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 established tumor suppressor genes, as well as previously unknown 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 could 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.

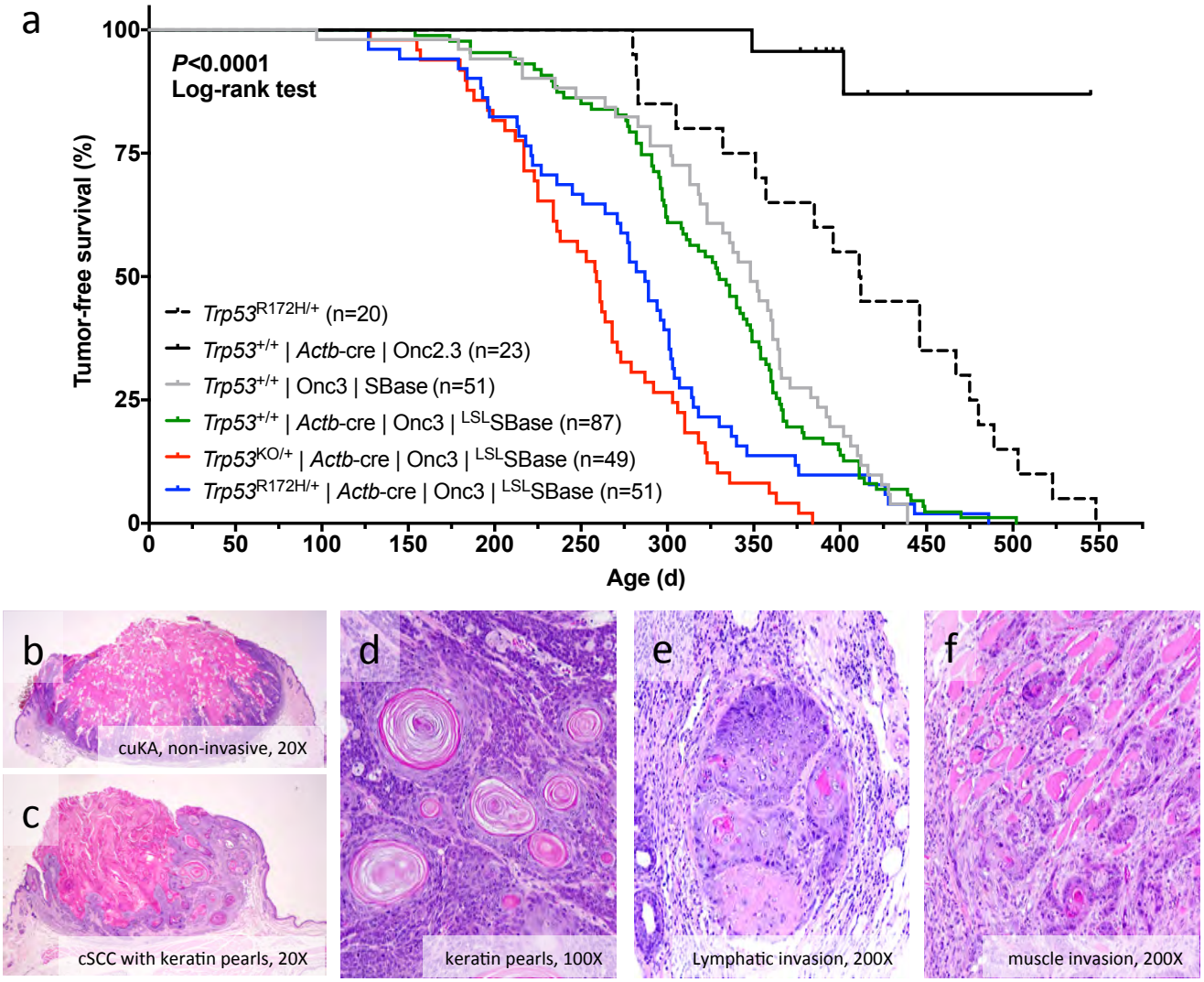
Cutaneous squamous cell carcinoma (cuSCC) is the second most common cancer in man, with approximately one million cases diagnosed annually in the United States. Although the majority of cuSCC are considered a low-risk neoplasm, up to 5% of high-risk cuSCCs are locally or distantly invasive and carry a poor prognosis due to a lack of biomarkers, therapeutic targets, or FDA-approved molecularly targeted therapies. This represents a substantial unmet need for approximately 50,000 patients per year with high-risk cuSCC, and an opportunity to identify new therapeutic modalities that could improve disease outcomes. All non-viral associated skin cancers are thought to require multiple cooperating mutations that deregulate distinct signaling pathways to initiate and progress the multi-step transformation of normal cells into a clinically significant neoplasm. Identifying cooperating mutations that drive malignant transformation is a prerequisite for developing better combinatorial therapies for managing and treating skin cancers. Most skin cancers, including cuSCC (Pickering et al., 2014; South et al., 2014), have the highest mutation rates among human cancers due to ultraviolet irradiation (UV-IR) induced damage from chronic, intermittent sun exposure. Thus, using human cancer sequencing data alone, with some of the highest mutational burdens of any cancer, poses challenges to identify cooperating, low-penetrant mutations that lead to cancer progression. This presents a need to develop *in vivo* model systems to help identify and prioritize novel cooperating candidate cancer drivers for keratinocyte transformation and subsequent progression to late-stage, invasive cuSCC. *Sleeping Beauty* (SB) insertional mutagenesis (Ivics et al., 1997) is a powerful tool used to perform genome-wide forward genetic screens in laboratory mice for cancer gene discovery (Collier and Largaespada, 2007; Copeland and Jenkins, 2010; Dupuy et al., 2005; Dupuy et al., 2009; Mann et al., 2013; Mann et al., 2016; Mann et al., 2012; Mann et al., 2015; Mann et al., 2014b; Rangel et al., 2016; Takeda et al., 2015) in animal models of both hematopoietic and solid tumors (Mann et al., 2014a; Mann et al., 2014b). SB transposons can identify early cancer progression drivers, which cooperate to initiate tumors (Mann et al., 2015; Takeda et al., 2015), and potentially drive metastasis

65 (Genovesi et al., 2013; Perez-Mancera et al., 2012). Importantly, SB insertions directly affect changes in gene
66 expression, thus providing information that is not easily obtained from current genomic sequencing technology
67 from chemical (Nassar et al., 2015) or chronic UV irradiation (Chitsazzadeh et al., 2016; Knatko et al., 2017)
68 carcinogenesis mouse models or limited patient samples. Here, we demonstrate that SB mobilization of a low-copy
69 T2/Onc3 transposon allele was sufficient to induce and progress a variety of cancers *in vivo*. We focused on the
70 analysis of skin tumors from these mice and identified several oncogenic and many tumor suppressor driver genes.
71 We profiled SB-induced keratinocyte transformation, including early- and late-stage cuSCC by high-throughput
72 sequencing approaches (Mann et al., 2016; Mann et al., 2015) to identify genome-wide SB mutations. Using our SB
73 Driver Analysis (Newberg et al., 2018b) statistical framework, we defined recurrently mutated, statistically
74 significant candidate cancer drivers (CCDs) from bulk tumors, as well as SB insertions from normal keratinocytes
75 and early stage tumors. We prioritized oncogenic and tumor suppressor candidates, and provide *in vitro* and *in vivo*
76 functional evidence of a role for these genes in the initiation and progression of cuSCC. Taken together, our study
77 provides a systematic evolutionary landscape of cuSCC genesis, and highlights key pathways promoting disease that
78 are potential therapeutic targets.

79 **RESULTS**

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

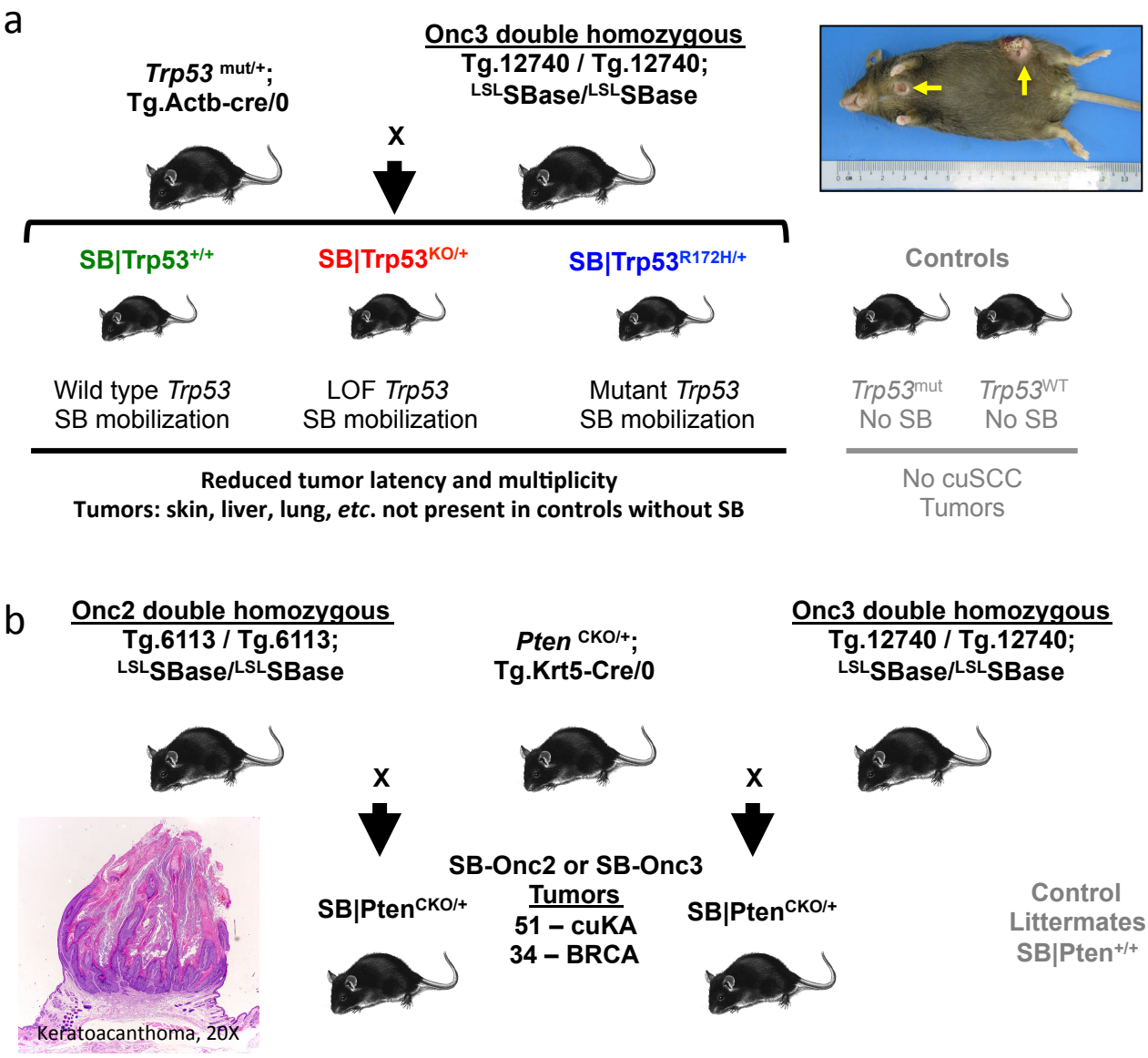
81 We performed a forward genetic screen to simultaneously drive tumorigenesis in multiple organ systems (**Fig. 1a**
82 and **Supplementary Fig. 1a**). Whole-body mutagenesis was achieved using low-copy, bi-functional SB transposons
83 from SB|Onc3 mice carrying an inducible SB transposase (Dupuy et al., 2009), Onc3 transposon (Dupuy et al., 2009),
84 and *Actb-Cre* transgene (Lewandoski et al., 1997). The addition of either a conditional *Trp53* null (*Trp53*$^{KO/+}$)
85 (Jonkers et al., 2001) or recurrent point mutant allele (*Trp53*$^{R172H/+}$) (Olive et al., 2004) accelerated tumor
86 progression and significantly reduced tumor-free survival relative to *Trp53* wild-type littermate controls (P<0.0001,
87 **Fig. 1a**). SB|*Trp53*$^{KO/+}$ mice were the earliest to reach endpoint, but had the lowest overall tumor burden,
88 presumably because these mice had the highest proportion of hematopoietic disease (**Supplementary Table 1**).
89 SB|*Trp53*$^{R172H/+}$ mice reached endpoint earlier than SB|*Trp53*$^{+/+}$ mice and had the widest spectrum of solid tumors
90 (**Supplementary Table 1**). This observation is consistent with previous studies showing that missense mutations in
91 *Trp53* promote malignant transformation by abrogating both the tumor suppressive and oncogenic (gain-of-
92 function and/or dominant negative) functions of TRP53 *in vivo* (Doyle et al., 2010; Hanel et al., 2013; Lang et al.,
93 2004; Morton et al., 2010; Olive et al., 2004). SB|Onc3 mice succumbed to a wide variety of solid and hematopoietic
94 tumor types with variable penetrance (**Supplementary Table 1**). The most prevalent of the more than 20 solid
95 tumor types observed were early-stage cutaneous keratoacanthoma (cuKA) (**Fig. 1b,g**) and late-stage, well
96 differentiated cutaneous squamous cell carcinoma (cuSCC) (**Fig. 1c-g**), consistent with their penetrance and latency
97 in Onc3 cohorts produced with constitutively active SBase (Dupuy et al., 2009). Importantly, these cuSCC/cuKAs
98 were absent in control mice with *Trp53* mutation without SB (**Fig. 1g**), strongly suggesting SB mutagenesis drives
99 initiation of cellular transformation and subsequent progression of a wide spectrum of tumor types *in vivo*.



g

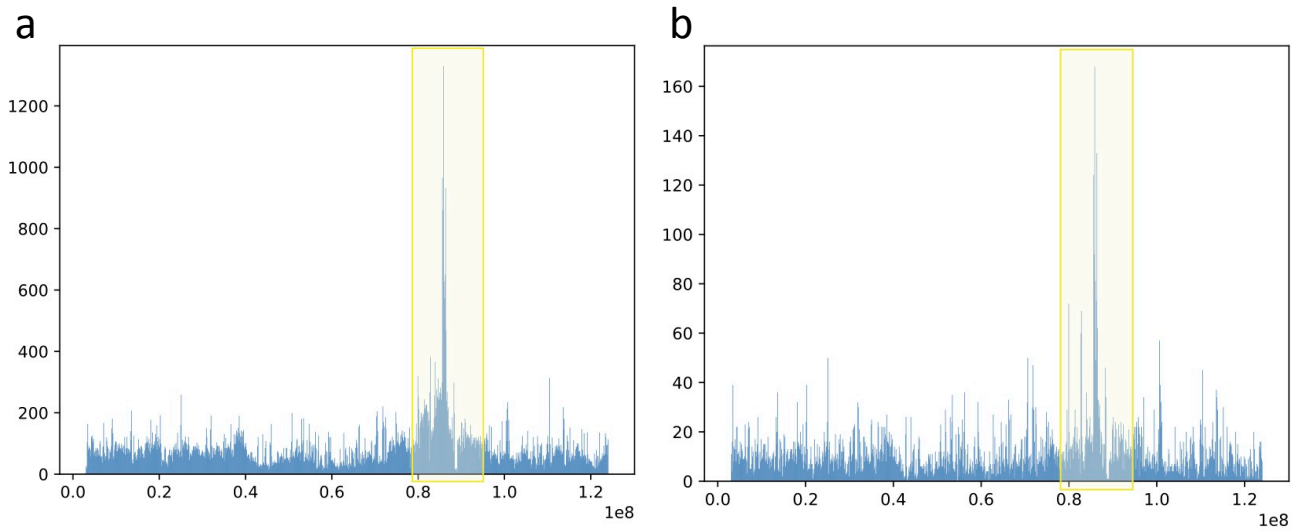
Cohort / Tumor classifications	SB-Onc3|Trp53-KO/+		SB-Onc3|Trp53-R172H/+		SB-Onc3|Trp53+/+		Trp53-R172H/+ Controls	
	# tumors	# mice	# tumors	# mice	# tumors	# mice	# tumors	# mice
Cutaneous Squamous Cell Carcinoma	17	10 (20%)	14	7 (15%)	52	20 (24%)	0	0 (0%)
Cutaneous Keratoacanthoma	3	3 (6%)	0	0 (0%)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

Figure 1: Whole-body SB transposon mutagenesis drives a diverse tumor spectrum in *Trp53* wildtype 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**).



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.

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

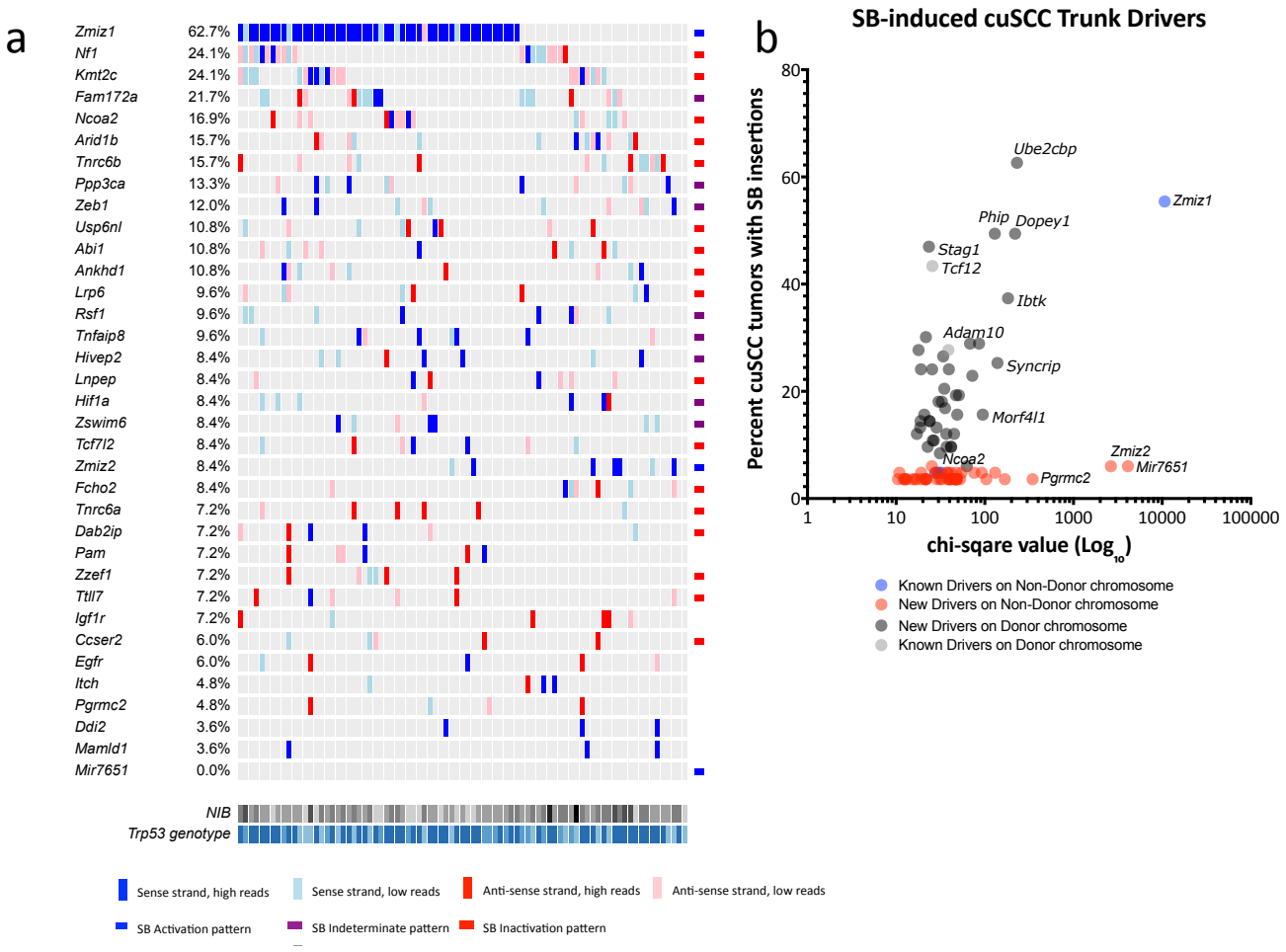


Supplementary Figure 2: SB T2/Onc3 TG.12740 allele donor position mapping and exclusion for SB Driver Analysis. Cumulative SB insertion sites at TA-dinucleotides from (a) 454_Splink and (b) Ion_SBCapSeq datasets showing the likely TG.12740 donor site at 87,000,000 bp and the exclusion region (yellow 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.*

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

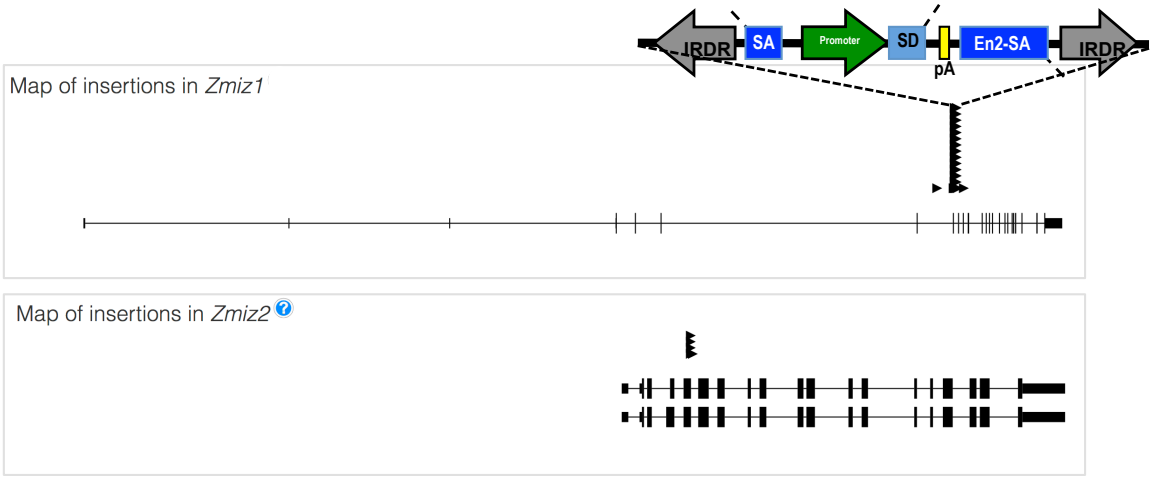
Next, we investigated keratinocyte transformation using keratinocyte-specific SB mutagenesis in *Pten*-sensitized mice. The screen originally was designed to activate an inducible allele of the SB transposase (Rosa26-LSL-SB11) using *Keratin5*-Cre (K5-Cre, Tg(KRT5-cre)5132Jlj, MGI:3050065) (Ramirez et al., 2004) to drive SB insertional

mutagenesis in stratified epithelium of the mammary ducts in a haploinsufficient *Pten*$^{floxed/+}$ sensitized background (Rangel et al., 2016). However, K5 also induces cre expression in cells of the basal epidermis. Using a two-generation mating scheme, we first introduced K5-Cre onto the background of a Cre-LoxP inducible *Pten* mutant strain, *Pten*$^{CKO/+}$ (*Pten*^{tm1Hwu} (MGI:2156086)). Compound heterozygous progeny (K5-Cre; *Pten*$^{CKO/+}$) were mated with double homozygous SB strains containing either a high-copy (T2/Onc2-TG.6113) or a low-copy (T2/Onc3-TG.12740) transposon donor concatemer and the inducible SB transposase (Rosa26-LSL-SBase or SBase^{LSL}) to generate experimental and control cohorts (Rangel et al., 2016) (**Supplementary Figure 1a**). Cutaneous keratoacanthomas (cuKA) were diagnosed from routine H&E sections and found to occur in 43% (13 of 30) of quadruple mutant mice (**Supplementary Table 1**), either K5-Cre; *Pten*$^{CKO/+}$; TG.6113; SBase^{LSL} or K5-Cre; *Pten*$^{CKO/+}$; TG.12740; SBase^{LSL}. Four of the mice with cuKA also developed a breast tumor and none of the mice in the control groups developed skin masses before reaching a predetermined experimental end-point at 450 days of age.

