Cameron Mroske[1], Jennifer Dickerson[2], Laurence Walsh[2], Yi Zheng[3], Sha Tang[1], Kelly D. Gonzalez[1]
1) Ambry Genetics Corporation, Aliso Viejo, CA
2) Child Neurology, Indiana University School of Medicine, Indianapolis, IN
3) Children's Hospital Research Foundation, Cincinnati, OH

Diagnostic Exome Sequencing (DES) in a Patient with Syndromic Intellectual Disability Identifies a Novel Gene, *CDC42*, which Lies at the 1p36 Deletion Syndrome Locus

BACKGROUND

- Diagnostic Exome Sequencing (DES) has emerged as a powerful new tool for the molecular diagnosis of genetic disease in families and patients in whom traditional genetic and biochemical testing has been uninformative and/or ambiguous.
- A 14 year-old male presenting with multiple neurologic symptoms and a negative family history was referred to Ambry Genetics for proband-parent trio DES in order to elucidate the underlying disease etiology.
- Proband symptoms included mental retardation (MR), developmental delay, cognitive impairment, scoliosis, microcephaly, seizures, absent speech, mild dysmorphism.
- Prior to DES, the patient had evaded diagnosis by means of clinical evaluation and extensive genetic testing. This included negative chromosome analysis; subtelomeric FISH; CGH/SNP microarray; Angelman/Prader Willi syndrome uniparental disomy analysis; *MECP2, CDKL5, FOXG1, ARX,* 90 gene X-linked intellectual disability gene panel; 53 gene seizure panel; 101 nuclear mitochondrial gene panel, as well as uninformative biochemical results.

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**
		Post-alteration review	Post-clinical association review			
		TOTAL	Characterized‡	Clinically novel‡	TOTAL	
Autosomal Dominant Genes (Alterations)	16(16)	2(2)	0(0)	2(2)	2(2)	1(1)
Autosomal Recessive Genes (Alterations)	13(27)	3(6)	0(0)	3(6)	3(6)	0(0)
X-linked Recessive Genes (Alterations)	4(4)	3(3)	0(0)	3(3)	3(3)	0(0)
X-linked Dominant Genes (Alterations)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
Y-linked Genes (Alterations)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
TOTAL GENES (Alterations)	33(47)	8(11)	0(0)	8(11)	8(11)	1(1)

*Post-medical review filtering involves the manual removal of genes unrelated to the patient's evaluated phenotype and alterations considered benign. ‡Characterized Genes: Genes known to be associated with a clinical phenotype based on the HGMD or OMIM-Morbid databases or the medical literature; a clinically novel gene is a gene that is not currently known to underlie a genetic condition (the alterations within these genes are located in parentheses). **Notable Candidate Genes: Gene alterations selected for co-segregation.

FIGURE 1. Exome-Trio Sequencing Identifies a *de novo* Missense Change, c.68A>G (p.Y23C), in *CDC42*

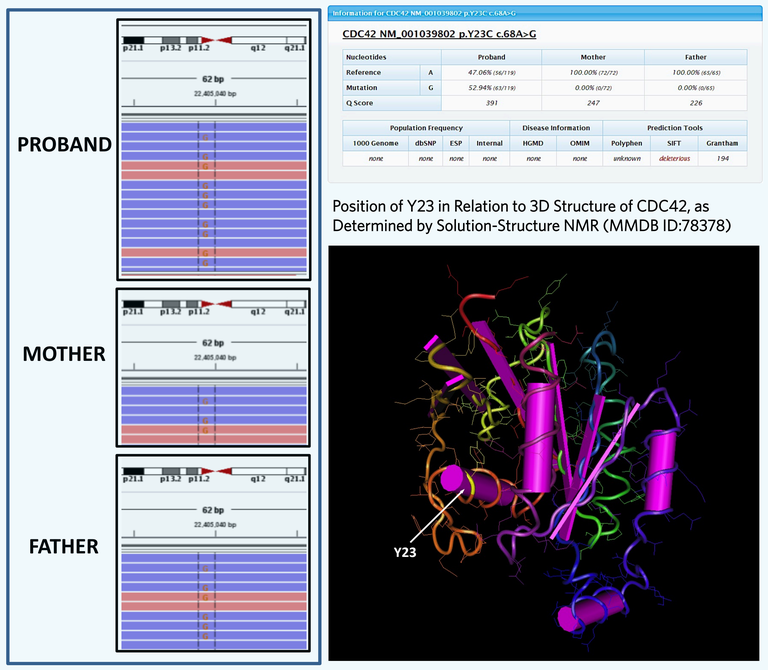
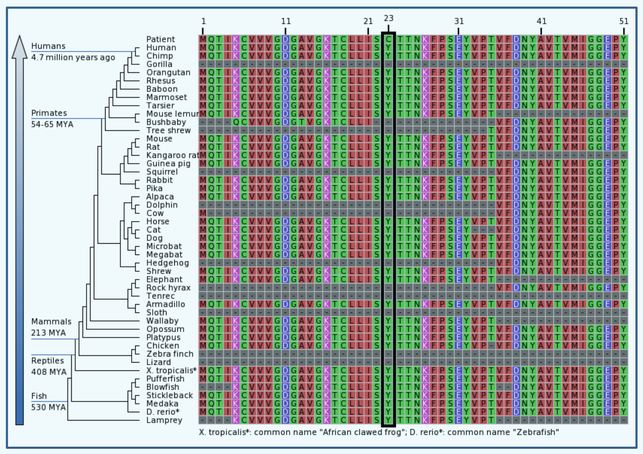


