Ruth Baxter[1], Kelly Farwell Gonzalez[1], Megan Truitt[2], Kwame Anyane-Yeboa[2], Wenqi Zeng[1]
1) Ambry Genetics Corporation, Aliso Viejo, CA
2) Columbia University Medical Center

Diagnostic Exome Sequencing Identifies a *de novo* Variant in a Consanguineous Family

BACKGROUND

- Diagnostic exome sequencing (DES) is increasingly being used for diagnosis in intractable cases where the underlying cause is thought be genetic.
- Clinical information is an important part of the exome analysis and the family history informs the likely mode of inheritance.
- The majority of DES tests performed in our laboratory involve exome sequencing of the proband plus two first-degree relatives.
- In some cases sequencing of the proband alone is expected to yield easily interpretable results. For example, children of parents with known consanguinity are more likely to harbor true homozygous variants.
- The proband was a boy with multiple congenital anomalies whose parents were acknowledged first cousins. Exome sequencing was performed on the proband sample only.
- A SNP array had identified multiple regions of homozygosity consistent with approximately 8.5% consanguinity.

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 (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.
- Statistical analyses were computed by chi^2 goodness of fit tests and Fisher's Exact Probability.

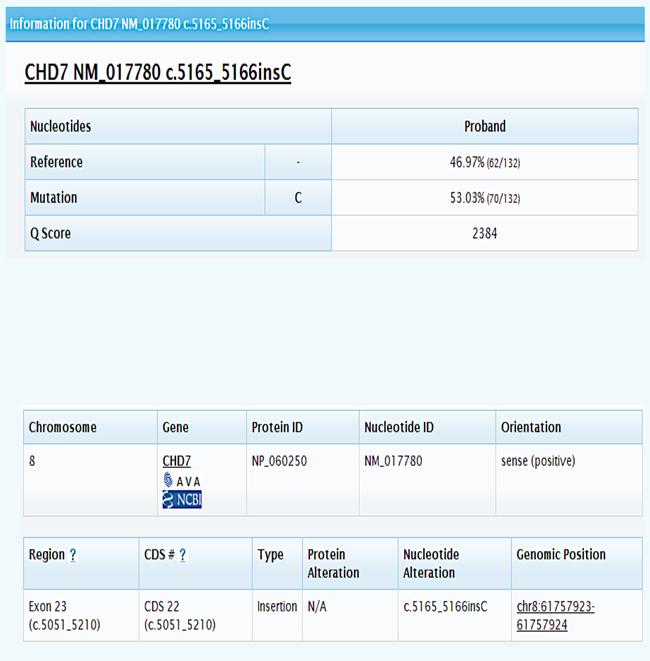
Table 1. Number of Genes and Alterations Identified

NUMBER OF GENES & ALTERATIONS IDENTIFIED BASED ON BIOINFORMATICS & INTERPRETATION									
	Post-Inheritance Model Filtering			Post-Medical Review			Notable Candidate Genes		
	HGMD/ OMIM-Morbid‡	Clinically novel**	TOTAL	HGMD/ OMIM-Morbid	Clinically novel	TOTAL	HGMD/ OMIM-Morbid	Clinicall y novel	TOTAL
Autosomal Dominant Genes (Alterations)	55 (57)	215 (244)	270 (281)	4 (4)	191 (200)	195 (204)	1 (1)	0 (0)	1 (1)
Autosomal Recessive Genes (Alterations)	13 (17)	58 (103)	71 (120)	1 (1)	51 (79)	52 (80)	0 (0)	0 (0)	0 (0)
X-linked Recessive Genes (Alterations)	1 (1)	7 (7)	8 (8)	0 (0)	6 (6)	6 (6)	0 (0)	0 (0)	0 (0)
X-linked Dominant Genes (Alterations)	0 (0)	2 (2)	2 (2)	0 (0)	2 (2)	2 (2)	0 (0)	0 (0)	0 (0)
Y-linked Genes (Alterations)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
TOTAL GENES (Alterations)	69 (75)	282 (356)	351 (411)	5 (5)	250 (287)	255 (292)	1 (1)	0 (0)	1 (1)

‡HGMD/ OMIM-Morbid: Listed in this category are genes that have been described in either the HGMD or OMIM-Morbid databases (the alterations within these genes are located in parentheses)

