A CCR4-NOT Transcription Complex, Subunit 1, *CNOT1*, Variant Associated with Holoprosencephaly

Paul Kruszka,[1] Seth I. Berger,[1,5] Karin Weiss,[1,6] Joshua L. Everson,[2,3] Ariel F. Martinez,[1] Sungkook Hong,[1] Kwame Anyane-Yeboa,[4] Robert J. Lipinski,[2,3] and Maximilian Muenke[1,*]

Holoprosencephaly is the incomplete separation of the forebrain during embryogenesis. Both genetic and environmental etiologies have been determined for holoprosencephaly; however, a genetic etiology is not found in most cases. In this report, we present two unrelated individuals with semilobar holoprosencephaly who have the identical *de novo* missense variant in the gene CCR4-NOT transcription complex, subunit 1 (*CNOT1*). The variant (c.1603C>T [p.Arg535Cys]) is predicted to be deleterious and is not present in public databases. *CNOT1* has not been previously associated with holoprosencephaly or other brain malformations. *In situ* hybridization analyses of mouse embryos show that *Cnot1* is expressed in the prosencephalic neural folds at gestational day 8.25 during the critical period for subsequent forebrain division. Combining human and mouse data, we show that *CNOT1* is associated with incomplete forebrain division.

Holoprosencephaly (HPE) is defined by varying degrees of separation of the embryonic forebrain. While occurring in approximately 1 in 10,000 live births, HPE is estimated to occur in 1 in 250 embryos, making it one of the most common human developmental abnormalities.[1] The etiology of HPE is complex and most likely involves the interaction of genetic and environmental factors. The most common cause is trisomy 13, but in cases not associated with aneuploidy, only a fraction of affected subjects have a known genetic etiology.[2,3] In this report, we describe the identical *de novo* missense variant in two unrelated families in the gene CCR4-NOT transcription complex, subunit 1 (*CNOT1* [MIM: 604917]) and show that this gene is expressed during early neurulation in the mouse embryo.

CNOT1 is one of at least nine components of the CCR4-NOT complex, which has an important role in posttranscriptional regulation and is conserved from yeast to mammals.[4] The CCR4-NOT complex is the main enzyme responsible for mRNA deadenylation, which shortens the poly(A) tail in mRNA, thus leading to mRNA degradation.[5] *Cnot1* is expressed in the embryonic brain of the mouse and has drastically decreased expression after gestational day 13.[4] Whether *Cnot1* is expressed during the critical period for induction of HPE, between gestational day 7.0 and 8.5, has not been studied previously.[6,7]

To expand the genetic etiology of HPE and uncover novel regulators of forebrain development, we have applied whole-exome sequencing (WES) to 134 trios (proband and parents) with holoprosencephaly in an ongoing HPE research protocol (Table S1). The individuals and families with HPE in this study are recruited from multiple clinical genetics centers from the United States. Within the participating institutions, the phenotype was evaluated by clinical exam and brain imaging (MRI or CT) or autopsy. The study was approved by National Human Genome Research Institute (NHGRI) Institutional Review Board (protocol 98-HG-0249); procedures followed were in accordance with the ethical standards of NHGRI for human experimentation, and proper consent was obtained.

DNA samples from study participants underwent WES at the National Intramural Sequencing Center (NISC) (Supplemental Material and Methods). The mean read depth for each sample was 79.8. Variant calling, annotation, and filtering is described in the Supplemental Material and Methods. Copy-number variation (CNV) prediction from exome data was done using the eXome-Hidden Markov Model (XHMM) caller (Supplemental Material and Methods).[8]

