Transposon mutagenesis identifies cooperating genetic drivers during keratinocyte transformation and cutaneous squamous cell carcinoma progression

Aziz Aiderus[1,*], Justin Y. Newberg[1,2,*], Liliana Guzman-Rojas[2,*], Ana M. Contreras-Sandoval[1], Amanda L. Meshey[1], Devin J. Jones[2], Felipe Amaya-Manzanares[2], Roberto Rangel[2], Jerrold M. Ward[3], Song-Choon Lee[3], Kenneth Hon-Kim Ban[3], Keith Rogers[3], Susan M. Rogers[3], Luxmanan Selvanesan[4], Leslie A. McNoe[4], Neal G. Copeland[2,3], Nancy A. Jenkins[2,3], Kenneth Y. Tsai[5,6,7], Michael A. Black[4], Karen M. Mann[1,2,3,7,8,9], and Michael B. Mann[1,2,3,6,7,9,10]

[1]Department of Molecular Oncology, Moffitt Cancer Center & Research Institute, Tampa, FL, USA.
[2]Cancer Research Program, Houston Methodist Research Institute, Houston, Texas, USA.
[3]Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Biopolis, Singapore, Republic of Singapore.
[4]Centre for Translational Cancer Research, Department of Biochemistry, University of Otago, Dunedin, New Zealand.
[5]Departments of Anatomic Pathology & Tumor Biology, Moffitt Cancer Center & Research Institute, Tampa, FL, USA.
[6]Donald A. Adam Melanoma and Skin Cancer Research Center of Excellence, Moffitt Cancer Center & Research Institute, Tampa, FL, USA.
[7]Department of Oncologic Sciences, Morsani College of Medicine, University of South Florida, Tampa, FL, USA.
[8]Departments of Gastrointestinal Oncology & Malignant Hematology, Moffitt Cancer Center & Research Institute, Tampa, FL, USA.
[9]Cancer Biology and Evolution Program, Moffitt Cancer Center & Research Institute, Tampa, FL, USA.
[10]Department of Cutaneous Oncology, Moffitt Cancer Center & Research Institute, Tampa, FL, USA.

*These authors contributed equally to this work.

Correspondence to M.B.M. (Michael.Mann@moffitt.org).

Present addresses:
Foundation Medicine, Inc., Cambridge, MA, USA (J.Y.N.); Houston Methodist Cancer Center, Houston Methodist Research Institute, Houston, TX, USA (L.G.-R.); Department of Genetics and Development, Columbia University, New York, NY, USA (D.J.J.); Monoclonal Antibody Core Facility, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA (F.A.-M); Department of Head & Neck Surgery, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA (R.R.); Global VetPathology, Montgomery Village, MD, USA (J.M.W.); Science Centre Singapore, Republic of Singapore (S.-C.L.); Department of Biochemistry, Yong Loo Lin School of Medicine, National University Singapore, Republic of Singapore (K.H.-K.B.); Pacific Edge Limited, Dunedin, Otago, New Zealand (L.S.); AgResearch Invermay Agricultural Centre, Mosgiel, Otago, New Zealand (L.A.M.); and Genetics Department, University of Texas M.D. Anderson Cancer Center, Houston, TX (N.G.C. and N.A.J.).

Abstract

The systematic identification of genetic events driving cellular transformation and tumor progression in the absence of a highly recurrent oncogenic driver mutation is a challenge in cutaneous oncology. In cutaneous squamous cell carcinoma (cuSCC), the high UV-induced mutational burden poses a hurdle to achieve a complete molecular landscape of this disease. Here, we utilized the *Sleeping Beauty* transposon mutagenesis system to statistically define drivers of keratinocyte transformation and cuSCC progression *in vivo* in the absence of UV-IR, and identified both known tumor suppressor genes and novel oncogenic drivers of cuSCC. Functional analysis confirms an oncogenic role for the *ZMIZ* genes, and tumor suppressive roles for *KMT2C, CREBBP* and *NCOA2*, in the initiation or progression of human cuSCC. Taken together, our *in vivo* screen demonstrates an extremely heterogeneous genetic landscape of cuSCC initiation and progression, which can be harnessed to better understand skin oncogenic etiology and prioritize therapeutic candidates.

Key words: *Sleeping Beauty* transposon insertional mutagenesis, cutaneous squamous cell carcinoma, keratinocyte transformation, cancer driver genes, cancer hallmarks, chromatin modification, ZMIZ paralogs.

Subject Areas: Cancer Biology, Genetics and Genomics.

44 **INTRODUCTION**

45 Cutaneous squamous cell carcinoma (cuSCC) is the second most common cancer in man, with approximately one
46 million cases diagnosed annually in the United States. Although the majority of cuSCC are considered a low-risk
47 neoplasm, up to 5% of high-risk cuSCCs are locally or distantly invasive and carry a poor prognosis due to a lack of
48 biomarkers, therapeutic targets, or FDA-approved molecularly targeted therapies. This represents a substantial
49 unmet need for approximately 50,000 patients per year with high-risk cuSCC, and an opportunity to identify new
50 therapeutic modalities that could improve disease outcomes. All non-viral associated skin cancers are thought to
51 require multiple cooperating mutations that deregulate distinct signaling pathways to initiate and progress the
52 multi-step transformation of normal cells into a clinically significant neoplasm. Indeed, identifying cooperating
53 mutations that drive malignant transformation is a prerequisite for developing better combinatorial therapies for
54 managing and treating skin cancers. Most skin cancers, including cuSCC (Pickering et al., 2014; South et al., 2014),
55 have the highest mutation rates among human cancers due to ultraviolet irradiation (UV-IR) induced damage from
56 chronic, intermittent sun exposure. Thus, using human cancer sequencing data alone, with some of the highest
57 mutational burdens of any cancer, poses challenges to identify cooperating, low-penetrant mutations that lead to
58 cancer progression. This presents a need to develop *in vivo* model systems to help identify and prioritize novel
59 cooperating candidate cancer drivers for keratinocyte transformation and subsequent progression to late-stage,
60 invasive cuSCC.

61 *Sleeping Beauty* (SB) insertional mutagenesis (Ivics et al., 1997) is a powerful tool used to perform genome-wide
62 forward genetic screens in laboratory mice for cancer gene discovery (Collier and Largaespada, 2007; Copeland and
63 Jenkins, 2010; Dupuy et al., 2005; Dupuy et al., 2009; Mann et al., 2013; Mann et al., 2016b; Mann et al., 2012; Mann
64 et al., 2015; Mann et al., 2014b; Rangel et al., 2016; Takeda et al., 2015) in animal models of both hematopoietic and
65 solid tumors (Mann et al., 2014a; Mann et al., 2014b). SB transposons can identify early cancer progression drivers
66 that cooperate to initiate tumors (Mann et al., 2015; Takeda et al., 2015), and potentially drive metastasis (Genovesi
67 et al., 2013; Perez-Mancera et al., 2012). Importantly, SB insertions induce changes in gene expression, thus
68 providing epigenetic information not easily obtained from carcinogenesis mouse models using chemical (Nassar et
69 al., 2015) or chronic UV irradiation (Chitsazzadeh et al., 2016; Knatko et al., 2017) carcinogenesis mouse models or
70 from limited patient samples. We demonstrate that SB mobilization of a low-copy T2/Onc3 transposon allele is
71 sufficient to induce and progress a variety of cancers *in vivo*. Here, we report our efforts for cancer gene discovery in
72 skin tumors. We focused on the analysis of skin tumors from these mice and identified several oncogenic and many
73 tumor suppressor driver genes. Using high-throughput sequencing approaches (Mann et al., 2016b; Mann et al.,
74 2015) to identify genome-wide SB mutations. Using our SB Driver Analysis (Newberg et al., 2018b) statistical
75 framework, we profiled genome-wide SB mutations from early- and late-stage cuSCC and defined recurrently
76 mutated, statistically significant candidate cancer drivers (CCDs) from bulk cuSCC tumors and from normal
77 keratinocytes and early stage tumors, identifying both known tumor suppressor genes and novel oncogenic drivers.
78 We further prioritized oncogenic and tumor suppressor candidates, and provide *in vitro* and *in vivo* functional
79 evidence for the roles of these genes in the initiation and progression of cuSCC. Taken together, our efforts provide a

80 systematic evolutionary landscape of cuSCC genesis, and highlights key pathways promoting disease that are
81 potential therapeutic targets.

82 **RESULTS**

83 **Candidate cancer driver discovery in SB-driven keratinocyte cancer models**

84 We performed a forward genetic screen to define the tumor spectrum driven by the cooperation of SB with Trp53
85 mutant alleles and the underlying cooperating genes (**Fig. 1a** and **Supplementary Fig. 1a**). Whole-body
86 mutagenesis was achieved using low-copy, bi-functional SB transposons from SB|Onc3 mice carrying an inducible
87 SB transposase (Dupuy et al., 2009), Onc3 transposon (Dupuy et al., 2009), and *Actb-Cre* transgene (Lewandoski et
88 al., 1997). The addition of either a conditional *Trp53* null (*Trp53*$^{KO/+}$) (Jonkers et al., 2001) or recurrent point mutant
89 allele (*Trp53*$^{R172H/+}$) (Olive et al., 2004) accelerated tumor progression and significantly reduced tumor-free survival
90 relative to *Trp53* wild-type littermate controls (P<0.0001, **Fig. 1a**). SB|Onc3 mice succumbed to a wide variety of
91 solid and hematopoietic tumor types with variable penetrance (**Supplementary Table 1**). SB|*Trp53*$^{KO/+}$ mice were
92 the earliest to reach endpoint, but had the lowest overall tumor burden, presumably because these mice had the
93 highest proportion of hematopoietic disease (**Supplementary Table 1**). SB|*Trp53*$^{R172H/+}$ mice reached endpoint
94 earlier than SB|*Trp53*$^{+/+}$ mice and had the widest spectrum of solid tumors (**Supplementary Table 1**). This
95 observation is consistent with previous studies showing that missense mutations in *Trp53* promote malignant
96 transformation by abrogating both the tumor suppressive and oncogenic (gain-of-function and/or dominant
97 negative) functions of TRP53 *in vivo* (Doyle et al., 2010; Hanel et al., 2013; Lang et al., 2004; Morton et al., 2010;
98 Olive et al., 2004). The most prevalent of the more than 20 solid tumor types observed were early-stage cutaneous
99 keratoacanthoma (cuKA) (**Fig. 1b,g**) and late-stage, well differentiated cutaneous squamous cell carcinoma (cuSCC)
100 (**Fig. 1c-g**), with a disease penetrance and latency similar to Onc3 cohorts produced with constitutively active SBase
101 (Dupuy et al., 2009). Importantly, these cuSCC/cuKAs were absent in mice with mutant *Trp53* lacking SB (**Fig. 1g**),
102 strongly suggesting SB mutagenesis drives initiation of cellular transformation and subsequent cancer progression
103 of a wide spectrum of tumor types *in vivo*.

104 To investigate SB-driven keratinocyte transformation and cuSCC progression in wildtype and Trp53 mutant mice,
105 we sequenced SB insertions sites using the 454_splink method (Mann et al., 2016b) from 83 individual skin masses
106 isolated from 65 SB|Onc3 mice (**Supplementary Table 2–3**), that were histologically confirmed as late-stage
107 cutaneous squamous cell carcinoma (cuSCC). Using gene-guided SB Driver Analysis statistical framework (Newberg
108 et al., 2018b), we identified 31,266 non-redundant transposon insertion sites from the individual (**Supplementary**
109 **Table 1**) or combined analysis of Trp53 mutant and wild type cuSCC cohorts (**Supplementary Table 4**) which
110 were used to statistically define 428 candidate cancer drivers (CCDs, **Supplementary Table 5**; see **Methods** and
111 **Supplementary Figure 2**). Nearly all CCDs were identified across all cohorts, suggesting that their role in driving
112 cuSCC is independent of *Trp53* mutation (**Supplementary Figure 3a**). For this reason, we focused subsequent data
113 analysis on the superset consisting of all SB|cuSCC masses with all *Trp53* alleles.

114 Previous studies suggest that most CCDs identified by SB mutagenesis function during tumor progression rather
115 than tumor initiation (Mann et al., 2015). We previously reported a workflow that allow us to distinguish initiating
116 and progression drivers within bulk tumor and single cell sequencing in an SB-driven myeloid leukemia model
117 (Mann et al., 2016b). Stringent filtering of the SB insertions to include only events with the highest sequencing read
118 depths defines trunk driver alterations that are predicted to be under positive-selection within the clonally
119 expanding tumors, where they participate in either initiation or early progression of cellular transformation (Mann
120 et al., 2016b; Mann et al., 2015). To identify CCDs that drive keratinocyte transformation, we selected the SB
121 insertion sites with the highest number of sequencing reads to define trunk drivers (see **Online Methods**),
122 reasoning that tumor initiating insertions would have among the highest frequency in tumor cells. Using SB
123 insertions sites by the 454_splink method, we identified 142 trunk drivers of initiation and/or early progression of
124 keratinocyte transformation *in vivo* (subsequently referred to as trunk drivers) (**Supplementary Table 6** and
125 **Supplementary Figure 3b**). SB insertion maps for representative cuSCC trunk drivers show the locations of
126 mapped SB insertions and define candidate cancer drivers as putative proto-oncogenes (**Supplementary Figure 3c**)
127 or tumor suppressor (**Supplementary Figure 3d**) candidate cuSCC driver genes.

128 Next, we investigated keratinocyte transformation using keratinocyte-specific SB mutagenesis in *Pten*-sensitized
129 mice. The screen originally was designed to activate an inducible allele of the SB transposase (Rosa26-LSL-SB11)
130 using *Keratin5*-Cre (K5-Cre, Tg(KRT5-cre)5132Jlj, MGI:3050065) (Ramirez et al., 2004) to drive SB insertional
131 mutagenesis in stratified epithelium of the mammary ducts in a haploinsufficient $Pten^{floxed/+}$ sensitized background
132 (Rangel et al., 2016). However, K5 also induces cre expression in cells of the basal epidermis. Using a two-generation
133 mating scheme, we first introduced K5-Cre onto the background of a Cre-LoxP inducible *Pten* mutant
134 strain, $Pten^{CKO/+}$ ($Pten^{tm1Hwu}$; MGI:2156086). Compound heterozygous progeny (K5-Cre; $Pten^{CKO/+}$) were mated with
135 double homozygous SB strains containing either a high-copy (T2/Onc2-TG.6113) or a low-copy (T2/Onc3-
136 TG.12740) transposon donor concatemer and the inducible SB transposase (Rosa26-LSL-SBase or $SBase^{LSL}$) to
137 generate experimental and control cohorts (Rangel et al., 2016) (**Supplementary Figure 1a**). Cutaneous
138 keratoacanthomas (cuKA) were diagnosed from routine H&E sections and found to occur in 43% (13 of 30) of
139 quadruple mutant mice (**Supplementary Table 1**), either K5-Cre; $Pten^{CKO/+}$; TG.6113; $SBase^{LSL}$ or K5-Cre; $Pten^{CKO/+}$;
140 TG.12740; $SBase^{LSL}$. Four of the mice with cuKA also developed a breast tumor and none of the mice in the control
141 groups developed skin masses before reaching a predetermined experimental end-point at 450 days of age.

142 To identify SB drivers of cuKA in *Pten*-sensitized mice, we used 45 flash frozen cuKA specimens from 13 quadruple
143 mutant mice (**Supplementary Table 2–3**) collected at necropsy to create genomic DNA and sequenced using the
144 454-Splink method to sequencing SB insertion events. Using the 454 sequencing data (**Supplementary Table 7**), we
145 performed gene-guided SB Driver Analysis (Newberg et al., 2018b) to define 302 statistically significant candidate
146 cancer drivers (**Supplementary Table 8**) and 29 Trunk drivers (**Supplementary Table 9**). Our top ranked genes
147 were *Chuk* (*IKKA*/*IKK1*), *Notch1*, and *Kctd15* (**Supplementary Figure 4a-b**). All genes were found to have
148 inactivating patterns of SB insertion events into each of their gene-coding regions, strongly suggesting that
149 cooperating tumor suppressors drive early keratinocyte transformation and cuKA initiation *in vivo*. However, three
150 genomes lacking *Chuk* insertions were found to have activating SB insertions into *Zmiz1* that were identical to those

151 observed in the cuSCC cohorts (**Supplementary Figure 4a**). This may suggest that the altered function(s) of *Chuk* in
152 cuKAs may be overlapping with those of the SB-induced *Zmiz1*$^{\Delta N185}$ (Rogers et al., 2013) allele.

153 **SBCapSeq enhances transposon driver detection in normal, premalignant, and tumor skin cell genomes**

154 To confirm and extend our SB analysis from 454 sequencing and define genetic drivers acquired during the
155 evolution of keratinocyte transformation, we developed an enhanced version of SB Capture Sequencing (SBCapSeq)
156 (Mann et al., 2016b) that improves transposon-based, liquid-phase capture experimental workflow optimized for
157 Ion Torrent sequencing from normal and tumor specimen genomes (Mann, *et al*., 2020, manuscript in preparation).
158 We used SBCapSeq to investigate keratinocyte mutational heterogeneity by histological stage.

159 We generated SBCapSeq libraries using genomic DNA isolated from the necropsy collected flash frozen specimens of
160 60 cuSCC, 11 cuKA and 32 normal skin genomes (**Fig. 2** and **Supplementary Table 2**). All cuSCC specimens were
161 from independent mice and 59 of 60 were previously sequenced using the Splink_454 method (**Supplementary**
162 **Table 10**). Comparisons between the sequencing methods allowed for platform-independent sequence validation of
163 clonal drivers and demonstrated the superiority of the SBCapSeq method to identify clonal and sub-clonal insertions
164 based on the total number and read-depth enrichment of genome-wide SB insertion events with biological and
165 technical reproducibility (**Supplementary Figure 5**). Using the SBCapSeq bioinformatics workflow (Mann et al.,
166 2016b) and gene-guided SB Driver Analysis statistical framework (Newberg et al., 2018b), we identified 11,113,694
167 reads at 59,337 non-redundant transposon insertion sites from cuSCCs (**Supplementary Table 11**) which were
168 used to statistically define 1,333 candidate cancer driver genes (**Supplementary Table 12**). Additionally, we
169 identified 86 trunk drivers (**Supplementary Figure 6a** and **Supplementary Table 13**). Using the same
170 methodology, we next analyzed cuKA specimens harvested either from mice with cuSCC or from an independent
171 age- and genotype-matched cohort (**Supplementary Table 10**) and identified 2,085,280 reads at 6,727 non-
172 redundant transposon insertion sites (**Supplementary Table 14**) defining 192 candidate cancer driver genes
173 (**Supplementary Table 15**) and 10 trunk drivers (**Supplementary Figure 6b** and **Supplementary Table 16**).
174 Finally, we identified 2,910,945 reads at 47,537 non-redundant transposon insertion sites (**Supplementary Table**
175 **17**) from normal skin specimens (**Supplementary Table 4**) that defined 547 candidate cancer drivers
176 (**Supplementary Table 18**). No trunk drivers were identified.

177 The histologically normal skin genomes from mice with SB mobilization had a mean of 1,320 SBCapSeq detected
178 insertion events (range 1–3,857; **Supplementary Table 2**), although only 0.82% (389/47,537) of the normal skin
179 insertions had read depths >300 reads per site (range 1–2,297; **Supplementary Table 17**). This is one order of
180 magnitude less than the 23.7% (range 1-19,798) of cuKA genomes or 8.5% (5,014/57,337) of cuSCC genomes with
181 SBCapSeq detected insertion events with >300 reads per specimen (cuKA range 1-19,798; **Supplementary Table**
182 **14** and cuSCC range 1–35,809; **Supplementary Table 11**). Clonal selection of recurrent, high-read depth SB
183 insertion sites is featured prominently within the cuSCC and cuKA genomes and almost entirely missing within the
184 normal skin genomes (**Fig. 2**), strongly suggesting that SBCapSeq read depths below ~300 represent background
185 events and those at 300 and substantially higher represent clonally selected foreground mutation events.

186 Further evidence for clonal selection associated with disease progression was provided by technical replication of
187 SBCapSeq libraries prepared from two independent cuSCC or normal skin specimens (**Supplementary Figure 7**).
188 Independent SBCapSeq libraries demonstrated excellent biological/technical reproducibility in clonal cuSCC
189 genomes (**Supplementary Figure 7a-b**) and no reproducibility in non-clonal normal skin genomes
190 (**Supplementary Figure 7c-d**). Importantly, sequence analysis of the same normal skin SBCapSeq library from
191 independent Ion Torrent runs demonstrated excellent technical reproducibility (**Supplementary Figure 7e-f**),
192 further supporting the robustness of our methodology to detect clonal selection and validating our observation that
193 normal skin does not exhibit clonal selection of SB insertions, despite the high insertion burden. These observations
194 are consistent with the large numbers of mutations observed during chronic UV-IR exposure in human eyelid
195 (Martincorena et al., 2015) and SKH hairless mouse (Chitsazzadeh et al., 2016; Knatko et al., 2017) skin and,
196 together with our SB data, indicates that keratinocytes within phenotypically normal skin can tolerate high
197 background mutagenic insults that may act as a mutational reservoir prior to clonal selection of cooperating driver
198 mutations that permit frank outgrown of cuSCC *in vivo*. It is for this reason that no trunk drivers could be identified
199 from the normal skin specimen genomes. The quantitative, stage-dependent insertional mutation burdens derived
200 from SBCapSeq in cutaneous specimens reveal incredible levels of tumor heterogeneity that has not previously been
201 observed for any SB-driven solid tumor.

202 **cuKA progression to late-stage cuSCC via clonal and sub-clonal intratumor heterogeneity**

203 We identified four skin masses with histologically continuous but distinct regions of cuKA and cuSCC, suggesting for
204 the first time that cuKA likely progresses directly to late-stage invasive cuSCC *in vivo* (**Supplementary Table 19**).
205 We performed SBCapSeq on representative, histologically distinct regions to investigate intratumor heterogeneity
206 and gain insight into the drivers likely to mediate cuKA to cuSCC progression (**Fig. 3a-c**, **Supplementary Figure 8a-**
207 **f**, and **Supplementary Table 20**). Here, the read-depth threshold was adjusted to 200 reads to reflect the lower
208 sequencing depth of the tumor regions. For several loci, we identified insertions at the same nucleotide address
209 across tumor regions from both cuKA and cuSCC in the same mouse, indicating clonal origins of these lesions.
210 Invariably, clonal insertions into *Zmiz1* were identified across each section of all 4 masses (**Fig. 3a**), at three
211 different TA-dinucleotide addresses (**Supplementary Figure 8a-d**), suggesting that *Zmiz1* was under positive
212 selection within the nascent transformed cells that gave rise to the distinct cuKA and cuSCC features. Additional
213 tumor-specific clonal loci include *Fer*, *Rere*, and *Usp34*, all of which were defined in the population analysis. We also
214 found evidence for sub-clonal insertion events identified at the same nucleotide address in some but not all regions
215 within the same tumor, demonstrating the evolution of the tumor. Some genes with sub-clonal insertion events with
216 read depths greater than 200 were defined as drivers in the bulk population, including *Tcf4* and *Pak2, while other*
217 *genes, such as Egr2* and *Rab3gap1*, were not (**Fig. 3b-c** and **Supplementary Figure 8e-f**). These data provide
218 evidence for a common keratinocyte progenitor that acquires SB insertions in genes that promote cuKA and are
219 maintained in cuSCC. This analysis also highlights the intra- and inter-tumor heterogeneity of insertions in SB cuSCC
220 tumors and importantly demonstrates that SB insertion profiles can be used to trace the evolution of cuKA to cuSCC
221 using both the insertion nucleotide address and read depth to define clonal relationships. Finally, this analysis lends
222 insight into potential cooperating drivers in keratinocyte transformation progression.

223 ***Pten cooperation with Chuk in early-stage cuKA***

224 In a previously published study, a surprising insight into genotype-phenotype relationships during keratinocyte
225 transformation was found by comparing the sequencing of early-stage cuKA masses that arose spontaneously in a
226 breast cancer cohort consisting of quadruple heterozygous mice carrying alleles for K5-Cre, Pten-floxed,
227 Onc3(TG.12740), ^{LSL}SBase to phenotypically indistinguishable cuKA masses observed in the SB|P53 cohorts (Rangel
228 et al., 2016; Stransky et al., 2011). In keratinocytes with SB mobilization and heterozygous loss of *Pten*, activating SB
229 insertions into *Zmiz1, Zmiz2,* and *Mamld1* were rarely, if ever, observed. Instead, inactivating insertions into *Chuk*
230 were statistically over-represented and, when they occurred, were mutually exclusive with *Zmiz1* or *Mamld1*
231 insertions, suggesting they may be within the same genetic pathway. Surprisingly, SBCapSeq analysis of normal skin
232 from 9 separate mice with the same quadruple heterozygous genotype revealed many low-read depth insertions
233 into the *Pten* locus, but no insertions in *Chuk* (**Supplementary Table 21-22**), suggesting that selection for *Chuk*
234 inactivation may be dependent upon haploinsufficient loss of *Pten*. In addition, using the datasets within the recently
235 published SBCDDB (Newberg et al., 2017), we found statistically significant evidence for co-occurring loss of *Pten*
236 (either via SB insertion or by breeding in a heterozygous null allele) and SB inactivation of *Chuk* across solid tumor
237 models (Fisher's exact test, $P=2.36 \times 10^{-7}$). Statistical significance for co-occurring mutations in *Pten* and *Chuk* in
238 solid cancers persisted even after removing all cuKA tumors from the analysis (Fisher's exact test, $P=3.78 \times 10^{-16}$),
239 suggesting this relationship may be generally observed in non-keratinocyte tumor lineages. Remarkably, querying
240 for co-occurrence of alterations in *PTEN* and *CHUK* via cBioPortal, across 41,834 patient samples from 170 studies,
241 revealed significant co-occurrence for these two genes in human cancers (Chi-square with Yates correction, $\chi^2=406$,
242 $P<0.001$; **Supplementary figure 4c-d**). Collectively, these data suggest that coordinated loss of both *Pten* and *Chuk*
243 may have a role in initiating and progressing oncogenic transformation in keratinocyte as well as non-keratinocyte
244 lineages. Given the recurring inactivating SB insertions in *Chuk* and activating *Zmiz1* observed in our *Pten*-loss
245 sensitized mice (**Supplementary Figure 4a**), it is curious that a recent separate *Pten*-sensitized SB screen for
246 cooperating TSGs failed to identify *Chuk* or *Zmiz1* entirely (de la Rosa et al., 2017). One possibility for this difference
247 is that the reported screen used a single-copy SB allele, while our screen used multi-copy SB alleles with multiple
248 independent genic insertion events per tumor, and the lack of *Chuk* or *Zmiz1* insertion events may suggest that
249 positive selection for variant *Chuk* or *Zmiz1* alleles in haploinsufficient *Pten*-mutant skin cells may require one or
250 more additional cooperating drivers.

251 **Comparative oncogenomic meta-analysis**

252 One hundred-seven of the SB candidate Trunk driver genes are direct orthologs of human cancer genes found in the
253 Cancer Gene Census (GRCh38 COSMICv86) (Futreal et al., 2004), a growing catalogue of mutations causally
254 implicated in cancer, which is an enrichment greater than expected by chance ($\chi^2=60.75$ with Yates correction,
255 $P<0.0001$; **Supplementary Tables 23–27**). Another 64 SB candidate driver genes, including *ZMIZ1*, *ZMIZ2*, and
256 *KMT2C,* have human orthologs with recurrent non-silent mutations identified by exome sequencing (Lee et al., 2014;
257 Pickering et al., 2014; South et al., 2014) of 68 cuSCC genomes ($\chi^2=23.70$ with Yates correction, $P<0.0001$). We

258 conclude from these findings that SB identifies the same genes mutated by the diversity of mutational processes
259 operating within end-stage human cuSCC genomes.

260 To gain insight into the biological functions of the candidate discovery drivers, we used a curated subset of the
261 Enrichr analysis platform (Chen et al., 2013; Kuleshov et al., 2016) gene-set libraries for pathways and gene
262 ontologies (see **Methods**) to perform biological pathway and process enrichment analysis from the cutaneous
263 specimen cohorts. Here we integrated the drivers identified by 454 and SBCapSeq for cuSCC in SB|Onc3 mice and
264 derived a composite list of cuKA drivers from two different cuKA datasets, one from cuKAs derived from SB|Onc3
265 mice sequenced by SBCapSeq and the second from cuKAs isolated from a previously published mouse model with
266 K5-Cre; Pten$^{CKO/+}$ and SB|Onc2 or SB|Onc3 mice (Rangel et al., 2016; Stransky et al., 2011) sequenced by 454. Using
267 these integrated driver lists, we discovered statistically significant enrichment of recurrent SB candidate drivers in
268 many signaling pathways and biological processes previously associated with human SCC, including EGFR, MAPK,
269 NOTCH, and WNT-TGFβ (**Supplementary Figure 9a-i** and **Supplementary Table 28**).

270 Several of the top 30 most frequent drivers in SB cuSCC are key players in multiple pathways, suggesting that SB
271 perturbation of multi-functional genes drives cuSCC. An oncoprint aligns the conserved drivers between SB cuSCC,
272 SB cuKA and mutated human homologs in cuSCC that map to enriched pathways (**Fig .4**). Further, the SB|cuSCC
273 Trunk drivers encode proteins with significantly more known protein-protein interactions than expected by chance
274 (P=8.55 × 10^{-6}, STRING enrichment analysis; number of nodes: 144, number of edges: 124, expected number of
275 edges: 82) as do the SB|cuKA trunk drivers (P=6.93 × 10^{-5}, STRING enrichment analysis; number of nodes: 37,
276 number of edges: 14, expected number of edges: 4; **Supplementary Figure 10**). PPIs among drivers identified in SB-
277 driven cuKAs suggests that perturbation of NOTCH signaling may be an initiating or acquired event in cuKA. Notably,
278 we also observed that all SB|cuSCC tumor genomes have at least one defined driver involved in chromatin
279 modification (consisting of chromatin and histone modification enzymes) and/or Hippo pathway genes (**Fig. 4** and
280 **Supplementary Table 28**). Comparative genomic integration of the SB defined cuSCC and cuKA drivers with human
281 cuSCC genomes reveals that while differences in the incidence of individual genes across cohorts can vary
282 substantially, the biological pathways are conserved (**Supplementary Figure 9a**). These differences may be due in
283 part to the types of alterations produced by SB transposition (gene activation/inactivation) versus UV-IR
284 (silent/non-silent mutations) or, alternatively, may reflect the complexity of the systems of genetic networks
285 (Nadeau and Dudley, 2011) that exists between driver genes and keratinocyte transformation phenotypes across
286 species.

287 **SB-driven gene-expression alterations in cuSCC genomes**

288 The most highly mutated oncogenic drivers in cuSCC genomes were the transcriptional co-factors *Zmiz1* and *Zmiz2*,
289 which are collectively and mutually exclusively mutated in 71% of end-stage SB|cuSCC genomes. High read depth
290 insertions are clustered into a single intron of each gene and occur in the same transcriptional orientation, where
291 they are predicted to drive expression of a truncated chimeric transcript containing the oncogenic MIZ domain (**Fig.**
292 **5a**). We confirmed this prediction by RT-PCR (not shown) and normalized microarray data (**Fig. 5b**) analysis, where
293 cuSCC tumors with *Zmiz1* insertions were found to have significantly higher mRNA abundance than tumors without

294 *Zmiz1* insertions (**Fig. 5b**). We also observed a positive correlation between high-read-depth SB insertions among
295 the 11 tumors harboring trunk *Zmiz1* insertions and *Zmiz1* gene expression (**Fig. 5c**).

296 Since ZMIZ1$^{\Delta N185}$ retains the DNA binding and transactivation domains, we reasoned that the *ZMIZ1*$^{\Delta N185}$ isoform
297 may drive cuSCC via alterations in gene expression. To identify genes that were differentially expressed between SB
298 cuSCC tumors with and without insertions in *Zmiz1*, we used Rstudio to create a Shiny app that used the normalized
299 microarray data (see **Supplementary Table 29-30**), to generate a 25- (**Fig. 6d** and **Supplementary Fig. 11a**), 50-
300 (**Supplementary Fig. 11b**), and 96-gene (**Supplementary Table 31**) differential expression signature (P<0.05, P_{FDR-}
301 $_{corrected}$ <0.12) from the top differentially expressed genes. These genes demonstrated enrichment for biological
302 pathways and processes related to gene expression, RNA processing and metabolism (**Supplementary Table 32-**
303 **33**).

304 We also conducted a secondary differential analysis in R and identified 289 genes that were differentially expressed
305 (P<0.05, $P_{FDR-corrected}$ <0.2) between the *Zmiz1*-positive and *Zmiz1*-negative SB-driven cuSCC tumor groups
306 (**Supplementary Table 34**). As expected, *Zmiz1* expression was significantly higher in tumors with activating
307 insertions. Among the differentially expressed genes, the expression levels of *Serpinb2*, *Spink12*, and *Stfa3* were
308 individually statistically significantly higher in cuSCC genomes with activating *Zmiz1* insertions than in cuSCC
309 genomes without *Zmiz1* insertions (**Supplementary Fig. 11c**), and *Mtdh*, *Anp32e*, *Ecd*, *Dcun1d5*, and *Kdm7a* were
310 also independently identified as cuSCC Progression Drivers. These statistically defined co-driver expression
311 relationships indicate that these cooperating mutagenic insertions may drive keratinocyte transformation and/or
312 progression in this mouse model of human cuSCC.

313 To identify potential oncogenic pathways and processes that were enriched among the 289 significantly
314 differentially expressed genes (DEGs), we conducted enrichment analysis using Enrichr analysis platform. Based on
315 the KEGG database, only the interleukin-17 signaling was significantly enriched among the DEGs, suggesting that
316 *ZMIZ1*$^{\Delta N185}$ isoform may not alter oncogenic pathways at the transcript level. However, gene ontology analysis
317 revealed a significant enrichment for biological processes involved in epidermis differentiation and development,
318 suggesting that *ZMIZ1*$^{\Delta N185}$ isoform drives the dysregulation of keratinocyte development during cuSCC
319 development. We observed enrichment of several paralog gene families, including keratin (*Krt1/10/15/16* and
320 *Krtdap*), stratum corneum-associated structural late cornified envelope (*Lce1f* and *Lce3c*), and keratinocyte envelop
321 cross-linking small proline rich (*Sprr1a* and *Sprr1b*) proteins, and key components of the epidermal differentiation
322 complex (EDC) (Kypriotou et al., 2012), including *S100a8*, *S100a9*, *Ivl*, *Flg* (**Table 1**). Based on these finding, we
323 asked whether these EDC genes are altered in human cuSCC. Using a published gene expression dataset (GSE32628
324 (Hameetman et al., 2013)), we demonstrate that EDC gene transcripts *IVL*, *LCE3C*, *S100A8*, *S100A9*, *SPRR1B*, and
325 *SPRR2B* were significantly upregulated in human AK and cuSCC, relative to matched normal epidermis
326 (**Supplementary Figure 12**)**.** Importantly, increased expression of EDC genes has been reported in human cuSCC
327 (Hudson et al., 2010). Taken together, our microarray gene expression analysis demonstrates that the *ZMIZ1*$^{\Delta N185}$
328 isoform is a critical mediator of gene expression changes essential for early keratinocyte transformation and
329 sustained during cuSCC progression to end-stage tumors.

330 To further analyze the impact of clonally selected trunk SB insertions on gene-level and transcript-level expression,
331 we sequenced ribo-depleted, whole-transcriptome RNA-seq (wtRNA-seq) libraries created from total RNA isolated
332 from 7 cuSCC genomes (**Supplementary Table 35**) with high-read depth activating SB insertions into either *Zmiz1*
333 (n=6) or *Zmiz2* (n=1) to identify SBfusion reads (Mann et al., 2016b). The wtRNA-seq reads contained spliced
334 chimeric transposon splice-donor or splice-acceptor sequences fused to adjacent known exons in the mouse
335 genome. In total, we identified 2,906 SBfusion wtRNA-seq reads from 889 unique RefSeq genes (**Supplementary**
336 **Table 35**), including 154 progression and 27 trunk cuSCC drivers, respectively (**Supplementary Table 37-38**),
337 which is substantially more than expected by chance (Chi-square test with Yates correction, χ^2= 159, P<0.0001).
338 Thirty-three genes with SBfusion events supported by ≥5 SBfusion reads, include our top trunk drivers *Zmiz1* and
339 *Zmiz2* and several keratins (**Fig. 6a-c**). We also observed a positive correlation between high-read-depth SB
340 insertions among the 6 cuSCC tumors harboring trunk *Zmiz1* insertions and exhibiting high *ZMIZ1* gene expression
341 (**Fig. 6d**). We next selected 10 genes with high-read depth SBCapSeq insertion sites and SBfusion reads detected by
342 RNA-seq within the 7 cuSCCs to investigate potential alterations in gene expression due to clonal SB insertion
343 events, including 4 activating (*Zmiz1*, *Zmiz2*, *Wnt7b*, and *Epha1*) and 6 inactivating (*Nf1*, *Adam10*, *Oxsr1*, *Map2k4*,
344 *Cul5*, and *Cep350*) SB insertion events (**Fig. 6e-f**, **Supplementary Figure 13a-f**, and **Supplementary Figure 14a-f**).

345 Importantly, in each tumor, SBfusion reads containing either *SB-Zmiz1* or *SB-Zmiz2* spliced-exons were represented
346 in the top 2% of SBfusion transcripts and were supported by multiple independent SBfusion reads (**Supplementary**
347 **Table 39-40**). *Zmiz1* or *Zmiz2* gene expression was significantly increased within each of the 7 cuSCC genomes
348 (≧100 and up to 300 times normal expression levels) and correlated with a highest SBCapSeq read depth sites with
349 activating SB insertions and activating SBfusion reads (**Figure 6e** and **Supplementary Figure 13a-d**). This strongly
350 suggests that the SB insertion events driving either *Zmiz1* or *Zmiz2* have been clonally selected to increase
351 oncogenic expression of the N-terminal truncated forms during keratinocyte transformation and cuSCC progression
352 *in vivo*. Similarly, *Epha1* and *Wnt7b* had high SBCapSeq read depth activating SB insertions and wtRNA-seq SBfusion
353 reads and significantly higher gene expression compared to basal expression levels (≧100 times) observed in
354 tumors without insertions or SBfusion reads (**Figure 6e** and **Supplementary Figure 13e-f**). Since SB insertions into
355 these genes were not recurrent at the population level, *Epha1* and *Wnt7b* were not identified as statistically
356 significant drivers in cuSCC. However, the integrated genomic and transcriptomic analysis strongly suggests that
357 each of these genes may have a 'private' role in driving tumor progression in the genomes in which they occur with
358 elevated expression levels. Both genes encode proteins that are linked to cancer hallmarks — *Epha1* is an ephrin
359 receptor subfamily of the protein-tyrosine kinase family (Potla et al., 2002) and *Wnt7b* is a WNT family member
360 (Huguet et al., 1994). For the 6 predicted inactivated genes, we observed reduced gene expression that correlated
361 with high SBCapSeq read depth of inactivating SB insertions and truncated SBfusion reads (**Figure 6f** and
362 **Supplementary Figures 14a-f**). *Adam10*, *Cul5*, and *Nf1* are Trunk Drivers and *Map2k4* is a Progression Driver in
363 cuSCC cohorts, further suggesting these events contribute to cuSCC progression *in vivo*. *Oxsr1* and *Cep350* were
364 private selected events in individual tumors that may operate in cuSCC progression within the tumors that harbor
365 clonal insertion events. Taken together, these data demonstrate that integrated genomic and transcriptome analyses
366 may provide a means to discover significant cooperating drivers in cuSCC genomes from bulk specimens.

