The American Journal of Human Genetics, Volume 90

Supplemental Data

GPSM2 Mutations Cause the Brain Malformations and Hearing Loss in Chudley-McCullough Syndrome

Dan Doherty, Albert E. Chudley, Gail Coghlan, Gisele E. Ishak, A. Micheil Innes, Edmond G. Lemire, R. Curtis Rogers, Aizeddin A. Mhanni, Ian G. Phelps, Steven J. M. Jones, Shing H. Zhan, Anthony P. Fejes, Hashem Shahin, Moien Kanaan, Hatice Akay, Mustafa Tekin, FORGE Canada Consortium, Barbara Triggs-Raine, and Teresa Zelinski

Supplemental Methods

Clinical Assessments and Study Subjects

Individuals with CMS were enrolled with informed consent under protocols approved by the Health Research Ethics Boards of the participating academic institutions, and/or of the FORGE Canada Consortium. Inclusion criteria included prelingual severe or profound sensorineural hearing loss and abnormal brain imaging (at least ventriculomegaly). Clinical features were abstracted from medical records and/or reported by clinicians seeing the subjects at each site. Brain MRIs for all subjects, except 2A, 5A and 5B were systematically reviewed by a neurogeneticist (DD) and a pediatric neuroradiologist (GEI). In addition to the MRI examination for clinical care, all MRIs were evaluated for the extent of corpus callosum agenesis, ventriculomegaly, frontal gray matter heterotopia, frontal polymicrogyria and presence of arachnoid cysts.

SNP Analysis

Genomic DNA from six CMS affected individuals and eight of their family members was isolated from whole blood and sent to The Centre for Applied Genomics, (The Hospital for Sick Children, Toronto, Canada) for SNP genotyping (250K *NspI* array). Affymetrix (Affymetrix Inc., Santa Clara, California, USA) Genotyping console software (version 4.1) was used to analyze SNP data. The allele call rate ranged from 94.66 - 98.91%.

Preparation and Sequencing of Agilent All-Exome Capture Library

For whole exome sequencing, genomic DNA was quantified using a Quant-iT dsDNA HS assay kit and a Qubit fluorometer (Invitrogen). Approximately 500ng of DNA was sheared for 75 seconds at duty cycle "20%" and intensity of "5" using a Covaris E210, and run on an 8% PAGE gel. A 200-500 bp DNA fraction was excised and eluted from the gel slice, and was ligated to Illumina paired-end adapters following a standard protocol as described previously[1]. The adapter ligated DNA was amplified for 10 cycles using the PE primer set (Illumina) and purified as a pre-exome capture library DNA. The DNA was assessed using an Agilent DNA 1000 Series II assay, and the 500ng library DNA was hybridized to the 50Mb exon probe using the Human All exon Kit (Cat#G3370) following the SureSelect Human All exon 50 Mb Target Enrichment kit protocol. The capture DNA was purified using a Qiagen (QIAGEN Sciences, Maryland, USA) MiniElute column, and amplified for 12 cycles using the standard Illumina PE primer set. The PCR products were run of an 8% PAGE gel, the desired size range (320 – 370 bp) was excised and purified, and was then assessed using an Agilent DNA 1000 series II assay and diluted to 10nM. The final library DNA concentration was confirmed using a Quanti-iT dsDNA HS assay kit and Qubit fluorometer for cluster generation and sequencing.

Bioinformatic Analysis

All samples were sequenced on Illumina HiSeq-2000 platforms following the manufacturer's recommendations. Chastity-passed reads were mapped to the reference human genome (hg18) using Burrows-Wheeler algorithm 0.5.7[2]. Duplicate reads were removed prior to SNV calling; duplicate reads and reads having mapping quality of 0 were removed prior to small indel calling.

The aligned reads were exported to pileup format and variants were called using SAMtools 0.1.13[3]. Single nucleotide variants (SNVs) were filtered, using SAMtools.pl varFilter (all default parameters, except -D 1000), retaining SNVs with a minimum SNP quality of 20 (awk '$6>=20'). Small indels were processed similarly using SAMtools varFilter default parameters, except -D 1000, -d2 and -1 1.

All SNVs and small indels were imported into a local PostgreSQL database used to store and process human variation data[4]. Ensembl 54 annotations were used to identify variations that cause non-synonymous changes in protein-coding regions and those that fall within two bases of exon boundaries (which, potentially interfere with intron splicing). Using the database API, common polymorphisms were filtered out by their presence in dbSNP129/130[5], genomes sequenced as part of the 1000 Genomes Project[6], and other genomes sequences at the BC Genome Sciences Centre, or gathered from public sources. This filtering procedure led to a tractable number of candidate disease-causing mutations in each study subject (**Table S1**).

DNA Sequencing

Intronic primers flanking the identified mutations (from whole exome sequencing) in *GPSM2* were designed using ExonPrimer (Institute of Human Genetics, Munich, Germany).

The sequence of the forward and reverse primers, the annealing temperatures and sizes of the amplified products is given in **Table S2**. 100ng of genomic DNA was used in each amplification reaction. All PCR reactions started with a 3 min denaturation step at 95°C, followed by 30 cycles of 1 min at 95°C, 1 min at the calculated annealing temperature, and 2 min at 72°C. At the end of the 30 cycles, a 10 min extension at 72°C was performed. The PCR products were separated by agarose gel electrophoresis, and the DNA fragments purified using a Qiagen QIAquick Gel Extraction Kit (QIAGEN Sciences, Maryland, USA) according to the manufacturer's instructions. Sanger sequencing was performed at The Centre for Applied Genomics, (The Hospital for Sick Children, Toronto, Canada), using the same primers as were used for PCR amplification. The DNA sequence was determined for both strands of all fragments. Sequence variation was identified using Basic Local Alignment Search Tool (See Supplemental Web Resources), and all variations reported on the forward strand were confirmed on the reverse strand.