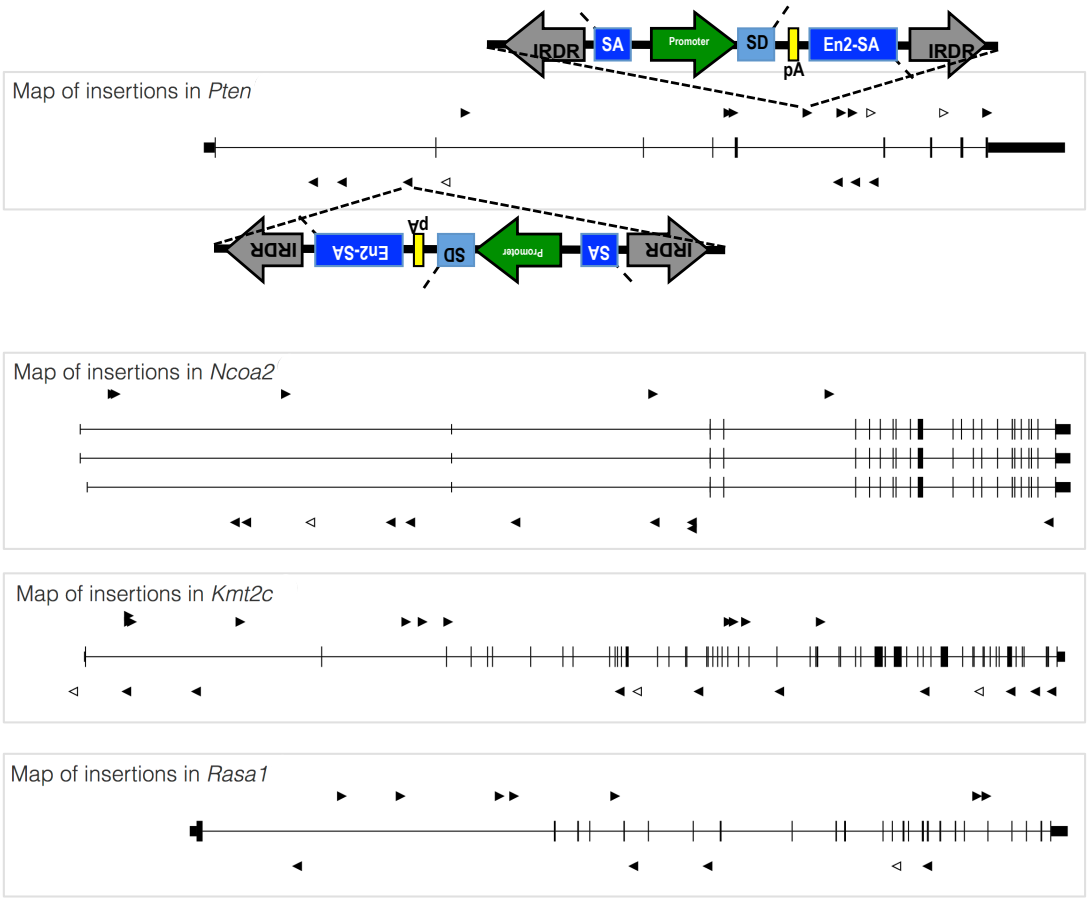


Supplementary Figure 3: Landscape of candidate truncal drivers mutated in SB-induced cuSCC with Splink_454 sequencing. SB Driver Analysis applied to Splink_454 SB insertion data (**Supplementary Table 4**) with read depth cutoff of 12 on 83 late-stage cuSCC genomes. (**a**) FWER significant driver genes (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**) Truncal cuSCC Driver genes containing SB insertion events with 10 or more 454 sequence reads per tumor. *(continues on next page)*

c



d

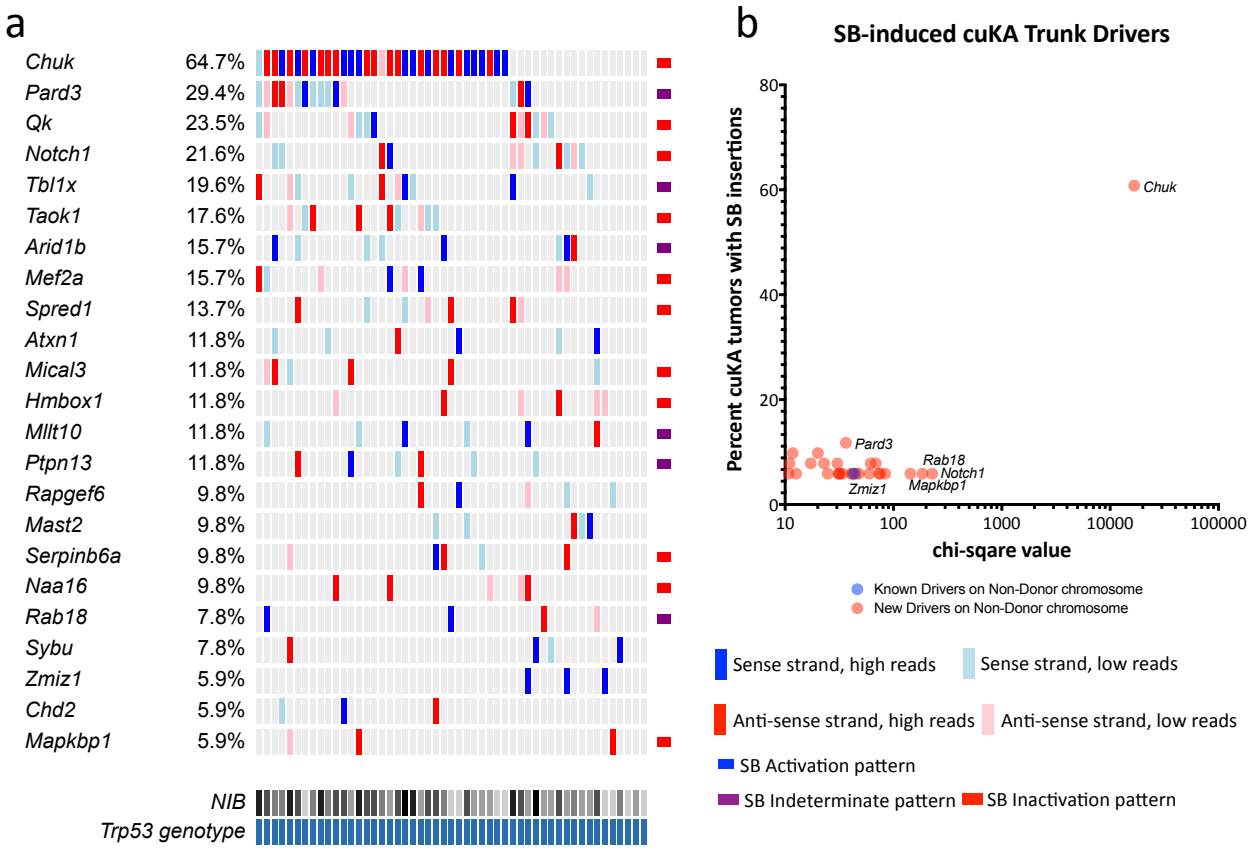


(continued from previous page) (**c-d**) Representative SB insertion maps in cuSCC trunk drivers showing the locations of mapped SB insertions (triangles) within (**c**) activating (oncogenic) pattern SB insertions into *Zmiz1* and *Zmiz2* or (**d**) Inactivating (tumor suppressive) pattern of SB insertions into *Pten*, *Ncoa2*, *Kmt2c*, and *Rasa1* from SB|cuSCC genomes.

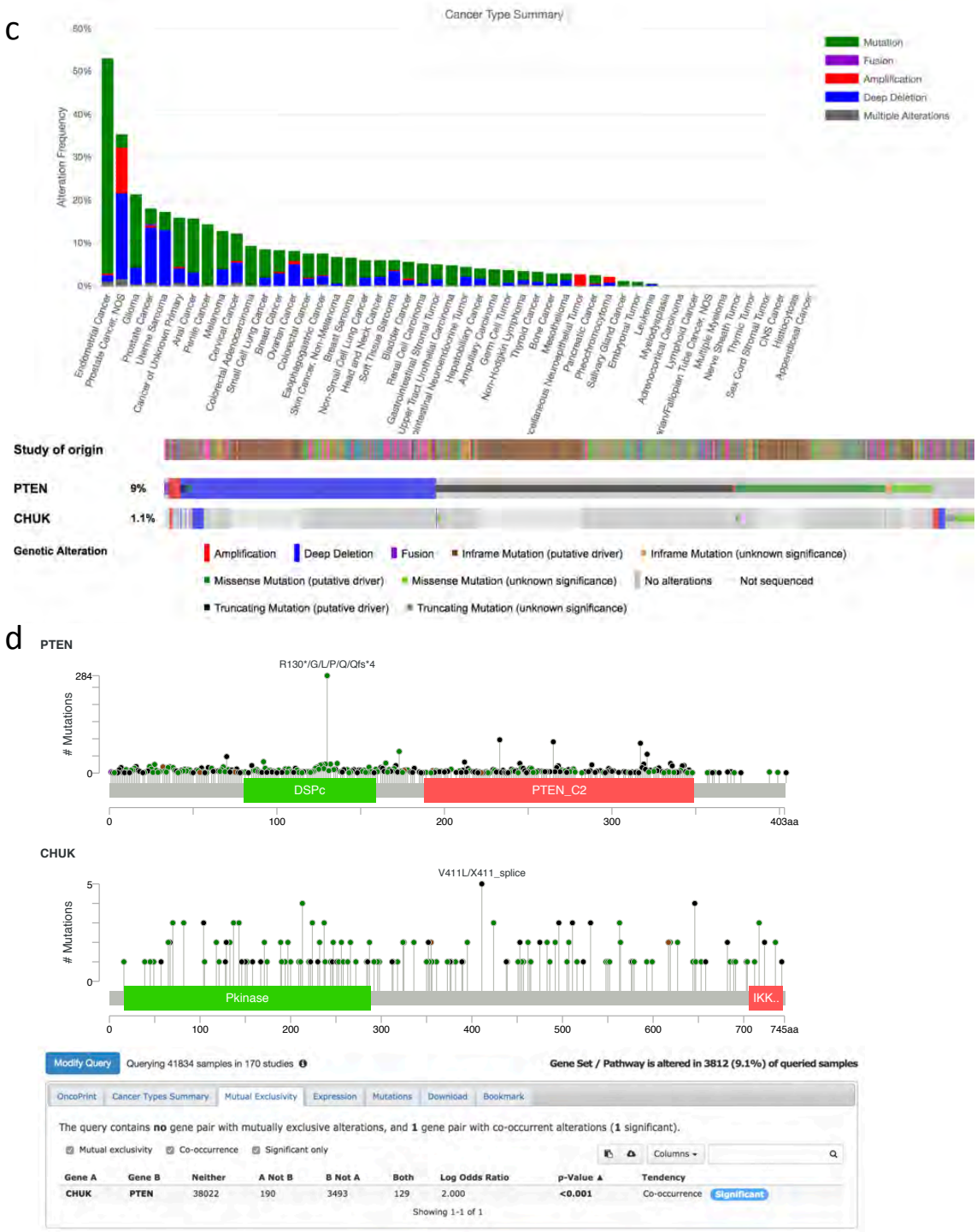
180 To identify SB drivers of cuKA in *Pten*-sensitized mice, we used 45 flash frozen cuKA specimens from 13 quadruple
181 mutant mice (**Supplementary Table 2–3**) collected at necropsy to create genomic DNA and sequenced using the
182 454-Splink method to sequencing SB insertion events. Using the 454 sequencing data (**Supplementary Table 7**), we
183 performed gene-guided SB Driver Analysis (Newberg et al., 2018b) to define 302 statistically significant candidate
184 cancer drivers (**Supplementary Table 8**) and 29 Trunk drivers (**Supplementary Table 9**). Our top ranked genes
185 were *Chuk* (*IKKA*/*IKK1*), *Notch1*, and *Kctd15* (**Supplementary Figure 4a-b**). All genes were found to have
186 inactivating patterns of SB insertion events into each of their gene-coding regions, strongly suggesting that
187 cooperating tumor suppressors drive early keratinocyte transformation and cuKA initiation *in vivo*. However, three
188 genomes lacking *Chuk* insertions were found to have activating SB insertions into *Zmiz1* that were identical to those
189 observed in the cuSCC cohorts (**Supplementary Figure 4a**). This may suggest that the altered function(s) of *Chuk* in
190 cuKAs may be overlapping with those of the SB-induced *Zmiz1*$^{\Delta N185}$ (Rogers et al., 2013) allele.

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

192 To confirm and extend our SB analysis from 454 sequencing, and achieve quantitative SB driver gene discovery
193 during the evolution of keratinocyte transformation, we developed an enhanced version of SB Capture Sequencing
194 (SBCapSeq) (Mann et al., 2016) that improves transposon-based, liquid-phase capture experimental workflow
195 optimized for Ion Torrent sequencing from normal and tumor specimen genomes (Mann, *et al*., 2019, co-submitted
196 *Nature Protocols* manuscript). We used SBCapSeq to investigate keratinocyte mutational heterogeneity by
197 histological stage.



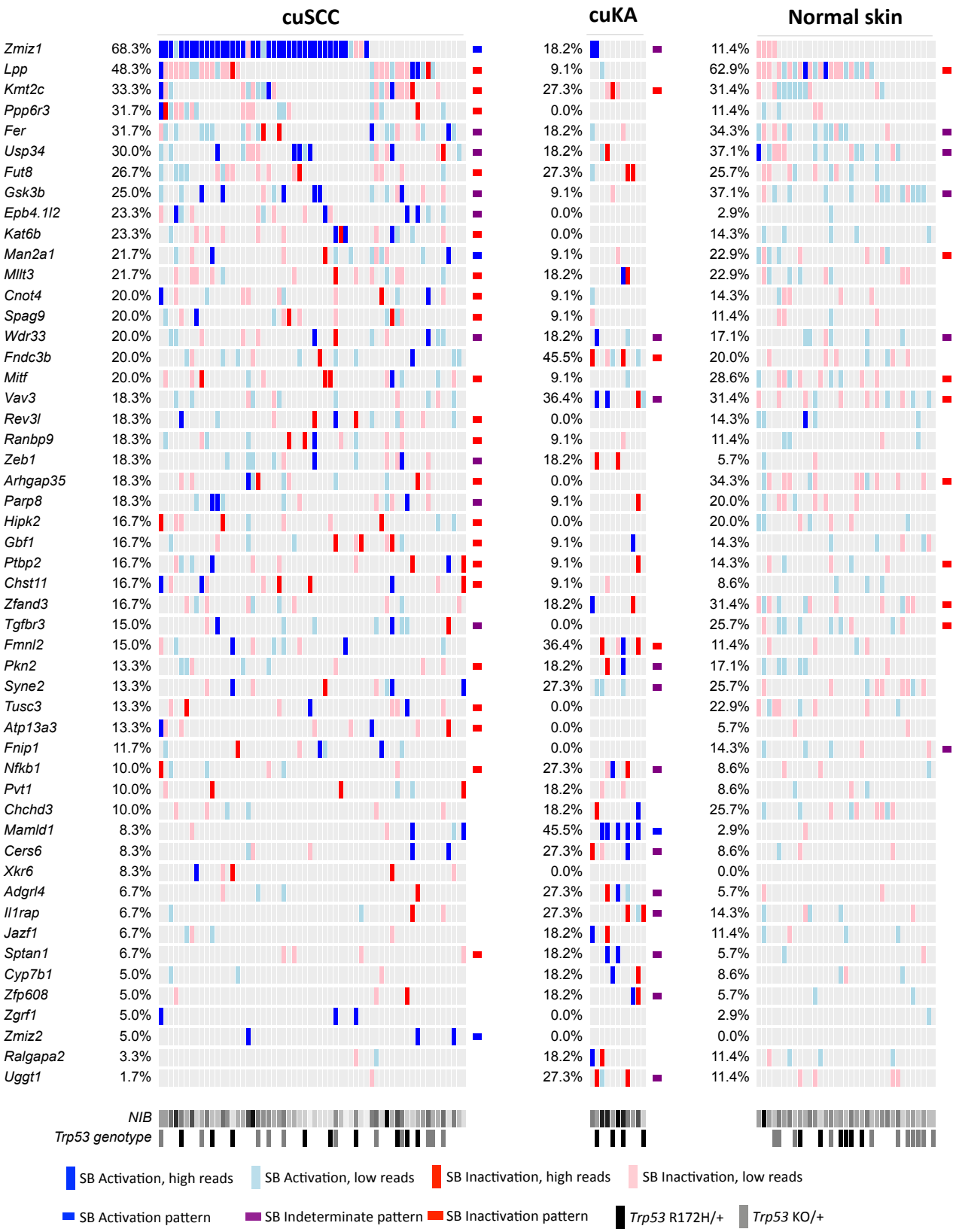
198



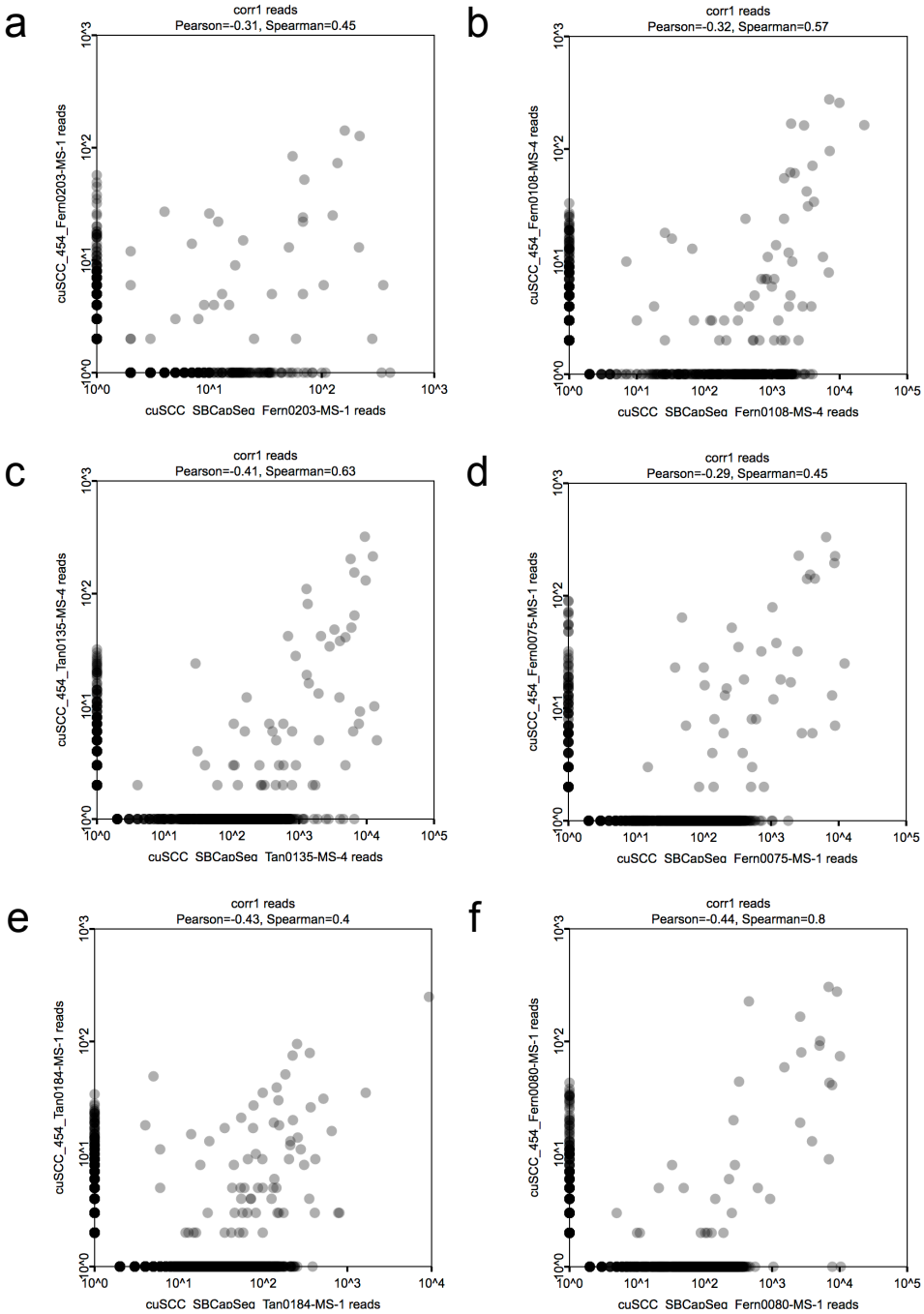
Supplementary Figure 4: Landscape of candidate truncal drivers mutated in SB-induced cuKA arising from the SB|*Krt5*-Cre|*Pten*$^{floxed/+}$ cohort with Splink_454 sequencing. SB Driver Analysis applied to Splink_454 SB insertion data (**Supplementary Table 7**) with read depth cutoff of 20 on 83 late-stage 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**) Truncal cuKA Drivers genes containing SB insertions events with 10 or more 454 sequence reads per tumor. (c) PTEN and CHUK alteration occurrences and Oncoprint across bBioPortal datasets. (d) PTEN and CHUK protein coding alterations and co-occurrence values from cBioPortal. *(Panels **a** and **b** on preceding page)*

209 We generated SBCapSeq libraries using genomic DNA isolated from the necropsy collected flash frozen specimens of
210 60 cuSCC, 11 cuKA and 32 normal skin genomes (**Fig. 2** and **Supplementary Table 2**). All cuSCC specimens were
211 from independent mice and 59 of 60 were previously sequenced using the Splink_454 method (**Supplementary**
212 **Table 10**). Comparisons between the sequencing methods allowed for platform-independent sequence validation of
213 clonal drivers and demonstrated the superiority of the SBCapSeq method based on the total number and read-depth
214 enrichment of genome-wide SB insertion events and biological and technical reproducibility (**Supplementary**
215 **Figure 5**). Using the SBCapSeq bioinformatics workflow (Mann et al., 2016) and gene-guided SB Driver
216 Analysis(Newberg et al., 2018b) statistical framework, we identified 11,113,694 reads at 59,337 non-redundant
217 transposon insertion sites (**Supplementary Table 11**) which were used to statistically define 1,333 candidate
218 cancer driver genes (**Supplementary Table 12**). Additionally, we identified 86 truncal drivers (**Supplementary**
219 **Figure 6a** and **Supplementary Table 13**). Using the same methodology, we next sequenced the cuKA specimens
220 harvested either from mice with cuSCC or from an independent age-/genotype-matched cohort (**Supplementary**
221 **Table 10**) and identified 2,085,280 reads at 6,727 non-redundant transposon insertion sites (**Supplementary**
222 **Table 14**) defining 192 candidate cancer driver genes (**Supplementary Table 15**) and 10 truncal drivers
223 (**Supplementary Figure 6b** and **Supplementary Table 16**). Finally, we identified 2,910,945 reads at 47,537 non-
224 redundant transposon insertion sites (**Supplementary Table 17**) from normal skin specimens (**Supplementary**
225 **Table 4**) that defined 547 candidate cancer drivers (**Supplementary Table 18**). No truncal drivers were identified.

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

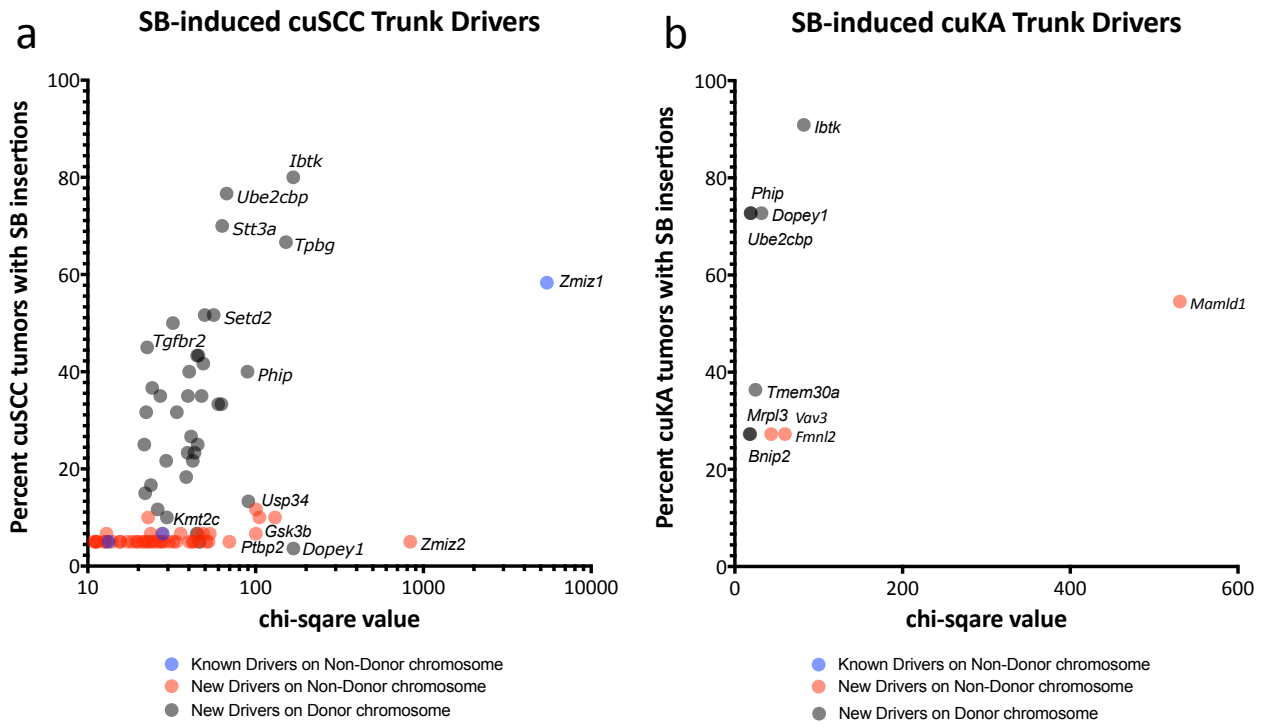

cuSCC
cuKA
Normal skin
Zmiz1 68.3% 18.2% 11.4%
Lpp 48.3% 9.1% 62.9%
Kmt2c 33.3% 27.3% 31.4%
Ppp6r3 31.7% 0.0% 11.4%
Fer 31.7% 18.2% 34.3%
Usp34 30.0% 18.2% 37.1%
Fut8 26.7% 27.3% 25.7%
Gsk3b 25.0% 9.1% 37.1%
Epb4.1l2 23.3% 0.0% 2.9%
Kat6b 23.3% 0.0% 14.3%
Man2a1 21.7% 9.1% 22.9%
Mllt3 21.7% 18.2% 22.9%
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/+

238



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.

250 The histologically normal skin genomes from mice with SB mobilization had a mean of 1,320 SBCapSeq detected
251 insertion events (range 1–3,857; **Supplementary Table 2**), although only 0.82% (389/47,537) of the normal skin
252 insertions had read depths >300 reads per site (range 1–2,297; **Supplementary Table 17**). This is one order of
253 magnitude less than the 23.7% (range 1-19,798) of cuKA genomes or 8.5% (5,014/57,337) of cuSCC genomes with
254 SBCapSeq detected insertion events with >300 reads per specimen (cuKA range 1-19,798; **Supplementary Table**
255 **14** and cuSCC range 1–35,809; **Supplementary Table 11**). Clonal selection of recurrent, high-read depth SB
256 insertion sites is featured prominently within the cuSCC and cuKA genomes and almost entirely missing within the
257 normal skin genomes (**Fig. 2**), strongly suggesting that SBCapSeq read depths below ~300 represent background
258 events and those at 300 and substantially higher represent clonally selected foreground mutation events.



259

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

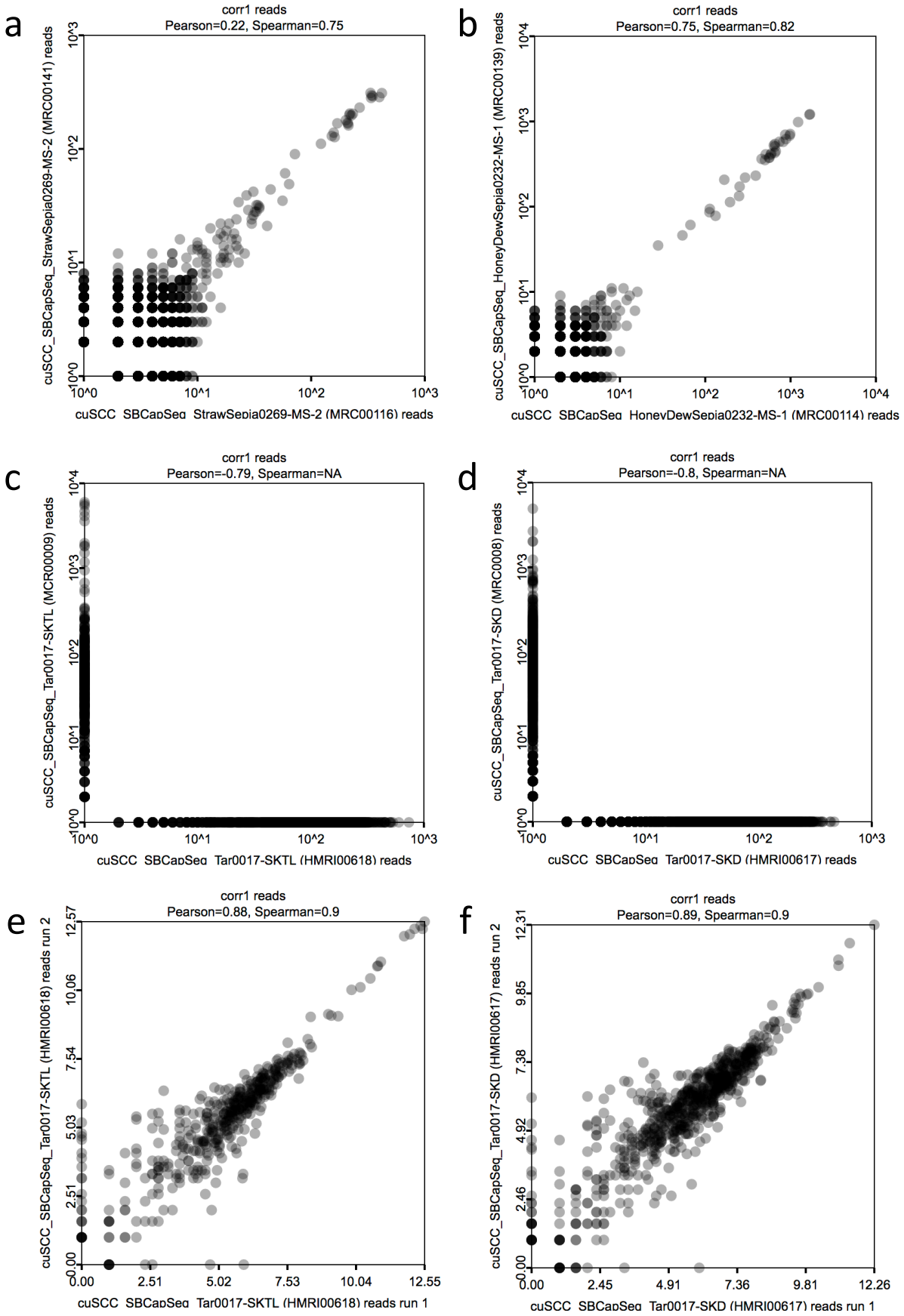
265 Further evidence for clonal selection associated with disease progression was provided by technical replication of
266 SBCapSeq libraries prepared from two independent cuSCC or normal skin specimens (**Supplementary Figure 7**).
267 Independent SBCapSeq libraries demonstrated excellent biological/technical reproducibility in clonal cuSCC
268 genomes (**Supplementary Figure 7a-b**) and no reproducibility in non-clonal normal skin genomes
269 (**Supplementary Figure 7c-d**). Importantly, sequence analysis of the same normal skin SBCapSeq library from
270 independent Ion Torrent runs demonstrated excellent technical reproducibility (**Supplementary Figure 7e-f**),
271 further supporting the robustness of our methodology to detect clonal selection and validating our observation that

272 normal skin does not exhibit clonal selection of SB insertions, despite the high insertion burden. These observations
273 are consistent with the large numbers of mutations observed during chronic UV-IR exposure in human eyelid
274 (Martincorena et al., 2015) and SKH hairless mouse (Chitsazzadeh et al., 2016; Knatko et al., 2017) skin and,
275 together with our SB data, indicates that keratinocytes within phenotypically normal skin can tolerate high
276 background mutagenic insults that may act as a mutational reservoir prior to clonal selection of cooperating driver
277 mutations that permit frank outgrown of cuSCC *in vivo*. It is for this reason that no truncal drivers could be identified
278 from the normal skin specimen genomes. The quantitative, stage-dependent insertional mutation burdens derived
279 from SBCapSeq in cutaneous specimens reveal incredible levels of tumor heterogeneity that has not previously been
280 observed for any SB-driven solid tumor.

