

1           The landscape of driver mutations in cutaneous squamous cell carcinoma  
2

3           Darwin Chang<sup>1,2</sup>, A. Hunter Shain<sup>1,2</sup>

4

5           <sup>1</sup> University of California San Francisco, Department of Dermatology

6           <sup>2</sup> University of California San Francisco, Helen Diller Family Comprehensive Cancer Center

7           Address correspondence to [Hunter.Shain@ucsf.edu](mailto:Hunter.Shain@ucsf.edu)

8

9

10 **Abstract**

11 Cutaneous squamous cell carcinoma is a form of skin cancer originating from keratinocytes in  
12 the skin. It is the second most common type of cancer and is responsible for an estimated 8000  
13 deaths per year in the United States. Compared to other cancer subtypes with similar  
14 incidences and death tolls, our understanding of the somatic mutations driving cutaneous  
15 squamous cell carcinoma is limited. The main challenge is that these tumors have high mutation  
16 burdens, primarily a consequence of UV-radiation-induced DNA damage from sunlight, making  
17 it difficult to distinguish driver mutations from passenger mutations. We overcame this challenge  
18 by performing a meta-analysis of publicly available sequencing data covering 105 tumors from  
19 10 different studies. Moreover, we eliminated tumors with issues, such as low neoplastic cell  
20 content, and from the tumors that passed quality control, we utilized multiple strategies to reveal  
21 genes under selection. In total, we nominated 30 cancer genes. Among the more novel genes,  
22 mutations frequently affected *EP300*, *PBRM1*, *USP28*, and *CHUK*. Collectively, mutations in the  
23 NOTCH and p53 pathways were ubiquitous, and to a lesser extent, mutations affected genes in  
24 the Hippo pathway, genes in the Ras/MAPK/PI3K pathway, genes critical for cell-cycle  
25 checkpoint control, and genes encoding chromatin remodeling factors. Taken together, our  
26 study provides a catalogue of driver genes in cutaneous squamous cell carcinoma, offering  
27 points of therapeutic intervention and insights into the biology of cutaneous squamous cell  
28 carcinoma.

29 **Introduction**

30 Over the past decade, large-scale DNA-sequencing studies have profiled a wide range of  
31 different cancers<sup>1,2</sup>. These studies have revealed candidate genes for targeted therapy and  
32 genetically distinct subtypes of cancer – information that has changed the way in which many  
33 cancers are treated. Moreover, at a basic science level, these studies have revealed  
34 fundamental insights into the biology of these cancers, often forming the basis of downstream  
35 hypothesis-driven work.

36 Given these achievements, there has been momentum to genomically profile the rarest of  
37 cancer subtypes<sup>1,2</sup>, yet cutaneous squamous cell carcinoma, the second most common form of  
38 cancer in the United States<sup>3</sup>, has largely been overlooked. Thirty-four cancer subtypes were  
39 included in The Cancer Genome Atlas program (TCGA) – a comprehensive effort to catalogue  
40 the driver genes in cancer – but regrettably, cutaneous squamous cell carcinoma was left out.  
41 Several individual laboratories have sequenced the exomes or genomes of cutaneous  
42 squamous cell carcinomas, examples of which are here<sup>4–13</sup>, but the small size of each study and  
43 difficulties in interpreting the high mutational loads in cutaneous squamous cell carcinomas  
44 have precluded the research community from settling upon a consensus set of driver genes.

45 One reason why large-scale sequencing consortiums have overlooked cutaneous squamous  
46 cell carcinomas is because cutaneous squamous cell carcinomas are thought of as non-life-  
47 threatening tumors, however, this reputation is misleading. Most cutaneous squamous cell  
48 carcinomas are caught at an early stage, reducing their mortality, but 8000 people per year still  
49 die from this disease in the United States<sup>3,14–16</sup>. To put this death toll in perspective, it is on par  
50 with that of melanoma<sup>17</sup>, for which nearly 1000 tumors have been sequenced, to date, at exome  
51 or genome resolution<sup>18</sup>.

52 A better understanding of the genetic drivers of cutaneous squamous cell carcinoma promises  
53 to improve treatment strategies. The current standard of care is for patients to receive immune-  
54 checkpoint blockade therapies, but roughly half do not respond and the responses are not  
55 always durable<sup>19</sup>. In addition, the risk of cutaneous squamous cell carcinoma is nearly 100-fold  
56 higher in immunosuppressed patients, such as organ transplant recipients, who are typically not  
57 eligible to receive immunotherapies<sup>20</sup>. Establishing the driver mutations in cutaneous squamous  
58 cell carcinoma promises to reveal new points for therapeutic intervention in this deadly tumor  
59 subtype. Towards this goal, we performed a meta-analysis of publicly available exome  
60 sequencing data from cutaneous squamous cell carcinomas.

## 61 Results

### 62 *Assembling a cohort of cutaneous squamous cell carcinomas*

63 We performed a literature search to identify whole-exome or whole-genome sequencing studies  
64 of cutaneous squamous cell carcinoma in which raw sequencing data was made publicly  
65 available. In total, we identified 105 tumors spanning 10 studies (Table 1, Table S1)<sup>4–12</sup>. We  
66 assessed the quality of sequencing data and removed 17 tumors from subsequent analyses  
67 (see methods for exclusion criteria). The remaining 88 tumors were retained, though we  
68 accounted for our ability to call somatic mutations in each tumor before comparing them.  
69

70 The main issues affecting our ability to detect somatic mutations in each tumor were: the  
71 neoplastic cell content, the mean sequencing coverage, and/or the variability in sequencing  
72 coverage. We bioinformatically quantified tumor cellularities, and they ranged from 12% to 99%.  
73 The mean sequencing coverages ranged from 12.4X to 498X. Finally, some tumors had high  
74 sequencing coverages, on average, but extreme variability in coverage, primarily linked to the  
75 GC content of their targets (see Fig. S1A for an example). To account for each of these  
76 potential issues, we used the Footprints software<sup>21</sup> to count the exact number of basepairs in  
77 each sample with sufficient sequencing coverage to make a mutation call. For the average  
78 sample, we could detect mutations at 91.2% of target bases, though this ranged from 52.3% to  
79 nearly 100% (Fig. S1B).

80 Establishing the extent to which we could detect mutations in each tumor allowed us to  
81 accurately calculate mutation burdens, irrespective of technical variables that are known to  
82 distort these measurements<sup>22</sup>. Moreover, the performance of cancer gene discovery tools  
83 deteriorates when there are large portions of the exome for which a mutation call cannot be  
84 made, and we were able to exclude problematic tumors from these analyses (Fig. S1B).  
85 Altogether, we improved the caliber of candidate cancer genes by aggregating a large cohort of  
86 tumors, reanalyzing the raw sequencing data, and applying rigorous quality control measures  
87 for sample inclusion.  
88

### 89 *Subtypes of cutaneous squamous cell carcinoma*

91 We calculated the mutation burden and the proportion of each tumor's mutations that were  
92 attributable to established mutational signatures (Fig. 1), revealing five distinct subtypes of  
93 cutaneous squamous cell carcinoma among the tumors analyzed in this study.

94 First, two cutaneous squamous cell carcinomas came from patients with xeroderma  
95 pigmentosum – a rare hereditary disorder characterized by extreme sensitivity to UV radiation  
96 and caused by germline mutations in genes involved in nucleotide excision repair<sup>23</sup>. As  
97 expected, these two tumors had high mutation burdens with a high frequency of cytosine to  
98 thymine transitions at the 3' basepairs of consecutive pyrimidines (the classic mutation that  
99 arises from UV radiation<sup>24</sup>). Interestingly, they did not have a high proportion of “signature 7”  
100 mutations (a mutational signature extracted from pan-cancer analyses and attributed to UV  
101 radiation<sup>25,26</sup>). The absence of signature 7 was due to differences in the trinucleotide contexts of  
102 mutations arising in wild-type versus mutant-XPC tumors (Fig. S2) and illustrates how the  
103 repertoire of known mutational signatures remains incomplete.

104 Second, there were cutaneous squamous cell carcinomas that arose sporadically in patients  
105 with no known comorbidities. These cutaneous squamous cell carcinomas had high mutation  
106 burdens, and the majority of mutations were attributable to UV radiation (Fig. 1).

107 Furthermore, there were two distinct types of cutaneous squamous cell carcinomas arising from  
108 immunosuppressed patients, who were primarily organ-transplant recipients. As previously  
109 reported<sup>7</sup>, cutaneous squamous cell carcinomas from patients treated with Azathioprine (as a  
110 means to prevent transplant rejection) had high mutation burdens with high proportions of  
111 signature 32 (Fig. 1). Azathioprine increases the risk of cutaneous squamous cell carcinoma  
112 beyond the risk conferred by other immunosuppressive agents<sup>27</sup> because Azathioprine is both  
113 immunosuppressive and a potent mutagen<sup>28,29</sup>. Patients on other immunosuppressive drug  
114 regimens had comparably lower mutation burdens, primarily attributable to UV radiation (Fig. 1).

115 Finally, there were cutaneous squamous cell carcinomas from patients with recessive  
116 dystrophic epidermolysis bullosa (RDEB), a rare hereditary disorder characterized by chronic  
117 blistering in the skin and caused by germline mutations in collagen VII (COL7A1). As previously  
118 noted<sup>5</sup>, these tumors had relatively low mutation burdens, primarily from APOBEC-mediated  
119 mutagenesis (Fig. 1).

#### 120 *Nomination of driver mutations in cutaneous squamous cell carcinoma*

121 Genes under positive selection in cancer are distinguished by having significantly more  
122 mutations than the background mutation rate at that locus would predict<sup>30</sup>. We utilized four  
123 cancer gene discovery tools to reveal such genes: MutSig<sup>31</sup>, dN/dS<sup>32</sup>, LOFsigrank<sup>33</sup>, and  
124 OncodriveFML<sup>34</sup>. Collectively, these tools nominated 12 genes total, including a subset of 7  
125 genes by at least two tools (Fig. 2, Table S2).

126 While these tools nominated a relatively small number of genes, the number of genes  
127 nominated is in line with our statistical power to detect cancer genes in this study. It is  
128 challenging to identify genes under selection in cutaneous squamous cell carcinoma because of  
129 the high background of passenger mutations. Lawrence and colleagues performed a power  
130 analysis to determine the detection limit of cancer genes given the mutation burden of the tumor  
131 subtype and the number of tumors analyzed<sup>35</sup>. Based on the parameters of our study, we were  
132 powered to detect genes under selection if their mutation frequencies were approximately 15%  
133 or greater.

134

141 To nominate driver mutations that were too infrequent to show evidence of positive selection in  
142 this dataset alone, we searched for overlapping mutations in the cancerhotspot.org database  
143 (Fig. 2B). This database contains mutations identified from pan-cancer analyses that cluster  
144 within genes<sup>36</sup> – a common pattern for gain-of-function mutations. We reasoned that if a  
145 mutation shows evidence of positive selection from pan-cancer analyses, and the exact same  
146 mutation was present in our dataset, then it deserves consideration as a driver of cutaneous  
147 squamous cell carcinoma.

148  
149 Finally, we nominated genes with focal copy number alterations. Focal, homozygous deletions  
150 affected the *CDKN2A* and *PTEN* tumor suppressor genes (Fig. S3). A focal, heterozygous  
151 deletion affected *AJUBA* in one tumor, and in the same tumor there was a point mutation  
152 affecting the other allele (Fig. S3B). Focal amplifications affected: *CCND1*, *MDM2*, *YAP1*, and  
153 *RAP1B* (Fig. S4). *RAP1B* is a ras-related-protein, and the tumor with amplification of *RAP1B*  
154 also had a point mutation affecting the amplified allele. This point mutation was analogous to  
155 mutations known to activate other Ras genes (Fig. S4 E,F).

156  
157 *Removal of false positive driver mutations*  
158 The earliest generation of algorithms to discover cancer genes assumed a background mutation  
159 rate that is uniform across the genome (an assumption that is not true<sup>37</sup>), resulting in the  
160 nomination poor quality candidates in high mutation burden cancers<sup>38,39</sup>. The algorithms, utilized  
161 here, have improved background mutation rate models, but the determinants of the mutation  
162 rate across the genome are complex and remain incompletely understood<sup>37</sup>, leaving open the  
163 possibility of false positives. We curated the nominated genes, as described below, to root out  
164 unlikely cancer genes, though we acknowledge that these are ultimately judgement calls.