367 **Clinical relevance of SB|cuSCC driver orthologs in human SCC**

368 The mammalian paralogous genes *Zmiz1* and *Zmiz2* are mutually exclusive trunk drivers activated in three quarters
369 of SB-driven cuSCC genomes in an SB-driven cuSCCs, suggesting that *Zmiz* proteins are proto-oncogenes in
370 cutaneous squamous keratinocyte transformation *in vivo*. *Zmiz1* and *Zmiz2* encode transcription factor proteins that
371 share core c-terminal features, including androgen receptor (AR) binding domain, proline-rich binding domain, and
372 MIZ-type zinc finger PIAS (protein inhibitor of activated STAT) domains, that coordinately act to both interact with
373 and regulate the activity of various cancer-associated proteins (Beliakoff and Sun, 2006; Huang et al., 2005; Kahyo et
374 al., 2001; Lee et al., 2013; Megidish et al., 2002; Schmidt and Muller, 2002, 2003; Sharma et al., 2003; Soler et al.,
375 2008). SB activating trunk insertions in *Zmiz1* or *Zmiz2* are unique to SB-induced cuSCC tumors and are predicted to
376 activate gene expression by driving expression of specific N-terminal truncated forms of ZMIZ1$^{\Delta N185}$ and ZMIZ2$^{\Delta N184}$,
377 a closely related paralog (**Figure 5a**). The likelihood of achieving unique, sense-strand oriented SB insertion events
378 into two related genes at the same protein coding region on two different chromosomes in dozens of tumors from
379 independent mice by chance is incredibly small (P<0.0001), suggesting both trunk driver ZMIZ oncoproteins
380 contribute directly to keratinocyte initiation and tumor progression *in vivo*. The ZMIZ1$^{\Delta N185}$ and ZMIZ2$^{\Delta N184}$ onco-
381 proteins share 62% sequence identity, retain the zinc finger PIAS, proline-rich transactivation, and AR binding
382 domains shared by both peptides. Importantly, both oncoproteins lack the N-terminus in the full-length protein,
383 which is thought to inhibit the intrinsic transcriptional activity of a C-terminal proline-rich transactivation domain
384 (Sharma et al., 2003). *ZMIZ1 (RAI17/ZIMP10)* and *ZMIZ2 (ZIMP7)* are collectively mutated or amplified in one third
385 of human cuSCC (Pickering et al., 2014; South et al., 2014) and 5% of hnSCC (Cancer Genome Atlas, 2015) genomes.
386 To investigate their oncogenic potential in human keratinocyte transformation, we performed a metagene
387 expression analysis of *ZMIZ1* and *ZMIZ2* in human TCGA head and neck SCC (hnSCC) datasets for which gene
388 expression (RNA-seq) and patient survival outcomes are available. A *ZMIZ1* metagene, with a correlation threshold
389 of 0.65, identifies a 23-gene signature that links high expression with poor outcomes in human hnSCC patients (Cox
390 Proportional Hazards Regression, P=0.0195; Log-rank Test, P=0.028 at 50% quintile; **Figure 5e; Supplementary**
391 **Figure 15a–c**).

392 Ectopic over-expression of N-terminus truncated ZMIZ1 in mouse keratinocytes leads to epithelial hyper-
393 proliferation but is not sufficient for progression to late-stage cuSCC (Rogers et al., 2013), suggesting that
394 cooperating mutations are required for frank cuSCC progression *in vivo*. Biological pathway enrichment in the *SB*
395 cuSCC genomes revealed several cancer-associated pathways, including the EGFR, NOTCH, and WNT signaling
396 pathways. However, we noted that all *Zmiz*-mutated cuSCC genomes contained one or more inactivating SB
397 mutations in chromatin-remodeling genes (**Supplementary Figure 9g**), suggesting this class of mutated tumor
398 suppressor genes may contribute significantly in driving cuSCC progression. Four of the top-ranked chromatin
399 remodeling trunk driver orthologs were also found to be significantly mutated in >15% of human cuSCC genomes
400 (Pickering et al., 2014; South et al., 2014). Importantly, as observed in mouse tumors, mutations in the orthologs
401 *KMT2C (MLL3)*, *ARID1B (BAF250B)*, *NCOA2 (KAT13C)*, and *CREBBP (CBP/KAT3A)* were also found in the same cuSCC
402 genomes harboring *ZMIZ* mutations (**Figure 2**), suggesting they might cooperate during human keratinocyte
403 transformation. Previously, *KMT2C* (*MLL3*) mutations were found to be associated with poor outcomes in aggressive

404 cuSCC (Pickering et al., 2014), suggesting mutations in this gene may be a biomarker for advanced stage disease. To
405 extend this result, we used survival information and RNA-seq expression from a hnSCC dataset, and observed that a
406 *KMT2C* metagene, with a correlation threshold of 0.66, identifies a 176-gene signature that links high expression
407 with poor outcomes in human hnSCC patients (Cox Proportional Hazards Regression, P=0.0498; Log-rank Test,
408 P=0.0093 at 75% quintile; **Supplementary Figure 16**). Similarly, consistent with hippo signaling in cuSCC, we
409 identified a 20-gene signature consisting of Hippo pathway orthologs predictive of patient outcome (Cox
410 Proportional Hazards Regression, P<0.0001; Log-rank Test, P=0.0002 at 50% quintile; **Supplementary Figure 17**).

411 **Chromatin modifying tumor suppressors drive human keratinocyte initiation**

412 Chromatin remodeling is one of the most frequently inactivated biological processes in SCC, and has been
413 demonstrated to be involved in either tumor initiation or progression in multiple cancer types (Andricovich et al.,
414 2018; Cho et al., 2018; Dhar et al., 2018; Jia et al., 2018; Mathur et al., 2017; Yu et al., 2016). In our SB cuSCC screen,
415 chromatin remodeling was the top biological process that was significantly enriched (**Supplementary Table 28**).
416 However, it remains unclear how expression alteration of chromatin remodeling genes feature in the cuSCC disease
417 trajectory remains unclear. We prioritized three tumor suppressor genes involved in histone modification (*KMT2C*
418 and *CREBBP*) or transcriptional regulation (*NCOA2*) with high concordance between SB-cuSCC and human cuSCC to
419 study their roles in keratinocyte transformation using immortalized and minimally transformed HaCaT keratinocyte
420 cell line (Boukamp et al., 1988). First, we confirmed expression of each gene within HaCaT cells (data not shown)
421 and established a baseline colony formation assay using enforced expression of oncogenic KRAS-G12D in HaCaT
422 cells (**Supplementary Figure 18**). Next, we used shRNAs targeting these genes to assess cellular transformation *in*
423 *vitro* as a surrogate of tumor initiation, and disease progression *in vitro* and *in vivo* (**Fig. 7**, 1-factor ANOVA, P<0.05).
424 To investigate whether knockdown of *KMT2C*, *CREBBP* or *NCOA2* can transform keratinocytes, we used HaCaT cells
425 to stably knock down expression of target genes. Forced depletion of *CREBBP* significantly increased cellular
426 proliferation relative to control, and a similar trend was observed with *KMT2C* knockdown (**Figure 7a**). No
427 significant difference in proliferation was observed between *NCOA2* knockdown and control (**Figure 7a**). To more
428 stringently investigate transformation, we assessed anchorage-independent growth, a hallmark of cellular
429 transformation, using a soft agar colony formation assay. Knockdown of *CREBBP* or *KMT2C* expression resulted in
430 significantly larger size and increased number of colonies, relative to control (**Figure 7b**), while no significant
431 difference was observed between *NCOA2* knockdown and control (**Figure 7b**). Taken together, these data suggest
432 that knockdown of *CREBBP* or *KMT2C* can transform the immortalized HaCaT cell system *in vitro*. Finally, we tested
433 whether *KMT2C*, *CREBBP*, or *NCOA2* single gene knockdown in HaCaT cells can initiate the formation of masses *in*
434 *vivo*. Cells were injected into flanks of immunodeficient *NSG* mice, and mass formation was monitored for 9 weeks.
435 At endpoint, no significant difference in tumor volume was observed between control and *shKMT2C, shCREBBP* or
436 *shNCOA2* (**Figure 7c-d**), suggesting that expression changes in more than one driver gene may be needed for *in vivo*
437 transformation.

438 **Functional validation of SB cutaneous candidate drivers in human SCC cell lines**

439 To investigate the role of SB-driven cuSCC drivers to promote cuSCC progression and provide functional evidence

440 that the SB candidate drivers are important for human SCC progression, we prioritized 12 top-ranked SB driver
441 orthologs from cuKA or cuSCC datasets (*CHUK*, *CUL1*, *EHMT1*, *FBXO11*, *KMT2C*, *LARP4B*, *MAMLD1*, *NCOA2*, *NCOR2*,
442 *NOTCH1*, *PTEN*, and *ZMIZ1*) for in vitro validation. Ten are predicted tumor suppressor drivers (*CHUK*, *CUL1*, *EHMT1*,
443 *FBXO11*, *KMT2C*, *LARP4B*, *NCOA2*, *NCOR2*, *NOTCH1*, and *PTEN*) that are mutated in human cuSCC and commonly
444 identified as drivers in other SB mutagenesis screens reported in the SBCDDB (Newberg et al., 2018c) and 2
445 predicted oncogene drivers, *ZMIZ1* and *MAMLD1*, that are mutated in human cuSCC but are not oncogenic drivers in
446 other SB mouse models of human cancers (Newberg et al., 2018c). We performed short-term siRNA-mediated gene
447 knockdown (3 siRNAs per gene, **Supplementary Table 41**) in two well-characterized SCC cell lines (A431 and
448 SCC13) and one lung SCC cell line (SW900) that each harbor *TP53* mutations. Cell lines with confirmed knockdown
449 of driver target gene expression were assayed for changes in proliferation rates. All 12 drivers tested had at least
450 one siRNA that significantly altered cellular proliferation compared to a scrambled control siRNA in at least one cell
451 line (one-factor ANOVA with multiple testing correction, FDR adjusted-P<0.1; **Supplementary Figure 19** and
452 **Supplementary Table 42**). We observed a modest but significant decrease in cellular proliferation with ZMIZ1
453 depletion and a dramatic decrease in cell proliferation in A431 cells with ZMIZ2 depletion. Likewise, knockdown of
454 the predicted tumor suppressive drivers consistently increased cell proliferation, supporting the role of these genes
455 in human SCCs (**Supplementary Table 42**). *CHUK*, *NOTCH1*, and *PTEN* validated in all three cell lines. Increased
456 proliferation in A431 cells, consistent with previously reported data (Leonard et al., 2011).

457 *NCOA2* validated in both cuSCC cell lines, and *FBXO11*, *KMT2C*, *LARP4B*, *MAMLD1* and *ZMIZ1* validated in one cuSCC
458 and the lung cell line. *CUL1*, *EHMT1*, *NCOR2* were significant only in the SW900 lung cell line. si*PTEN* knockdown
459 Additionally, 3 independent siRNAs targeting *ZMIZ2*, an independent driver and paralog of *ZMIZ1*, were tested and
460 found to grow significantly slower compared to a scrambled control siRNA in the A431 cell line (**Supplementary**
461 **Figure 19**).

462 To both confirm and extend these findings, we next stably depleted each of these 12 drivers in in A431 cells using
463 lentiviral pools containing three non-overlapping, short hairpin RNAs (shRNAs) to individually knockdown each of
464 the 13 target candidate drivers and one non-targeting control delivered to silence target gene expression in A431
465 cells (**Supplementary Figure 20**) which was verified by qPCR (**Figure 8a**). Seven of the drivers displayed
466 statistically significant changes in proliferation of A431 cells: *shZMIZ1*, *shZMIZ2*, and *shMAMLD1* decreased
467 proliferation; and *shKMTC2*, *shNCOA2, shFBXO11* and *shCHUK*, accelerated proliferation, consistent with their
468 predicted roles as proto-oncogenes, while shRNA-mediated knockdown of tumor suppressors KMTC2, NCOA2,
469 FBXO11 and CHUK, accelerated proliferation (**Figure 8a-c**, two-factor ANOVA with multiple comparisons, FDR-
470 adjusted-P<0.05). While knockdown of the *ZMIZ* paralogs in the A431 cell line demonstrated significant reduction in
471 proliferation, we sought to verify that this observation was not cell-line specific. Using the COLO16 cell line, we
472 achieved stable knockdown of *ZMIZ* expression, and observed decreased proliferation, relative to control. Taken
473 together, our functional shRNA screen has yielded strong evidence for the *ZMIZ* oncogenes identified from our SB
474 screen to be involved in human cuSCC progression *in vitro*, and prioritized several tumor suppressor genes for
475 further characterization.

476 **Chromatin modifying tumor suppressors drive human keratinocyte initiation and cuSCC progression**

477 Since we demonstrated that knockdown of *KMT2C* or *NCOA2* increases the proliferation rates of A431 cell line *in*
478 *vitro*, we asked whether this finding could be extended to other cell lines. Consistent with the observation made in
479 A431 cells, we observed increased proliferation in COLO16 and SCC13 cells with decreased expression of *KMT2C or*
480 *NCOA2*, relative to control (**Figure 8**, $P<0.05$, 1-factor ANOVA). We then determined whether *KMT2C and NCOA2* are
481 cuSCC TSGs *in vivo*, and performed xenograft assays by subcutaneous injection of cells for each of the three cuSCC
482 cell lines into *NSG* immunodeficient mice. All three cuSCC cell lines demonstrated enhanced *in vivo* tumor growth
483 with *KMT2C*, *NCOA2,* or *CREBBP* depletion compared to cells with a non-targeting control shRNA (**Figure 8d-i**, 1-
484 factor ANOVA, $q<0.05$ and **Supplementary Figure 21-22**), providing functional evidence that *KMT2C and NCOA2*,
485 but not *CREBBP*, are human TSGs that promote *in vivo* cuSCC progression.

486 Finally, we explored the mechanistic basis of how decreased expression of *KMT2C* or *NCOA2* may progress cuSCC.
487 The MAPK pathway has been shown to be operational in cuSCC (Lambert et al., 2014), and melanoma patients
488 receiving BRAF inhibitor can develop cuSCC via hyperactivation of the MAPK pathway (Zhang et al., 2015). Since
489 *KMT2C* and *NCOA2* can dynamically regulate global changes in gene expression, we hypothesized that knockdown of
490 these two genes progress cuSCC via altering the MAPK flux. To test this, we immunoblotted for total and
491 phosphorylated ERK (phospho-ERK) expression in our stable cell systems. In A431 and SCC13 cell lines with high
492 basal phospho-ERK expression, knockdown of *KMT2C* or *NCOA2* did not alter phospho-ERK levels. In contrast, in
493 COLO16 cell line with low phospho-ERK expression in control, knockdown of either *KMT2C* or *NCOA2* resulted
494 increased ERK activation (**Figure 8j**). We pursued this finding further and used the flash-frozen tissues from our
495 xenograft assay. Whole cell lysates from the three largest tumors for each condition (i.e., NTC, *shKMT2C*, *shNCOA2*)
496 were harvested, and probed for phospho-ERK expression (**Figure 8k**). Consistent with our *in vitro* data, we
497 observed a trend of increased phospho-ERK expression in tissues with *KMT2C* or *NCOA2* knockdown. Taken
498 together, these data suggest that increased MAPK signaling may be one mechanism through which *KMT2C* or *NCOA2*
499 knockdown drives to cuSCC progression.

500 **DISCUSSION**

501 In this study, we sought to define the cooperating genetic events required for keratinocyte transformation and
502 progression to frank cuSCC using the *SB* transposon mutagenesis system *in vivo*. Using an *Actb-Cre* transgene to
503 activate systemic expression of the transposon, we observed tumor lesions in multiple organs. The kinetics of
504 tumorigenesis were further accelerated in mice harboring either null ($Trp53^{KO/+}$) (Jonkers et al., 2001) or recurrent
505 point mutant ($Trp53^{R172H/+}$) (Olive et al., 2004) alleles. This observation is consistent with previous studies showing
506 that missense mutations in *Trp53* promote malignant transformation by abrogating both the tumor suppressive and
507 oncogenic (gain-of-function and/or dominant negative) functions of Trp53 *in vivo* (Doyle et al., 2010; Hanel et al.,
508 2013; Lang et al., 2004; Morton et al., 2010; Olive et al., 2004). While the SB|Onc3 mice developed a spectrum of
509 tumor types, including of the hematopoietic lineage, we observed an enrichment for late-stage, well-differentiated
510 cuSCC.

511 Despite the high curative rate for cuSCC, there is a significant population of patients who develop advanced disease,
512 for which there are no defined therapeutic targets. Due to the high UV-induced mutation burden, combined with the
513 scarcity of clinical samples, it is challenging to define the recurrent and druggable targets within these tumors. To
514 address this need, we isolated early and advanced lesions from a SB mouse model of cuSCC to define recurrent
515 drivers and potential new avenues for therapeutic intervention. From our in-depth sequencing analysis of SB
516 insertions in advanced cuSCC genomes, we identified clonal and sub clonal drivers involved in the progression and
517 maintenance of cuSCC. 64 drivers have human orthologs with non-silent mutations in human cuSCC genomes.
518 Intriguingly, the majority of drivers in SB-driven cuSCC lesions were predicted tumor suppressor genes, suggesting
519 that cumulative loss of these genes is a central feature of cuSCC. We showed that a few key TSGs, including *KMT2C*
520 are differentially expressed in cuSCC with prognostic implications for patients with hnSCC.

521 We subsequently focused on understanding the genetic events required for keratinocyte transformation. Our
522 SBCapSeq and SB Driver analysis identified 349 CCDs, regardless of *Trp53* status, of which 107 have been implicated
523 in human cancers. Intriguingly, we observed that the SB-driven cuSCC lesions were driven almost exclusively by SB
524 inactivating insertions in tumor suppressor genes, suggesting that cumulative loss of these TSG drivers is a central
525 feature of cuSCC. Indeed, recent analysis from our lab, using a complimentary promoterless SB transposon forward
526 genetic screen, revealed that tumor suppressor drivers alone can lead to keratinocyte initiation and cuSCC
527 progression *in vivo* (Aiderus et al., 2020). Furthermore, in cuSCC lesions that arose from K5-Cre; *Pten*$^{CKO/+}$ model
528 previously reported, our SB Driver Analysis workflow revealed inactivating insertion patterns in all 300 CCDs, which
529 corroborates the Actin-Cre driven model utilized in our study. Notably, in the single allele *Pten* loss model, we
530 identified inactivation in *Chuk* (also known as *Ikka*) as a significant driver of cuSCC, and this gene has previously
531 been reported to have a tumor suppressive role in cuSCC via regulating the EGFR signaling axis (Liu et al., 2008).
532 However, since *Chuk* inactivation only occurs in a haploinsufficient *Pten* background, these data suggest that
533 inactivation of these genes alters different pathways that cooperate to drive cuSCC. Tantalizingly, PanCancer
534 analysis of over 40,000 tumors from multiple tissue types using cBioPortal revealed a significant co-occurrence of
535 both genes, suggesting that the association between *Pten* and *Chuk* may be extended to cancer derived from
536 different cellular lineages. Our data provide a rich resource to rationally identify, prioritize, and credential likely
537 cooperating mutational events in cuSCC (**Fig. 3**).

538 Our report is the first to identify *Zmiz2* as a significant candidate cancer gene in any SB study and the first to
539 discover mutual exclusivity among activated *Zmiz1*, *Zmiz2*, and *Mamld1* in early keratinocyte transformation and
540 cuSCC progression. Insertions into *Zmiz1* (Dupuy et al., 2009; Friedel et al., 2013; Rogers et al., 2013) and *Mamld1*
541 (Friedel et al., 2013) have been previously observed in skin tumors induced by transposon insertional mutagenesis.
542 The possible redundancy of ZMIZ oncoproteins in the keratinocyte transformation program, by mutual exclusivity of
543 trunk insertions within cuSCC genomes, strongly suggests that the ZMIZ1$^{\Delta N185}$ and ZMIZ2$^{\Delta N184}$ oncoproteins may
544 have common neomorphic properties that contribute to the hallmarks of skin cancer. The work of Rogers *et al*.
545 (Rogers et al., 2013), and more recently Mathios *et al.* (Mathios et al., 2019), on the dynamic regulation of *ZMIZ1*
546 expression in human cancer, together with our finding that *ZMIZ1* and *ZMIZ2* are mutated with at least on non-silent
547 alteration in at least one third of published human cuSCC genomes (Pickering et al., 2014; South et al., 2014) and,

548 suggest that this gene may be an important initiating trunk mutation in human keratinocyte initiation and cuSCC
549 progression. Transcriptome analysis on SB-driven cuSCCs with high read depth insertions in *Zmiz1* or *Zmiz2* to
550 detected SB transposon fusions to both drivers. Of note, the N-terminal truncated versions of both *Zmiz* genes were
551 selected in a mutually exclusive manner. The Zmiz1$^{\Delta N185}$ and Zmiz2$^{\Delta N184}$ onco-proteins share 62% sequence identity,
552 and share the zinc finger PIAS, proline-rich transactivation, and AR binding domains. Importantly, both oncoproteins
553 lack the N-terminus in the full-length protein, which is thought to inhibit the intrinsic transcriptional activity C-
554 terminal proline-rich transactivation domain (Sharma et al., 2003). *ZMIZ1* and *ZMIZ2* are mutated in one third of
555 human cuSCC genomes sequenced to date (Pickering et al., 2014; South et al., 2014), however, its functional
556 significance remains unclear. *Zmiz1* has been shown to be involved in T cell development and leukemogenesis via
557 interacting with NOTCH1 (Pinnell et al., 2015), but how it is involved in the tumorigenesis of other cell lineages is
558 unclear. We have provided two preliminary lines of evidence with regards to how *ZMIZ* may be involved in
559 squamous cancers. First, shRNA-mediated knockdown of *ZMIZ1* and *ZMIZ2* in A431 and COLO16 human cuSCC cell
560 lines significantly decreased their growth rates, compared to non-targeting controls. While ZMIZ1 overexpression
561 has been shown to sufficient for disease initiation, our cell proliferation data suggests that pathways regulated by
562 ZMIZ1/2 continue to be active in advanced disease. Further, the association of high expression of a *ZMIZ*-centric
563 metagene derived from TCGA hnSCC cohort, with poor patient survival supports an oncogenic role for *ZMIZ* in
564 tumors of squamous lineage.

565 We identified potential signaling pathways, such as EGFR, NOTCH and WNT, that may cooperate with ZMIZ to drive
566 cuSCC *in vivo*. Intriguingly, all SB cuSCC tumors with *Zmiz1/2* insertions had inactivating insertions in at least one
567 gene involved in chromatin remodeling, which suggests that changes in gene expression via epigenetic regulation
568 are necessary to progress a benign lesion to frank cuSCC. Some examples of the chromatin remodeler genes that
569 were recurrently altered in our screen include members of the complex of proteins associated with Set1 (COMPASS)
570 such as *Kmt2c* and *Kdm6a*, and the switch/sucrose non-fermenting (SWI/SNF) such as *Arid1b, Arid1a* and *Pbrm1*.
571 Although the COMPASS and SWI/SNF pathway members have been reported to be mutated in various cancers
572 (Gozdecka et al., 2018; Kadoch et al., 2013; Wang et al., 2018), whether alterations in one or cooperation of both
573 pathways is required for the etiology of the disease remains unclear. Our functional data showing that knockdown of
574 *KMT2C,* a key molecule in the COMPASS complex, in human cuSCC cell lines accelerated *in vitro* proliferation and *in*
575 *vivo* xenograft growth, supporting a role for this gene as a tumor suppressor in advanced disease. Of clinical
576 significance, several recent studies have highlighted potential dependencies amenable to pharmacologic inhibition
577 in cancer cells with aberrations in COMPASS members (Wang et al., 2018).

578 While mutations in chromatin remodelers have been reported in the literature, only recently have studies provided
579 functional evidence that this family of genes can function as tumor suppressor in the initiation and progression of
580 different cancers. Our *in vivo* screen identified inactivation of at least one gene involved in chromatin remodeling in
581 every tumor, suggesting that this family of genes may have an important role at some stage during cuSCC genesis.
582 Our functional data suggests that *CREBBP* is involved in the earlier stages of keratinocyte transformation, while
583 decreased *NCOA2* expression can drive disease progression. Interestingly, *KMT2C* knockdown drives keratinocyte
584 transformation or cuSCC progression in different cell systems utilized. Finally, while ongoing work is focused on

585 elucidating the mechanistic basis of how alterations in chromatin remodeler gene expression affect cuSCC biology,
586 we demonstrate that one route of cuSCC progression via decreased *NCOA2* or *KMT2C* expression is activation of the
587 MAPK pathway. Notably, this was observed in one of three cell lines assessed, suggesting that depending on the
588 genetic background, chromatin remodelers influence disease progression via different mechanisms.

589 In a similar genetic screen for myeloid leukemia drivers (Mann et al., 2016b; Newberg et al., 2018a), we identified
590 *Ncoa2* as an oncogene and *Zmiz1* as a tumor suppressor, in contrast to the tumor suppressor of *Ncoa2* and oncogenic
591 *Zmiz1* in the cuSCC model described in this study. These two examples, and several other from our various studies
592 cataloged in the SBCDDB (Newberg et al., 2018c), support the dual roles many candidate cancer driver genes have in
593 both promoting and inhibiting various cancer hallmark features depending upon the complexities of network of
594 interactions operating to control genotype-phenotype outcomes in disease. In a related example, it was surprising
595 that the frequency of *Zmiz1/2* insertions were low within the K5-cre|Pten cuKA masses given that our study in
596 SB|Onc3 mice, and several previous transposon studies, have found activating *Zmiz1* insertions with high frequency
597 in skin tumors. This may result from the use of the *Krt5*-cre, as the *Keratin5* promoter is principally active within
598 basal skin cells but not keratinocytes. This raises the possibility that the tumor cell of origin is another important
599 factor for skin cancer progression (as it is in breast cancer), and suggests that masses arising from basal cells can
600 readily form frank cuKAs but not cuSCCs for unknown reasons. Additional experiments are needed and warranted to
601 explore this possibility.

602 Our experimental approach for driving high mutational burdens by SB mutagenesis, recapitulates the sporadic,
603 stepwise evolutionary selection of cooperating drivers in non-melanoma skin cancers in humans exposed to solar
604 UV. Despite these divergent mutagens, we observed remarkable overlap of driver genes, pathways, and networks
605 between the mouse and human cuSCCs, suggesting a deep mutual biology between these shared drivers and
606 cutaneous oncogenesis. We hope that by defining the cooperative oncogenic and tumor suppressor networks that
607 operate during keratinocyte transformation and subsequent cuSCC progression, the results of our screen will
608 provide a foundation for exploring new therapeutic strategies.

609 **Author Contributions**

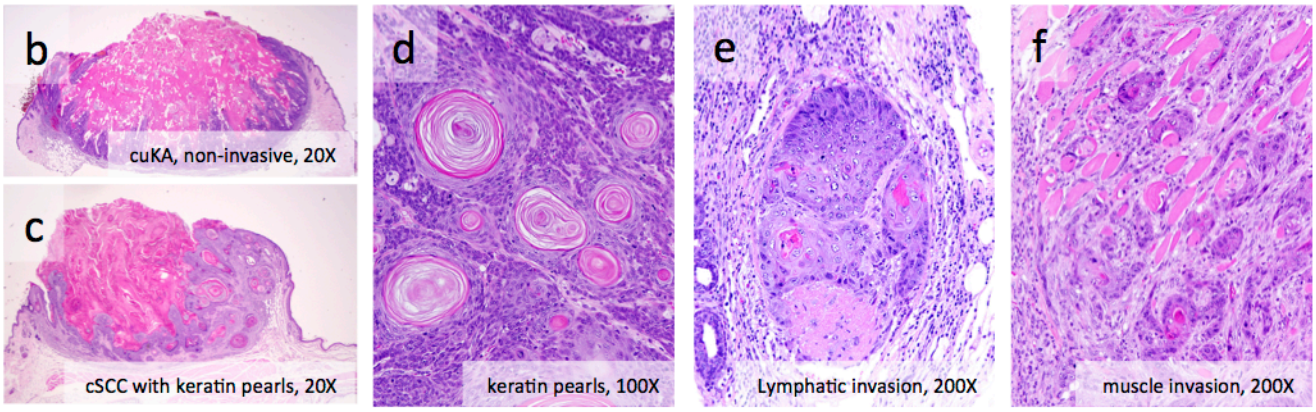
610 M.B.M., K.M.M. N.G.C., and N.A.J. conceived the project and provided laboratory resources and personnel for animal
611 husbandry, specimen archiving, sequencing and compute management. M.B.M. and K.M.M. designed the studies,
612 directed the research, interpreted the data, performed experimental work, designed the SBCapSeq oligos,
613 coordinated sequencing efforts and analyzed the data. A.A., L.G.-R., A.M.C.-S., A.L.M., D.J.J., F.A.-M., R.R., S.-C.L., K.H.-
614 K.B., and K.Y.T. provided reagents, performed and/or analyzed experimental datasets. A.A., J.Y.N., M.A.B., and M.B.M.
615 performed statistical and bioinformatics analyses. K.R. and S.M.R. performed and directed necropsy and
616 histotechnology services. J.M.W. performed and directed veterinary pathology analysis, including tumor grading and
617 diagnosis. L.M. and L.S. isolated RNA and performed microarray hybridizations. L.G.-R., A.M.C.-S., D.J.J., and F.A.-M.
618 performed and optimized capture hybridizations for the SBCapSeq method and/or performed Ion Torrent
619 sequencing for SBCapSeq, RNA-Seq experiments. M.B.M., L.G.-R., A.A., and A.L.M performed *in vitro* experiments.
620 M.B.M. and A.L.M performed *in vivo* xenograft experiments. A.A., K.M.M., and M.B.M. wrote the manuscript with

621 contributions from N.G.C., N.A.J., K.Y.T., and M.A.B. All co-authors approved the manuscript prior to submission.

622 **Competing financial interests.** The authors declare no competing financial interests.

Biological Process (GO term ID)	**Adjusted *P*-value**	**Gene symbols**
Skin development (GO:0043588)	3.40×10^{-13}	*Flg, Krt10, Prr9, Lce1f, Ereg, Hrnr, Lce3c, Krt16, Sprr1a, Rptn, Tgm3, Ivl, Sprr1b*
Keratinocyte differentiation (GO:0030216)	1.45×10^{-12}	*Flg, Lce3c, Krt16, Krt10, Prr9, Lce1f, Sprr1a, Tgm3, Sprr1b, Ivl, Ereg*
Peptide cross-linking (GO:0018149)	3.91×10^{-12}	*Flg, Lce3c, Krt1, Krt10, Prr9, Lce1f, Sprr1a, Tgm3, Sprr1b, Ivl*
Epidermal cell differentiation (GO:0009913)	4.32×10^{-12}	*Flg, Lce3c, Krt16, Krt10, Prr9, Lce1f, Sprr1a, Tgm3, Sprr1b, Ivl, Ereg*
Epidermis development (GO:0008544)	3.91×10^{-8}	*Krtdap, Flg, Casp14, Krt16, Krt15, Sprr1a, Rptn, Sprr1b, Hrnr*

623 **Table 1. Gene Ontology biological processes enriched in SB-driven cuSCC tumors with *Zmiz1* insertions.**

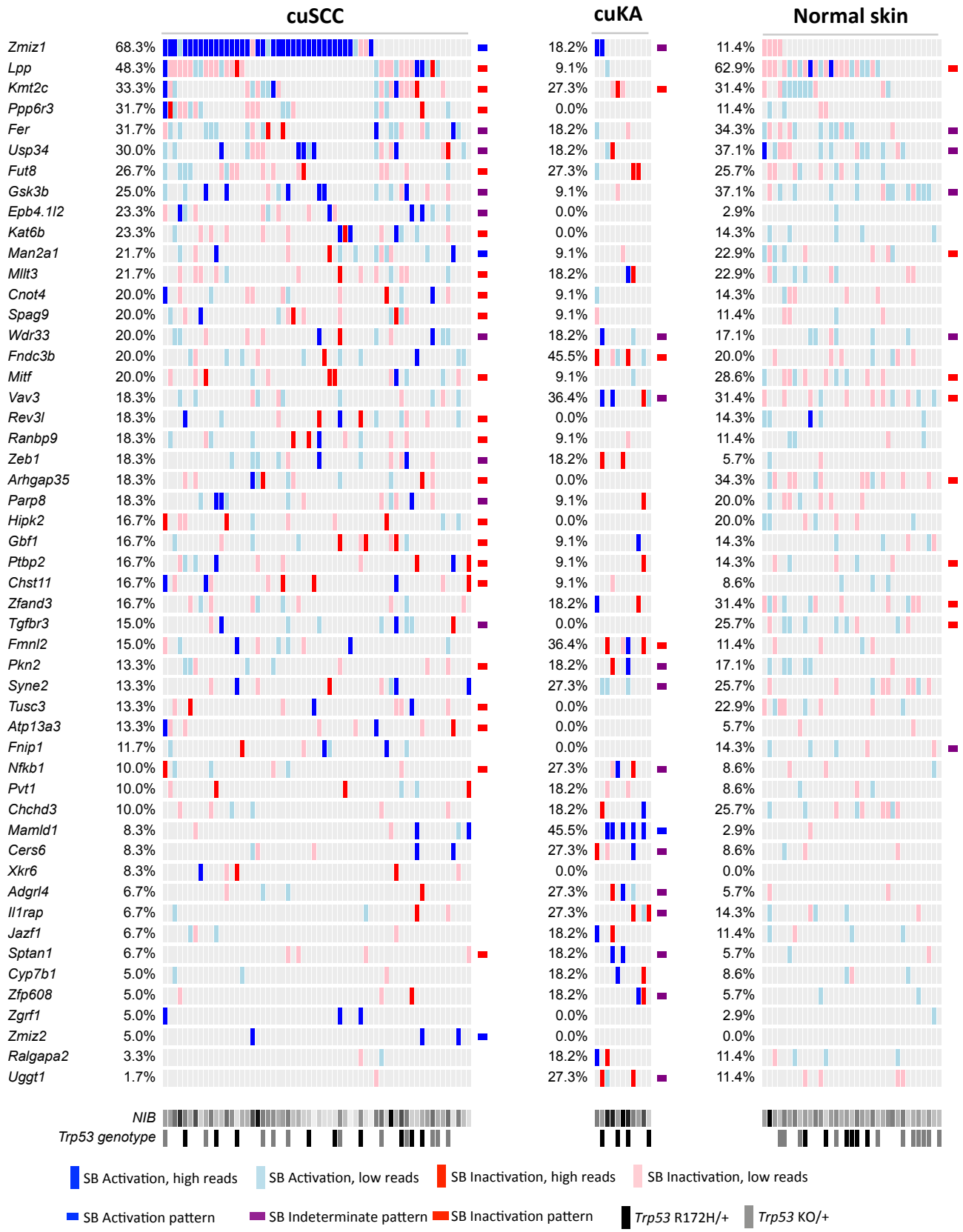


g

Cohort / Tumor classifications	SB-Onc3|Trp53-KO/+		SB-Onc3|Trp53-R172H/+		SB-Onc3|Trp53+/+		Trp53-R172H/+ Controls	
	# tumors	# mice	# tumors	# mice	# tumors	# mice	# tumors	# mice
Cutaneous Squamous Cell Carcinoma	17	10 (20%)	14	7 (15%)	52	20 (24%)	0	0 (0%)
Cutaneous Keratoacanthoma	3	3 (6%)	0	0 (0%)	9	8 (10%)	0	0 (0%)

Figure 1: Whole-body SB transposon mutagenesis drives a diverse tumor spectrum in *Trp53* wild type and mutant mice. (**a**) Kaplan-Meier survival curves comparing experimental SB and non-SB control mouse cohorts (log-rank test, $P < 0.0001$, **Supplementary Figure 1**). Wild type *Trp53* cohorts with SB-Onc3 mobilized by a constitutively active *Rosa26*-SBase allele (Gray line, reported in (Dupuy et al., 2009)) do not differ significantly from SB-Onc3 mice mobilized by a conditionally activated *Rosa26*-LSL-SBase allele to the *Rosa26*-1lox-SBase allele by whole body *Cre* expression (log-rank test, $P>0.05$). Mice in all SB-Onc3 cohorts developed solid tumors, including cuSCC and HCA, but *Trp53*$^{+/-}$ (red line) and *Trp53*$^{R172H/+}$ (blue line) mice had significantly decreased survival compared to the WT cohort (Green line, $P<0.0001$, log-rank). Non-SB control Trp53 R172H/+ (broken black line) mice had significantly decreased survival compared to the WT cohort (solid black line, $P<0.0001$, log-rank). (**b-c**) Histology and tumor classification from sections of skin masses stained with hematoxylin and eosin. (**c**) Early-stage, non-invasive mass with cutaneous keratoacanthoma-like morphology in SB-Onc3|*Trp53*$^{+/+}$ mouse (20×). (**d**) Late-stage, invasive mass with cuSCC displaying keratin pearl morphology in SB-Onc3|*Trp53*$^{KO/+}$ mouse (20×). (**f**) cuSCC keratin pearls in SB-Onc3|*Trp53*$^{+/+}$ mouse (100×). (**e**) cuSCC displaying lymphatic invasion in SB-Onc3|*Trp53*$^{R172H/+}$ mouse (200×). (**e**) cuSCC displaying muscle invasion in SB-Onc3|*Trp53*$^{+/+}$ mouse (200×). (**g**) Cutaneous tumor spectrum in SB-Onc3 mice. *Trp53* mutation shortens tumor latency, broadens solid tumor spectrum, and reduces overall tumor burden/multiplicity induced by whole body SB transposon mutagenesis compared to wild type *Trp53* cohorts (**Supplementary Table 1**).

