The American Journal of Human Genetics, Volume *90*

Supplemental Data

Somatic Mosaic Activating Mutations

in *PIK3CA* Cause CLOVES Syndrome

Kyle C. Kurek, Valerie L. Luks, Ugur M. Ayturk, Ahmad I. Alomari, Steven J. Fishman, Samantha A. Spencer, John B. Mulliken, Margot E. Bowen, Guilherme L. Yamamoto, Harry P.W. Kozakewich, and Matthew L. Warman

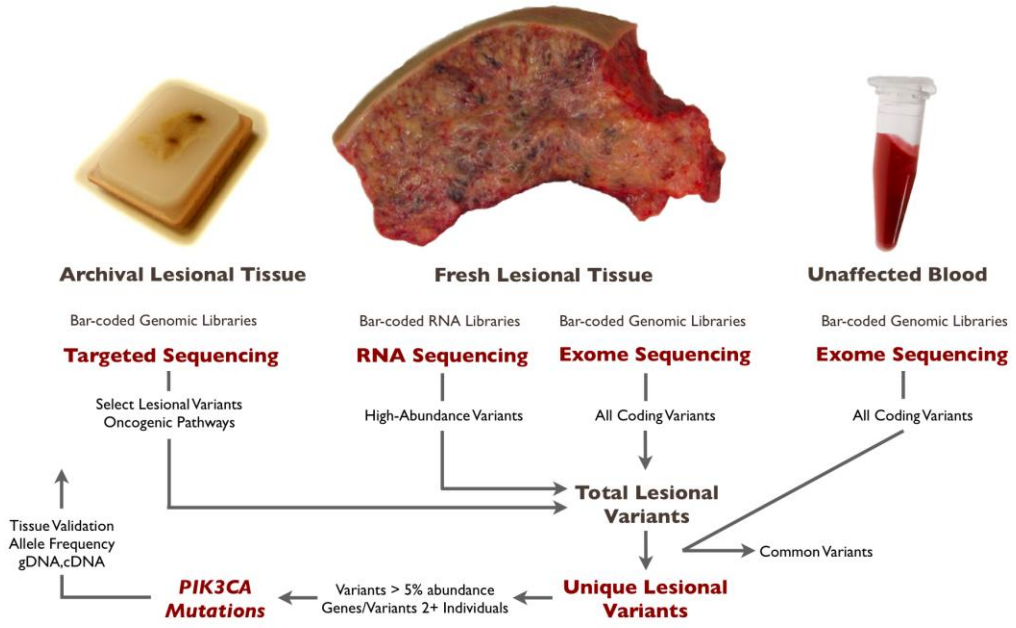

Figure S1. Sequencing Approach to Identify Somatic Mutations in CLOVES Syndrome

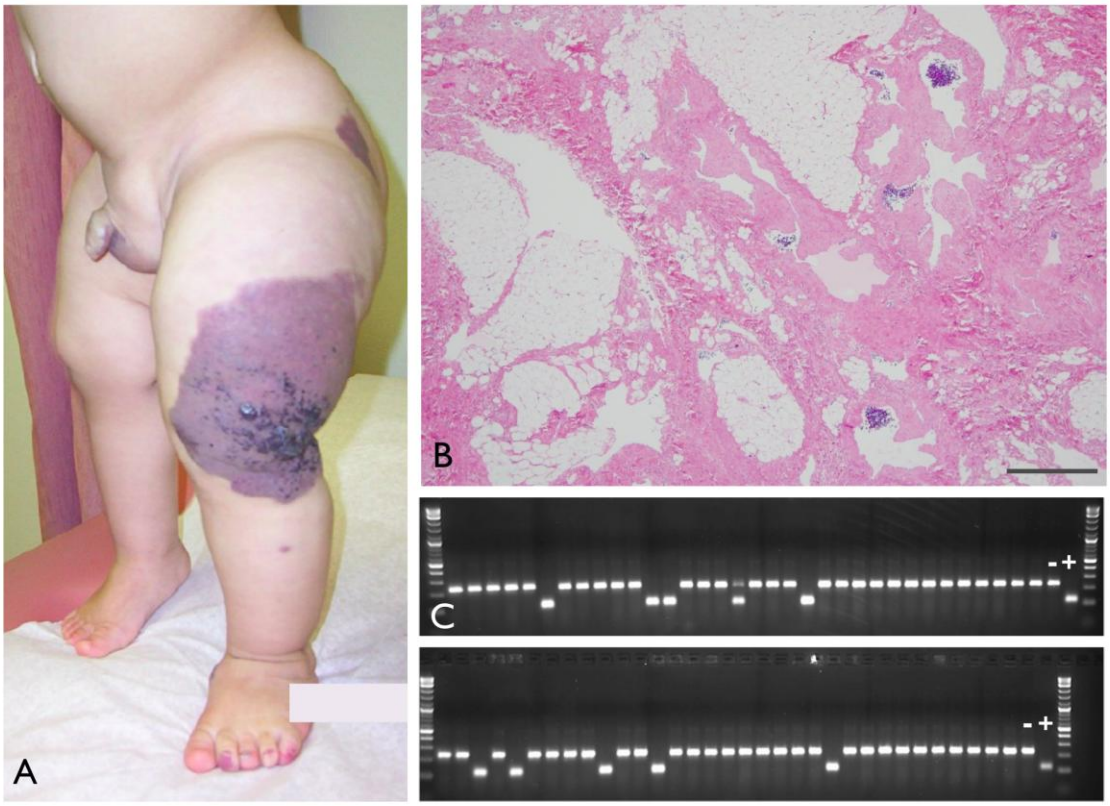

Figure S2. Identification of a *PIK3CA* Mutation in an Individual with Klippel-Trenaunay Syndrome

(A) Photograph of an individual with Klippel-Trenaunay Syndrome at age 13 months showing overgrowth of the left buttock and lower extremity, and vascular staining over the genital area, buttock, thigh, and digits. The right lower extremity and trunk are unaffected.

(B) Photomicrograph of a hematoxylin and eosin stained section of subcutaneous tissue resected from the lateral thigh at age 4.5 years showing a combination of venous and lymphatic malformation with admixed adipose tissue. The scale bar represents 500 μm.

(C) Photograph of a 3% agarose gel showing PCR amplimers that contain one of the *PIK3CA* mutation sites. Genomic DNA was extracted from FFPE lesional tissue and used as PCR template to produce amplimers, which were then subcloned. Sixty-eight subclones were PCR amplified and digested with the restriction enzyme HpyCH4IV, which cleaves only the mutant allele. In total, 10/68 (~15%) of subclones contained the c.3140A>G mutation. DNA size standards are included in the outer lanes, and positive (+) and negative (-) restriction enzyme digest controls are indicated.

Supplementary Table 1. Exome Sequence Data Filtering Analysis						
Participant	CL6	CL6	CL4	CL4	CL5	CL3
	Lesion	Saliva	Lesion	Blood	Lesion	Lesion
Total number of 50 bp paired-end reads	38,651,718	86,624,682	73,886,659	80,611,990	40,455,784	17,177,763
Total called variants	1,295,862	1,622,764	2,039,289	2,092,662	1,315,956	991,785
Variants in >1 read	18,120	20,997	23,919	24,389	18,212	15,479
Variants in >5% of reads	12,333	16,820	16,554	14,547	12,238	10,371
Variants after SNP filtering	690	763	797	776	830	664
Variants enriched in lesional tissue compared to blood or saliva	343		415			
Variants unique to lesional tissue	48		83			
Variants in >5% of reads in lesional tissue	31		20			
Genes with variants in >2 individuals	***PIK3CA***		***PIK3CA***		***PIK3CA***	***PIK3CA***

