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

Negative Diagnostic Exome Sequencing Results: A Retrospective Analysis of the Phenotypic Spectrum of Patients with Negative Exome Sequencing Results

BACKGROUND

- Diagnostic exome sequencing (DES) has been instrumental in providing a molecular diagnosis in patients with previously undiagnosed disease. Depending on criteria used in selecting patients for DES and associated disease categories, the likelihood of a positive finding can be quite different.
- The detection rate of DES in patient cohorts with a wide range of phenotypes is reportedly 25-30%.
- Data from the first 500 cases submitted for DES at a single laboratory suggests that 54% (274) of cases are negative, meaning no causative mutation was identified to explain the patient's phenotype.
- Among the reported cases, 30% (151) identified a positive finding among characterized genes, 9%(44) resulted in an uncertain finding, and 7% (31) identified a finding in a novel gene (manuscript in progress).
- A retrospective analysis of the clinical presentation of the patients with negative exome sequencing results reveals a spectrum of phenotypes, with a majority of patients presenting with neurological or musculoskeletal disease.

METHODS

- **Patients/study population:** 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). 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 (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 nformatics 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, patient age, diagnosis and/or clinical description, and exome sequencing results, including gene(s), alteration(s), gene category (novel, characterized), alteration interpretation (pathogenic, likely pathogenic, uncertain, likely benign), and clinical overlap of gene-association and patient phenotype (positive, uncertain, partial).
- Statistical analyses were computed by chi^2 goodness of fit tests and Fisher's Exact Probability.

Figure 1. Phenotypic Distribution

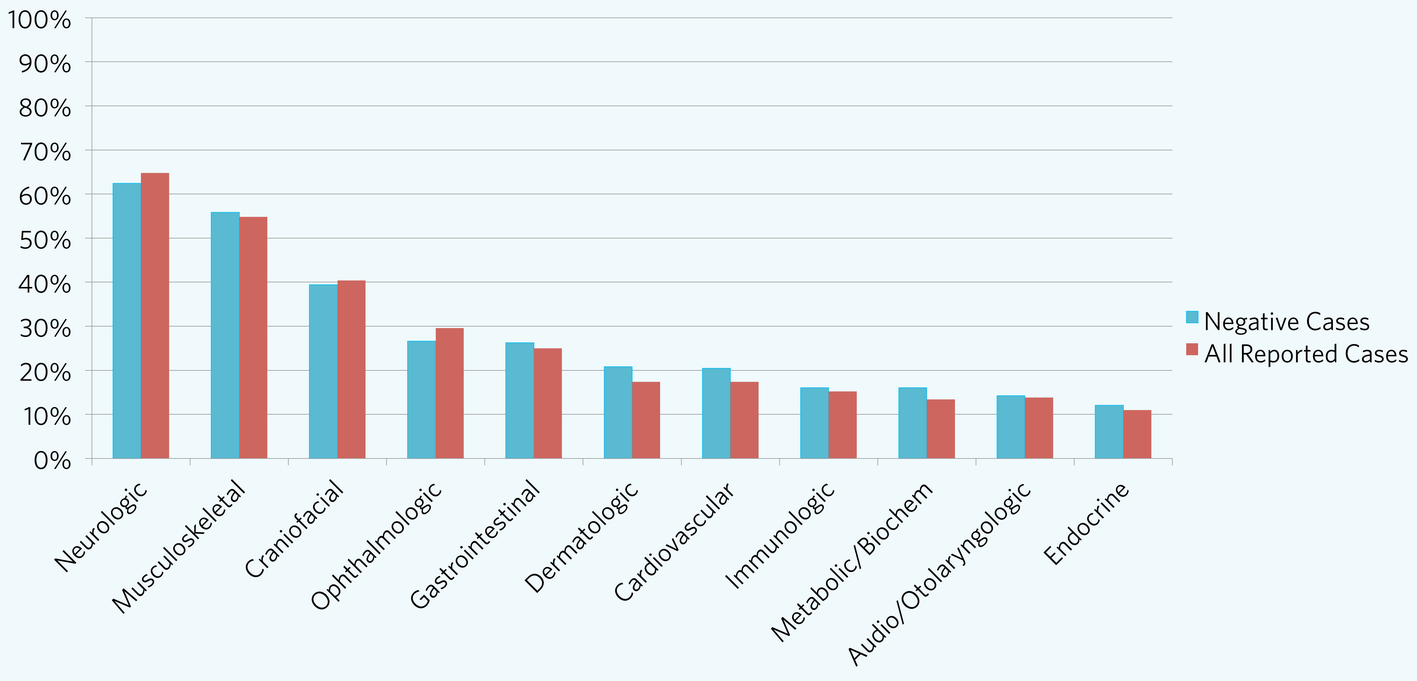
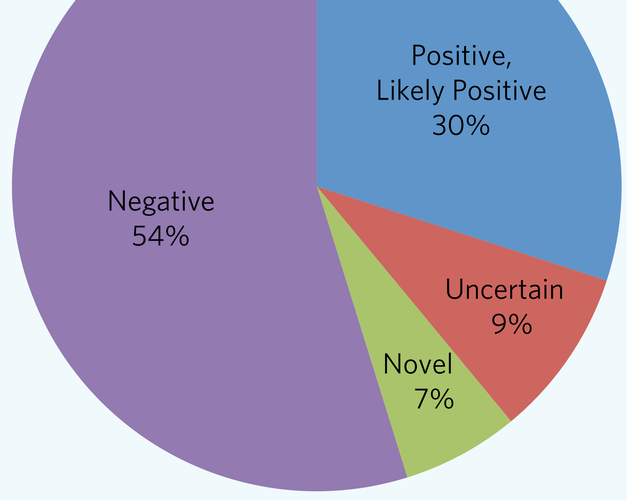