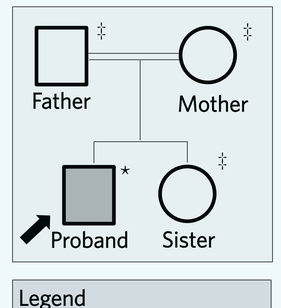
**A clinically novel gene is a gene which is not currently known to underlie a genetic condition

Figure 1. Pedigree and Sequencing Data for *CHD7* Variant



Father: wild type

Proband: p.F1722SfsX15



Mother: wild type

Sister: wild type

Legend
*Whole exome sequencing
‡ Co-segregation analysis

SEQUENCING AND ANALYSIS

- Full exome sequencing, bioinformatics analysis, and filtering based on autosomal and X-linked dominant and recessive inheritance revealed 351 genes (411 unique alterations).
- Manual review to rule out sequencing artifacts and polymorphisms along with medical interpretation to rule out genes lacking clinical overlap with the patient's evaluated phenotype resulted in 255 genes (292 unique alterations) (**Table 1**).
- Due to the degree of consanguinity of the family analysis initially focused on the autosomal recessive model, with particular focus on homozygous alterations.
- We identified 13 homozygous variants in known genes (reported in HGMD or OMIM morbid) and an additional 28 homozygous variants in novel genes not currently known to underlie Mendelian disorders.
- None of the known genes with homozygous variants fit well with the clinical phenotype of the patient.
- **Examination of genes** in additional models with one candidate variant identified (i.e. potentially autosomal dominant inheritance) revealed a frameshift change in the *CHD7* gene leading to a premature stop (c.5165_5166insC, p.F1722SfsX15).
- This alteration was confirmed in the proband by Sanger sequencing (Figure 1).

De novo Variant in *CHD7* Supports the Clinical Findings in This Patient

- Sanger sequencing in additional familial samples showed that the variant was not observed in either of the unaffected parents, or the unaffected sibling, consistent with a *de novo* occurrence (Figure 1).
- Mutations in the *CHD7* gene have been identified in patients with phenotypic similarities with the proband. The patient's characteristics which are consistent with CHARGE syndrome include cleft lip, hearing loss, tetralogy of fallot, short stature, micropenis, undescended testicle, hydronephrosis, developmental delay, webbed neck, and speech delay. Additionally, there are reported cases of CHARGE syndrome patients with seizures (Jongmans, 2008), failure to thrive (Blake, 1998; Jongmans, 2008), and myopia (McMain, 2008).
- The majority of CHARGE syndrome cases arise from a *de novo* alteration within the *CHD7* gene (Lalani, 2006). To date, mutations in the *CHD7* gene are 100% penetrant (i.e., all individuals who are heterozygous for a *CHD7* mutation have some features of CHARGE syndrome) and account for approximately 58-65% of individuals affected with CHARGE syndrome (Jongmans, 2006; Lalani, 2006; Sanlaville, 2007).
- The c.5165_5166insC mutation, located in exon 23 of the *CHD7* gene, results from a insertion of one nucleotide between position 5165 and 5166, causing a translational frameshift with a predicted premature stop codon. Since frameshifts are typically deleterious in nature, this mutation is interpreted as disease-causing (ACMG Recommendations for Standards for Interpretation and Reporting of Sequence Variations. Revision 2007. Richards, 2008).

TAKE-HOME POINTS

- *De novo* dominant changes occur and can be disease causing even when a proband has large regions of homozygosity.
- It is important to consider all possible inheritance patterns when analyzing consanguineous exome cases.

REFERENCES

- 1000 genomes database: A map of human genome variation from population-scale sequencing. *Nature* 467,:1061-1073.
- Blake K & Prasad C (2006) *Orph J Rare Diseases* **1**(34):1-8
- International HapMap (2003) The International HapMap Project. *Nature* 426:789-796.
- Jongmans MC, *et al.* (2008) *Am J Med Genet A* **146A**:43-50.
- Lalani SR, *et al.* (2006) *Amer J Hum Genet* **78**:303-314.
- McMain K, *et al.* (2008) *J AAPOS* **12(6)**:460-465.
- 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/
- Richards, CS, et al (2008) Genetics in Medicine 10:294-300
- Sanlaville D, *et al.* (2007) *Europ J of Hum Genet* **15**:389-399.
- Sherry ST, et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29,:308-311.
- Stenson PD, *et al.* (2009) The Human Gene Mutation Database: 2008 update. Genome medicine 1 :13.