All probands were first searched for four common genes known to cause HPE—*SHH* (MIM: 600725) on 7q36, *ZIC2* (MIM: 603073) on 13q32, *SIX3* (MIM: 603714) on 2p21, and *TGIF1* (MIM: 602630)—on 18p11.3 using Sanger sequencing as recommended.[3] 20% (27 probands) of the discovery cohort had damaging variants in these genes. With the goal of gene discovery, minimizing false positives, and sacrificing sensitivity, the discovery cohort was filtered with stringent criteria including *de novo* inheritance in genes intolerant of variation,[9] variant absence in the Exome Aggregation Consortium (ExAC) database,[9] and Combined Annotation-Dependent Depletion (CADD) scores above 20.[10] Variants that met these criteria were considered deleterious. An identical *de novo* missense variant (c.1603C>T [p.Arg535Cys]) in *CNOT1* (GenBank: NM_001265612.1) was found in two unrelated families by WES and verified by Sanger sequencing (Supplemental Material and Methods). In proband 1, the WES alternate allele frequency

[1]Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; [2]Department of Comparative Biosciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA; [3]Molecular and Environmental Toxicology Center, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI 53706, USA; [4]Division of Clinical Genetics, Department of Pediatrics, Columbia University Medical Center, New York, NY 10032, USA; [5]Rare Disease Institute, Genetics and Metabolism, Children's National Health System, Washington, DC 20036, USA
[6]Present address: Genetics Institute, Rambam Health Care Campus, Haifa 3109601, Israel
*Correspondence: mamuenke@mail.nih.gov
https://doi.org/10.1016/j.ajhg.2019.03.017.

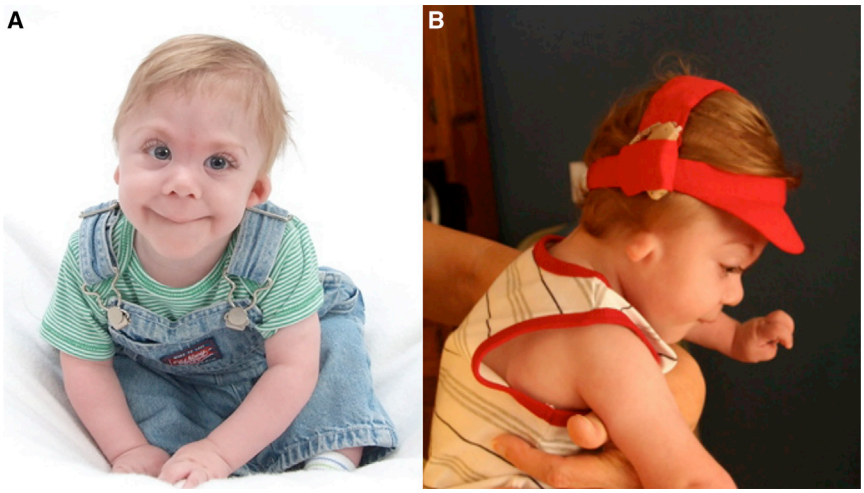

Figure 1. Patient Images
(A) Proband 1 at age 12 months; facial characteristics include hypotelorism, epicanthal folds, depressed nasal bridge, and long philtrum.
(B) Proband 1 at 15 months, note right ear microtia.
Photo not available for proband 2.

for the c.1603C>T (p.Arg535Cys) variant was 39% (read depth 57), and for proband 2 the alternate allele frequency was 54% (read depth 54). Proband 1 (Figure 1) is a male born at 33 weeks' gestation after a pregnancy complicated by intrauterine growth restriction (IUGR). Semilobar holoprosencephaly was confirmed by brain MRI. Medical problems (Table 1) included bilateral microtia, hearing loss, diabetes insipidus, neonatal diabetes mellitus requiring insulin, pancreatic exocrine insufficiency requiring enzyme therapy, and global developmental delay. Facial characteristics (Figure 1) include epicanthal folds, depressed nasal bridge, hypotelorism, and long philtrum. Proband 1 died at age 16 months. Proband 2 is female and was a term uncomplicated pregnancy without IUGR. Semilobar holoprosencephaly was confirmed by brain MRI postnatally and other medical problems included severe bilateral sensorineural hearing loss, global developmental delay, and hypertonia. Proband 2 was last seen in clinic at age 6.5 years, and her exam (Table 1) was significant for microcephaly, epicanthal folds, and long philtrum (photo unavailable). With the exception of insulin-requiring diabetes in proband 1, both probands have similar phenotypes including semilobar HPE, similar facial features, hearing loss, and global developmental delay.