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

282 We identified four skin masses with histologically continuous but distinct regions of cuKA and cuSCC, suggesting for
283 the first time that cuKA likely progresses directly to late-stage invasive cuSCC *in vivo* (**Supplementary Table 19**).
284 We performed SBCapSeq on representative histologically distinct regions to investigate intratumor heterogeneity
285 and gain insight into the drivers likely to mediate cuKA to cuSCC progression (**Fig. 3a-c**, **Supplementary Figure 8a-**
286 **f**, and **Supplementary Table 20**). Invariably, clonal insertions into *Zmiz1* were identified across each section of all 4
287 masses (**Fig. 3a**), at three different TA-dinucleotide addresses (**Supplementary Figure 8a-d**), suggesting that *Zmiz1*
288 was under positive selection within the nascent transformed cells that gave rise to the distinct cuKA and cuSCC
289 features. Additional tumor-specific clonal (including *Fer*, *Rere*, *Tcf4*, and *Tcf12*) and sub-clonal (including *Arhgap10*,
290 *Egr2*, *Pak2*, and *Rab3gap1*) SB insertions sites were also identified **Fig. 3b-c** and **Supplementary Figure 8e-f**),
291 lending insight into potential cooperating drivers in keratinocyte transformation progression.

292 **Supplementary Figure 7: Evaluating the reproducibility of SBCapSeq results from bulk cuSCC and normal skin**
293 **specimens.** Representative plots of the individual SB insertion sites based on read depth from two individual cuSCC genomes (**a-**
294 **b**) and two unselected skin cell genomes from histologically normal skin (**c-d**) from genomic DNAs created from bulk tumor
295 specimens from biological replicate libraries (independent library workflows applied to the same biological specimen isolate)
296 comparing library 1 (*x*-axis) to library 2 (*y*-axis). Biological reproducibility of cuSCC specimen libraries is high, supported by both
297 Pearson's and Spearman's correlation metrics, using all of SB insertion sites with read depths of 2 or higher (**a-b**). In contrast,
298 biological reproducibility of normal skin specimen libraries is exceptionally low, supported by both Pearson's and Spearman's
299 correlation metrics, and fails entirely to identify the same SB insertion sites (**c-d**), indicating that these are likely background
300 passenger insertion events that are not reproducibly sequenced. However, plotting of the individual SB insertion sites based on
301 read depth from two technical replicates (the same library preparation) from two separate sequencing runs confirms very high
302 reproducibility between individual libraries at high-to-moderate read depths, supported by both Pearson's (r= 0.48) and
303 Spearman's (ρ=0.49 or undetermined) correlation metrics (**e-f**), further supporting the lack of biological replication is most likely
304 due to non-clonal background passenger insertion events and do not result from flaws in the SBCapSeq protocol *per se*. *(Image on*
305 *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

306

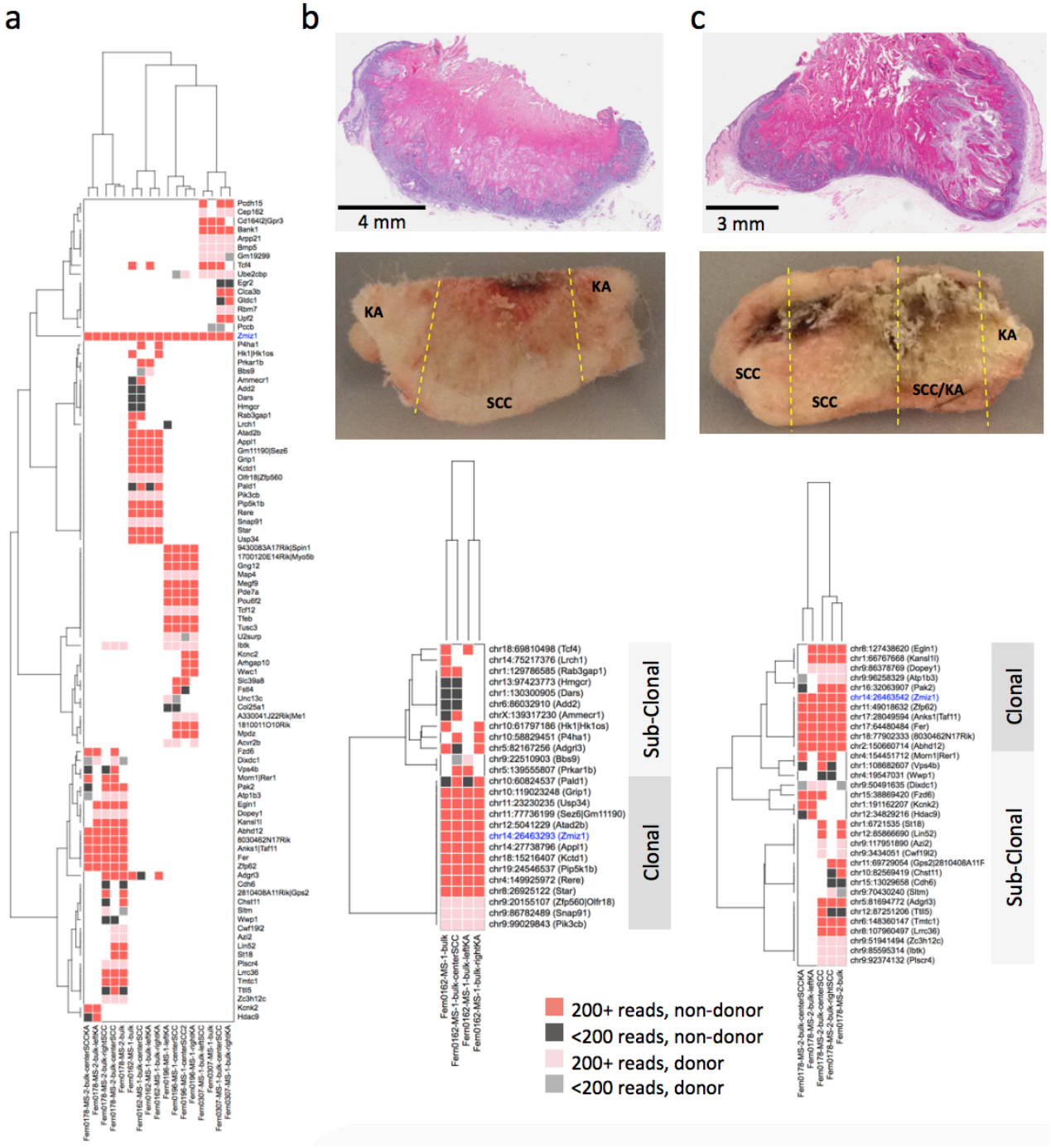
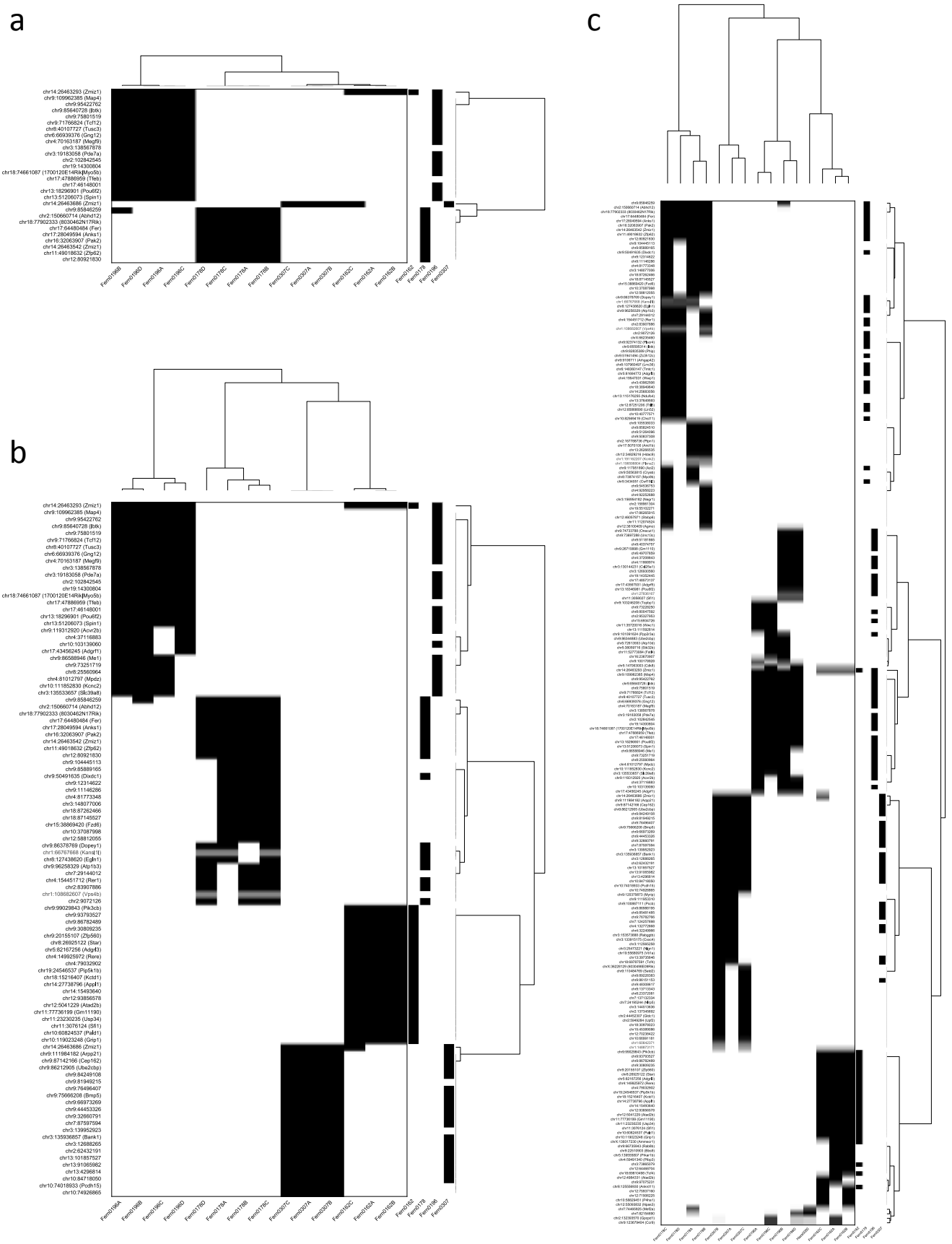
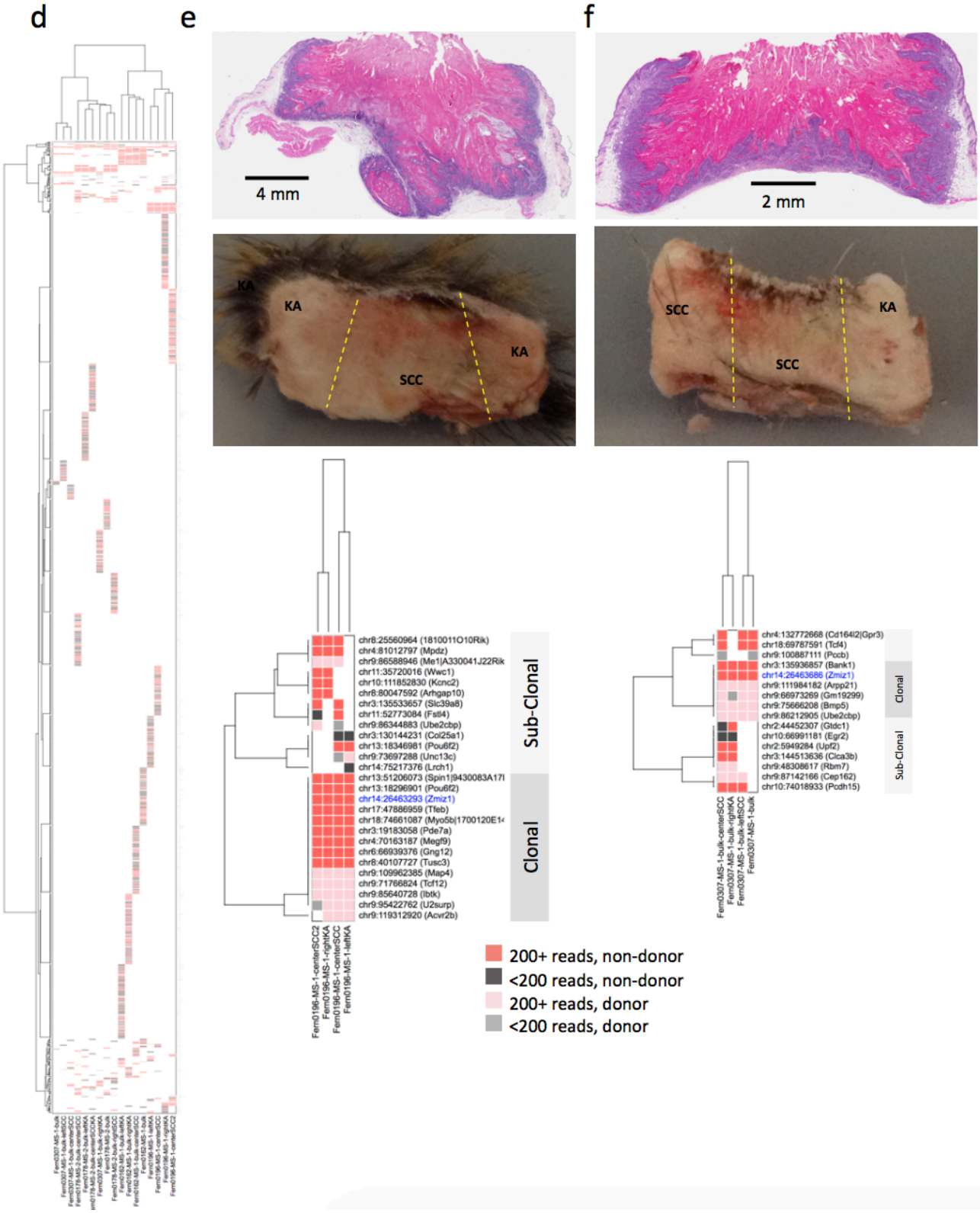


Figure 3: Multi-region SBCapSeq analysis of dual histology keratinocyte-derived skin masses containing distinct cuKA and cuSCC histological regions. (**a**) Hierarchical two-dimensional clustering (Hamming distance with the Ward method of agglomeration) of recurrent genic SBCapSeq insertion data from 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 truncal (common to all samples) and sub-clonal, lesion-specific SB insertions.



320



Supplementary Figure 8: Multi-region SBCapSeq analysis of bi-lesional keratinocyte-derived skin masses containing distinct cuKA and cuSCC histologies. Hierarchical two-dimensional clustering (Hamming distance with the Ward method) of recurrent genic SBCapSeq insertion data from four skin masses containing distinct cuKA and cuSCC regions with ≥4 (**a**), ≥3 or more (**b**), or ≥2 (**c**), recurrent SB insertions. At each recurrent threshold, four distinct specimen (*x*-axis) and gene (*y*-axis) clades, each pertaining to a single bi-lesional mass, were detected, thus demonstrating clonal outgrowth. Genomic DNA was isolated from adjacent, non-overlapping serial sections and three or four regions dissected from histologically distinct regions; the tumor

328 reference, consisting of a cross-section of each mass (including all regions), is plotted as independent tracks on the right of each
329 graph (see **panels e-f** and **Fig. 3 a-b**). (**d**) Hierarchical two-dimensional clustering (Hamming distance with the Ward method) of
330 recurrent genic SBCapSeq insertion data from four skin masses containing distinct cuKA and cuSCC regions. Four distinct
331 specimen (*x*-axis) and gene (*y*-axis) clades, each pertaining to a single bi-lesional mass, were detected, demonstrating clonal
332 identities. *Zmiz1* was the only gene recurrently mutated across all samples with high read depths; most SB insertions were
333 private to a single sample and consistently had lower read depths suggesting these are unselected passenger insertions. Genomic
334 DNA was isolated from adjacent serial sections and included a "bulk" reference consisting of a cross-section of the entire mass,
335 and three or four regions dissected from histologically distinct regions (see **panels e-f**). (**e-f**) Two masses with Aperio slide scans
336 of H&E stained FFPE specimens showing distinct cuKA and cuSCC histologies suggestive of clonal evolution of cuSCC
337 differentiation within cuKA masses and the adjacent flash frozen tissue prior to sectioning for genomic DNA isolation.
338 Hierarchical two-dimensional clustering (Hamming distance with the Ward method) of recurrent genic SBCapSeq insertion data
339 shown below each specimen demonstrates truncal (common to all samples) and sub-clonal, lesion-specific SB insertions.

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

341 In a previously published study, a surprising insight into genotype-phenotype relationships during keratinocyte
342 transformation was found by comparing the sequencing of early-stage cuKA masses that arose spontaneously in a
343 breast cancer cohort consisting of quadruple heterozygous mice carrying alleles for K5-Cre, Pten-floxed,
344 Onc3(TG.12740), ^{LSL}SBase to phenotypically indistinguishable cuKA masses observed in the SB|P53 cohorts (Rangel
345 et al., 2016; Stransky et al., 2011). In keratinocytes with SB mobilization and heterozygous loss of *Pten*, activating SB
346 insertions into *Zmiz1, Zmiz2,* and *Mamld1* were rarely, if ever, observed. Instead, inactivating insertions into *Chuk*
347 were statistically over-represented and, when they occurred, were mutually exclusive with *Zmiz1* or *Mamld1*
348 insertions, suggesting they may be within the same genetic pathway. Surprisingly, SBCapSeq analysis of normal skin
349 from 9 separate mice with the same quadruple heterozygous genotype revealed many low-read depth insertions
350 into the *Pten* locus, but no insertions in *Chuk* (**Supplementary Table 21-22**), suggesting that selection for *Chuk*
351 inactivation may be dependent upon haploinsufficient loss of *Pten*. In addition, using the datasets within the recently
352 published SBCDDB (Newberg et al., 2017), we found statistically significant evidence for co-occurring loss of *Pten*
353 (either via SB insertion or by breeding in a heterozygous null allele) and SB inactivation of *Chuk* across solid tumor
354 models (Fisher's exact test, $P=2.36 \times 10^{-7}$). Statistical significance for co-occurring mutations in *Pten* and *Chuk* in
355 solid cancers persisted even after removing all cuKA tumors from the analysis (Fisher's exact test, $P=3.78 \times 10^{-16}$),
356 suggesting this relationship may be generally observed in non-keratinocyte tumor lineages. Remarkably, querying
357 for co-occurrence of alterations in *PTEN* and *CHUK* via cBioPortal, across 41,834 patient samples from 170 studies,
358 revealed significant co-occurrence for these two genes in human cancers (Chi-square with Yates correction, $\chi^2=406$,
359 $P<0.001$; **Supplementary figure 4c-d**). Collectively, these data suggest that coordinated loss of both *Pten* and *Chuk*
360 may have a role in initiating and progressing oncogenic transformation in keratinocyte as well as non-keratinocyte
361 lineages. Given the recurring inactivating SB insertions in *Chuk* and activating *Zmiz1* observed in our *Pten*-loss
362 sensitized mice (**Supplementary Figure 4a**), it is curious that a recent separate *Pten*-sensitized SB screen for
363 cooperating TSGs failed to identify *Chuk* or *Zmiz1* entirely (de la Rosa et al., 2017). One possibility for this difference
364 is that the reported screen used a single-copy SB allele, while our screen used multi-copy SB alleles with multiple
365 independent genic insertion events per tumor, and the lack of *Chuk* or *Zmiz1* insertion events may suggest that
366 positive selection for variant *Chuk* or *Zmiz1* alleles in haploinsufficient *Pten*-mutant skin cells may require one or
367 more additional cooperating drivers.

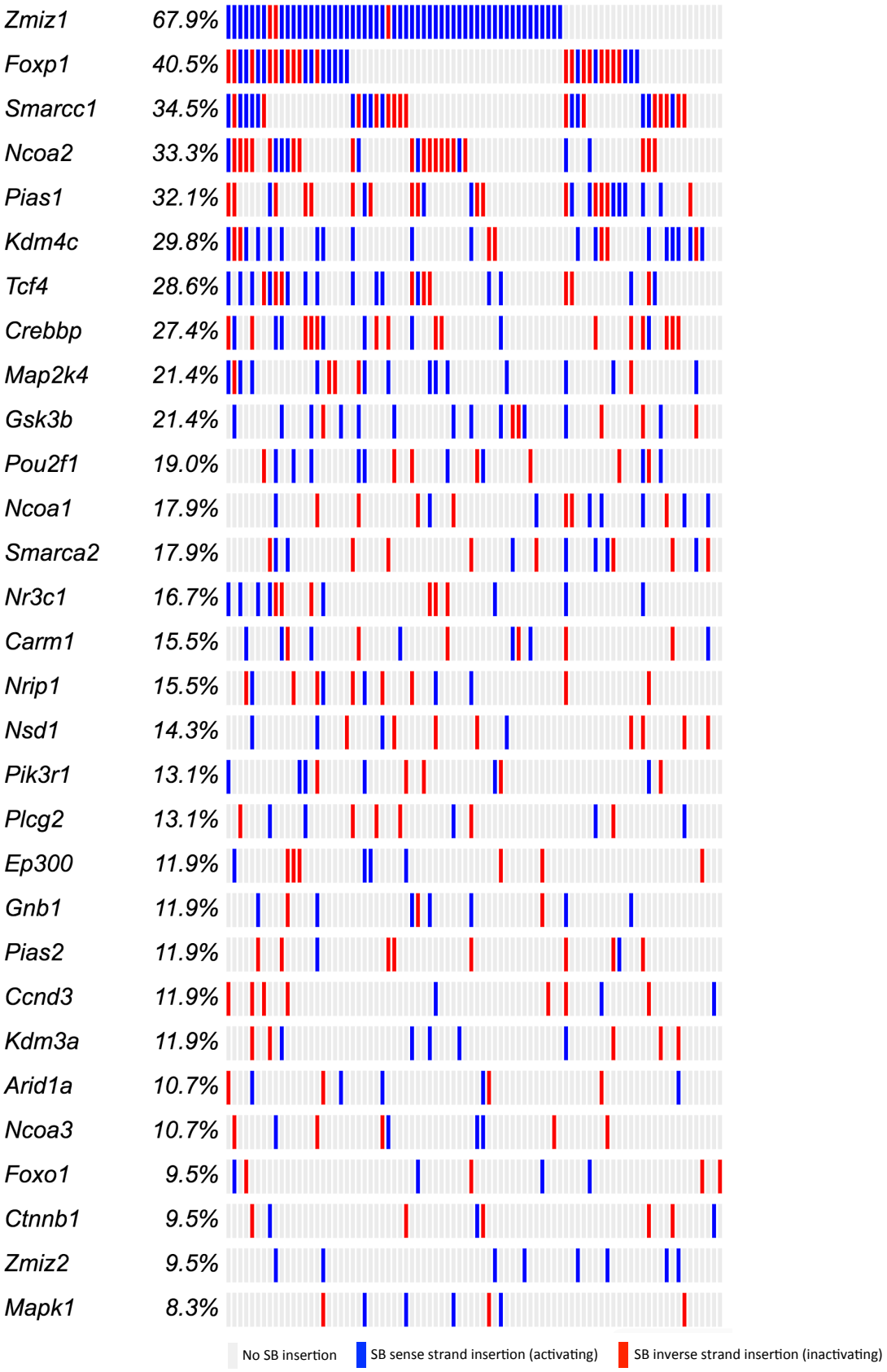
a

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, $q<0.001$

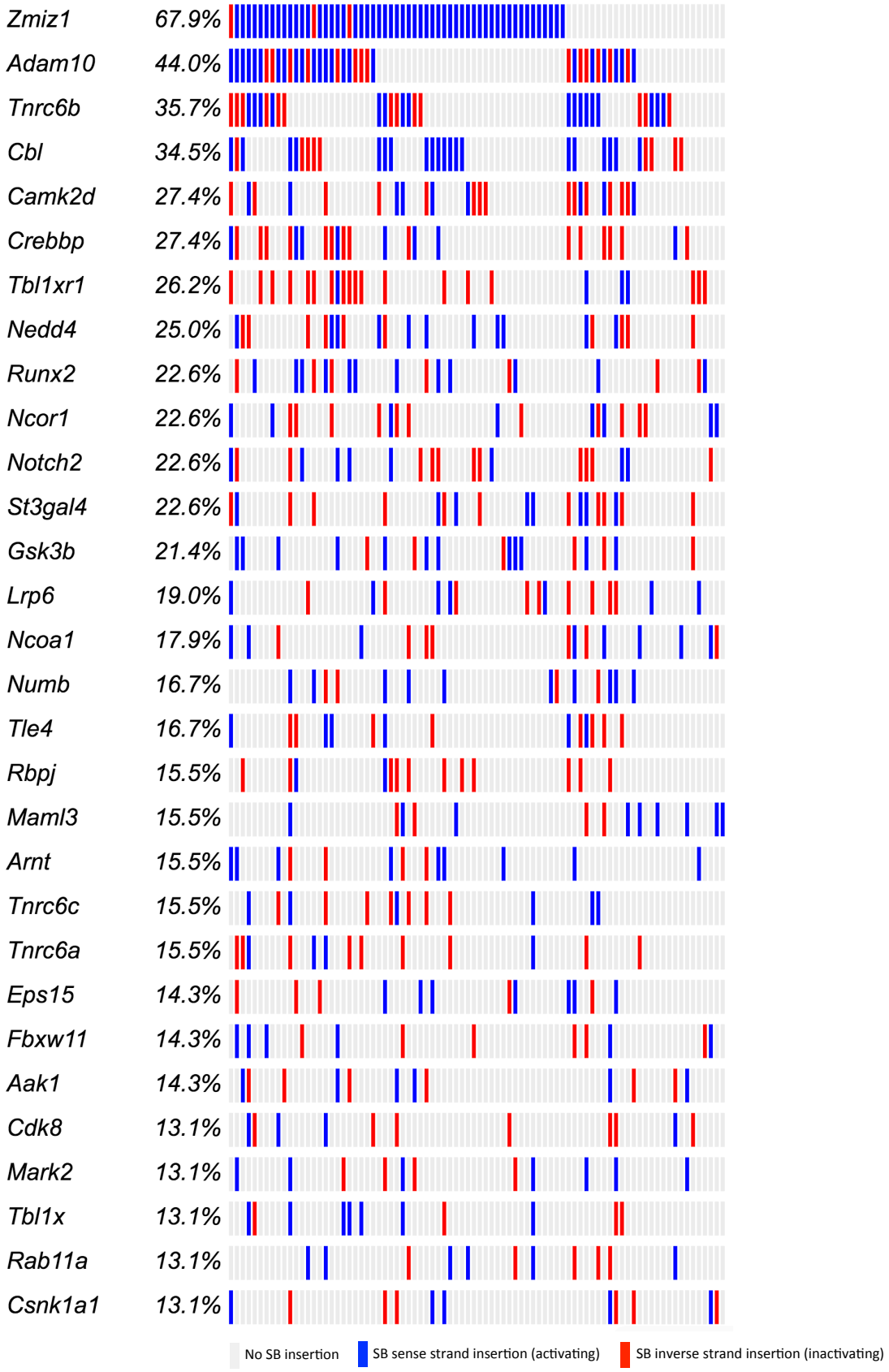
Supplementary Figure 9: Curated biological pathways and processes enriched within SB-induced cuSCC. (**a**) Table of significant pathways collated from pathway enrichment categories in **Supplementary Table 28**. Oncoprints of the top 30 mutated cuSCC candidate drivers enriched in (**b**) Androgen Receptor 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)