Figure 2: Landscape of trunk drivers mutated during SB-induced keratinocyte transformation and cuSCC progression using SBCapSeq. SB Trunk Driver Analysis, with read depth cutoff of 300, was performed using SBCapSeq insertion data from the SB|cuSCC and SB|cuKA tumor cohorts (the normal skin cohort did not identify any significant trunk drivers). Drivers with significant family-wise error rate (FWER) adjusted *P*-values were merged into a single list and plotted from each of the cuSCC, cuKA and normal skin specimens by occurrence within the cuSCC cohort. Vertical colored bars represent genomes (columns) with SB insertion events that occur within the same (sense, blue) or opposite (anti-sense, red) DNA strand relative to transcription of a driver gene (row). High read depth sites (>299 reads supporting an SB insertion event, blue or red) are distinguished from low read depth sites (<300 reads supporting an SB insertion event, light blue or pink) to denote clonal and sub-clonal SB insertion events, respectively. Driver gene classifications (indeterminate, activating or Inactivating) are shown as a bar at the end of a gene on a cohort specific manner where calculation was possible. NIB, normalized insertion burden from highest (black) to lowest (light gray). *Trp53* genotypes: *Trp53* $^{+/+}$ (white), *Trp53* $^{R172H/+}$ (black), and *Trp53* $^{KO/+}$ (gray). *(Image on next page)*


cuSCC
cuKA
Normal skin
Zmiz1 68.3% 18.2% 11.4%
Lpp 48.3% 9.1% 62.9%
Kmt2c 33.3% 27.3% 31.4%
Ppp6r3 31.7% 0.0% 11.4%
Fer 31.7% 18.2% 34.3%
Usp34 30.0% 18.2% 37.1%
Fut8 26.7% 27.3% 25.7%
Gsk3b 25.0% 9.1% 37.1%
Epb4.1l2 23.3% 0.0% 2.9%
Kat6b 23.3% 0.0% 14.3%
Man2a1 21.7% 9.1% 22.9%
Mllt3 21.7% 18.2% 22.9%
Cnot4 20.0% 9.1% 14.3%
Spag9 20.0% 9.1% 11.4%
Wdr33 20.0% 18.2% 17.1%
Fndc3b 20.0% 45.5% 20.0%
Mitf 20.0% 9.1% 28.6%
Vav3 18.3% 36.4% 31.4%
Rev3l 18.3% 0.0% 14.3%
Ranbp9 18.3% 9.1% 11.4%
Zeb1 18.3% 18.2% 5.7%
Arhgap35 18.3% 0.0% 34.3%
Parp8 18.3% 9.1% 20.0%
Hipk2 16.7% 0.0% 20.0%
Gbf1 16.7% 9.1% 14.3%
Ptbp2 16.7% 9.1% 14.3%
Chst11 16.7% 9.1% 8.6%
Zfand3 16.7% 18.2% 31.4%
Tgfbr3 15.0% 0.0% 25.7%
Fmnl2 15.0% 36.4% 11.4%
Pkn2 13.3% 18.2% 17.1%
Syne2 13.3% 27.3% 25.7%
Tusc3 13.3% 0.0% 22.9%
Atp13a3 13.3% 0.0% 5.7%
Fnip1 11.7% 0.0% 14.3%
Nfkb1 10.0% 27.3% 8.6%
Pvt1 10.0% 18.2% 8.6%
Chchd3 10.0% 18.2% 25.7%
Mamld1 8.3% 45.5% 2.9%
Cers6 8.3% 27.3% 8.6%
Xkr6 8.3% 0.0% 0.0%
Adgrl4 6.7% 27.3% 5.7%
Il1rap 6.7% 27.3% 14.3%
Jazf1 6.7% 18.2% 11.4%
Sptan1 6.7% 18.2% 5.7%
Cyp7b1 5.0% 18.2% 8.6%
Zfp608 5.0% 18.2% 5.7%
Zgrf1 5.0% 0.0% 2.9%
Zmiz2 5.0% 0.0% 0.0%
Ralgapa2 3.3% 18.2% 11.4%
Uggt1 1.7% 27.3% 11.4%
NIB
Trp53 genotype
SB Activation, high reads
SB Activation, low reads
SB Inactivation, high reads
SB Inactivation, low reads
SB Activation pattern
SB Indeterminate pattern
SB Inactivation pattern
Trp53 R172H/+
Trp53 KO/+

652

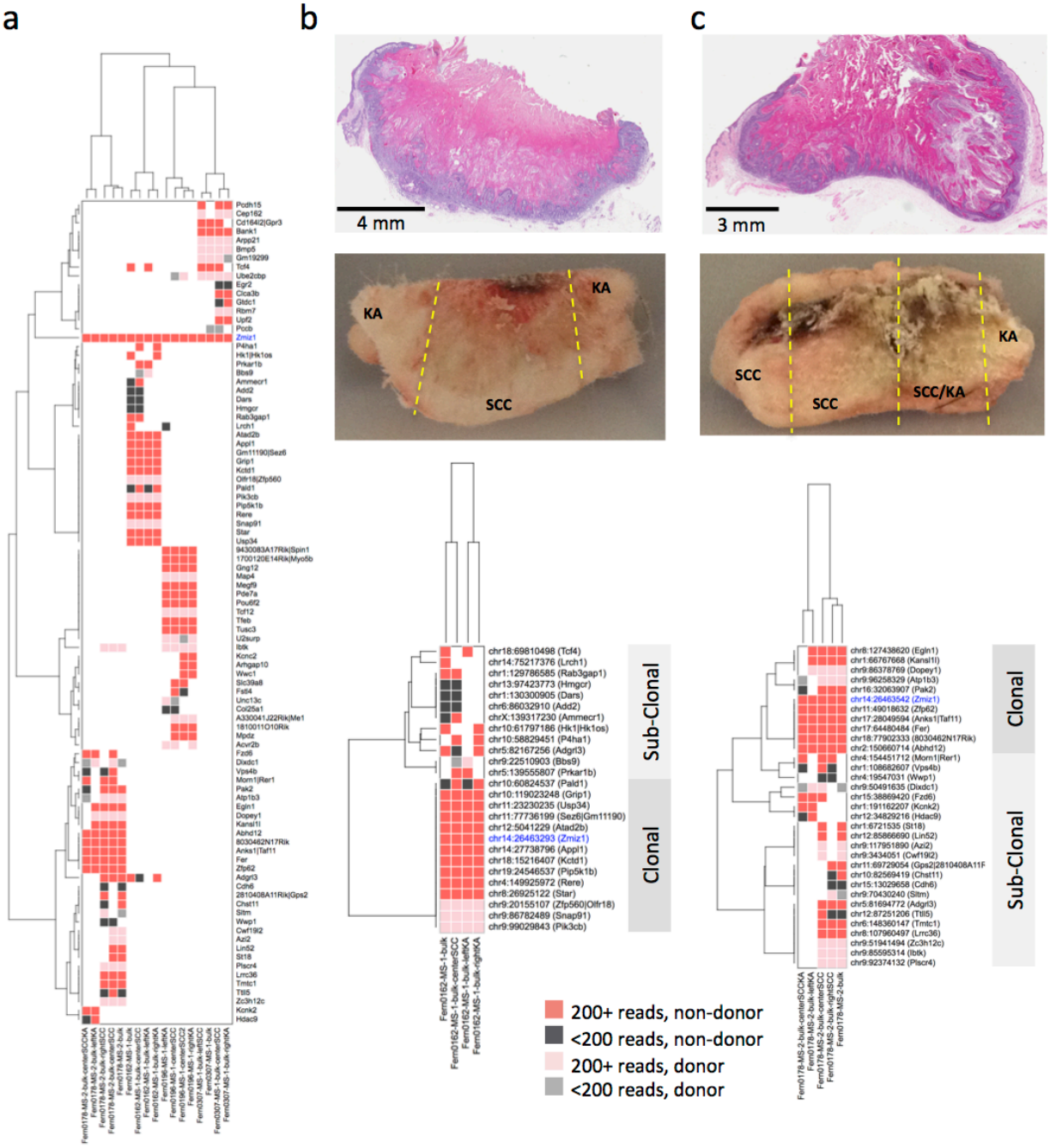


Figure 3: Multi-region SBCapSeq analysis of dual histology keratinocyte-derived skin masses containing distinct cuKA and cuSCC histological regions. (**a**) Hierarchical two-dimensional clustering (Hamming distance with the Ward method of agglomeration) of recurrent genic SBCapSeq insertion data from four skin masses containing distinct cuKA and cuSCC regions. Four distinct specimen (*x*-axis) and gene (*y*-axis) clades, each pertaining to a single bi-lesional mass, were detected demonstrating clonal identities. *Zmiz1* was the only gene recurrently mutated across all samples with high read depths. Genomic DNA was isolated from adjacent serial sections and included a "bulk" reference consisting of a cross-section of entire mass, and three or four regions dissected from histologically distinct regions (see **panels b-c** and **Supplementary Figure 8**). (**b**) Two masses with Aperio slide scans of H&E stained FFPE specimens showing distinct cuKA and cuSCC histologies suggestive of clonal evolution of cuSCC differentiation within cuKA masses and the adjacent flash frozen tissue prior to sectioning for genomic DNA isolation. Hierarchical two-dimensional clustering (Hamming distance with the Ward method) of recurrent genic SBCapSeq insertion data shown below each specimen demonstrates trunk (common to all samples) and sub-clonal, lesion-specific SB insertions.

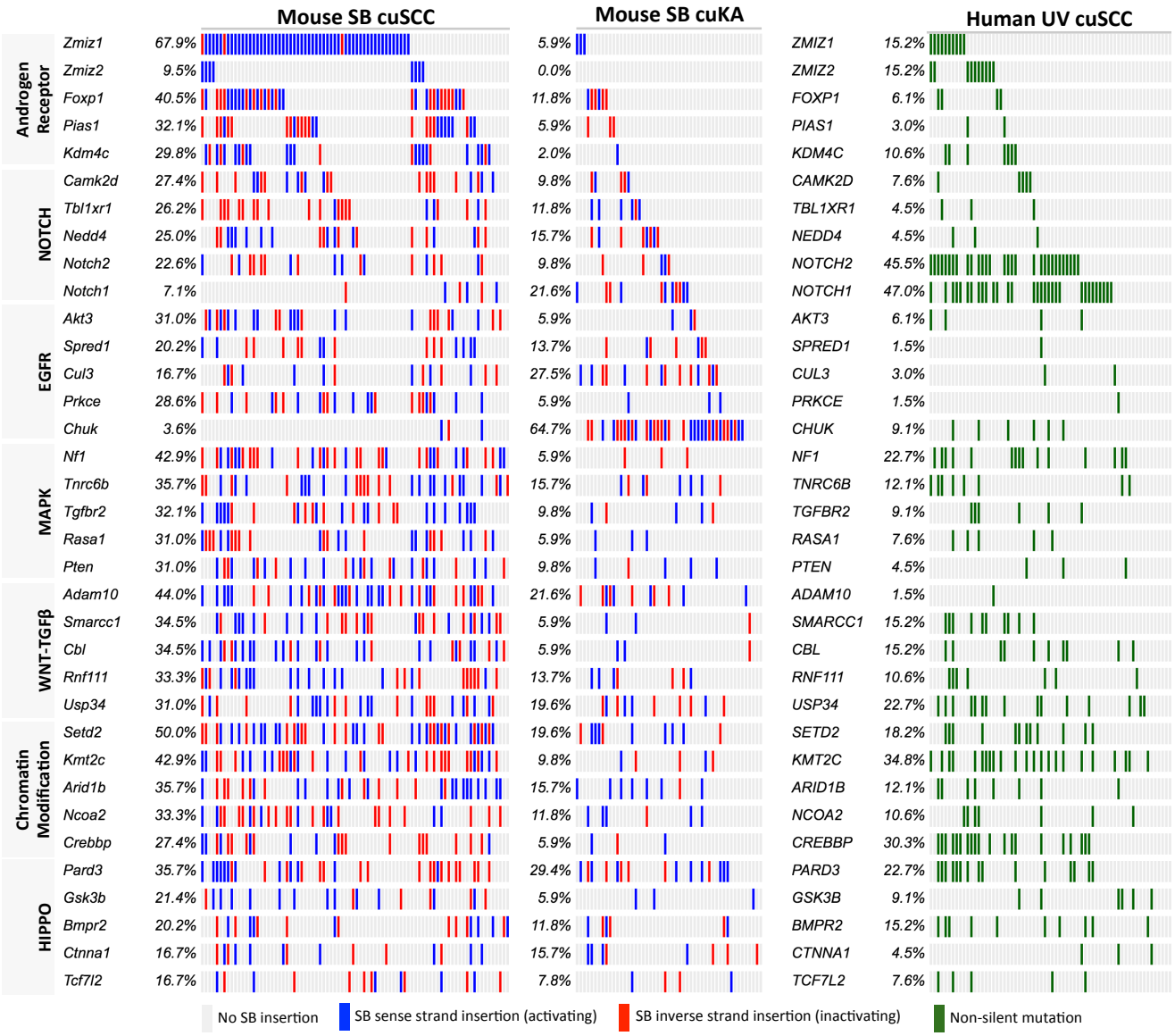


Figure 4: Comparative genomic landscape of candidate driver mutations and common altered pathways in mouse and human cuSCC genomes. Integrated Oncoprints sorted by biological pathway or process discovery significance within the SB|cuSCC dataset (**Supplementary Table 28**) and displaying the mutation burden across SB-induced mouse cuSCC (n=84) or cuKA (n=62) and human UV-IR-induced cuSCC genomes.

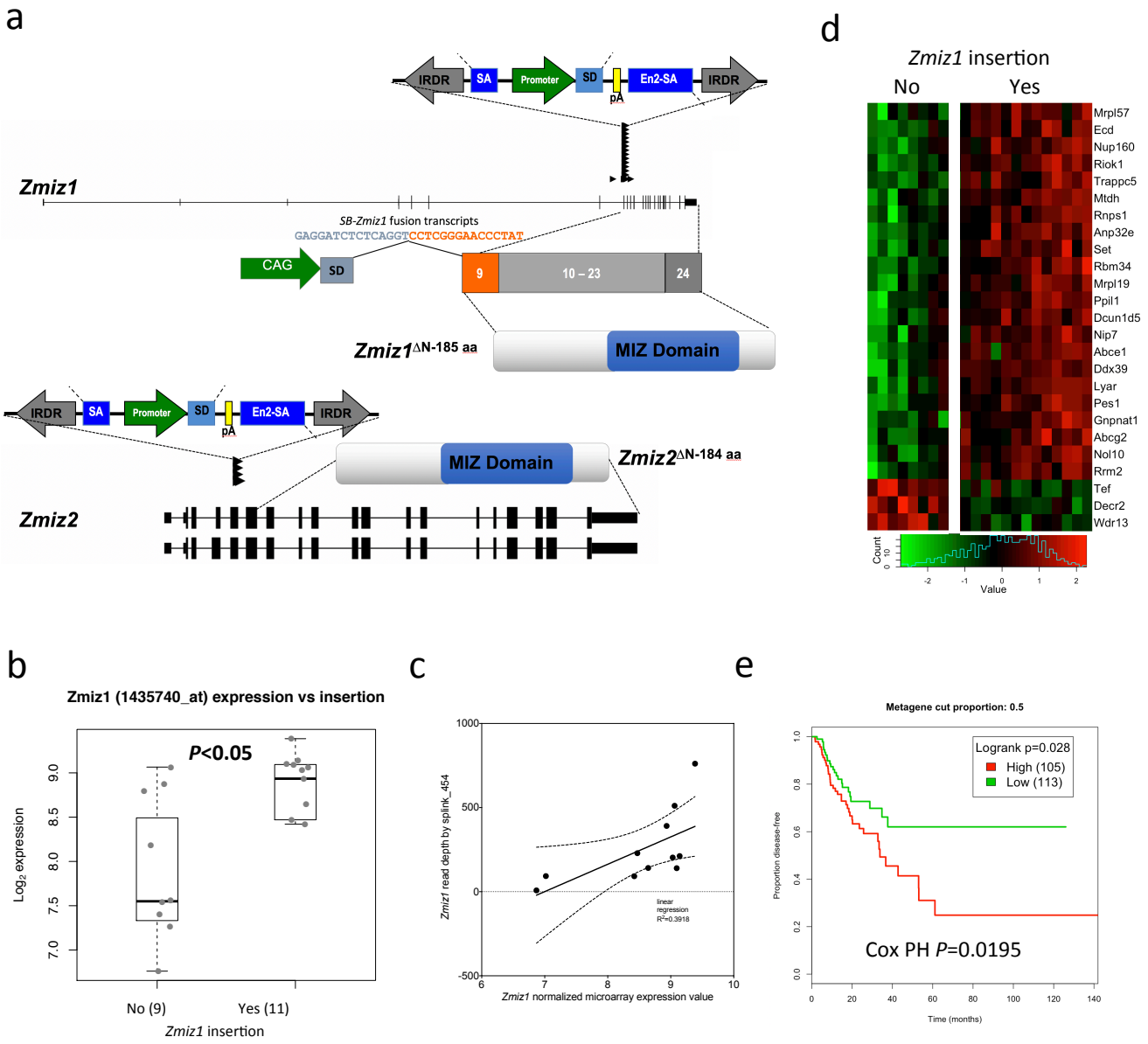
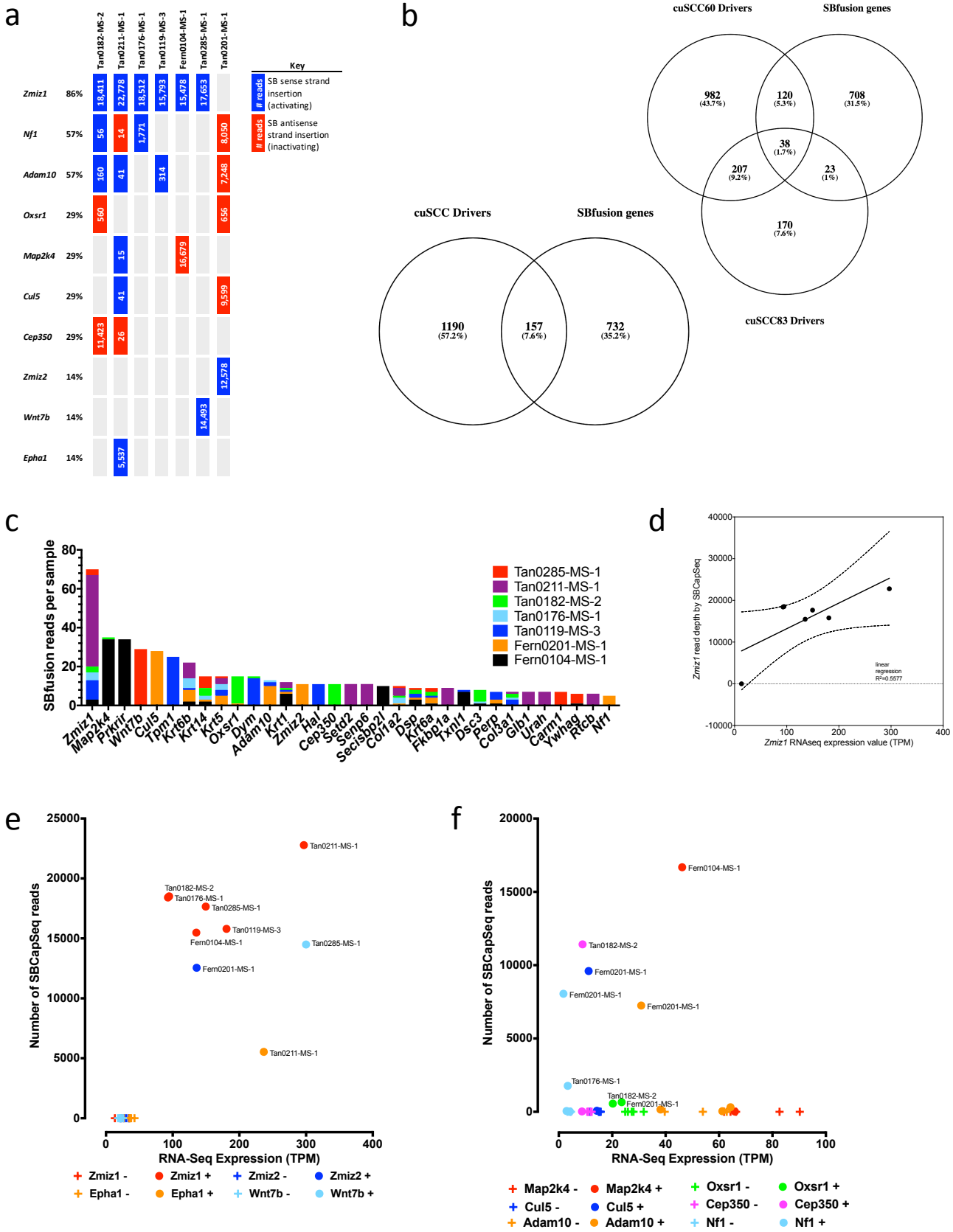


Figure 5: SB-mediated mutagenesis promotes N-terminal truncated ZMIZ paralog activation during cuSCC progression in mice. (**a**) Recurrent TA-dinucleotide SB insertion events from individual cuSCC genomes reveal activating sense strand SB insertions into Zmiz paralogs Zmiz1 and Zmiz2. Chimeric fusion between the SB transposon splice donor (SD) and splice acceptor sites at exon 9 of *Zmiz1* or exon 6 of *Zmiz2* resulting in a CAG-promoter driven transcription of a truncated mRNA that is predicted to encode N-terminally truncated proteins *Zmiz1*$^{\Delta N\text{-}185\ aa}$ or *Zmiz1*$^{\Delta N\text{-}184\ aa}$ that contain functional MIZ-type zinc finger domains, containing a Siz/PIAS RING finger (SP-RING). (**b**) Microarray gene expression analysis reveal that tumors with *Zmiz1* SB insertions (yes) had significantly higher expression of *Zmiz1*, compared to tumors without insertions (no). Box boundaries indicate interquartile range; whiskers indicate maximum and minimum values; center lines indicate medians. (**c**) High positive correlation between Splink_454 read depth and *Zmiz1* expression by microarray analysis. (**d**) Top 25 differentially expressed genes between genomes with (yes) or without (no) *Zmiz1* SB insertions (see **Supplementary Figure 11a**). (**e**) Survival plots for patients with head and neck type SCC (hnSCC) based on the expression of a *ZMIZ1*–centric metagene constructed using singular value decomposition of human hnSCC RNA-seq dataset from TCGA. These reveal poor patient survival outcomes predicted by increasing expression of *ZMIZ1* and closely regulated genes consisting of a 23-gene signature (**Supplementary Figures 14-16**).

685 **Figure 6: Clonally selected SB insertion events affect trunk driver gene expression in SB-induced cuSCC genomes.** (**a**)
686 Seven late-stage cuSCC SB-Onc3 genomes with mutually exclusive high-read depth *Zmiz1* or *Zmiz2* SB insertions selected for
687 RNA-seq analysis of SB-fusion transcripts and global gene expression levels. (**b**) Venn diagrams comparing the overlap between
688 cuSCC driver genes and detection of SBfusion transcripts by whole transcriptome RNA sequencing (wtRNA-seq) performed on
689 bulk populations of cells isolated from 7 late-stage cuSCC genomes. (**c**) Genes with the most abundant SBfusion reads detected by
690 wtRNA-seq analysis performed on bulk populations of cells isolated across the 7 late-stage cuSCC genomes. (**d**) High positive
691 correlation between SBCapSeq read depth and *Zmiz1* expression by wtRNA-seq analysis. (**e**) Multifold induction of gene
692 expression in cuSCC masses with (**+**) high read depth activating SB insertion events among 4 candidate oncogenic drivers
693 compared with normal gene expression levels in cuSCC tumors without (**–**) SB insertions. Individual gene plots may be found in
694 (**Supplementary Tables 39-40**). (**f**) Reduced gene expression in cuSCC masses with (**+**) high read depth inactivating SB insertion
695 events among 6 candidate tumor suppressor drivers compared with normal gene expression levels in cuSCC tumors without (**–**)
696 SB insertions. Individual gene plots may be found in (**Supplementary Tables 39-40**). *(Image on next page)*


a
Tan0182-MS-2
Tan0211-MS-1
Tan0176-MS-1
Tan0119-MS-3
Fern0104-MS-1
Tan0285-MS-1
Tan0201-MS-1
Zmiz1 86% 18,411 22,778 18,512 15,793 15,478 17,653
Nf1 57% 56 14 1,771 8,050
Adam10 57% 160 41 314 7,248
Oxsr1 29% 560 656
Map2k4 29% 15 16,679
Cul5 29% 41 9,599
Cep350 29% 11,423 26
Zmiz2 14% 12,578
Wnt7b 14% 14,403
Epha1 14% 5,537
Key
reads SB sense strand insertion (activating)
reads SB antisense strand insertion (inactivating)
b
cuSCC60 Drivers
SBfusion genes
982 (43.7%)
120 (5.3%)
708 (31.5%)
38 (1.7%)
207 (9.2%)
23 (1%)
170 (7.6%)
cuSCC83 Drivers
cuSCC Drivers
SBfusion genes
1190 (57.2%)
157 (7.6%)
732 (35.2%)
c
SBfusion reads per sample
Tan0285-MS-1
Tan0211-MS-1
Tan0182-MS-2
Tan0176-MS-1
Tan0119-MS-3
Fern0201-MS-1
Fern0104-MS-1
Zmiz1 Map2k4 Prkrir Wnt7b Cul5 Tpm1 Krt6b Krt14 Krt5 Oxsr1 Dym Adam10 Krt1 Zmiz2 Hal Cep350 Setd2 Senp6 Secisbp2l Col1a2 Dsp Krt6a Fkbp1a Txnl1 Dsc3 Perp Col3a1 Glb1 Urah Carm1 Ywhag Rtcb Nf1
d
Zmiz1 read depth by SBCapSeq
Zmiz1 RNAseq expression value (TPM)
linear regression
R²=0.5577
e
Number of SBCapSeq reads
RNA-Seq Expression (TPM)
Tan0211-MS-1
Tan0182-MS-2
Tan0176-MS-1
Tan0285-MS-1
Tan0119-MS-3
Fern0104-MS-1
Tan0285-MS-1
Fern0201-MS-1
Tan0211-MS-1
Zmiz1 -
Zmiz1 +
Zmiz2 -
Zmiz2 +
Epha1 -
Epha1 +
Wnt7b -
Wnt7b +
f
Number of SBCapSeq reads
RNA-Seq Expression (TPM)
Fern0104-MS-1
Tan0182-MS-2
Fern0201-MS-1
Fern0201-MS-1
Fern0201-MS-1
Tan0176-MS-1
Tan0182-MS-2
Fern0201-MS-1
Map2k4 -
Map2k4 +
Oxsr1 -
Oxsr1 +
Cul5 -
Cul5 +
Cep350 -
Cep350 +
Adam10 -
Adam10 +
Nf1 -
Nf1 +

697

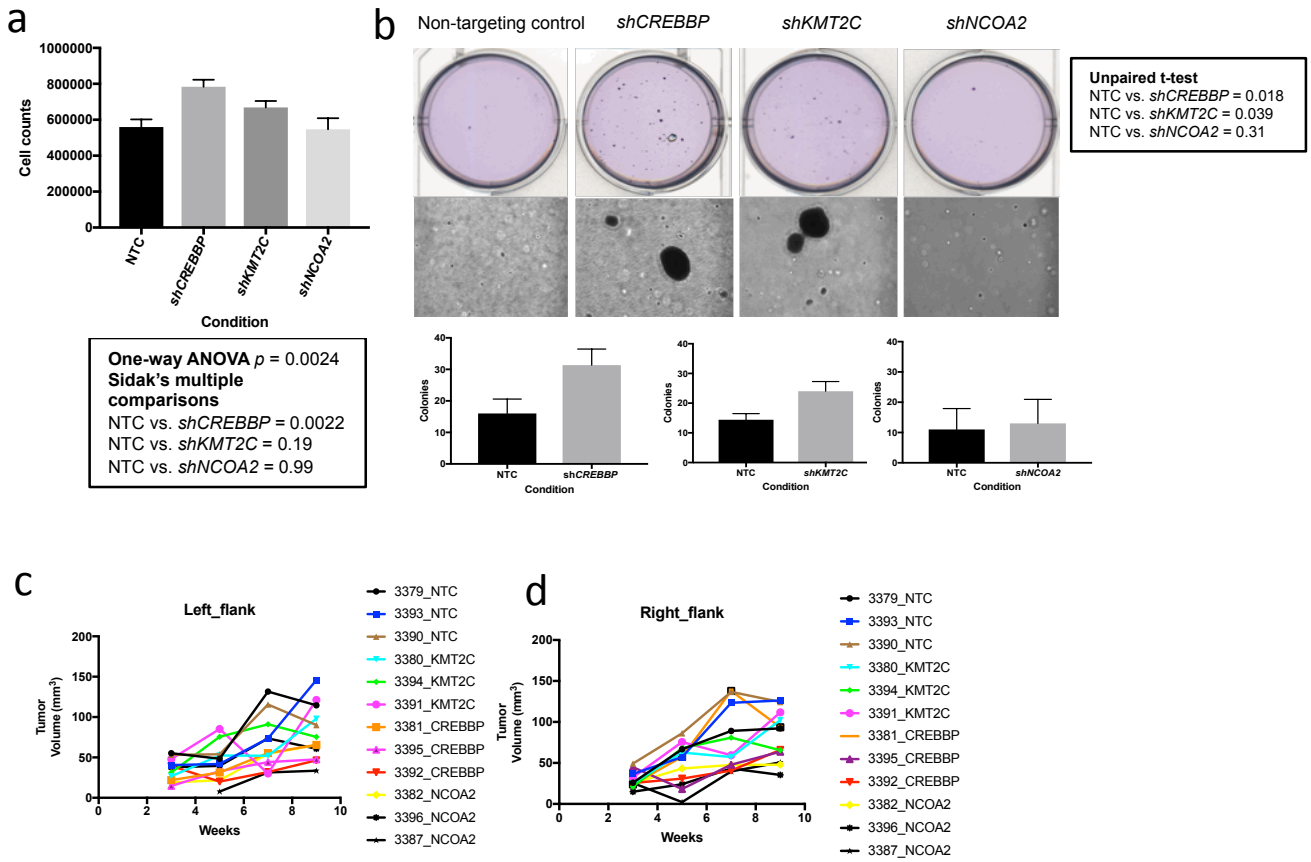
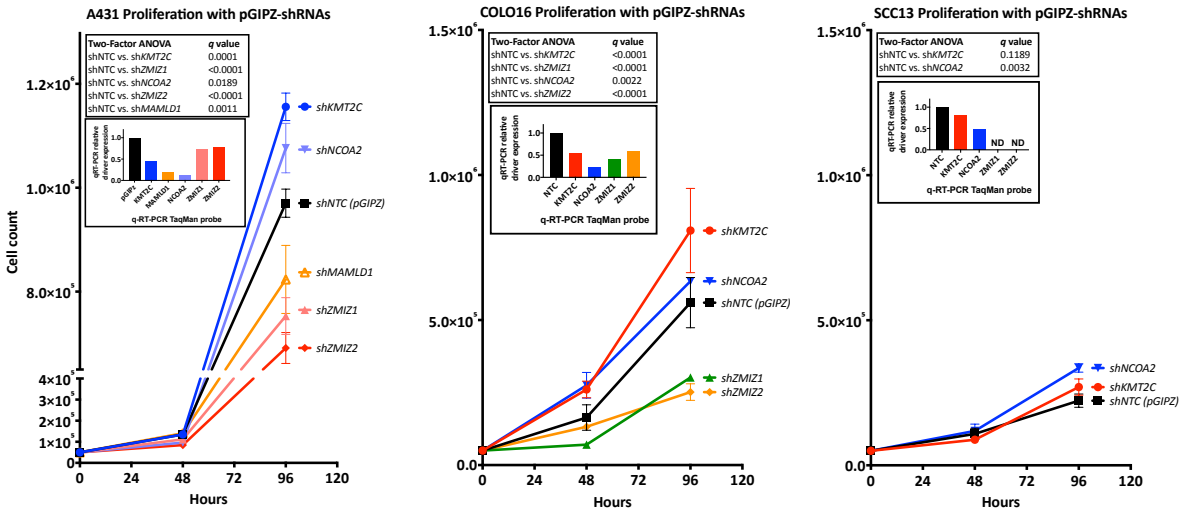
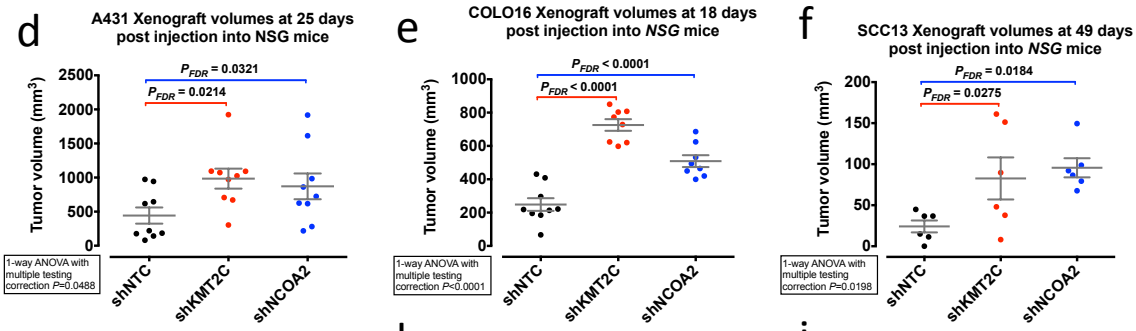
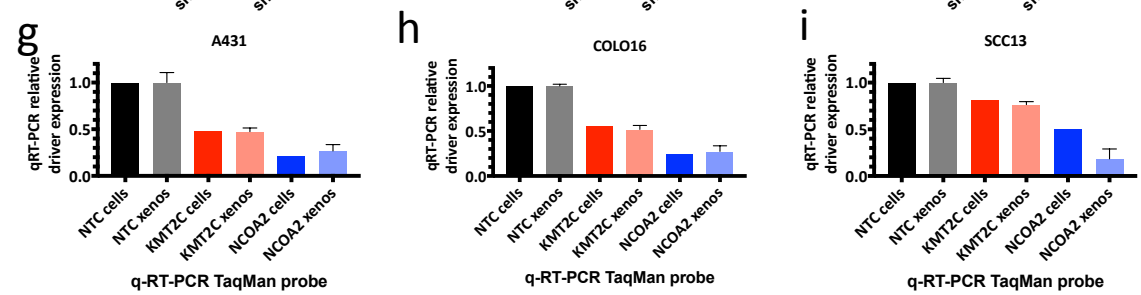


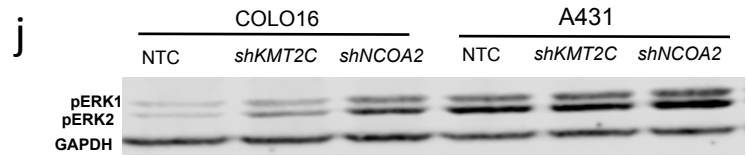
Figure 7: *In vitro* and *in vivo* functional validation of chromatin remodeler in immortal keratinocytes. (**a**-**b**) Assessment of keratinocyte transformation in response to knockdown of *CREBBP*, *KMT2C* and *NCOA2* by proliferation assay or anchorage-independent soft agar assay. For proliferation assay (n=6), statistical significance was tested by one-way ANOVA followed by Sidak's multiple comparisons adjustment. For soft agar assay (n=3), statistical significance was tested by unpaired t-test. (**c**-**d**) In vivo assessment of mass formation in HaCaT cells with knockdown of KMT2C, CREBBP, or NCOA2. Three million cells were injected into left and right flanks of NSG immunodeficient mice (n=3 per condition), and growth monitored up to nine weeks.

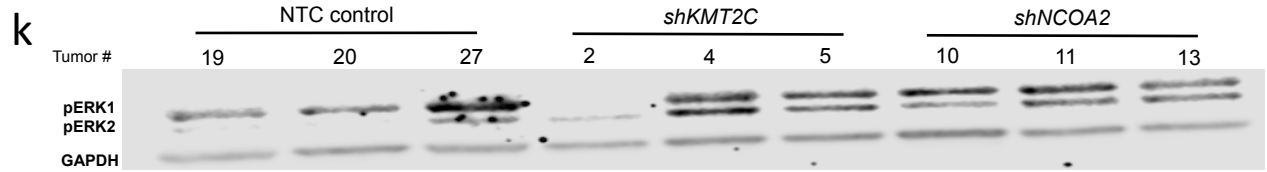
Figure 8: *In vitro* and *in vivo* functional validation of chromatin remodeler and ZMIZ drivers in cuSCC cell lines. (**a**-**c**) Knockdown of *KMT2C, NCOA2* or *ZMIZ* genes alter proliferation rates of human cuSCC cell lines (n=3; error bars, SEM; statistical significance measured by one-way ANOVA followed by multiple pairwise comparisons and FDR-adjusted *P*-values). (**d**-**f**) *KMT2C* and *NCOA2* knockdown accelerates cuSCC xenograft progression *in vivo.* (**g**-**h**) qRT-PCR values in 'cells' at the time of injection into NSG mice and in 'xenos' from specimens collected at necropsy. N=1 for 'cells' and N=6 or 8 (duplicates from 3 (COLO16) or 4 (A431, SCC13) independent masses) from xeno specimens. (**j**) Immunoblot for phospho-ERK expression in COLO16 and A431 cell lines with *KMT2C* or *NCOA2* knockdown. (**k**) Immunoblot for phospho-ERK expression from three largest tumors for each condition from the COLO16 xenograft assay shown in (**d**). (**l**) Summary of genes investigated with regards to their knockdown effects in human immortal keratinocyte and cuSCC cell line phenotypes. *(Image on next page)*







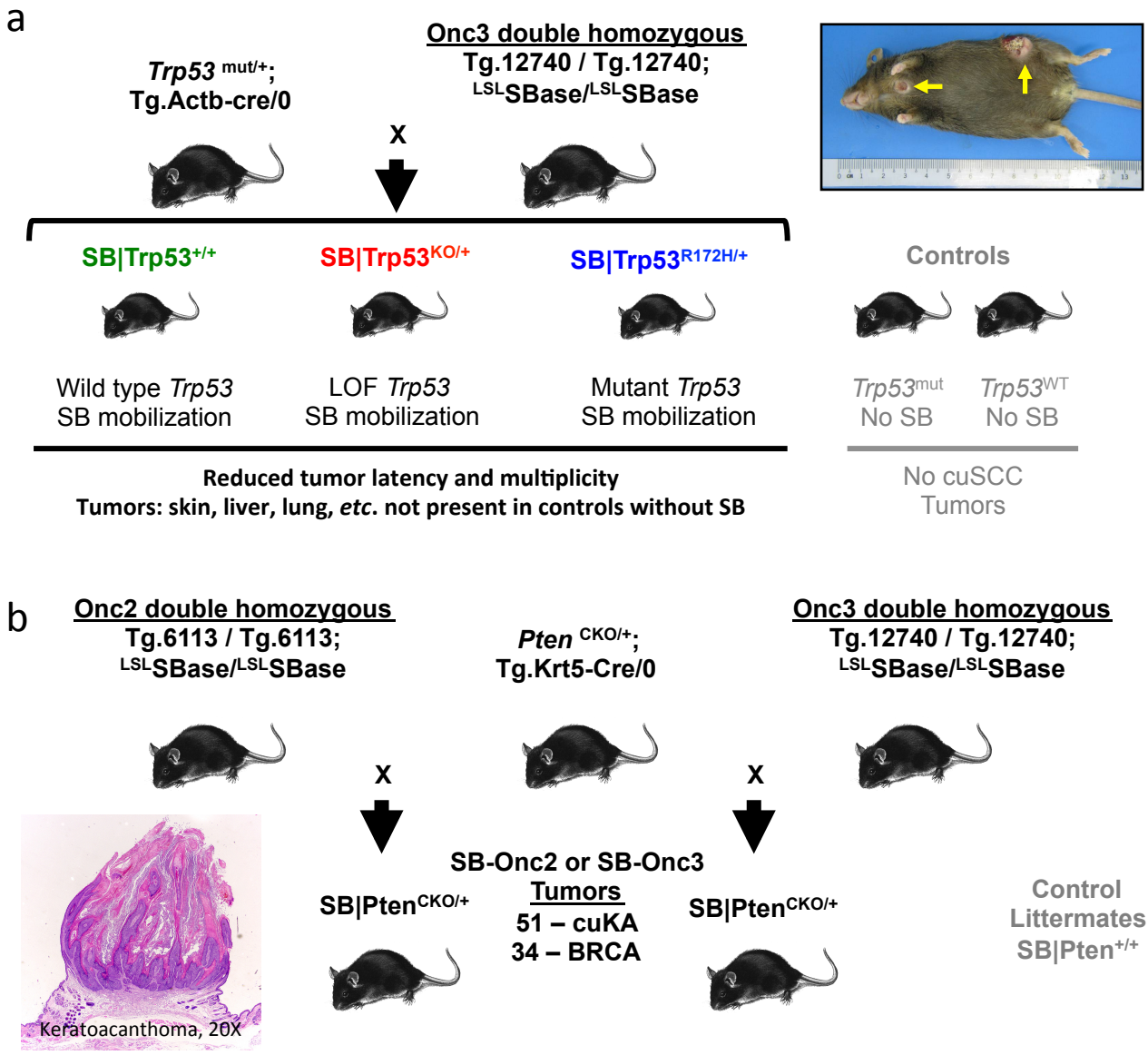




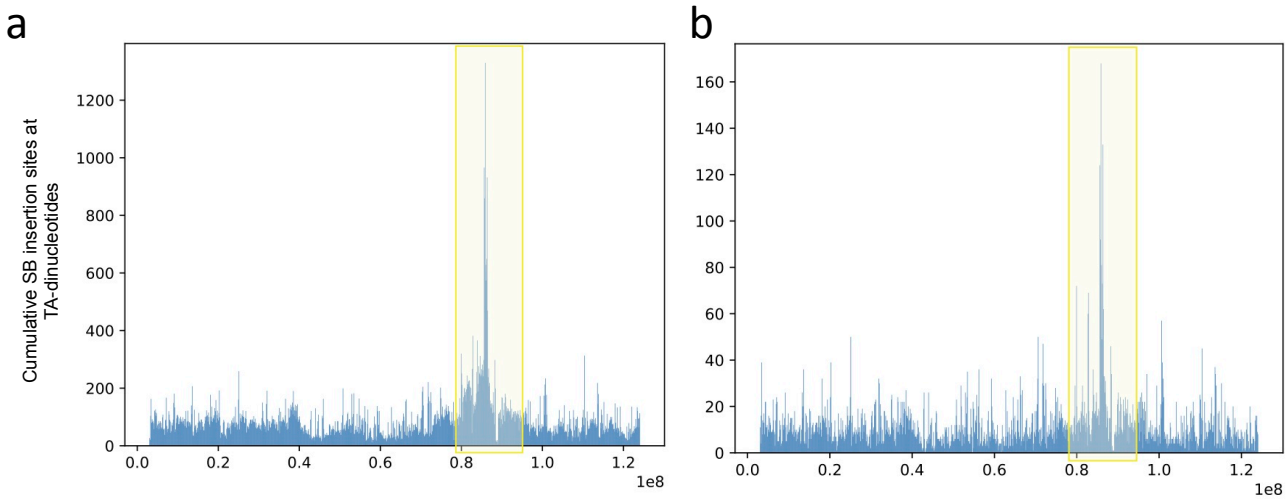
l

Driver Gene	Keratinocyte Initiation (HaCaT cells)	cuSCC Progression (cuSCC cell lines A431/COLO16/SCC13)
KMT2C	✓	✓
CREBBP	✓	✗
NCOA2	✗	✓
ZMIZ1	Not tested, no expression	✓
ZMIZ2	Not tested, no expression	✓

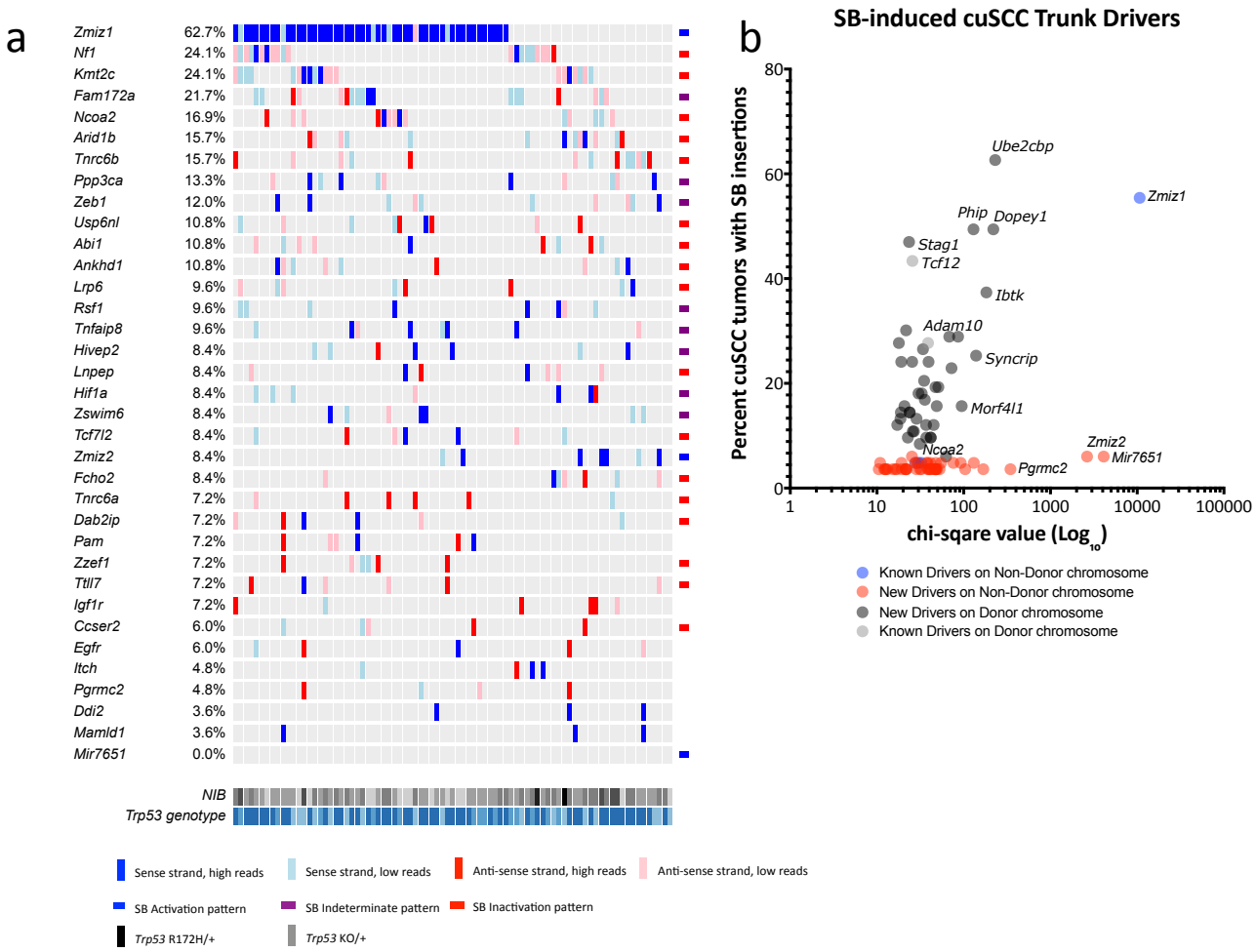
714



Supplementary Figure 1: Overview of SB-induced tumor studies. (**a**) SB|Trp53|Onc3, (**b**) SB|Pten|Onc2, and (**c**) SB|Pten|Onc3 cohorts, from (Rangel et al., 2016), used in this work. Yellow arrow indicates cuSCC masses at necropsy.