The c.1603C>T (p.Arg535Cys) variant is located in the conserved HEAT domain in the N-terminal (Figure 2). The N-terminal of CNOT1 associates with another complex protein, CNOT11.[11] The CADD score for the *CNOT1* c.1603C>T (p.Arg535Cys) variant is 35, and it is not present in the ExAC database (accessed January 18, 2019). XHMM analysis of exomes showed no copy number variations in either proband. Two identical variants in unrelated probands with holoprosencephaly is unlikely by chance in a relatively small cohort of 134 trios with HPE; especially, given that *CNOT1* is intolerant of both missense change ($z = 7.44$) and loss of function ($pLi = 1.00$) (constraint

Table 1. Summary of Clinical Characteristics

	Proband 1	Proband 2
Age at last exam	16 months	6.5 years
Gender	male	female
Prenatal history	IUGR	increased risk for Down syndrome on prenatal quad screen; amniocentesis not done
Birth history	Cesarean section at 35 weeks for IUGR	term vaginal delivery
Brain MRI	semilobar holoprosencephaly	semilobar holoprosencephaly
Craniofacial exam	microcephaly, epicanthal folds, long philtrum	microcephaly, epicanthal folds, long philtrum
Ears/hearing	bilateral microtia, bilateral conductive and sensorineural hearing loss with right ear worse than left, CT scan showed ossicle anomalies	severe bilateral sensorineural hearing loss
Seizure history	isolated seizure associated with fentanyl administration, normal EEG	none
Diabetes insipidus	present, treated with desmopressin	none
Neurologic history	global developmental delay, low muscle tone, non-ambulatory	global developmental delay, muscle spasticity, non-ambulatory
Other anomalies	pancreatic insufficiency: neonatal diabetes mellitus requiring insulin therapy and pancreatic exocrine deficiency treated with enzyme therapy	none

Abbreviations: IUGR, intrauterine growth restriction; MRI, magnetic resonance imaging; CT, computed tomography; EEG, electroencephalogram.

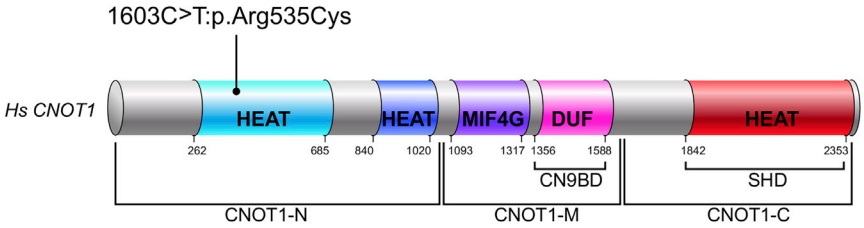

Figure 2. *CNOT1* Protein Domains
CNOT1 c.1603C>T (p.Arg535Cys) (GenBank: NM_001265612.1) variant is located in the conserved HEAT domain

metrics accessed from ExAC database on January 18, 2019).[9] There were two other *de novo* variants in *CNOT1* in two other unrelated individuals in our HPE cohort, but neither met the above criteria for a deleterious variant. One variant was a synonymous change (c.6057C>T [p.(=)]) in *CNOT1* (GenBank: NM_001265612.1) and the other variant was a missense change (c.1394A>C [p.Gln465Pro]) in *CNOT1* (GenBank: NM_001265612.1) with an ExAC database allele frequency of 8.25E−06 and a CADD score of 17.9. Additionally, Table S2 lists all *de novo* variants found in proband 1 and proband 2. The *CNOT1* c.1603C>T (p.Arg535Cys) was the only *de novo* variant found for proband 2. Proband 1 had three additional *de novo* variants; two were synonymous (Table S2) and a missense variant occurred in the gene *RNF150* (MIM: not listed) (GenBank: NM_020724.2; c.510G>A [p.Met170Ile]). The variant in *RNF150* is a variant of unknown significance. *RNF150* is not intolerant of variation based on ExAC constraint metrics: z = 2.15 for missense and pLI = 0.01 (accessed from ExAC database on January 18, 2019).[9]

