Cell of origin and mutation pattern define three clinically distinct classes of sebaceous carcinoma

Jeffrey P. North1,2, Justin Golovato3, Charles J. Vaske3, J. Zachary Sanborn3, Andrew Nguyen3, Wei Wu4, Benjamin Goode2, Meredith Stevers2, Kevin McMullen1, Bethany E. Perez White4, Eric A. Colisson5, Michele Bloomer2,6, David A. Solomon2, Stephen C. Benz3 & Raymond J. Cho1

Sebaceous carcinomas (SeC) are cutaneous malignancies that, in rare cases, metastasize and prove fatal. Here we report whole-exome sequencing on 32 SeC, revealing distinct mutational classes that explain both cancer ontogeny and clinical course. A UV-damage signature predominates in 10/32 samples, while nine show microsatellite instability (MSI) profiles. UV-damage SeC exhibited poorly differentiated, infiltrative histopathology compared to MSI signature SeC (p = 0.003), features previously associated with dissemination. Moreover, UV-damage SeC transcriptomes and anatomic distribution closely resemble those of cutaneous squamous cell carcinomas (SCC), implicating sun-exposed keratinocytes as a cell of origin. Like SCC, this UV-damage subclass harbors a high somatic mutation burden with >50 mutations per Mb, predicting immunotherapeutic response. In contrast, ocular SeC acquires far fewer mutations without a dominant signature, but show frequent truncations in the ZNF750 epidermal differentiation regulator. Our data exemplify how different mutational processes convergently drive histopathologically related but clinically distinct cancers.

1 Department of Dermatology, University of California, San Francisco, CA 94115, USA. 2 Department of Pathology, University of California, San Francisco, CA 94143, USA. 3 NanOlimics, LLC, Culver City, CA 90232, USA. 4 Skin Tissue Engineering Core and Department of Dermatology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. 5 Department of Hematology and Oncology, University of California, San Francisco, CA 94158, USA. 6 Department of Ophthalmology, University of California, San Francisco, CA 94143, USA. Correspondence and requests for materials should be addressed to R.J.C. (email: Raymond.cho@ucsf.edu)
sebaceous carcinoma (SeC) accounts for 0.7% of skin cancers and carries a cancer-specific mortality rate of 3–6.7%1,2. SeC are believed to arise from sebaceous glands, ostensibly explaining their occurrence in sebocyte-dense sites such as the eyelids and head and neck regions3. Poorly differentiated ocular SeC has been reported to metastasize more frequently, often with fatal outcomes4. In all locations, SeC is considered a potential cutaneous marker for Muir–Torre disease, a cancer syndrome associated with germline mutations in mismatch-repair pathway components MLH1, MSH2, MSH6, and PMS25–7. Muir–Torre is a variant of hereditary non-polyposis colon cancer, also known as Lynch syndrome, and causes patients to develop endometrial, ovarian, gastric, biliary, and genitourinary tract cancers, as well as benign and malignant sebaceous neoplasms and keratoacanthomas. Tumors defective in this pathway show errors in replicating microsatellite repeat sequences, generating a mutagenesis pattern known as microsatellite instability (MSI).

Recently, Tetzlaff et al.8 applied targeted sequencing in ocular SeC, reporting that 52% (14 of 27) harbored somatic mutations in PI3K signaling components. TP53 and RB1 mutations were detected in ocular SeC, while four extra-ocular SeC showed additional mutations affecting DNA repair and chromatin remodeling pathways. Three extra-ocular SeC demonstrated a high level of MSI with somatic mutations in the mismatch-repair genes MLH1 and MSH2, suggesting that mutations outside the germline may drive mutation and tumorigenesis in these tumors. However, because of the targeted nature of the sequencing in this study, the broader patterns of mutation in SeC remain unknown.

In this study, we report that SeC fall into three distinct subtypes based on mutational genetics, and are heavily influenced by anatomic site of origin. Two mutually exclusive classes are dominated by either mismatch-repair-derived insertions and deletions or ultraviolet (UV) signature single-nucleotide mutations. In contrast, ocular SeC harbor fewer mutations but acquire recurrent truncating mutations in the ZNF750 transcription factor.

