Kelly Gonzalez[1], Sha Tang[1], Cameron Mroske[1], Layla Shahmirzadi[1], Dima El-Khechen[1], Zoe Powis[1], Ruth Baxter[1], Brigette Tippin[1], Wenqi Zeng[1], Elizabeth C. Chao[1,2]
1) Ambry Genetics, Aliso Viejo, CA
2) Department of Pediatrics, University of California, Irvine, CA 92697

Diagnostic Exome Sequencing (DES) Unravels Novel Gene Findings in a Significant Portion of Previously Undiagnosed Patients

BACKGROUND

- Diagnostic exome sequencing (DES) is successful in solving the diagnostic odyssey for 30-40% of undiagnosed patients with underlying Mendelian disorders.
- DES is uniquely useful in overcoming limitations posed by traditional molecular diagnostic strategies in the identification of novel gene findings.
- The evidence used for the evaluation of the involvement of the novel gene findings with the patient's phenotypes most commonly included animal model phenocopies and/or genes located within described microdeletion syndromes.

METHODS

- **Patients/study population and whole exome sequencing: :** Genomic deoxyribonucleic acid (gDNA) was isolated from whole blood from probands and relatives referred to Ambry Genetics (Aliso Viejo, CA) for diagnostic exome sequencing (DES). 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 (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 human 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 mutation 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.
- Patient demographics (age, gender, referral indication) were collected from required testing documents supplied to the laboratory with the requisition form and the biospecimens. Patient identifiers were removed. Data curation included the primary exome test option ordered, diagnosis and/or clinical description, and exome sequencing results, including gene(s) and alteration(s). Statistical analyses were computed by chi^2 goodness of fit tests and Fisher's Exact Probability.

Figure 1: DES Detection Rates

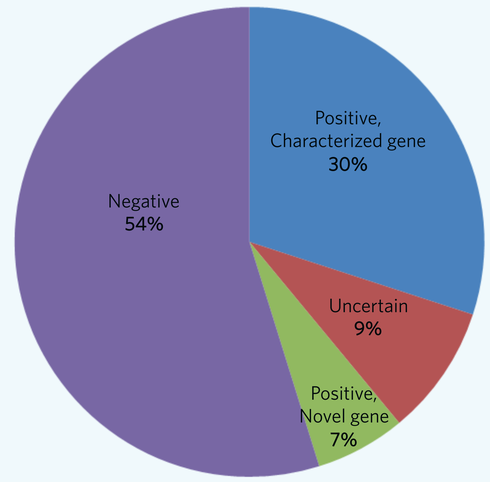
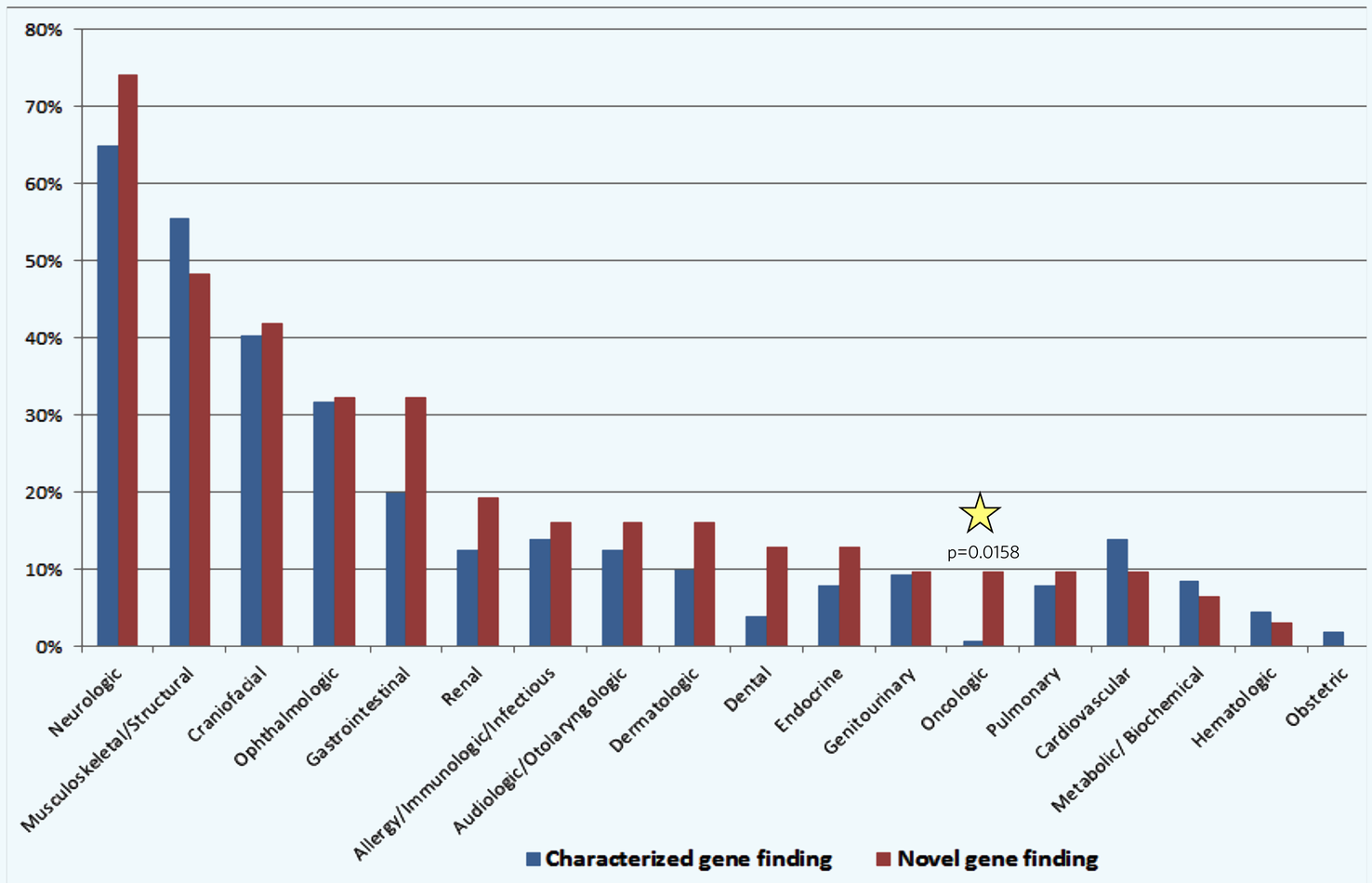