165  
166 We concluded that *COL11A1* (Fig. 2A) requires additional evidence to be considered a driver  
167 gene. *COL11A1* has a borderline significant q-value (Table S2); it is poorly expressed in the  
168 keratinocyte lineage (Fig. S5); and it has no known role in keratinocyte biology. We also  
169 determined that *DIS3*, *HIST1H3C*, *KDR*, and *MAP2K4* (Fig. 2B) need additional evidence to be  
170 considered driver genes. The hotspot mutations affecting these genes had q-values that were  
171 low in comparison to others in the cancerhotspot.org database, and OncoKb, a precision  
172 oncology consortium at Memorial Sloan Kettering Cancer Center<sup>40</sup>, classifies these hotspot  
173 mutations as unlikely to be oncogenic.

174  
175 Finally, we do not believe there is sufficient evidence to classify the hotspot mutation affecting  
176 *KNSTRN* as oncogenic. The mutation is annotated as coding, but this appears to be based on  
177 an erroneous gene model. From RNA-sequencing data of normal skin and cutaneous  
178 squamous cell carcinoma, expression of *KNSTRN* begins downstream of the mutation site (Fig.  
179 S5B), implying that the mutation affects the promoter of *KNSTRN*. Promoter mutations are  
180 ubiquitous in sun-exposed cancers<sup>41,42</sup> because transcription factors at the promoter can bend  
181 DNA in ways that render their binding elements vulnerable to mutagenesis by UV radiation<sup>43</sup>.  
182 Regrettably, these types of annotation errors are not uncommon – many hotspot mutations in  
183 melanoma, which were initially thought to be coding mutations, were subsequently revealed to  
184 be promoter mutations after further studies<sup>18,33,44</sup>. While there is a study suggesting *KNSTRN*<sup>S24F</sup>  
185 is oncogenic<sup>45</sup>, the supporting evidence presumes the mutation is coding.

186  
187 *Novel candidates in cutaneous squamous cell carcinoma*  
188  
189 We compared the genes nominated in our meta-analysis to those from eight other studies that  
190 have proposed cancer genes in cutaneous squamous cell carcinoma<sup>5–10,46,47</sup> (Fig. 3). *TP53*,  
191 *NOTCH1*, *NOTCH2*, *CDKN2A*, and *HRAS* were proposed by a majority of the other studies and

192 also nominated by us. Moreover, *FAT1*, *ARID2*, *CASP8*, *CREBBP*, *AJUBA*, *PTEN*, *PIK3CA*,  
193 *RAC1*, *EZH2*, *KRAS*, *NFE2L2*, and *MTOR* were nominated in 1-3 studies each as well as by us,  
194 lending credibility to their pathogenic roles in cutaneous squamous cell carcinoma.

195  
196 We nominated 12 genes that were not implicated in the other studies. Many of these genes  
197 harbored hotspot mutations that occurred relatively infrequently (see Fig. 2B for the full list),  
198 likely explaining why they were not noted in other analyses. However, four genes were mutated  
199 in greater than 10% of tumors: *EP300*, *PBRM1*, *USP28*, and *CHUK*.

200  
201 *EP300* (p300) encodes a histone acetyltransferase that is a critical transcriptional co-activator of  
202 NOTCH<sup>48</sup>. *EP300* had frequent loss-of-function mutations, including missense mutations that  
203 clustered in the histone acetyltransferase domain (Fig. 3B). Several of these missense  
204 mutations have been functionally confirmed to eliminate histone acetyltransferase activity of the  
205 protein<sup>49</sup>. *EP300* has also been implicated as a tumor suppressor gene in esophageal  
206 squamous cell carcinoma<sup>50</sup>.

207  
208 *PBRM1* encodes a subunit of the SWI/SNF chromatin remodeling complex and has been  
209 implicated as a tumor suppressor gene in a wide range of other cancers<sup>51</sup>. *PBRM1* had  
210 deleterious mutations occurring throughout the length of the protein (Fig. 3C). Of note, another  
211 member of the SWI/SNF chromatin remodeling complex, *ARID2*, was also implicated as a tumor  
212 suppressor gene in cutaneous squamous cell carcinoma.

213  
214 *USP28* encodes a deubiquitinase that stabilizes key proteins involved in DNA repair<sup>52</sup>. It is  
215 required for DNA-damage-induced apoptosis mediated through the Chk2-p53-PUMA pathway<sup>52</sup>.  
216 *USP28* was nominated here because of its high frequency of truncating mutations (Fig. 3D).

217  
218 *CHUK* encodes a protein, also known as IκB Kinase α (IKKα), that is involved in the NFκB  
219 signaling pathway. *CHUK* knockout mice are born with thickened skin, and their cutaneous  
220 keratinocytes are unable to differentiate, resulting in death shortly after birth<sup>53</sup>. An identical  
221 phenotype has been observed in humans with a defective *CHUK* gene<sup>54</sup>. Collectively, these  
222 studies implicate *CHUK* as a key factor governing growth and differentiation of keratinocytes in  
223 skin. In cutaneous squamous cell carcinoma, *CHUK* had a high frequency of truncating  
224 mutations, and somatic alterations affected both alleles in most tumors (Fig. 3E).

225  
226 We also checked for genes nominated in other studies but not by our analyses. *KMT2D* was the  
227 only gene implicated in more than one of the other studies interrogated here (Fig. S6A). The  
228 majority of *KMT2D* mutations were silent or missense mutations predicted to be benign (Fig.  
229 S6B). Even under relaxed thresholds of significance (Table S2), *KMT2D* would not have been  
230 nominated here, though future studies may reveal more compelling evidence of selection.

231  
232 *Recurrent pathways disrupted in cutaneous squamous cell carcinoma*

233  
234 The individual genes, nominated here, encode proteins that participate in a core set of signaling  
235 pathways perturbed in cutaneous squamous cell carcinoma (Fig. 4). Mutations in genes  
236 encoding proteins involved in the NOTCH and p53 pathways were ubiquitous in cutaneous  
237 squamous cell carcinomas. The NOTCH pathway had loss-of-function mutations occurring in  
238 80% of tumors, and the p53 pathway had loss-of-function mutations occurring in 71% of tumors.  
239 Mutations in these pathways appear to be defining features of cutaneous squamous cell  
240 carcinoma.

241

242 Other pathways were recurrently disrupted, albeit to a lesser extent. Mutations that disrupt cell-  
243 cycle-checkpoint control occurred in 39% of tumors, primarily affecting the *CDKN2A* gene.  
244 Mutations that disrupt the SWI/SNF chromatin remodeling complex occurred in 38% of tumors.  
245 Mutations that activate the Hippo pathway occurred in 37% of tumors. We broadly grouped  
246 together *CASP8*, *CHUK*, and *NFE2L2*, which were collectively mutated in 33% of tumors. These  
247 genes mediate cellular responses to stress, such as inflammation and oxidative stress. More  
248 work will be needed to determine whether and how these genes are related. Finally, mutations  
249 that activate the Mitogen Activated Protein Kinase (MAPK) and/or Phospholnositide 3-Kinase  
250 (PI3K) pathways occurred in 31% of tumors.

251  
252 We next interrogated whether mutations affecting specific genes, pathways, or tumor subtypes  
253 overlapped more or less than would be expected by chance. Cutaneous squamous cell  
254 carcinomas from patients with recessive dystrophic epidermolysis bullosa (RDEB) were  
255 enriched with mutations affecting *CASP8* (Fig. 4A, Table S3). *CASP8* mediates cellular  
256 apoptosis in response to inflammatory cytokines<sup>55,56</sup>. Skin from patients with RDEB is  
257 chronically blistering and inflamed, likely explaining the selective pressure to accumulate  
258 *CASP8* mutations in this subtype of cutaneous squamous cell carcinoma. Other comparisons  
259 did not reach statistical significance after accounting for multiple hypothesis testing (see Table  
260 S3 for a complete list of comparisons).

261  
262 Most of the studies in this meta-analysis were exome-sequencing studies, prohibiting us from  
263 analyzing mutations in non-coding portions of the genome. *TERT* promoter mutations are  
264 common in many cancer subtypes, and while we were unable to investigate the locus, other  
265 studies have reported a high frequency of *TERT* promoter mutations in cutaneous squamous  
266 cell carcinoma<sup>57</sup>, prompting us to include *TERT* among our final list of cancer genes (Fig. 4B).  
267 Future studies will be needed to more systematically interrogate the role of non-coding  
268 mutations in cutaneous squamous cell carcinoma.

269  
270 **Discussion**

271  
272 In this meta-analysis of exome sequencing data, we analyzed the largest cohort of cutaneous  
273 squamous cell carcinomas to date, upheld rigorous quality control standards for sample  
274 inclusion, and utilized state-of-the-art cancer gene discovery algorithms to nominate cancer  
275 genes. In total, we nominated 30 cancer genes, known to operate in a core set of signaling  
276 pathways, that were perturbed in cutaneous squamous cell carcinoma. Our study suggests new  
277 cancer genes and helps clarify which candidates from previous studies are likely *bona fide*  
278 driver genes in cutaneous squamous cell carcinoma.

279  
280 Future work is still needed to understand the driver genes in cutaneous squamous cell  
281 carcinoma. Cancer gene discovery studies have likely reached a saturation point for many  
282 cancers, but this is not the case for cutaneous squamous cell carcinoma. Despite the size of our  
283 meta-analysis, we could only detect cancer genes with mutations in 15% or more of tumors. We  
284 overcame this limitation, in part, by identifying genes with well-characterized hotspot mutations  
285 and/or focal copy number alterations; however, there are likely many cancer genes in cutaneous  
286 squamous cell carcinoma that have yet to be discovered. Our study also focused on the exome,  
287 prohibiting us from identifying driver mutations in non-coding regions of the genome or from  
288 identifying structural variants and viral integrations that may play a pathogenic role.

289  
290 Taken together, our study provides the most detailed catalogue of driver genes in cutaneous  
291 squamous cell carcinoma to date, offers points of therapeutic vulnerability, and reveals critical  
292 insights into the basic biology of cutaneous squamous cell carcinoma.

293

## 294 Methods

295

### 296 *Selection of studies*

297 We performed a literature search to identify whole-exome or whole-genome sequencing studies  
 298 of cutaneous squamous cell carcinoma that made their raw sequencing data publicly available  
 299 as of September 1<sup>st</sup>, 2020. The studies meeting this inclusion criteria are summarized in Table  
 300 1. The number of samples shown in Table 1 may not match the reported numbers in each study  
 301 because some studies re-analyzed previously published data, or we were unable to retrieve the  
 302 entirety of the raw sequencing dataset.

303

### 304 *Removal of 17 samples*

305 We assessed the quality of sequencing data and removed 17 tumors from all analyses. Thirteen  
 306 of these tumors had few, if any, discernible point mutations, and among the point mutations  
 307 detected, their mutant allele frequencies were close to our detection limit. These patterns  
 308 suggest poor sampling of the neoplastic cells. Two cases had less than five-fold coverage in the  
 309 reference tissues, making it difficult to confidently distinguish somatic mutations from germline  
 310 single nucleotide polymorphisms (SNPs). One tumor and reference pair were not properly  
 311 matched, which was evident from their patterns of germline SNPs. Finally, one reference tissue  
 312 had high levels of tumor contamination, prohibiting us from sensitively detecting somatic  
 313 mutations.

314

### 315 *Calling somatic point mutations*

316 We collected either fastq or bam files from each study. Fastq files underwent quality checks  
 317 using FastQC and were subsequently aligned to the hg19 reference genome using the BWA-  
 318 MEM algorithm (v0.7.13). These were further groomed and deduplicated using Genome  
 319 Analysis Toolkit (v4.1.2.0) and Picard (v4.1.2.0).

320

321 Somatic point mutations were called using Mutect2 (v4.1.2.0) by comparing each tumor bam to  
 322 a corresponding reference bam, thus producing an initial set of candidate somatic mutations.  
 323 The variants were annotated using Funcotator (v4.1.2.0) and further filtered to remove  
 324 suspected sequencing artifacts, extremely subclonal mutations, and/or mutations from unrelated  
 325 clones of keratinocytes. In parallel, indels were called using Pindel (v0.2.5) and further filtered.  
 326 Our filtering scripts are available here: <https://github.com/darwinchangz/ShainMutectFilter>.

327

328 To provide an overview, the script uses samtools mpileup to count the number of reference and  
 329 mutant reads for each variant. Variants with low overall coverage were removed, and variants  
 330 with few supporting mutant reads were also removed. Finally, we calculated tumor cellularity in  
 331 each sample, and removed variants that were not predicted to be in at least 40% of tumor cells.  
 332 The main reason we removed these variants is because it was difficult to distinguish whether  
 333 they were from subclones within the tumor or from unrelated clones of mutant keratinocytes.  
 334 Normal skin is comprised of clones of keratinocytes, many of which harbor pathogenic  
 335 mutations<sup>6,58</sup>. We have observed that these clones commingle with adjacent skin tumors and  
 336 are often unintentionally included in microdissections<sup>59</sup>.