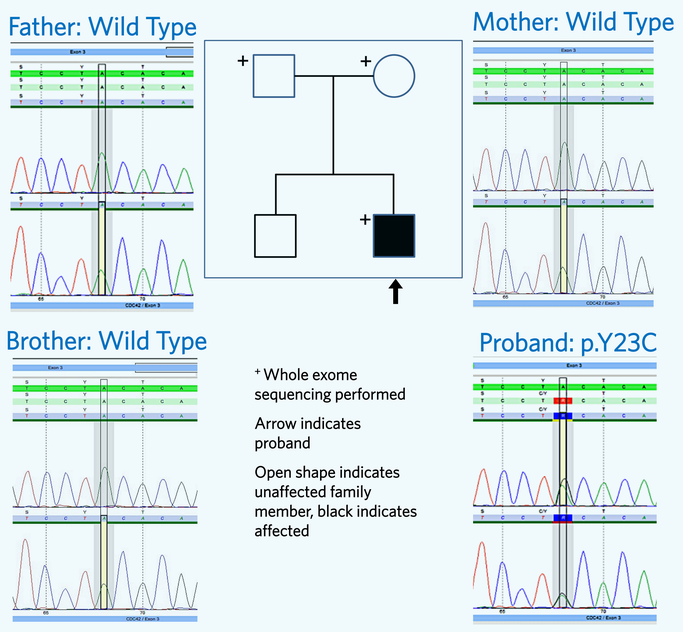
FIGURE 2. Amino Acid Y23 is Highly Conserved Across Evolution and Within Certain Human Rho SubFamilies



