Supplemental Text

Mosaic activating *RAS* mutations in nevus sebaceus and nevus sebaceus syndrome

Bryan K. Sun, Andrea Saggini, Kavita Y. Sarin, Jinah Kim, Latanya Benjamin, Philip E. LeBoit, and Paul A. Khavari.

Exome sequencing

Genomic DNA was isolated using the Qiagen Blood/Tissue kit (Qiagen, Valencia, CA). Exome capture was performed with the Agilent SureSelect All-Exon v2 kit, and 100 bp paired-end library reads were generated on Illumina HiSeq 2000 to a 50-fold target average depth of coverage. Alignment of reads to the human genome was performed with DNAnexus (www.dnanexus.com) and bwa (Li and Durbin, 2009), and small nucleotide variants called by the DNAnexus cancer nucleotide variation algorithm using default parameters, as well as Seqgene (Deng, 2011) with the following parameters: minimum read coverage = 5, minimum percent for variant call = 0.05, minimum Phred-based cutoff for variant bases = 20.

Variant filtration strategy

All non-synonymous coding and splice-site variants from lesional tissue were filtered against patient-matched skin controls that were isolated from clinically normal surgical excision tips. We generated an initial candidate list consisting of genes that were mutated in the index case and at least 2 of the 4 sporadic nevus sebaceus. This list contained 83 genes. The resulting mutation list was further filtered against a set of control

internal exomes, against genes not expressed in the epidermis, and then annotated for predicted mutation deleteriousness.

Due to a large number of candidate mutations, we aimed to rationally prioritize more likely candidates based on the known clinical, histologic, and biological features of nevus sebaceus. We compiled a high-likelihood gene list from gene sets in the Molecular Signatures Database (http://www.broadinstitute.org/gsea/msigdb/index.jsp), which include data from Gene Ontogeny (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Reactome. Our set included 379 unique genes from the following gene sets: GO-sebaceous gland development, GO-androgen receptor signaling pathway, GO-Ras protein signal transduction, GO-hair follicle development, Reactome-Signaling to Ras, KEGG-basal cell carcinoma, KEGG-Hedgehog pathway, and KEGG-Wnt pathway.

Intersection of recurrent mutations with our high-likelihood gene list yielded 11 candidates: *APC2, BMP8A, CTBP2, FGD2, FZD1, HRAS, LAMA5, MAP2K2, MAPK12, NOTCH2,* and *SUFU*. Candidates were manually evaluated by predicted mutation deleteriousness and presence or absence of the mutation in the Exome Variant Server (http://evs.gs.washington.edu/EVS/). Targeted Sanger re-sequencing was used to verify candidate mutations identified by exome sequencing, and a subset of candidate genes (*APC2, CTBP2, LAMA5, NOTCH2,* and *SUFU*) were subsequently excluded because mutations could not be verified by Sanger sequencing, with the stipulation that a low prevalence mutation might be below the level of Sanger sequencing detection.

During the course of candidate mutation resequencing, positive identification of the recurrent *HRAS* p.Gly13Arg mutation in all five discovery set samples, as well as the absence of the mutation in all five matched controls, led to further focus on this

candidate, as described in the main text. In the exome data, the *HRAS* c.37G>C, p.Gly13Arg mutation was noted at the following frequencies:

Sample	Reads with *HRAS* c.37G>C (%)
Index case, lesional	2/12 (17%)
Index case, control	0/16 (0%)
NS #1, lesional	6/27 (22%)
NS #1, control	0/36 (0%)
NS #2, lesional	0/16 (0%)
NS #2, control	0/52 (0%)
NS #3, lesional	0/19 (0%)
NS #3, control	0/21 (0%)
NS #4, lesional	9/21 (43%)
NS #4, control	0/25 (0%)

Supplemental Text References

Deng X (2011). SeqGene: a comprehensive software solution for mining exome- and transcriptome- sequencing data. *BMC Bioinformatics*, *12*, 267.

Li H and R Durbin (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760.