337

### 338 *Calling Heterozygous SNPs*

339 We also identified a high-confidence set of germline heterozygous SNPs from the reference  
 340 bams corresponding to each patient. Knowing these SNPs allowed us to measure allelic  
 341 imbalance, thereby revealing tumor cellularity (detailed below) and corroborating copy number  
 342 alterations within tumors. To identify heterozygous, germline SNPs, we called variants in the  
 343 reference tissue as compared to the reference genome using FreeBayes (v1.3.1). Next, we

344 filtered these variants to include only those that overlapped known 1000 genomes sites and  
345 which had 40-60% variant allele frequency (VAF).

346

### 347 *Inferring Tumor Cellularity*

348 We used multiple methods, if possible, to infer the neoplastic cell content from each tumor. The  
349 methods used for each tumor are listed in Table S1 and further described below.

350

351 “Allelic Imbalance of SNPs over Deletions”: We calculated tumor cellularity from the degree of  
352 allelic imbalance of heterozygous, germline SNPs over chromosomal arms with deletions in the  
353 tumor. This strategy assumes the deletions are fully clonal and there remains only one copy of  
354 the remaining chromosome in each tumor cell. A deletion results in a complete loss of an allele  
355 within the tumor cells. As a result, sequencing reads from the deleted allele are assumed to  
356 come from non-neoplastic cells. Tumor cellularity can therefore be calculated from ratio of reads  
357 mapping to the A and B alleles as described<sup>59</sup>.

358

359 “Allelic Imbalance of SNPs over Copy Number Neutral LOH”: Similar to the above strategy, we  
360 calculated tumor cellularity from the degree of allelic imbalance of heterozygous, germline SNPs  
361 over chromosomal arms with copy-number-neutral loss-of-heterozygosity (LOH). This strategy  
362 assumes that copy number neutral LOH is fully clonal and there are two copies of the remaining  
363 allele in each tumor cell. Copy number neutral LOH results in complete loss of an allele within  
364 the tumor cells. As a result, sequencing reads mapping to the lost allele are assumed to come  
365 from the non-neoplastic cells. Tumor cellularity can therefore be calculated from ratio of reads  
366 mapping to the A and B alleles as described<sup>59</sup>.

367

368 “Modal Somatic MAF”: In addition to investigating the variant allele frequencies of heterozygous,  
369 germline SNPs, we also used the mutant allele frequencies (MAFs) of somatic mutations. The  
370 mutant allele frequency of a somatic mutation that is fully-clonal and heterozygous should be  
371 50%, but will decrease with stromal contamination. For each tumor, we plotted a histogram of  
372 mutant allele frequencies and determined the “peak” or “modal” mutant allele frequency, and we  
373 doubled these values to infer tumor cellularity.

374

375 “Median Somatic MAF”: For a small number of tumors, the density of somatic mutations was  
376 insufficient to produce a smooth histogram. In these cases, we determined the median MAF  
377 from all somatic mutations, and we doubled this value to infer tumor cellularity. If the patient was  
378 male, we separately calculated the median MAF of somatic mutations on the sex chromosomes  
379 and incorporated these values without doubling.

380

### 381 *Determining statistical power to call somatic mutations (related to figure S1)*

382 To identify a somatic mutation, there must be sufficient coverage in both the reference and the  
383 tumor. Therefore, for each tumor/reference pair, we calculated the footprint for which  
384 sequencing coverage was sufficient to call somatic mutations.

385

386 In the reference, sufficient coverage is necessary to detect both alleles, thus ensuring that a  
387 variant in the tumor is a somatic mutation and not a germline SNP. We required at least 6-fold  
388 coverage in the reference to call a somatic mutation. Assuming each allele is randomly sampled  
389 during sequencing, the probability of both alleles being sampled at least once with 6-fold  
390 coverage is 96.9% (two-tailed binomial test). We used the Footprints software<sup>21</sup> to calculate the  
391 precise number of base-pairs that achieved 6-fold coverage (or greater) in each reference bam  
392 and designated this value as the “call-able” footprint for each reference bam (Table S1).

393

394 In the tumor, there needs to be sufficient coverage to detect the mutant allele. We required our  
395 somatic mutation calls to have at least 4 unique reads. Some mutation callers, including  
396 MuTect2, which was used in this study, will attempt to call mutations with fewer reads, but in  
397 practice, we found those calls to be of poor quality and filtered them out. We considered a site  
398 to be “call-able” if it had 8-fold effective tumor coverage. “Effective tumor coverage” refers to the  
399 sequencing coverage derived from the tumor after discounting the proportion of reads from non-  
400 tumor cells. For example, if a tumor sample has 100-fold total coverage and 8% tumor  
401 cellularity, then the effective tumor coverage would be 8-fold. Assuming that the alleles are  
402 randomly sampled during sequencing, their relative coverages will fit a binomial distribution and  
403 8-fold effective tumor coverage is sufficient to call a heterozygous somatic mutation 50% of the  
404 time (two-tailed binomial test). We used the Footprints software<sup>21</sup> to calculate the precise  
405 number of base-pairs that achieved 8-fold effective tumor coverage or greater in each tumor  
406 bam and designated this value as the “call-able” footprint for each tumor bam (Table S1).

407  
408 For each tumor/reference pair, we took the minimum “call-able” footprint between the tumor and  
409 the reference and designated that value to be the “call-able” footprint for that sample. We  
410 subsequently divided the “call-able” footprint by the bait territory that was targeted to indicate  
411 the fraction of target basepairs for which we were statistically powered to recognize mutations.  
412 These numbers are reported in figure S1A.

413  
414 There were primarily three variables that reduced statistical power to recognize mutations: 1.  
415 Low overall coverage, 2. Low tumor cellularity, 3. Extreme variability in coverage (e.g. from GC-  
416 selection biases introduced during hybridization).

417  
418 *Calling Copy Number Alterations*

419 Copy number alterations were inferred from the DNA- sequencing data using CNVkit<sup>60</sup>.  
420 CNVkit can be run in reference or reference-free mode. We elected to run CNVkit in reference  
421 mode using the panel of normals from each study. This approach consistently produced the  
422 least noisy copy number profiles, as compared to reference-free mode or a universal reference.  
423 All other parameters were run on their default settings.

424  
425 *Calculating tumor mutation burden and inferring mutational signature (related to figure 1)*

426 When calling somatic point mutations, we only considered mutations that were estimated to be  
427 in at least 40% of tumor cells. This was helpful in comparing the mutation burdens from tumors  
428 across the different studies for which there was considerable variability in sequencing  
429 coverages. High-sequencing coverage permits the detection of subclonal mutations, which  
430 would artificially inflate the mutation burden of a tumor, compared to another with lower  
431 coverage, if subclonal mutations are counted. There were also differences in our ability to detect  
432 clonal mutations in each tumor (described in more detail in the “*Determining statistical power to*  
433 *call somatic mutations*” section above). To address this issue, we divided the number of clonal  
434 mutations in each tumor by the footprint which we were statistically powered to detect mutations  
435 in each tumor.

436  
437 To perform mutational signature analysis, surrounding genomic contexts were applied to single  
438 nucleotide variants identified in each clone using the Biostrings hg19 human genome sequence  
439 package (BSgenome.Hsapiens.UCSC.hg19 v1.4.0). Variant contexts were used to assess the  
440 proportion of each clone’s mutational landscape that could be attributed to a mutagenic process  
441 using the deconstructSigs R package (v1.9.0). A set of 48 signatures recently described<sup>26</sup> were  
442 analysed. The results of these analyses are shown in the “Signature Proportion” stacked barplot  
443 of figure 1. In parallel, we performed a simpler analysis of dinucleotide contexts to identify  
444 cytosine to thymine transitions at the 3’ basepair of dipyrimidines or cytosine-cytosine to

445 thymine-thymine mutations (see the “UV” column of Table S4) – these are the classic mutation  
446 types attributed to UV radiation, and the results are shown in the “Fraction UV Mutations”  
447 stacked barplot of figure 1.

448

#### 449 *Nomination of driver genes*

450 We used four cancer gene discovery programs to nominate cancer genes: MutSig<sup>31</sup>,  
451 LOFsigrank<sup>33</sup>, dN/dS<sup>32</sup>, and OncodriveFML<sup>34</sup>. dN/dS was run in covariate value mode, which  
452 combines the synonymous mutations in a gene with its epigenomic covariates to determine the  
453 background mutation rate. The other programs were run on their default settings. All genes with  
454 q-values of less than 0.2 are shown in Table S2. For the purposes of this study, a gene was  
455 considered significant if its q-value was less than 0.05 (Figure 2).

456

457 Next, we cross-referenced the cancerhotspot.org database to identify mutations found in our  
458 study. We included mutations with a q-value of less than 0.01, but relaxed this threshold for all  
459 secondary hotspots within a gene. For example, EP300 had two hotspot mutations with q-  
460 values of 6.9E-22 and 4.8E-07, but we also show additional hotspots in the vicinity, one with a  
461 q-value of 1.9E-02 and one that was a predicted hotspot using the 3D hotspot algorithm.

462

463 Finally, driver genes were curated as described in the main text to root out potential false  
464 positives.

#### 465 *RNA-Sequencing analysis (related to figure S5)*

466 Two of the studies analyzed in this meta-analysis had RNA-sequencing data available<sup>5,6</sup>. These  
467 datasets covered both normal skin (n=17 samples) and cutaneous squamous cell carcinoma  
468 (n=17 samples). We downloaded the raw sequencing data, aligned with STAR, and quantified  
469 gene expression with RSEM, as previously described<sup>61</sup>. In figure S5, we show the Fragment Per  
470 Kilobase of transcript per Million reads (FPKM) values for candidate genes (see figure 2 for a list  
471 of candidates). FPKM values normalize for gene length and read depth, allowing the  
472 comparison of gene expression levels across genes. In figure S5, we combined RNA-  
473 sequencing reads from all 34 samples over the *KNSTRN* gene, demonstrating that the mutant  
474 hotspot is not expressed.

475

#### 476 *Mutational overlap analysis (related to Table S3)*

477 We interrogated whether mutations affecting specific genes, pathways, or tumor subtypes  
478 overlapped more or less than would be expected by chance. We restricted our analyses to  
479 genes, pathways, or tumor subtypes with at least 16 mutant tumors – the minimum number that  
480 could reach statistical significance with our sample size. P-values for individual comparisons  
481 were calculated using the Fisher’s exact test. We corrected for multiple hypothesis testing by  
482 computing false discovery rates (q-values) using the Benjamini-Hochberg procedure. A full list  
483 of p- and q- values for each comparison is shown in Table S3.

484

#### 485 **Figure and Table legends**

486

487 **Figure 1. Subtypes of cutaneous squamous cell carcinoma.** Each vertical bar corresponds  
488 to a single tumor. Top track: The mutation burden (mutations per megabase) of each tumor.  
489 Middle track: The fraction of mutations within each tumor matching the canonical UV radiation  
490 signature. Bottom track: The fraction of mutations attributable to established mutational  
491 signatures within each tumor. The proposed etiology of select signatures is indicated. XP =  
492 tumors from patients with xeroderma pigmentosum. Sporadic = tumors from patients with no  
493 known comorbidities. Immunosuppressed: tumors from immunosuppressed patients -- this

494 group is further stratified by the usage or absence of azathioprine as an immunosuppressive  
495 drug. RDEB = tumors from patients with recessive dystrophic epidermolysis bullosa.

496  
497 **Figure 2. Nomination of cancer genes in cutaneous squamous cell carcinoma. A.** A Venn  
498 diagram depicting nominated cancer genes from four separate cancer gene discovery  
499 programs, each designed to identify genes under positive selection in cancer. The set of  
500 candidate genes were further curated, as described, to nominate candidates for which additional  
501 evidence is warranted (red text) or not (blue text). **B.** A list of mutations in our study that overlap  
502 mutations in the cancerhotspot.org database. Mutations are grouped by gene and ordered by  
503 their q-values (lowest to highest). These mutations were also curated, as described, to nominate  
504 candidates for which additional evidence is warranted (Possible False Positive) or not (Likely  
505 Driver).

506  
507 **Figure 3. Candidate cancer genes in cutaneous squamous cell carcinoma. A.** The genes  
508 nominated in our meta-analysis are stratified by their mutation frequency (x-axis) and how often  
509 they were nominated in 8 previous studies (y-axis) that catalogued drivers of cutaneous  
510 squamous cell carcinoma. EP300, PBRM1, USP28, and CHUK were mutated in greater than  
511 10% of tumors but not implicated by other studies. Lollipop diagrams portray the spectrum of  
512 mutations in each of these four genes in panels **B-E**.