Results

Whole-exome sequencing of 32 SeC. In order to study the genetic alterations that characterize SeC, we performed whole-exome sequencing and selected transcriptome analysis on a cohort of 32 SeC. To identify possible associations between mutation profile and anatomic site, 23 SeC were sequenced from the head and neck (nine of which were ocular) and 9 were analyzed from the trunk. Supplementary Data 1 details all samples used for this study. Sequencing was performed after targeted library capture of 62.52 Mb of coding nucleotides. The average reads for tumor and normal samples were 2,230,611,589 and 117,499,836, respectively, generating an average coverage of ×84 and ×46 across 99% of captured sequence. The number of somatic single-nucleotide variants (SSNVs) identified in SeC exomes ranged unexpectedly widely, from 73 to 36,659 (Fig. 1a), reflecting mutation prevalences from 1.2 to 536 mutations per Mb (Supplementary Data 2). We found that SeC showed low incidence of CNVs; only 12/32 harbored more than five events. The most common aberration, single-copy loss of chromosome 17p, where TP53 is located, occurred in nine tumors. No biallelic deletions or focal high-level amplifications were identified. Supplementary Data 3 lists all copy-number aberrations >2.5 megabases in size.

Mutational classification of sebaceous carcinoma. Thirteen SeC, including all nine ocular SeC, demonstrated a low mutation prevalence between 1.2 and 5.2 per Mb, distinguishing them sharply from the high mutation burden in other tumors (samples 1–13, Fig. 1a). Of the remaining 19 tumors, 9 (28% overall) exhibited a somatic mutation pattern composed of at least 30% indels, while the other 10 highly mutated cancers were predominated by SSNVs (samples 14–22 vs. 23–32, Supplementary Data 2). None of the patients in our study carried known germline mismatch-repair defects, but it was possible that their SeC marked an initial manifestation of Muir–Torre syndrome. We thus analyzed uninvolved normal skin in these patients to assess for possible germline mutations in MLH1, MSH2, MSH6, and PMS2, DNA repair genes whose inactivation is causative of Muir–Torre syndrome. A germline nonsense mutation was found in only one case, SeC sample 17 (MSH6 p.R911*), which harbored high numbers of indels. Truncating mutations or likely damaging mutations in POLE or APOBEC, other potential sources of high mutation burden, were not detected.

These observations led us to suspect that three distinct mutational mechanisms had given rise to the tumors of our cohort. We therefore examined our mutations using established algorithms capable of distinguishing known mutagenic processes.9 This analysis structure assesses a somatically altered base and surrounding sequence context to deduce a likely cause of each mutation. These deduced mutational origins are themselves classified into a series of discrete categories or “signatures.” The analysis revealed that the tumors in our study fall into three distinct classes (Fig. 1b).

Ten different tumors were defined by mutations harboring a UV-damage signature (signatures 7 and 11) with high numbers of CC > TT dinucleotide mutations induced by exposure to ultraviolet radiation. These SeC developed on sun-exposed areas and showed scarce evidence of mismatch-repair errors (Fig. 1b). Isolated signature single-nucleotide mutations (signature 1) were also found in 11 tumors, representing an analysis signature subset of the remaining cohort of tumors. These deduced mutational origins are themselves classifiable into two distinct classes (Fig. 1b).

Interestingly, the remaining 13 of our SeC, including all nine ocular cancers, were pauci-mutational (the group described above with mutation prevalences between 1.2 and 5.5 per Mb), a phenomenon occurring much less frequently in other cancers of sun-exposed skin12. This range is on average 30-fold fewer than in the MSI and UV classes. Tumor purity, estimated based on mutant allele frequency, exceeded 70% in these cancers, confirming an authentic low mutation prevalence. While almost all tumors (88%) showed some amount of signature 1 mutation, eight of the pauci-mutational tumors were found to have SSNVs primarily consistent with this signature, which is typified by spontaneous deamination of 5-methylcytosine and thought to be related to natural aging. Staining of six UV class and six pauci-mutational SeC revealed MLH1 and MSH2 in all 12
tumors, confirming these subgroups maintain expression of mismatch-repair proteins (Fig. 2).

UV-damage SeC show more aggressive histopathologic features. We examined the clinical and histopathologic characteristics of tumors in each of these three classes. The mean age of patients developing UV SeC was 83.1 (7.82), compared to 69.6 (15.2) for MSI and 72.7 (7.26) for pauci-mutational SeC. These differences are significant using a one-way analysis of variance at a p value of 0.017. Uniformly, MSI tumors showed greater differentiation (six well-differentiated, three moderately differentiated, Fig. 3a, b) and all nine had well-circumscribed (non-infiltrative) borders. We also found that seven of nine MSI tumors arose on the trunk, a site almost completely spared in the other two classes (Fig. 4). In contrast, UV class tumors were more poorly differentiated (eight moderate, two poor, Supplementary Data 1) and 70% exhibited an infiltrative growth pattern (χ² test p value = 0.003, Fig. 3c, d).