Genes that regulate forebrain patterning and play a role in HPE pathogenesis are expected to be expressed in the prosencephalic neural folds that give rise to the forebrain during primary neurulation.[12] We therefore conducted *in situ* hybridization on mouse embryos at GD8.25, a stage representing early neurulation and within the critical period for HPE genesis.[7] Mouse *in situ* hybridization studies were conducted in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the University of Wisconsin-Madison School of Veterinary Medicine Institutional Animal Care and Use Committee (protocol number G005396). CD-1 mice (*Mus musculus*) were purchased from Charles River and C57BL/6J mice from The Jackson Laboratory. Timed pregnancies were established as previously described.[13] Embryos were dissected at GD8.25 and fixed overnight in 4% PFA. *In situ* hybridization (ISH) was conducted on whole C57BL/6J embryos or 50 μm sections cut from CD-1 embryos with a vibrating microtome in the transverse plane along the anterior-posterior axis. ISH was conducted as previously described and analysis was limited to the prosencephalic regions of the neural fold from which the forebrain will develop.[14] As seen in Figure 3, *Cnot1* expression is detectable in both the neuroectoderm and the mesenchyme of the neural folds but not in extra-embryonic membranes (Figure 3). Specificity of staining is additionally shown by staining for *Foxa2* (*Hnf-3β*), which is expressed in the ventral neuroectoderm.[15]

Disruption of the sonic hedgehog signaling pathway is known to result in holoprosencephaly.[16] Using human osteosarcoma cells, Cheng et al. showed that knockdown of CNOT1 using short hairpin (sh) RNA inhibited the sonic hedgehog signaling pathway based on decreased expression of genes downstream of *SHH* including *GLI1* and *PTCH1*.[17] These experiments in osteosarcoma cells have established a link between CNOT1 knockdown and

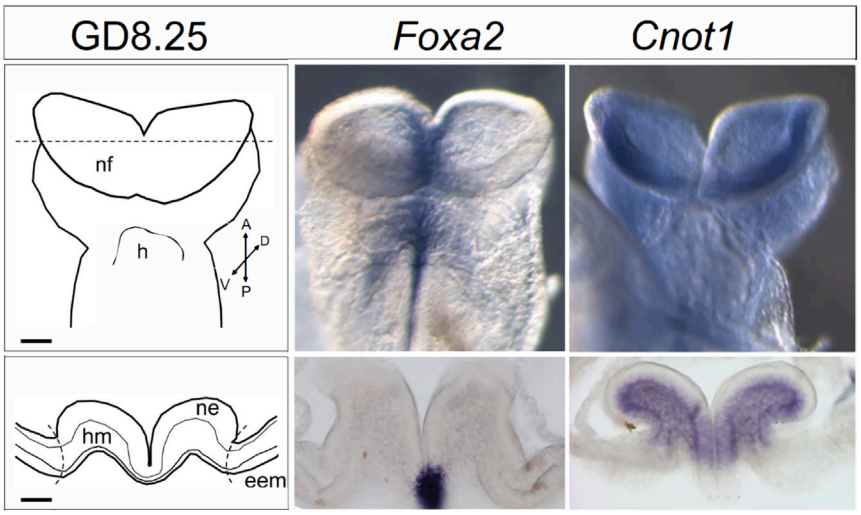

Figure 3. Gestational Day (GD) 8.25 Mouse Embryos
A ventral view (top) is shown for whole mounts. Transverse sections (bottom) through the prosencephalic neural folds (at the level of the dashed line in schematic) were stained to visualize gene expression in specific cellular compartments. Abbreviations: nf, neural folds; h, heart; ne, neuroectoderm; hm, head mesenchyme; eem, extra-embryonic membranes. Scale bar = 100 μm.

decreased sonic hedgehog signaling and raise the possibility that the *CNOT1* p.Arg535Cys variant may inhibit Sonic hedgehog signaling by a loss-of-function or a dominant-negative mechanism. This report allows for further research into the molecular mechanisms involved in CNOT1 and Hedgehog signaling.

