members involved in the testing process.
- **Whole exome sequencing:** Samples were prepared using the SureSelect Target Enrichment System (Agilent Technologies, Santa Clara, CA) or or SeqCap EZ VCRome 2.0 (Roche NimbleGen, Madison, WI). The enriched exome libraries were sequenced using paired-end, 100-cycle chemistry on the Illumina HiSeq 2000 or 2500 (Illumina, San Diego, CA).
- **Characterized and Disease-causing (ChAD) and Novel gene databases**: The Characterized and Disease-causing (ChAD) gene database was curated on a weekly basis to include genes currently known to be responsible for causing Mendelian disease. The ChAD database included genes which are associated with syndromes listed in the Human Gene Mutation Database (HGMD) (Stenson, 2009) and the Online Mendelian Inheritance in Man (OMIM) database. Novel genes were defined as those not known to underlie a Mendelian condition at the time of data analysis. Any RefSeq gene not included in the ChAD database was included in the novel gene database.
- **Bioinformatics annotation, filtering of variants, and Family history Inheritance-based Detection (FIND):** HGMD, OMIM, the Single Nucleotide Polymorphism database (dbSNP) (Sherry, 2001), 1000 genomes, HapMap data (International HapMap, 2003) and online search engines (e.g., PubMed) were used to search for previously described gene mutations and polymorphisms. Stepwise filtering included the removal of common SNPs, intergenic and 3′/5′ UTR variants, non-splice-related intronic variants, and lastly synonymous variants. Variants were then filtered further based family history and possible inheritance models using the informatics program "FIND" (Family history Inheritance-based Detection).
- **Personalized Medical Review with Enhanced and Comprehensive Assessment (PRECISE) of potentially causal variants:** Each candidate alteration was assessed by a molecular geneticist to identify the most likely causative mutation(s) using the "PRECISE" (Personalized Medical Review with Enhanced and Comprehensive Assessment) analysis method. In brief, interpretive filtering was based on the deleterious nature of the candidate alterations, literature search, and analysis of the relevance of the candidate genes' function in relation to the patient's phenotype. Most candidate alterations undergo Sanger sequencing confirmation and familial co-segregation analysis.
- **Statistical Analyses:** Statistical significance was computed by chi^2 goodness of fit tests and Fisher's Exact Probability.

Design and Algorithm for Characterized Gene Database



Ambry Characterized Gene Database

DATA IN:

- HGMD database
- OMIM-Morbid disease association
- Genes with clinical characterizations reported in peer-review publications **(Weekly Updates)**

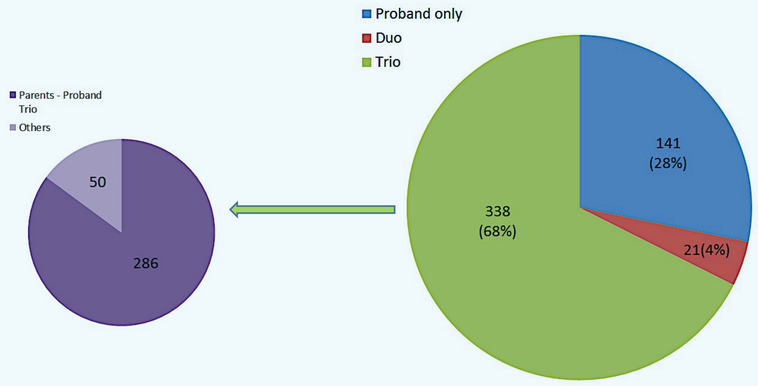
Ambry Characterized Gene Database

DATA OUT:

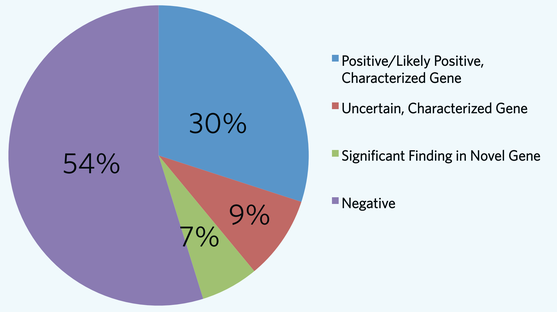
- Association risk alleles
- Genes with no clear clinical association in humans
- Genes with somatic mutation reports only

Summaries for the First 500 Exome Cases at Ambry Genetics

152 patients (30.4%) received a molecular diagnosis in characterized genes; significant findings in clinically novel genes identified in 7% patients



67.6% of the exome cases analyzed at Ambry Genetics were based on whole exome sequencing of the proband plus two first-degree relatives (Trio sequencing), mostly parents-proband Trio



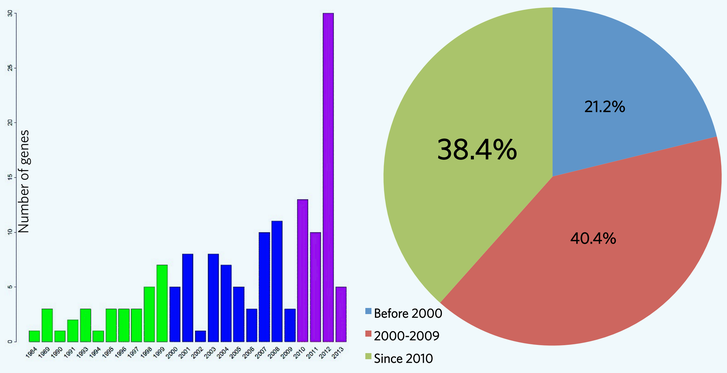
Characterized Gene(s)