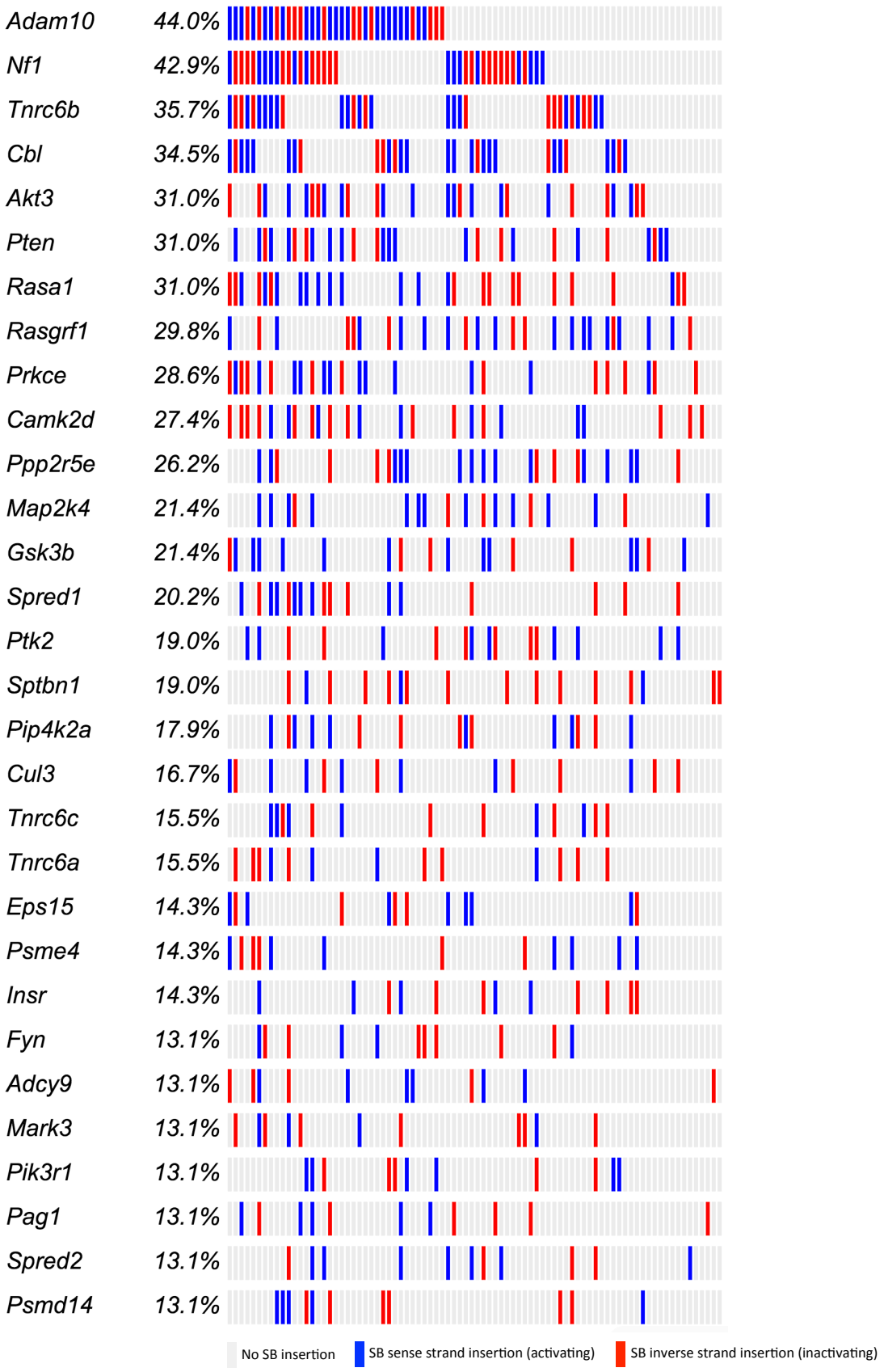
374

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)

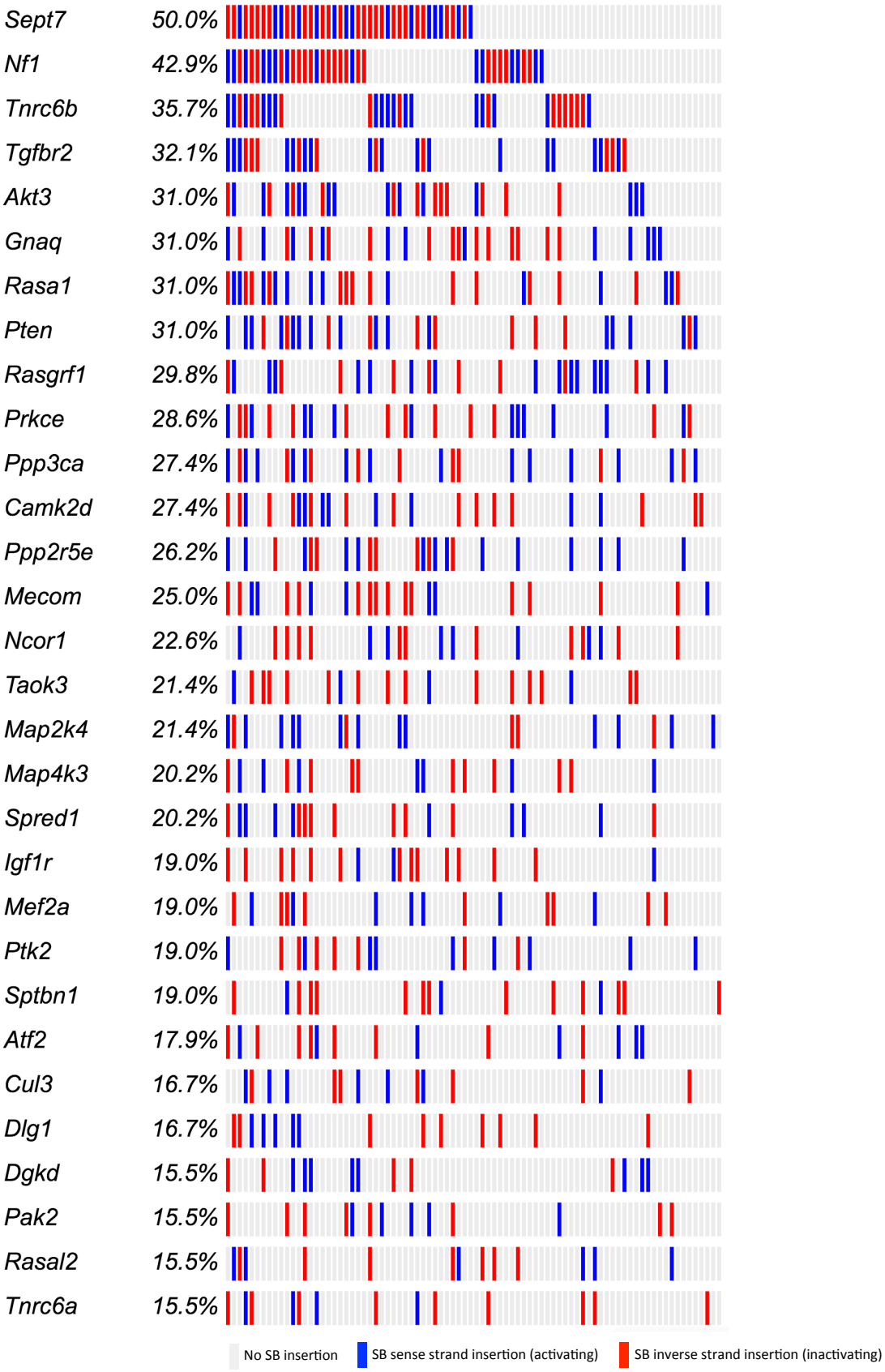
375

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)

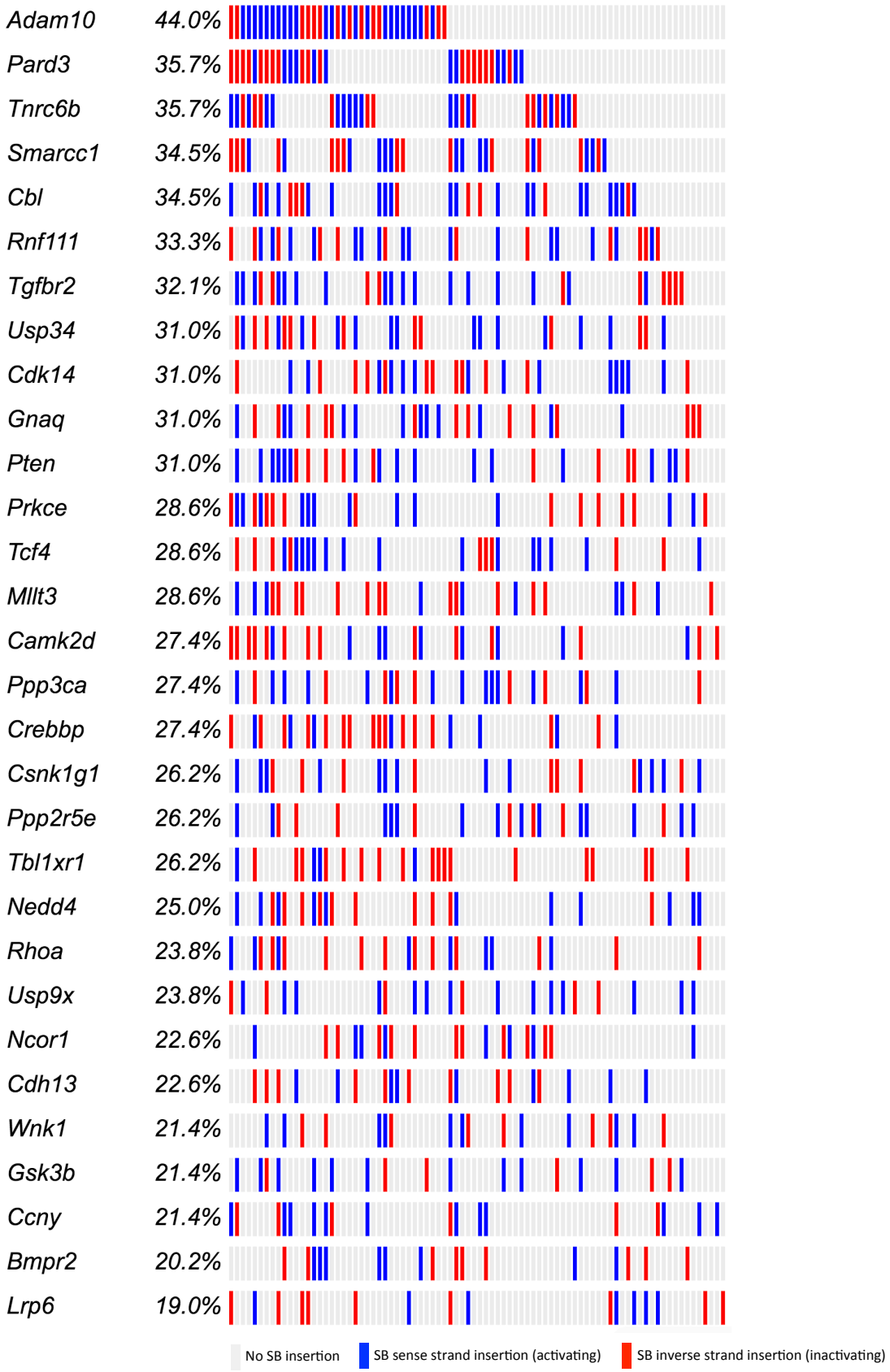
376

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)

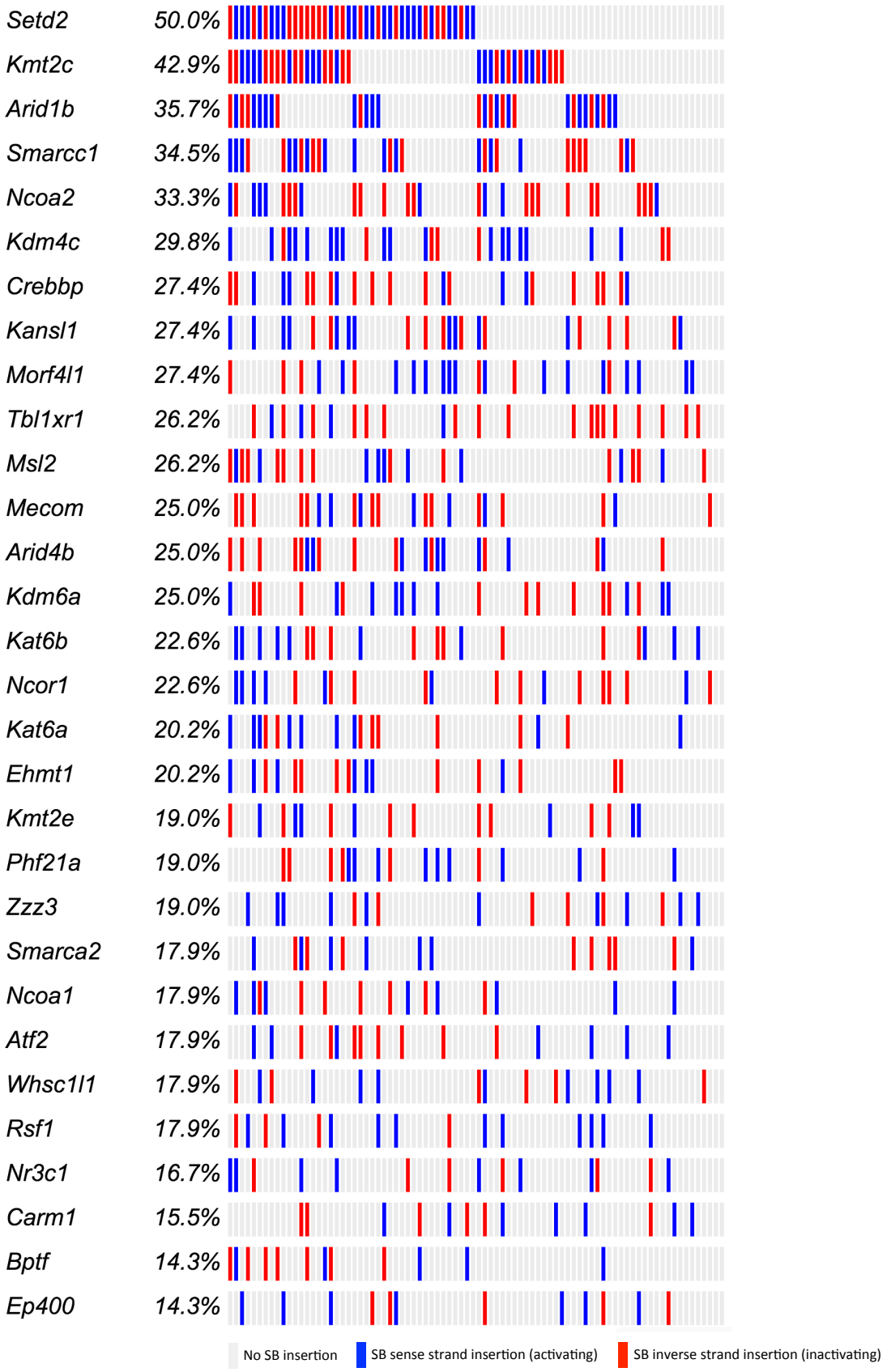
377

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)

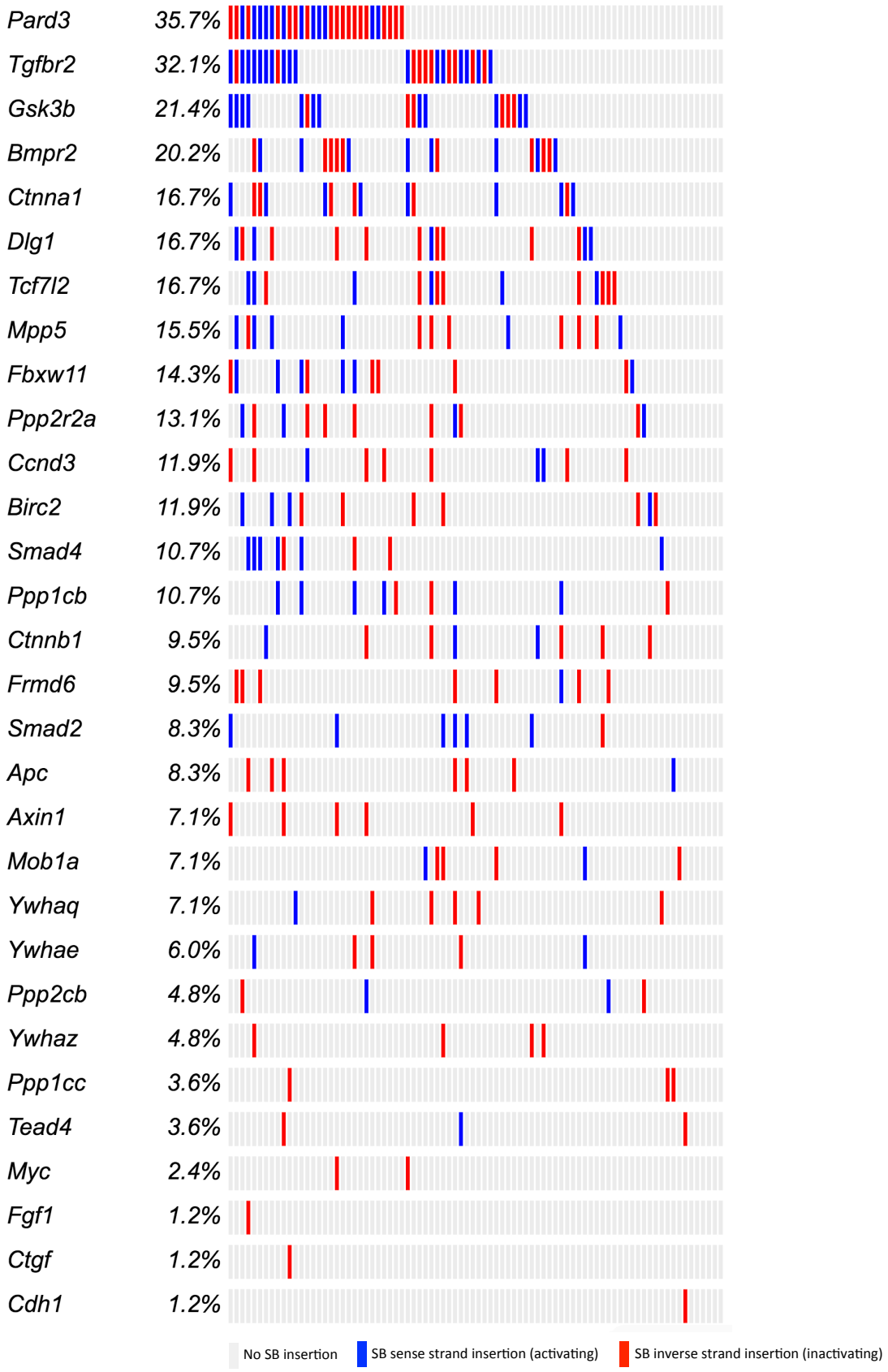
378

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)

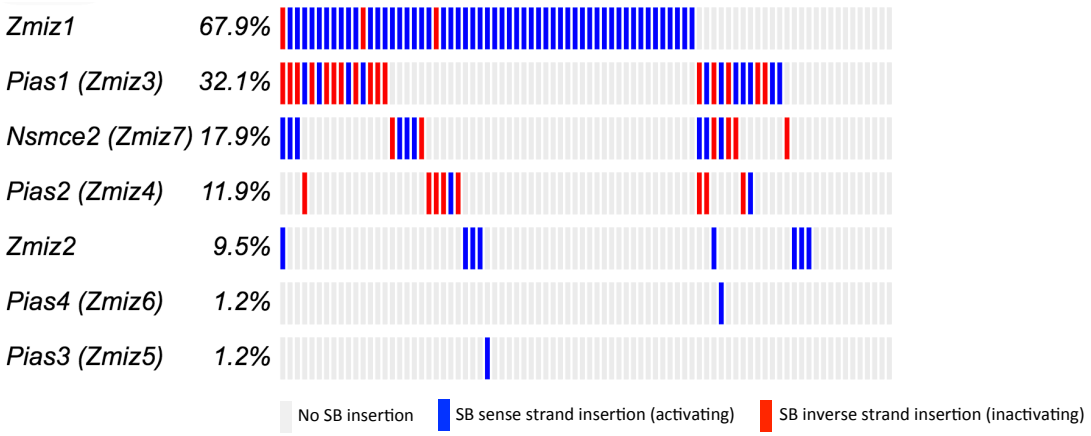
379

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)

380

i

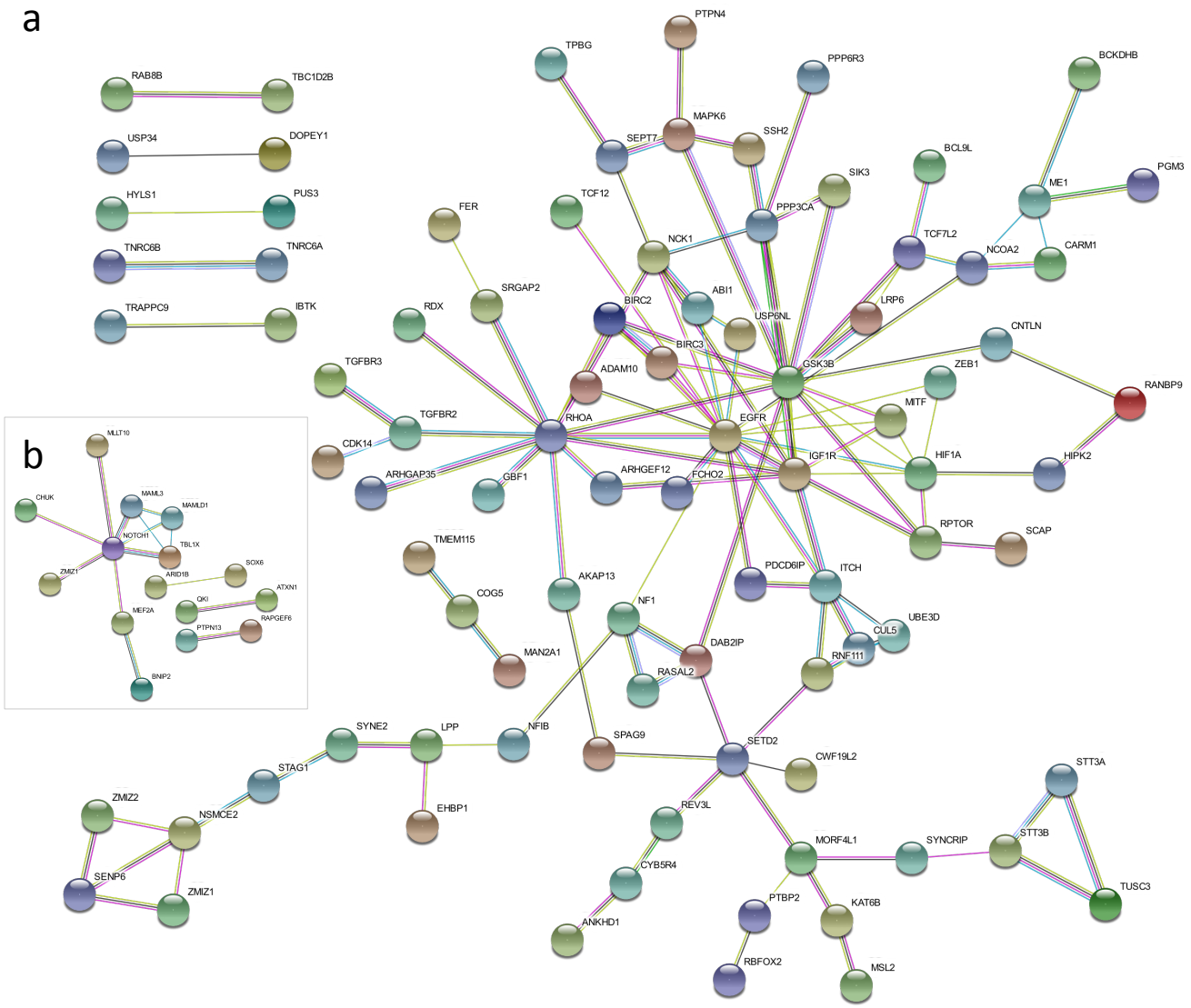


Comparative oncogenomic meta-analysis

One hundred-seven of the SB candidate driver genes are direct orthologs of human cancer genes found in the Cancer Gene Census (GRCh38 COSMICv86) (Futreal et al., 2004), a growing catalogue of mutations causally implicated in cancer, which is an enrichment greater than expected by chance (χ^2=60.75 with Yates correction, P<0.0001; **Supplementary Tables 23–27**). Another 64 SB candidate driver genes, including *ZMIZ1*, *ZMIZ2*, and *KMT2C,* have human orthologs with recurrent non-silent mutations identified by exome sequencing (Lee et al., 2014; Pickering et al., 2014; South et al., 2014) of 68 cuSCC genomes (χ^2=23.70 with Yates correction, P<0.0001). We conclude from these findings that SB identifies the same genes mutated by the diversity of mutational processes operating within end-stage human cuSCC genomes.