Figure 2. Detection Rates Among First 500 Cases



RESULTS

- A retrospective analysis of the negative exome cases revealed that the majority of patients presented with phenotypes involving features in multiple organs. The most common (>12% of cases) organ systems involved patients with the following phenotypes: Neurologic (62.4%), Musculoskeletal/Structural (55.8%), Craniofacial (39.4%), Ophthalmologic (26.6%), Gastrointestinal (26.3%), Dermatologic (20.8%), Cardiovascular (20.4%), Immunologic (16.1%), Metabolic/Biochemical (16.1%), Audiologic/Otolaryngologic (14.2%) and Endocrine (12%).
- Interestingly, when compared to the most common phenotypes observed among all reported exome cases (positive, uncertain, novel and negative), the phenotypic spectrum of patients with negative findings showed a similar distribution of phenotypes (FIGURE 1). (manuscript in progress).
  - Neurologic and musculoskeletal findings are among the majority of phenotypes observed.
  - Ophthalmologic indications for testing were more commonly seen among all reported cases than among negative cases alone, however, there are no statistically significant differences.
  - Dermatologic and cardiovascular involvement was more common in the negative cases than among all reported cases, however, there are no statistically significant differences.
- Of the negative patients with a neurologic phenotype 60.2% (165) presented with some type of MR/ID/DD, 32.5% had a positive brain MRI , and 19.7% had seizures/epilepsy.
- The majority of patients with negative findings on DES had a microarray prior to diagnostic exome sequencing.
- ~23% of negative reported cases involved multiple affected first-degree relatives with similar presentations.

Table 1. Distribution of Phenotypes Among All Reported Cases

Phenotype	% Negative Cases	% All Cases	Difference
Neurologic	62.4%	64.8%	2.4%
Musculoskeletal/Structural	55.8%	54.8%	1.0%
Craniofacial	39.4%	40.4%	1.0%
Ophthalmologic	26.6%	29.6%	3.0%
Gastrointestinal	26.3%	25.0%	1.3%
Dermatologic	20.8%	17.4%	3.4%
Cardiovascular	20.4%	17.4%	3.0%
Allergy/Immunologic/Infectious	16.1%	15.2%	0.9%
Metabolic/ Biochemical	16.1%	13.4%	2.7%
Audiologic/Otolaryngologic	14.2%	13.8%	0.4%
Endocrine	12.0%	11.0%	1.0%
Genitourinary	10.9%	10.6%	0.3%
Pulmonary	10.2%	9.8%	0.4%
Renal	10.2%	11.8%	1.6%
Hematologic	9.5%	8.2%	1.3%
Dental	4.4%	4.6%	0.2%
Oncologic	3.6%	3.0%	0.6%
Obstetric	1.1%	1.4%	0.3%

TAKE-HOME POINTS

- These results suggest that disease category and organ system involvement are not informative indicators for predicting the likelihood of a positive finding in DES.
- These findings also highlight the fact that although DES is a powerful tool for molecular diagnosis among patients with neurologic, musculoskeletal and a spectrum of other undiagnosed diseases, there is still much to learn about the genome and its application in undiagnosed disease.

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.