Somatic *HRAS* p.G12S Mutation Causes Woolly Hair and Epidermal Nevi

Jonathan L. Levinsohn[1], Joyce Teng[2,3], Brittany G. Craiglow[1], Erin S. Loring[4], T. Andrew Burrow[5], Shrikant S. Mane[4,6], John D. Overton[4,6], Richard P. Lifton[4,6,7], Jennifer M. McNiff[1,8], Anne W. Lucky[9], Keith A. Choate[1,8]

Table of Contents:

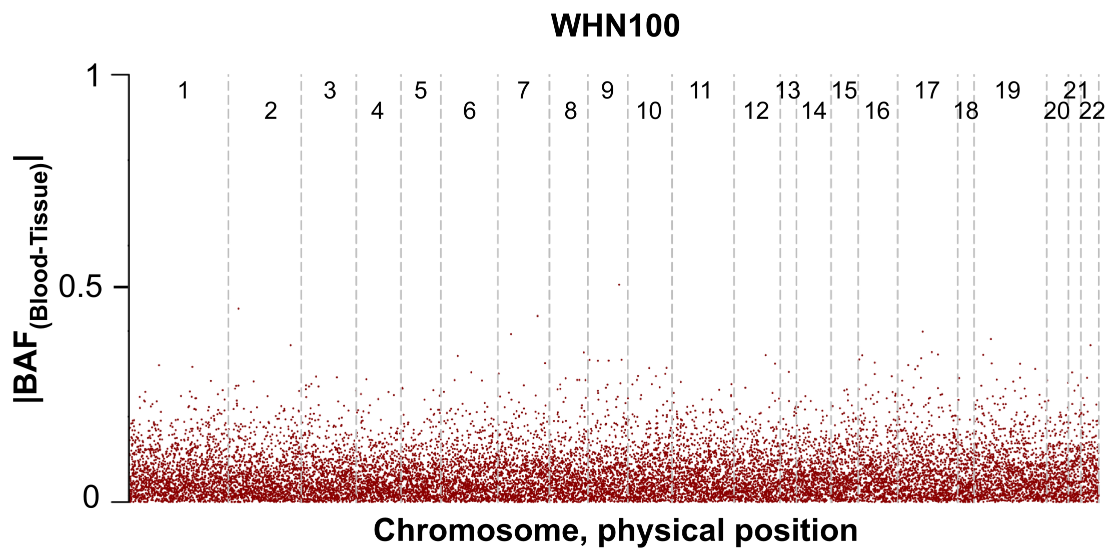
Supplementary Table 1……………………….2
Supplementary Table 2……………………….3
Supplementary Figure 1……………………...4
Supplementary Figure 2……………………...5
Supplementary Methods……………………..6
Supplementary References…………………...8

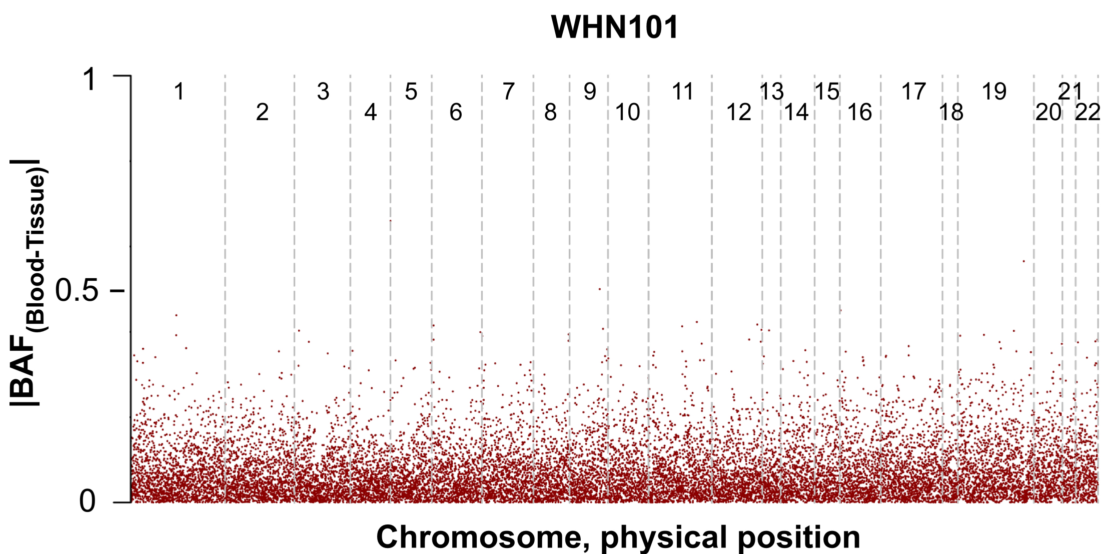
Sample	Mean Coverage	Bases Covered ≥8x	Bases Covered ≥20x	Mean Read Length
WHN100 Blood	60.1x	93.7%	82.0%	74 bases
WHN100 Tissue	154.1x	96.8%	93.4%	74 bases
WHN101 Blood	54.9x	93.4%	80.7%	74 bases
WHN101 Tissue	96.5x	95.0%	87.9%	74 bases

Supplementary Table 1. Exome Coverage Statistics. Whole exome sequencing was performed in cases WHN100 and WHN101 with at least 95% of bases covered at least 8x in tissue and 93.4% in blood. The average coverage of a given base ranged from 97-154X in tissue and 55-60X in blood.