513  
514 **Figure 4. The landscape of driver mutations in cutaneous squamous cell carcinoma. A.** Tiling plot of the genetic alterations (rows) in each tumor (columns). Genes and tumors are  
515 further organized into pathways and clinical subtypes. The percentages of samples harboring  
516 pathogenic alterations are indicated. Mut, mutation; Amp, amplification; XPC, xeroderma  
517 pigmentosum; RDEB, recessive dystrophic epidermolysis bullosa. **B.** Pathways affected. TERT  
518 was not implicated in this study, which focused on coding mutations, but is included here  
519 because it is known to have promoter mutations, as described.

520  
521 **Table 1. Summary of exome or genome sequencing studies analyzed in this meta-**  
522 **analysis.** The number of tumors, listed here, corresponds to unique tumors, whose data was  
523 made publicly available, and thus may not match the reported size from the original studies.

524  
525 **Figure S1. Ability to detect mutations in each tumor varies substantially. A.** Some tumors  
526 had high overall coverage but extreme variability in coverage, primarily linked to GC content,  
527 thus diminishing our ability to detect mutations within large portions of the bait territory. An  
528 example of a tumor in the top 10% of coverage variability (left panel) and a tumor in the bottom  
529 10% of coverage variability (right panel). Each datapoint corresponds to a bait interval, stratified  
530 by its GC content and sample-normalized coverage. **B.** The percentage of base pairs within the  
531 target bait territory with sufficient coverage to call a mutation. Tumors with low overall  
532 sequencing coverage, extreme variability in sequencing coverage, and/or low tumor cellularity  
533 tended to have a lower proportion of “callable” mutations. Tumors with greater than 75% “call-  
534 able” mutations were included in all analyses, whereas those with fewer were only included in  
535 certain analyses, as described.

536  
537 **Figure S2. Somatic mutations in Xeroderma Pigmentosum tumors occur in a distinct tri-**  
538 **nucleotide context.** The tri-nucleotide context of mutations in two XPC-/ tumors as compared  
539 to signature 7. C>T transitions at dipyrimidines were common in both, however the basepair  
540 downstream (three prime) to the mutation site differed with thymines more common in XPC-/–  
541 tumors and cytosines more common in signature 7.

542  
543

544 **Figure S3. Focal deletions affecting CDKN2A, AJUBA, and PTEN.** Copy number over  
545 individual genomic positions (data points) with copy number segments (yellow lines) overlaid.  
546 Colored data points correspond to probes covering CDKN2A (panel A), AJUBA (panel B), and  
547 PTEN (panel C).

548  
549 **Figure S4. Focal amplification affecting CCCND1, YAP1, MDM2, and/or RAP1B in six**  
550 **tumors.** Copy number over individual genomic positions (data points) with copy number  
551 segments (yellow lines) overlaid. Colored data points correspond to probes covering CCND1  
552 (panels A and B), YAP1 (panel B), MDM2 (panels C and D), and RAP1B (panel D). Note that  
553 amplifications converge upon the noted genes across these six tumors. The tumor with  
554 amplification of RAP1B (shown in panel D) also had a point mutation affecting the 12th codon of  
555 RAP1B (panel E). The mutant allele frequency of this point mutation was higher than all other  
556 mutations in that tumor, consistent with the mutation affecting the amplified allele. RAP1B is a  
557 ras-like protein with recurrent mutations affecting codons 12 and 61 across cancer (panel F),  
558 analogous to mutations known to activate other ras-like proteins.

559  
560 **Figure S5. Gene expression of candidate genes. A.** The Fragments Per Kilobase of  
561 transcript per Million reads (FPKM) values -- a measurement that normalizes for both gene  
562 length and sequencing depth in RNA-sequencing data -- of candidate genes. RNA-sequencing  
563 data derived from both normal skin samples (n=17 samples) and cutaneous squamous cell  
564 carcinomas (n=17 tumors). This data was used to flag potential false positive candidates based  
565 on their poor expression in normal and neoplastic keratinocytes. **B.** Transcription of KNSTRN  
566 begins downstream of the hotspot mutation site. RNA-sequencing data was aggregated from  
567 normal skin (n=17) and cutaneous squamous cell carcinoma (n=17). Top panel -- RNA-  
568 sequencing coverage over the KNSTRN gene. Coverage was approximately 2,000 to 3,000-fold  
569 over most exons with the exception of exon1. Bottom panel -- A zoomed inset of RNA-  
570 sequencing coverage over exon1. While sequencing coverage surpassed 1,000-fold at the  
571 three-prime junction of exon1, there were no reads spanning the hotspot mutation site from the  
572 any of the 34 samples, in-line with our prediction that the mutation is non-coding.

573  
574 **Figure S6. KMT2D has a benign spectrum of mutations. A.** Genes nominated by other  
575 studies, but not our study, are stratified by their mutation frequency (x-axis) and how often they  
576 were nominated in 8 previous studies (y-axis) that catalogued drivers of cutaneous squamous  
577 cell carcinoma. KMT2D was the only gene recurrently implicated in other studies but not ours.  
578 **B.** Lollipop diagram portrays the spectrum of mutations. KMT2D was not nominated by cancer  
579 gene discovery algorithms because of its high frequency of silent mutations and missense  
580 mutations predicted to be benign.

581  
582 **Table S1. A summary of samples included in this study and their sequencing metrics.** We  
583 identified 105 cutaneous squamous cell carcinomas from 101 patients (4 patients had multiple  
584 tumors) for which sequencing data was publicly available. This table summarizes available  
585 information pertaining to these patients and their tumors as well as relevant sequencing metrics.  
586 Note that some tumors had whole genome sequencing, but we restricted our analyses to the  
587 exome. Methods to infer tumor cellularity are explained in more detail under *Inferring Tumor*  
588 *Cellularity* in Methods. RDEB = Recessive Dystrophic Epidermolysis Bullosa. XPC = Xeroderma  
589 Pigmentosum.

590  
591 **Table S2. Genes nominated to be under positive selection in cutaneous squamous cell**  
592 **carcinoma.** All genes nominated by each of the four cancer gene discovery tools with a q-value  
593 below 0.2 are shown. For the purposes of this study, we considered a gene to significant if it  
594 had a q-value of less than 0.05. In the first tab, we show genes nominated from an analysis of

595 all tumors passing quality control, and in the subsequent tabs, we show the genes nominated  
596 from analysis of clinically distinct types of cutaneous squamous cell carcinoma.  
597

598 **Table S3. Genes, pathways, and tumor subtypes with significantly overlapping (or non-**  
599 **overlapping) mutations.** P-values reflect the results of a two-tailed Fisher exact test, and q-  
600 values account for multiple hypothesis testing using the Benjamini-Hochberg procedure.  
601

602 **Table S4. Somatic mutations in cutaneous squamous cell carcinoma.** Somatic mutations  
603 that passed filtering were annotated using Funcotator and shown here.  
604

## 605 Acknowledgements

606 We wish to acknowledgement funding support from the University of California Cancer  
607 Research Coordinating Committee, the UCSF Resource Allocation Program (Mt. Zion Health  
608 Fund), the UCSF Department of Dermatology, and the National Cancer Institute (K22 award).  
609

## 610 References

- 611 1. Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell* **153**, 17–37 (2013).
- 612 2. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
- 613 3. Nehal, K. S. & Bichakjian, C. K. Update on Keratinocyte Carcinomas. *N. Engl. J. Med.* **379**, 363–374  
614 (2018).
- 615 4. Yilmaz, A. S. *et al.* Differential mutation frequencies in metastatic cutaneous squamous cell  
616 carcinomas versus primary tumors. *Cancer* **123**, 1184–1193 (2017).
- 617 5. Cho, R. J. *et al.* APOBEC mutation drives early-onset squamous cell carcinomas in recessive  
618 dystrophic epidermolysis bullosa. *Sci. Transl. Med.* **10**, (2018).
- 619 6. Chitsazzadeh, V. *et al.* Cross-species identification of genomic drivers of squamous cell carcinoma  
620 development across preneoplastic intermediates. *Nat. Commun.* **7**, 12601 (2016).
- 621 7. Inman, G. J. *et al.* The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine  
622 associated mutational signature. *Nat. Commun.* **9**, 3667 (2018).
- 623 8. Cammareri, P. *et al.* Inactivation of TGF $\beta$  receptors in stem cells drives cutaneous squamous cell  
624 carcinoma. *Nat. Commun.* **7**, 12493 (2016).
- 625 9. Durinck, S. *et al.* Temporal dissection of tumorigenesis in primary cancers. *Cancer Discov.* **1**, 137–  
626 143 (2011).
- 627 10. South, A. P. *et al.* NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. *J.*  
628 *Invest. Dermatol.* **134**, 2630–2638 (2014).
- 629 11. Wang, N. J. *et al.* Loss-of-function mutations in Notch receptors in cutaneous and lung squamous  
630 cell carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17761–17766 (2011).
- 631 12. Zheng, C. L. *et al.* Transcription restores DNA repair to heterochromatin, determining regional  
632 mutation rates in cancer genomes. *Cell Rep.* **9**, 1228–1234 (2014).
- 633 13. Ji, A. L. *et al.* Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell  
634 Carcinoma. *Cell* **182**, 1661–1662 (2020).
- 635 14. Karia, P. S., Han, J. & Schmults, C. D. Cutaneous squamous cell carcinoma: estimated incidence of  
636 disease, nodal metastasis, and deaths from disease in the United States, 2012. *J. Am. Acad.*  
637 *Dermatol.* **68**, 957–966 (2013).
- 638 15. Wu, W. & Weinstock, M. A. Trends of keratinocyte carcinoma mortality rates in the United States as  
639 reported on death certificates, 1999 through 2010. *Dermatol. Surg. Off. Publ. Am. Soc. Dermatol.*  
640 *Surg. Al* **40**, 1395–1401 (2014).
- 641 16. Mansouri, B. & Housewright, C. D. The Treatment of Actinic Keratoses-The Rule Rather Than the  
642 Exception. *JAMA Dermatol.* **153**, 1200 (2017).
- 643 17. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2016. *CA. Cancer J. Clin.* **66**, 7–30 (2016).

- 644 18. Multi-omic analysis reveals significantly mutated genes and DDX3X as a sex-specific tumor  
645 suppressor in cutaneous melanoma | *Nature Cancer*. <https://www.nature.com/articles/s43018-020-0077-8>.
- 646 19. Migden, M. R. *et al.* PD-1 Blockade with Cemiplimab in Advanced Cutaneous Squamous-Cell  
647 Carcinoma. *N. Engl. J. Med.* **379**, 341–351 (2018).
- 648 20. Madeleine, M. M. *et al.* Epidemiology of keratinocyte carcinomas after organ transplantation. *Br. J.*  
649 *Dermatol.* **177**, 1208–1216 (2017).
- 650 21. Tang, J. *et al.* The genomic landscapes of individual melanocytes from human skin. *bioRxiv*  
651 2020.03.01.971820 (2020) doi:10.1101/2020.03.01.971820.
- 652 22. Stenzinger, A. *et al.* Tumor mutational burden standardization initiatives: Recommendations for  
653 consistent tumor mutational burden assessment in clinical samples to guide immunotherapy  
654 treatment decisions. *Genes. Chromosomes Cancer* **58**, 578–588 (2019).
- 655 23. DiGiovanna, J. J. & Kraemer, K. H. Shining a light on xeroderma pigmentosum. *J. Invest. Dermatol.*  
656 **132**, 785–796 (2012).
- 657 24. Brash, D. E. UV Signature Mutations. *Photochem. Photobiol.* **91**, 15–26 (2015).
- 658 25. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421  
659 (2013).
- 660 26. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–  
661 101 (2020).
- 662 27. Jiyad, Z., Olsen, C. M., Burke, M. T., Isbel, N. M. & Green, A. C. Azathioprine and Risk of Skin Cancer  
663 in Organ Transplant Recipients: Systematic Review and Meta-Analysis. *Am. J. Transplant. Off. J. Am.*  
664 *Soc. Transplant. Am. Soc. Transpl. Surg.* **16**, 3490–3503 (2016).
- 665 28. Harwood, C. A. *et al.* PTCH mutations in basal cell carcinomas from azathioprine-treated organ  
666 transplant recipients. *Br. J. Cancer* **99**, 1276–1284 (2008).
- 667 29. O'Donovan, P. *et al.* Azathioprine and UVA light generate mutagenic oxidative DNA damage. *Science*  
668 **309**, 1871–1874 (2005).
- 669 30. Martínez-Jiménez, F. *et al.* A compendium of mutational cancer driver genes. *Nat. Rev. Cancer*  
670 (2020) doi:10.1038/s41568-020-0290-x.
- 671 31. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated  
672 genes. *Nature* **499**, 214–218 (2013).
- 673 32. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* **171**, 1029–  
674 1041.e21 (2017).
- 675 33. Shain, A. H. *et al.* Exome sequencing of desmoplastic melanoma identifies recurrent NFKBIE  
676 promoter mutations and diverse activating mutations in the MAPK pathway. *Nat. Genet.* **47**, 1194–  
677 1199 (2015).
- 678 34. Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. OncodriveFML: a  
679 general framework to identify coding and non-coding regions with cancer driver mutations. *Genome*  
680 *Biol.* **17**, 128 (2016).
- 681 35. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types.  
682 *Nature* **505**, 495–501 (2014).
- 683 36. Chang, M. T. *et al.* Accelerating Discovery of Functional Mutant Alleles in Cancer. *Cancer Discov.* **8**,  
684 174–183 (2018).
- 685 37. Gonzalez-Perez, A., Sabarinathan, R. & Lopez-Bigas, N. Local Determinants of the Mutational  
686 Landscape of the Human Genome. *Cell* **177**, 101–114 (2019).
- 687 38. Shain, A. H. & Bastian, B. C. Raising the bar for melanoma cancer gene discovery. *Pigment Cell*  
688 *Melanoma Res.* **25**, 708–709 (2012).
- 689 39. Hodis, E. *et al.* A Landscape of Driver Mutations in Melanoma. *Cell* **150**, 251–263 (2012).