All UV signature tumors showed histopathologic evidence of moderate to severe sun damage. Prominent squamous differentiation was present in only two tumors, both of the UV signature class. Of the pauci-mutational class, nine were ocular and four were cutaneous on facial skin (Fig. 4). Two (both extra-ocular) were well-differentiated, similar to SeC in the MSI class. Two of the extra-ocular pauci-mutational tumors had histopathologic evidence of moderate to severe sun damage and infiltrative growth patterns. The nine pauci-mutational ocular SeC were all moderately differentiated and showed scant evidence of chronic sun damage, consistent with development on sun-protected epithelium of the inner eyelid.

SeC driver genes are specified by mutational class. In order to identify the recurrently mutated genes in our SeC, the MuSIC...
We restricted this analysis to somatic changes occurring at a minimum allele frequency of 20% to reduce the impact of the high background mutation rate and enhance for positively selected variants. Only 11 genes with a convolution test false discovery rate of <1 × 10^{-5} showed both multiple point mutations and indels in our series, including TP53, NOTCH1, NOTCH2, ZNF750, RREB1, KMT2D, and FAT3 (Fig. 5, Supplementary Data 4). KM2TD and NOTCH1 are known epithelial tumor suppressors affecting epigenetic programming and differentiation, and mutation of TP53 and NOTCH1/2 is classically associated with SCC. Although the sebaceous gland epithelium believed to give rise to SeC resembles the germinative differentiation of basal cell carcinomas (BCC), the mutation spectrum of SeC more closely resembles that of cutaneous SCC. No nonsense mutations were observed in PTCH1, SMO, SUFU, or other Hedgehog signaling pathway genes typically mutated in BCC.

We then re-examined individual gene mutations in the context of mutational classes we discovered in SeC (Fig. 5). Truncating NOTCH1 mutations were notably excluded from ocular SeC (1/9 harbored non-synonymous mutations) while 8/23 of the remaining tumors acquired a nonsense or frameshift mutation (Fig. 5). In contrast, 9/13 pauci-mutational SeC acquired mutations in ZNF750, including 6 nonsense and frameshift mutations, compared to just 2/19 truncations in the remaining SeC. Only 2/19 of the UV signature and ocular SeC harbored activating hotspot KRAS or HRAS mutations, as opposed to 10/13 of the MSI and pauci-mutational cutaneous classes (χ² test, p value = 0.0002). Similarly, 4/9 MSI tumors harbored truncating mutations in RREB1^{14}, compared to only 2/23 of the non-MSI cancers.

UV SeC transcriptomes resemble those of SCC and BCC. All classes of SeC acquire recurrent mutations identified in cutaneous SCC, including TP53, KMT2D^{15, 16}, and NOTCH1/NOTCH2^{17}. However the squamous differentiation and infrequency of KRAS/HRAS activation in the UV subclass led us to hypothesize that these tumors may display epigenetic similarity with SCC, which share transcriptional patterns across their many subtypes^{18}. RNA sequencing was performed on five UV class, four MSI, and four pauci-mutational SeC, along with six each of spontaneously arising cutaneous SCCs and BCCs (Fig. 6). Strikingly, heavily UV-damaged samples SeC 24, 25, and Y shared transcriptomic patterns resembling those found in SCC and BCC, and none of the five clustered with MSI or pauci-mutational SeC. SeC 24 RNA expression was more strongly correlated with all six SCCs than any SeC (t test p value = 0.0002). Similarly, SeC 25 RNA expression correlated with all six BCCs more closely than with any SeC (t test p value = 0.0002). SeC Y had higher average correlation with SCC samples than SeC; its transcripts also differentiated it significantly from BCC (t test p value = 0.009). Pauci-mutational SeC 3 and 9 also showed some similarity to SCC, but clustered most strongly with MSI SeC. None of the eight pauci-mutational and MSI SeC profiled showed sufficient transcriptional similarities to classify them with keratinocytic malignancies.
UV and MSI SeC harbor high neoantigen burdens. The different mutation rates between SeC subclasses seemed likely to proportionately impact predicted neoantigen load, potentially affecting immunotherapy response. As expected, UV-damage tumors harbored the largest numbers of predicted neoantigens, followed by MSI tumors, then the pauci-mutational SeC (Supplementary Data 5). The total tumor mutation burden per sample exceeded 100 for all 32 tumors, which has been reported as a threshold for response to CTLA-4 blockade in melanoma. The total predicted neoantigen burden exceeded 100 for 24 of 32 samples, the threshold for clinical benefit reported in non-small cell lung cancer.