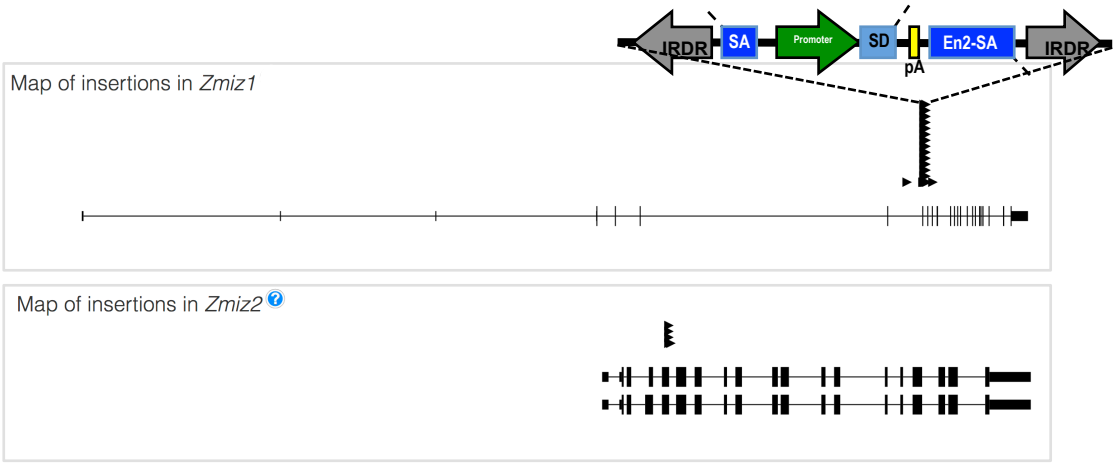


Supplementary Figure 2: SB T2/Onc3 TG.12740 allele donor position mapping and exclusion for SB Driver Analysis. Cumulative SB insertion sites at TA-dinucleotides from (a) 454_Splink and (b) Ion_SBCapSeq datasets showing the likely TG.12740 donor site at 87,000,000 bp and the exclusion region (yellow box) between 79,000,000 and 95,000,000 bp applied when SB Driver Analysis was run on chromosome 9. The 81 genes between *Filip1* at chr9:79663368-79825689 and *Tfdp2* at chr9:96096693-96224065 were excluded from SB Driver Analysis. The complete list of 81 genes excluded from SB Driver Analysis from tumors harboring the SB T2/Onc3 TG.12740 donor allele: *1190002N15Rik, 1700034K08Rik, 1700057G04Rik, 1700065D16Rik, 4930524O08Rik, 4930554C24Rik, 4933400C23Rik, 9330159M07Rik, 9430037G07Rik, A330041J22Rik, Adamts7, AF529169, Ankrd34c, Atr, B430319G15Rik, Bckdhb, Bcl2a1a, Bcl2a1b, Bcl2a1d, Cep162, Chst2, Ctsh, Cyb5r4, D430036J16Rik, Dopey1, Elovl4, Fam46a, Gk5, Hmgn3, Htr1b, Ibtk, Impg1, Irak1bp1, Lca5, Me1, Mei4, Mir184, Mir6386, Mir7656, Morf4l1, Mrap2, Mthfs, Mthfsl, Myo6, Nt5e, Paqr9, Pcolce2, Pgm3, Phip, Plod2, Pls1, Plscr1, Plscr2, Plscr4, Plscr5, Prss35, Rasgrf1, Ripply2, Rwdd2a, Senp6, Sh3bgrl2, Slc9a9, Snap91, Snhg5, Snx14, Syncrip, Tbc1d2b, Tbx18, Tmed3, Tpbg, Trim43a, Trim43b, Trim43c, Trpc1, Ttk, U2surp, Ube2cbp, Xrn1, Zfp949, Zic1, Zic4.*

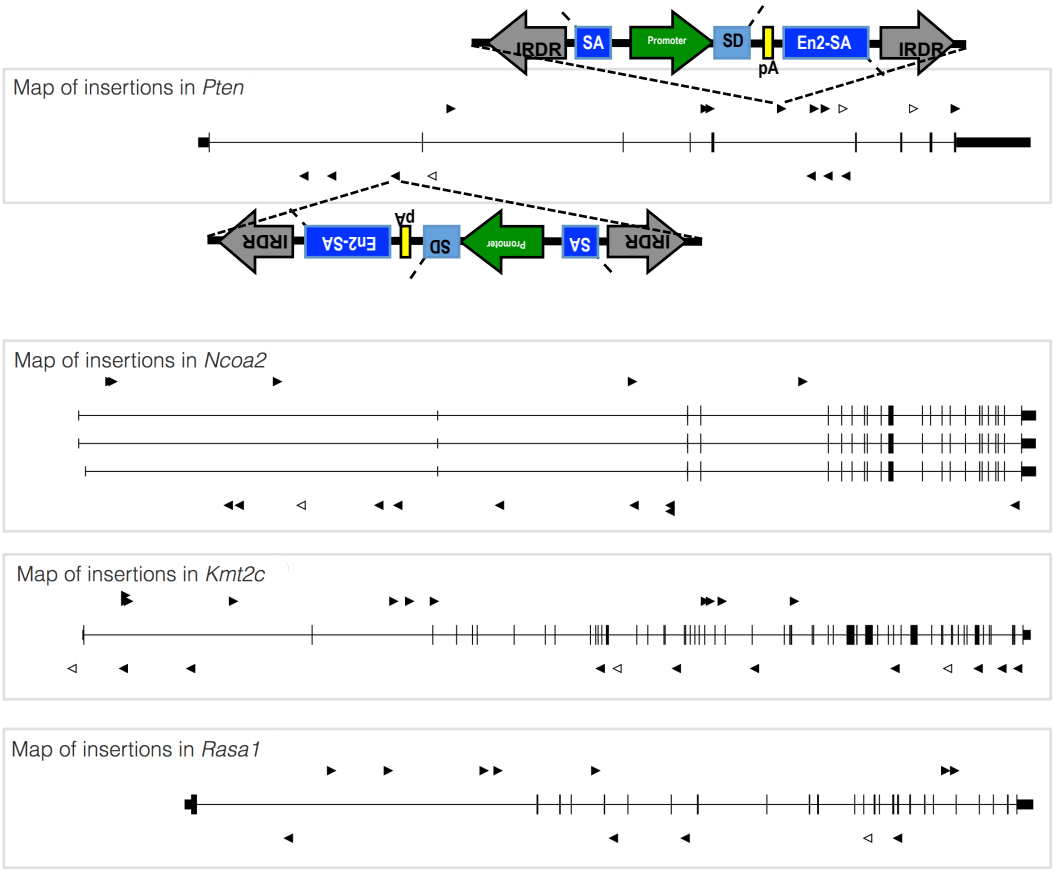


Supplementary Figure 3: Landscape of candidate trunk drivers mutated in SB-induced cuSCC with Splink_454 sequencing. SB Driver Analysis applied to Splink_454 SB insertion data (**Supplementary Table 4**) with read depth cutoff of 12 on 83 late-stage cuSCC genomes. (**a**) FWER significant driver genes (rows) were sorted by occurrence within the cohort genomes (columns). Waterfall plots were generated on all insertions, and insertion patterns determined by progression driver analysis are shown to the side. Normalized insertion burden (NIB) from lowest (light gray) to highest (black). *Trp53* genotype shown from Trp53 $^{+/+}$ (light blue), Trp53 $^{R172H/+}$ (blue) and Trp53 $^{KO/+}$ (dark blue). (**b**) Trunk cuSCC Driver genes containing SB insertion events with 10 or more 454 sequence reads per tumor. (**c-d**) Representative SB insertion maps in cuSCC trunk drivers showing the locations of mapped SB insertions (triangles) within (**c**) activating (oncogenic) pattern SB insertions into *Zmiz1* and *Zmiz2* or (**d**) Inactivating (tumor suppressive) pattern of SB insertions into *Pten*, *Ncoa2*, *Kmt2c*, and *Rasa1* from SB|cuSCC genomes. *(panels **c** and **d** continue on next page)*

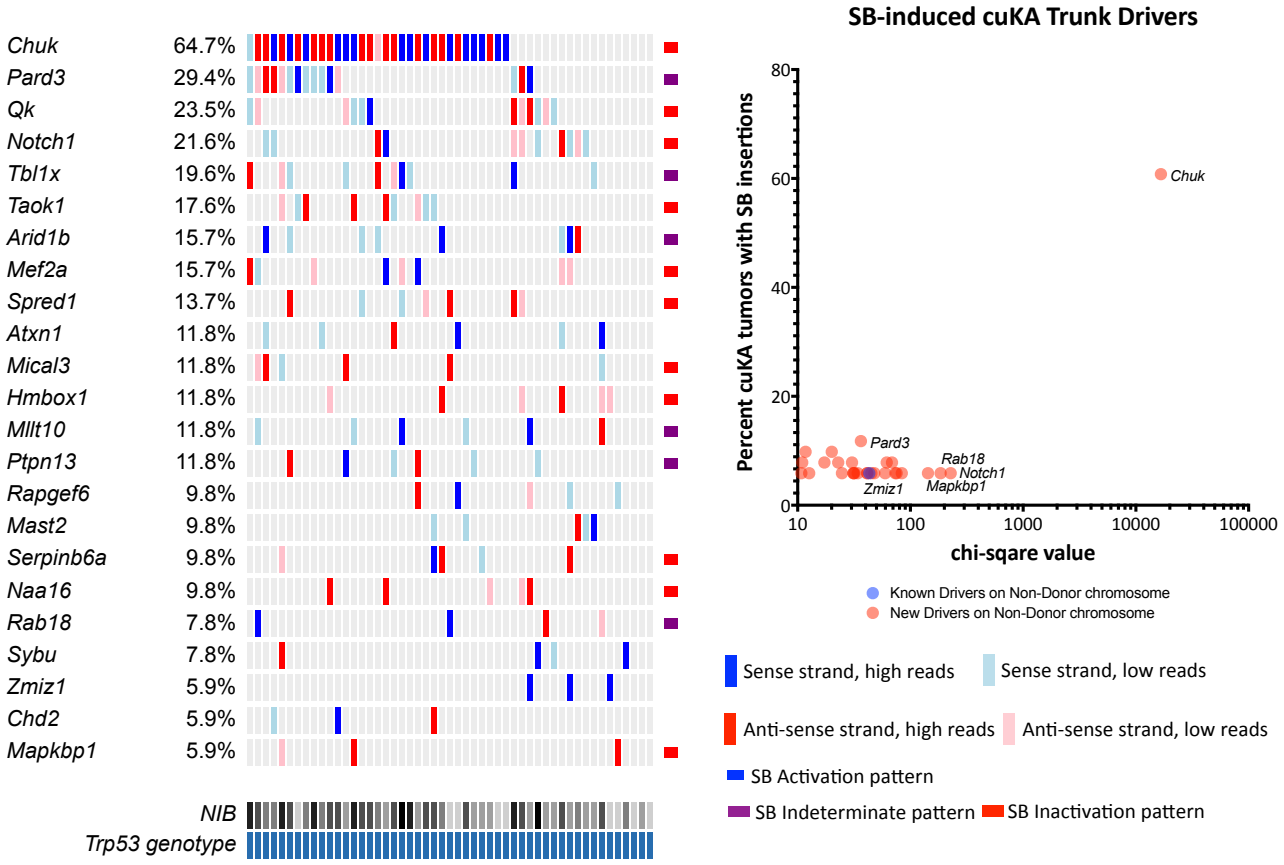
c


IRDR
SA
Promoter
SD
pA
En2-SA
IRDR
Map of insertions in *Zmiz1*
Map of insertions in *Zmiz2*

d

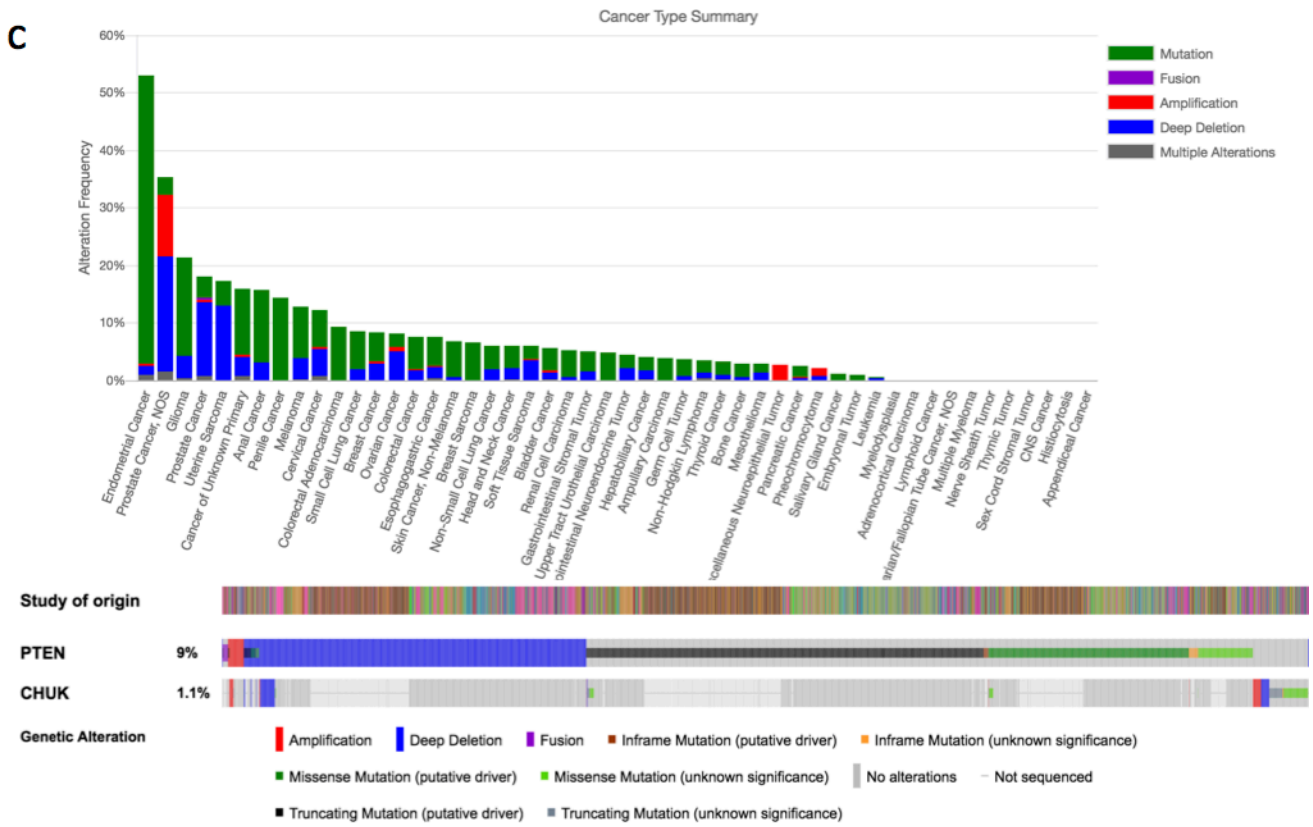

IRDR
SA
Promoter
SD
pA
En2-SA
IRDR
Map of insertions in *Pten*
Map of insertions in *Ncoa2*
Map of insertions in *Kmt2c*
Map of insertions in *Rasa1*

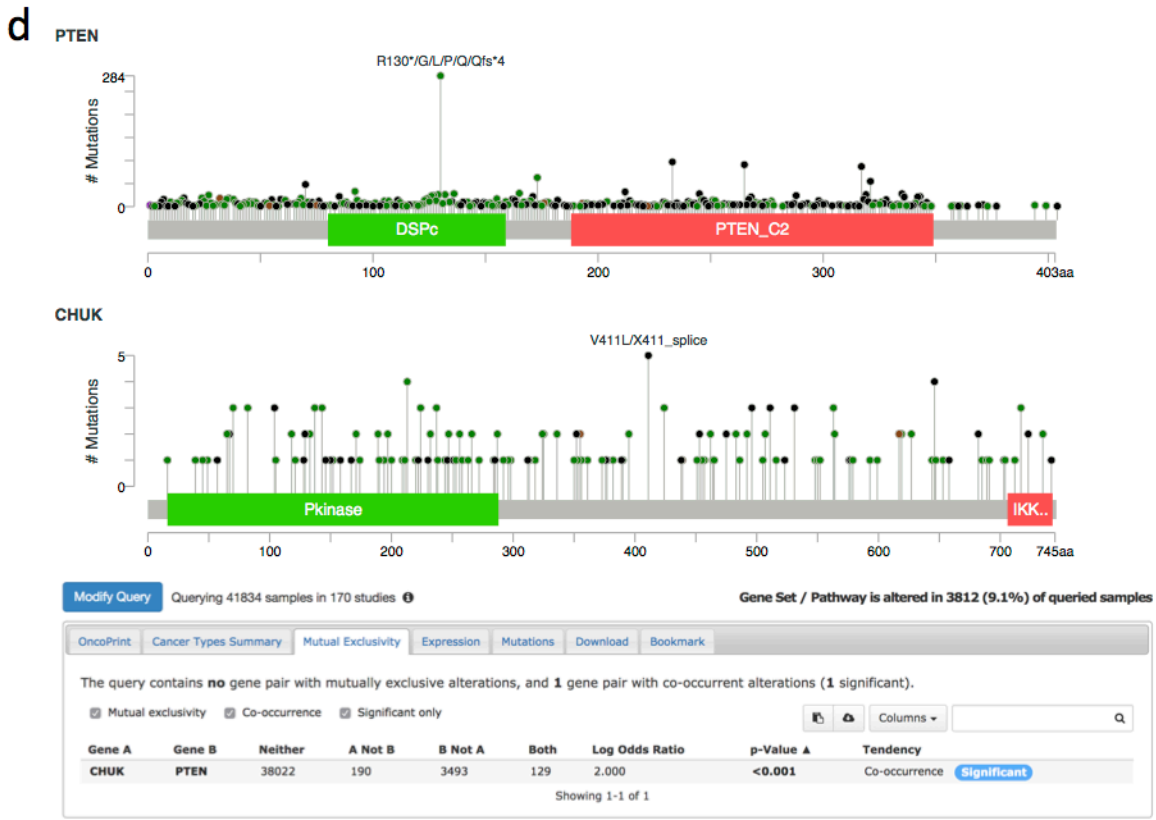
742



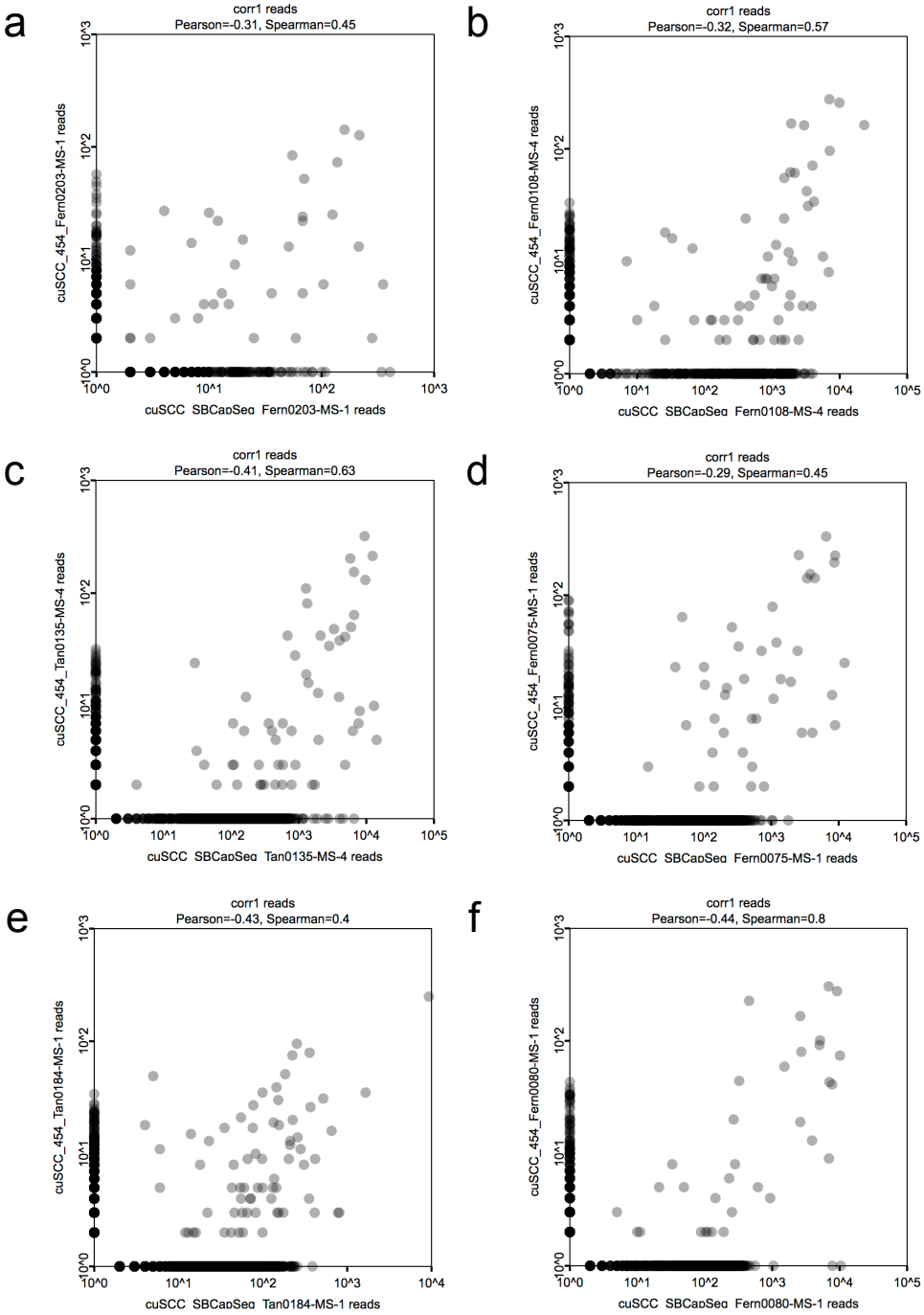
743

744 **Supplementary Figure 4: Landscape of candidate trunk drivers mutated in SB-induced cuKA arising from the SB|*Krt5-***
745 **Cre|*Pten*$^{floxed/+}$ cohort with Splink_454 sequencing.** SB Driver Analysis applied to Splink_454 SB insertion data
746 (**Supplementary Table 7**) with read depth cutoff of 20 on 83 late-stage cuSCC genomes. (**a**) FWER significant driver genes
747 (rows) were sorted by occurrence within the cohort genomes (columns). Waterfall plots were generated on all insertions, and
748 insertion patterns determined by progression driver analysis are shown to the side. Normalized insertion burden (NIB) from
749 lowest (light gray) to highest (black). *Trp53* genotype shown from *Trp53* $^{+/+}$ (light blue), *Trp53* $^{R172H/+}$ (blue) and *Trp53* $^{KO/+}$ (dark
750 blue). (**b**) Trunk cuKA Drivers genes containing SB insertions events with 10 or more 454 sequence reads per tumor. (**c**) PTEN
751 and CHUK alteration occurrences and Oncoprint across bBioPortal datasets. (**d**) PTEN and CHUK protein coding alterations and
752 co-occurrence values from cBioPortal. *(panels **c** and **d** continue on next page)*

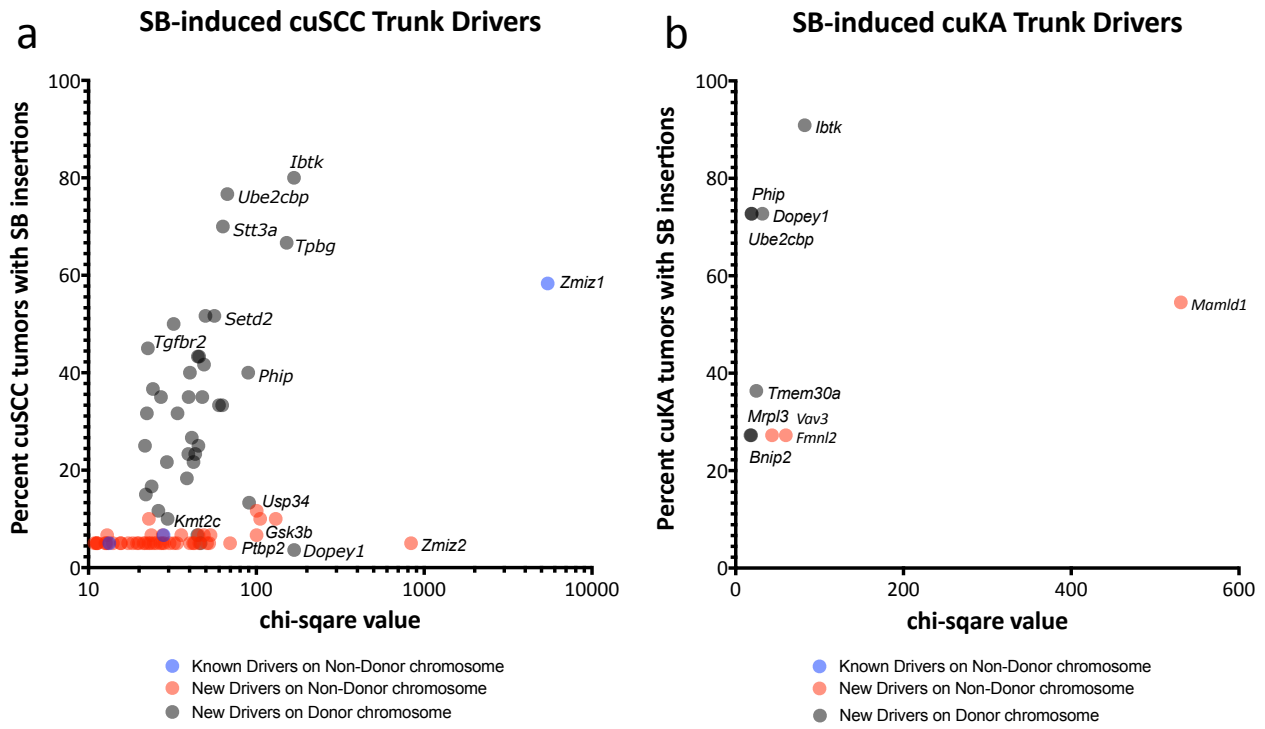

c
Cancer Type Summary
Alteration Frequency
60%
50%
40%
30%
20%
10%
0%
Mutation
Fusion
Amplification
Deep Deletion
Multiple Alterations
Endometrial Cancer
Prostate Cancer, NOS
Glioma
Prostate Cancer
Uterine Sarcoma
Cancer of Unknown Primary
Anal Cancer
Penile Cancer
Melanoma
Cervical Cancer
Colorectal Adenocarcinoma
Small Cell Lung Cancer
Breast Cancer
Ovarian Cancer
Colorectal Cancer
Esophagogastric Cancer
Skin Cancer, Non-Melanoma
Non-Small Cell Lung Cancer
Breast Sarcoma
Head and Neck Cancer
Soft Tissue Sarcoma
Bladder Cancer
Renal Cell Carcinoma
Gastrointestinal Stromal Tumor
Upper Tract Urothelial Carcinoma
Hepatobiliary Cancer
Ampullary Carcinoma
Germ Cell Tumor
Non-Hodgkin Lymphoma
Thyroid Cancer
Bone Cancer
Mesothelioma
Pancreatic Cancer
Pheochromocytoma
Salivary Gland Cancer
Embryonal Tumor
Leukemia
Myelodysplasia
Adrenocortical Carcinoma
Lymphoid Cancer
Multiple Myeloma
Nerve Sheath Tumor
Thymic Tumor
Sex Cord Stromal Tumor
CNS Cancer
Histiocytosis
Appendiceal Cancer
Study of origin
PTEN
9%
CHUK
1.1%
Genetic Alteration
Amplification
Deep Deletion
Fusion
Inframe Mutation (putative driver)
Inframe Mutation (unknown significance)
Missense Mutation (putative driver)
Missense Mutation (unknown significance)
No alterations
Not sequenced
Truncating Mutation (putative driver)
Truncating Mutation (unknown significance)


d
PTEN
R130*/G/L/P/Q/Qfs*4
Mutations
284
0
DSPc
PTEN_C2
0
100
200
300
403aa
CHUK
V411L/X411_splice
Mutations
5
0
Pkinase
IKK..
0
100
200
300
400
500
600
700
745aa
Modify Query
Querying 41834 samples in 170 studies
Gene Set / Pathway is altered in 3812 (9.1%) of queried samples
OncoPrint
Cancer Types Summary
Mutual Exclusivity
Expression
Mutations
Download
Bookmark
The query contains no gene pair with mutually exclusive alterations, and 1 gene pair with co-occurrent alterations (1 significant).
Mutual exclusivity
Co-occurrence
Significant only
Columns
Gene A
Gene B
Neither
A Not B
B Not A
Both
Log Odds Ratio
p-Value
Tendency
CHUK
PTEN
38022
190
3493
129
2.000
<0.001
Co-occurrence
Significant
Showing 1-1 of 1

753

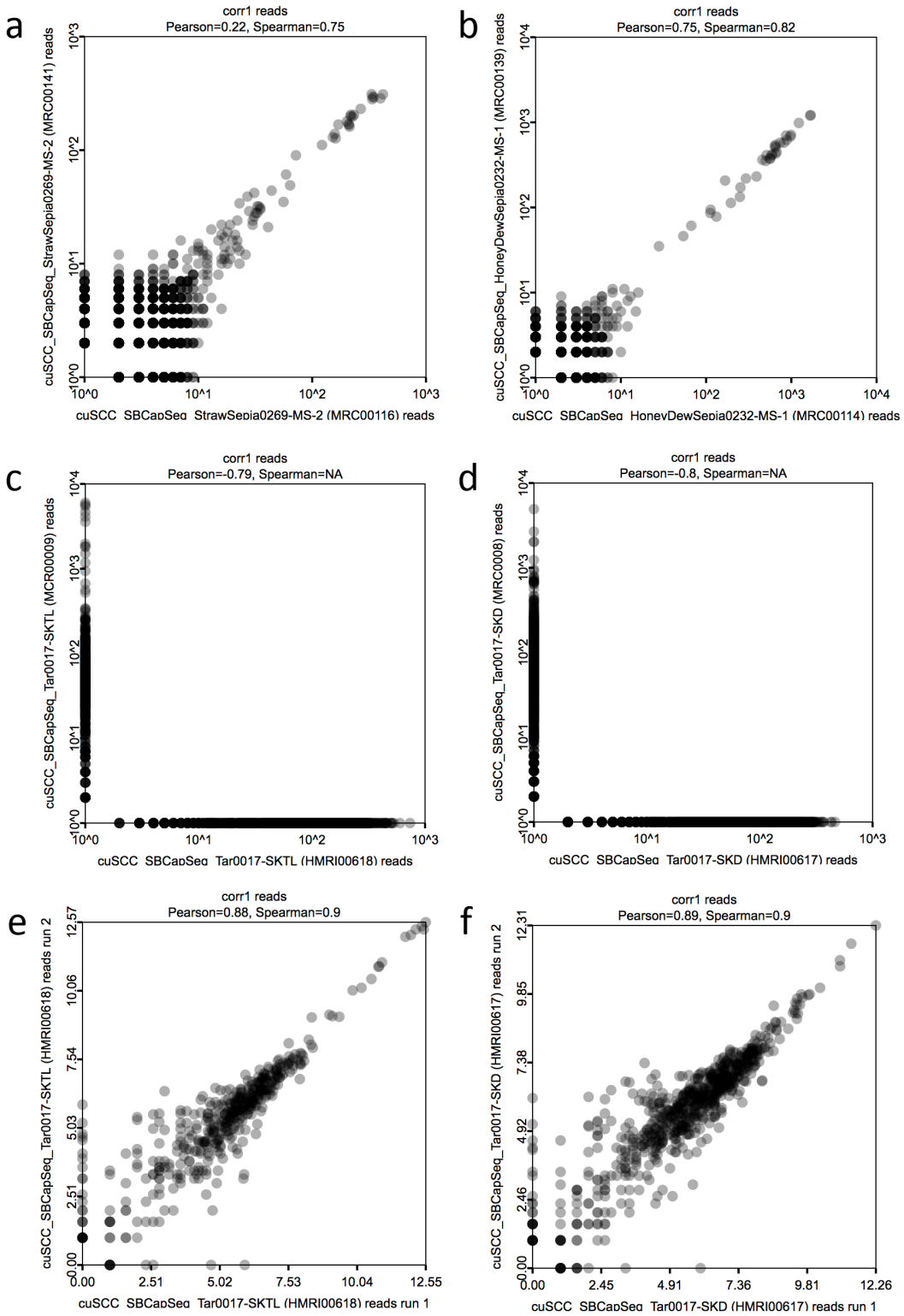


Supplementary Figure 5: Comparison of SB insertion results using SBCapSeq versus splink_454 methods. Representative plots of the individual SB insertion sites based on read depth from six individual cuSCC genomes (**a-f**) from genomic DNAs created from bulk tumor specimens from biological replicate libraries (independent library workflows applied to the same biological specimen isolate) comparing SBCapSeq (*x*-axis) versus splink_454 (*y*-axis) methods. Compared to Splink_454, the SBCapSeq method identifies more SB insertion events and overall higher read depth sites, however many of the high read depth sites were highly reproducible. Across the individual cuSCC genomes in the cohort, we observe highly variable reproducibility (from low to moderate) between the sequencing methods, supported by both Pearson's and Spearman's correlation metrics (**a-f**). We find SBCapSeq to be universally superior to 454 sequencing with little difference between the various SBCapSeq protocol variants tested, including SBCapSeq v1 (**a-b**) or SBCapSeq v2 (**c-d**) with individual captured libraries or SBCapSeq v2 (**e-f**) with multiplex pooled captured libraries.