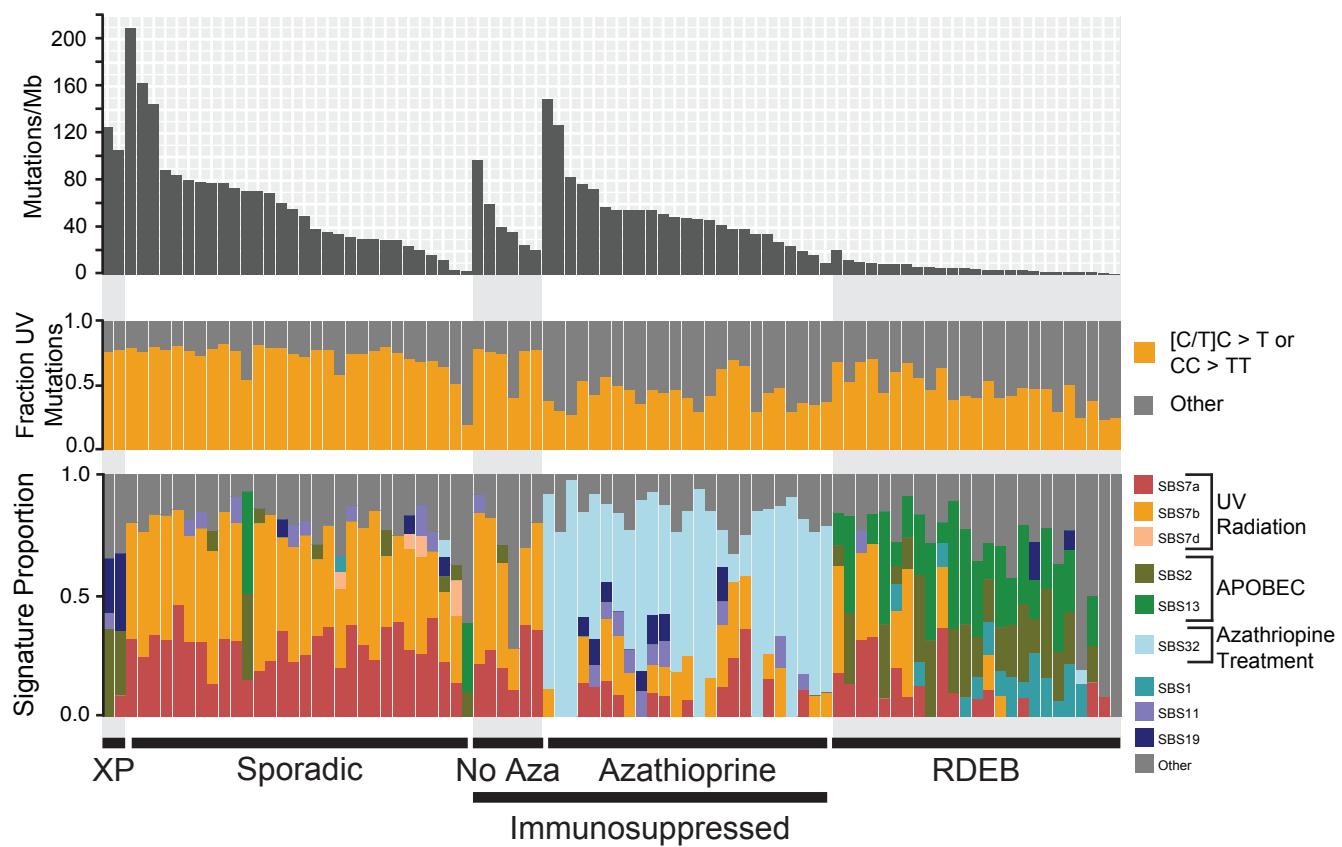
- 691 40. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* **2017**,  
692 (2017).
- 693 41. Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. Nucleotide  
694 excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016).
- 695 42. Perera, D. *et al.* Differential DNA repair underlies mutation hotspots at active promoters in cancer  
696 genomes. *Nature* **532**, 259–263 (2016).
- 697 43. Mao, P. *et al.* ETS transcription factors induce a unique UV damage signature that drives recurrent  
698 mutagenesis in melanoma. *Nat. Commun.* **9**, 2626 (2018).
- 699 44. Rodríguez-Martínez, M. *et al.* Evidence That STK19 Is Not an NRAS-dependent Melanoma Driver.  
700 *Cell* **181**, 1395-1405.e11 (2020).
- 701 45. Lee, C. S. *et al.* Recurrent point mutations in the kinetochore gene KNSTRN in cutaneous squamous  
702 cell carcinoma. *Nat. Genet.* **46**, 1060–1062 (2014).
- 703 46. Pickering, C. R. *et al.* Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin.*  
704 *Cancer Res. Off. J. Am. Assoc. Cancer Res.* **20**, 6582–6592 (2014).
- 705 47. Li, Y. Y. *et al.* Genomic analysis of metastatic cutaneous squamous cell carcinoma. *Clin. Cancer Res.*  
706 *Off. J. Am. Assoc. Cancer Res.* **21**, 1447–1456 (2015).
- 707 48. Oswald, F. *et al.* p300 acts as a transcriptional coactivator for mammalian Notch-1. *Mol. Cell. Biol.*  
708 **21**, 7761–7774 (2001).
- 709 49. Duex, J. E. *et al.* Functional Impact of Chromatin Remodeling Gene Mutations and Predictive  
710 Signature for Therapeutic Response in Bladder Cancer. *Mol. Cancer Res. MCR* **16**, 69–77 (2018).
- 711 50. Gao, Y.-B. *et al.* Genetic landscape of esophageal squamous cell carcinoma. *Nat. Genet.* **46**, 1097–  
712 1102 (2014).
- 713 51. Shain, A. H. & Pollack, J. R. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS*  
714 *One* **8**, e55119 (2013).
- 715 52. Zhang, D., Zaugg, K., Mak, T. W. & Elledge, S. J. A role for the deubiquitinating enzyme USP28 in  
716 control of the DNA-damage response. *Cell* **126**, 529–542 (2006).
- 717 53. Li, Q. *et al.* IKK1-deficient mice exhibit abnormal development of skin and skeleton. *Genes Dev.* **13**,  
718 1322–1328 (1999).
- 719 54. Lahtela, J. *et al.* Mutant CHUK and severe fetal encasement malformation. *N. Engl. J. Med.* **363**,  
720 1631–1637 (2010).
- 721 55. Boldin, M. P., Goncharov, T. M., Goltsev, Y. V. & Wallach, D. Involvement of MACH, a novel  
722 MORT1/FADD-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death. *Cell* **85**,  
723 803–815 (1996).
- 724 56. Muzio, M. *et al.* FLICE, a novel FADD-homologous ICE/CED-3-like protease, is recruited to the CD95  
725 (Fas/APO-1) death-inducing signaling complex. *Cell* **85**, 817–827 (1996).
- 726 57. Griewank, K. G. *et al.* TERT promoter mutations are frequent in cutaneous basal cell carcinoma and  
727 squamous cell carcinoma. *PLoS One* **8**, e80354 (2013).
- 728 58. Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic  
729 mutations in normal human skin. *Science* **348**, 880–886 (2015).
- 730 59. Shain, A. H. *et al.* The Genetic Evolution of Melanoma from Precursor Lesions. *N. Engl. J. Med.* **373**,  
731 1926–1936 (2015).
- 732 60. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection  
733 and Visualization from Targeted DNA Sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
- 734 61. Shain, A. H. *et al.* Genomic and Transcriptomic Analysis Reveals Incremental Disruption of Key  
735 Signaling Pathways during Melanoma Evolution. *Cancer Cell* **34**, 45-55.e4 (2018).
- 736

**Table 1.**

<b>Studies</b>	<b>Sample Size</b>
Durinck, S. et. al., <i>Cancer Disc.</i> , 2011 <sup>9</sup>	8
Wang, N.J. et. al., <i>PNAS</i> , 2011 <sup>11</sup>	4
South, A. P. et. al., <i>JID</i> , 2014 <sup>10</sup>	20
Zheng C. L. et. al., <i>Cell Reports</i> , 2014 <sup>12</sup>	4
Cammameri et. al., <i>Nature Comm.</i> , 2016 <sup>8</sup>	10
Chitsazzadeh et. al., <i>Nature Comm.</i> , 2016 <sup>6</sup>	7
Yilmaz, A. S. et. al., <i>Cancer</i> , 2017 <sup>4</sup>	6
Cho, R. J. et. al., <i>Sci. Trans. Med.</i> , 2018 <sup>5</sup>	27
Inman, G. J. et. al., <i>Nature Comm.</i> , 2018 <sup>7</sup>	10
Ji, A. L. et. al., <i>Cell</i> , 2020 <sup>13</sup>	9
Total	105

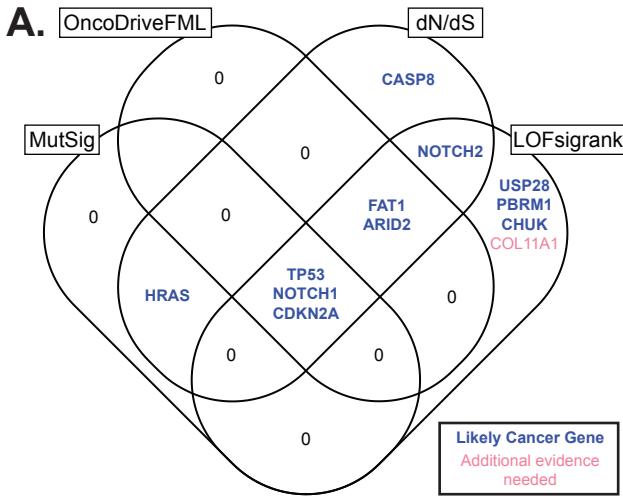
**Table 1. Summary of exome or genome sequencing studies analyzed in this meta-analysis.** The number of tumors, listed here, corresponds to unique tumors, whose data was made publicly available, and thus may not match the reported size from the original studies.

**Figure 1.**



**Figure 1. Subtypes of cutaneous squamous cell carcinoma.** Each vertical bar corresponds to a single tumor. Top track: The mutation burden (mutations per megabase) of each tumor. Middle track: The fraction of mutations within each tumor matching the canonical UV radiation signature. Bottom track: The fraction of mutations attributable to established mutational signatures within each tumor. The proposed etiology of select signatures is indicated. XP = tumors from patients with xeroderma pigmentosum. Sporadic = tumors from patients with no known comorbidities. Immunosuppressed: tumors from immunosuppressed patients -- this group is further stratified by the usage or absence of azathioprine as an immunosuppressive drug. RDEB = tumors from patients with recessive dystrophic epidermolysis bullosa.