Clonality of predicted neoantigens ranged widely in our SeC series, from 0 to 100% and did not show differences by mutational class. Eleven out of 32 samples harbored predicted neoantigens that were at least 50% clonal.

Discussion

Here, we report three distinct mechanisms by which SeC accumulate somatic mutations. Indel-rich MSI SeC appear to be initiated by inactivating mutations in mismatch-repair genes (which can be either germline or somatic), while SeC with UV-damage-associated mutational signatures develop on heavily sun-damaged skin. Although sebaceous neoplasms have been reported anecdotally to harbor somatic mutations in mismatch-repair genes, our results show that virtually all MSI class SeC arise from either a germline or somatic mutation in this pathway. A third, previously undescribed pauci-mutational class is primarily represented by facial SeC. In fact, all nine of the ocular SeC in our series displayed this pattern.

Remarkable mutual exclusivity characterizes the mutation signatures in SeC. Even in MSI SeC that arise on sun-exposed skin, mutations characteristically associated with UV damage are virtually absent. It is conceivable that all SeC originate from sebocytes, with the UV class undergoing dedifferentiation and acquiring keratinocytic features. However, the considerable UV-damage burden in this subset (Fig. 1) militates against this hypothesis. UV class SeC are not found more superficially in the dermis than MSI tumors, where they could acquire UV mutations during a later growth phase. In fact, MSI SeC, which show the most distinctive features of sebaceous differentiation, acquire much lower levels of UV-associated mutagenesis, as one would expect of sebocytes located in the mid-dermis. A hypothesis of divergent evolution of SeC subtypes also fails to explain why, like MSI SeC, UV subclass tumors acquire truncating NOTCH1 mutations, aberrations that should restrict keratinocytic differentiation.

We therefore favor an alternative model in which poorly differentiated UV-damage SeC, like cutaneous SCC, arise from a subpopulation of more superficial keratinocytes vulnerable to UV mutagenesis, either from the epidermis or superficial portion of the folliculosebaceous unit. In this model, UV class SeC may acquire sebocytic differentiation because of specific somatic alterations or the epigenetic state of the cell of origin. This hypothesis is consistent with the scarcity of UV-damage SeC on...
the trunk (Fig. 4) and helps explain why tumors with prominent squamous differentiation are found exclusively in the UV class. A keratinocytic origin of UV SeC also accounts for relatedness of transcriptomes of UV SeC and SCC and BCC, which diverge substantially from MSI and pauci-mutational SeC.

In contrast, MSI SeC invariably arise through inactivation of mismatch-repair genes. The predilection of these cancers for sebocyte-rich skin (Fig. 4) raises the question: do all cells of epithelial lineage acquire stochastic mutations in mismatch-repair genes, with sebocytes particularly susceptible to subsequent hypermutation and tumor evolution? Or do sebocytes permit higher mutation rates that stochastically inactivate repair genes? The answer is germane to the anatomic distribution of other skin cancers. For example, in multiple hereditary infundibulocystic BCC syndrome, a germline mutation inactivates one copy of the SUFU repressor of the Hedgehog signaling pathway. Basal cell cancers arise in affected individuals when the wild-type trans copy is inactivated by small deletions. However, these cancers arise predominantly in the central face and genital area. Either cells at these sites are somehow more prone to deletion events, or they are more vulnerable to tumorigenesis once such events occur. We also report the RREB1 gene is frequently mutated in MSI SeC. A transcription factor that functions in adipocytic differentiation, RREB1 has not been previously implicated as a tumor suppressor gene in human cancer. We detected robust RREB1 protein expression in adult sebaceous glands; its role in differentiation and the consequences of its deactivation in SeC remain to be investigated.