To gain insight into the biological functions of the candidate discovery drivers, we used a curated subset of the Enrichr analysis platform (Chen et al., 2013; Kuleshov et al., 2016) gene-set libraries for pathways and gene ontologies (see **Methods**) to perform biological pathway and process enrichment analysis from the cutaneous specimen cohorts (cuSCC_Splink, cuKA_Splink, cuSK_SBCapSeq, cuKA_SBCapSeq, cuSCC_SBCapSeq). We discovered statistically significant enrichment of recurrent SB candidate drivers in many signaling pathways and biological processes previously associated with human SCC, including EGFR, MAPK, NOTCH, and WNT-TGFβ (**Supplementary Figure 9a-i**). The combined SB|cuSCC Trunk Driver network has significantly more known protein-protein interactions than expected by chance (P=8.55e-06, STRING enrichment analysis; number of nodes: 144, number of edges: 124, expected number of edges: 82; **Supplementary Figure 10a**). Similarly, the combined SB|cuKA Trunk Driver network also has significantly more known protein-protein interactions than expected by chance (P=6.93e-05, STRING enrichment analysis; number of nodes: 37, number of edges: 14, expected number of edges: 4; **Supplementary Figure 10b**). Notably, we also observed that all SB|cuSCC tumor genomes have a mutation in at least one gene involved in either the chromatin modification (consisting of chromatin and histone modification enzymes) and/or Hippo pathway genes (**Fig. 4** and **Supplementary Tables 28**). Comparative genomic integration of the SB defined cuSCC and cuKA drivers with human cuSCC genomes reveals that while differences in the incidence of individual genes across cohorts can vary substantially, the biological pathways are conserved (**Supplementary**

407 **Figure 9a**). These differences may be due in part to the types of alterations produced by SB transposition (gene
408 activation/inactivation) versus UV-IR (silent/non-silent mutations) or, alternatively, may reflect the complexity of
409 the systems of genetic networks (Nadeau and Dudley, 2011) that exists between driver genes and keratinocyte
410 transformation phenotypes across species.



411

412 **Supplementary Figure 10: Truncal driver STRING network analysis.** (**a**) The SB|cuSCC Trunk Driver (n=84) network has
413 significantly more known protein-protein interactions than expected by chance (P=8.55e-06, STRING enrichment analysis;
414 number of nodes: 144, number of edges: 124, expected number of edges: 82 (**b**) The SB|cuKA Trunk Driver (n=62) network also
415 has significantly more known protein-protein interactions than expected by chance (P=6.93e-05, STRING enrichment analysis;
416 number of nodes: 37, number of edges: 14, expected number of edges: 4.

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

418 The most highly mutated oncogenic drivers in cuSCC genomes were the transcriptional co-factors *Zmiz1* and *Zmiz2*,
419 which are collectively and mutually exclusively mutated in 71% of end-stage SB|cuSCC genomes. High read depth
420 insertions are clustered into a single intron of each gene and occur in the same transcriptional orientation, where

421 they are predicted to drive expression of a truncated chimeric transcript containing the oncogenic MIZ domain (**Fig.**
422 **5a**). We confirmed this prediction by RT-PCR (not shown) and normalized microarray data (**Fig. 5b**) analysis, where
423 cuSCC tumors with *Zmiz1* insertions were found to have significantly higher mRNA abundance than tumors without
424 *Zmiz1* insertions (**Fig. 5b**). We also observed a positive correlation between high-read-depth SB insertions among
425 the 11 tumors harboring truncal *Zmiz1* insertions and *Zmiz1* gene expression (**Fig. 5c**).

426 To identify genes that were differentially expressed between SB cuSCC tumors with and without insertions in *Zmiz1*,
427 we used the normalized microarray data (see **Supplementary Table 29-30**), and generated a 25- (**Fig. 6d** and
428 **Supplementary Fig. 11a**), 50- (**Supplementary Fig. 11b**), and 96-gene (**Supplementary Table 31**) differential
429 expression signature ($P<0.05$, $P_{FDR\text{-}corrected}<0.12$) from the top differentially expressed genes. These genes
430 demonstrated enrichment for biological pathways and processes related to gene expression, RNA processing and
431 metabolism (**Supplementary Table 32-33**). Among the differentially expressed genes, the expression levels of
432 *Serpinb2*, *Spink12*, and *Stfa3* were individually statistically significantly higher in cuSCC genomes with activating
433 *Zmiz1* insertions than in cuSCC genomes without *Zmiz1* insertions (**Supplementary Fig. 11c** and **Supplementary**
434 **Table 34**), and *Mtdh*, *Anp32e*, *Ecd*, *Dcun1d5*, and *Kdm7a* were also independently identified as cuSCC Progression
435 Drivers. These statistically defined co-driver expression relationships indicate that these cooperating mutagenic
436 insertions may drive keratinocyte transformation and/or progression in this mouse model of human cuSCC.

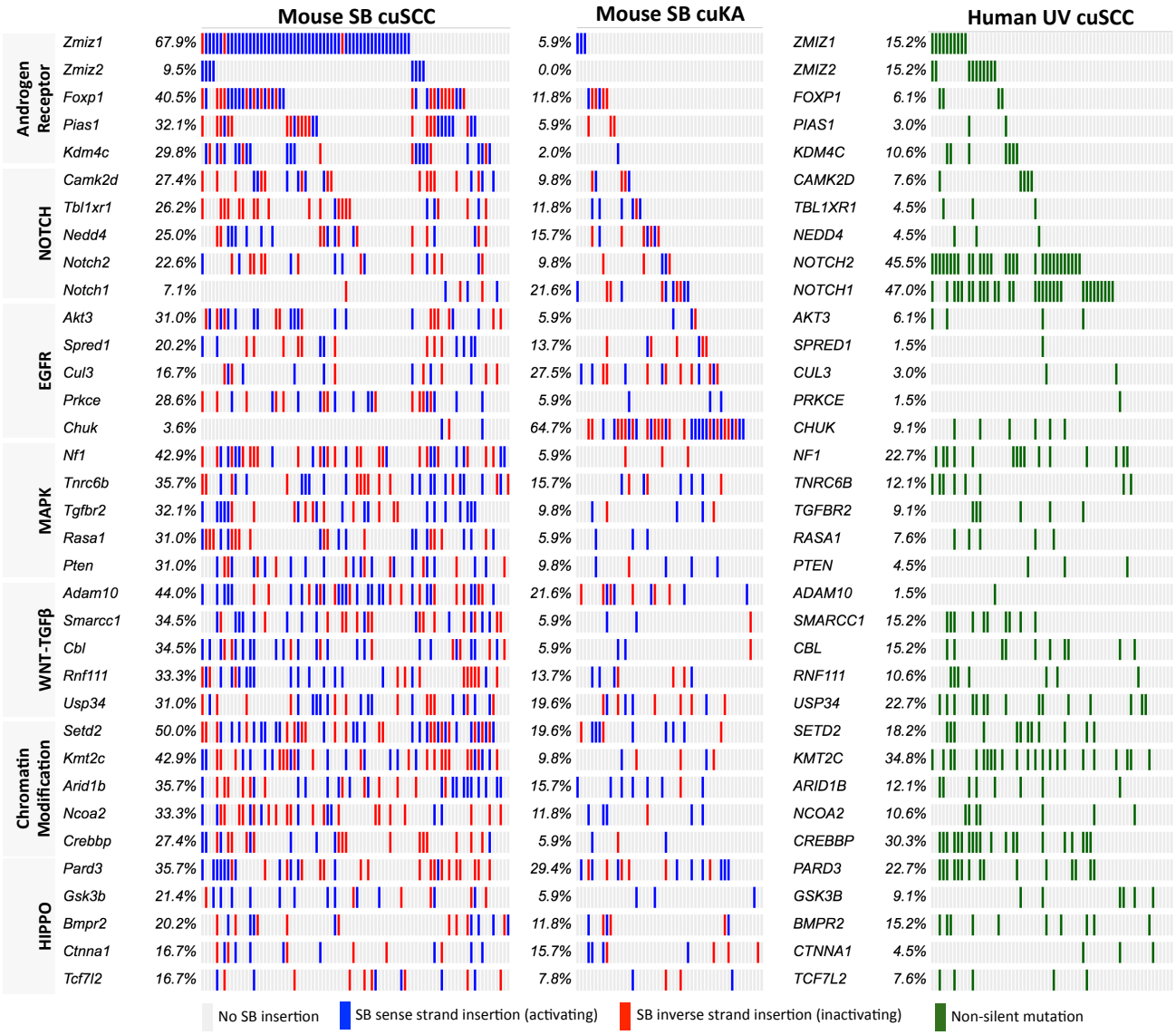


Figure 4: Comparative genomic landscape of candidate driver mutations and common altered pathways in mouse and human cuSCC genomes. Integrated Oncoprints sorted by biological pathway or process discovery significance within the SB|cuSCC dataset (**Supplementary Table 28**) and displaying the mutation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

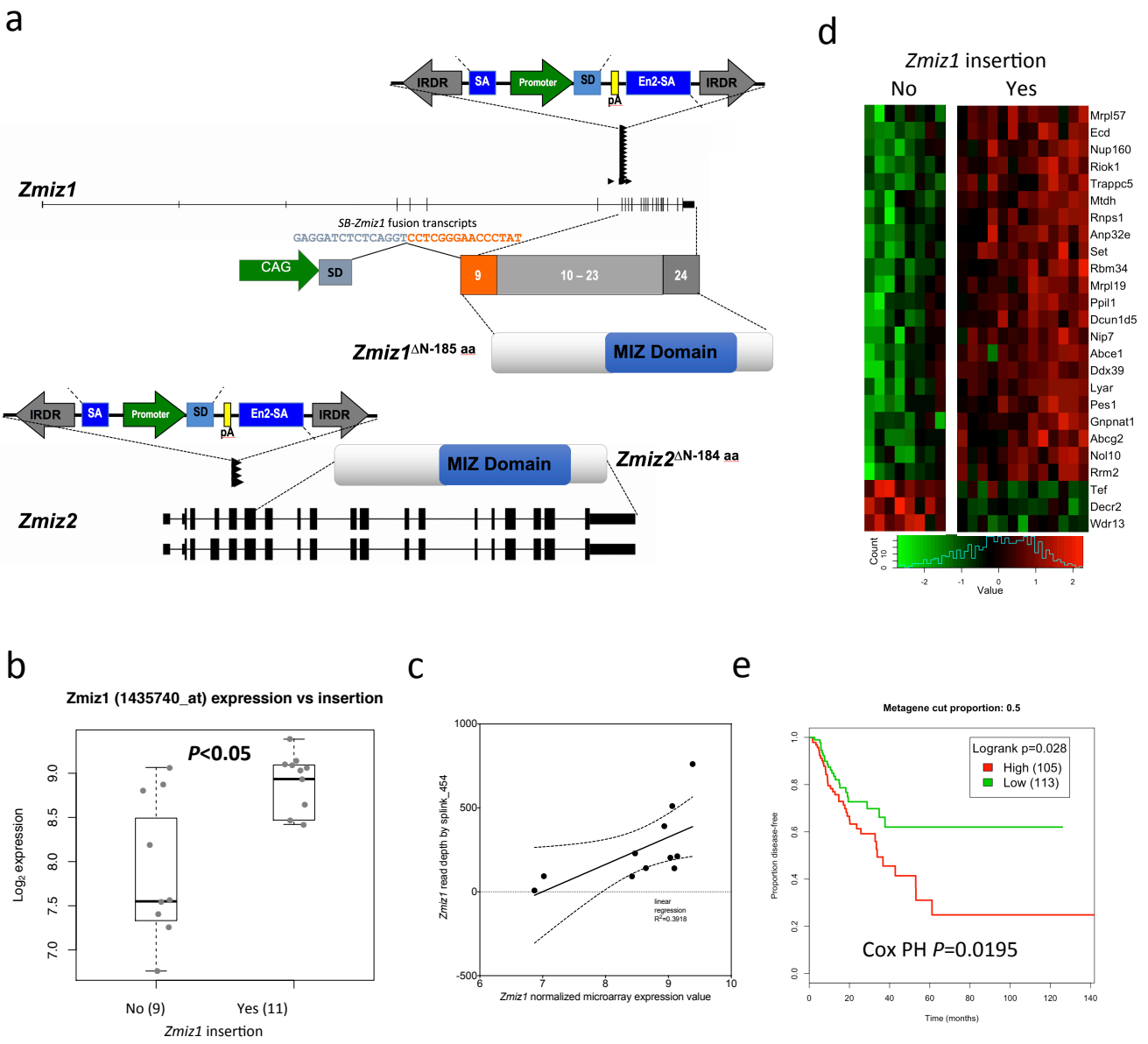
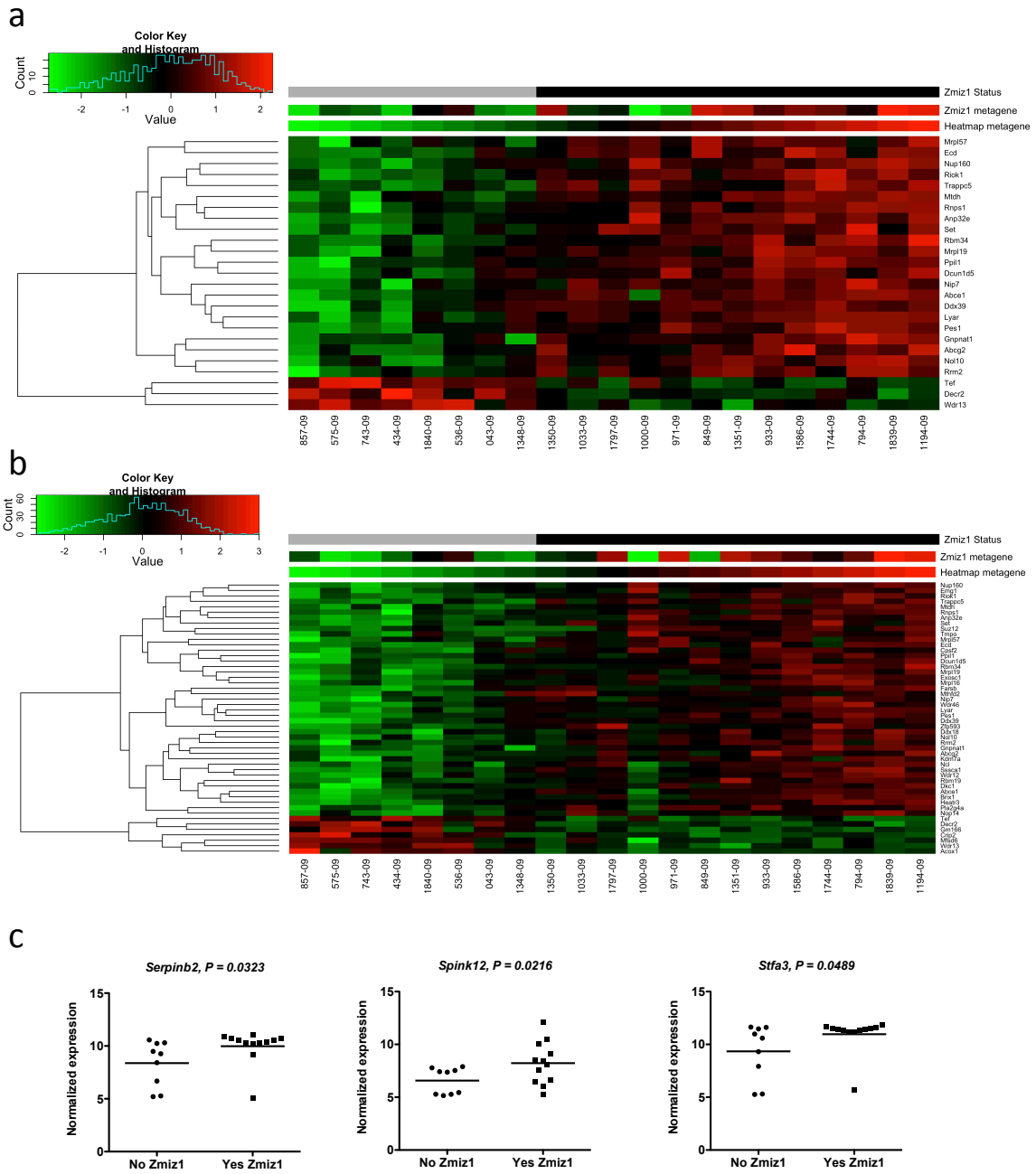


Figure 5: SB-mediated mutagenesis promotes N-terminal truncated ZMIZ paralog activation during cuSCC progression in mice. (**a**) Recurrent TA-dinucleotide SB insertion events from individual cuSCC genomes reveal activating sense strand SB insertions into Zmiz paralogs Zmiz1 and Zmiz2. Chimeric fusion 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**).

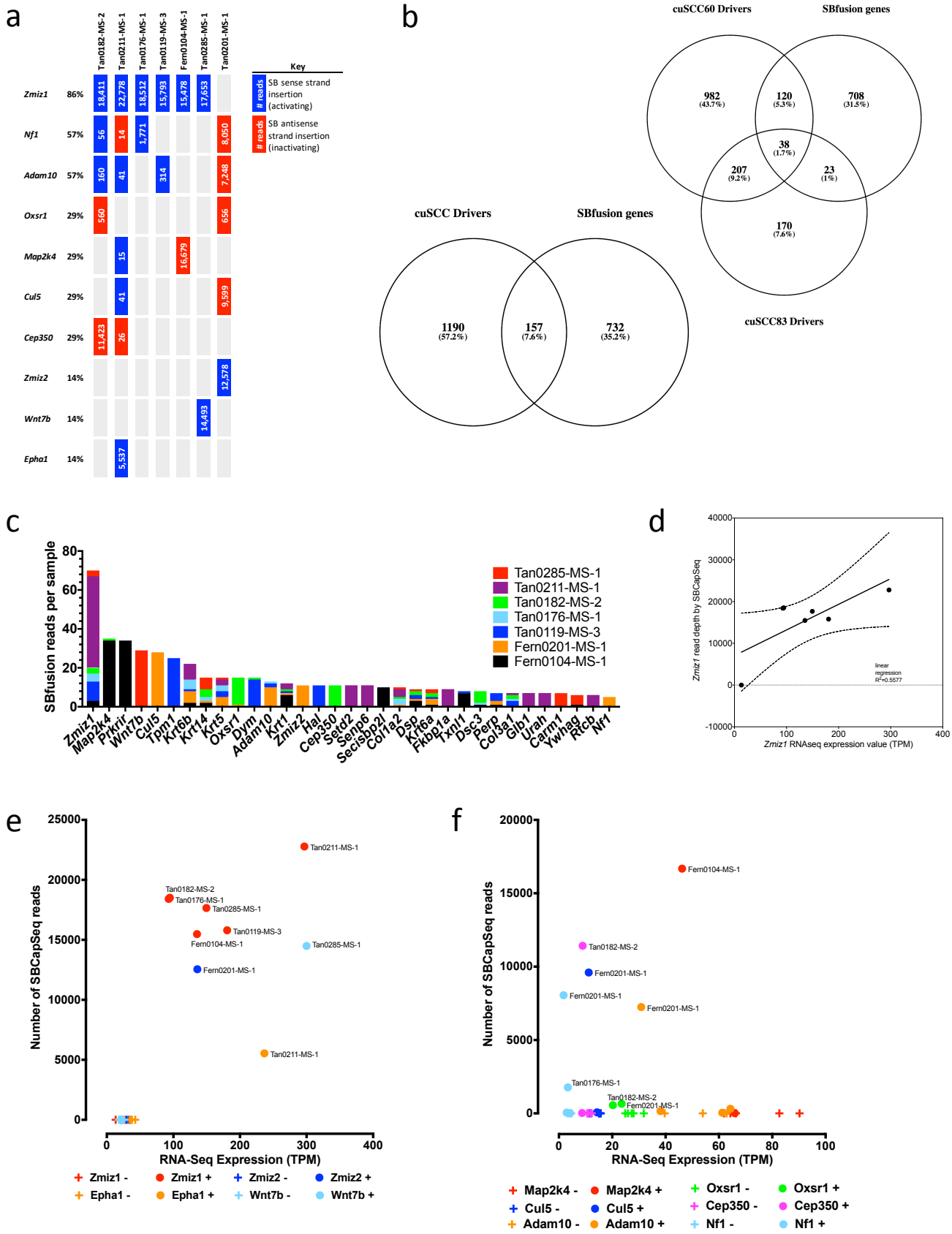


Supplementary Figure 11: SB effect on gene expression by microarray analysis. (**a**) Top 25 DE genes: *Mtdh*, *Anp32e*, *Ecd* are cuSCC progression drivers. (**b**) Top 50 DE genes: *Mtdh*, *Anp32e*, *Ecd*, *Dcun1d5*, and *Kdm7a* are cuSCC progression drivers. (**c**) Differentially expressed genes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

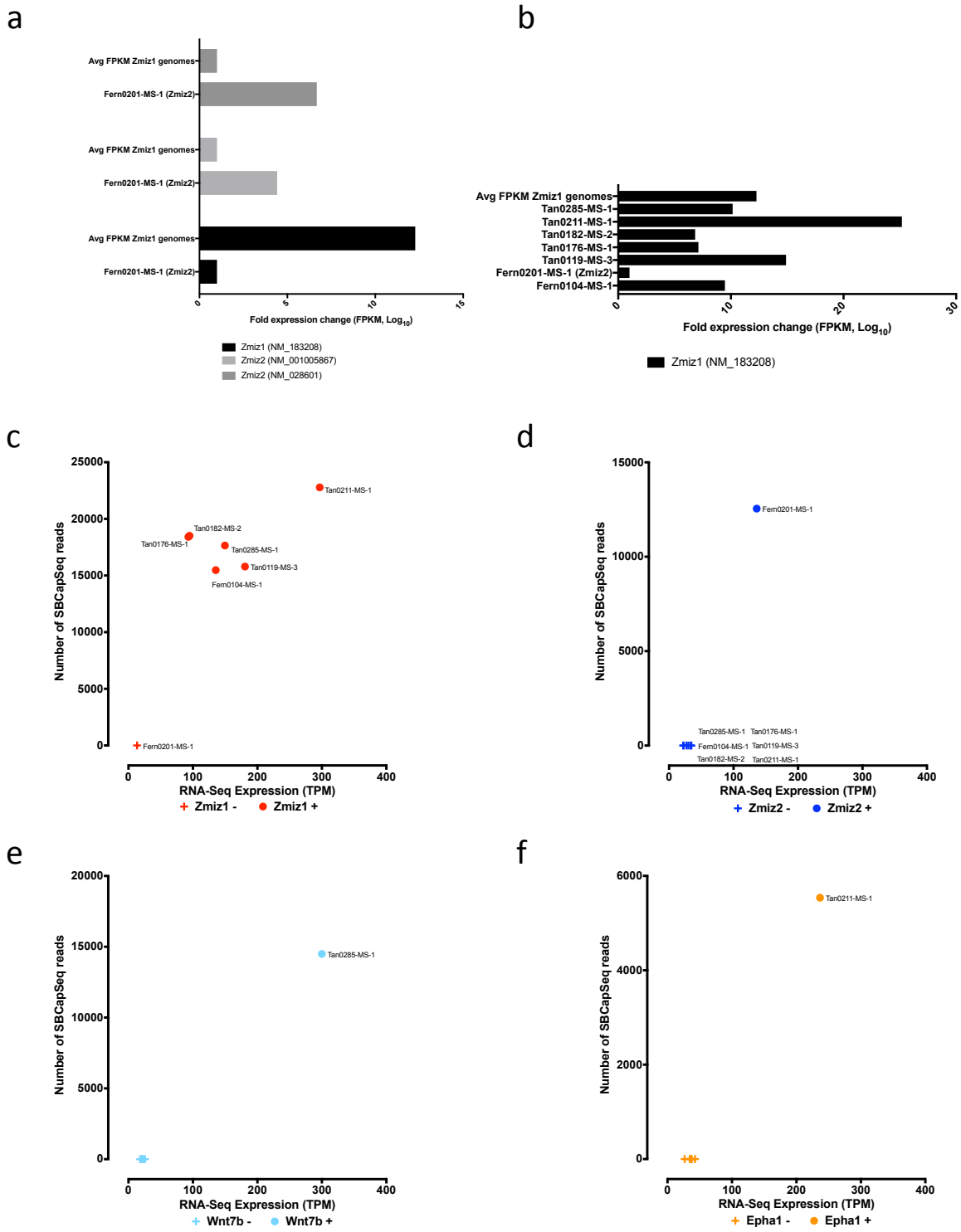
To further analyze the impact of clonally selected truncal SB insertions on gene-level and transcript-level expression, we sequenced ribo-depleted, whole-transcriptome RNA-seq (wtRNA-seq) libraries created from total RNA isolated from 7 cuSCC cuSCC genomes (**Supplementary Table 35**) with high-read depth activating SB insertions into either *Zmiz1* (n=6) or *Zmiz2* (n=1) to identify SBfusion reads (Mann et al., 2016). The wtRNA-seq reads contained spliced chimeric transposon splice-donor or splice-acceptor sequences fused to adjacent known

467 exons in the mouse genome. In total, we identified 2,906 SBfusion wtRNA-seq reads from 889 unique RefSeq genes
468 (**Supplementary Table 35**), including 154 progression and 27 truncal cuSCC drivers, respectively (**Supplementary**
469 **Table 37-38**), which is substantially more than expected by chance (Chi-square test with Yates correction, χ^2= 159,
470 P<0.0001). Thirty-three genes with SBfusion events supported by ≥5 SBfusion reads, include our top trunk drivers
471 *Zmiz1* and *Zmiz2* and several keratins (**Fig. 6a-c**). We also observed a positive correlation between high-read-depth
472 SB insertions among the 6 cuSCC tumors harboring truncal *Zmiz1* insertions and exhibiting high *Zmiz1* gene
473 expression (**Fig. 6d**). We next selected 10 genes with high-read depth SBCapSeq insertion sites and SBfusion reads
474 detected by RNA-seq within the 7 cuSCCs to investigate potential alterations in gene expression due to clonal SB
475 insertion events, including 4 activating (*Zmiz1*, *Zmiz2*, *Wnt7b*, and *Epha1*) and 6 inactivating (*Nf1*, *Adam10*, *Oxsr1*,
476 *Map2k4*, *Cul5*, and *Cep350*) SB insertion events (**Fig. 6e-f**, **Supplementary Figure 12a-f**, and **Supplementary**
477 **Figure 13a-f**).