In summary, we report identical *de novo* missense variants in *CNOT1* in two unrelated individuals with semilobar holoprosencephaly and show in the mouse model that *Cnot1* is expressed during the critical period for holoprosencephaly.

Accession Numbers

The accession number for the *CNOT1* c.1603C>T (p.Arg535Cys) variant is ClinVar: SUB5130764.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.03.017.

Acknowledgments

This work was supported by the National Human Genome Research Institute Intramural Research program. Work in the R.J.L. lab was supported by the National Institute of Environmental Health Sciences of the National Institutes of Health under award numbers R01ES026819 and T32ES007015.

Declaration of Interests

The authors declare no competing interests.

Received: December 7, 2018
Accepted: March 18, 2019
Published: April 18, 2019

Web Resources

ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/
ExAC Browser, http://exac.broadinstitute.org/
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
OMIM, http://www.omim.org/

References

1. Matsunaga, E., and Shiota, K. (1977). Holoprosencephaly in human embryos: epidemiologic studies of 150 cases. Teratology *16*, 261–272.
2. Kruszka, P., and Muenke, M. (2018). Syndromes associated with holoprosencephaly. Am. J. Med. Genet. C. Semin. Med. Genet. *178*, 229–237.
3. Kruszka, P., Martinez, A.F., and Muenke, M. (2018). Molecular testing in holoprosencephaly. Am. J. Med. Genet. C. Semin. Med. Genet. *178*, 187–193.
4. Chen, C., Ito, K., Takahashi, A., Wang, G., Suzuki, T., Nakazawa, T., Yamamoto, T., and Yokoyama, K. (2011). Distinct expression patterns of the subunits of the CCR4-NOT deadenylase complex during neural development. Biochem. Biophys. Res. Commun. *411*, 360–364.
5. Bartlam, M., and Yamamoto, T. (2010). The structural basis for deadenylation by the CCR4-NOT complex. Protein Cell *1*, 443–452.
6. Heyne, G.W., Everson, J.L., Ansen-Wilson, L.J., Melberg, C.G., Fink, D.M., Parins, K.F., Doroodchi, P., Ulschmid, C.M., and Lipinski, R.J. (2016). Gli2 gene-environment interactions contribute to the etiological complexity of holoprosencephaly: evidence from a mouse model. Dis. Model. Mech. *9*, 1307–1315.
7. Heyne, G.W., Melberg, C.G., Doroodchi, P., Parins, K.F., Kietzman, H.W., Everson, J.L., Ansen-Wilson, L.J., and Lipinski, R.J. (2015). Definition of critical periods for Hedgehog pathway antagonist-induced holoprosencephaly, cleft lip, and cleft palate. PLoS ONE *10*, e0120517.
8. Fromer, M., Moran, J.L., Chambert, K., Banks, E., Bergen, S.E., Ruderfer, D.M., Handsaker, R.E., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., et al. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am. J. Hum. Genet. *91*, 597–607.
9. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
10. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.
11. Bawankar, P., Loh, B., Wohlbold, L., Schmidt, S., and Izaurralde, E. (2013). NOT10 and C2orf29/NOT11 form a conserved module of the CCR4-NOT complex that docks onto the NOT1 N-terminal domain. RNA Biol. *10*, 228–244.
12. Geng, X., and Oliver, G. (2009). Pathogenesis of holoprosencephaly. J. Clin. Invest. *119*, 1403–1413.
13. Heyne, G.W., Plisch, E.H., Melberg, C.G., Sandgren, E.P., Peter, J.A., and Lipinski, R.J. (2015). A simple and reliable method for early pregnancy detection in inbred mice. J. Am. Assoc. Lab. Anim. Sci. *54*, 368–371.
14. Everson, J.L., Fink, D.M., Yoon, J.W., Leslie, E.J., Kietzman, H.W., Ansen-Wilson, L.J., Chung, H.M., Walterhouse, D.O., Marazita, M.L., and Lipinski, R.J. (2017). Sonic hedgehog regulation of *Foxf2* promotes cranial neural crest mesenchyme proliferation and is disrupted in cleft lip morphogenesis. Development *144*, 2082–2091.
15. Sasaki, H., and Hogan, B.L. (1993). Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. Development *118*, 47–59.
16. Roessler, E., and Muenke, M. (2010). The molecular genetics of holoprosencephaly. Am. J. Med. Genet. C. Semin. Med. Genet. *154C*, 52–61.
17. Cheng, D.D., Li, J., Li, S.J., Yang, Q.C., and Fan, C.Y. (2017). CNOT1 cooperates with LMNA to aggravate osteosarcoma tumorigenesis through the Hedgehog signaling pathway. Mol. Oncol. *11*, 388–404.