Pauci-mutational SeC occur exclusively on the face (13/13), and show no dominant mutational signature. They may represent relatively sun-shielded epithelial cells that by chance acquire driver mutations because of the modest mutagenesis of normal aging. Interestingly, the ZNF750 transcription factor acquired truncating mutations in 6/13 of these tumors. ZNF750 is required for normal epithelial homeostasis and has recently been explored as a lineage-specific tumor suppressor in SCC. Its striking rate of inactivation (frameshift or nonsense mutations) in ocular SeC (5/9 samples) has not been reported. The recent Tetzlaff et al. targeted sequencing study in ocular SeC did not appear to assess for ZNF750 mutations in their panel of genes. Our data suggest that ZNF750 inactivation is critical for tumorigenesis in ocular, and perhaps facial SeC more generally. Only two of the 19 UV and MSI SeC were mutated for ZNF750. Conversely, 0/9 ocular SeC acquired truncating NOTCH1 mutations, compared to 5/9 of the MSI class. The class specificity of NOTCH1, RREB1, and ZNF750 mutation supports the theory that epigenetic differences in epithelial sites underlie distinct mechanisms of SeC tumorigenesis. We did not detect activating PIK3C or inactivating PTEN mutations in either ocular or extraocular SeC, but such mutations may contribute a selective advantage in a minority of cases.

Our data illustrates how cancers grouped by histopathologic similarity actually arise from different mutational mechanisms, likely acting on distinct cells of origin. These differences reframe clinical approach. Investigations for germline mismatch-repair defects will rarely prove successful in the context of UV class SeC.
MSI SeC may also behave indolently, much like their visceral counterparts, mitigating the urgency of follow-up and surveillance. This classification also impacts therapeutic approaches to those cancers that metastasize. Given the reported correlation between poor differentiation and metastasis, it is likely that disseminated disease either falls into a highly mutated (UV-damage) class amenable to immunotherapeutics, or a pauci-mutational class, including most ocular SeC, predicted to respond less robustly. In our series, poorly differentiated UV SeC showed predicted neoantigen burdens at times exceeding 500 per sample, >50% of them clonal in some cases, well above published thresholds for reliable treatment response.

**Immunostaining.** Formalin-fixed paraffin-embedded (FFPE) sections of 4-µm thickness were stained with the following immunoperoxidase stains: adipophilin immunoperoxidase (393A-18, Cell Marque, Rocklin, CA, predilute on Leica Bond); MLH1 (PA0610, Leica); MSH2 (IR08, Dako).

**DNA extraction and exome sequencing.** Ten 5-µm tissue sections were cut using a board-certified pathologist who identified and marked regions of malignancy. Using the marked-up image as a guide, tissue sections were macerodissected to collect the desired regions. Each collected cell population was then extracted using the Qiagen QiAamp DNA FFPE Tissue Kit for DNA and RNeasy FFPE Kit for RNA. Extracted material was then quantified using a Qubit fluorometer. DNA-seq libraries were captured to exome regions using xGen Exome Research Panel v1.0 (IDT), and libraries were prepared using the KAPA Hyper prep kit. DNA libraries were sequenced to a target depth of ×200 for tumor sample, ×100 for normal samples on the Illumina HiSeq platform.

RNA-seq libraries were prepared using the KAPA Stranded RNA-Seq Kit with RibonEx (Kapa Biosystems, Wilmington, MA) and sequenced to a target depth of 200-M reads on the Illumina HiSeq platform (Illumina, San Diego, CA). RNA samples were aligned to RefSeq build 73 transcriptome using Bowtie2 v2.2.6 and quantified using RSEM v1.2.25.

**Exome variant calling.** Raw sequencing data in Illumina’s fastq format was converted into fastq files with base quality scores encoded in the Sanger basecall format. Next, the reads were aligned using the Burrows-Wheeler Aligner (BWA)29. This aligner is based on the Burrows-Wheeler transformation, aligns paired-end reads and handles indels robustly. The output of BWA are the aligned reads in sequence alignment/map (SAM) format. Reads stored in SAM format were then converted to the binary SAM (BAM) format using the samtools software30. Once reads were in the sorted and indexed BAM file format, position-based retrieval of reads was expedited and data storage requirements were minimized. Next, to remove erroneous mutation calls due to PCR duplication, all duplicate reads were removed using samblaster31. After removal of