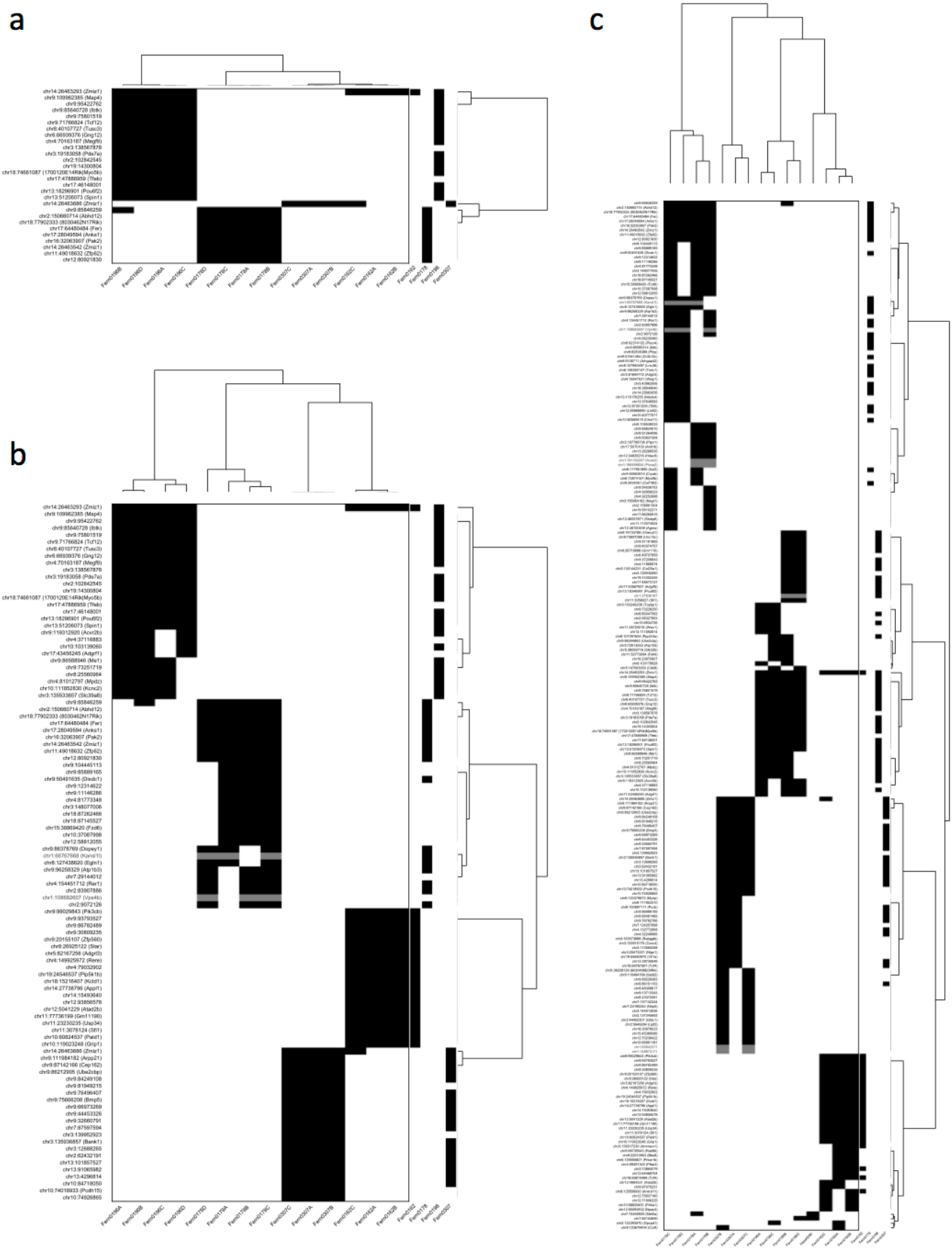
Supplementary Figure 6: Candidate trunk drivers mutated in SB-induced cuSCC with Splink_454 sequencing. (**a**) Trunk cuSCC Drivers genes from SB Driver Analysis applied to SBCapSeq insertion data (**Supplementary Table 4**) with read depth cutoff of 300 on 84 late-stage cuSCC genomes. Trunk cuSCC Driver genes. (**b**) Trunk cuKA Drivers genes from SB Driver Analysis applied to SBCapSeq insertion data (**Supplementary Table 7**) with read depth cutoff of 300 on 11 early-stage cuKA genomes.

Supplementary Figure 7: Evaluating the reproducibility of SBCapSeq results from bulk cuSCC and normal skin specimens. Representative plots of the individual SB insertion sites based on read depth from two individual cuSCC genomes (**a-b**) and two unselected skin cell genomes from histologically normal skin (**c-d**) from genomic DNAs created from bulk tumor specimens from biological replicate libraries (independent library workflows applied to the same biological specimen isolate) comparing library 1 (*x*-axis) to library 2 (*y*-axis). Biological reproducibility of cuSCC specimen libraries is high, supported by both Pearson's and Spearman's correlation metrics, using all of SB insertion sites with read depths of 2 or higher (**a-b**). In contrast, biological reproducibility of normal skin specimen libraries is exceptionally low, supported by both Pearson's and Spearman's correlation metrics, and fails entirely to identify the same SB insertion sites (**c-d**), indicating that these are likely background passenger insertion events that are not reproducibly sequenced. However, plotting of the individual SB insertion sites based on read depth from two technical replicates (the same library preparation) from two separate sequencing runs confirms very high reproducibility between individual libraries at high-to-moderate read depths, supported by both Pearson's (r= 0.48) and Spearman's (ρ=0.49 or undetermined) correlation metrics (**e-f**), further supporting the lack of biological replication is most likely due to non-clonal background passenger insertion events and do not result from flaws in the SBCapSeq protocol *per se*. *(Image on next page)*

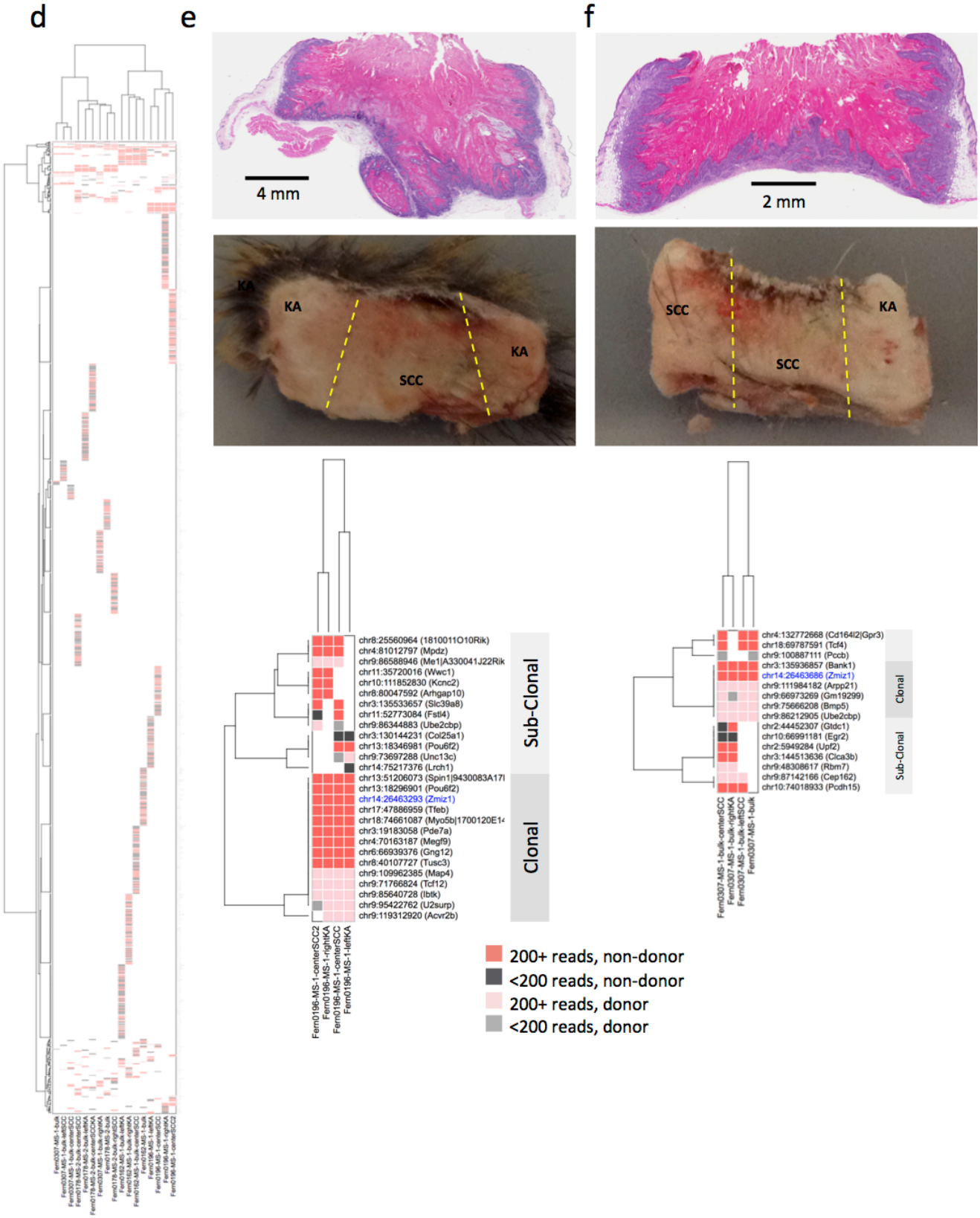

a
corr1 reads
Pearson=0.22, Spearman=0.75
cuSCC_SBCapSeq_StrawSepia0269-MS-2 (MRC00141) reads
cuSCC_SBCapSeq_StrawSepia0269-MS-2 (MRC00116) reads
b
corr1 reads
Pearson=0.75, Spearman=0.82
cuSCC_SBCapSeq_HoneyDewSepia0232-MS-1 (MRC00139) reads
cuSCC_SBCapSeq_HoneyDewSepia0232-MS-1 (MRC00114) reads
c
corr1 reads
Pearson=-0.79, Spearman=NA
cuSCC_SBCapSeq_Tar0017-SKTL (MCR00009) reads
cuSCC_SBCapSeq_Tar0017-SKTL (HMRI00618) reads
d
corr1 reads
Pearson=-0.8, Spearman=NA
cuSCC_SBCapSeq_Tar0017-SKD (MRC0008) reads
cuSCC_SBCapSeq_Tar0017-SKD (HMRI00617) reads
e
corr1 reads
Pearson=0.88, Spearman=0.9
cuSCC_SBCapSeq_Tar0017-SKTL (HMRI00618) reads run 2
cuSCC_SBCapSeq_Tar0017-SKTL (HMRI00618) reads run 1
f
corr1 reads
Pearson=0.89, Spearman=0.9
cuSCC_SBCapSeq_Tar0017-SKD (HMRI00617) reads run 2
cuSCC_SBCapSeq_Tar0017-SKD (HMRI00617) reads run 1

784

785 **Supplementary Figure 8: Multi-region SBCapSeq analysis of bi-lesional keratinocyte-derived skin masses containing**
786 **distinct cuKA and cuSCC histologies.** Hierarchical two-dimensional clustering (Hamming distance with the Ward method) of
787 recurrent genic SBCapSeq insertion data from four skin masses containing distinct cuKA and cuSCC regions with ≥4 (**a**), ≥3 or
788 more (**b**), or ≥2 (**c**), recurrent SB insertions. At each recurrent threshold, four distinct specimen (*x*-axis) and gene (*y*-axis) clades,
789 each pertaining to a single bi-lesional mass, were detected, thus demonstrating clonal outgrowth. Genomic DNA was isolated
790 from adjacent, non-overlapping serial sections and three or four regions dissected from histologically distinct regions; the tumor
791 reference, consisting of a cross-section of each mass (including all regions), is plotted as independent tracks on the right of each
792 graph (see **panels e-f** and **Fig. 3 a-b**). (**d**) Hierarchical two-dimensional clustering (Hamming distance with the Ward method) of
793 recurrent genic SBCapSeq insertion data from four skin masses containing distinct cuKA and cuSCC regions. Four distinct
794 specimen (*x*-axis) and gene (*y*-axis) clades, each pertaining to a single bi-lesional mass, were detected, demonstrating clonal
795 identities. *Zmiz1* was the only gene recurrently mutated across all samples with high read depths; most SB insertions were
796 private to a single sample and consistently had lower read depths suggesting these are unselected passenger insertions. Genomic
797 DNA was isolated from adjacent serial sections and included a "bulk" reference consisting of a cross-section of the entire mass,
798 and three or four regions dissected from histologically distinct regions (see **panels e-f**). (**e-f**) Two masses with Aperio slide scans
799 of H&E stained FFPE specimens showing distinct cuKA and cuSCC histologies suggestive of clonal evolution of cuSCC
800 differentiation within cuKA masses and the adjacent flash frozen tissue prior to sectioning for genomic DNA isolation.
801 Hierarchical two-dimensional clustering (Hamming distance with the Ward method) of recurrent genic SBCapSeq insertion data
802 shown below each specimen demonstrates trunk (common to all samples) and sub-clonal, lesion-specific SB insertions. *(Images*
803 *on next 2 pages)*



804


d
e
f
4 mm
2 mm
KA
KA
KA
SCC
SCC
SCC
KA
Sub-Clonal
Clonal
Clonal
Sub-Clonal
200+ reads, non-donor
<200 reads, non-donor
200+ reads, donor
<200 reads, donor

805

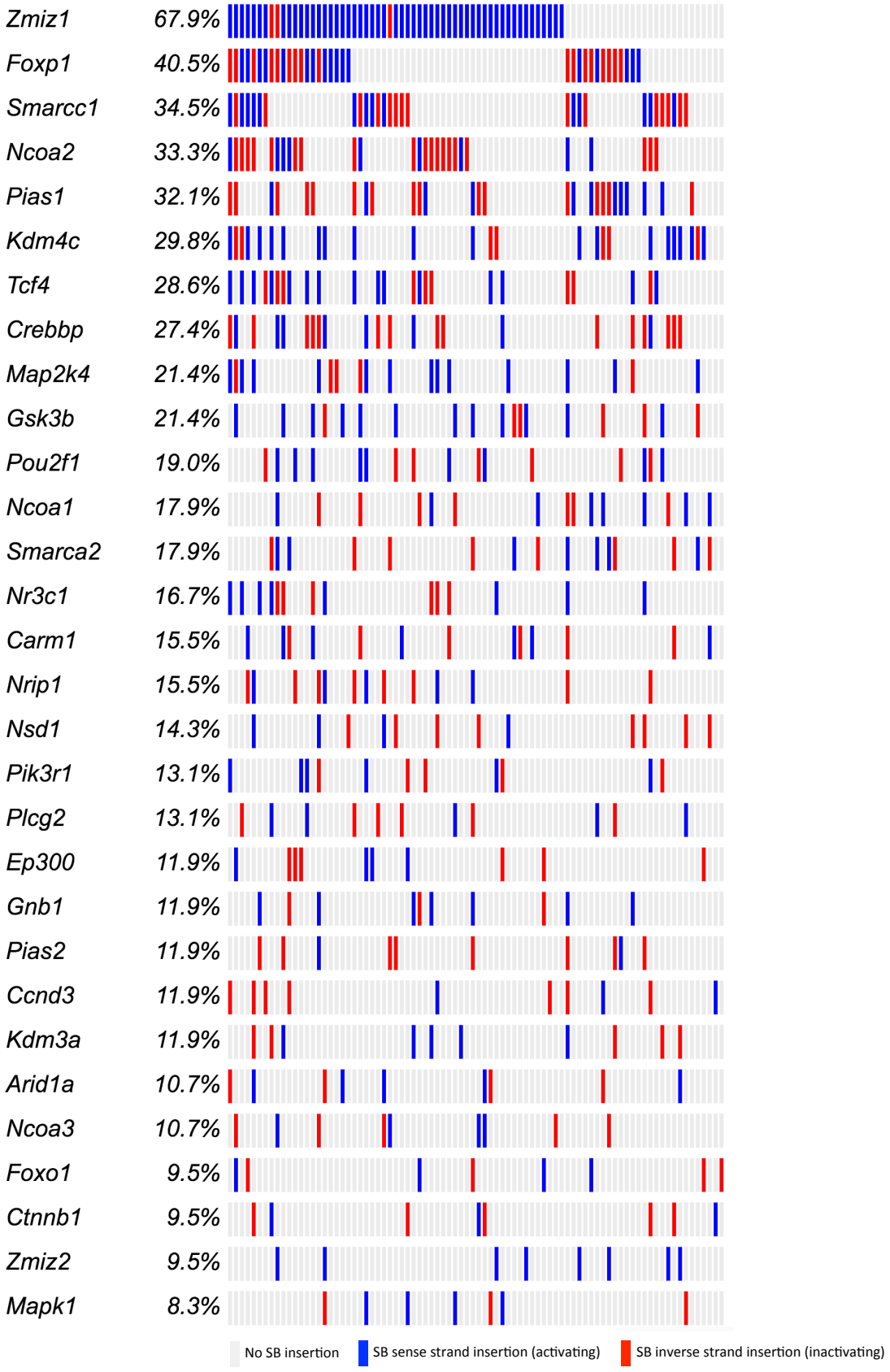
a

Curated Biological Pathways or Processes via Enrichr*	SB|cuSCC genomes altered	Number of Genes	Representative human orthologs of SB|cuSCC drivers mutated in >10% of human cuSCC genomes
Androgen Receptor Signaling	100%	40	*ARID1A, **CREBBP**, EP300, FOXO1, GATA2, HRAS, KDM3A, KDM4C, NCOA1, **NCOA2**, NCOA3, NRIP1, NSD1, PKN1, PLCG2, SMARCC1, **ZMIZ1**, **ZMIZ2***
NOTCH Signaling	100%	62	*AAK1, APC, CBL, CREBBP, EP300, EPN2, GATA2, ITCH, LRP6, MOV10, NCOA1, NCOR1, NOTCH1, NOTCH2, STAT3, TLE1, TNRC6A, TNRC6B, TNRC6C, **ZMIZ1***
EGFR Signaling	100%	72	*ADCY7, ADCY9, CBL, FYN, HRAS, IQGAP1, MAP3K1, MOV10, NF1, PDE1A, PRKAR1A, PTK2, RASA3, RASGRF1, SPTAN1, SPTBN1, SYNGAP1, TNRC6A, TNRC6B, TNRC6C*
MAPK Signaling	100%	131	*APBB1IP, CACNA1D, DAB2IP, FLNB, FOXO1, FYN, HRAS, IGF1R, IQGAP1, ITCH, LRRK2, MAP3K1, MAP4K4, MECOM, MINK1, MOV10, NBR1, NCOA3, NCOR1, NF1, PAK2, PLCG2, PTK2, RASA3, RASAL2, RASGRF1, SETX, SPTAN1, SPTBN1, STAT3, SYNGAP1, TNRC6A, TNRC6B, TNRC6C*
TGFβ-WNT Signaling	100%	147	*APC, ARID1A, BCL9L, BICC1, BMPR2, CAPRIN2, CBL, CDC73, CDH1, CDH13, CDK14, CREBBP, CTNND1, DAB2IP, EP300, EP400, FAT1, FAT2, HRAS, INO80, ITCH, LRP6, LRRK2, MOV10, NCOR1, NOTCH1, PARD3, PLCG2, RNF111, ROCK2, SMARCC1, SPTBN1, TIAM1, TLE1, TNRC6A, TNRC6B, TNRC6C, UBR5, USP34, USP47, USP9X, WNK1*
Chromatin and Histone Modification	98%	64	*ARID1A, ARID1B, ATRX, BAZ2A, BPTF, **CREBBP**, EP300, EP400, KANSL1, KAT6A, KAT6B, KDM3A, KDM3B, KDM4C, KDM5A, KDM5B, KDM6A, KMT2C, MECOM, NCOA1, **NCOA2**, NCOA3, NCOR1, NSD1, NSD2, NSD3, PBRM1, PHF20, SETD1A, SETD2, SETD3, SMARCC1, TOX3, ZZZ3*
HIPPO Signaling	93%	30	*APC, BMPR2, CDH1, PARD3*
ZMIZ paralog Family	87%	30	*PIAS3 (ZMIZ5), **ZMIZ1, ZMIZ2***

*Significant enrichment of cuSCC candidate drivers in selected pathways, $q<0.001$

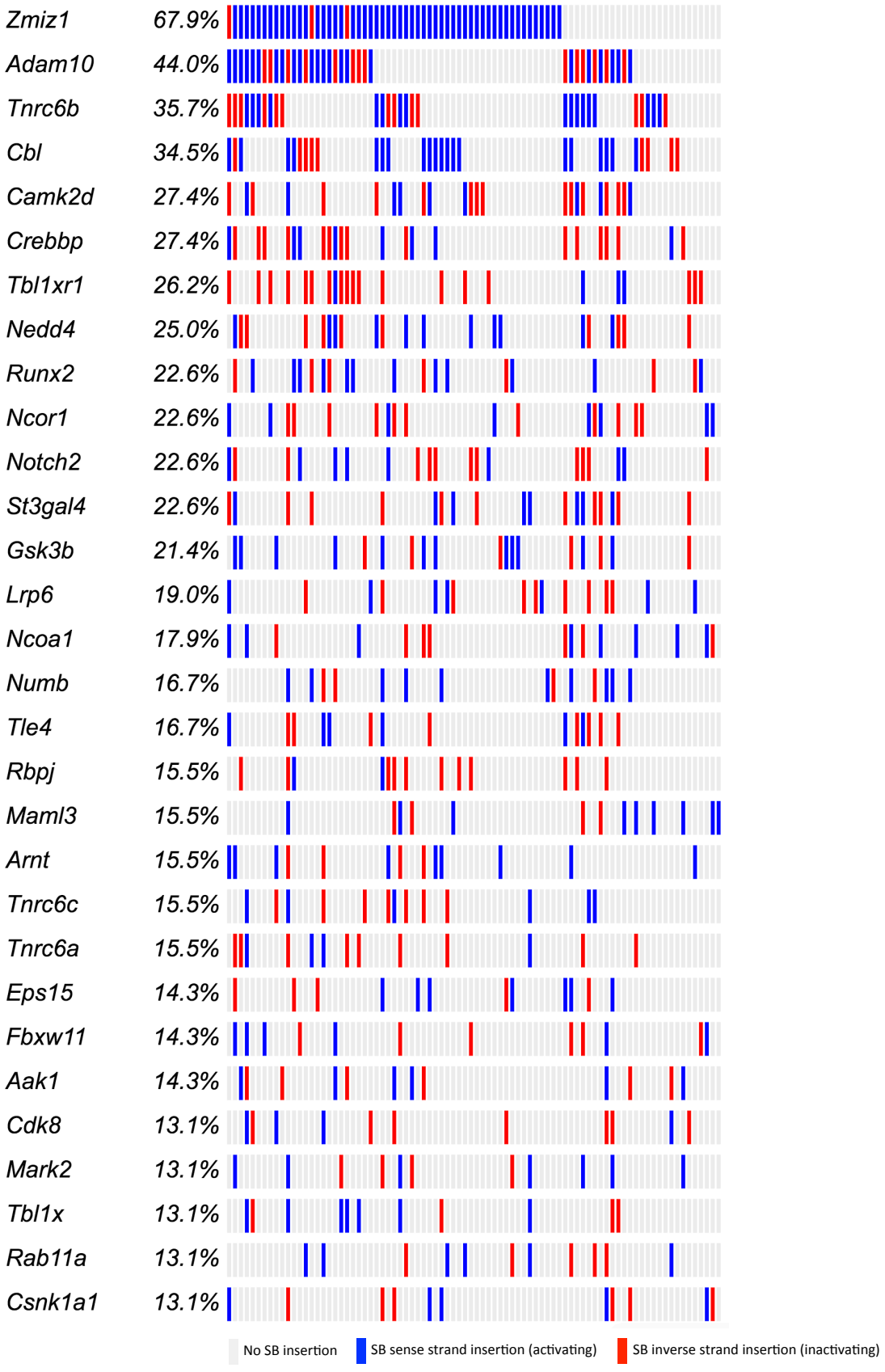
Supplementary Figure 9: Curated biological pathways and processes enriched within SB-induced cuSCC. (**a**) Table of significant pathways collated from pathway enrichment categories in **Supplementary Table 28**. Oncoprints of the top 30 mutated cuSCC candidate drivers enriched in (**b**) Androgen Receptor Signaling, (**c**) NOTCH Signaling, (**d**) EGFR Signaling, (**e**) MAPK Signaling, (**f**) TGFβ-WNT Signaling, (**g**) Chromatin and Histone Modification, (**h**) HIPPO Signaling, and (**i**) ZMIZ paralog family. *(Panels b–i on next 8 pages)*

b


Zmiz1 67.9%
Foxp1 40.5%
Smarcc1 34.5%
Ncoa2 33.3%
Pias1 32.1%
Kdm4c 29.8%
Tcf4 28.6%
Crebbp 27.4%
Map2k4 21.4%
Gsk3b 21.4%
Pou2f1 19.0%
Ncoa1 17.9%
Smarca2 17.9%
Nr3c1 16.7%
Carm1 15.5%
Nrip1 15.5%
Nsd1 14.3%
Pik3r1 13.1%
Plcg2 13.1%
Ep300 11.9%
Gnb1 11.9%
Pias2 11.9%
Ccnd3 11.9%
Kdm3a 11.9%
Arid1a 10.7%
Ncoa3 10.7%
Foxo1 9.5%
Ctnnb1 9.5%
Zmiz2 9.5%
Mapk1 8.3%
No SB insertion
SB sense strand insertion (activating)
SB inverse strand insertion (inactivating)

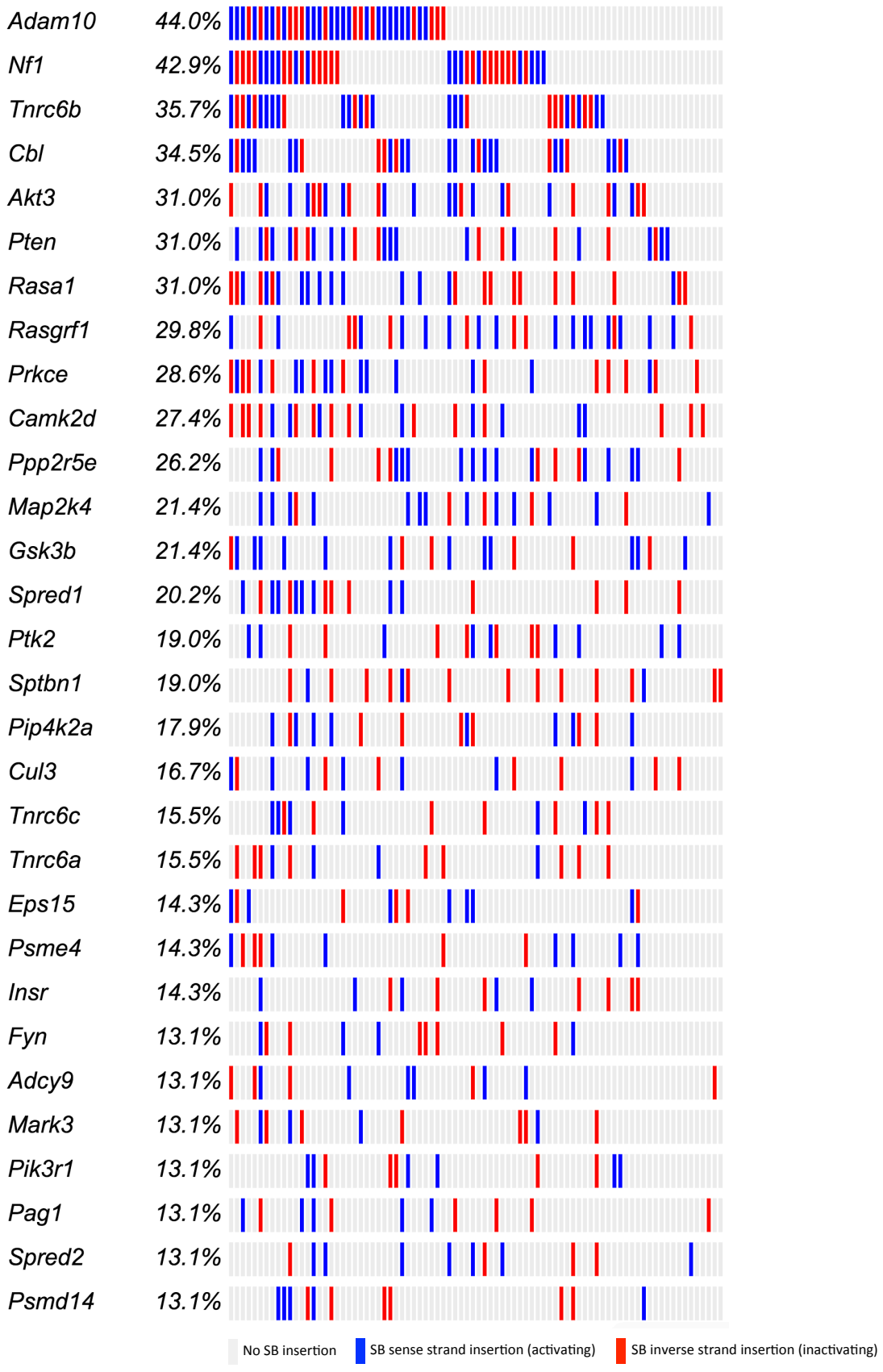
812

C


Zmiz1 67.9%
Adam10 44.0%
Tnrc6b 35.7%
Cbl 34.5%
Camk2d 27.4%
Crebbp 27.4%
Tbl1xr1 26.2%
Nedd4 25.0%
Runx2 22.6%
Ncor1 22.6%
Notch2 22.6%
St3gal4 22.6%
Gsk3b 21.4%
Lrp6 19.0%
Ncoa1 17.9%
Numb 16.7%
Tle4 16.7%
Rbpj 15.5%
Maml3 15.5%
Arnt 15.5%
Tnrc6c 15.5%
Tnrc6a 15.5%
Eps15 14.3%
Fbxw11 14.3%
Aak1 14.3%
Cdk8 13.1%
Mark2 13.1%
Tbl1x 13.1%
Rab11a 13.1%
Csnk1a1 13.1%
No SB insertion
SB sense strand insertion (activating)
SB inverse strand insertion (inactivating)

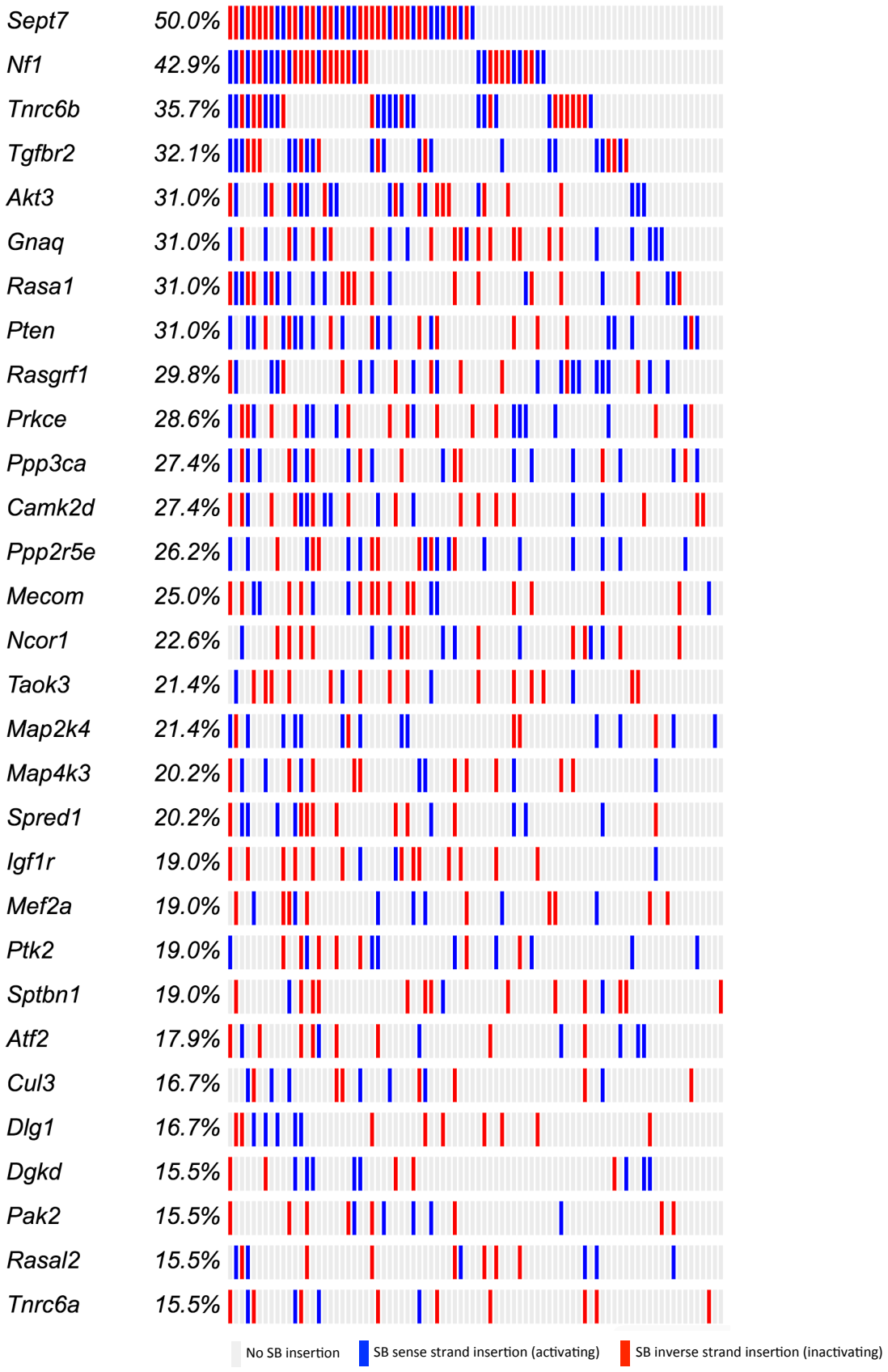
813

d


Adam10 44.0%
Nf1 42.9%
Tnrc6b 35.7%
Cbl 34.5%
Akt3 31.0%
Pten 31.0%
Rasa1 31.0%
Rasgrf1 29.8%
Prkce 28.6%
Camk2d 27.4%
Ppp2r5e 26.2%
Map2k4 21.4%
Gsk3b 21.4%
Spred1 20.2%
Ptk2 19.0%
Sptbn1 19.0%
Pip4k2a 17.9%
Cul3 16.7%
Tnrc6c 15.5%
Tnrc6a 15.5%
Eps15 14.3%
Psme4 14.3%
Insr 14.3%
Fyn 13.1%
Adcy9 13.1%
Mark3 13.1%
Pik3r1 13.1%
Pag1 13.1%
Spred2 13.1%
Psmd14 13.1%
No SB insertion
SB sense strand insertion (activating)
SB inverse strand insertion (inactivating)

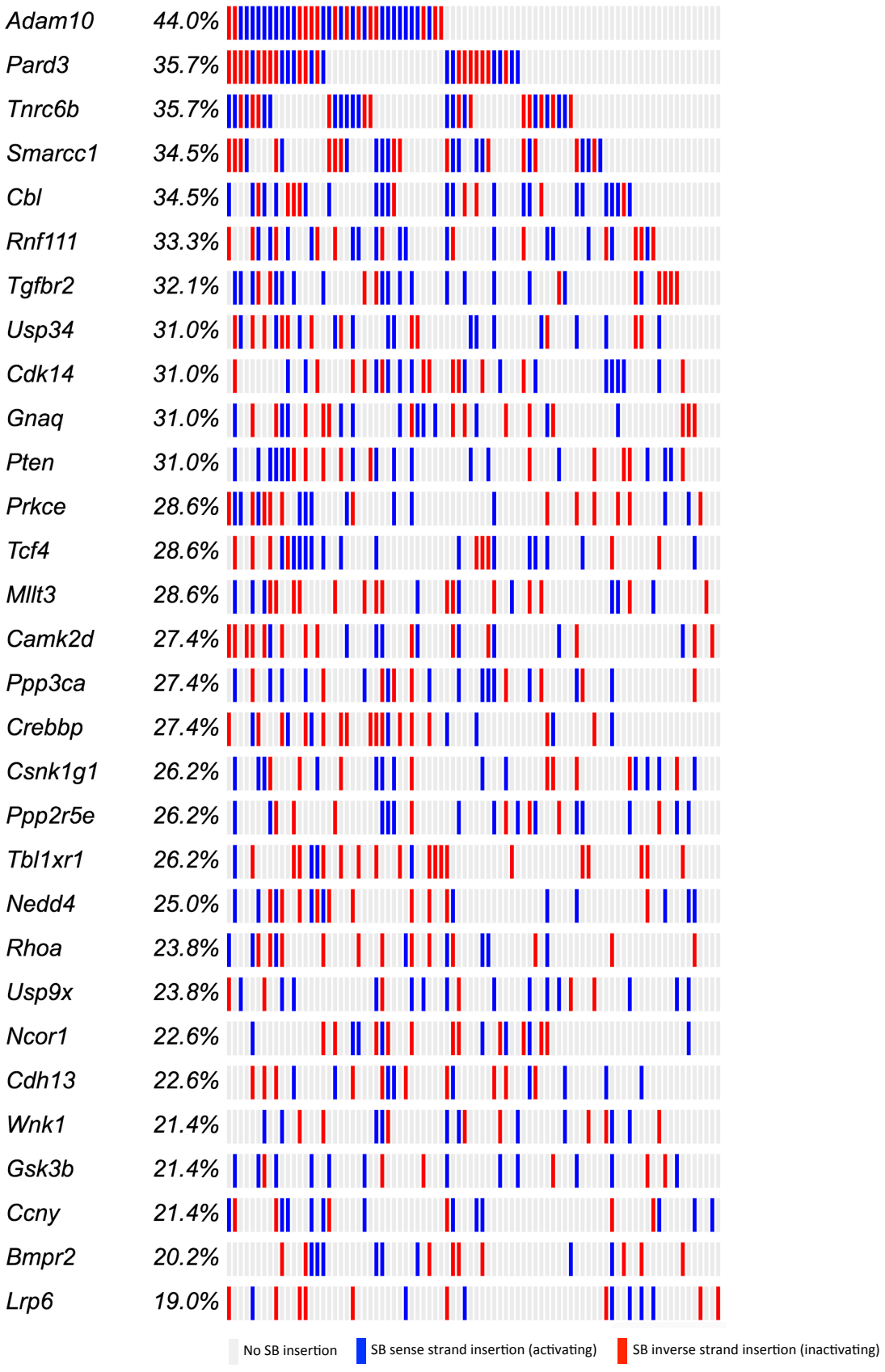
814

e


Sept7 50.0%
Nf1 42.9%
Tnrc6b 35.7%
Tgfbr2 32.1%
Akt3 31.0%
Gnaq 31.0%
Rasa1 31.0%
Pten 31.0%
Rasgrf1 29.8%
Prkce 28.6%
Ppp3ca 27.4%
Camk2d 27.4%
Ppp2r5e 26.2%
Mecom 25.0%
Ncor1 22.6%
Taok3 21.4%
Map2k4 21.4%
Map4k3 20.2%
Spred1 20.2%
Igf1r 19.0%
Mef2a 19.0%
Ptk2 19.0%
Sptbn1 19.0%
Atf2 17.9%
Cul3 16.7%
Dlg1 16.7%
Dgkd 15.5%
Pak2 15.5%
Rasal2 15.5%
Tnrc6a 15.5%
No SB insertion
SB sense strand insertion (activating)
SB inverse strand insertion (inactivating)

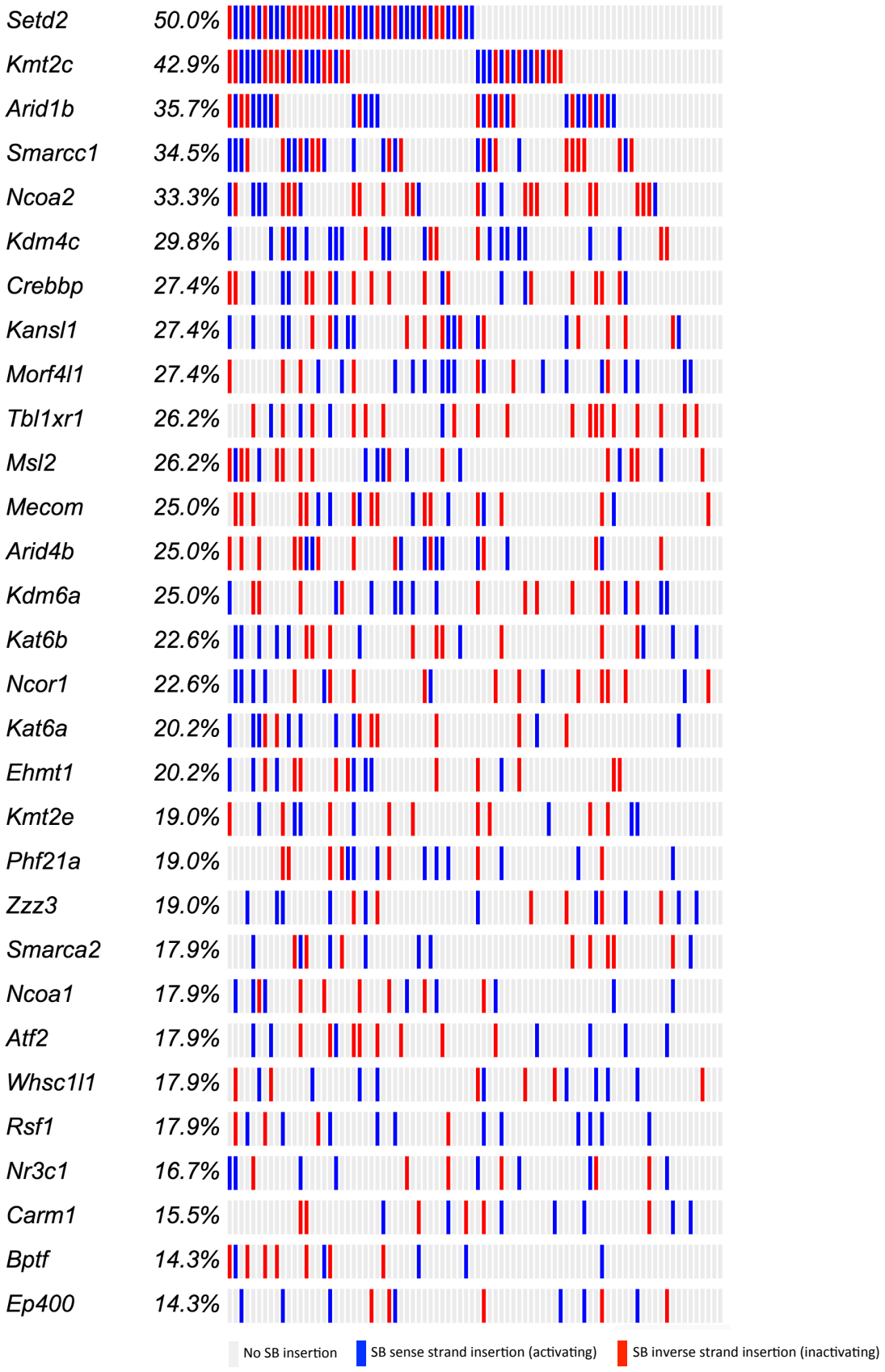
815

f


Adam10 44.0%
Pard3 35.7%
Tnrc6b 35.7%
Smarcc1 34.5%
Cbl 34.5%
Rnf111 33.3%
Tgfbr2 32.1%
Usp34 31.0%
Cdk14 31.0%
Gnaq 31.0%
Pten 31.0%
Prkce 28.6%
Tcf4 28.6%
Mllt3 28.6%
Camk2d 27.4%
Ppp3ca 27.4%
Crebbp 27.4%
Csnk1g1 26.2%
Ppp2r5e 26.2%
Tbl1xr1 26.2%
Nedd4 25.0%
Rhoa 23.8%
Usp9x 23.8%
Ncor1 22.6%
Cdh13 22.6%
Wnk1 21.4%
Gsk3b 21.4%
Ccny 21.4%
Bmpr2 20.2%
Lrp6 19.0%
No SB insertion
SB sense strand insertion (activating)
SB inverse strand insertion (inactivating)