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

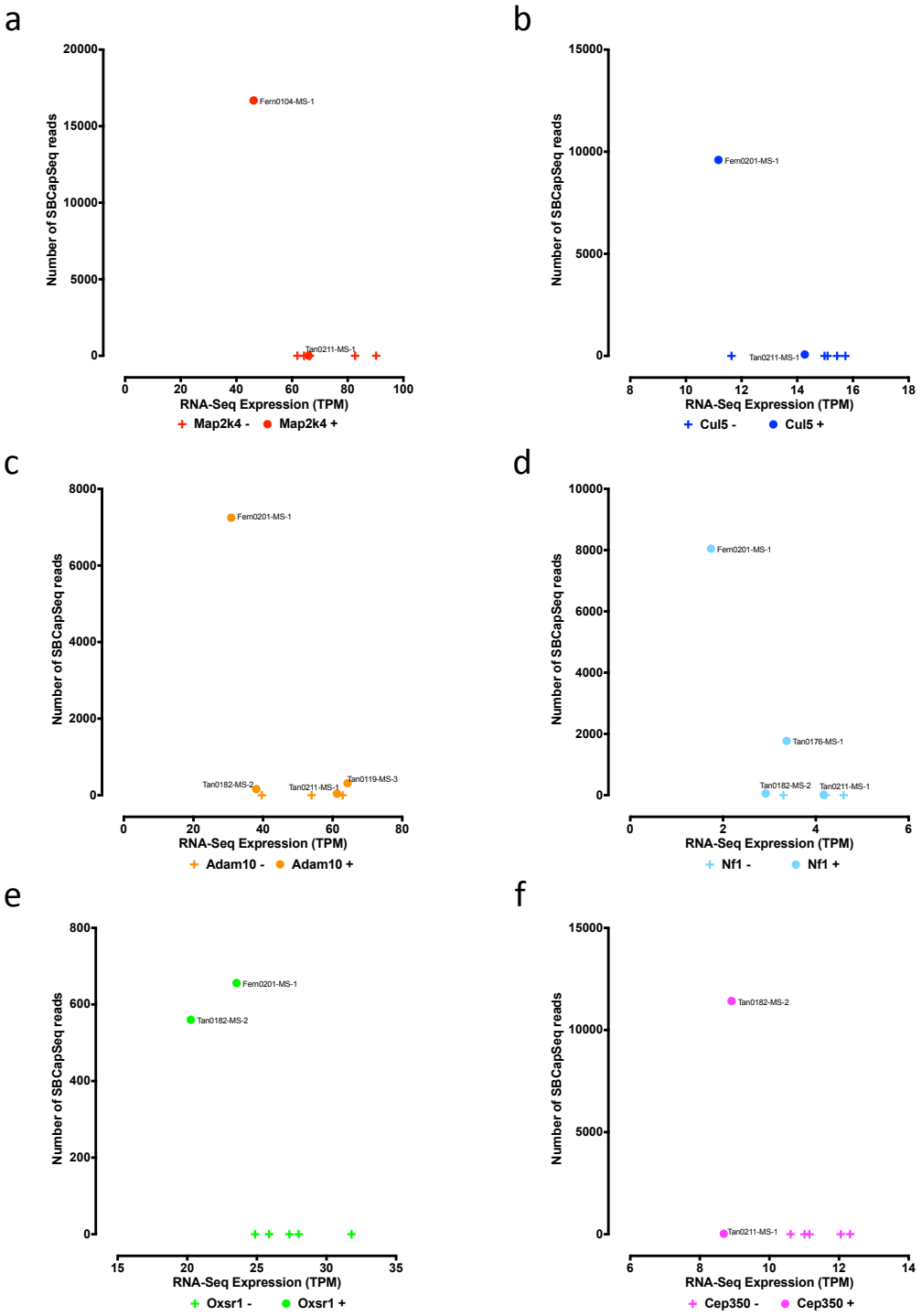

a
Tan0182-MS-2
Tan0211-MS-1
Tan0176-MS-1
Tan0119-MS-3
Fern0104-MS-1
Tan0285-MS-1
Tan0201-MS-1
Zmiz1 86% 18,411 22,778 18,512 15,793 15,478 17,653
Nf1 57% 56 14 1,771 8,050
Adam10 57% 160 41 314 7,248
Oxsr1 29% 560 656
Map2k4 29% 15 16,679
Cul5 29% 41 9,599
Cep350 29% 11,423 26
Zmiz2 14% 12,578
Wnt7b 14% 14,493
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 Serd2 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 +

490



Supplementary Figure 12: 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.

501



502

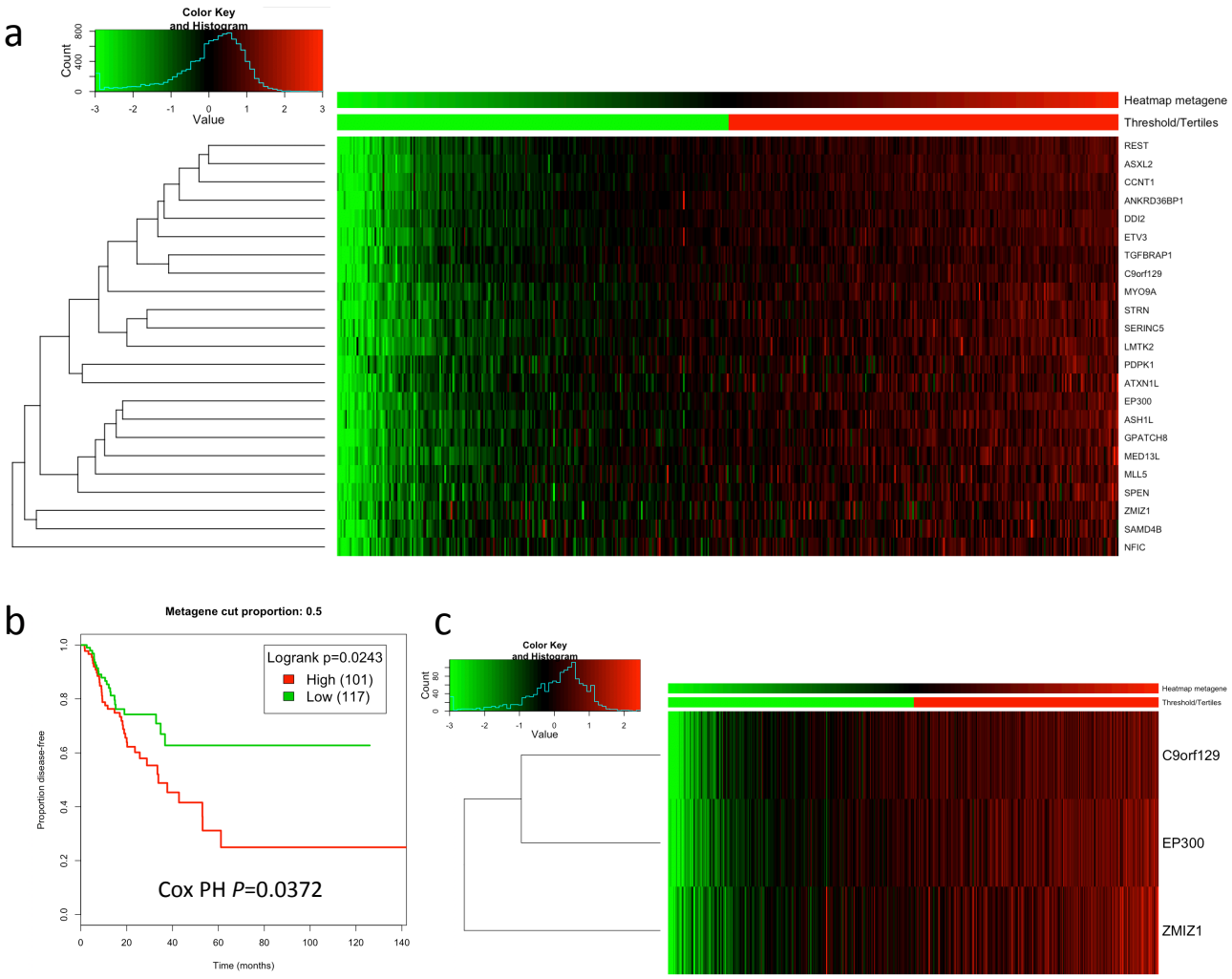
503 **Supplementary Figure 13: Clonally selected SB insertions affect trunk driver genes by inactivating expression in SB-**
504 **cuSCC genomes.** (**a-f**) Individual panels from **Figure 5f**: reduced gene expression in cuSCC masses with (**+**) high read depth
505 inactivating SB insertion events among 6 candidate tumor suppressor drivers compared with normal gene expression levels in
506 cuSCC tumors without (**–**) SB insertions.

507 Importantly, in each tumor, SBfusion reads containing either *SB-Zmiz1* or *SB-Zmiz2* spliced-exons were represented
508 in the top 2% of SBfusion transcripts and were supported by multiple independent SBfusion reads (**Supplementary**
509 **Table 39-40**). *Zmiz1* or *Zmiz2* gene expression was significantly increased within each of the 7 cuSCC genomes
510 (≧100 and up to 300 times normal expression levels) and correlated with a highest SBCapSeq read depth sites with
511 activating SB insertions and activating SBfusion reads (**Figure 6e** and **Supplementary Figure 12a-d**). This strongly
512 suggests that the SB insertion events driving either *Zmiz1* or *Zmiz2* have been clonally selected to increase
513 oncogenic expression of the N-terminal truncated forms during keratinocyte transformation and cuSCC progression
514 *in vivo*. Similarly, *Epha1* and *Wnt7b* had high SBCapSeq read depth activating SB insertions and wtRNA-seq SBfusion
515 reads and significantly higher gene expression compared to basal expression levels (≧100 times) observed in
516 tumors without insertions or SBfusion reads (**Figure 6e** and **Supplementary Figure 12e-f**). Since SB insertions into
517 these genes were not recurrent at the population level, *Epha1* and *Wnt7b* were not identified as statistically
518 significant drivers in cuSCC. However, the integrated genomic and transcriptomic analysis strongly suggests that
519 each of these genes may have a 'private' role in driving tumor progression in the genomes in which they occur with
520 elevated expression levels. Both genes encode proteins that are linked to cancer hallmarks — *Epha1* is an ephrin
521 receptor subfamily of the protein-tyrosine kinase family (Potla et al., 2002) and *Wnt7b* is a WNT family member
522 (Huguet et al., 1994). For the 6 predicted inactivated genes, we observed reduced gene expression that correlated
523 with high SBCapSeq read depth of inactivating SB insertions and truncated SBfusion reads (**Figure 6f** and
524 **Supplementary Figures 21a-f**). *Adam10*, *Cul5*, and *Nf1* are Trunk Drivers and *Map2k4* is a Progression Driver in
525 cuSCC cohorts, further suggesting these events contribute to cuSCC progression *in vivo*. *Oxsr1* and *Cep350* were
526 private selected events in individual tumors that may operate in cuSCC progression within the tumors that harbor
527 clonal insertion events. Taken together, these data demonstrate that integrated genomic and transcriptome analyses
528 may provide a means to discover significant cooperating drivers in cuSCC genomes from bulk specimens.

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

530 The mammalian paralogous genes *Zmiz1* and *Zmiz2* are mutually exclusive trunk drivers activated in three quarters
531 of SB-driven cuSCC genomes in an SB-driven cuSCCs, suggesting that *Zmiz* proteins are proto-oncogenes in
532 cutaneous squamous keratinocyte transformation *in vivo*. *Zmiz1* and *Zmiz2* encode transcription factor proteins that
533 share core c-terminal features, including androgen receptor (AR) binding domain, proline-rich binding domain, and
534 MIZ-type zinc finger PIAS (protein inhibitor of activated STAT) domains, that coordinately act to both interact with
535 and regulate the activity of various cancer-associated proteins (Beliakoff and Sun, 2006; Huang et al., 2005; Kahyo et
536 al., 2001; Lee et al., 2013; Megidish et al., 2002; Schmidt and Muller, 2002, 2003; Sharma et al., 2003; Soler et al.,
537 2008). SB activating truncal insertions in *Zmiz1* or *Zmiz2* are unique to SB-induced cuSCC tumors and are predicted
538 to activate gene expression by driving expression of specific N-terminal truncated forms of ZMIZ1$^{\Delta N185}$ (Dupuy et al.,
539 2009) and ZMIZ2$^{\Delta N184}$, a closely related paralog (**Figure 5a**). The likelihood of achieving unique, sense-strand
540 oriented SB insertion events into two related genes at the same protein coding region on two different
541 chromosomes in dozens of tumors from independent mice by chance is incredibly small (P<0.0001), suggesting both
542 trunk driver ZMIZ oncoproteins contribute directly to keratinocyte initiation and tumor progression *in vivo*. The
543 ZMIZ1$^{\Delta N185}$ and ZMIZ2$^{\Delta N184}$ onco-proteins share 62% sequence identity, retain the zinc finger PIAS, proline-rich

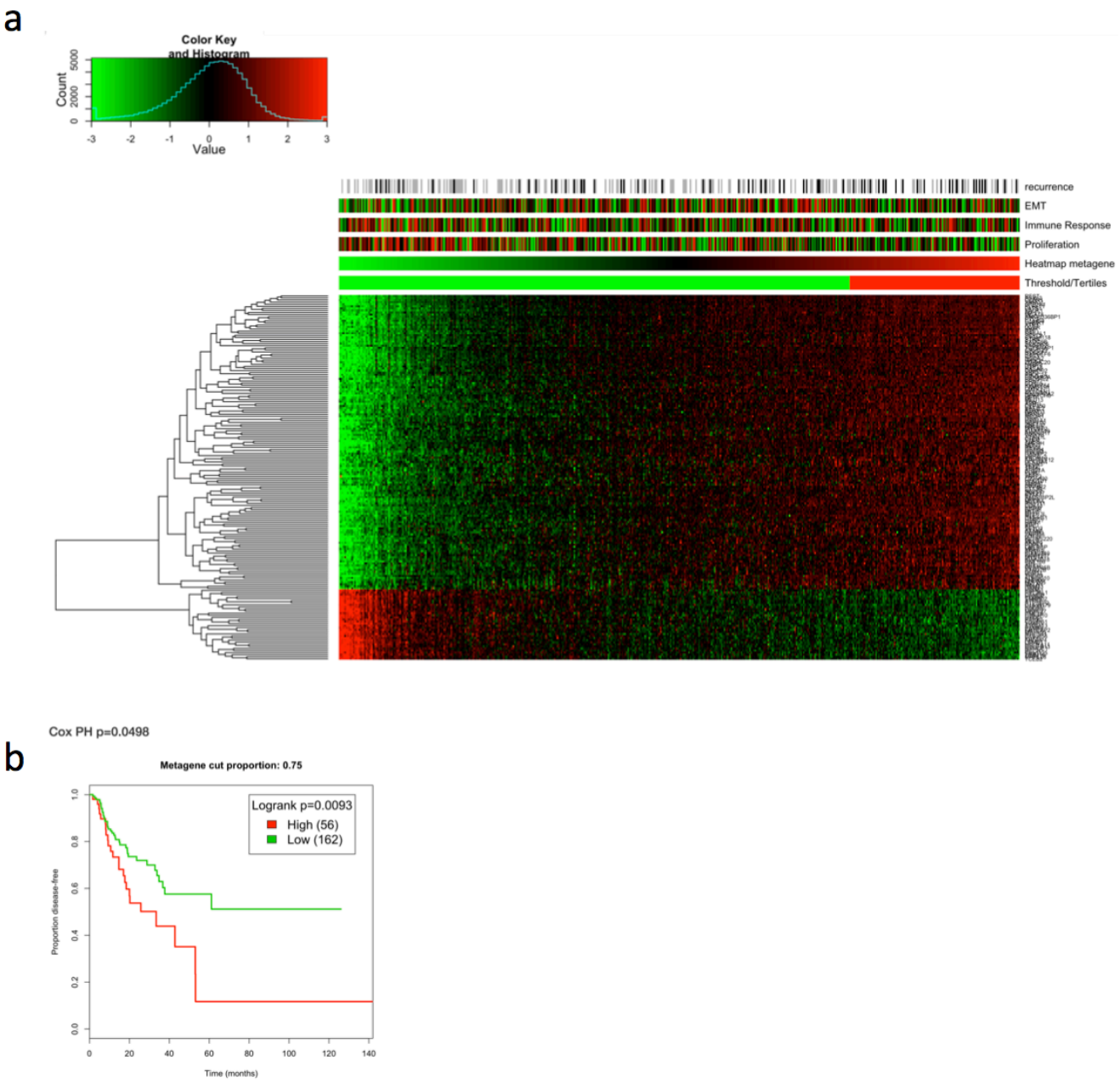
544 transactivation, and AR binding domains shared by both peptides. Importantly, both oncoproteins lack the N-
545 terminus in the full-length protein, which is thought to inhibit the intrinsic transcriptional activity of a C-terminal
546 proline-rich transactivation domain (Sharma et al., 2003). *ZMIZ1 (RAI17/ZIMP10)* and *ZMIZ2 (ZIMP7)* are
547 collectively mutated or amplified in one third of human cuSCC (Pickering et al., 2014; South et al., 2014) and 5% of
548 hnSCC (Cancer Genome Atlas, 2015) genomes. To investigate their oncogenic potential in human keratinocyte
549 transformation, we performed a metagene expression analysis of *ZMIZ1* and *ZMIZ2* in human TCGA head and neck
550 SCC (hnSCC) datasets for which gene expression (RNA-seq) and patient survival outcomes are available. A *ZMIZ1*
551 metagene, with a correlation threshold of 0.65, identifies a 23-gene signature that links high expression with poor
552 outcomes in human hnSCC patients (Cox Proportional Hazards Regression, *P*=0.0195; Log-rank Test, *P*=0.028 at
553 50% quintile; **Figure 5e; Supplementary Figure 14a–c**).



554

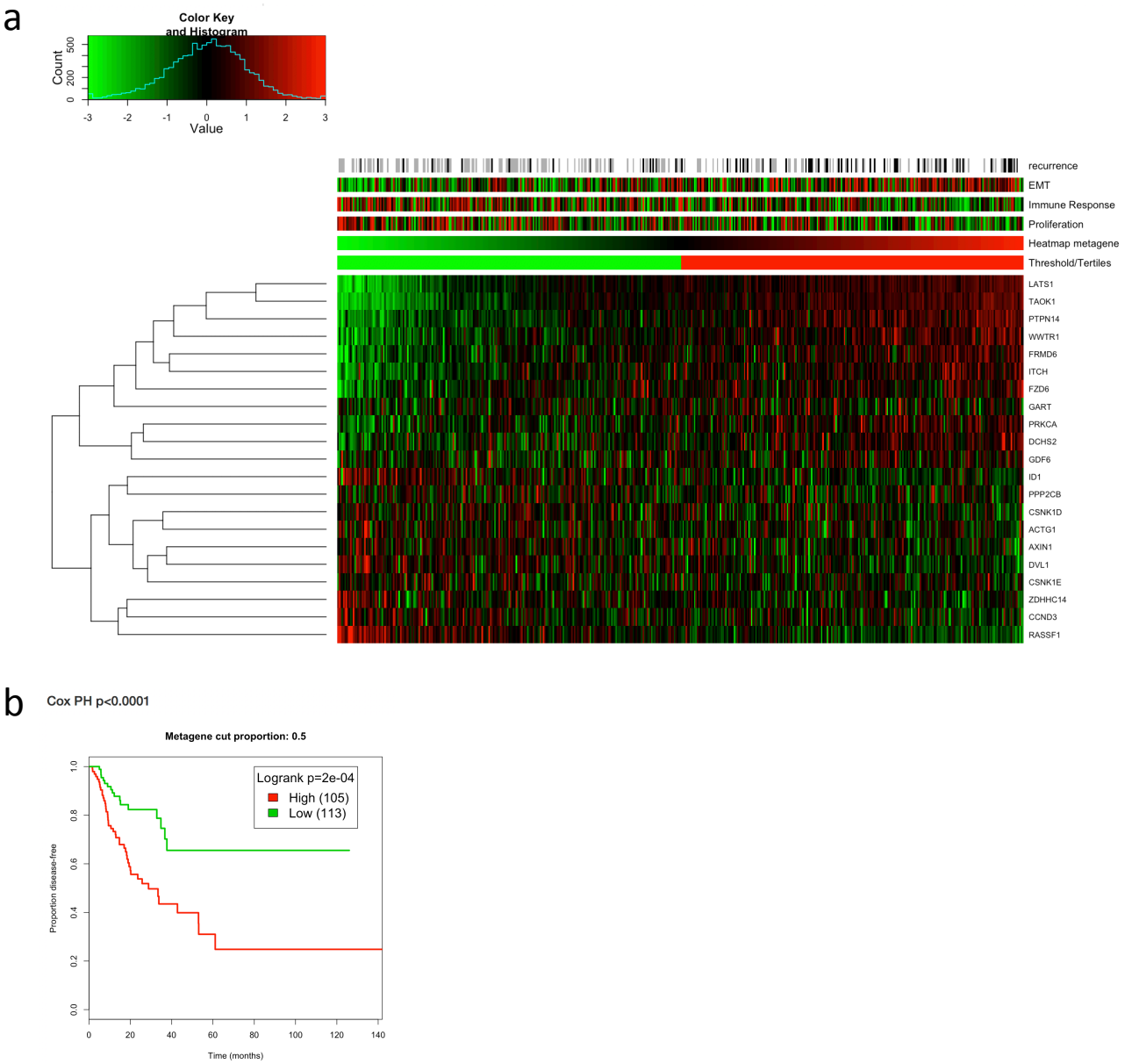
555 **Supplementary Figure 14: *ZMIZ1* metagene within the TCGA Head & Neck Squamous Cell Carcinoma (hnSCC) RNA-seq**
556 **dataset.** (**a**) *ZMIZ1* metagene heatmap constructed using singular value decomposition of human hnSCC RNA-seq dataset from
557 TCGA consisting of a 23-gene signature from **Figure 5e**. (**b**) Survival plot of patient outcomes in human hnSCC (**b**) based on the
558 expression of a *ZMIZ1*–centric metagene heatmap (**c**) constructed using singular value decomposition of human hnSCC RNA-seq
559 dataset from TCGA, reveal poor patient survival outcomes predicted by increasing expression of *ZMIZ1* and closely regulated
560 genes consisting of a 3-gene signature: *ZMIZ1*, *EP300*, and *C9orf129*.

561 Ectopic over-expression of N-terminus truncated Zmiz1 in mouse keratinocytes leads to epithelial hyper-
562 proliferation but is not sufficient for progression to late-stage cuSCC (Rogers et al., 2013), suggesting that
563 cooperating mutations are required for frank cuSCC progression *in vivo*. Biological pathway enrichment in the *SB*
564 cuSCC genomes revealed several cancer-associated pathways, including the EGFR, NOTCH, and WNT signaling
565 pathways. However, we noted that all *Zmiz*-mutated cuSCC genomes contained one or more inactivating SB
566 mutations in chromatin-remodeling genes (**Supplementary Figure 9g**), suggesting this class of mutated tumor
567 suppressor genes may contribute significantly in driving cuSCC progression. Four of the top-ranked chromatin
568 remodeling trunk driver orthologs were also found to be significantly mutated in >15% of human cuSCC genomes
569 (Pickering et al., 2014; South et al., 2014). Importantly, as observed in mouse tumors, mutations in the orthologs
570 *KMT2C (MLL3)*, *ARID1B (BAF250B)*, *NCOA2 (KAT13C)*, and *CREBBP (CBP/KAT3A)* were also found in the same cuSCC
571 genomes harboring *ZMIZ* mutations (**Figure 2**), suggesting they might cooperate during human keratinocyte
572 transformation. Previously, *KMT2C* (*MLL3*) mutations were found to be associated with poor outcomes in aggressive
573 cuSCC (Pickering et al., 2014), suggesting mutations in this gene may be a biomarker for advanced stage disease. To
574 extend this result, we used survival information and RNA-seq expression from a hnSCC dataset, and observed that a
575 *KMT2C* metagene, with a correlation threshold of 0.66, identifies a 176-gene signature that links high expression
576 with poor outcomes in human hnSCC patients (Cox Proportional Hazards Regression, P=0.0498; Log-rank Test,
577 P=0.0093 at 75% quintile; **Supplementary Figure 15**). Similarly, consistent with hippo signaling in cuSCC, we
578 identified a 20-gene signature consisting of Hippo pathway orthologs predictive of patient outcome (Cox
579 Proportional Hazards Regression, P<0.0001; Log-rank Test, P=0.0002 at 50% quintile; **Supplementary Figure 16**).



580

581 **Supplementary Figure 15: *KMT2C (MLL3)* metagene within the TCGA Head & Neck Squamous Cell Carcinoma (hnSCC)**
582 **RNA-seq dataset.** (**a**) *KMT2C* metagene heatmap and (**b**) patient survival outcomes.



Supplementary Figure 16: 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.

Chromatin modifying tumor suppressors drive human keratinocyte initiation

Chromatin remodeling is one of the most frequently inactivated process in SCC, and has been demonstrated to be involved in either tumor initiation or progression in multiple with cancer types (Andricovich et al., 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, chromatin remodeling was the top biological process that was significantly enriched (**Supplementary Table 28**). However, how alterations in expression levels of chromatin remodeling genes feature in cuSCC disease trajectory remains unclear. Thus, based on our shRNA screen, we prioritized three tumor suppressor genes involved in histone modification (*KMT2C* and *CREBBP*) or transcriptional regulation (*NCOA2*) to study their roles in cuSCC initiation or

594 progression. First, we confirmed expression of each gene within HaCaT cells (data not shown). Next, we used
595 shRNAs targeting these genes to assess cellular transformation *in vitro* as a surrogate of tumor initiation, and
596 disease progression *in vitro* and *in vivo* (**Fig. 7**, 1-factor ANOVA, $P<0.05$). To investigate whether knockdown of
597 *KMT2C*, *CREBBP* or *NCOA2* can transform keratinocytes, we used the immortalized, minimally transformed HaCaT
598 keratinocyte cell line (Boukamp et al., 1988), to stably knock down expression of target genes. Forced depletion of
599 *CREBBP* significantly increased cellular proliferation relative to control, and a similar trend was observed with
600 *KMT2C* knockdown (**Figure 7a**). No significant difference in proliferation was observed between *NCOA2* knockdown
601 and control (**Figure 7a**). To more stringently investigate transformation, we assessed anchorage-independent
602 growth, a hallmark of cellular transformation, using a soft agar colony formation assay. Knockdown of *CREBBP* or
603 *KMT2C* expression resulted in significantly larger size and multiplicity of colonies, relative to control (**Figure 7b**),
604 while no significant difference was observed between *NCOA2* knockdown and control (**Figure 7b**). Taken together,
605 these data suggest that knockdown of *CREBBP* or *KMT2C* can transform the immortalized HaCaT cell system *in vitro*.
606 Finally, we tested whether *KMT2C*, *CREBBP*, or *NCOA2* single gene knockdown in HaCaT cells can initiate formation
607 of masses *in vivo*. Cells were injected into flanks of immunodeficient *NSG* mice, and mass formation was monitored
608 for 9 weeks. At endpoint, no significant difference in tumor volume was observed between control and *shKMT2C,*
609 *shCREBBP* or *shNCOA2* (**Figure 7c-d**), suggesting that expression changes in more than one driver gene may be
610 needed for *in vivo* transformation.

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

612 To provide functional evidence that the SB-identified 12 candidate drivers are important for human SCC
613 progression, we selected 12 top-ranked SB cuSCC/cuKA driver orthologs (*CHUK*, *CUL1*, *EHMT1*, *FBXO11*, *KMT2C*,
614 *LARP4B*, *MAMLD1*, *NCOA2*, *NCOR2*, *NOTCH1*, *PTEN*, and *ZMIZ1*) for validation by using 3 siRNAs per gene
615 (**Supplementary Table 41**) to reduce RNA abundance in three well-characterized SCC cell lines, two skin (A431 and
616 SCC13) and one lung (SW900), which each harbor *TP53* mutations. We included 10 predicted tumor suppressor
617 drivers, *CHUK*, *CUL1*, *EHMT1*, *FBXO11*, *KMT2C*, *LARP4B*, *NCOA2*, *NCOR2*, *NOTCH1*, and *PTEN*, from the SB mouse skin
618 tumor datasets that are mutated in human cuSCC and commonly identified as drivers in other SB mutagenesis
619 screens reported in the SBCDDB (Newberg et al., 2018c) and 2 predicted oncogene drivers, *ZMIZ1* and *MAMLD1*, that
620 are mutated in human cuSCC but not identified as drivers in other SB mutagenesis screens reported in the SBCDDB
621 (Newberg et al., 2018c). First, we verified endogenous expression of the driver targets in each cell line and that
622 significant knockdown of RNA abundance was achieved in the target genes for each siRNA (**Supplementary Figure**
623 **17**). Next, we assessed the proliferation rates of cell lines with transient knockdown in target gene expression. All 12
624 drivers tested had at least one siRNA that significantly altered cellular proliferation compared to a scrambled
625 control siRNA from at least one cell line (one-factor ANOVA with multiple testing correction, FDR adjusted-$P<0.1$;
626 **Supplementary Table 42**). We observed a modest but significant decrease in cellular proliferation with ZMIZ1
627 depletion and increased cell proliferation with depletion of the predicted tumor suppressor driver candidates,
628 supporting the role of these genes in human SCCs (**Supplementary Table 42**). si*CHUK*, si*NOTCH1*, and si*PTEN* were
629 significant in all three cell lines, *NCOA2* was significant in both cuSCC cell lines, and si*FBXO11*, si*KMT2C*, si*LARP4B*,
630 si*MAMLD1* and si*ZMIZ1* were significant in one cuSCC and the lung cell line. si*CUL1*, si*EHMT1*, si*NCOR2* were

631 significant only in the SW900 lung cell line. si*PTEN* knockdown increased proliferation in A431 cells, consistent with
632 previously reported data (ref: Leonard et al., 2011). Additionally, 3 independent siRNAs targeting *ZMIZ2*, an
633 independent driver and paralog of *ZMIZ1*, were tested and found to grow significantly slower compared to a
634 scrambled control siRNA in the A431 cell line (**Supplementary Figure 17**).