The American Journal of Human Genetics, Volume *104*

Supplemental Data

A CCR4-NOT Transcription Complex, Subunit 1, *CNOT1*, Variant Associated with Holoprosencephaly

Paul Kruszka, Seth I. Berger, Karin Weiss, Joshua L. Everson, Ariel F. Martinez, Sungkook Hong, Kwame Anyane-Yeboa, Robert J. Lipinski, and Maximilian Muenke

SUPPLEMENTAL MATERIAL AND METHODS

Table S1. Holoprosencephaly cohort.

	Trios (n=134)
Average age	2.8 years
Gender	42% male
HPE type	
Alobar	18%
Semilobar	49%
Lobar	23%
Middle interhemispheric variant (MIHV)	5%
Microform	4%
Ethnicity	
Caucasian	61%
African American	1%
Latin American	25%
Asian	6%
Native American	0%
Middle Eastern	3%

Table S2. *De novo* variants in proband 1 and 2.

	Chromosome	Position (hg19)	Variant type	Gene	Variant	ExAC frequency	CADD score
Proband 2	Chr16	58610468	nonsynonymous	*CNOT1*	c.1603C>T:p.Arg535Cys (GenBank: NM_001265612.1)	0	35
Proband 1	Chr16	58610468	nonsynonymous	*CNOT1*	c.1603C>T:p.Arg535Cys	0	35
Proband 1	Chr17	78210857	synonymous	*SLC26A11*	c.867A>G:p.(=) (GenBank: NM_000199.4)	0	
Proband 1	Chr19	48722162	synonymous	*CARD8*	c.1119T>C:p.(=) (GenBank: NM_001184902.1)	0	
Proband 1	Chr4	141889002	nonsynonymous	*RNF150*	c.510G>A:p.Met170Ile (GenBank: NM_020724.2)	0	29

SUPPLEMENTAL METHODS

Exome Sequencing

Exome sequencing, assembly, genotyping, and annotation were carried out by the National

Intramural Sequencing Center (NISC). Genomic DNA (approximately 1 µg) was fragmented to an average size of 150 bp and subjected to DNA library creation using established Illumina paired-end protocols. Capture utilized the NimbleGen SeqCap EZ Version 3.0+ UTR (Roche NimbleGen, Madison, WI). Captured regions totaled approximately 96 Mb. Flow cell preparation and 125-bp paired end read sequencing were performed as per the HiSeq2000 Sequencer protocol (Illumina, San Diego, CA).

Read mapping, variant calling and annotation. Fastq files were then aligned to reference genome human_g1k_v37_decoy using bwa mem and sam output was compressed to bam format using picard SamFormatConverter. The aligned bamfile was sorted and indexed using samtools and ReadGroups based on Sample ID were added with Picard AddOrReplaceReadGroups command. Picard's MarkDuplicates command was then applied. At this point the file was processed through a GATK 3.6 pipeline based on the recommended best practices using the genome capture intervals utilized by the EXaC consortium for exome targets with interval padding of 100 basepairs. Targets were realigned using the RealignerTargetCreator tools with the GATK resource bundle's 1000G_phase1.indels.b37.vcf and Mills_and_1000G_gold_standard.indels.b37.vcf. IndelRealigner was then applied with the intervals identified. The BaseRecalibrator was then applied using the GATK resource bundle's known sites from dbsnp_138.hg19.vcf.gz, 000G_phase1.indels.b37.vcf, and Mills_and_1000G_gold_standard.indels.b37.vcf. PrintReads was then used to generate the recalibrated file utilizing the data table generated from the previous step. GATK HaplotypeCaller was then used to generate g.vcf files for each reprocessed bam file. All g.vcf files were simultaneously passed to GATK's GenotypeGVCFs to generate a combined joint called vcf file. Variant Quality Score Recalibration pipeline was then applied. First the SNP VQSR was performed using the GATK VariantRecalibrator using annotations of QD, MQRankSum, ReadPosRankSum, FS, MQ, and InbreedingCoeff. Resources utilized from the GATK resource bundle included hapmap_3.3.b37.vcf , 1000G_omni2.5.b37.vcf,