---

**Methods**

**Sample inclusion and histopathologic classification.** With UCSF IRB approval and a waiver of informed consent, we screened the UCSF dermatopathology tissue archives by diagnostic code for SeC from 2007 to 2015 that had available tumor and normal skin remaining in the paraffin block(s). The diagnosis of SeC was confirmed by a board-certified dermatopathologist (JPN). For any tumors in which sebocyte differentiation was questionable, adipophilin immunoperoxidase staining was used to confirm sebaceous differentiation. Any tumors in which the diagnosis was uncertain after adipophilin stainings were excluded. Each tumor was classified by the degree of differentiation (e.g., poorly, moderately, well differentiated) as manifested by sebaceous gland organization and the number of mature sebocytes present. High-grade cytologic atypia and squamous differentiation were noted if present. Solar elastosis was graded as minimal or moderate to severe, and tumors were assessed for infiltrative growth pattern at the edges of the neoplasm.
the duplicate reads, the base quality scores were recalibrated using the CountCovariates and TableRecalibration tools included in the Genome Analysis Toolkit software, also developed by the Broad Institute (https://www.broadinstitute.org/gatk/). The tumor and matched-normal BAM files were analyzed for SNVs with the following method. Sequenced bases with ≥8 unique (nonduplicate reads) with mapping quality ≥20 in both tumor and matched-normal were used to compute the likelihoods of all possible genotypes (AA, AT, AC, and so on) using a mapping quality-based error model, available in the samtools source code. If fewer than two reads support any non-reference allele at the current position, then the position was deemed homozgyous reference and no further analysis was performed. The genotype likelihoods were used in a Bayesian model used by the SomaticSniper method\(^2\),\(^3\),\(^4\), incorporating a prior probability on the reference, the fraction of heterozygous positions in the human genome, the probability to convert the normal genotype to the tumor genotype. Each tumor/nonormal genotype pair was scored using this model. The genotype pair with the highest likelihood, given the data, was chosen as the most likely tumor and normal genotype. Any position that was determined to be homozgyous for the reference allele in both tumor and normal was not further analyzed.

If tumor and normal genotypes were identical, then the variant was classified as germline. Instances where the normal genotype was heterozygous and the tumor genotype was homozgyous suggest new point mutations but regions of loss-of-heterozygosity (LOH), and such variants were so classified. For variants classified as either germline or LOH, the log-likelihood of the paired genotype was used to compute a Phred-scaled quality/convidence of the germline variant. All other variants were classified as a somatic mutation and their somatic score (SS) was calculated. For any position where the tumor and/or normal genotype was not homozgyous for the reference allele, a number of metrics were computed, including number of total reads, number of allelic reads, average base and mapping quality, number of reads with mapping quality, number and quality sum of mismatches in m, or variant or reference allele, the distance of the variant to the 3′ end of the read, and number of reads aligned to the forward/reverse strand. All putative variants and their associated metrics were converted to the variant call format (VCF) and the following filters applied:\(^5\): conf: genotype quality or SS >15 dp; depth (DP) of normal + primary) ≥2; m oligot: number of mismatch = DP reads <5; ab male allele strand bias p value >0.05 (binomial test); mmsq: quality sum of mismatches (per read) ≤20; am: average number of mismatches (per read) ≤1.5; dpf: fractional distance to 3′ <0.2 or >0.8; ad: mutant allele depth in tumor ≥4; gad: mutant allele depth in normal ≤3; ma: read support ≥2 for two or more allelic alleles at this position\(^6\). A single VCF file was produced for each tumor sample.

### Calling copy-number variants.

For copy-number variants (CNVs), referred to here as CNVs, average tumor vs. matched-normal relative coverage and standard deviation were calculated for each captured exon by dividing read depth measured in the tumor by the read depth in the matched-normal for each position within the exon. Exons that were insufficiently covered, average read depth <5 reads, in two tumor and matched-normal were removed from the remaining analysis.

Average majority allele fraction in both tumor and matched-normal was computed for each segment with at least one heterozygous SNP in the normal tissue: $\text{AF}_\text{maj} = \frac{\text{DP}_\text{maj}}{\text{DP}_\text{total}}, \text{AF}_\text{maj} = \frac{\text{DP}_\text{maj}}{\text{DP}_\text{total}},$ where $\text{DP}_\text{maj}$ and $\text{DP}_\text{maj}$ are the read depths of the germline allele with the greatest read support and $\text{DP}_\text{total}$ and $\text{DP}_\text{maj}$ are the total read depths at the position in the tumor and matched-normal, respectively. Relative coverage is determined simply by dividing the coverage observed in tumor by the matched-normal’s coverage, i.e. $RC = \frac{\text{DP}_\text{read}}{\text{DP}_\text{read}}$. The standard deviation of $\text{AF}_\text{maj}$ estimates was computed for segments featuring at least three heterozygous SNPs. Only dbSNP sites with sufficient read support in the tumor and matched-normal ($\text{DP}_\text{maj} \geq 10, \text{DP}_\text{maj} \geq 20$) deemed heterozygous (0.25 ≤ $\text{AF}_\text{maj} ≤ 0.75$) are considered when computing majority allele fraction estimates.