GPSM2 Transcript Analysis

Total mRNA was isolated from lymphoblastoid cell lines using the RNeasy Mini Kit (QIAgen, Valencia, CA), and polyA cDNA was generated using the RevertAid cDNA Synthesis kit (Fermentas Life Sciences, Glen Burnie, MD). To detect changes in exon 9 splicing, we performed RT-PCRs under standard conditions using forward primers in exons 8 and 9, and a reverse primer in exon 10. Primer sequences and conditions are listed in **Table S2**.

Supplemental Web Resources

Basic Local Alignment Search Tool, http://blast.ncbi.nlm.nih.gov

Supplemental References

1. Morin, R.D., Johnson, N.A., Severson T.M., Mungall, A.J., An, J., Goya, R., Paul, J.E., Boyle, M., Woolcock, B.W., Kuchenbauer, F., et al. (2010). Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat. Genet. *42*, 181-185.

2. Li, H., and Durbin, R. (2009). Fast and accurate short read alignments with Burrows-Wheeler Transform. Bioinformatics *25*, 1754-1760.

3. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map (SAM) format and SAMtools. Bioinformatics *25*, 2078-2079.

4. Fejes, A.P., Hadj Khodabakhshi, A., Birol, I., Jones, S.J.M. (2011). Human variation database: and open-source database template for genomic discovery. Bioinformatics *27*, 1155-1156.

5. Sherry, S.T., Ward, M-H., Kholodov, M., Baker, J., Phan, L., Smiergiski, E.M., Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308-311.

6. Marth, G. T., Yu, F., Indap, A.R., Garimella, K., Gravel, S., Leong, W.F., Tyler-Smith, C., Bainbridge, M., Blackwell, T., Zheng-Bradley, X., et al. (2011). The 1000 Genomes Project. Genome Biol. *12*:R84.

Table S1. Summary Statistics of Exome Resequencing

Subject	1A	5A	6A	7B	8A	8B
Total sequence yield (Gb) [a]	1.88	1.60	2.02	1.86	1.61	1.65
Total reads	437987248	407113542	384057734	356305352	294591038	353360082
Chastity-passed reads	365156554	348046754	356941878	330042772	255355748	271506376
Reads aligned to hg18	360384273	341102621	353113118	325932253	252648165	268356244
Mean exome coverage [b]	394.05	375.67	372.74	346.03	251.32	275.48
Total exonic variants [c]	24253	23951	24353	23992	24273	24434
Exonic synonymous SNVs	11542	11315	11426	11381	11432	11526
Exonic non-synonymous (ns) variants [d]	12711	12636	12927	12611	12841	12908
Splice-site SNVs [also nsSNVs]	904 [264]	857 [263]	897 [272]	910 [268]	925 [289]	916 [290]
Insertions/deletions [e]	518/552	499/524	547/604	544/558	540/591	546/561
ns variants not in dbSNP129/130	2607	2568	2667	2579	2782	2762
Novel ns variants [f]	220	214	221	182	179	202

[a] – span of the human genome (hg18) covered by ≥1 read aligned with Phred-scaled mapping quality of ≥10
[b] – average read depth of exons annotated in Ensembl 54; (sum of the number of reads aligned per site for all exonic sites) / (total number of exonic sites)
[c] – variants including SNVs as output from [SAMtools.pl varFilter –D 1000 | awk '$6>=20] and small indels as output from [SAMtools.pl varFilter –D 1000 –d 2 –l 1| awk '$6>=20], excluding those in 5'UTR, 3'UTR, introns and intergenic regions
[d] – variants include nsSNVs, splice-site SNVs within 2bp of exon boundaries and small indels
[e] – coding region small indels supported by ≥7 aligned reads
[f] – not previously reported in dbSNP129/130, 1000 Genomes Project, or other non-cancer genomes collected in the Michael Smith Genome Sciences Center local database

Table S2. ***GPSM2*** **Primers and Amplification Conditions**

Mutation	Direction	Primer Sequence	Annealing Temperature (°C)	Product Size (bp)
c.741delC	F	5’-TTA TGG CTG TAA GCT TCC CTA C-3’	57.5	575
	R	5’-GCT TAG ATG CCA AAA GTT CTG C-3’		
c.1062+1G>T	F	5’-TTC TTT ATC CCT TTA GTT CCT GC-3’	58.0	279
	R	5’-AGA AAT AGG GTT CCC TCC CC-3’		
c.1471delG	F	5’-TCC TCA TCC TTT ACC TTT TGC-3’	57.5	341
	R	5’-TTC AAT TCT CAG AAA TAA GCC AC-3’		
c.1661C>A	F	5’-CAA GGC CGA AAA GAT CTA GG-3’	59.0	410
	R	5’-AAA ATA AGC CCT TCC TGA ATG AG-3’		
PCR product				
Exon 8-10	8F	5’-GCA CAG TCT TGT TAC AGT CTT GG-3’	53.0	291
	10R	5’-GCT CAG ACC AAG AAC CAT TTG-3’		
Exon 9-10	9F	5’-GGA AGA GCA TGT TGG AGC TT-3’	53.0	171
	10R	5’-GCT CAG ACC AAG AAC CAT TTG-3’		

F, forward; R, reverse.