**Figure 2.**



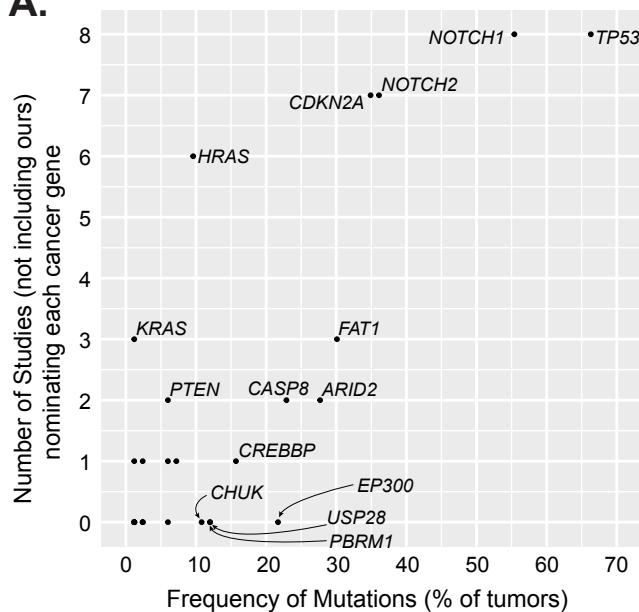
**Figure 2. Nomination of cancer genes in cutaneous squamous cell carcinoma. A.** A Venn diagram depicting nominated cancer genes from four separate cancer gene discovery programs, each designed to identify genes under positive selection in cancer. The set of candidate genes were further curated, as described, to nominate candidates for which additional evidence is warranted (red text) or not (blue text). **B.** A list of mutations in our study that overlap mutations in the cancerhotspot.org database. Mutations are grouped by gene and ordered by their q-values (lowest to highest). These mutations were also curated, as described, to nominate candidates for which additional evidence is warranted (Possible False Positive) or not (Likely Driver).

**B.**

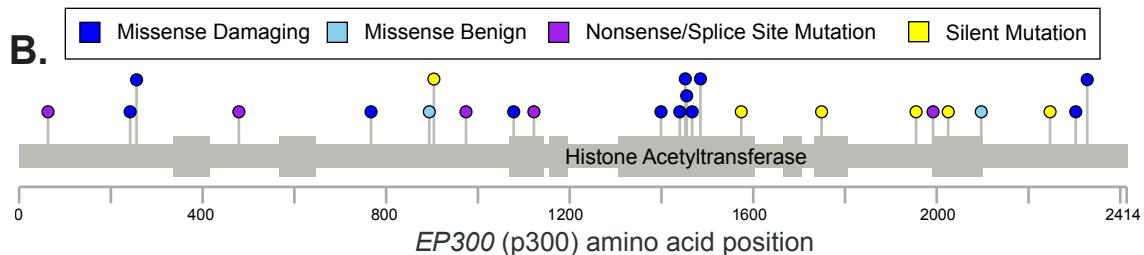
Hugo Symbol	Protein Change	Q-value	Number of Tumors	Call
KRAS	p.G12D	0.0	1	Likely Driver
	p.E542K	0.0	1	Likely Driver
	p.E545K	0.0	1	Likely Driver
PIK3CA	p.G1049D	2.8E-20	1	Likely Driver
	p.P104L	1.6E-07	1	Likely Driver
	p.P471L	2.3E-03	1	Likely Driver
GNAS	p.R187C	7.5E-257	1	Likely Driver
	p.Q61L	8.1E-210	2	Likely Driver
	p.Q61K	8.1E-210	1	Likely Driver
HRAS	p.G13D	5.5E-80	2	Likely Driver
	p.G13R	5.5E-80	1	Likely Driver
	p.G12D	4.3E-56	2	Likely Driver
EZH2	p.Y646N	6.9E-69	2	Likely Driver
RAC1	p.P29S	1.6E-65	1	Likely Driver
	p.P29L	1.6E-65	1	Likely Driver
NFE2L2	p.R18G	9.7E-44	1	Likely Driver
EP300	p.D1399H	6.9E-22	1	Likely Driver
	p.Y1467H	4.8E-07	1	Likely Driver
	p.Q1455K	1.9E-02	1	Likely Driver
	p.P1452Y	3D Hotspot	1	Likely Driver
SPOP	p.F102S	1.6E-18	1	Likely Driver
MTOR	p.S2215F	1.2E-16	1	Likely Driver
RRAS2	p.Q72L	6.6E-15	1	Likely Driver
ERBB3	p.G284R	4.1E-12	1	Likely Driver
KNSTRN	p.S24F	1.5E-10	3	Possible False Positive
DIS3	p.D458H	3.5E-09	1	Possible False Positive
HIST1H3C	p.K37I	4.3E-05	1	Possible False Positive
EGFR	p.P551L	1.3E-04	1	Possible False Positive
CREBBP	p.D1435N	1.2E-03	1	Likely Driver
KDR	p.R1032Q	5.8E-03	2	Possible False Positive
GNA11	p.R183C	6.0E-03	1	Likely Driver
MAP2K4	p.R145W	7.4E-03	1	Possible False Positive