RESULTS/DISCUSSION

- Full exome sequencing, bioinformatics analysis, and filtering based on autosomal and X-linked dominant and recessive inheritance models of the proband, mother, and father revealed **47** unique alterations in **32** genes. Manual review to remove sequencing artifacts and polymorphisms, and medical interpretation to remove alterations in genes lacking clinical overlap with the patient's evaluated phenotype, resulted in **11** unique alterations in **8** genes (**TABLE 1**).
- The autosomal dominant (AD) filter prioritized a novel *de novo* change, c.68 A>G (p.Y23C), in the ***CDC42*** gene of the proband (**FIGURE 1**). Because this alteration was deemed clinically relevant, ***CDC42*** was classed as a "Notable Candidate Gene" and co-segregation analysis was performed. Sanger sequencing confirmed the presence of this **heterozygous** alteration solely in the proband, thereby elucidating the *de novo* nature of ***CDC42*** c.68 A>G (p.Y23C) (**FIGURE 3**).
- ***CDC42*** lies within the common locus of 1p36 deletion syndrome. Virtually all patients suffering from this disease exhibit language and developmental delay, dysmorphic features, and MR. Other features include microcephaly, EEG and structural brain abnormalities, as well as skeletal defects like scoliosis (OMIM #607872).
- Widely expressed in brain, **CDC42** is one of the best characterized members of the Rho GTPase family of intracellular switches (Etienne-Manneville & Hall, 2002). Through its regulation of both the actin and microtubule cytoskeleton, CDC42 is known to play a crucial role in neuronal morphogenesis and polarization (Gonzalez-Billault, 2012).
- Animal studies support the importance of **CDC42** in neurodevelopment. In the developing mouse embryo, complete *CDC42* ablation results in ***(1)*** abnormal neuronal precursor proliferation, ***(2)*** a decrease in the number of neuronal precursor cells, and ***(3)*** embryonically lethality (Chen, 2000). Conditional deletion of ***CDC42*** from developing mouse forebrains causes the animals to suffer holoprosencephaly (Chen, 2006). **CDC42** expression is increased in rats induced to undergo mesial temporal lobe epilepsy with kainic acid (Sharma, 2009).
- Amino acid Y23 is highly conserved throughout evolution (**FIGURE 2A**) and the alteration identified herein, p.Y23C, is predicted to be damaging and deleterious by the *in silico* modeling algorithms PolyPhen and SIFT (Adzhubei, 2010; Ng, 2006). Further, in humans Y23 is conserved within RHO subfamilies 4, 5 and 6, which together constitute the second of four structural RHO clusters (**FIGURE 2B**) (Boureux, 2006). Solution-structure studies performed at a resolution of 3.5 Å have revealed that residue Y23 forms part of ***A)*** the Pa6 cell polarity protein interaction site and ***B)*** the ACK tyrosine kinase interaction site (Mott, 1999; Watts, 1996).
- CDC42, like the other RHO family members, is a relatively small protein (191 AA, approx. 21 kDa) characterized by an N-terminal RhoGTPase Domain, a Hypervariable Region, and a conserved C-terminal CAAX region (Gonzalez-Billault, 2012). It is plausible that Y23C, which is situated in the GTPase Domain (**FIG 1**), is more likely to exert a disproportionally deleterious effect on such a small protein compared to a much larger molecule.
- Structure simulations of wild type CDC42 reveal that 2Y23 stacks with F169; as a consequence, the Y23C mutation may increase GTP loading and thereby cause a gain of function.

FIGURE 3. Sanger Sequencing Confirms the *de Novo* Nature of the Mutation



CONCLUSIONS

- Clinical DES of a parents-proband trio has revealed a "*de novo*" mutation in the *CDC42* gene of the affected child.
- Multiple lines of evidence including CNV, sequence and structural conservation, as well as *in vivo* studies, strongly suggest that the mutation identified herein, c.68 A>G (p.Y23C), is disease causing.
- These results demonstrate the clinical utility of DES for detecting rare variants in patients for whom more traditional genetic and biochemical testing methods have failed to elucidate an underlying disease etiology.
- Our findings establish *CDC42* to be a novel gene associated with neurological symptoms and seizures in human patients. This finding will allow for the targeted sequencing of *CDC42* in patients with similar clinical presentations.
- Ongoing functional studies will further define the impact of *CDC42* alterations on neurodevelopment and disease.

REFERENCES

- 1000 genomes database: A map of human genome variation from population-scale sequencing. *Nature* **467**,:1061-1073.
- Boureux A, *et al.* (2006) Mol Biol Evol. **24**:203-216.
- Chen F, *et al.* (2000) *Curr Biol.* **10**:758-765.
- Chen L, *et al.* (2006) *Proc Natl Acad Sci USA.* **103**:16520-16525.
- Etienne-Manneville S & Hall A. (2002) *Nature.* **420**:629-634.
- Gonzalez-Billault C, *et al.* (2012) *Cytoskel.* **69**:464-485.
- Madej T, *et al.* *"MMDB: 3D structures and macromolecular interactions."* *Nucleic Acids Res. 2012 Jan; 40(Database issue)*:D461-4
- International HapMap (2003) The International HapMap Project. *Nature* **426**:789-796.
- Mott HR, *et al.* (1999) *Nature.* **399**:384-387.
- 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/
- PolyPhen [Internet]: Adzhubei IA, *et al.* (2010) *Nat Methods* **7**(4):248-249.
- Sharma AK, *et al.* (2009) *Toxicol Pathol.* **37**:776-789
- Sherry ST, *et* al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**,:308-311.
- SIFT [Internet]: Ng PC & Henikoff S (2006) *Hum Genet.* **7**:61-80
- Stenson PD, *et al.* (2009) The Human Gene Mutation Database: 2008 update. Genome medicine **1**:13.
- Watts JL, *et al.* (1996) *Develop.* **122**:3133-3140.