635 Next, to both confirm and extend the siRNA studies and assess whether a selected panel of prioritized candidate
636 proto-oncogene (*ZMIZ1*, *ZMIZ2*, and *MAMLD1*) or tumor suppressor (*CHUK, CUL1, EHMT1, FBXO11*, *KMT2C*, *LARP4B,*
637 *NCOA2, NCOR2*, and *NOTCH1*) are involved in human cuSCC progression, we used lentivirus pools containing three
638 non-overlapping, short hairpin RNAs (sh) targeting the 13 candidate drivers and one non-targeting control
639 delivered to silence target gene expression in A431 cells (**Figure 8** and **Supplementary Figure 18**). Decreased
640 expression of target genes relative to NTC was verified by qPCR (**Figure 8a**). Seven of the drivers displayed
641 statistically significant changes in proliferation of A431 cells: *shZMIZ1*, *shZMIZ2*, and *shMAMLD1* decreased
642 proliferation; and *shKMTC2*, *shNCOA2, shFBXO11* and *shCHUK*, accelerated proliferation, consistent with their
643 predicted roles as proto-oncogenes and tumor suppressors, respectively (**Figure 8a-c**, two-factor ANOVA with
644 multiple comparisons, FDR-adjusted-$P<0.05$). While knockdown of the *ZMIZ* paralogs in the A431 cell line
645 demonstrated significant reduction in proliferation, we sought to verify that this observation was not cell-line
646 specific. Using the COLO16 cell line, we achieved stable knockdown of *ZMIZ* expression, and observed decreased
647 proliferation, relative to control. Taken together, our functional shRNA screen has yielded strong evidence for the
648 *ZMIZ* oncogenes identified from our SB screen to be involved in human cuSCC progression *in vitro*, and prioritized
649 several tumor suppressor genes for further characterization.

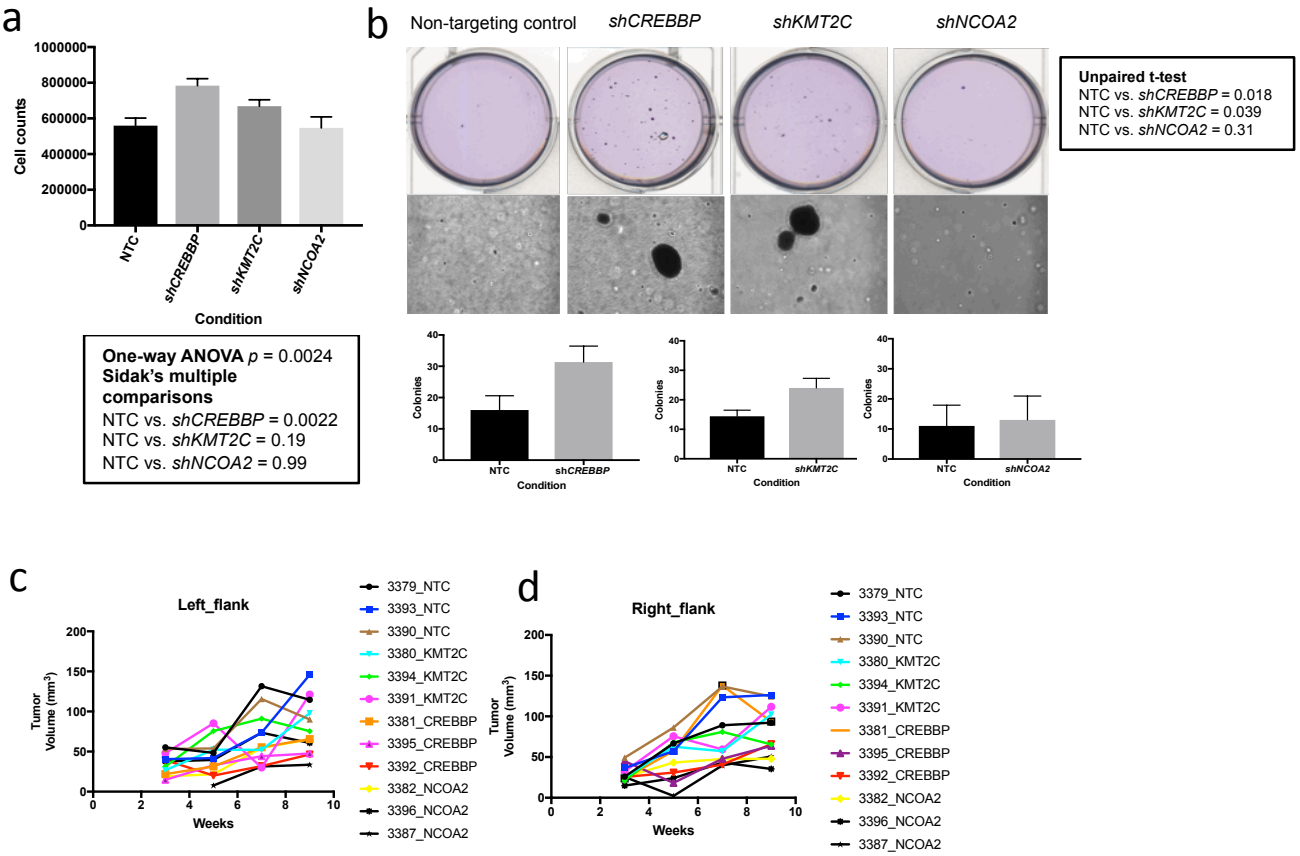
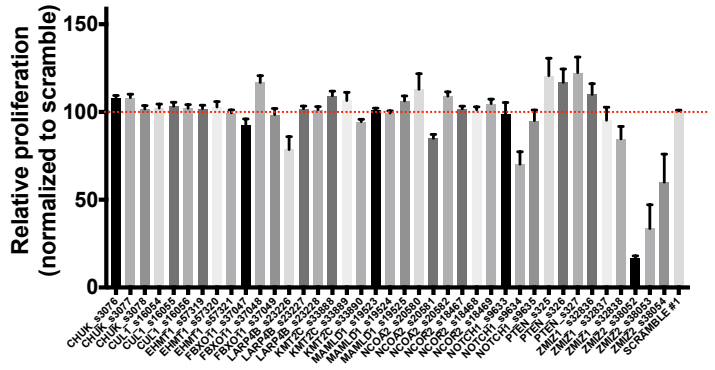
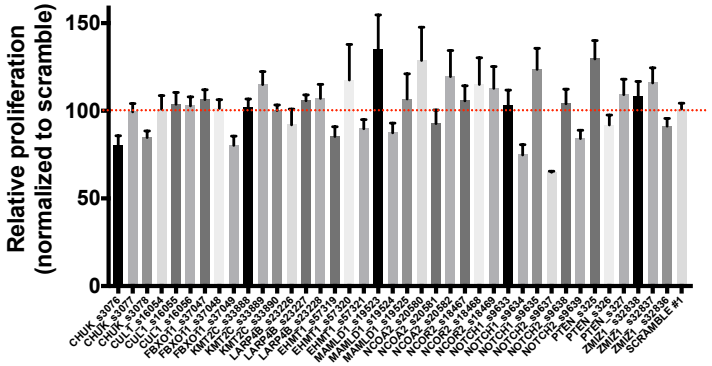


Figure 7: *In vitro* and *in vivo* functional validation of chromatin remodeler in immortal keratiocytes. (**a**-**b**) Assessment of keratinocyte transformation in response to knockdown of *CREBBP*, *KMT2C* and *NCOA2* by proliferation assay or anchorage-independent soft agar assay. For proliferation 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.

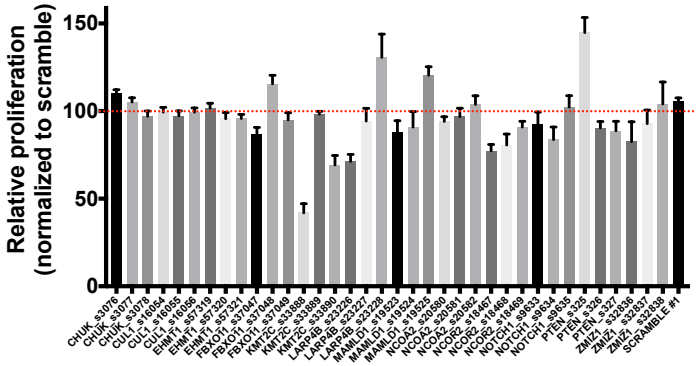
a



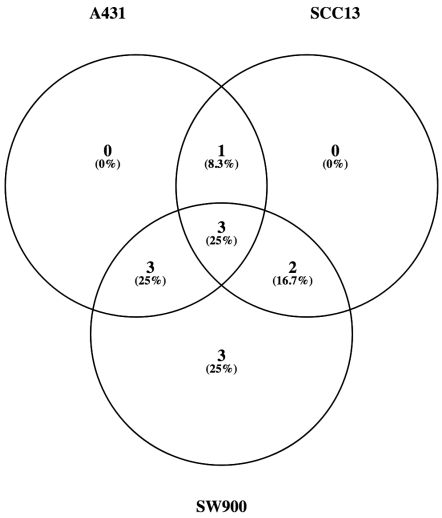
b



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*

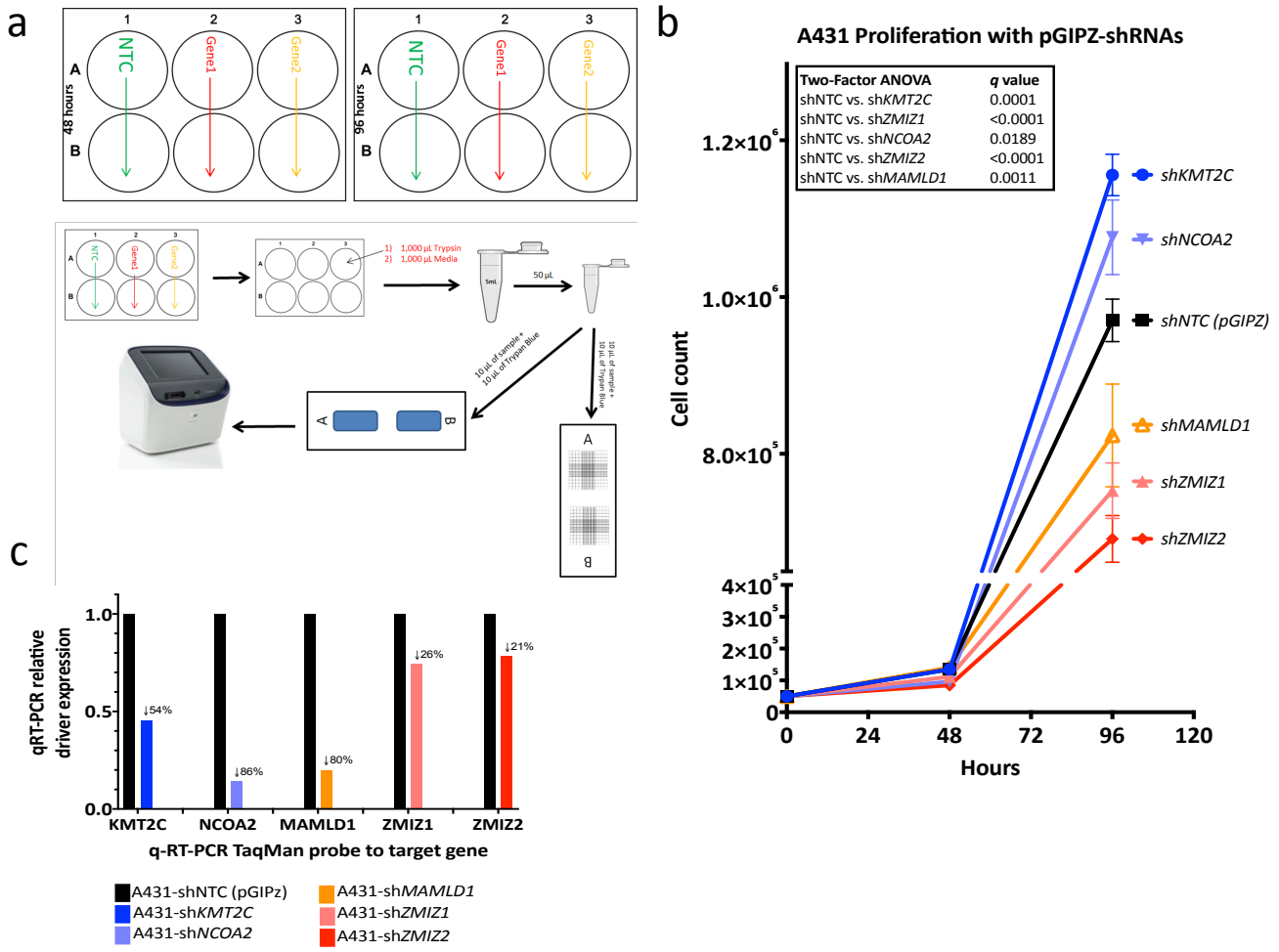
In "SCC13" and "SW900" only: *KMT2C,* and *MAMLD1*

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 17: 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 18: 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).

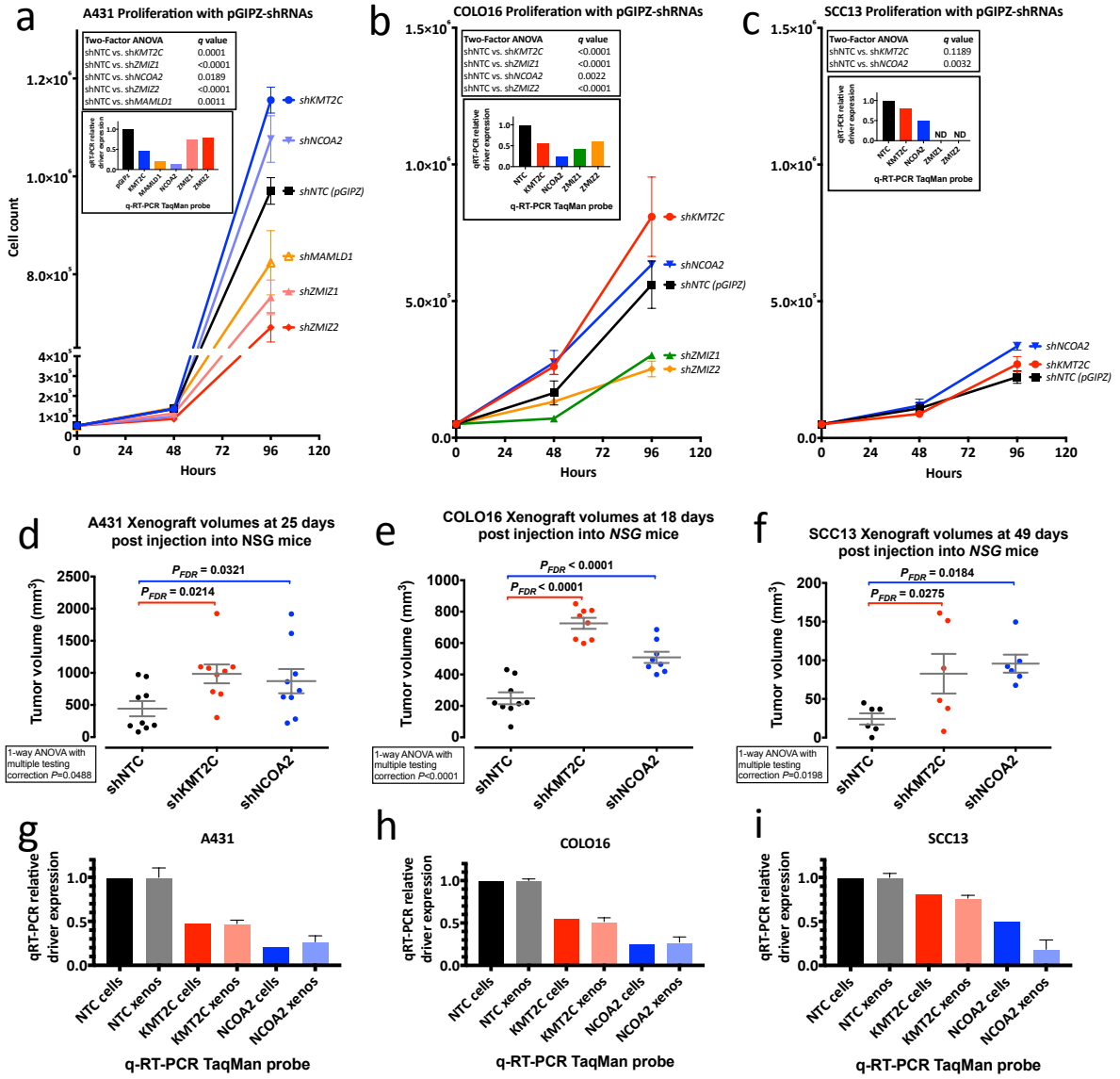
Chromatin modifying tumor suppressors drive human keratinocyte initiation and cuSCC progression

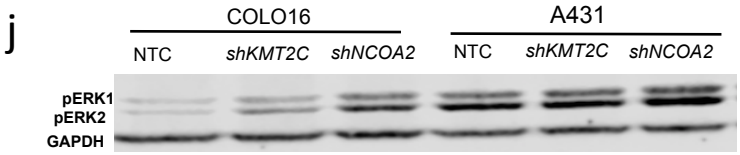
Since we demonstrated that knockdown of *KMT2C* or *NCOA2* increases the proliferation rates of A431 cell line *in vitro*, we asked whether this finding could be extended to other cell lines. Consistent with the observation made in A431 cells, we observed increased proliferation in COLO16 and SCC13 cells with decreased expression of *KMT2C or NCOA2*, relative to control (**Figure 8**, *P*<0.05, 1-factor ANOVA). We then determined whether *KMT2C and NCOA2* are cuSCC TSGs *in vivo*, and performed xenograft assays by subcutaneous injection of cells for each of the three cuSCC cell lines into *NSG* immunodeficient mice. All three cuSCC cell lines demonstrated enhanced *in vivo* tumor growth with *KMT2C*, *NCOA2,* or *CREBBP* depletion compared to cells with a non-targeting control shRNA (**Figure 8d-i**, 1-factor ANOVA, *q*<0.05 and **Supplementary Figure 19-20**), providing functional evidence that *KMT2C and NCOA2*, but not *CREBBP*, are human TSGs that promote *in vivo* cuSCC progression.

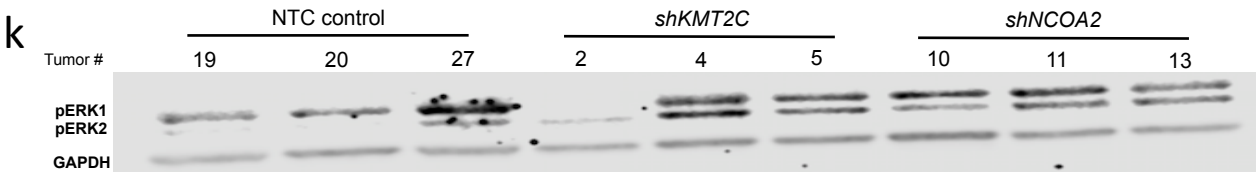
Finally, we explored the mechanistic basis of how decreased expression of *KMT2C* or *NCOA2* may progress cuSCC.

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

692 **Figure 8: *In vitro* and *in vivo* functional validation of chromatin remodeler and ZMIZ drivers in cuSCC cell lines.** (**a-c**)
693 Knockdown of *KMT2C, NCOA2* or *ZMIZ* genes alter proliferation rates of human cuSCC cell lines (n=3; error bars, SEM; statistical
694 significance measured by one-way ANOVA followed by multiple pairwise comparisons and FDR-adjusted *P*-values). (**d-f**) *KMT2C*
695 and *NCOA2* knockdown accelerates cuSCC xenograft progression *in vivo.* (**g-h**) qRT-PCR values in 'cells' at the time of injection
696 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
697 (A431, SCC13) independent masses) from xeno specimens. (**j**) Immunoblot for phospho-ERK expression in COLO16 and A431 cell
698 lines with *KMT2C* or *NCOA2* knockdown. (**k**) Immunoblot for phospho-ERK expression from three largest tumors for each
699 condition from the COLO16 xenograft assay shown in (**d**). (**l**) Summary of genes investigated with regards to their knockdown
700 effects in human immortal keratinocyte and cuSCC cell line phenotypes. *(Image on next page)*



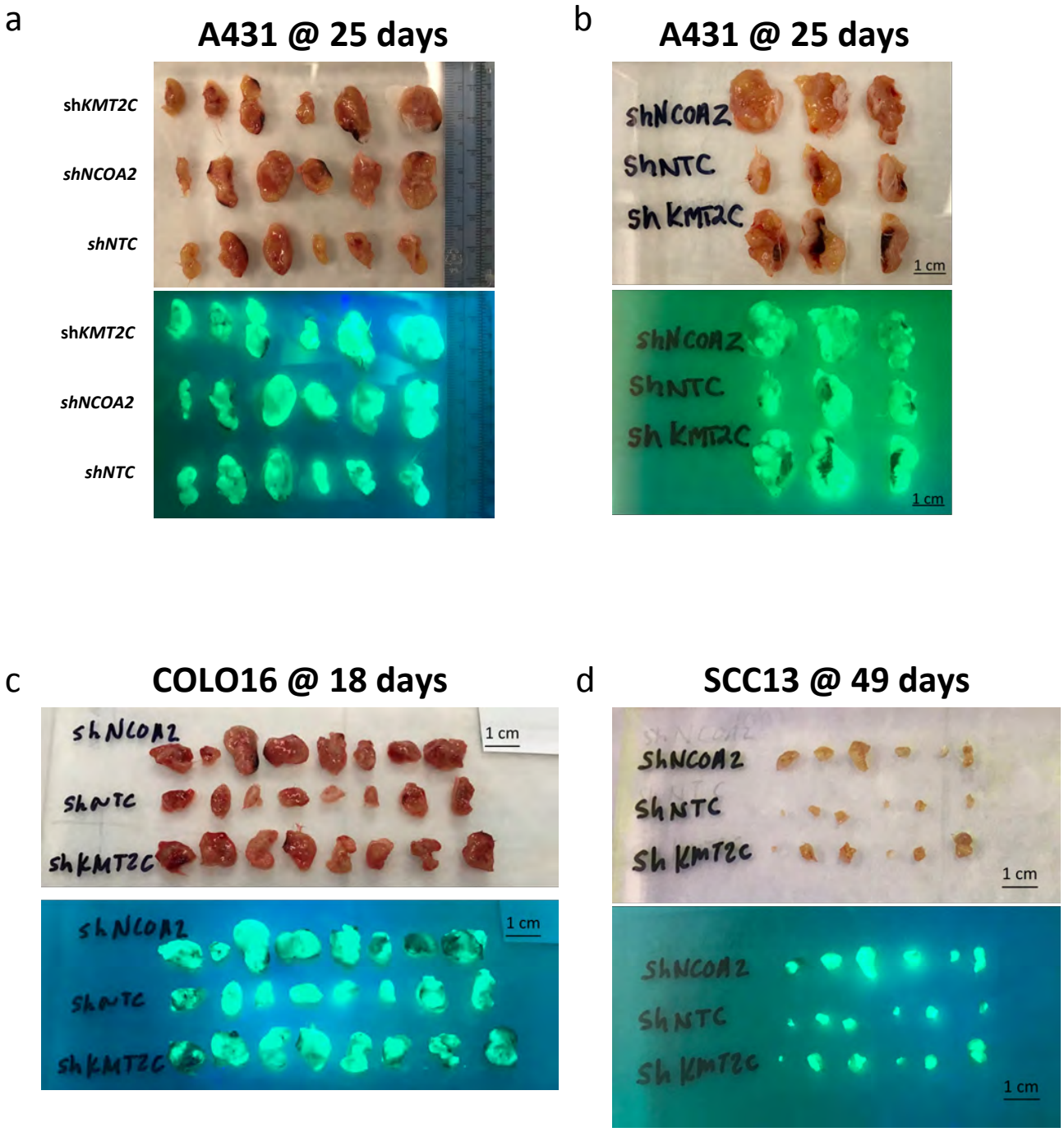




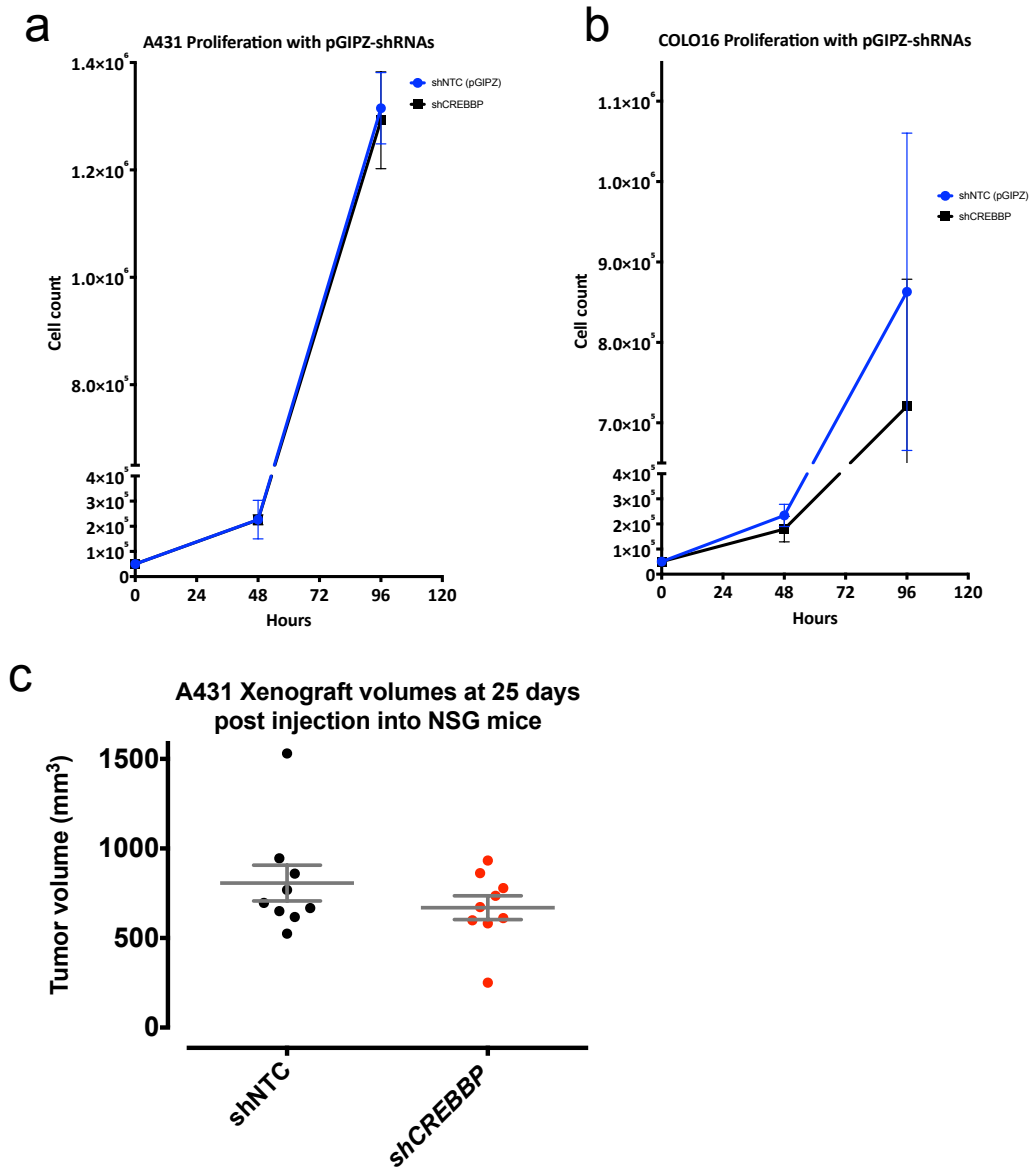
l

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	✓

701



Supplementary Figure 19: 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* immuno-deficient mice.