We enriched custom bar-coded genomic DNA libraries made from frozen tissue, blood, and saliva for exome sequence using the Agilent SureSelect human exome kit (Agilent Technologies, Santa Clara, CA USA). Three libraries were combined into each of 2 captures, and the samples were pooled post-capture and sequenced on 2 lanes of an Illumina HiSeq using 50 bp paired-end reads. The 3 targeted capture samples were combined with 4 samples from a different project and sequenced on an additional lane. We sorted the reads using Novobarcode, and aligned them to the human genome [UCSC Genome Browser hg19, GRCh37 (February, 2009)] using the Burrows-Wheeler Aligner (BWA). There was an average of 50 million reads per barcode. After merging BAM files using SAMtools, we removed PCR duplicates (10-18%) using Picard, as previously described,[a] and were left with an average of 36 million unique mapped reads for each sample. We performed variant calling using the Genome Analysis Toolkit (GATK), and filtered variants for dbSNP 132, 1000 Genomes, and the NHBLI whole exome database using Annovar. We initially considered variants present in greater than 5% of reads in affected tissue and ranked these variants with respect to the fold-coverage for that nucleotide. Finally, we focused on highly ranked variants that were either more abundant in affected tissue compared to unaffected tissue (blood or saliva) or seen solely in affected tissue compared to sequence from unaffected tissue. From this list, we first examined genes with variants present in multiple individuals. Only *PIK3CA* had variants in more than 2 participants (based on NCBI Reference Sequence NM_006218.2). The numbers of variants filtered at each step are detailed above.

Web Resources

Novobarcode, http://www.novocraft.com/main/page.php?s=novobarcode

Burrows-Wheeler Aligner, http://bio-bwa.sourceforge.net/

SAMtools, http://samtools.sourceforge.net/

Picard, http://sourceforge.net/projects/picard/?source=directory

Genome Analysis Toolkit, http://www.broadinstitute.org/gsa/wiki/index.php/Home_Page

Annovar, www.openbioinformatics.org/annovar/

References

1. Bowen, M.E., Boyden, E.D., Holm, I.A., Campos-Xavier, B., Bonafé, L., Superti-Furga, A., Ikegawa, S., Cormier-Daire, V., Bovée, J.V., Pansuriya, T.C., et al. (2011). Loss-of-function mutations in PTPN11 cause metachondromatosis, but not Ollier disease or Maffucci syndrome. PLoS Genet. *7*, e1002050, 1-11.