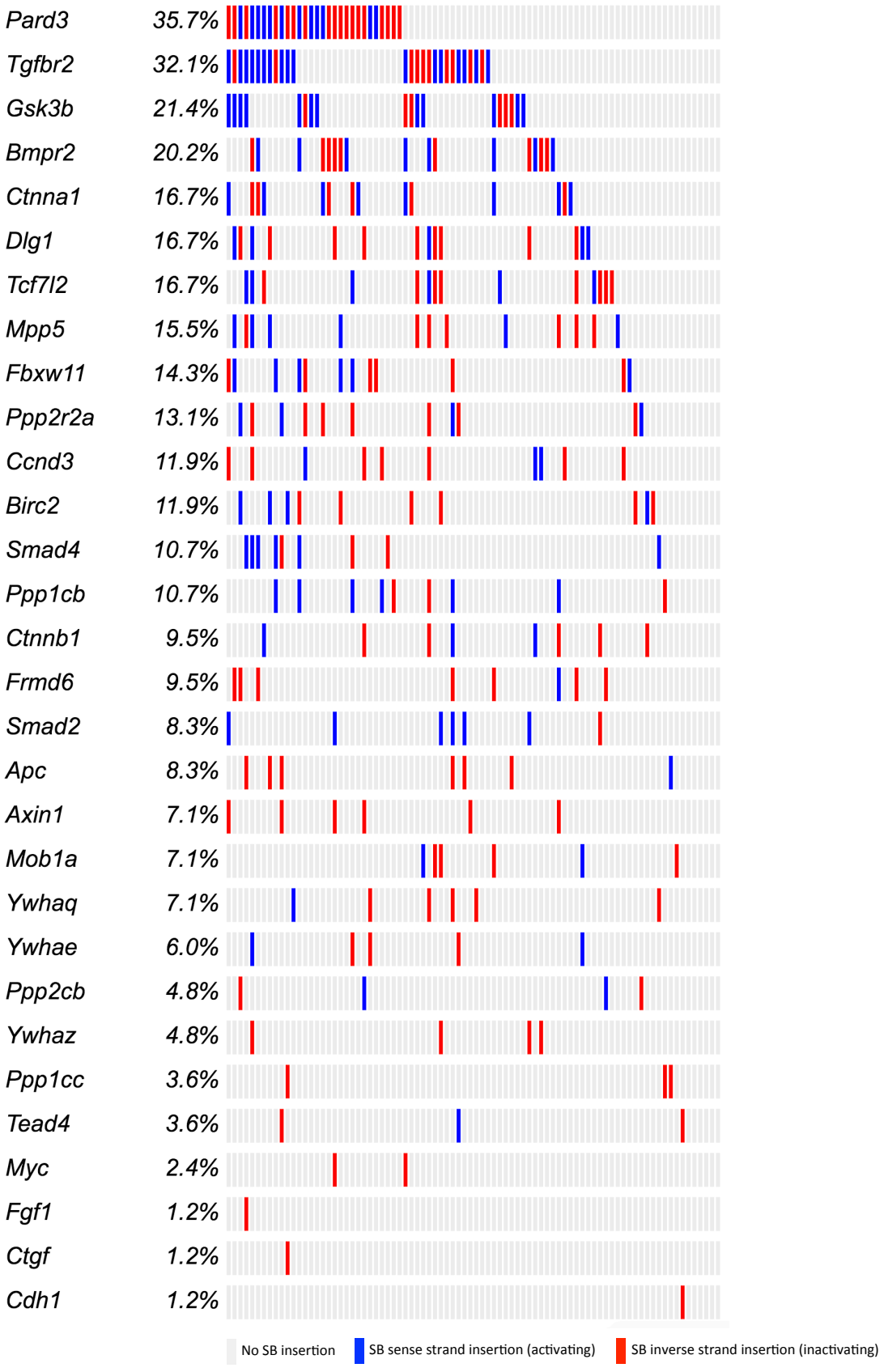
816

g


Setd2 50.0%
Kmt2c 42.9%
Arid1b 35.7%
Smarcc1 34.5%
Ncoa2 33.3%
Kdm4c 29.8%
Crebbp 27.4%
Kansl1 27.4%
Morf4l1 27.4%
Tbl1xr1 26.2%
Msl2 26.2%
Mecom 25.0%
Arid4b 25.0%
Kdm6a 25.0%
Kat6b 22.6%
Ncor1 22.6%
Kat6a 20.2%
Ehmt1 20.2%
Kmt2e 19.0%
Phf21a 19.0%
Zzz3 19.0%
Smarca2 17.9%
Ncoa1 17.9%
Atf2 17.9%
Whsc1l1 17.9%
Rsf1 17.9%
Nr3c1 16.7%
Carm1 15.5%
Bptf 14.3%
Ep400 14.3%
No SB insertion
SB sense strand insertion (activating)
SB inverse strand insertion (inactivating)

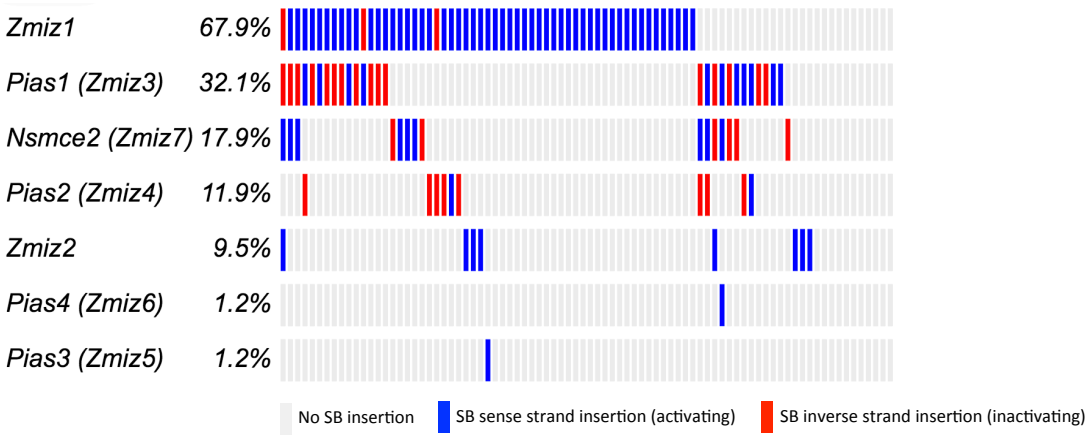
817

h

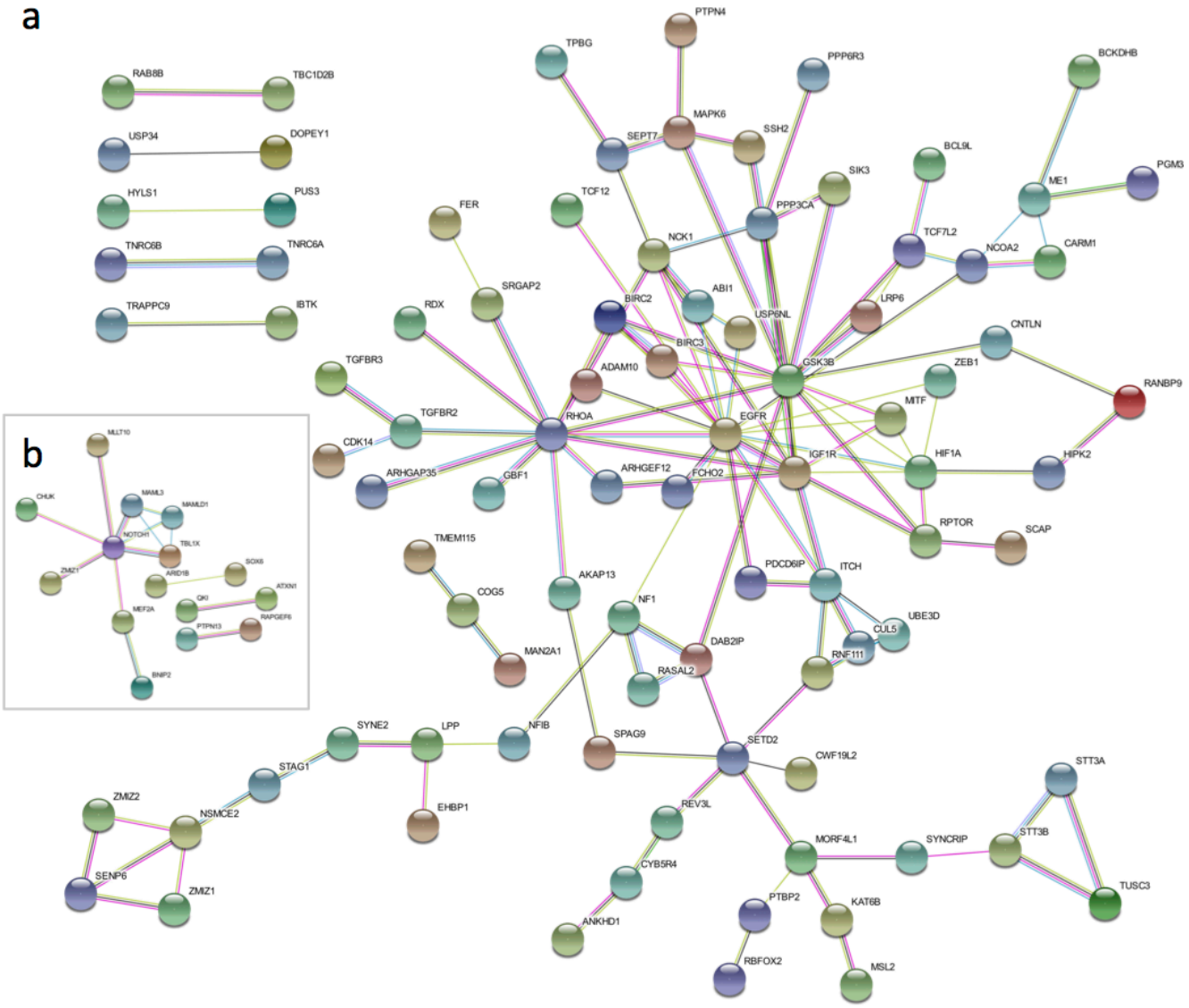

Pard3 35.7%
Tgfbr2 32.1%
Gsk3b 21.4%
Bmpr2 20.2%
Ctnna1 16.7%
Dlg1 16.7%
Tcf7l2 16.7%
Mpp5 15.5%
Fbxw11 14.3%
Ppp2r2a 13.1%
Ccnd3 11.9%
Birc2 11.9%
Smad4 10.7%
Ppp1cb 10.7%
Ctnnb1 9.5%
Frmd6 9.5%
Smad2 8.3%
Apc 8.3%
Axin1 7.1%
Mob1a 7.1%
Ywhaq 7.1%
Ywhae 6.0%
Ppp2cb 4.8%
Ywhaz 4.8%
Ppp1cc 3.6%
Tead4 3.6%
Myc 2.4%
Fgf1 1.2%
Ctgf 1.2%
Cdh1 1.2%
No SB insertion
SB sense strand insertion (activating)
SB inverse strand insertion (inactivating)

818

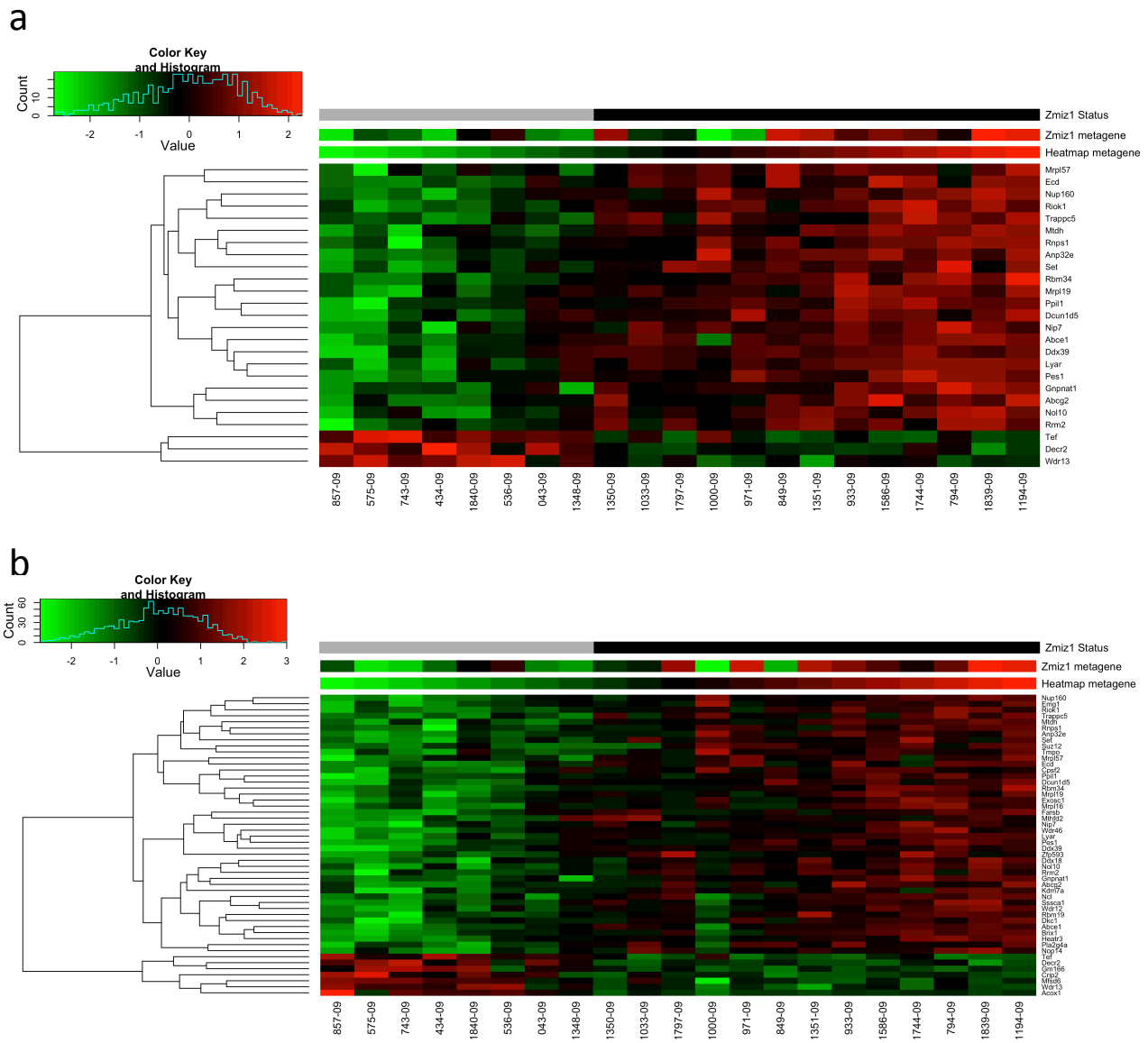
i


Zmiz1 67.9%
Pias1 (Zmiz3) 32.1%
Nsmce2 (Zmiz7) 17.9%
Pias2 (Zmiz4) 11.9%
Zmiz2 9.5%
Pias4 (Zmiz6) 1.2%
Pias3 (Zmiz5) 1.2%
No SB insertion
SB sense strand insertion (activating)
SB inverse strand insertion (inactivating)

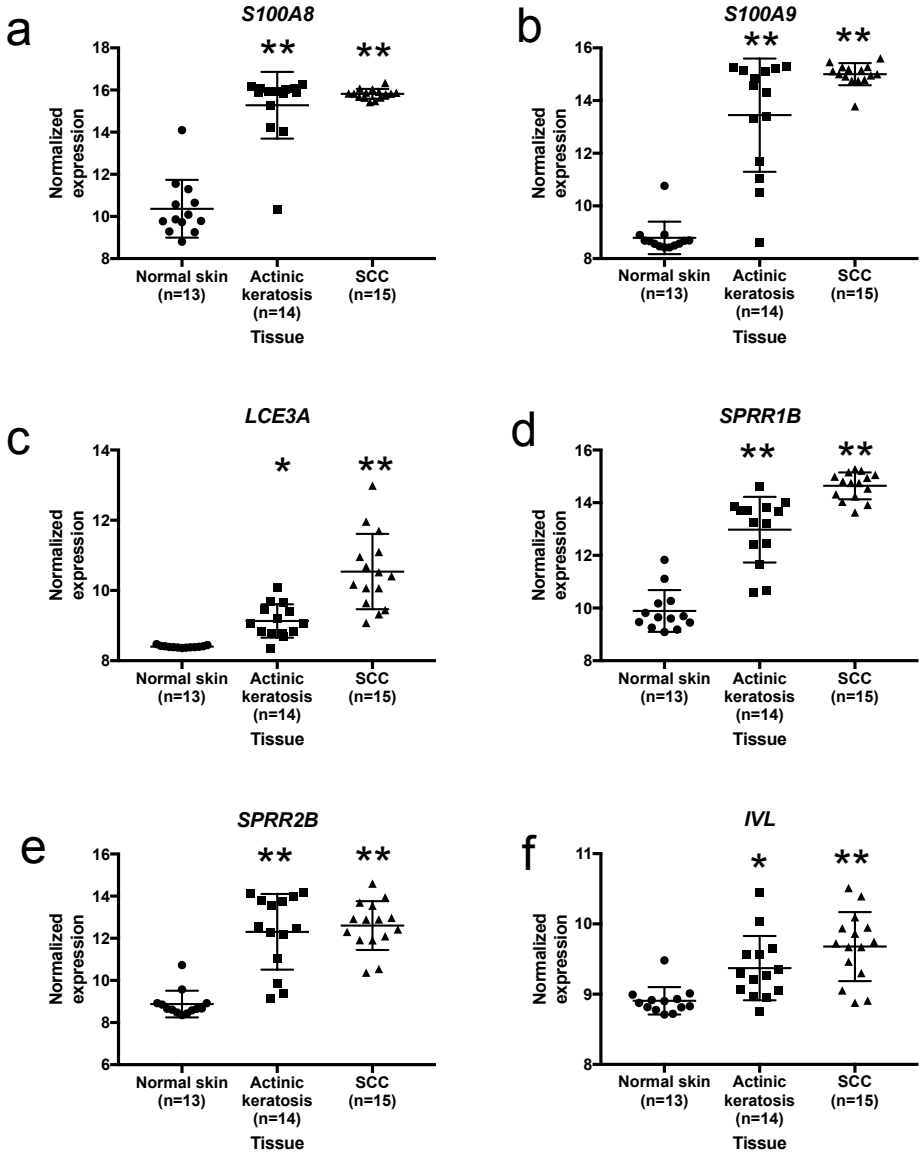
819



Supplementary Figure 10: Trunk driver STRING network analysis. (**a**) The SB|cuSCC Trunk Driver (n=84) network has significantly more known protein-protein interactions than expected by chance (P=8.55 × 10^{-6}8.55e-06, STRING enrichment analysis; number of nodes: 144, number of edges: 124, expected number of edges: 82 (**b**) The SB|cuKA Trunk Driver (n=62) network also has significantly more known protein-protein interactions than expected by chance (P=6.93 × 10^{-5}, STRING enrichment analysis; number of nodes: 37, number of edges: 14, expected number of edges: 4.

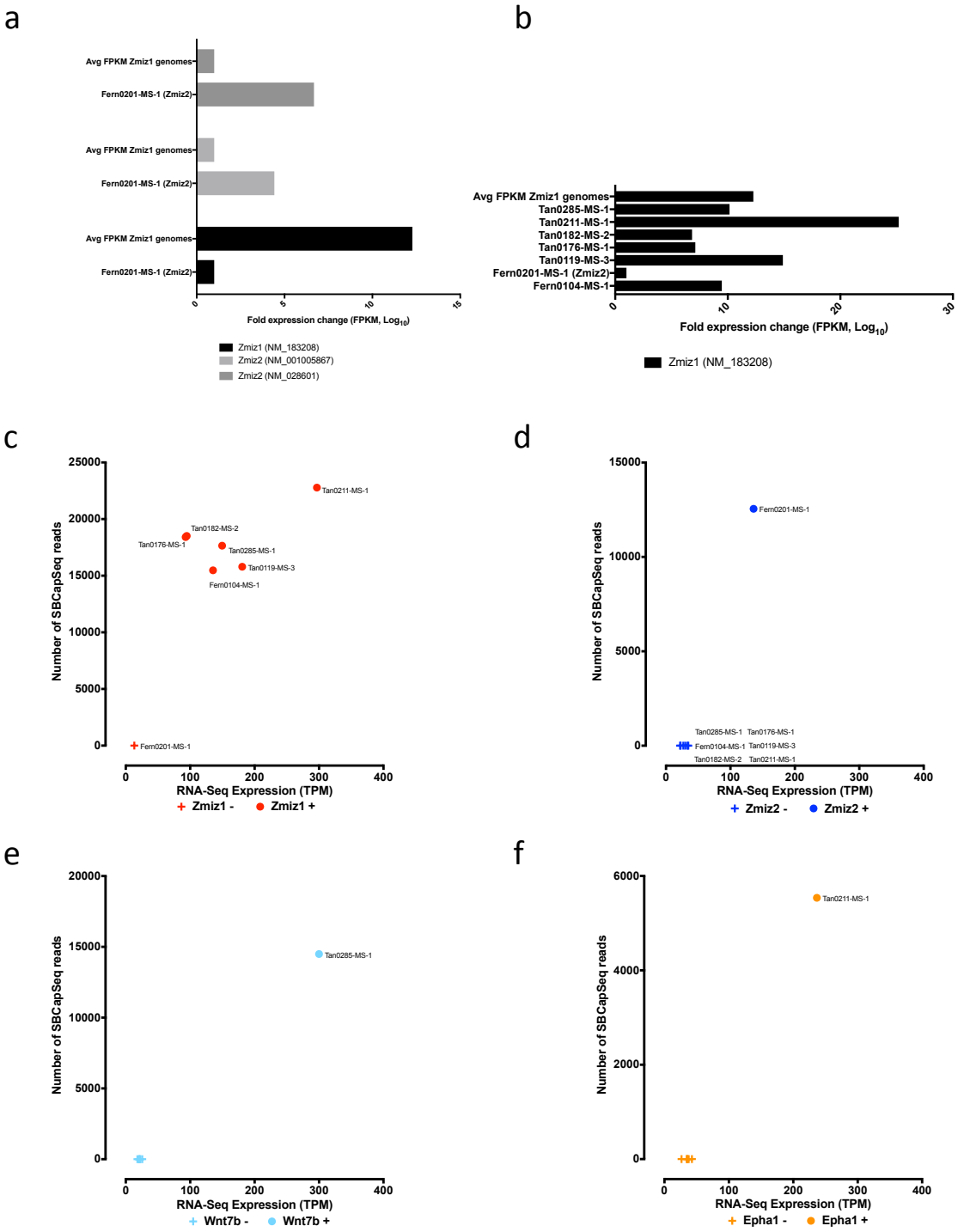


Supplementary Figure 11: SB effect on gene expression by microarray analysis. (**a**) Top 25 DE genes: *Mtdh*, *Anp32e*, *Ecd* are cuSCC progression drivers. (**b**) Top 50 DE genes: *Mtdh*, *Anp32e*, *Ecd*, *Dcun1d5*, and *Kdm7a* are cuSCC progression drivers. (**c**) Differentially expressed genes reported between KA and normal skin after Zmiz1 ectopic expression (Rogers et al.), confirmed with microarray analysis from SB|cuSCC tumors.

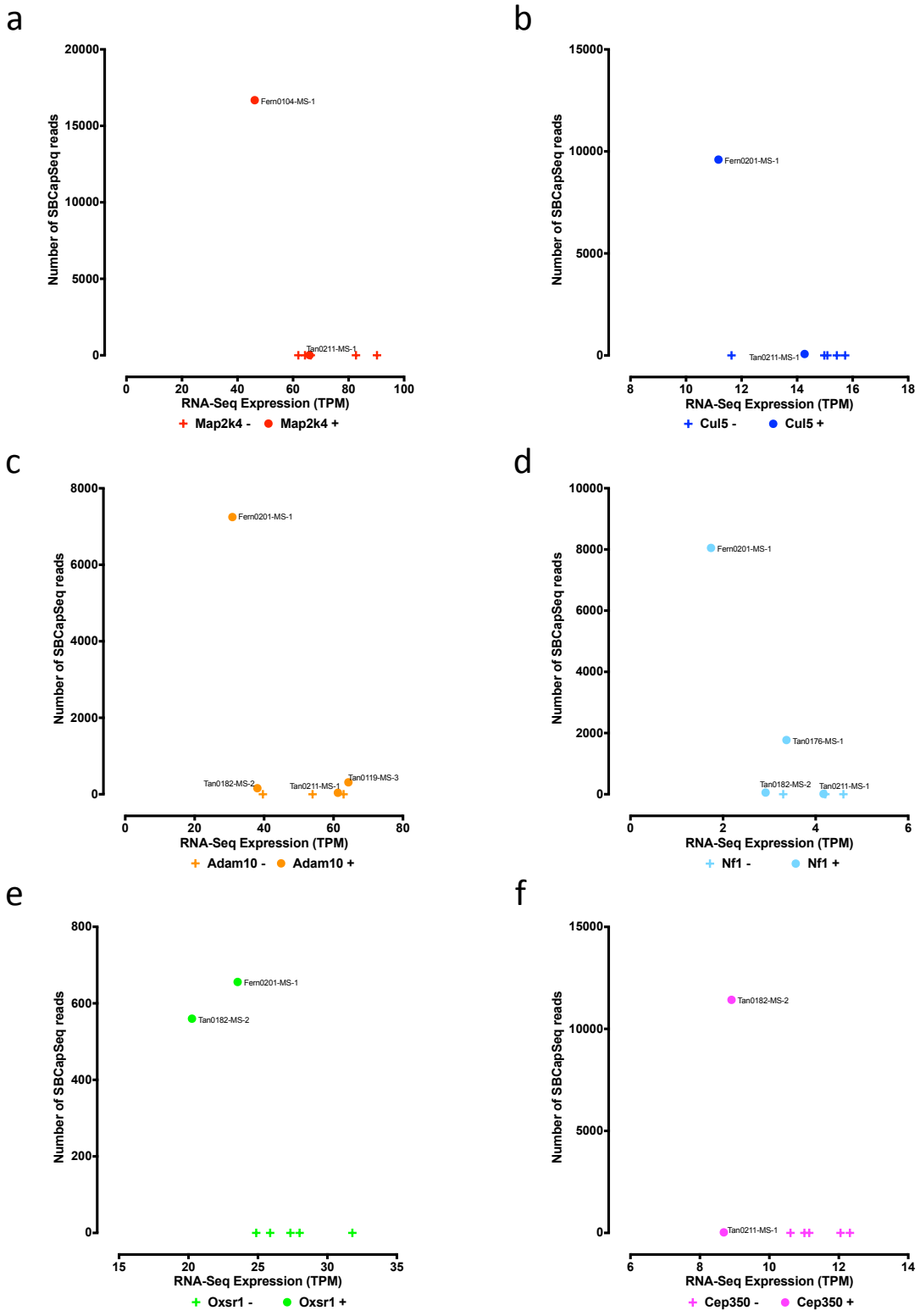


832

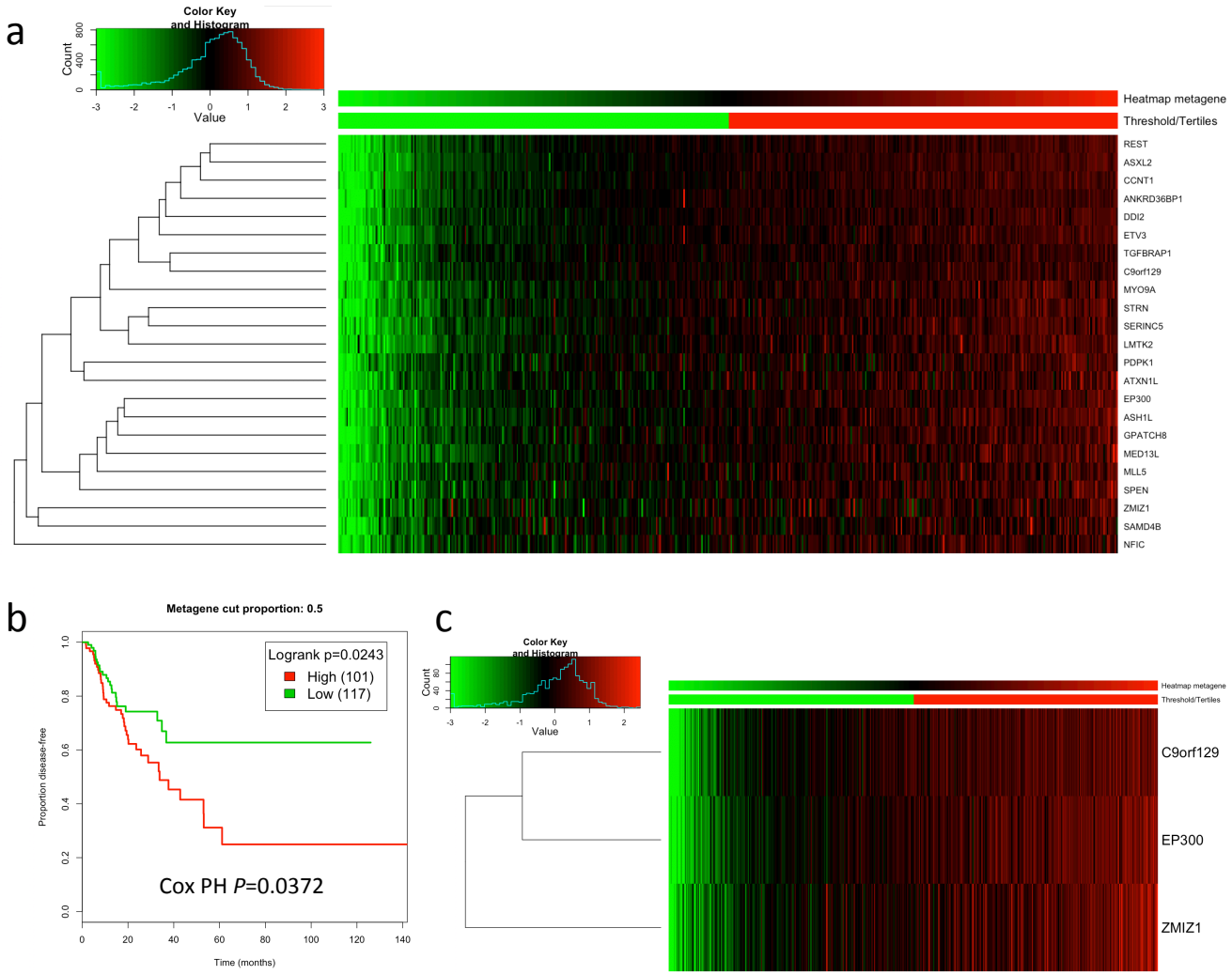
833 **Supplementary Figure 12: Zmiz1$^{\Delta N185}$ driven epidermal differentiation complex genes orthologs are significantly**
834 **unregulated during human keratinocyte transformation and cuSCC progression. (a-f)** Expression of selected EDC genes are
835 significantly unregulated in human actinic keratosis (n=14) and cuSCC (n=15) tissues, relative to normal skin (n=13). * adjusted *P*
836 <0.05 vs normal skin; **adjusted *P* < 0.0001 versus normal skin. GEO dataset: GSE32628 (Hameetman et al., 2013).



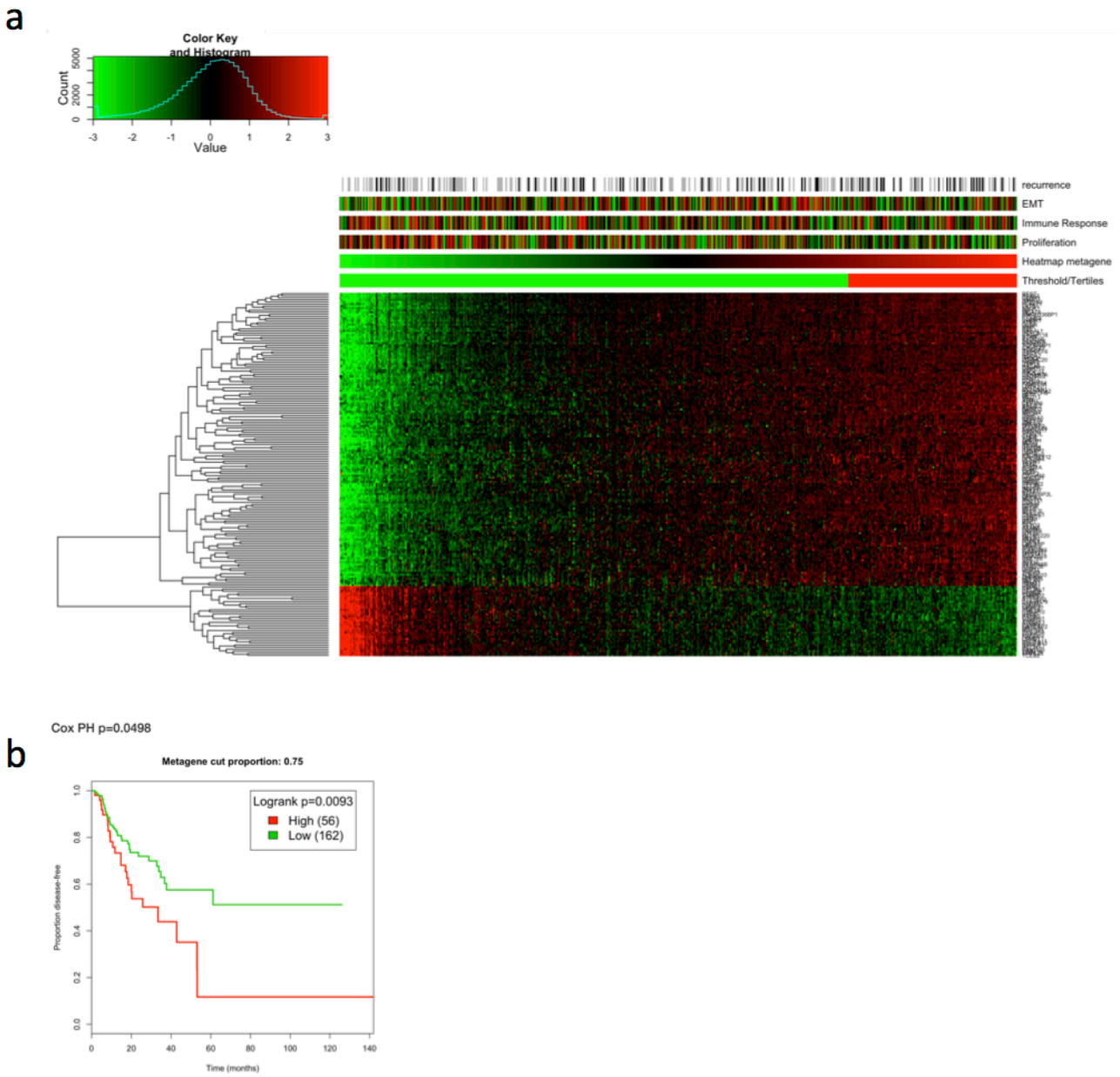
Supplementary Figure 13: Clonally selected SB insertions affect trunk driver proto-oncogene expression in SB-cuSCC genomes. (**a**) Summary of *Zmiz1* and *Zmiz2* transcripts identified from bulk analysis of cuSCC cells by wtRNA-seq are represented as the $\log_{10}$ ratio of fragments per kilobase of transcript per million (FPKM) from the average of 6 genomes with SB insertions into *Zmiz1* compared to the transcripts from the single genome with *Zmiz2* SB insertion, defined by FPKM mapped reads from ribo-depleted RNA. (**b**) Transcripts identified from bulk analysis of cuSCC cells by wtRNA-seq are represented as the $\log_{10}$ ratio of SB insertion containing compared to wild-type transcripts, defined by fragments per kilobase of transcript per million mapped reads from ribo-depleted. (c-f) Individual panels from **Figure 5e**: multifold induction of gene expression in cuSCC masses with (**+**) high read depth activating SB insertion events among 4 candidate oncogenic drivers compared with normal gene expression levels in cuSCC tumors without (**–**) SB insertions.



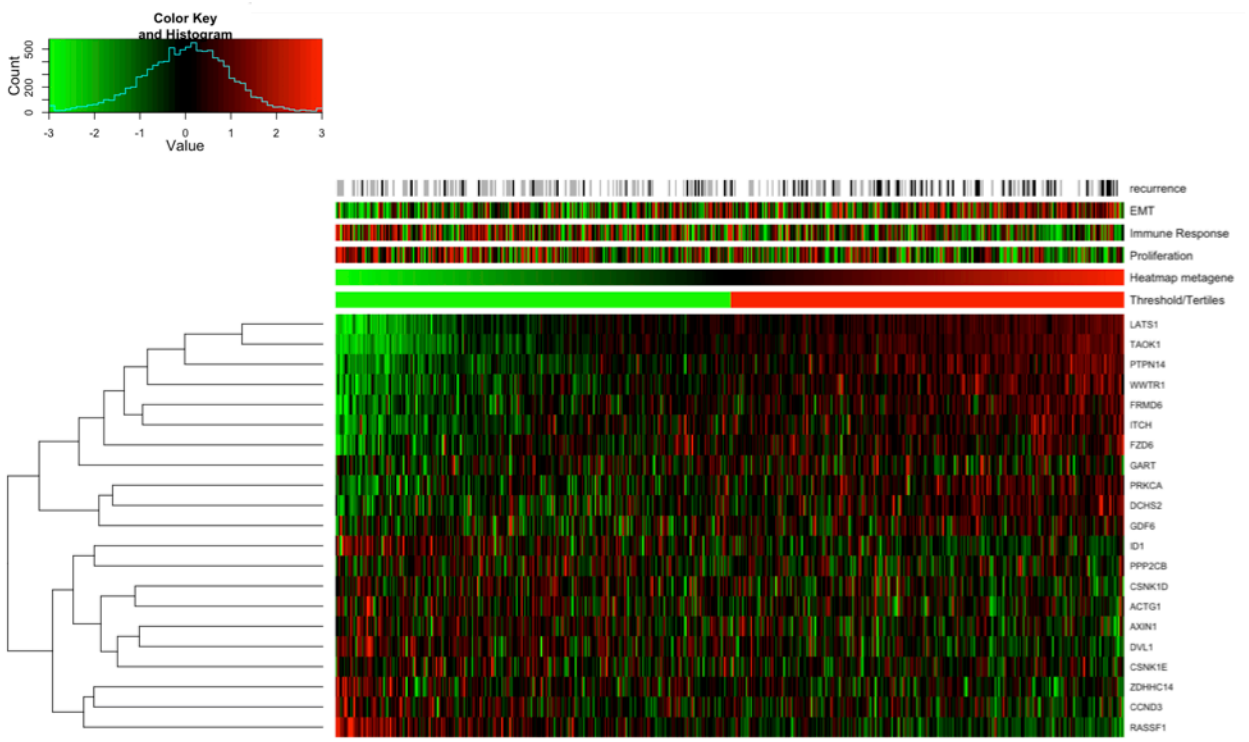
Supplementary Figure 14: Clonally selected SB insertions affect trunk driver genes by inactivating expression in SB-cuSCC genomes. (**a-f**) Individual panels from **Figure 5f**: reduced gene expression in cuSCC masses with (**+**) high read depth inactivating SB insertion events among 6 candidate tumor suppressor drivers compared with normal gene expression levels in cuSCC tumors without (**–**) SB insertions.