**Figure 3.**

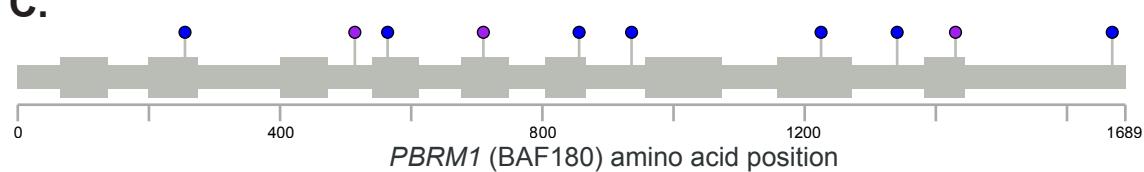
**A.**



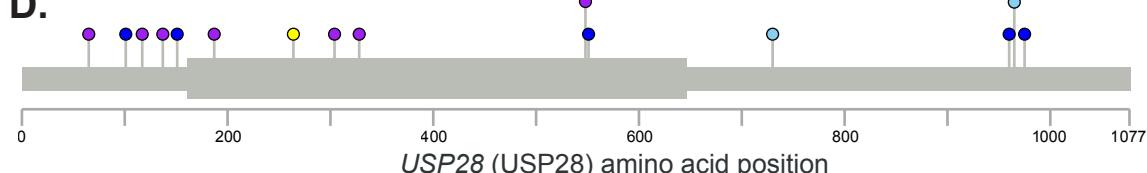
**B.**



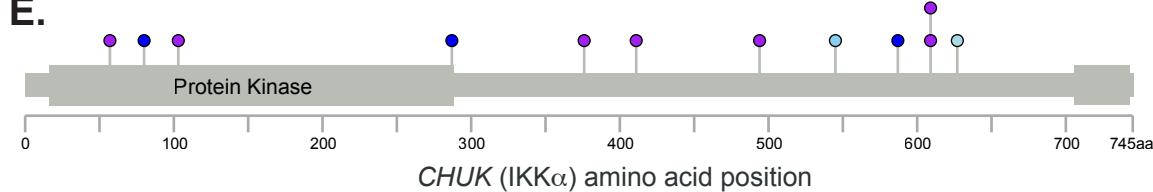
**C.**



**D.**

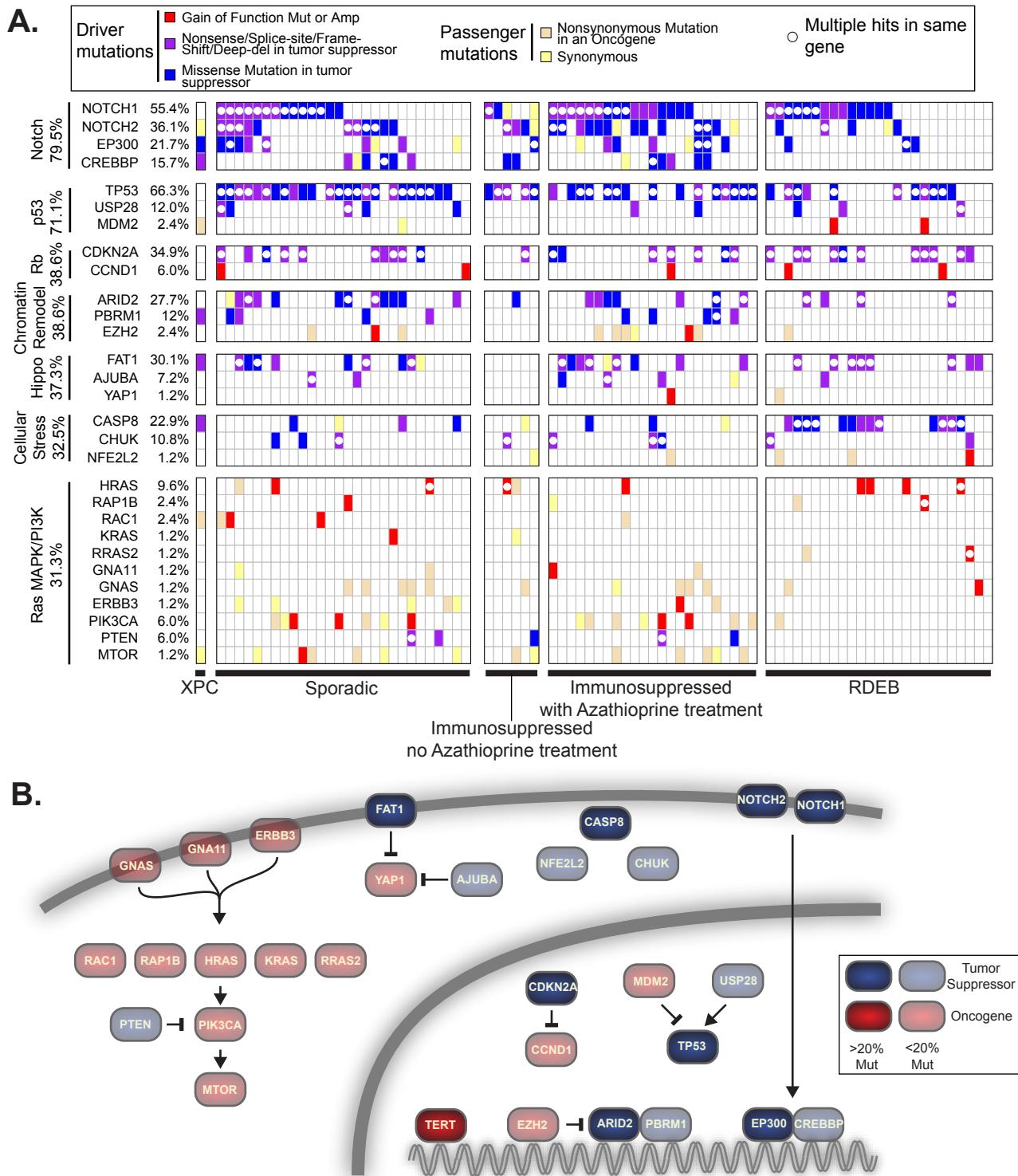


**E.**



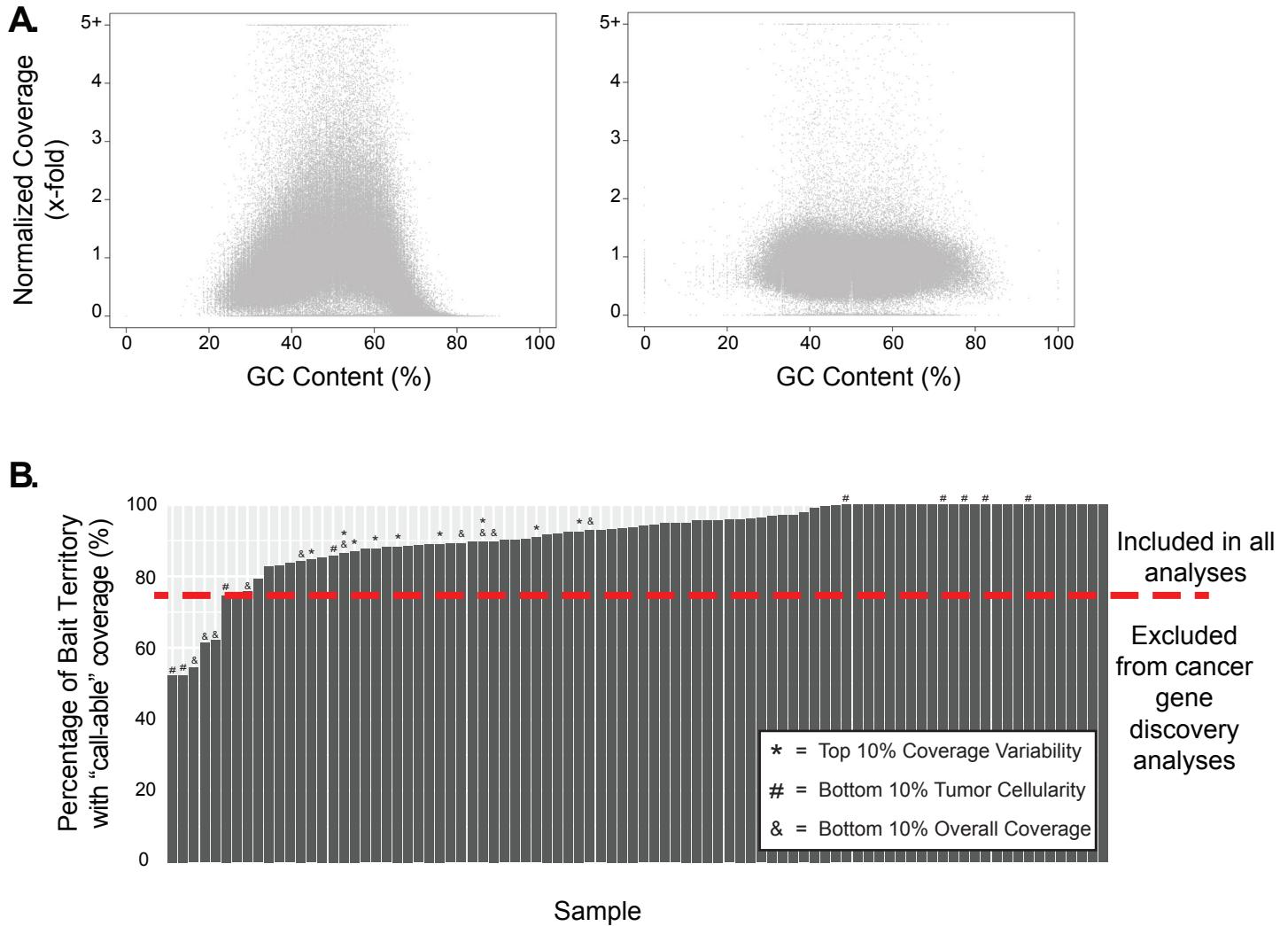
**Figure 3. Candidate cancer genes in cutaneous squamous cell carcinoma. A.** The genes nominated in our meta-analysis are stratified by their mutation frequency (x-axis) and how often they were nominated in 8 previous studies (y-axis) that catalogued drivers of cutaneous squamous cell carcinoma. EP300, PBRM1, USP28, and CHUK were mutated in greater than 10% of tumors but not implicated by other studies. Lollipop diagrams portray the spectrum of mutations in each of these four genes in panels **B-E**.

**Figure 4.**



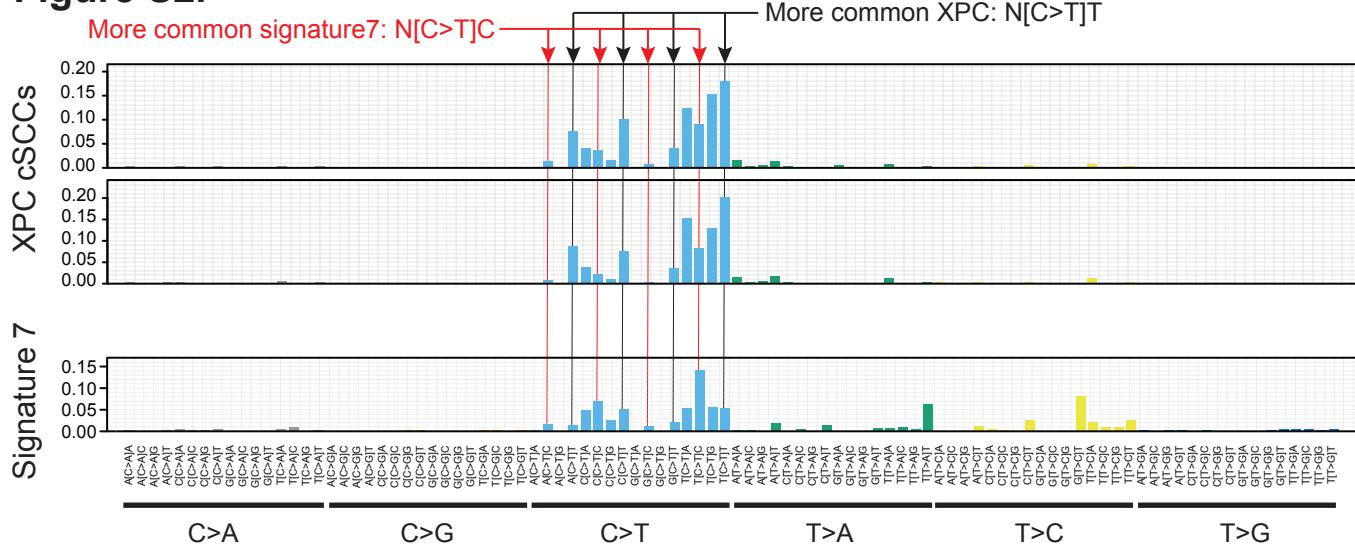
**Figure 4. The landscape of driver mutations in cutaneous squamous cell carcinoma. A.** Tiling plot of the genetic alterations (rows) in each tumor (columns). Genes and tumors are further organized into pathways and clinical subtypes. The percentages of samples harboring pathogenic alterations are indicated. Mut, mutation; Amp, amplification; XPC, xeroderma pigmentosum; RDEB, recessive dystrophic epidermolysis bullosa. **B.** Pathways affected. TERT was not implicated in this study, which focused on coding mutations, but is included here because it is known to have promoter mutations, as described.

**Figure S1.**



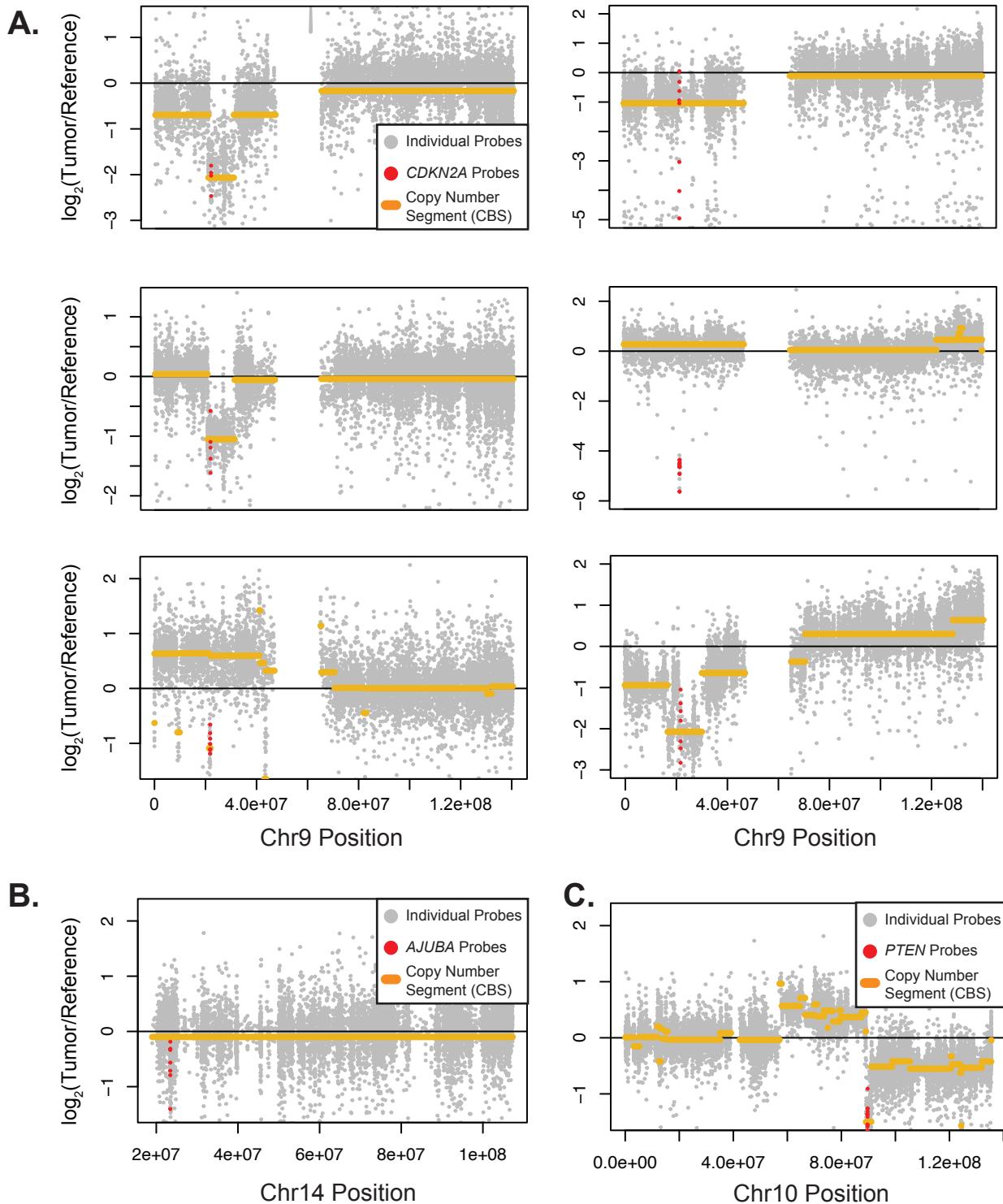
**Figure S1. Ability to detect mutations in each tumor varies substantially.** **A.** Some tumors had high overall coverage but extreme variability in coverage, primarily linked to GC content, thus diminishing our ability to detect mutations within large portions of the bait territory. An example of a tumor in the top 10% of coverage variability (left panel) and a tumor in the bottom 10% of coverage variability (right panel). Each datapoint corresponds to a bait interval, stratified by its GC content and sample-normalized coverage. **B.** The percentage of base pairs within the target bait territory with sufficient coverage to call a mutation. Tumors with low overall sequencing coverage, extreme variability in sequencing coverage, and/or low tumor cellularity tended to have a lower proportion of "callable" mutations. Tumors with greater than 75% "callable" mutations were included in all analyses, whereas those with fewer were only included in certain analyses, as described.

## Figure S2.



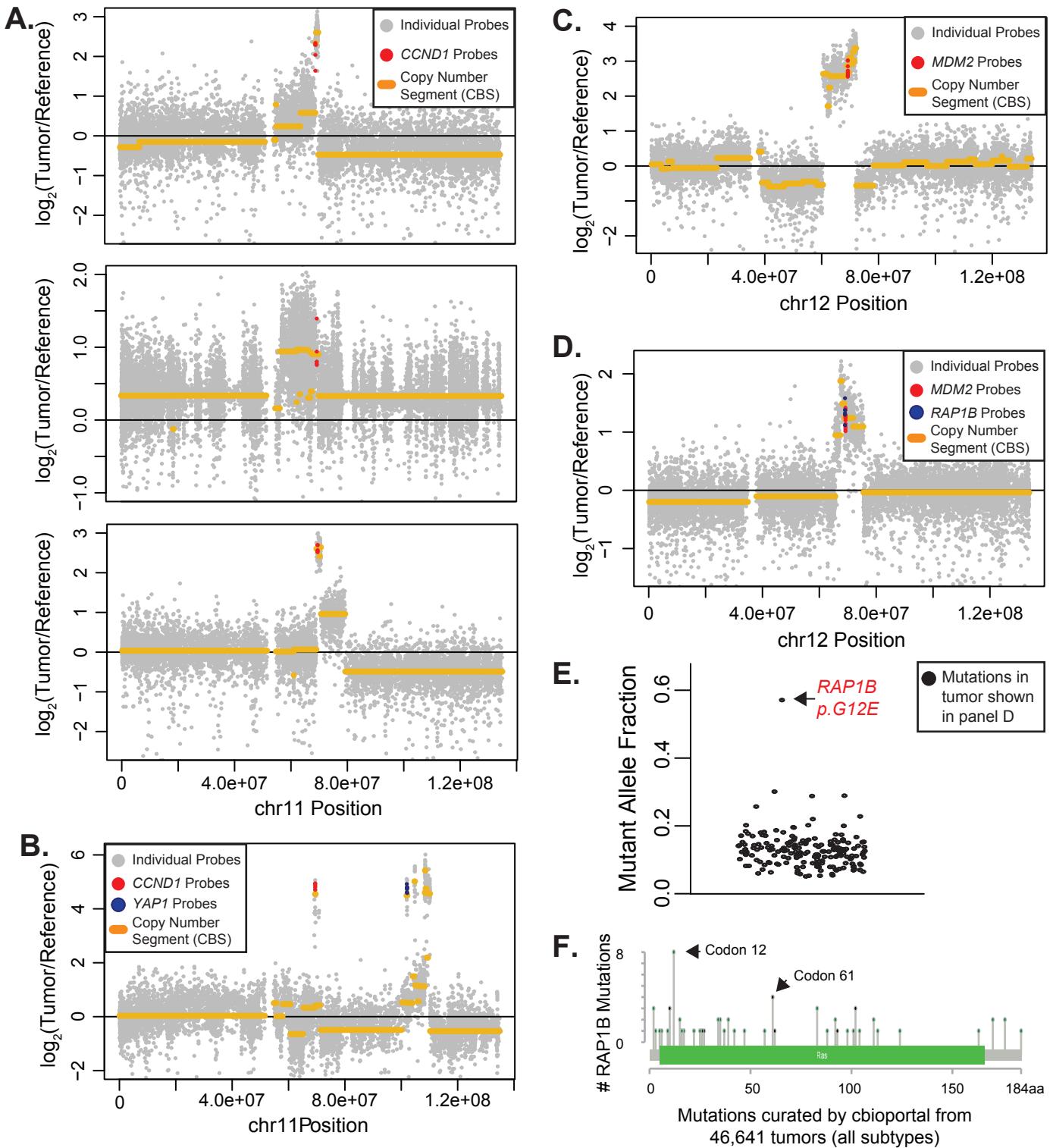
**Figure S2. Somatic mutations in Xeroderma Pigmentosum tumors occur in a distinct tri-nucleotide context.** The tri-nucleotide context of mutations in two XPC-/- tumors as compared to signature 7. C>T transitions at dipyrimidines were common in both, however the basepair downstream (three prime) to the mutation site differed with thymines more common in XPC-/- tumors and cytosines more common in signature 7.