To determine CNVs, the exon-level statistics computed above were iteratively aggregated into larger segments in an agglomerative process similar in spirit to hierarchical clustering. In the first round, every pair of neighboring exons was analyzed. Neighboring exons that did not have significantly different relative coverage were merged into a single segment. The average relative coverage for the new segment was calculated as the base-pair count adjusted averages and standard deviations of the two individual exons measurements, while $\text{AF}_\text{maj}$ and its standard deviation were recomputed for the new segment. The procedure described above was continued with neighboring segments using the above method for three rounds.

The relative coverage estimates of all segments were centered to the median of the entire genome, which was assigned the value of 1.0, signifying the normal copy-number state. The skew in $\text{AF}_\text{maj}$ estimates caused when the majority allele’s read support was too low was instead of an underlying imbalance in copy number was corrected by subtracting the $\text{AF}_\text{maj}$ estimated in the matched-normal sample. However, since this bias is only present in regions where both alleles have equal copy number, the correction is made using the following equation:

$$\text{AF}_\text{maj} = \text{AF}_\text{maj} - \left( \text{AF}_\text{maj} - 0.5 \right) e^{-\frac{800}{\text{read depth}^2}}.$$
transformed via log2 (1 + TPM) before further correlation analysis. The Pearson correlations reported here were calculated on the 500 genes with highest variance. The correlation trends were verified on two further gene sets: all genes with variance >1.0, and the full transcriptome. Hierarchical clustering was performed on the correlation matrix in R, using the default parameters of the heatmap.2 package from gplots version 3.0.1.

**Data availability**. Sequence data has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS0000102869.

Further information about EGA can be found on https://ega-archive.org and “The European Genome-phenome Archive of human data consented for biomedical research”36. All other data are available from the authors upon reasonable request.