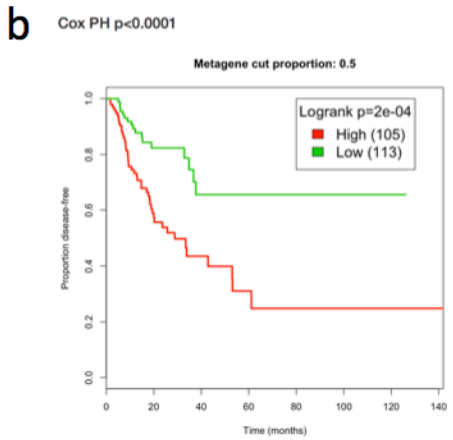


Supplementary Figure 15: *ZMIZ1* metagene within the TCGA Head & Neck Squamous Cell Carcinoma (hnSCC) RNA-seq dataset. (**a**) *ZMIZ1* metagene heatmap constructed using singular value decomposition of human hnSCC RNA-seq dataset from TCGA consisting of a 23-gene signature from **Figure 5e**. (**b**) Survival plot of patient outcomes in human hnSCC (**b**) based on the expression of a *ZMIZ1*–centric metagene heatmap (**c**) constructed using singular value decomposition of human hnSCC RNA-seq dataset from TCGA, reveal poor patient survival outcomes predicted by increasing expression of *ZMIZ1* and closely regulated genes consisting of a 3-gene signature: *ZMIZ1*, *EP300*, and *C9orf129*.

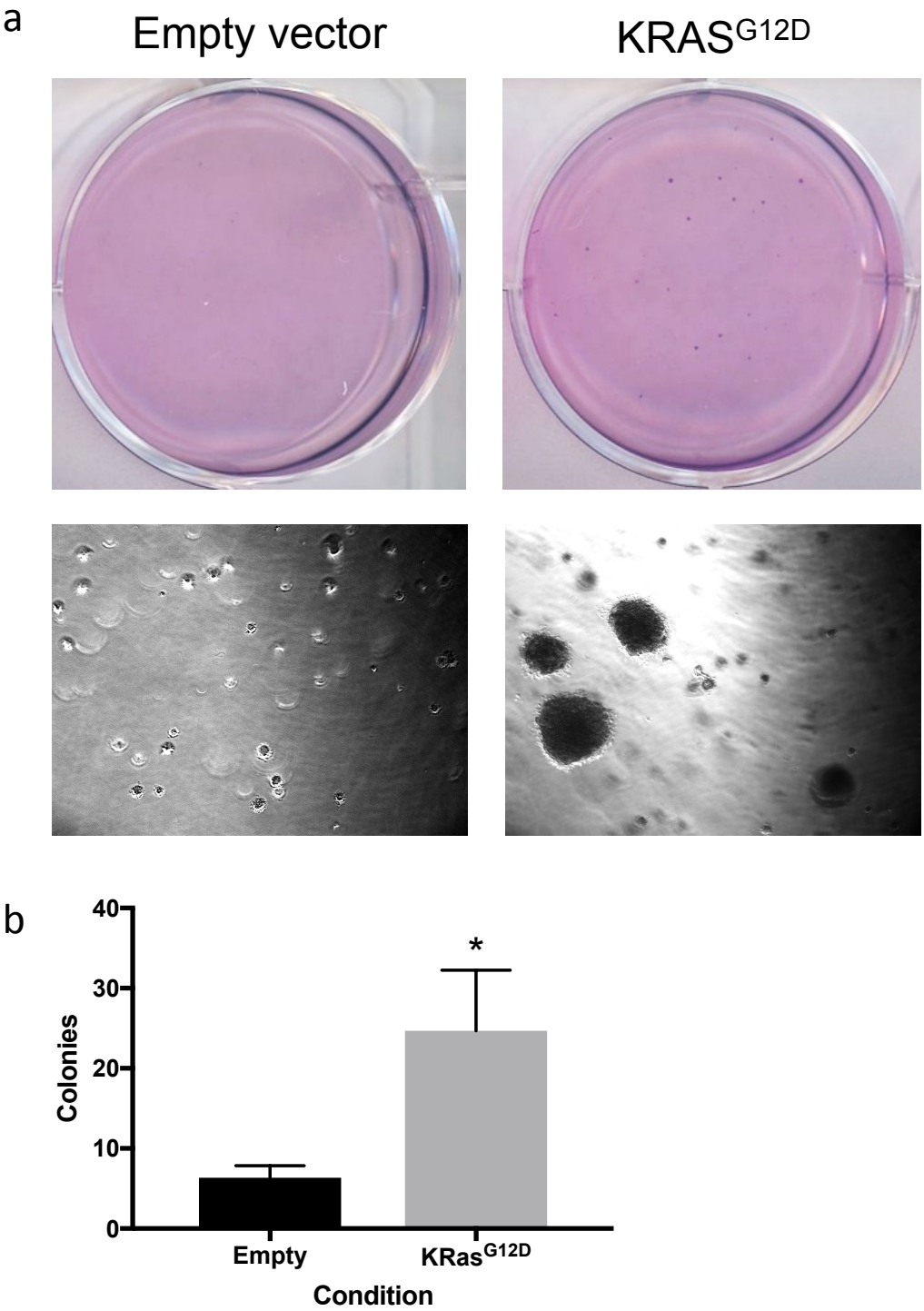


Supplementary Figure 16: *KMT2C (MLL3)* metagene within the TCGA Head & Neck Squamous Cell Carcinoma (hnSCC) RNA-seq dataset. (**a**) *KMT2C* metagene heatmap and (**b**) patient survival outcomes.





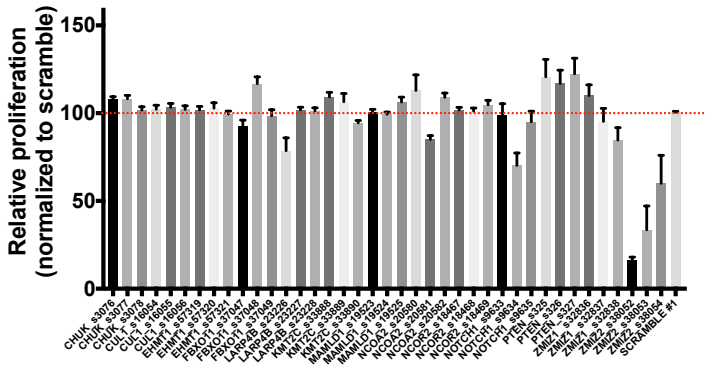
Supplementary Figure 17: HIPPO pathway metagene within the TCGA Head & Neck Squamous Cell Carcinoma (hnSCC) RNA-seq dataset. (**a**) HIPPO pathway metagene heatmap and (**b**) patient survival outcomes.



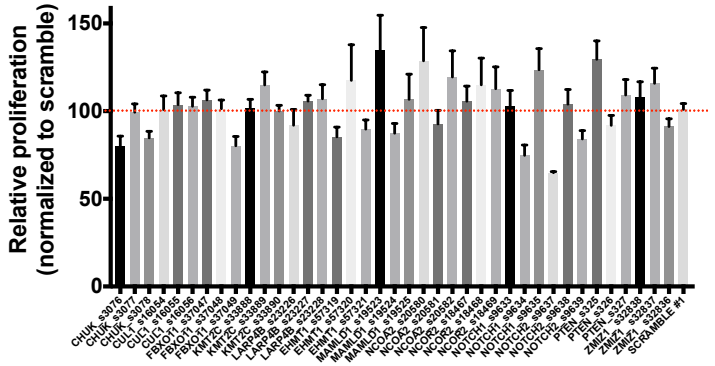
865

866 **Supplementary Figure 18: *In vitro* functional validation of oncogenic KRAS^{G12D} in immortal keratinocytes.** (**a-b**)
867 Assessment of keratinocyte transformation in response to expression of oncogenic KRAS^{G12D} by anchorage-independent soft agar
868 assay (n=3). (**a**) Images of colonies stained with crystal violet solution. (**b**) Statistical significance was tested by unpaired t-test.
869 *P = 0.015.

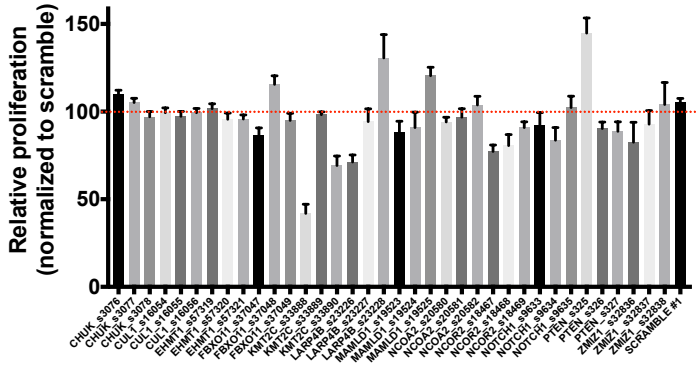
a



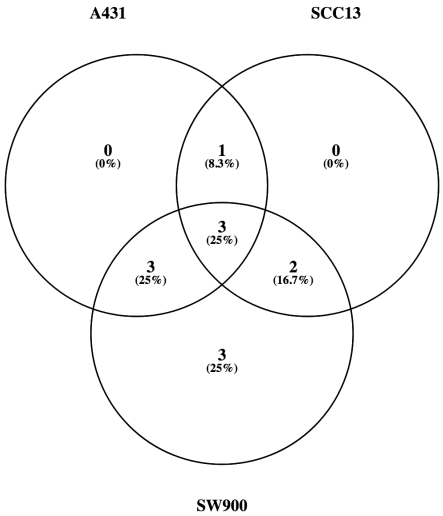
b



c



d



In "A431", "SCC13" and "SW900": *CHUK, NOTCH1,* and *PTEN*

In "A431" and "SCC13" only: *NCOA2*

In "A431" and "SW900" only: *FBXO11, LARP4B,* and *ZMIZ1*

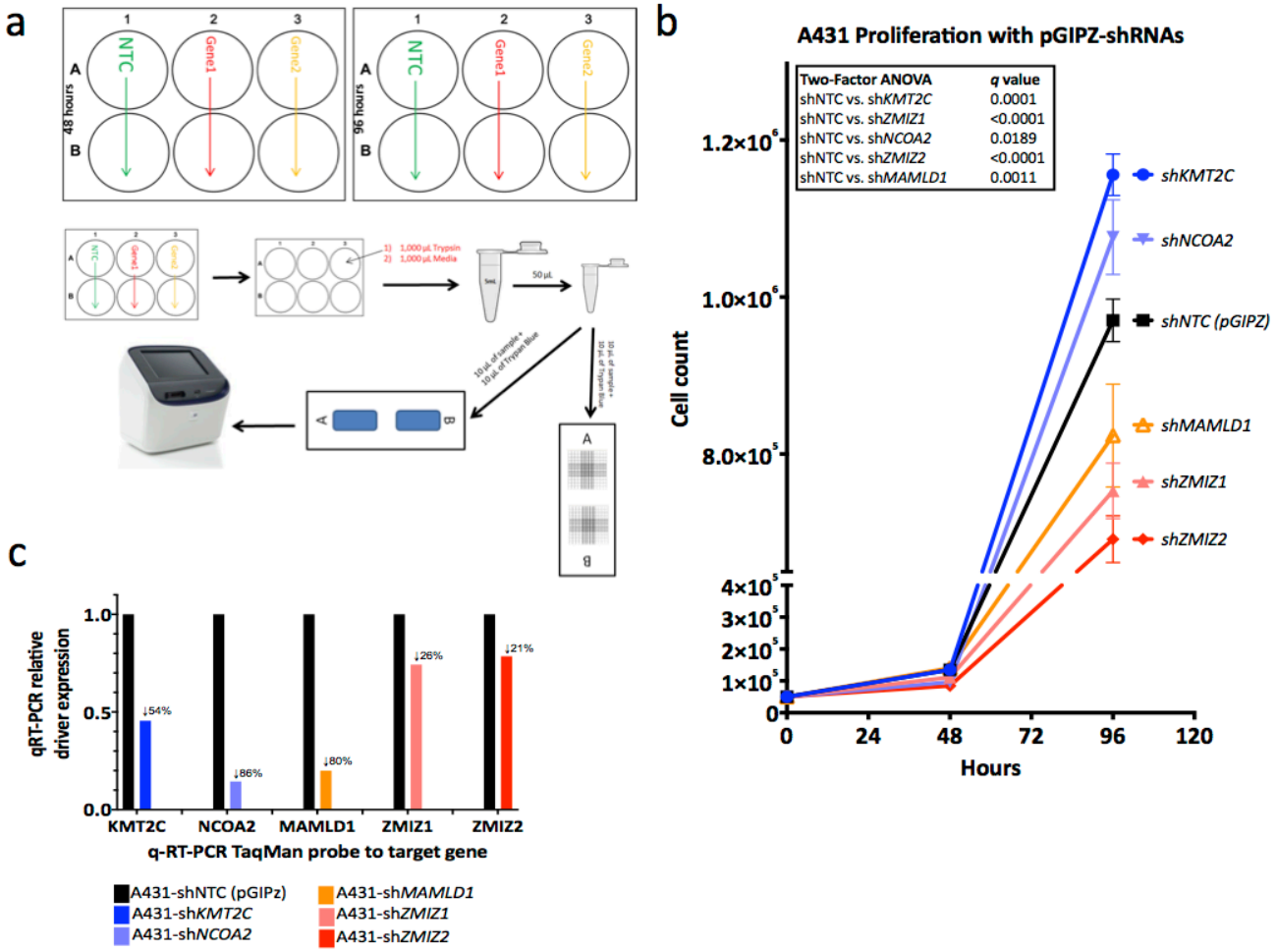
In "SCC13" and "SW900" only: *KMT2C,* and *MAMLD1*

In "SW900" only: *CUL1, EHMT1,* and *NCOR2*

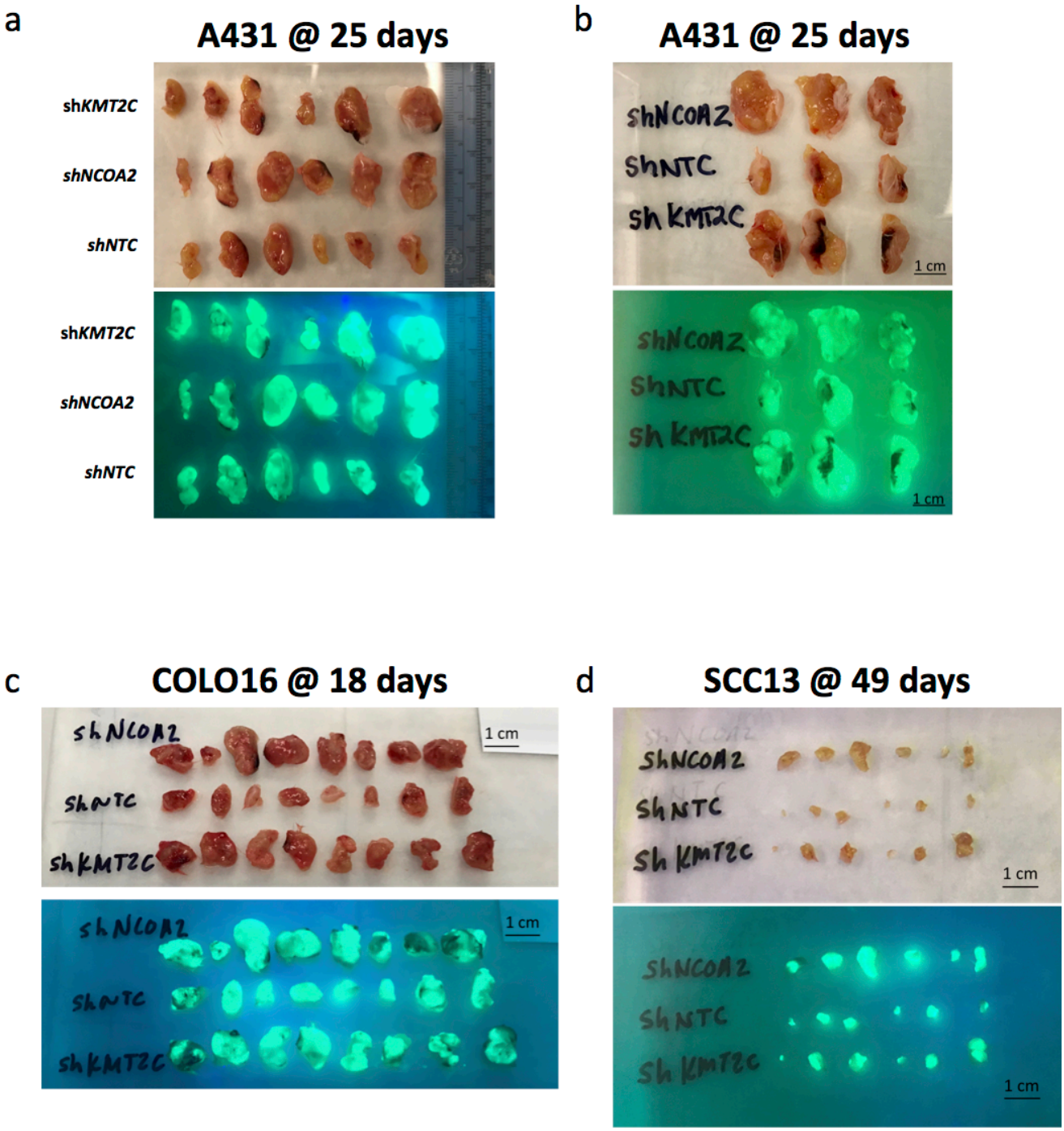
e

Driver gene	A431	SCC13	SW900
CHUK	2	1	1
CUL1	0	0	1
EHMT1	0	0	2
FBXO11	2	0	3
KMT2C	0	1	2
LARP4B	1	0	2
MAMLD1	0	1	3
NCOA2	1	1	0
NCOR2	0	0	3
NOTCH1	1	1	1
PTEN	3	1	3
ZMIZ1	1	0	1
ZMIZ2	3	nd	nd

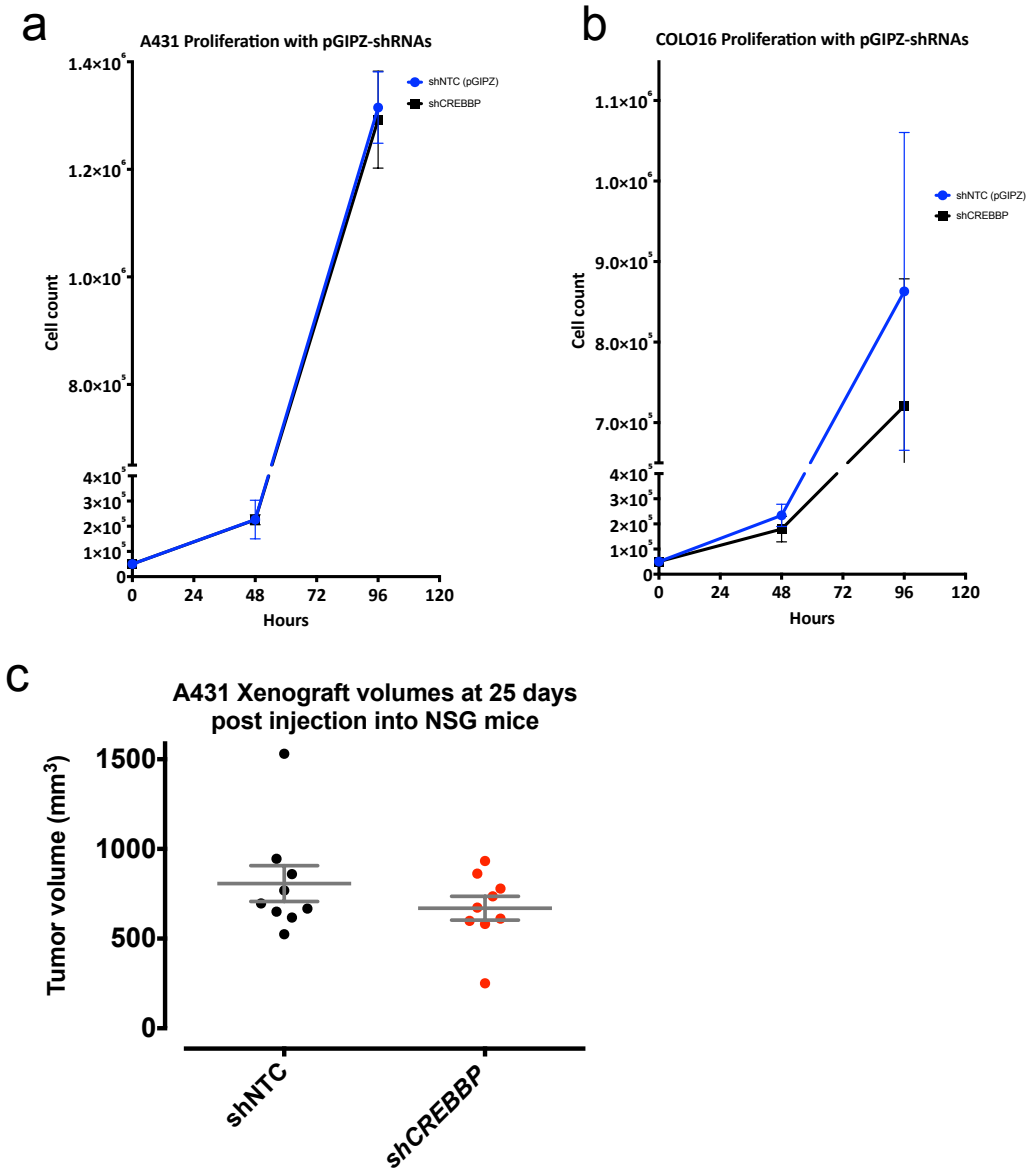
Supplementary Figure 19: siRNA-mediated driver gene silencing in SCC cell lines. Cellular proliferation assay (WST-1) using an siRNA screen with 3 non-overlapping siRNAs from 13 genes and 2 non-targeting scramble controls (41 siRNAs, **Supplementary Table 42**) was conducted in (**a**) A431 cells (cuSCC), (**b**) SCC13 cells (cuSCC), (**c**) SW900 cells (lung SCC) at 72-hour time point. (**d**) Venn diagram of overlap between SCC cell lines, (**e**) SCC cell lines with number of statistically significant siRNAs (two-factor ANOVA, FDR adjusted-*P* <0.1).



Supplementary Figure 20: Lentiviral-shRNA-mediated driver gene silencing in cuSCC cell lines. (**a**) Experimental workflow for cellular proliferation in cuSCC cell line A431. Representative proliferation (**b**) and RNA abundance (**c**) of A431 cells with pools of 3 non-overlapping shRNAs directed against two tumor suppressor and three oncogenic drivers relative to control (shNTC) cells (two-factor ANOVA, FDR adjusted-*P* <0.1).



Supplementary Figure 21: Gross photographs of cuSCC xenograft masses collected at necropsy showing robust TurboGFP expression. Xenografts of cuSCC human cell lines A431, COLO16 and SCC13 with stable lentiviral constructs expressing shRNAs directed knockdown of target drivers *KMT2C* (shKMT2C) or *NCOA2* (shNCOA2) relative to a non-targeting control (shNTC) grown in *NSG* immunodeficient mice.



886

887 **Supplementary Figure 22: *CREBBP* knockdown does not alter proliferation rate in cuSCC cell lines.** (**a**) 96-hour
888 proliferation assay (n=3, error bars SEM). (**b**) *In vivo* xenograft assay of COLO16 shNTC or shCREBBP into NSG immunodeficient
889 mice (n=9 per condition).

890 **Methods**

891 **Mice used for SB screens.** The following alleles were used to construct the SB-driven mouse model of multiple solid
892 tumor histologies: *Actb-Cre* (FVB/N-Tg(ACTB-cre)2Mrt/J) (Lewandoski et al., 1997);*Trp53*$^{flox/+}$ (FVB.129P2-
893 *Trp53*^{tm1Brn}/Nci) (Jonkers et al., 2001); *Trp53*$^{LSL\text{-}R172H/+}$ (129S4-*Trp53*^{tm2Tyj}/Nci) (Olive et al., 2004);
894 T2/Onc2(TG.12740) (TgTn(sb-T2/Onc3)12740Njen) (Dupuy et al., 2009)); and *Rosa26*-LSL SBase or SBase^{LSL};
895 (Gt(ROSA)26Sor$^{tm2(sb11)Njen}$) (Starr et al., 2009)). The resulting cohorts of mice were on mixed genetic backgrounds
896 consisting of C57BL/6J, 129, C3H and FVB. Genotyping by PCR assays with primers specific to the alleles was
897 performed. No sample size estimate was used to determine the number of mice for aging. Mice were bred and
898 maintained in accordance with approved procedures directed by the respective Institutional Animal Care and Use
899 Committees in the National Cancer Institute Frederick National Lab, A*STAR Biological Resource Centre, Houston
900 Methodist Research Institute, and Moffitt Cancer Center. Animals were co-housed for the duration the experiment,
901 except in rare cases where single mice were separated based on vet recommendation due to fighting or near tumor
902 burden end-point. Gross necropsies were performed and all masses were documented and prepared for subsequent
903 analysis. See **Supplementary Figure 1** for details of mouse crosses. Both sexes were used for experiments. No
904 randomization was performed; mice were assigned to groups based on genotype. No blinding was performed.

905 **Histological analysis.** Histological analysis of spleen was performed on 5-μm sections of formalin-fixed, paraffin-
906 embedded (FFPE) specimens stained with hematoxylin and eosin. As expected, robust nuclear staining of SB
907 transposase (SBase) was confirmed by immunohistochemistry on FFPE tissues after antigen retrieval (pH 9) and
908 endogenous peroxidase inhibition followed by overnight incubation with mouse antibody to SBase (anti-SBase; R&D
909 Systems; pH 9; 1:200 dilution). After incubation with primary antibody, chromogen detection (with HRP polymer,
910 anti-rabbit or anti-mouse, with Envision System from Dako) and hematoxylin counterstaining were performed per
911 manufacturer's instructions. Genomic DNA (gDNA) was isolated from flash frozen necropsy specimens using Qiagen
912 Gentra® Puregene® DNA isolation kit protocol for tissue.

913 **Mapping transposon insertion sites using the splink_454 method.** SB insertion reads were generated by 454 GS
914 Titanium sequencing (Roche) of pooled splinkerette PCR reactions with nested, barcoded primers was
915 performed(Mann et al., 2012; Mann et al., 2015; March et al., 2011). Pre- and post-processing of 454 reads to assign
916 sample DNA barcodes, filter out local hopping events from donor chromosomes, and map and orient the SB insertion
917 sites across the entire nuclear genome of the mouse was performed. All SB insertions from donor chromosomes
918 were filtered out prior to identification of common insertion sites using the Gaussian kernel convolution (GKC) (de
919 Ridder et al., 2006; March et al., 2011) and SB driver analysis (Newberg et al., 2018b) methods.

920 **Mapping transposon insertion sites using the SBCapSeq method.** Full details for the SBCapSeq protocol (Mann
921 et al., 2016b) optimized for sequencing from solid tumors, will be published elsewhere (Mann *et al.*, in preparation;
922 for general protocol and concept, see (Mann et al., 2016a; Mann et al., 2016b; Mann et al., 2019b)). Briefly, for
923 selective SB insertion site sequencing by liquid hybridization capture, gDNA (0.5 μg per sample) of either bulk tumor
924 specimens or single cell WGA genomes was used for library construction using the AB Library Builder™ System,
925 including random fragmentation and ligation of barcoded Ion Xpress™ sequencing adapters. Adapter-ligated

926 templates were purified by Agencourt® AMPure beads® and fragments with insert size of 200±30bp were excised,
927 purified by Agencourt® AMPure® beads, amplified by 8 cycles of adapter-ligation-mediated polymerase chain
928 reaction (aLM-PCR), and purified by Agencourt® AMPure® beads with elution in 50 µl of TE (1X Tris-EDTA
929 [Ethylenediaminetetraacetic acid], pH8). Capture hybridization of single or multiplexed up to 12 barcoded libraries
930 (60 ng per sample) was performed using custom xGen® Lockdown® Probes (IDT; full details available at
931 https://doi.org/10.35092/yhjc.11441001.v1 (Mann et al., 2019b)). All 120-mer capture and blocking
932 oligonucleotide probes were purchased from IDT as Ultramer DNA Oligos with Standard Desalting. Bar-coded single
933 and multiplex captured library fragments were further amplified by 12 cycles of LM-PCR and run using an Agilent
934 2100 Bioanalyzer or TapeStation to estimate enrichment. Bar-coded single and multiplex captured libraries were
935 quantified by Qubit® Fluorometer and quantitative Real Time-PCR (qRT-PCR) were used to dilute libraries for
936 template preparation and Ion Sphere™ Particle (ISP) loading using Ion Chef™ System and sequencing on the Ion
937 Proton™ platform with PI_{v3} semiconductor wafer chips per manufacturers recommended instructions. High-
938 throughput sequencing of up to 39 multiplex captured libraries was carried out per PI_{v3} chip to achieve at least 1.5
939 million reads per barcode. Reads containing the transposon IRDR element were processed using the SBCapSeq
940 bioinformatic workflow as described(Mann et al., 2016b).

941 **SB Driver Analysis.** BED formatted files containing SB insertions from each of the histologically verified SB-cuSCC,
942 SB-cuKA, and SB-cuSK cohort specimens (**Supplementary Tables 2,4,7,11,14,17,21**) were used to perform SB
943 Driver Analysis to identify statistically significant discovery, progression, and trunk driver genes that contain more
944 SB insertions than expected by chance and were recurrently altered in three or more tumors (Newberg et al.,
945 2018b). Discovery and progression SB Driver Analysis considered all SB insertion events; trunk SB Driver Analysis
946 considered only insertions represented by 5 or more reads. Statistically significant discovery drivers were defined
947 as adjusted *p*-values using false discovery rate (FDR) multiple testing correction (*q*-value<0.5). Statistically
948 significant progression and trunk drivers were defined as adjusted *p*-values using family-wise error rate (FWER)
949 multiple testing correction (FWER adjusted-*P*<0.05). Statistically significant drivers on donor and non-donor
950 chromosomes analyses were performed separately and combined into a single driver lists (**Supplementary Tables**
951 **2,5-6,12-13,15-16**). Due to a local-hopping phenomenon known to occur with SB, where transposition events are
952 biased to occur in cis along the donor chromosome more frequently than in trans to non-donor chromosomes in the
953 genome, insertions from the donor chromosome are typically filtered away computationally before candidate cancer
954 genes are identified. However, doing so in cohorts derived from mice with only a single SB donor allele (like our
955 cuSCC cohorts) means that genome-wide driver analysis is not possible, because all genes on the donor chromosome
956 are censored and are not reported even though they may contribute to the overall tumor burden. Since we obtained
957 quantitative SBCapSeq datasets in this study, we evaluated whether all genes on the donor chromosome should be
958 censored or if perhaps, as observed in *PiggyBac* transposon genomes, only a relatively small portion of the donor
959 chromosome locus, occurring in close proximity to the donor concatemer insertion site, demonstrate significant bias
960 compared with the frequency of SB insertions into non-donor chromosomes. In the cuSCC cohorts sequenced with
961 SBCapSeq, we found that donor chromosome SB insertion site frequencies are significantly higher between
962 chr9:79,000,000 and chr9:95,000,000 and drop to the genome wide average for non-donor chromosomes over the
963 rest of chromosome 9 (**Supplementary Figure 2b**), suggesting that filtering all SB donor chromosome insertion

964 events may not be warranted. To confirm and extend this observation, we used all tumors within the SBCDDB with
965 chromosome 9 donor alleles and identified that the same region between chr9:79,000,000 and chr9:95,000,000 had
966 a higher frequency of SB insertions than the rest of chromosome 9 loci, which again matched non-donor
967 chromosome insertion frequencies (**Supplementary Figure 2a**). Thus, we ran SB Driver Analysis on tumors derived
968 from mouse cohorts harboring an SB T2/Onc3 TG.12740 donor allele chromosome 9 (cuSCC83_454, cuSCC60_SBC,
969 and cuKA11_SBC), by excluding the 81 genes that map between *Filip1* at chr9:79663368-79825689 and *Tfdp2* at
970 96096693-96224065 from the driver gene output files.

971 **Microarray gene expression analysis**. Gene expression profiling of histologically confirmed SB-cuSCC masses
972 selected by presence or absence of *Zmiz1* insertion by 454Splink sequencing was performed using Affymetrix
973 microarrays, as described previously (Mann et al., 2016b). Briefly, 100ng of total RNA for each sample was extracted
974 using a NORGEN Biotek Animal Tissue RNA Purification kit (Cat #25700) followed by labeled with an Affymetrix 3'
975 IVT Express kit (Cat # 901229) using the manufacturer's instructions. Labeled samples were hybridized to
976 Affymetrix GeneChip Mouse Genome 430 2.0 Arrays, and scanned at the University of Otago Genomics &
977 Bioinformatics Facility. Raw data processing used R (version 2.15) (Team, 2013) with the "rma" function of the
978 "affy" package (Gautier et al., 2004), including quantile normalization but no background correction. Quality
979 assessment of the microarray data was performed using R (R_Core_Team, 2008; 2013) or the "affyQCReport"
980 package (Parman et al., 2005). Data analysis and differential gene expression analysis were performed using Python,
981 R (R_Core_Team, 2008; 2013) and visualized using the R Shiny package in RStudio.

982 **Transcriptome Sequencing.** Total RNA was isolated from flash frozen necropsy specimens using *mir*Vana™ miRNA
983 Isolation Kit (Ambion® by Life Technologies, AM1560) from a shaved portion of each pathology verified cuSCC mass
984 from 7 tumors for which sufficient tissues were available and the presence of a high read depth *Zmiz1* or *Zmiz2*
985 trunk driver insertion was identified by SBCapSeq. Whole transcriptome RNA-Seq (wtRNA-Seq) libraries were
986 prepared from total RNA (5 µg per sample) plus ERCC RNA Spike-In (Ambion® by Life Technologies, 4456739)
987 followed by selective ribosomal RNA (rRNA) depletion using the RiboMinus™ Eukaryote System v2 (Ambion® by
988 Life Technologies, A15026). rRNA-depleted total RNA (500 ng per sample) was used for whole transcriptome library
989 construction according to the Ion Total RNA-Seq Kit for the AB Library Builder™ System (Life Technologies,
990 4482416) protocol for barcoded libraries, Ion Sphere™ Particle (ISP) loading using Ion Chef™ System and
991 sequencing on the Ion Proton™ platform with PI_{v3} semiconductor wafer chips per manufacturers recommended
992 instructions. Up to 4 RNA-Seq libraries were multiplex sequenced on a PI_{v3} chip twice to achieve >3 GB (~40 million
993 reads) per specimen. The Life Technologies Torrent Suite software was used to perform checking of the raw
994 sequence data before the generation of sequencing read files in FASTQ format and minimally processed mRNA-Seq
995 and wtRNA-Seq reads were processed using the Bowtie2 (Langmead and Salzberg, 2012) and Tophat (Trapnell et
996 al., 2009) algorithms to align reads to a custom version of the mouse mm9+pT2/Onc genome, as described (Mann et
997 al., 2016b). The detection of SB fusion events, defining novel SB fusion transcripts, and measuring transcript
998 abundance in SB cuSCC tumor specimens from RNA-Seq data were performed as described (Mann et al., 2016b).

999 **Biological Pathway and Process Enrichment Analysis.** Enrichr (Chen et al., 2013; Kuleshov et al., 2016), an online
1000 analysis tool for human and mouse gene-set enrichment, was used to identify specific signaling pathways and
1001 processes enrichment using the various cohort candidate cancer driver genes and/or their human orthologs
1002 (**Supplementary Tables 28,32-33**). The online STRING network enrichment analysis tool (Szklarczyk et al., 2019)
1003 was also used to determine functional connections between trunk drivers from the cuSCC and cuKA cohorts.

1004 **Human cancer cell lines.** The cutaneous squamous cell carcinoma A431(CRL–1555™) and grade IV lung squamous
1005 cell carcinoma SW900 (HTB–59™) cell lines, were purchased from American Type Culture Collection (ATCC®) and
1006 grown according to the manufacturer's suggested conditions in complete medium (1× DMEM (ATCC, 30–2002) for
1007 A431cells or 1× ATCC-formulated Leibovitz's L–15 Medium, Catalog No. 30–2008 medium (ATCC, 30–2001) for
1008 SW900 cells) supplemented with 10% FBS and 1× penicillin-streptomycin grown at 37 °C in 5% CO2. The cutaneous
1009 squamous cell carcinoma SCC13 cell line (Rheinwald and Beckett, 1981; Rheinwald and Green, 1977), were obtained
1010 from Harvard Skin Disease Research Center (HSDRC) and grown using the human keratinocyte culture methods
1011 provided by the HSDRC. The cutaneous squamous cell carcinoma COLO16 cell line (Vin et al., 2013) were a gift from
1012 Dr. K. Y. Tsai and grown in DMEM/Ham's F12 50/50 media supplemented with cholera toxin (Sigma, C0852), insulin
1013 (Sigma, I5500-50MG), epidermal growth factor (Serotec/Bio-Rad, EGF-1), hydrocortisone (Sigma, H4881-1G),
1014 liothyronine (Sigma, T6397-100MG) and apo-transferrin (Sigma, T2252-100MG). All cell lines were verified human
1015 origin and free from pathogens and mycoplasma. Mycoplasma infection monitoring was performed using MycoAlert
1016 Detection Kit (Lonza) and only mycoplasma-free cultures were used.

1017 **siRNA screening**. Pre-designed siRNA oligonucleotides were purchased from Thermo Fisher Scientific and are
1018 listed in **Supplementary Table 41**. For each gene, three non-overlapping siRNA oligos were transfected using
1019 Lipofectamine RNAiMAX transfection reagent (Thermo Fisher Scientific) according to the manufacturer's
1020 instructions. Briefly, 4×10^3 cells were seeded into 96-well plates and cultured for 24 hours at 37 ºC in complete
1021 culture medium. At 24 hours, single siRNAs were transfected and incubated at 37 ºC. Cellular proliferation was
1022 measured with a WST–1 cell proliferation assay system (Takara). WST–1 reagent was added to each well and
1023 incubated for 2 hours at 37 ºC. The colorimetric absorbance at 450 nm and 650 nm (reference) was measured with
1024 an Infinite 200 pro multimode reader (TECAN).

1025 **Generation of stable shRNA-expressing cell lines.** High titer lentiviral particles for pGIPZ-shRNA constructs
1026 targeting each of the 12 cutaneous candidate driver genes, and one non-targeting control, were purchased from
1027 Thermo Scientific Open Biosystems (**Supplementary Table 43**). Cells were plated at a density of 5×10^4 cells per
1028 well in a 24-well plate in complete media 24 hours prior to infection. The following day, cells were plated with
1029 serum-free culture medium containing 8 µg/mL polybrene (Millipore), and transduced as pools of three
1030 independent shRNAs to different exons of each target gene, or individually for the control, at multiplicity of infection
1031 of 6. Puromycin selection was added the following day at concentrations of 1.0 to 3.0 µg/mL puromycin (Thermo
1032 Fisher Scientific) in complete media, and replaced every three until stable lines were achieved.