1000G_phase1.snps.high_confidence.b37.vcf, dbsnp_138.b37.vcf, and dbsnp_138.b37.excluding_sites_after_129.vcf. Indel VariantRecalibrator was performed using annotations of FS, ReadPosRankSum, InbreedingCoeff, MQRankSum, and QD. Resources used from GATK resource bundle included Mills_and_1000G_gold_standard.indels.b37.vcf, Axiom_Exome_Plus.genotypes.all_populations.poly.vcf, and dbsnp_138.b37.vcf. The SNP recalibration was then applied using APplyRecalibration with a filter level of 99.6 and the INDEL recalibration was applied with a filter level of 95.0. Genomic posteriors were calculated for each call using the GATK CalculateGenotypePosteriors with supporting data from GATK resource bundle 1000G_phase3_v4_20130502.sites.vcf.gz and a pedigree file containing relationships between parents and probands in the trios. Genotypes with GQ less than 20 were labeled with lowGQ using GATK VariantFiltration. GATK VariantAnnotator was then applied to label PossibleDeNovo variants.

Variant sites with multiple alleles were split into single line entries in the vcf file using bcftools norm. Indels were left aligned and normalized using bcftools. VCF file was then annotated using Annovar's table_annovar command to annotated with refGene annotations, frequency information from exac, 1000 genomes, kaviar, and haplotype reference consortium, and scores from GERP, CADD 1.3, DANN, FATHMM, EIGEN, GWAVA, and DBNSFP30a.

A custom perl script then processed the table output to label inheritance calls and filter out low quality and common variants. Filter settings included QD>2, DP>5930 (to restrict to reads with an average call depth of 10x across all samples), ExcessiveHets<10, Maximum frequency in Exac or other population database or subpopulation database of 0.001, Passing the VQSR filter, Function of Exonic or splicing, and exonic function not being a synonymous SNV. This variant list was further filtered to identify presumed de novo variants where the trio genotypes are called such that both parents are homozygous reference, and the proband was called heterozygous. Further quality filtering was applied to

ensure that all genotypes in the trio have a QD greater than 20, Depth greater than 10. Also required that both parents have a variant call depth of 0 while the proband have a variant call depth greater than 4, a reference call depth greater than 4, and the reference calls must make up between 20% to 80% of the calls that that locus. To remove calls resulting from systematic sequencing noise, we excluded variants with an allele count greater than 3 in our samples.

Copy number variation calling

Copy number variation (CNV) prediction from exome data was done using the XHMM (eXome-Hidden Markov Model) caller. We used GATK to generate the depth of coverage statistics required for XHMM from the BAM files of our HPE cohort and a control set. GATK output was then run through the XHMM pipeline, generating a VCF file containing each predicted CNV. We then annotated each CNV for genes contained and cytogenetic region using Annovar. A custom perl script was used to evaluate for any de-novo CNVs in the HPE probands. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).

Sanger Sequencing

Variant sequence verification was performed using standard methods (Sanger et al. 1977). Sequencing was performed with v3.1 BigDye Terminator Cycle Sequencing Kit (Life Technologies, Grand Island, NY) in the ABI 3730xl Sequencer (Life Technologies) according to the manufacturer's protocol. Sequence data were aligned to the published reference genomic sequences (GenBank) using Sequencher 5.0.1 (Gene Codes Corp., Ann Arbor, MI).