Received: 20 July 2017 Accepted: 28 March 2018

Published online: 14 May 2018

**References**

1. Kyllo, R. L., Brady, K. L. & Hurst, E. A. Sebaceous carcinoma: review of the literature. Dermatol. Surg. 41, 1–15 (2015).
2. Warren, S. & Warvi, W. N. Tumors of sebaceous glands. Am. J. Pathol. 19, 441–459 (1943).
3. Nelson, B. R., Hamlet, K. R., Gillard, M., Rulau, D. & Johnson, T. M. Sebaceous carcinoma. J. Am. Acad. Dermatol. 33, 1–15 (1995).
4. Rao, N. A., Hidayat, A. A., McLean, I. W. & Zimmerman, L. E. Sebaceous carcinomas of the ocular adnexa: a clinicopathologic study of 104 cases, with five-year follow-up data. Hum. Pathol. 13, 113–122 (1982).
5. Shalin, S. C. et al. P53 staining correlates with tumor type and location in sebaceous neoplasms. Am. J. Dermatopathol. 34, 129–138 (2012).
6. Kruse, R. et al. Is the mismatch repair deficient type of Muir-Torre syndrome confined to mutations in the hMSH2 gene? Hum. Genet. 98, 747–750 (1996).
7. Ponti, G. & Longo, C. Microsatellite instability and mismatch repair protein expression in sebaceous tumors, keratoacanthoma, and basal cell carcinomas with sebaceous differentiation in Muir-Torre syndrome. J. Am. Acad. Dermatol. 68, 509–510 (2013).
8. Tetzlaff, M. T. et al. Next-generation sequencing identifies high frequency of mutations in potentially clinically actionable genes in sebaceous carcinoma. J. Pathol. 240, 84–95 (2016).
9. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. Nature 500, 415–421 (2013).
10. Salipante, S. J., Scruggs, S. M., Hampel, H. L., Turner, E. H. & Pritchard, C. C. Microsatellite instability detection by next generation sequencing. Clin. Chem. 60, 1192–1199 (2014).
11. Shalin, S. C., Lyle, S., Calonje, E. & Lazar, A. J. F. Sebaceous neoplasia and the Muir-Torre syndrome: important connections with clinical implications. Histopathology 56, 133–147 (2010).
12. Cancer Genome Atlas Network. Genomic classification of cutaneous melanoma. Cell 161, 1681–1696 (2015).
13. Dees, N. D. et al. MuSiC: identifying mutational significance in cancer genomes. Genome Res. 22, 1589–1598 (2012).
14. Chu, A. Y. et al. Multiethnic genome-wide meta-analysis of ectopic fat depots identifies loci associated with adipocyte development and differentiation. Nat. Genet. 49, 125–130 (2017).
15. Lin, D.-C. et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. Nat. Genet. 46, 467–473 (2014).
16. Durinck, S. et al. Temporal dissection of tumorigenesis in primary cancers. Cancer Discov. 1, 137–143 (2011).
17. Wang, N. et al. Loss-of-function mutations in Notch receptors in cutaneous and lung squamous cell carcinoma. Proc. Natl Acad. Sci. USA doi: https://doi.org/10.1073/pnas.1114669108 (2011).
18. Chisatsazadeh, V. et al. Cross-species identification of genomic drivers of squamous cell carcinoma development across preneoplastic intermediates. Nat. Commun. 7, 12601 (2016).
19. Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N. Engl. J. Med. 371, 2189–2199 (2014).
20. Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 348, 124–128 (2015).
21. McGranahan, N. et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science 351, 1463–1469 (2016).
22. Joly, M.-O. et al. Somatic MMR gene mutations as a cause for MSI-H sebaceous neoplasms in Muir-Torre syndrome-like patients. Hum. Mutat. 36, 292–295 (2015).
23. Lefort, K. & Dotto, G. P. Notch signaling in the integrated control of keratinocyte growth/differentiation and tumor suppression. Semin. Cancer Biol. 14, 374–386 (2004).
24. Schulman, J. M. et al. Multiple hereditary infundibulocystic basal cell carcinoma syndrome associated with a germline SUFU mutation. JAMA Dermatol. 152, 323–327 (2016).
25. Sen, G. L. et al. ZNF750 is a p63 target gene that induces KLF4 to drive terminal epidermal differentiation. Dev. Cell. 22, 669–677 (2012).
26. Hazawa, M. et al. ZNF750 is a lineage-specific tumour suppressor in squamous cell carcinoma. Oncogene 36, 2243–2254 (2017).
27. Frei, J. V. Hereditary nonpolyposis colorectal cancer (Lynch syndrome II): diploid malignancies with prolonged survival. Cancer 69, 1108–1111 (1992).
28. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform. 12, 323 (2011).
29. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 18, 1853–1858 (2008).
30. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
31. Faust, G. G., Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinforma. Oxf. Engl. 30, 2503–2505 (2014).
32. Larson, D. E. et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics 28, 311–317 (2012).
33. Welch, J. S. et al. The origin and evolution of mutations in acute myeloid leukemia. Cell 150, 264–278 (2012).
34. Ellia, M. J. et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 486, 353–360 (2012).
35. Robinson, J. et al. The IPD and IMGT/HLA database: allele variant databases. Nucleic Acids Res. 43, D423–D431 (2015).
36. Lappalainen, I. et al. The European Genome-phenome Archive of human data consented for biomedical research. Nat. Genet. 47, 692–695 (2015).

**Acknowledgements**

We appreciate support received from the UCSF Department of Dermatology. D.A.S. is supported by the NIH Director’s Early Independence Award (DP5 OD021403) and the UCSF Physician-Scientist Scholar Program. We thank Sarah Pyle and Rachel Seye for constructing Figures 1, 2, and 4 using Adobe Illustrator CC.

**Author contributions**

J.P.N. and R.J.C. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: J.P.N. and R.J.C. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: J.P.N. and R.J.C. Acquisition, analysis, and interpretation of data: S.C.B., M.B., R.J.C., E. A.C., J.G., B.G., K.M., A.N., J.P.N., B.E.P.W., Z.S., D.A.S., M.S., C.J.V., and W.W. Drafting of the manuscript: J.P.N. and R.J.C. Critical revision of the manuscript for important intellectual content: J.P.N., D.A.S., C.I.V., and R.J.C. Statistical analysis: J.P.N. Obtained funding: R.J.C. Administrative, technical, or material support: S.C.B. and R.J.C. Study supervision: J.P.N., S.C.B., and R.J.C.

**Additional information**

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04008-y.

**Competing interests**: The authors declare no competing interests.

**Reprints and permission information** is available online at http://www.nature.com/reprintsandpermissions.

**Publisher’s note**: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2018