**Figure S3.**



**Figure S3. Focal deletions affecting *CDKN2A*, *AJUBA*, and *PTEN*.** Copy number over individual genomic positions (data points) with copy number segments (yellow lines) overlaid. Colored data points correspond to probes covering *CDKN2A* (panel A), *AJUBA* (panel B), and *PTEN* (panel C).

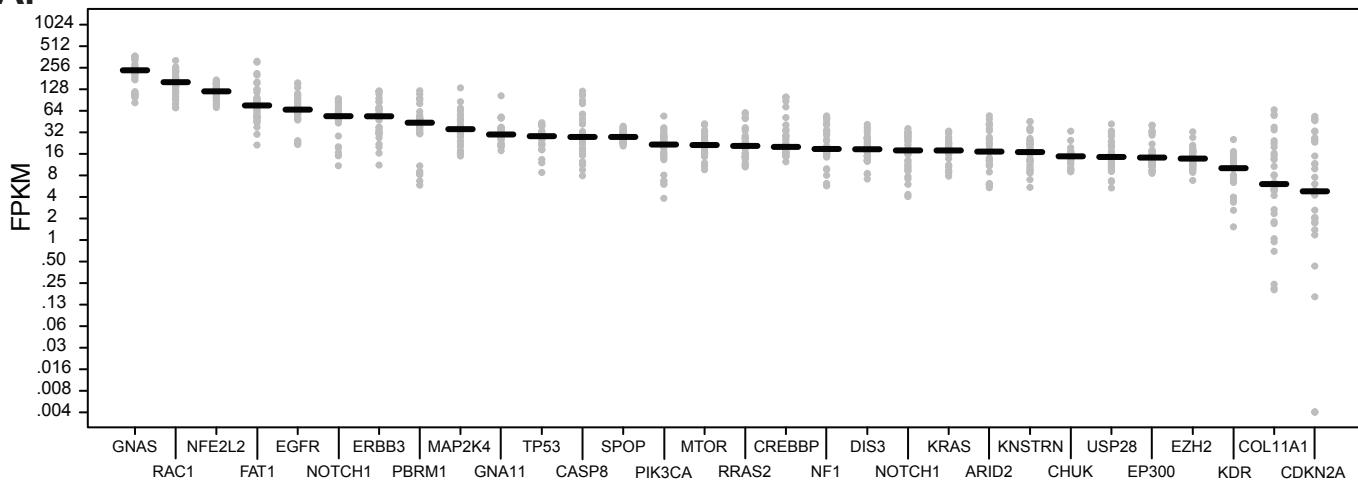
**Figure S4.**



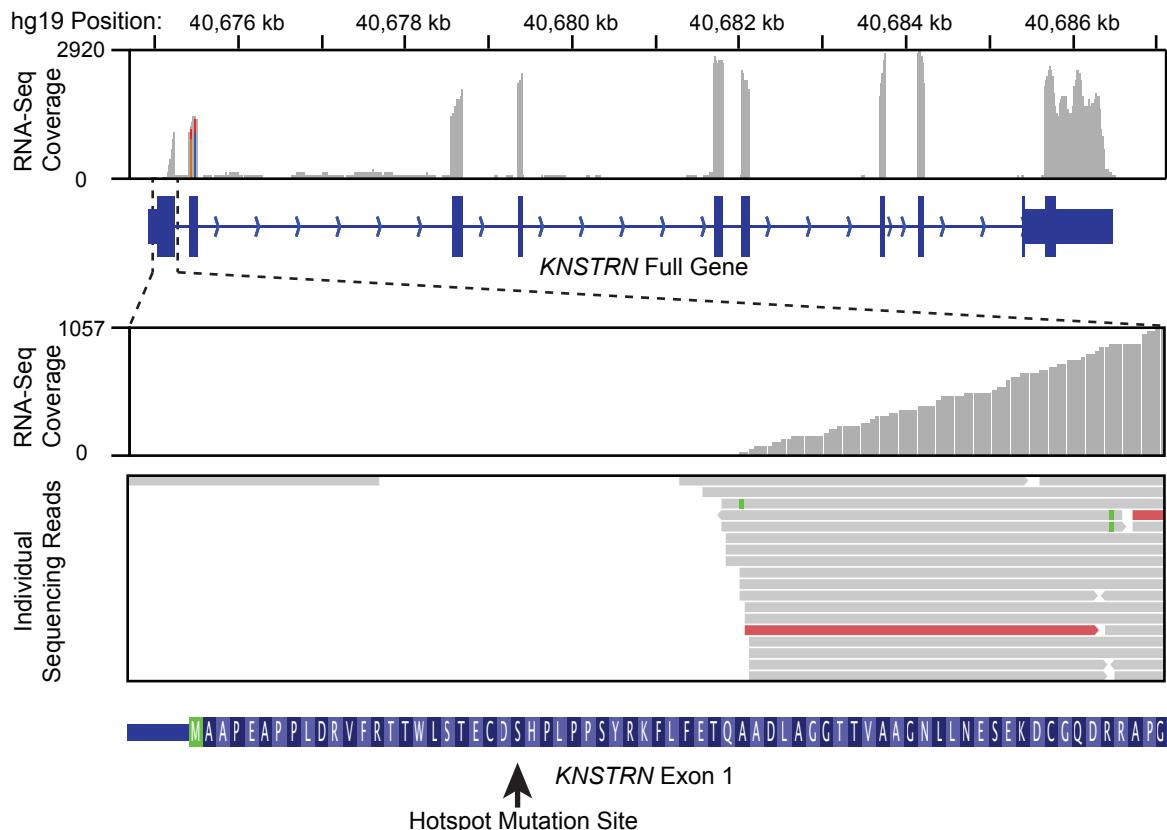
**Figure S4. Focal amplification affecting CCCND1, YAP1, MDM2, and/or RAP1B in six tumors.** Copy number over individual genomic positions (data points) with copy number segments (yellow lines) overlaid. Colored data points correspond to probes covering CCND1 (panels A and B), YAP1 (panel B), MDM2 (panels C and D), and RAP1B (panel D). Note that amplifications converge upon the noted genes across these six tumors. The tumor with amplification of RAP1B (shown in panel D) also had a point mutation affecting the 12th codon of RAP1B (panel E). The mutant allele frequency of this point mutation was higher than all other mutations in that tumor, consistent with the mutation affecting the amplified allele. RAP1B is a ras-like protein with recurrent mutations affecting codons 12 and 61 across cancer (panel F), analogous to mutations known to activate other ras-like proteins.

## Figure S5.

**A.**



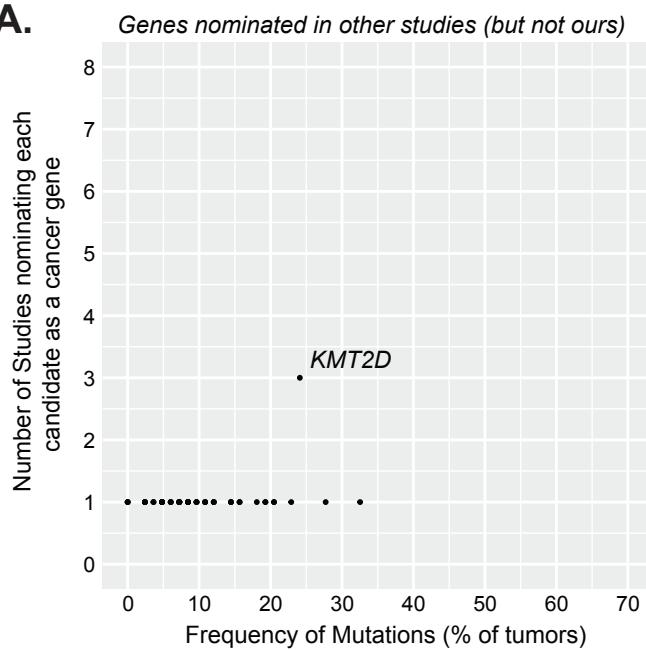
**B.**



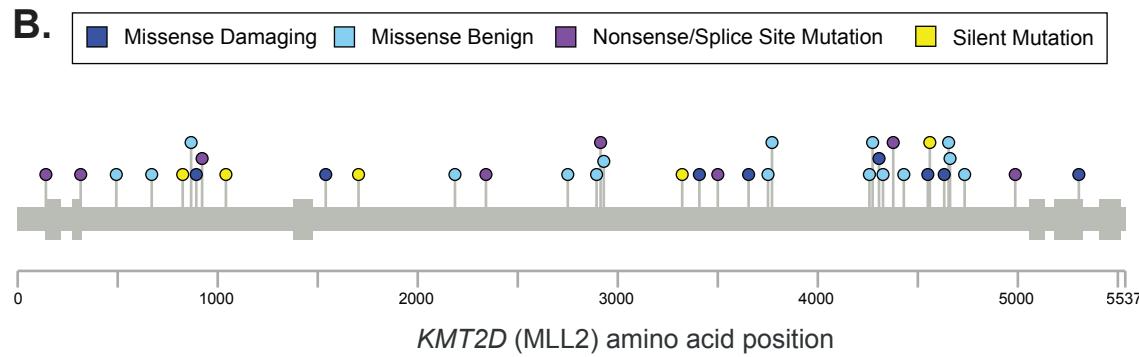
**Figure S5. Gene expression of candidate genes. A.** The Fragments Per Kilobase of transcript per Million reads (FPKM) values -- a measurement that normalizes for both gene length and sequencing depth in RNA-sequencing data -- of candidate genes. RNA-sequencing data derived from both normal skin samples (n=17 samples) and cutaneous squamous cell carcinomas (n=17 tumors). This data was used to flag potential false positive candidates based on their poor expression in normal and neoplastic keratinocytes. **B.** Transcription of KNSTRN begins downstream of the hotspot mutation site. RNA-sequencing data was aggregated from normal skin (n=17) and cutaneous squamous cell carcinoma (n=17). Top panel -- RNA-sequencing coverage over the KNSTRN gene. Coverage was approximately 2,000 to 3,000-fold over most exons with the exception of exon1. Bottom panel -- A zoomed inset of RNA-sequencing coverage over exon1. While sequencing coverage surpassed 1,000-fold at the three-prime junction of exon1, there were no reads spanning the hotspot mutation site from any of the 34 samples, in-line with our prediction that the mutation is non-coding.

## Figure S6.

A.



B.



**Figure S6. KMT2D has a benign spectrum of mutations.** A. Genes nominated by other studies, but not our study, are stratified by their mutation frequency (x-axis) and how often they were nominated in 8 previous studies (y-axis) that catalogued drivers of cutaneous squamous cell carcinoma. KMT2D was the only gene recurrently implicated in other studies but not ours. B. Lollipop diagram portrays the spectrum of mutations. KMT2D was not nominated by cancer gene discovery algorithms because of its high frequency of silent mutations and missense mutations predicted to be benign.