POSITIVE	Relevant Alteration(s) Detected
LIKELY POSITIVE	Relevant Alteration(s) Detected
UNCERTAIN	Alteration(s) of Uncertain Clinical Significance Detected
NEGATIVE	No Relevant Alterations Detected

Novel Gene

LIKELY POSITIVE	Relevant Alteration(s) Detected (Novel Gene)
POSSIBLY POSITIVE	Alteration(s) of Uncertain Clinical Significance Detected (Novel Gene)

38.4% Positive Cases Harbored Mutations in Genes Discovered Since 2010



- Since the proof-of-concept results of the clinical utility of whole exome sequencing was published in 2009, the pace for novel disease-causing gene identification has been rapid.
- When plotting the numbers of genes implicated in our first 152 Positive/ Likely Positive cases according to the year they were first published to cause Mendelian disorders, we found that 38.4% are genes discovered since 2010.

Mutations in Newly Characterized Genes Detected in our Cohort

35/152 (23%) positive cases are associated with genes first published since 2012

Gene	RefSeq	Year First Published	Occurrence	Associated Clinical Syndrome(s)
ACO2	NM_001098	2012	1	Infantile cerebellar-retinal degeneration
ACTG2	NM_001615	2012	3	Berdon Syndrome (Microcolon intestinal hypoperistalsis)
ADCY5	NM_183357	2012	1	Autism and Familial dyskinesia with facial myokymia
ALG13	NM_001099922	2012	1	Congenital disorder of glycosylation, type Is
ANK2	NM_001148	2012 (for autism)	1	*ANK2*-Related Autism
*ANO3**	NM_031418	2012	1	Dystonia 24
ARID1A	NM_006015	2012	1	Coffin-Siris syndrome
ARID1B	NM_020732	2012	2	Coffin-Siris syndrome
BRAT1	NM_152743	2012	1	Rigidity and multifocal seizure syndrome
CHD2	NM_001271	2012 (for autism)	1	Epileptic encephalopathy, childhood-onset and Autism
CHD8	NM_001170629	2012	1	Autism
EXOSC3	NM_016042	2012	1	Pontocerebellar hypoplasia, type 1B
HCFC1	NM_005334	2012	1	Mental retardation, X-linked 3
KMT2A	NM_001197104	2012	4	Wiedemann-Steiner Syndrome
NGLY1	NM_018297	2012	1	Rigidity and multifocal seizure syndrome
PACS1	NM_018026	2012	2	Mental retardation, autosomal dominant 17
PIK3R2	NM_005027	2012	1	Megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome
SF3B4	NM_005850	2012	1	Acrofacial dysostoses, Nager Syndrome
SLC52A2	NM_024531	2012	1	Brown-Vialetto-Van Laere Syndrome
TMEM231	NM_001077416	2012	1	Joubert syndrome
TUBB4A	NM_006087	2012	2	Hypomyelinating Leukodystrophy 6
UBE3B	NM_130466	2012	2	Blepharophimosis-Ptosis-Intellectual Disability Syndrome
AMPD2	NM_004037	2013	1	*AMPD2*-related neurodegenerative brainstem disorder
GNAO1	NM_020988	2013	2	Hypomyelinating Leukodystrophy with Atrophy of the Basal Ganglia and Cerebellum
*NALCN**	NM_052867	2013	1	Neuroaxonal neurodegeneration, infantile, with facial dysmophism
ZNF335	NM_022095	2013	1	Microcephaly 10, primary, autosomal recessive

* in the same patient .

Genes in Purple: No clinical tests, either single gene or panel sequencing, are available. Genetic Testing Registry accessed 03/10/2014.
In these 21 cases (21/152 = 13.8% of positive cases) , DES represents the sole available clinical approach to identify the causative molecular etiology.

TAKE-HOME POINTS

- Newly characterized genes account for a significant portion of the positive cases in our cohort. The prevalence of some of the newly discovered genes may be pretty frequent – recurrence has been observed in many of the newly characterized genes.
- The rapid pace of new disease gene discovery, the significant ratio of positive findings in new disease genes, and the recurrence of many of the just characterized loci illustrated in our cohort accentuate the importance of curating an up-to-date disease gene database. Our observations also highlight the superiority of DES over gene panel test in clinical sensitivity, especially in conditions of genetic heterogeneity. The ongoing efforts to "fill in the gaps" not well covered in current DES platforms may further justify improved DES as the first-tier test.

REFERENCES

- 1000 genomes database: A map of human genome variation from population-scale sequencing. *Nature* **467**,:1061-1073.
- International HapMap (2003) The International HapMap Project. *Nature* **426**:789-796.
- Sherry ST, *et al.* (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**,:308-311.
- Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), [date]. World Wide Web URL: http://omim.org/
- Stenson PD, *et al.* (2009) The Human Gene Mutation Database: 2008 update. Genome medicine **1**:13.