Gene	Protein Effect	Samples with mutation
HRAS	G13R	85
KRAS	G12D	5
KRAS	G12V	3
HRAS	G13V	2
HRAS	G12C	1
KRAS	G12R	1

Supplementary Table 2. *HRAS* and *KRAS* mutations identified in archival NS samples. 97 histologically-diagnosed, archival scalp NS were screened for mutations in the first coding exon in *HRAS* and *KRAS* by RFLP analysis and Sanger sequencing. No *HRAS* p.G12S mutations were found.





Supplementary Figure 1. WHN100 and WHN101 demonstrate no evidence of LOH. The absolute value of B-allele frequency (ExomeCNV) was plotted against chromosomal position. No regions of B-allele frequency deviation were found, indicating that LOH is not present.

Supplementary Figure 2. Morphology of woolly and straight hair in WHN101. Hair was plucked from regions of wooly and straight hair in WHN101. Woolly hair is short and tightly curled. Adjacent straight hair appears morphologically normal.

SUPPLEMENTARY METHODS

Human subjects

Parental permission and child assent was obtained in writing for each of our cases and our study was approved by the Yale Human Investigational Committee and complies with the Declaration of Helsinki Principles.

DNA isolation

Genomic DNA was isolated from venous blood using standard phenol-chloroform extraction. Tissue DNA was isolated using the Qiagen FFPE DNA extraction kit (Qiagen, Valencia, CA).

Whole exome sequencing

DNA from tissue and blood was sheared, bar-coded libraries prepared, and whole exome capture performed (EZ Exome 2.0, Roche) by the Yale Center for Genome analysis. An Illumina HiSeq 2000 instrument was used for sequencing with blood samples pooled 6 per lane and tissue samples pooled 4 per lane. Resulting sequence was aligned to the hg19 human reference using ELAND (Illumina, San Diego, CA). A perl script was employed to trim sequence to the targeted intervals, and remove PCR duplicates. SNVs, deletions and insertions were identified using SAMtools software (Li *et al.*, 2009). Perl scripts were used to exclude common variants in dbSNP (build 135) and 1000 Genomes (release 05/2011) from further analysis, and to annotate the remaining variants for functional impact (Choi *et al.*, 2009). To identify somatic mutations, a perl script was used to evaluate all SNVs with > 8 total reads in both tissue and blood that had >1 non-reference calls in tissue. A Fisher's exact test was used to compare mutant and wild-type read numbers in tissue and blood (genome-wide threshold for significance $\sim 1.7*10^{-6}$, after Bonferroni correction for multiple testing of ~30,000 genes). Reads were examined to ensure that they had a Samtools quality score of at least 50. SNVs were ranked by Fisher's exact p-value and then filtered to examine only coding mutations (missense, nonsense, and splice site SNVs and indels) that were not present in 2577 control exomes. Aligned reads at these positions for the highest ranked candidates were then examined manually with the Broad Institute Integrative Genome Viewer (IGV) (Thorvaldsdottir *et al.*, 2012) to ensure that the reads were not mis-mapped.

Analysis of exome data for LOH

Output from the exome pipeline was used to identify LOH segments of ≥1Mbp with a p-value of 0.015 using the exomeCNV script (Sathirapongsasuti *et al.*, 2011). To illustrate the LOH data, the number of B-allele (non-reference) reads was divided by the total number of reads independently for tissue and blood at each SNV position, and the absolute value of the B-allele frequency difference between tissue and blood was plotted against the genomic location.

Screening of archival NS samples for *RAS* mutations

After isolation of DNA from archival tissue, samples were first screened via restriction fragment length polymorphism (RFLP) using the EagI enzyme for *HRAS*

c37.G>C (p.G13R). Samples that were not positive for the RFLP were PCR amplified and Sanger sequenced for the first coding exon of *HRAS* and *KRAS*.

Sanger sequencing
PCR was performed using Kapa 2G polymerase (Kapa Biosystems, Woburn, MA).
Primers listed below:
HRAS exon2 Forward: CTCCTTGGCAGGTGGGGCAG
HRAS exon2 Reverse: AGCCCTATCCTGGCTGTGTCCTG
KRAS exon2 Forward: TGAGTTTGTATTAAAAGGTACTGGTGGAG
KRAS exon2 Reverse: AACTTGAAACCCAAGGTACATTTCAG

SUPPLEMENTARY REFERENCES

Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P*, et al.* (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* 106:19096-101.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-9.

Sathirapongsasuti JF, Lee H, Horst BA, Brunner G, Cochran AJ, Binder S*, et al.* (2011) Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics* 27:2648-54.

Thorvaldsdottir H, Robinson JT, Mesirov JP (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*.