Table 1: Novel Gene Findings

Gene	Genotype	Alteration Type
ADNP	Heterozygous, *de novo*	Nonsense
AIMP2	Compound heterozygous	Frameshift, Splice
ARHGAP11A	Heterozygous	Frameshift
CACNA1E	Compound heterozygous	Missense, Missense
CDC42	Heterozygous, *de novo*	Missense
DGKZ	Homozygous	Missense
DPYSL2	Heterozygous, *de novo*	Missense
EMILIN1	Heterozygous	Missense
ETV6	Heterozygous	Splice
HDAC1	Heterozygous, *de novo*	Missense
HNRNPK	Heterozygous, *de novo*	Nonsense
IL21R	Compound heterozygous	Missense, Missense
ITGA11	Compound heterozygous	Missense, Missense
ITSN1	Heterozygous, *de novo*	Missense
LAMA1	Heterozygous, *de novo*	Frameshift
LAS1L	Hemizygous, *de novo*	Missense
LRFN5	Compound heterozygous	Missense, Missense
MAPK1	Heterozygous, *de novo*	Missense
MTOR	Hemizygous	Missense
MYH10	Heterozygous, *de novo*	Nonsense
NID1	Heterozygous	Missense
OGDHL	Compound heterozygous	Missense, Missense
PURA [1]	Heterozygous, *de novo*	Frameshift
RAD54L [1]	Heterozygous, *de novo*	Missense
RAD54L	Heterozygous	Missense
RPUSD3	Compound heterozygous	Missense, Missense
SETD5	Heterozygous, *de novo*	Frameshift
SLIT2	Heterozygous, *de novo*	Missense
SNAP25	Heterozygous, *de novo*	Missense
SV2A	Homozygous	Missense
ZNF238	Heterozygous, *de novo*	Missense
ZNF302	Heterozygous	Frameshift

[1]Gene identified in two unrelated families

Figure 2: Comparison of Organ System Involvement Among Patients with Characterized and Novel Gene Findings



RESULTS

- Among the first 500 reported DES cases, analysis of novel genes was elected for 416 probands. Among these 416 patients, 31 (7.5%) were positive for a novel gene finding (**Figure 1**), accounting for 13.7% of all positive results.
- The novel gene alterations followed all inheritance patterns, roughly half were *de novo*, and the majority were missense alterations (**Table 1**).
- Among the 31 patients with novel gene findings, the most common organ systems involved in the patients' phenotype included neurologic (74.2%), musculoskeletal/structural (48.4%), craniofacial (41.9%), ophthalmologic (32.2%), gastrointestinal (32.3%), renal (19.4%), and immunologic/ infectious (16.1%).
- Compared to patients with characterized gene findings, clinical characteristics more common among patients with novel gene findings include neurologic (74 vs 65%), gastrointestinal (32 vs 20%), neurological, renal (19 vs 13%), and oncologic (10 vs 1%).
  - Patients with cancer phenotypes were significantly more likely to have a novel gene finding (p=0.0158).
  - All of the patients with cancer phenotypes and a novel gene positive finding had previously tested negative for multiple single gene and/or panel tests.

TAKE-HOME POINTS

- Novel genes account for ~14% of positive findings from diagnostic exome sequencing.
- Novel gene findings are significantly more likely in patients with cancer phenotypes.
- These data highlight the utility of DES in providing the most comprehensive molecular diagnosis in that no other clinical test available would have identified these results observed in the 31 patients. Moreover, these data have significant implications for genetic counseling and clinical management.

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.