Table S2. RNA Sequence Coverage Data for *PIK3CA*

Participant	Depth of Sequence Coverage at Specific *PIK3CA* Mutation			*PIK3CA* Expression	
	p.Cys420Arg c.1258T>C	p.Glu542Lys c.1624G>A	p.His1047Arg c.3140A>G	RPKM[a]	Rank[b]
CL3	2	2	0	1.22	22,494
CL4	1	0	2	1.22	19,661
CL5	0	0	0	1.46	17,917
CL6	0	0	2	1.29	20,459

[a]RPKM= Reads per kilobase of exon model per million.
[b]Rank out of 54,607 transcripts.

RNA library preparation methods were adapted from a previously published protocol with minor modifications.[a] We homogenized tissue samples (~20-30 mg) in 1 ml Trizol, and separated total RNA by adding 0.2 ml of chloroform and centrifuging for 15 min at 4°C. We purified extracted RNA samples using the Purelink RNA Mini Kit (Life Technologies, Grand Island, NY USA), also treating with DNAse I to remove residual DNA. We then isolated mRNA by two rounds of poly-A bead-purification (Dynabeads, Life Technologies), reverse transcribed using random hexamers, and Illumina multiplexing adapters via ligation. We purified libraries on a 2% TAE agarose gel by extracting 150-300 bp fragments. The libraries were PCR amplified after determining the optimum number of amplification cycles by qPCR, and purified using Agencourt AMPure XP beads (Beckman Coulter, Danvers, MA). We evaluated completed libraries on a 4-20% TBE gel prior to sequencing the libraries on a single lane of an Illumina HiSeq 2000 flow cell with 50 bp paired-end reads.

We obtained a total of 158 million paired reads, with the number of reads per library ranging between 7.9 and 40.3 million. We mapped libraries to the reference genome (hg19) using the short-read alignment software RUM,[b] and then removed PCR duplicates using Picard. We quantified the expression level of each gene based on the number of mapped reads per kilobase of exon model per million (RPKM). Reference data for all genes and individual exons were downloaded from the UCSC Genome Server and Ensembl database with the R packages GenomicFeatures and biomaRt, respectively. *PIK3CA* was expressed at very low levels (~19,000 most common transcript), with 0-2X coverage in areas with mutations. Only wild-type sequence was detected in participants CL3, CL4, and CL6.

Web Resources

BiomaRt, http://www.bioconductor.org/packages/2.2/bioc/html/biomaRt.html

GenomicFeatures, http://www.bioconductor.org/packages/2.9/bioc/

html/GenomicFeatures.html

References

1. Christodoulou D.C., Gorham J.M., Herman D.S., Seidman J.G. (2011). Construction of normalized RNA-seq libraries for next-generation sequencing using the crab duplex-specific nuclease. Curr. Protoc. Mol. Biol. *94*, 4.12.1-4.12.11.
2. Grant G.R., Farkas M.H., Pizarro A., Lahens N., Schug J., Stoeckert Jr J.B., Hogenesch J.B., Pierce E.A. (2011). Comparative analysis of RNA-seq alignment algorithms and the RNA-seq unified mapper (RUM). Bioinformatics *27*, 2518-28.

Table S3. Primers Used for PCR Amplification of *PIK3CA* for Mutation Detection and Confirmation

Mutation	Source	Forward	Reverse
p.Cys420Arg/ c.1258T>C[a]	gDNA	ccttttggggaagaaaagtg	caaactccaactctaagcatgg
	cDNA	ttcctgatcttcctcgtgct	caaactccaactctaagcatgg
p.Glu542Lys/c.1624G>A[b]	gDNA	tcagcagttactattctgtgactgg	tttagcacttacctgtgactcca
	cDNA	ccacgcaggactgagtaaca	ggccaatcttttaccaagca
p.His1047Arg/c.3140A>G[c]	gDNA	ttgatgacattgcatacattcg	ggaatccagagtgagctttca
	cDNA	cgtgtgccatttgttttgac	ttttcagttcaatgcatgctg

[a]The c.1258T>C mutation introduces a novel BslI restriction enzyme site that was used to screen subclones for the presence of mutant alleles following initial confirmation by Sanger sequence analysis.
[b]The c.1624G>A mutation was confirmed by PCR amplification followed by Sanger sequence analysis.
[c]The c.3140A>G mutation introduces a novel HpyCH4IV restriction enzyme site that was used to screen subclones for the presence of mutant alleles following initial confirmation by Sanger sequence analysis.