1033 **Proliferation assay**. To assess cuSCC progression, cuSCC cell lines were seeded in 6 well plates at a density of $5 \times$
1034 10^4 cells per well in complete media. Cell numbers for each condition were counted after 48- and 96-hour post-

1035 seeding. To assess transformation by proliferation, HaCaT cells were seeded at a density of 5×10^3 cells per well and
1036 cell counts performed after seven days.

1037 **Soft agar assay**. To assess anchorage-independent growth, HaCaT cells were seeded at a density of 1×10^4 cells per
1038 well in a six well plate in 0.3% agar, over a layer 0.6% agar. Complete media was added to each well the following
1039 day, and replaced every two days for four weeks. At end point, cells were fixed in 20% methanol and 0.0025%
1040 crystal violet. Washes were done with milliQ water until low background was achieved, and plates were scanned,
1041 and colonies counted for each condition. $KRAS^{G12D}$ expression vector (Origene cat# RC400104) was a kind gift from
1042 Dr Karen Mann. The empty vector (pCMV6-Entry) was generated by removing the $KRAS^{G12D}$ cDNA using the *SgfI* and
1043 *MluI* restriction enzymes, gel purified, and ligated to obtain an empty plasmid. HaCaT cells were transfected with the
1044 expression plasmids and selected with 800ug/mL G418 to obtain stable cell lines. For the soft agar assay was
1045 performed as described above.

1046 **qRT-PCR.** Total RNA was purified and DNase treated using the RNeasy Mini Kit (Qiagen). Synthesis of cDNA was
1047 performed using SuperScript VILO Master Mix (Life Technologies). Quantitative PCR analysis was performed on the
1048 QuantStudio 12K Flex System (Life Technologies) or 7900HT Sequence Detection System (Applied Biosystem). All
1049 signals were normalized to the levels of GAPDH TaqMan probes. TaqMan probes were obtained from Life
1050 Technologies (**Supplementary Table 41**).

1051 **Xenografts.** One million cells were prepared for injection into the left flank of randomized selected male and female
1052 immunodeficient NSG (NOD.Cg-Prkdcscid;Il2rgtm1Wjl/SzJ; JAX, 005557; 6–15 weeks old) mice. Allocation to study
1053 groups was random. Xenograft measurements were taken twice weekly using digital calipers while mice were
1054 conscious but restrained by an experimenter familiar with collecting caliper measurements of xenografts and
1055 blinded to the experimental group designations. GFP fluorescence was visualized at the time of caliper
1056 measurement. Ellipsoid tumor volumes were calculated as volume (mm3) = 0.52(length (mm) × width2 (mm2)),
1057 where the two longest axes, length and width, were the major and minor diameter measurements, respectively;
1058 width2 represents an assumption that the xenograft depth was equivalent to the diameter of the minor axis.
1059 Statistical significance of tumor volumes at defined time points was determined by either one-way t test for all
1060 cohorts relative to shNTC or by two-way, repeated-measures ANOVA with Bonferroni correction for multiple
1061 comparisons, as indicated.

1062 **Western blot.** Whole cell or tissue lysates were prepared in RIPA buffer supplemented with protease and
1063 phosphatase inhibitors, and samples sonicated to achieve optimum lysis. Protein concentrations were quantitated
1064 using the BCA assay, and 20 µg of lysates were loaded into 8% SDS-polyacrylamide gels. Gels were resolved at 80V
1065 for 120 mins, transferred onto nitrocellulose membrane for 70 mins at 0.35A with transfer buffer containing 20%
1066 methanol, and blocked for one hour at room temperature with 5% w/v BSA in 0.1% TBST. Membranes were
1067 incubated with primary antibodies in 5% w/v BSA in 0.1% TBST. Antibodies used in this study are: Total ERK
1068 (1:1000, Cell Signaling #9102), phospho ERK (1:1000, Cell Signaling #4695), GAPDH (1:5000, Santa Cruz sc-69778),
1069 anti-mouse and anti-goat secondary antibodies (1:10000, LiCOR IRDye 680/800 Cat. 925-68070/926-32211). Blots
1070 were imaged using the LiCOR Odyssey Fc system.

1071 **Software.** Unless otherwise noted, bioinformatic analysis pipelines, report generation, and figure visualization
1072 performed using bash, R (version 3.5.2), RStudio (Version 1.1.463), Python scripts, and GraphPad Prism 8 software
1073 (Version 8.1.1). Hierarchical clustering was performed in Python 2.7.10 with the scipy 0.13.0b1 toolbox using the
1074 Hamming distance metric with Ward's linkage method. The R packages ggplot2 (Wickham, 2009) and Gviz (Henare
1075 et al., 2012) were used to generate the graphics in various figure panels.

1076 **URLs.** TgTn(sb-T2/Onc3)12740Njen; Gt(ROSA)26Sortm2(sb11)Njen double mutant mice,
1077 https://ncifrederick.cancer.gov/Lasp/MouseRepository/MouseModels/StrainDetails.aspx?StrainNum=01XGB&g=R
1078 OSA26; FVB/N-Tg(ACTB-cre)2Mrt/J mice, http://jaxmice.jax.org/strain/003376.html; Cancer Gene Census,
1079 http://cancer.sanger.ac.uk/census/; Mouse Genome Informatics (MGI)
1080 database, ftp://ftp.informatics.jax.org/pub/reports/index.html; Enrichr, http://amp.pharm.mssm.edu/Enrichr/,
1081 STRING, https://string-db.org/, The Cancer Genome Atlas (TCGA), http://cancergenome.nih.gov/; FVB.129P2-
1082 Trp53tm1Brn/Nci mice,
1083 https://ncifrederick.cancer.gov/Lasp/MouseRepository/MouseModels/StrainDetails.aspx?StrainNum=01XC2&g=Tr
1084 p53; 129S4-Trp53tm2Tyj/Nci mice,
1085 https://ncifrederick.cancer.gov/Lasp/MouseRepository/MouseModels/StrainDetails.aspx?StrainNum=01XM2&g=T
1086 rp53.

1087 **Accession Codes.** NCBI BioProject accessions: PRJNA580460 for whole-transcriptome RNA-Seq data; PRJNA580462
1088 for microarray expression data.

1089 **Supplementary Tables.** Supplementary Table datasets may be downloaded from NIH.figshare.com at
1090 (http://dx.doi.org/10.35092/yhjc.11441130) (Mann et al., 2019a). Brief descriptions of each table are provided for
1091 reference.

1092 **Supplementary Table 1**: Tumor incidence and subgroup classifications by cohort

1093 **Supplementary Table 2**: Sequencing projects summary by cohort

1094 **Supplementary Table 3**: Specimen metafile data for projects sequenced using Splink_PCR protocol with
1095 454 GS-FLX Titanium pyrosequencer

1096 **Supplementary Table 4**: BED file of SB insertions for cuSCC83_454

1097 **Supplementary Table 5**: Discovery and progression SB Driver Analysis for cuSCC83_454

1098 **Supplementary Table 6**: Trunk SB Driver Analysis for cuSCC83_454

1099 **Supplementary Table 7**: BED file of SB insertions for cuKA51_454

1100 **Supplementary Table 8**: Discovery and progression SB Driver Analysis for cuKA51_454

1101 **Supplementary Table 9**: Trunk SB Driver Analysis for cuKA51_454

1102 **Supplementary Table 10**: Specimen metafile data for projects sequenced using SBCapSeq protocol with
1103 Ion Torrent Proton sequencer

1104 **Supplementary Table 11**: BED file of SB insertions for cuSCC60_SBC.

1105 **Supplementary Table 12**: Discovery and progression SB Driver Analysis for cuSCC60_SBC.

1106 **Supplementary Table 13**: Trunk SB Driver Analysis for cuSCC60_SBC.

1107 **Supplementary Table 14**: BED file of SB insertions for cuKA11_SBC

1108 **Supplementary Table 15**: Discovery and progression SB Driver Analysis for cuKA11_SBC

1109 **Supplementary Table 16**: Trunk SB Driver Analysis for cuKA11_SBC

1110 **Supplementary Table 17**: BED file of SB insertions for cuSK32_SBC

1111 **Supplementary Table 18**: Discovery and progression SB Driver Analysis for cuSK32_SBC

1112 **Supplementary Table 19**: BED file of SB insertions for 4 cuSCC genomes selected for multi-region
1113 resequencing because they had intermixing of cuSCC and cuKA histologies

1114 **Supplementary Table 20**: SBCapSeq read depth and analysis for 4 cuSCC genomes selected for multi-
1115 region resequencing because they had intermixing of cuSCC and cuKA histologies

1116 **Supplementary Table 21**: BED file of SB insertions for cuSK10_SBC

1117 **Supplementary Table 22**: Discovery and progression SB Driver Analysis for cuSK10_SBC

1118 **Supplementary Table 23**: Contingency Tables comparing Cancer Gene Census (GRCh38 COSMICv86)
1119 driver enrichment with cuSCC cohorts

1120 **Supplementary Table 24**: Venn diagram for overlap of Cancer Gene Census (GRCh38 COSMICv86) driver
1121 enrichment with all SB|cuSCC Trunk Driver and SB|cuKA Trunk Driver lists

1122 **Supplementary Table 25**: Venn diagram for overlap of Cancer Gene Census (GRCh38 COSMICv86) driver
1123 enrichment with all SB|cuSCC Progression Driver and SB|cuKA Progression Driver lists

1124 **Supplementary Table 26**: Venn diagram for overlap of Cancer Gene Census (GRCh38 COSMICv86) driver
1125 enrichment with cuSCC60_SBC Trunk Driver, cuSCC83_454 Trunk Driver, and cuKA51_454 Trunk Driver
1126 lists

1127 **Supplementary Table 27**: Venn diagram for overlap of Cancer Gene Census (GRCh38 COSMICv86) driver
1128 enrichment with cuSCC60_SBC Progression Driver, cuSCC83_454 Progression Driver, and cuKA51_454
1129 Progression Driver lists

1130 **Supplementary Table 28**: Enrichr gene set pathway enrichment analysis of cuSCC drivers

1131 **Supplementary Table 29**: Normalized microarray values per gene from RNA isolated from cuSCC genomes
1132 with and without Zmiz1 insertions

1133 **Supplementary Table 30**: Normalized microarray values per probe from RNA isolated from cuSCC
1134 genomes with and without Zmiz1 insertions

1135 **Supplementary Table 31**: Top 96 genes with differential expression analysis from microarray data from
1136 RNA isolated from cuSCC genomes with and without Zmiz1 insertions

1137 **Supplementary Table 32**: Enrichr gene set pathway enrichment analysis of the top 96 differentially
1138 expressed genes from microarray data from RNA isolated from cuSCC genomes with and without Zmiz1
1139 insertions

1140 **Supplementary Table 33**: Enrichr gene set gene ongology term enrichment analysis of the top 96
1141 differentially expressed genes from microarray data from RNA isolated from cuSCC genomes with and
1142 without Zmiz1 insertions

1143 **Supplementary Table 34**: All 289 genes with differential expression analysis from microarray data from
1144 RNA isolated from cuSCC genomes with and without Zmiz1 insertions with FDR adjusted-$P<0.0001$

1145 **Supplementary Table 35**: Summary of 7 cuSCC transcriptomes selected for whole transcriptome RNA-seq
1146 analysis

1147 **Supplementary Table 36**: BED file of SBfusion insertions in 7 cuSCC genomes by whole transcriptome
1148 RNA-seq analysis

1149 **Supplementary Table 37**: Venn diagram for overlap of genes with SBfusion reads detected by whole
1150 transcriptome RNA-seq analysis and cuSCC60_SBC discovery drivers

1151 **Supplementary Table 38**: Venn diagram for overlap of genes with SBfusion reads detected by whole
1152 transcriptome RNA-seq analysis and all cuSCC drivers

1153 **Supplementary Table 39**: Transcripts per million (TPM) normalized whole transcriptome RNA-seq values
1154 per gene from RNA isolated from cuSCC genomes with and without Zmiz1 insertions

1155 **Supplementary Table 40**: Fragments Per Kilobase of Transcripts per Million (FPKM) normalized whole
1156 transcriptome RNA-seq values per gene transcript from RNA isolated from cuSCC genomes with and
1157 without Zmiz1 insertions

1158 **Supplementary Table 41**: siRNA and TaqMan probes used in this study

1159 **Supplementary Table 42**: siRNA primary screen results

1160 **Supplementary Table 43**: Lentiviral vectors containing shRNAs used in this study

1161 **Acknowledgments**

1162 M.B.M. and K.M.M. are supported by start-up funds provided by the H. Lee Moffitt Cancer Center and Research
1163 Institute, a National Cancer Institute-designated Comprehensive Cancer Center through the Cancer Center Support
1164 Grant (NCI P30-CA76292; PI: J. Cleveland). M.B.M. was also supported by Career Enhancement Program Grant
1165 Award from the NCI Moffitt Skin Cancer SPORE (5P50CA168536-05; PI: V. Sondak). N.G.C. and N.A.J. were supported
1166 by the Biomedical Research Council, Agency for Science, Technology, and Research, Singapore and as CPRIT Scholars
1167 in Cancer Research from the Cancer Prevention Research Institute of Texas. The authors wish to thank the
1168 Copeland–Jenkins labs in Singapore and Houston for helpful discussions; we thank Tyler Jacks (MIT) for P53 LSL
1169 R172H mice; Anton Berns (NKI) for P53 floxed mice; Gail Martin (UCSF) for Actin-beta-Cre mice; James G. Rheinwald
1170 and the Harvard Skin Disease Research Center for the SCC13 cell line; David Adams, Theo Whipp, Richard Rance,
1171 Alistair Rust, and the Wellcome Trust Sanger Institute sequencing and informatics teams for 454 sequencing; the
1172 Elsa Flores Lab for use of the LiCOR imaging system; Sean Yoder (Moffitt Cancer Center) and Dr. Chaomei Zhang
1173 (Moffitt Cancer Center) for providing sequencing services; Pearlyn Cheok, Nicole Lim, Dorothy Chen and Cherylin
1174 Wee for assistance with tumor monitoring and animal husbandry at IMCB (Singapore, Republic of Singapore),
1175 Hubert Lee and Erik Freiter for assistance with animal husbandry at HMRI (Houston, TX, USA), and Crystal Reed and
1176 Bethanie Gore for assistance with animal husbandry at MCC/USF (Tampa, FL, USA). Necropsy and histology were
1177 performed by the Advanced Molecular Pathology Laboratory, Institute of Molecular and Cell Biology (A*STAR),
1178 Singapore. Necropsy and histology were also performed by the Tissue Core Facility and Ion Torrent sequencing was
1179 performed by the Molecular Genomics Core at the H. Lee Moffitt Cancer Center & Research Institute, a
1180 comprehensive cancer center designated by the National Cancer Institute, and funded in part by Moffitt's Cancer
1181 Center Support Grant (P30-CA076292; PI: J. Cleveland).

1182 **References**

1183 Aiderus, A., Contreras-Sandoval, A.M., Meshey, A.L., Newberg, J.Y., Ward, J.M., Swing, D., Copeland, N.G., Jenkins, N.A.,
1184 Mann, K.M., and Mann, M.B. (2020). Promoterless Transposon Mutagenesis Drives Solid Cancers via Tumor
1185 Suppressor Inactivation. bioRxiv, 2020.2008.2017.254565.

1186 Andricovich, J., Perkail, S., Kai, Y., Casasanta, N., Peng, W., and Tzatsos, A. (2018). Loss of KDM6A Activates Super-
1187 Enhancers to Induce Gender-Specific Squamous-like Pancreatic Cancer and Confers Sensitivity to BET
1188 Inhibitors. Cancer Cell *33*, 512-526 e518.

1189 Beliakoff, J., and Sun, Z. (2006). Zimp7 and Zimp10, two novel PIAS-like proteins, function as androgen receptor
1190 coregulators. Nucl Recept Signal *4*, e017.

1191 Boukamp, P., Petrussevska, R.T., Breitkreutz, D., Hornung, J., Markham, A., and Fusenig, N.E. (1988). Normal
1192 keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. J Cell Biol *106*, 761-771.

1193 Cancer Genome Atlas, N. (2015). Comprehensive genomic characterization of head and neck squamous cell
1194 carcinomas. Nature *517*, 576-582.

1195 Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr:
1196 interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

1197 Chitsazzadeh, V., Coarfa, C., Drummond, J.A., Nguyen, T., Joseph, A., Chilukuri, S., Charpiot, E., Adelmann, C.H., Ching,
1198 G., Nguyen, T.N.*, et al.* (2016). Cross-species identification of genomic drivers of squamous cell carcinoma
1199 development across preneoplastic intermediates. Nat Commun *7*, 12601.

1200 Cho, S.J., Yoon, C., Lee, J.H., Chang, K.K., Lin, J.X., Kim, Y.H., Kook, M.C., Aksoy, B.A., Park, D.J., Ashktorab, H.*, et al.*
1201 (2018). KMT2C Mutations in Diffuse-Type Gastric Adenocarcinoma Promote Epithelial-to-Mesenchymal
1202 Transition. Clin Cancer Res *24*, 6556-6569.

1203 Collier, L.S., and Largaespada, D.A. (2007). Transposons for cancer gene discovery: Sleeping Beauty and beyond.
1204 Genome Biol *8 Suppl 1*, S15.

1205 Copeland, N.G., and Jenkins, N.A. (2010). Harnessing transposons for cancer gene discovery. Nat Rev Cancer *10*, 696-
1206 706.

1207 de la Rosa, J., Weber, J., Friedrich, M.J., Li, Y., Rad, L., Ponstingl, H., Liang, Q., de Quiros, S.B., Noorani, I., Metzakopian,
1208 E.*, et al.* (2017). A single-copy Sleeping Beauty transposon mutagenesis screen identifies new PTEN-cooperating
1209 tumor suppressor genes. Nat Genet *49*, 730-741.

1210 de Ridder, J., Uren, A., Kool, J., Reinders, M., and Wessels, L. (2006). Detecting statistically significant common
1211 insertion sites in retroviral insertional mutagenesis screens. PLoS Comput Biol *2*, e166.

1212 Dhar, S.S., Zhao, D., Lin, T., Gu, B., Pal, K., Wu, S.J., Alam, H., Lv, J., Yun, K., Gopalakrishnan, V.*, et al.* (2018). MLL4 Is
1213 Required to Maintain Broad H3K4me3 Peaks and Super-Enhancers at Tumor Suppressor Genes. Mol Cell *70*,
1214 825-841 e826.

1215 Doyle, B., Morton, J.P., Delaney, D.W., Ridgway, R.A., Wilkins, J.A., and Sansom, O.J. (2010). p53 mutation and loss
1216 have different effects on tumourigenesis in a novel mouse model of pleomorphic rhabdomyosarcoma. J Pathol
1217 *222*, 129-137.

1218 Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G., and Jenkins, N.A. (2005). Mammalian mutagenesis using a
1219 highly mobile somatic Sleeping Beauty transposon system. Nature *436*, 221-226.

1220 Dupuy, A.J., Rogers, L.M., Kim, J., Nannapaneni, K., Starr, T.K., Liu, P., Largaespada, D.A., Scheetz, T.E., Jenkins, N.A.,
1221 and Copeland, N.G. (2009). A modified sleeping beauty transposon system that can be used to model a wide
1222 variety of human cancers in mice. Cancer Res *69*, 8150-8156.

1223 Friedel, R.H., Friedel, C.C., Bonfert, T., Shi, R., Rad, R., and Soriano, P. (2013). Clonal expansion analysis of transposon
1224 insertions by high-throughput sequencing identifies candidate cancer genes in a PiggyBac mutagenesis screen.
1225 PLoS One *8*, e72338.

1226 Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A
1227 census of human cancer genes. Nat Rev Cancer *4*, 177-183.

1228 Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe
1229 level. Bioinformatics *20*, 307-315.

1230 Genovesi, L.A., Ng, C.G., Davis, M.J., Remke, M., Taylor, M.D., Adams, D.J., Rust, A.G., Ward, J.M., Ban, K.H., Jenkins, N.A.,
1231 *et al.* (2013). Sleeping Beauty mutagenesis in a mouse medulloblastoma model defines networks that
1232 discriminate between human molecular subgroups. Proc Natl Acad Sci U S A *110*, E4325-4334.

1233 Gozdecka, M., Meduri, E., Mazan, M., Tzelepis, K., Dudek, M., Knights, A.J., Pardo, M., Yu, L., Choudhary, J.S.,
1234 Metzakopian, E., *et al.* (2018). UTX-mediated enhancer and chromatin remodeling suppresses myeloid
1235 leukemogenesis through noncatalytic inverse regulation of ETS and GATA programs. Nat Genet *50*, 883-894.

1236 Hameetman, L., Commandeur, S., Bavinck, J.N., Wisgerhof, H.C., de Gruijl, F.R., Willemze, R., Mullenders, L., Tensen,
1237 C.P., and Vrieling, H. (2013). Molecular profiling of cutaneous squamous cell carcinomas and actinic keratoses
1238 from organ transplant recipients. BMC Cancer *13*, 58.

1239 Hanel, W., Marchenko, N., Xu, S., Yu, S.X., Weng, W., and Moll, U. (2013). Two hot spot mutant p53 mouse models
1240 display differential gain of function in tumorigenesis. Cell Death Differ *20*, 898-909.

1241 Henare, K., Wang, L., Wang, L.C., Thomsen, L., Tijono, S., Chen, C.J., Winkler, S., Dunbar, P.R., Print, C., and Ching, L.M.
1242 (2012). Dissection of stromal and cancer cell-derived signals in melanoma xenografts before and after treatment
1243 with DMXAA. Br J Cancer *106*, 1134-1147.

1244 Huang, C.Y., Beliakoff, J., Li, X., Lee, J., Li, X., Sharma, M., Lim, B., and Sun, Z. (2005). hZimp7, a novel PIAS-like protein,
1245 enhances androgen receptor-mediated transcription and interacts with SWI/SNF-like BAF complexes. Mol
1246 Endocrinol *19*, 2915-2929.

1247 Hudson, L.G., Gale, J.M., Padilla, R.S., Pickett, G., Alexander, B.E., Wang, J., and Kusewitt, D.F. (2010). Microarray
1248 analysis of cutaneous squamous cell carcinomas reveals enhanced expression of epidermal differentiation
1249 complex genes. Mol Carcinog *49*, 619-629.

1250 Huguet, E.L., McMahon, J.A., McMahon, A.P., Bicknell, R., and Harris, A.L. (1994). Differential expression of human
1251 Wnt genes 2, 3, 4, and 7B in human breast cell lines and normal and disease states of human breast tissue.
1252 Cancer Res *54*, 2615-2621.

1253 Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvak, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like
1254 transposon from fish, and its transposition in human cells. Cell *91*, 501-510.

1255 Jia, D., Augert, A., Kim, D.W., Eastwood, E., Wu, N., Ibrahim, A.H., Kim, K.B., Dunn, C.T., Pillai, S.P.S., Gazdar, A.F., *et al.*
1256 (2018). Crebbp Loss Drives Small Cell Lung Cancer and Increases Sensitivity to HDAC Inhibition. Cancer Discov
1257 *8*, 1422-1437.

1258 Jonkers, J., Meuwissen, R., van der Gulden, H., Peterse, H., van der Valk, M., and Berns, A. (2001). Synergistic tumor
1259 suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. Nat Genet *29*, 418-425.

1260 Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and
1261 bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat
1262 Genet *45*, 592-601.

1263 Kahyo, T., Nishida, T., and Yasuda, H. (2001). Involvement of PIAS1 in the sumoylation of tumor suppressor p53. Mol
1264 Cell *8*, 713-718.

1265 Knatko, E.V., Praslicka, B., Higgins, M., Evans, A., Purdie, K.J., Harwood, C.A., Proby, C.M., Ooi, A., and Dinkova-Kostova,
1266 A.T. (2017). Whole-Exome Sequencing Validates a Preclinical Mouse Model for the Prevention and Treatment of
1267 Cutaneous Squamous Cell Carcinoma. Cancer Prev Res (Phila) *10*, 67-75.

1268 Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M.,
1269 Lachmann, A., *et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.
1270 Nucleic Acids Res *44*, W90-97.

1271 Kypriotou, M., Huber, M., and Hohl, D. (2012). The human epidermal differentiation complex: cornified envelope
1272 precursors, S100 proteins and the 'fused genes' family. Exp Dermatol *21*, 643-649.

1273 Lambert, S.R., Mladkova, N., Gulati, A., Hamoudi, R., Purdie, K., Cerio, R., Leigh, I., Proby, C., and Harwood, C.A. (2014).
1274 Key differences identified between actinic keratosis and cutaneous squamous cell carcinoma by transcriptome
1275 profiling. Br J Cancer *110*, 520-529.

1276 Lang, G.A., Iwakuma, T., Suh, Y.A., Liu, G., Rao, V.A., Parant, J.M., Valentin-Vega, Y.A., Terzian, T., Caldwell, L.C., Strong,
1277 L.C., *et al.* (2004). Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. Cell
1278 *119*, 861-872.

1279 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

1280 Lee, C.S., Bhaduri, A., Mah, A., Johnson, W.L., Ungewickell, A., Aros, C.J., Nguyen, C.B., Rios, E.J., Siprashvili, Z., Straight,
1281 A., *et al.* (2014). Recurrent point mutations in the kinetochore gene KNSTRN in cutaneous squamous cell
1282 carcinoma. Nat Genet *46*, 1060-1062.

1283 Lee, S.H., Zhu, C., Peng, Y., Johnson, D.T., Lehmann, L., and Sun, Z. (2013). Identification of a novel role of ZMIZ2
1284 protein in regulating the activity of the Wnt/beta-catenin signaling pathway. J Biol Chem *288*, 35913-35924.

1285 Lewandoski, M., Meyers, E.N., and Martin, G.R. (1997). Analysis of Fgf8 gene function in vertebrate development.
1286 Cold Spring Harb Symp Quant Biol *62*, 159-168.

1287 Liu, B., Xia, X., Zhu, F., Park, E., Carbajal, S., Kiguchi, K., DiGiovanni, J., Fischer, S.M., and Hu, Y. (2008). IKKalpha is
1288 required to maintain skin homeostasis and prevent skin cancer. Cancer Cell *14*, 212-225.

1289 Mann, K.M., Guzman-Rojas, L., Amaya-Manzanares, F., Jones, D.J., Newberg, J.Y., Jenkins, N.A., Copeland, N.G., and
1290 Mann, M.B. (2016a). SBCapSeq Protocol: a method for selective cloning of _Sleeping Beauty_ transposon
1291 insertions using liquid capture hybridization and Ion Torrent semiconductor sequencing.

1292 Mann, K.M., Jenkins, N.A., Copeland, N.G., and Mann, M.B. (2013). Transposon Insertional Mutagenesis Models of
1293 Cancer. Cold Spring Harb Protoc.

1294 Mann, K.M., Jenkins, N.A., Copeland, N.G., and Mann, M.B. (2014a). Transposon insertional mutagenesis models of
1295 cancer. Cold Spring Harb Protoc *2014*, 235-247.

1296 Mann, K.M., Newberg, J.Y., Black, M.A., Jones, D.J., Amaya-Manzanares, F., Guzman-Rojas, L., Kodama, T., Ward, J.M.,
1297 Rust, A.G., van der Weyden, L., *et al.* (2016b). Analyzing tumor heterogeneity and driver genes in single myeloid
1298 leukemia cells with SBCapSeq. Nat Biotechnol *34*, 962-972.

1299 Mann, K.M., Ward, J.M., Yew, C.C., Kovochich, A., Dawson, D.W., Black, M.A., Brett, B.T., Sheetz, T.E., Dupuy, A.J., Chang,
1300 D.K., *et al.* (2012). Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic
1301 adenocarcinoma. Proc Natl Acad Sci U S A *109*, 5934-5941.

1302 Mann, M.B., Aiderus, A., Newberg, J.Y., Guzman-Rojas, L., Contreras-Sandoval, A.M., Meshey, A.L., Jones, D.J., Amaya-
1303 Manzanares, F., Rangel, R., Ward, J.M., *et al.* (2019a). Supplementary Tables1-43 for 'Transposon mutagenesis
1304 identifies cooperating genetic drivers during keratinocyte transformation and cutaneous squamous cell
1305 carcinoma progression'.

1306 Mann, M.B., Black, M.A., Jones, D.J., Ward, J.M., Yew, C.C., Newberg, J.Y., Dupuy, A.J., Rust, A.G., Bosenberg, M.W.,
1307 McMahon, M., *et al.* (2015). Transposon mutagenesis identifies genetic drivers of Braf(V600E) melanoma. Nat
1308 Genet *47*, 486-495.

1309 Mann, M.B., Jenkins, N.A., Copeland, N.G., and Mann, K.M. (2014b). Sleeping Beauty mutagenesis: exploiting forward
1310 genetic screens for cancer gene discovery. Curr Opin Genet Dev *24*, 16-22.

1311 Mann, M.B., Mann, K.M., Contreras-Sandoval, A.M., Guzman-Rojas, L., Newberg, J.Y., Meshey, A.L., Jones, D.J., Amaya-
1312 Manzanares, F., Copeland, N.G., and Jenkins, N.A. (2019b). SBCapSeq Protocol manuscript files for 'Quantifying
1313 tumor heterogeneity, clonal dynamics, and cancer driver gene evolution from Sleeping Beauty transposon
1314 mutagenesis models using SBCapSeq'.

1315 March, H.N., Rust, A.G., Wright, N.A., ten Hoeve, J., de Ridder, J., Eldridge, M., van der Weyden, L., Berns, A., Gadiot, J.,
1316 Uren, A., *et al.* (2011). Insertional mutagenesis identifies multiple networks of cooperating genes driving
1317 intestinal tumorigenesis. Nature genetics *43*, 1202-1209.

1318 Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B.,
1319 Tubio, J.M., *et al.* (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in
1320 normal human skin. Science *348*, 880-886.

1321 Mathios, D., Hwang, T., Xia, Y., Phallen, J., Rui, Y., See, A.P., Maxwell, R., Belcaid, Z., Casaos, J., Burger, P.C., *et al.* (2019).
1322 Genome-wide investigation of intragenic DNA methylation identifies ZMIZ1 gene as a prognostic marker in
1323 glioblastoma and multiple cancer types. Int J Cancer.

1324 Mathur, R., Alver, B.H., San Roman, A.K., Wilson, B.G., Wang, X., Agoston, A.T., Park, P.J., Shivdasani, R.A., and Roberts,
1325 C.W. (2017). ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. Nat
1326 Genet *49*, 296-302.

1327 Megidish, T., Xu, J.H., and Xu, C.W. (2002). Activation of p53 by protein inhibitor of activated Stat1 (PIAS1). J Biol
1328 Chem *277*, 8255-8259.

1329 Morton, J.P., Timpson, P., Karim, S.A., Ridgway, R.A., Athineos, D., Doyle, B., Jamieson, N.B., Oien, K.A., Lowy, A.M.,
1330 Brunton, V.G., *et al.* (2010). Mutant p53 drives metastasis and overcomes growth arrest/senescence in
1331 pancreatic cancer. Proc Natl Acad Sci U S A *107*, 246-251.

1332 Nadeau, J.H., and Dudley, A.M. (2011). Genetics. Systems genetics. Science *331*, 1015-1016.

1333 Nassar, D., Latil, M., Boeckx, B., Lambrechts, D., and Blanpain, C. (2015). Genomic landscape of carcinogen-induced
1334 and genetically induced mouse skin squamous cell carcinoma. Nat Med *21*, 946-954.

1335 Newberg, J.Y., Black, M.A., Jenkins, N.A., Copeland, N.G., Mann, K.M., and Mann, M.B. (2018a). SB Driver Analysis: a
1336 Sleeping Beauty cancer driver analysis framework for identifying and prioritizing experimentally actionable
1337 oncogenes and tumor suppressors. Nucleic Acids Res *46*, e94.

1338 Newberg, J.Y., Black, M.A., Jenkins, N.A., Copeland, N.G., Mann, K.M., and Mann, M.B. (2018b). SB Driver Analysis: a
1339 Sleeping Beauty cancer driver analysis framework for identifying and prioritizing experimentally actionable
1340 oncogenes and tumor suppressors. Nucleic Acids Res.

1341 Newberg, J.Y., Mann, K.M., Mann, M.B., Jenkins, N.A., and Copeland, N.G. (2017). SBCDDB: Sleeping Beauty Cancer
1342 Driver Database for gene discovery in mouse models of human cancers. Nucleic Acids Res.

1343 Newberg, J.Y., Mann, K.M., Mann, M.B., Jenkins, N.A., and Copeland, N.G. (2018c). SBCDDB: Sleeping Beauty Cancer
1344 Driver Database for gene discovery in mouse models of human cancers. Nucleic Acids Res *46*, D1011-D1017.

1345 Olive, K.P., Tuveson, D.A., Ruhe, Z.C., Yin, B., Willis, N.A., Bronson, R.T., Crowley, D., and Jacks, T. (2004). Mutant p53
1346 gain of function in two mouse models of Li-Fraumeni syndrome. Cell *119*, 847-860.

1347 Parman, C., Halling, C., and Gentleman, R. (2005). affyQCReport: QC Report Generation for affyBatch objects. R
1348 package version *1*, 1.

1349 Perez-Mancera, P.A., Rust, A.G., van der Weyden, L., Kristiansen, G., Li, A., Sarver, A.L., Silverstein, K.A., Grutzmann, R.,
1350 Aust, D., Rummele, P., *et al.* (2012). The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma.
1351 Nature *486*, 266-270.

1352 Pickering, C.R., Zhou, J.H., Lee, J.J., Drummond, J.A., Peng, S.A., Saade, R.E., Tsai, K.Y., Curry, J.L., Tetzlaff, M.T., Lai, S.Y.,
1353 *et al.* (2014). Mutational landscape of aggressive cutaneous squamous cell carcinoma. Clin Cancer Res *20*, 6582-
1354 6592.

1355 Pinnell, N., Yan, R., Cho, H.J., Keeley, T., Murai, M.J., Liu, Y., Alarcon, A.S., Qin, J., Wang, Q., Kuick, R., *et al.* (2015). The
1356 PIAS-like Coactivator Zmiz1 Is a Direct and Selective Cofactor of Notch1 in T Cell Development and Leukemia.
1357 Immunity *43*, 870-883.

1358 Potla, L., Boghaert, E.R., Armellino, D., Frost, P., and Damle, N.K. (2002). Reduced expression of EphrinA1 (EFNA1)
1359 inhibits three-dimensional growth of HT29 colon carcinoma cells. Cancer Lett *175*, 187-195.

1360 R_Core_Team (2008; 2013). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation
1361 for Statistical Computing).

1362 Ramirez, A., Page, A., Gandarillas, A., Zanet, J., Pibre, S., Vidal, M., Tusell, L., Genesca, A., Whitaker, D.A., Melton, D.W.,
1363 *et al.* (2004). A keratin K5Cre transgenic line appropriate for tissue-specific or generalized Cre-mediated
1364 recombination. Genesis *39*, 52-57.

1365 Rangel, R., Lee, S.C., Hon-Kim Ban, K., Guzman-Rojas, L., Mann, M.B., Newberg, J.Y., Kodama, T., McNoe, L.A.,
1366 Selvanesan, L., Ward, J.M., *et al.* (2016). Transposon mutagenesis identifies genes that cooperate with mutant
1367 Pten in breast cancer progression. Proc Natl Acad Sci U S A *113*, E7749-E7758.

1368 Rheinwald, J.G., and Beckett, M.A. (1981). Tumorigenic keratinocyte lines requiring anchorage and fibroblast
1369 support cultured from human squamous cell carcinomas. Cancer Res *41*, 1657-1663.

1370 Rheinwald, J.G., and Green, H. (1977). Epidermal growth factor and the multiplication of cultured human epidermal
1371 keratinocytes. Nature *265*, 421-424.

1372 Rogers, L.M., Riordan, J.D., Swick, B.L., Meyerholz, D.K., and Dupuy, A.J. (2013). Ectopic expression of Zmiz1 induces
1373 cutaneous squamous cell malignancies in a mouse model of cancer. J Invest Dermatol *133*, 1863-1869.

1374 Schmidt, D., and Muller, S. (2002). Members of the PIAS family act as SUMO ligases for c-Jun and p53 and repress p53
1375 activity. Proc Natl Acad Sci U S A *99*, 2872-2877.

1376 Schmidt, D., and Muller, S. (2003). PIAS/SUMO: new partners in transcriptional regulation. Cell Mol Life Sci *60*, 2561-
1377 2574.

1378 Sharma, M., Li, X., Wang, Y., Zarnegar, M., Huang, C.Y., Palvimo, J.J., Lim, B., and Sun, Z. (2003). hZimp10 is an
1379 androgen receptor co-activator and forms a complex with SUMO-1 at replication foci. EMBO J *22*, 6101-6114.

1380 Soler, G., Radford-Weiss, I., Ben-Abdelali, R., Mahlaoui, N., Ponceau, J.F., Macintyre, E.A., Vekemans, M., Bernard, O.A.,
1381 and Romana, S.P. (2008). Fusion of ZMIZ1 to ABL1 in a B-cell acute lymphoblastic leukaemia with a
1382 t(9;10)(q34;q22.3) translocation. Leukemia *22*, 1278-1280.

1383 South, A.P., Purdie, K.J., Watt, S.A., Haldenby, S., den Breems, N.Y., Dimon, M., Arron, S.T., Kluk, M.J., Aster, J.C.,
1384 McHugh, A.*, et al.* (2014). NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. J
1385 Invest Dermatol *134*, 2630-2638.

1386 Starr, T.K., Allaei, R., Silverstein, K.A., Staggs, R.A., Sarver, A.L., Bergemann, T.L., Gupta, M., O'Sullivan, M.G., Matise, I.,
1387 Dupuy, A.J.*, et al.* (2009). A transposon-based genetic screen in mice identifies genes altered in colorectal cancer.
1388 Science *323*, 1747-1750.

1389 Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G.V., Lawrence, M.S.,
1390 Sougnez, C., McKenna, A.*, et al.* (2011). The mutational landscape of head and neck squamous cell carcinoma.
1391 Science *333*, 1157-1160.

1392 Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H.,
1393 Bork, P.*, et al.* (2019). STRING v11: protein-protein association networks with increased coverage, supporting
1394 functional discovery in genome-wide experimental datasets. Nucleic Acids Res *47*, D607-D613.

1395 Takeda, H., Wei, Z., Koso, H., Rust, A.G., Yew, C.C., Mann, M.B., Ward, J.M., Adams, D.J., Copeland, N.G., and Jenkins, N.A.
1396 (2015). Transposon mutagenesis identifies genes and evolutionary forces driving gastrointestinal tract tumor
1397 progression. Nat Genet *47*, 142-150.

1398 Team, R.C. (2013). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for
1399 Statistical Computing).

1400 Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics
1401 *25*, 1105-1111.

1402 Vin, H., Ojeda, S.S., Ching, G., Leung, M.L., Chitsazzadeh, V., Dwyer, D.W., Adelmann, C.H., Restrepo, M., Richards, K.N.,
1403 Stewart, L.R.*, et al.* (2013). BRAF inhibitors suppress apoptosis through off-target inhibition of JNK signaling.
1404 Elife *2*, e00969.

1405 Wang, L., Zhao, Z., Ozark, P.A., Fantini, D., Marshall, S.A., Rendleman, E.J., Cozzolino, K.A., Louis, N., He, X., Morgan,
1406 M.A.*, et al.* (2018). Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer
1407 therapy. Nat Med *24*, 758-769.

1408 Wickham, H. (2009). Ggplot2 : elegant graphics for data analysis (New York: Springer).

1409 Yu, J., Wu, W.K., Liang, Q., Zhang, N., He, J., Li, X., Zhang, X., Xu, L., Chan, M.T., Ng, S.S.*, et al.* (2016). Disruption of
1410 NCOA2 by recurrent fusion with LACTB2 in colorectal cancer. Oncogene *35*, 187-195.

1411 Zhang, C., Spevak, W., Zhang, Y., Burton, E.A., Ma, Y., Habets, G., Zhang, J., Lin, J., Ewing, T., Matusow, B.*, et al.* (2015).
1412 RAF inhibitors that evade paradoxical MAPK pathway activation. Nature *526*, 583-586.

1413