Supplementary Figure 20: *CREBBP* knockdown does not alter proliferation rate in cuSCC cell lines. (**a**) 96-hour proliferation assay (n=3, error bars SEM). (**b**) *In vivo* xenograft assay of COLO16 shNTC or shCREBBP into NSG immunodeficient mice (n=9 per condition).

DISCUSSION

In this study, we sought to define the cooperating genetic events required for keratinocyte transformation and progression to frank cuSCC using the *SB* transposon mutagenesis system *in vivo*. Using an *Actb-Cre* transgene to activate systemic expression of the transposon, we observed tumor lesions in multiple organs. The kinetics of tumorigenesis were further accelerated in mice harboring either null (*Trp53*$^{KO/+}$) (Jonkers et al., 2001) or recurrent point mutant (*Trp53*$^{R172H/+}$) (Olive et al., 2004) alleles. This observation is consistent with previous studies showing

717 that missense mutations in *Trp53* promote malignant transformation by abrogating both the tumor suppressive and
718 oncogenic (gain-of-function and/or dominant negative) functions of Trp53 *in vivo* (Doyle et al., 2010; Hanel et al.,
719 2013; Lang et al., 2004; Morton et al., 2010; Olive et al., 2004). While the SB|Onc3 mice developed a spectrum of
720 tumor types, including of the hematopoietic lineage, we observed an enrichment for late-stage, well-differentiated
721 cuSCC.

722 We subsequently focused on understanding the genetic events required for keratinocyte transformation. Our
723 SBCapSeq and SB Driver analysis identified 349 CCDs, regardless of *Trp53* status, of which 107 have been implicated
724 in human cancers. Intriguingly, we observed that the SB-driven cuSCC lesions were driven almost exclusively by SB
725 inactivating insertions in tumor suppressor genes, suggesting that cumulative loss of these genes is a central feature
726 of cuSCC. Furthermore, in cuSCC lesions that arose from K5-Cre; *Pten*$^{CKO/+}$ model previously reported, our SB Driver
727 Analysis workflow revealed inactivating insertion patterns in all 300 CCDs, which corroborates the Actin-Cre driven
728 model utilized in our study. Notably, in the single allele *Pten* loss model, we identified inactivation in *Chuk* (also
729 known as *Ikka*) as a significant driver of cuSCC, and this gene has previously been reported to have a tumor
730 suppressive role in cuSCC via regulating the EGFR signaling axis (Liu et al., 2008). However, since *Chuk* inactivation
731 only occurs in a haploinsufficient *Pten* background, these data suggest that inactivation of these genes alters
732 different pathways that cooperate to drive cuSCC. Tantalizingly, PanCancer analysis of over 40,000 tumors from
733 multiple tissue types using cBioPortal revealed a significant co-occurrence of both genes, suggesting that the
734 association between *Pten* and *Chuk* may be extended to cancer derived from different cellular lineages. Our data
735 provide a rich resource to rationally identify, prioritize, and credential likely cooperating mutational events in cuSCC
736 (**Fig. 3**).

737 Our report is the first to identify *Zmiz2* as a significant candidate cancer gene in any SB study and the first to
738 discover mutual exclusivity among activated *Zmiz1*, *Zmiz2*, and *Mamld1* in early keratinocyte transformation and
739 cuSCC progression. Insertions into *Zmiz1* (Dupuy et al., 2009; Friedel et al., 2013; Rogers et al., 2013) and *Mamld1*
740 (Friedel et al., 2013) have been previously observed in skin tumors induced by transposon insertional mutagenesis.
741 The possible redundancy of ZMIZ oncoproteins in the keratinocyte transformation program, by mutual exclusivity of
742 truncal insertions within cuSCC genomes, strongly suggests that the ZMIZ1$^{\Delta N185}$ and ZMIZ2$^{\Delta N184}$ oncoproteins may
743 have common neomorphic properties that contribute to the hallmarks of skin cancer. The work of Rogers *et al*.
744 (Rogers et al., 2013), and more recently Mathios *et al.* (Mathios et al., 2019), on the dynamic regulation of *ZMIZ1*
745 expression in human cancer, together with our finding that *ZMIZ1* and *ZMIZ2* are mutated with at least on non-silent
746 alteration in at least one third of published human cuSCC genomes (Pickering et al., 2014; South et al., 2014) and,
747 suggest that this gene may be an important initiating truncal mutation in human keratinocyte initiation and cuSCC
748 progression. Transcriptome analysis on SB-driven cuSCCs with high read depth insertions in *Zmiz1* or *Zmiz2* to
749 detected SB transposon fusions to both drivers. Of note, the N-terminal truncated versions of both *Zmiz* genes were
750 selected in a mutually exclusive manner. The Zmiz1$^{\Delta N185}$ and Zmiz2$^{\Delta N184}$ onco-proteins share 62% sequence identity,
751 and share the zinc finger PIAS, proline-rich transactivation, and AR binding domains. Importantly, both oncoproteins
752 lack the N-terminus in the full-length protein, which is thought to inhibit the intrinsic transcriptional activity C-
753 terminal proline-rich transactivation domain (Sharma et al., 2003). *ZMIZ1* and *ZMIZ2* are mutated in one third of

754 human cuSCC genomes sequenced to date (Pickering et al., 2014; South et al., 2014), however, its functional
755 significance remains unclear. *Zmiz1* has been shown to be involved in T cell development and leukemogenesis via
756 interacting with NOTCH1 (Pinnell et al., 2015), but how it is involved in the tumorigenesis of other cell lineages is
757 unclear. We have provided two preliminary lines of evidence with regards to how *ZMIZ* may be involved in
758 squamous cancers. First, shRNA-mediated knockdown of *ZMIZ1* and *ZMIZ2* in A431 and COLO16 human cuSCC cell
759 lines significantly decreased their growth rates, compared to non-targeting controls. While ZMIZ1 overexpression
760 has been shown to sufficient for disease initiation, our cell proliferation data suggests that pathways regulated by
761 ZMIZ1/2 continue to be active in advanced disease. Further, the association of high expression of a *ZMIZ*-centric
762 metagene derived from TCGA hnSCC cohort, with poor patient survival supports an oncogenic role for *ZMIZ* in
763 tumors of squamous lineage.

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

777 While mutations in chromatin remodelers have been reported in the literature, only recently have studies provided
778 functional evidence that this family of genes can function as tumor suppressor in the initiation and progression of
779 different cancers. Our *in vivo* screen identified inactivation of at least one gene involved in chromatin remodeling in
780 every tumor, suggesting that this family of genes may have an important role at some stage during cuSCC genesis.
781 Our functional data suggests that *CREBBP* is involved in the earlier stages of keratinocyte transformation, while
782 decreased *NCOA2* expression can drive disease progression. Interestingly, *KMT2C* knockdown drives keratinocyte
783 transformation or cuSCC progression in different cell systems utilized. Finally, while ongoing work is focused on
784 elucidating the mechanistic basis of how alterations in chromatin remodeler gene expression affect cuSCC biology,
785 we demonstrate that one route of cuSCC progression via decreased *NCOA2* or *KMT2C* expression is activation of the
786 MAPK pathway. Notably, this was observed in one of three cell lines assessed, suggesting that depending on the
787 genetic background, chromatin remodelers influence disease progression via different mechanisms.

788

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

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

809 **Acknowledgments**

810 M.B.M. and K.M.M. are supported by start-up funds provided by the H. Lee Moffitt Cancer Center and Research
811 Institute, a National Cancer Institute-designated Comprehensive Cancer Center through the Cancer Center Support
812 Grant (NCI P30-CA76292; PI: T.A. Sellers). M.B.M. was also supported by Career Enhancement Program Grant Award
813 from the NCI Moffitt Skin Cancer SPORE (5P50CA168536-05; PI: V. Sondak). N.G.C. and N.A.J. were supported by the
814 Biomedical Research Council, Agency for Science, Technology, and Research, Singapore and as CPRIT Scholars in
815 Cancer Research from the Cancer Prevention Research Institute of Texas. The authors wish to thank the Copeland–
816 Jenkins labs in Singapore and Houston for helpful discussions; we thank Tyler Jacks (MIT) for P53 LSL R172H mice;
817 Anton Berns (NKI) for P53 floxed mice; Gail Martin (UCSF) for Actin-beta-Cre mice; James G. Rheinwald and the
818 Harvard Skin Disease Research Center for the SCC13 cell line; David Adams, Theo Whipp, Richard Rance, Alistair
819 Rust, and the Wellcome Trust Sanger Institute sequencing and informatics teams for 454 sequencing; the Elsa Flores
820 Lab for use of the LiCOR imaging system; Sean Yoder (Moffitt Cancer Center) and Dr. Chaomei Zhang (Moffitt Cancer
821 Center) for providing sequencing services; Pearlyn Cheok, Nicole Lim, Dorothy Chen and Cherylin Wee for assistance
822 with tumor monitoring and animal husbandry at IMCB (Singapore, Republic of Singapore), Hubert Lee and Erik
823 Freiter for assistance with animal husbandry at HMRI (Houston, TX, USA), and Crystal Reed and Bethanie Gore for
824 assistance with animal husbandry at MCC/USF (Tampa, FL, USA). Necropsy and histology were performed by the

825 Advanced Molecular Pathology Laboratory, Institute of Molecular and Cell Biology (A*STAR), Singapore. Necropsy
826 and histology were also performed by the Tissue Core Facility and Ion Torrent sequencing was performed by the
827 Molecular Genomics Core at the H. Lee Moffitt Cancer Center & Research Institute, a comprehensive cancer center
828 designated by the National Cancer Institute, and funded in part by Moffitt's Cancer Center Support Grant (P30-
829 CA076292; PI: T.A. Sellers).

830 **Author Contributions**

831 M.B.M., K.M.M. N.G.C., and N.A.J. conceived the project and provided laboratory resources and personnel for animal
832 husbandry, specimen archiving, sequencing and compute management. M.B.M. and K.M.M. designed the studies,
833 directed the research, interpreted the data, performed experimental work, designed the SBCapSeq oligos,
834 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.-
835 K.B., and K.Y.T. provided reagents, performed and/or analyzed experimental datasets. J.Y.N., M.A.B., and M.B.M.
836 performed statistical and bioinformatics analyses. K.R. and S.M.R. performed and directed necropsy and
837 histopathological analysis. J.M.W. performed and directed veterinary pathology analysis, including tumor grading
838 and 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.-
839 M. performed and optimized capture hybridizations for the SBCapSeq method and/or performed Ion Torrent
840 sequencing for SBCapSeq, RNA-Seq experiments. M.B.M., L.G.-R., A.A., and A.L.M performed *in vitro* experiments.
841 M.B.M. and A.L.M performed *in vivo* xenograft experiments. A.A. and M.B.M. wrote the manuscript with contributions
842 from N.G.C., N.A.J., K.Y.T., M.A.B., and K.M.M. All co-authors approved the manuscript prior to submission.

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

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

846 **Methods**

847 **Mice used for SB screens.** The following alleles were used to construct the SB-driven mouse model of multiple solid
848 tumor histologies: *Actb-Cre* (FVB/N-Tg(ACTB-cre)2Mrt/J, ref. 8(Lewandoski et al., 1997));*Trp53*$^{flox/+}$ (FVB.129P2-
849 *Trp53*^{tm1Brn}/Nci, ref. 9(Jonkers et al., 2001)); *Trp53*$^{LSL\text{-}R172H/+}$ (129S4-*Trp53*^{tm2Tyj}/Nci; ref. 10(Olive et al., 2004));
850 T2/Onc2(TG.12740) (TgTn(sb-T2/Onc3)12740Njen; ref. 54 (Dupuy et al., 2009)); and *Rosa26*-LSL SBase or
851 SBase^{LSL}; (Gt(ROSA)26Sor$^{tm2(sb11)Njen}$; ref. 55(Starr et al., 2009)). The resulting cohorts of mice were on mixed genetic
852 backgrounds consisting of C57BL/6J, 129, C3H and FVB. Genotyping by PCR assays with primers specific to the
853 alleles was performed. No sample size estimate was used to determine the number of mice for aging. Mice were bred
854 and maintained in accordance with approved procedures directed by the respective Institutional Animal Care and
855 Use Committees in the National Cancer Institute Frederick National Lab, A*STAR Biological Resource Centre,
856 Houston Methodist Research Institute, and Moffitt Cancer Center. Animals were co-housed for the duration the
857 experiment, except in rare cases where single mice were separated based on vet recommendation due to fighting or
858 near tumor burden end-point. Gross necropsies were performed and all masses were documented and prepared for

859 subsequent analysis. See **Supplementary Figure 1** for details of mouse crosses. Both sexes were used for
860 experiments. No randomization was performed; mice were assigned to groups based on genotype. No blinding was
861 performed.

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

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

877 **Mapping transposon insertion sites using the SBCapSeq method.** Full details for the SBCapSeq protocol (Mann
878 et al., 2016) optimized for sequencing from solid tumors, will be published elsewhere (Mann *et al.*, in preparation;
879 for general protocol and concept, see also ref. (Mann et al., 2016)). Briefly, for selective SB insertion site sequencing
880 by liquid hybridization capture, gDNA (0.5 µg per sample) of either bulk tumor specimens or single cell WGA
881 genomes was used for library construction using the AB Library Builder™ System, including random fragmentation
882 and ligation of barcoded Ion Xpress™ sequencing adapters. Adapter-ligated templates were purified by Agencourt®
883 AMPure beads® and fragments with insert size of 200±30bp were excised, purified by Agencourt® AMPure® beads,
884 amplified by 8 cycles of adapter-ligation-mediated polymerase chain reaction (aLM-PCR), and purified by
885 Agencourt® AMPure® beads with elution in 50 µl of TE (1X Tris-EDTA [Ethylenediaminetetraacetic acid], pH8).
886 Capture hybridization of single or multiplexed up to 12 barcoded libraries (60 ng per sample) was performed using
887 custom xGen® Lockdown® Probes (IDT). All 120-mer capture and blocking oligonucleotide probes were purchased
888 from IDT as Ultramer DNA Oligos with Standard Desalting. Bar-coded single and multiplex captured library
889 fragments were further amplified by 12 cycles of LM-PCR and run using an Agilent 2100 Bioanalyzer or TapeStation
890 to estimate enrichment. Bar-coded single and multiplex captured libraries were quantified by Qubit® Fluorometer
891 and quantitative Real Time-PCR (qRT-PCR) were used to dilute libraries for template preparation and Ion Sphere™
892 Particle (ISP) loading using Ion Chef™ System and sequencing on the Ion Proton™ platform with PI_{v3} semiconductor
893 wafer chips per manufacturers recommended instructions. High-throughput sequencing of up to 39 multiplex
894 captured libraries was carried out per PI_{v3} chip to achieve at least 1.5 million reads per barcode. Reads containing

895 the transposon IRDR element were processed using the SBCapSeq bioinformatic workflow as described(Mann et al.,
896 2016).

897 **SB Driver Analysis.** BED formatted files containing SB insertions from each of the histologically verified SB-cuSCC,
898 SB-cuKA, and SB-cuSK cohort specimens (**Supplementary Tables 2,4,7,11,14,17,21**) were used to perform SB
899 Driver Analysis to identify statistically significant discovery, progression, and truncal driver genes that contain more
900 SB insertions than expected by chance and were recurrently altered in three or more tumors (Newberg et al.,
901 2018b). Discovery and progression SB Driver Analysis considered all SB insertion events; truncal SB Driver Analysis
902 considered only insertions represented by 5 or more reads. Statistically significant discovery drivers were defined
903 as adjusted *p*-values using false discovery rate (FDR) multiple testing correction (*q*-value<0.5). Statistically
904 significant progression and truncal drivers were defined as adjusted *p*-values using family-wise error rate (FWER)
905 multiple testing correction (FWER adjusted-*P*<0.05). Statistically significant drivers on donor and non-donor
906 chromosomes analyses were performed separately and combined into a single driver lists (**Supplementary Tables**
907 **2,5-6,12-13,15-16**). Due to a local-hopping phenomenon known to occur with SB, where transposition events are
908 biased to occur in cis along the donor chromosome more frequently than in trans to non-donor chromosomes in the
909 genome, insertions from the donor chromosome are typically filtered away computationally before candidate cancer
910 genes are identified. However, doing so in cohorts derived from mice with only a single SB donor allele (like our
911 cuSCC cohorts) means that genome-wide driver analysis is not possible, because all genes on the donor chromosome
912 are censored and are not reported even though they may contribute to the overall tumor burden. Since we obtained
913 quantitative SBCapSeq datasets in this study, we evaluated whether all genes on the donor chromosome should be
914 censored or if perhaps, as observed in *PiggyBac* transposon genomes, only a relatively small portion of the donor
915 chromosome locus, occurring in close proximity to the donor concatemer insertion site, demonstrate significant bias
916 compared with the frequency of SB insertions into non-donor chromosomes. In the cuSCC cohorts sequenced with
917 SBCapSeq, we found that donor chromosome SB insertion site frequencies are significantly higher between
918 chr9:79,000,000 and chr9:95,000,000 and drop to the genome wide average for non-donor chromosomes over the
919 rest of chromosome 9 (**Supplementary Figure 2b**), suggesting that filtering all SB donor chromosome insertion
920 events may not be warranted. To confirm and extend this observation, we used all tumors within the SBCDDB with
921 chromosome 9 donor alleles and identified that the same region between chr9:79,000,000 and chr9:95,000,000 had
922 a higher frequency of SB insertions than the rest of chromosome 9 loci, which again matched non-donor
923 chromosome insertion frequencies (**Supplementary Figure 2a**). Thus, we ran SB Driver Analysis on tumors derived
924 from mouse cohorts harboring an SB T2/Onc3 TG.12740 donor allele chromosome 9 (cuSCC83_454, cuSCC60_SBC,
925 and cuKA11_SBC), by excluding the 81 genes that map between *Filip1* at chr9:79663368-79825689 and *Tfdp2* at
926 96096693-96224065 from the driver gene output files.

927 **Microarray gene expression analysis**. Gene expression profiling of histologically confirmed SB-cuSCC masses
928 selected by presence or absence of *Zmiz1* insertion by 454Splink sequencing was performed using Affymetrix
929 microarrays, as described previously (Mann et al., 2016). Briefly, 100ng of total RNA for each sample was extracted
930 using a NORGEN Biotek Animal Tissue RNA Purification kit (Cat #25700) followed by labeled with an Affymetrix 3'
931 IVT Express kit (Cat # 901229) using the manufacturer's instructions. Labeled samples were hybridized to

932 Affymetrix GeneChip Mouse Genome 430 2.0 Arrays, and scanned at the University of Otago Genomics &
933 Bioinformatics Facility. Raw data processing used R (version 2.15) (Team, 2013) with the "rma" function of the
934 "affy" package (Gautier et al., 2004), including quantile normalization but no background correction. Quality
935 assessment of the microarray data was performed using R (R_Core_Team, 2008; 2013) or the "affyQCReport"
936 package (Parman et al., 2005). Data analysis and differential gene expression analysis were performed using Python,
937 R (R_Core_Team, 2008; 2013) and visualized using the R Shiny package in RStudio.

938 **Transcriptome Sequencing.** Total RNA was isolated from flash frozen necropsy specimens using *mir*Vana™ miRNA
939 Isolation Kit (Ambion® by Life Technologies, AM1560) from a shaved portion of each pathology verified cuSCC mass
940 from 7 tumors for which sufficient tissues were available and the prescence of a high read depth *Zmiz1* or *Zmiz2*
941 trunk driver insertion was identified by SBCapSeq. Whole transcriptome RNA-Seq (wtRNA-Seq) libraries were
942 prepared from total RNA (5 µg per sample) plus ERCC RNA Spike-In (Ambion® by Life Technologies, 4456739)
943 followed by selective ribosomal RNA (rRNA) depletion using the RiboMinus™ Eukaryote System v2 (Ambion® by
944 Life Technologies, A15026). rRNA-depleted total RNA (500 ng per sample) was used for whole transcriptome library
945 construction according to the Ion Total RNA-Seq Kit for the AB Library Builder™ System (Life Technologies,
946 4482416) protocol for barcoded libraries, Ion Sphere™ Particle (ISP) loading using Ion Chef™ System and
947 sequencing on the Ion Proton™ platform with PI_{v3} semiconductor wafer chips per manufacturers recommended
948 instructions. Up to 4 RNA-Seq libraries were multiplex sequenced on a PI_{v3} chip twice to achieve >3 GB (~40 million
949 reads) per specimen. The Life Technologies Torrent Suite software was used to perform checking of the raw
950 sequence data before the generation of sequencing read files in FASTQ format and minimally processed mRNA-Seq
951 and wtRNA-Seq reads were processed using the Bowtie2 (Langmead and Salzberg, 2012) and Tophat (Trapnell et
952 al., 2009) algorithms to align reads to a custom version of the mouse mm9+pT2/Onc genome, as described (Mann et
953 al., 2016). The detection of SB fusion events, defining novel SB fusion transcripts, and measuring transcript
954 abundance in SB cuSCC tumor specimens from RNA-Seq data were performed as described (Mann et al., 2016).

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

960 **Human cancer cell lines.** The cutaneous squamous cell carcinoma A431(CRL–1555™) and grade IV lung squamous
961 cell carcinoma SW900 (HTB–59™) cell lines, were purchased from American Type Culture Collection (ATCC®) and
962 grown according to the manufacturer's suggested conditions in complete medium (1× DMEM (ATCC, 30–2002) for
963 A431cells or 1× ATCC-formulated Leibovitz's L–15 Medium, Catalog No. 30–2008 medium (ATCC, 30–2001) for
964 SW900 cells) supplemented with 10% FBS and 1× penicillin-streptomycin grown at 37 °C in 5% CO2. The cutaneous
965 squamous cell carcinoma SCC13 cell line (Rheinwald and Beckett, 1981; Rheinwald and Green, 1977), were obtained
966 from Harvard Skin Disease Research Center (HSDRC) and grown using the human keratinocyte culture methods
967 provided by the HSDRC. The cutaneous squamous cell carcinoma COLO16 cell line (Vin et al., 2013) were a gift from

968 Dr. K. Y. Tsai and grown in DMEM/Ham's F12 50/50 media supplemented with cholera toxin (Sigma, C0852), insulin
969 (Sigma, I5500-50MG), epidermal growth factor (Serotec/Bio-Rad, EGF-1), hydrocortisone (Sigma, H4881-1G),
970 liothyronine (Sigma, T6397-100MG) and apo-transferrin (Sigma, T2252-100MG). All cell lines were verified human
971 origin and free from pathogens and mycoplasma. Mycoplasma infection monitoring was performed using MycoAlert
972 Detection Kit (Lonza) and only mycoplasma-free cultures were used.

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

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

989 **Proliferation assay**. To assess cuSCC progression, cuSCC cell lines were seeded in 6 well plates at a density of $5 \times$
990 10^4 cells per well in complete media. Cell numbers for each condition were counted after 48- and 96-hour post-
991 seeding. To assess transformation by proliferation, HaCaT cells were seeded at a density of 5×10^3 cells per well and
992 cell counts performed after seven days.

993 **Soft agar assay**. To assess anchorage-independent growth, HaCaT cells were seeded at a density of 1×10^4 cells per
994 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
995 day, and replaced every two days for four weeks. At end point, cells were fixed in 20% methanol and 0.0025%
996 crystal violet. Washes were done with milliQ water until low background was achieved, and plates were scanned,
997 and colonies counted for each condition.

998 **qRT-PCR.** Total RNA was purified and DNase treated using the RNeasy Mini Kit (Qiagen). Synthesis of cDNA was
999 performed using SuperScript VILO Master Mix (Life Technologies). Quantitative PCR analysis was performed on the
1000 QuantStudio 12K Flex System (Life Technologies) or 7900HT Sequence Detection System (Applied Biosystem). All
1001 signals were normalized to the levels of GAPDH TaqMan probes. TaqMan probes were obtained from Life
1002 Technologies (**Supplementary Table 41**).

1003 **Xenografts.** One million cells were prepared for injection into the left flank of randomized selected male and female
1004 immunodeficient NSG (NOD.Cg-Prkdcscid;Il2rgtm1Wjl/SzJ; JAX, 005557; 6–15 weeks old) mice. Allocation to study
1005 groups was random. Xenograft measurements were taken twice weekly using digital calipers while mice were
1006 conscious but restrained by an experimenter familiar with collecting caliper measurements of xenografts and
1007 blinded to the experimental group designations. GFP fluorescence was visualized at the time of caliper
1008 measurement. Ellipsoid tumor volumes were calculated as volume (mm3) = 0.52(length (mm) × width2 (mm2)),
1009 where the two longest axes, length and width, were the major and minor diameter measurements, respectively;
1010 width2 represents an assumption that the xenograft depth was equivalent to the diameter of the minor axis.
1011 Statistical significance of tumor volumes at defined time points was determined by either one-way t test for all
1012 cohorts relative to shNTC or by two-way, repeated-measures ANOVA with Bonferroni correction for multiple
1013 comparisons, as indicated.

1014 **Western blot.** Whole cell or tissue lysates were prepared in RIPA buffer supplemented with protease and
1015 phosphatase inhibitors, and samples sonicated to achieve optimum lysis. Protein concentrations were quantitated
1016 using the BCA assay, and 20 μg of lysates were loaded into 8% SDS-polyacrylamide gels. Gels were resolved at 80V
1017 for 120 mins, transferred onto nitrocellulose membrane for 70 mins at 0.35A with transfer buffer containing 20%
1018 methanol, and blocked for one hour at room temperature with 5% w/v BSA in 0.1% TBST. Membranes were
1019 incubated with primary antibodies in 5% w/v BSA in 0.1% TBST. Antibodies used in this study are: Total ERK
1020 (1:1000, Cell Signaling #9102), phospho ERK (1:1000, Cell Signaling #4695), GAPDH (1:5000, Santa Cruz sc-69778),
1021 anti-mouse and anti-goat secondary antibodies (1:10000, LiCOR IRDye 680/800 Cat. 925-68070/926-32211). Blots
1022 were imaged using the LiCOR Odyssey Fc system.

1023 **Software.** Unless otherwise noted, bioinformatic analysis pipelines, report generation, and figure visualization
1024 performed using bash, R (version 3.5.2), RStudio (Version 1.1.463), Python scripts, and GraphPad Prism 8 software
1025 (Version 8.1.1). Hierarchical clustering was performed in Python 2.7.10 with the scipy 0.13.0b1 toolbox using the
1026 Hamming distance metric with Ward's linkage method. The R packages ggplot2 (Wickham, 2009) and Gviz (Henare
1027 et al., 2012) were used to generate the graphics in various figure panels.

1028 **URLs.** TgTn(sb-T2/Onc3)12740Njen; Gt(ROSA)26Sortm2(sb11)Njen double mutant mice,
1029 https://ncifrederick.cancer.gov/Lasp/MouseRepository/MouseModels/StrainDetails.aspx?StrainNum=01XGB&g=R
1030 OSA26; FVB/N-Tg(ACTB-cre)2Mrt/J mice, http://jaxmice.jax.org/strain/003376.html; FVB.129P2-
1031 Trp53tm1Brn/Nci mice,
1032 https://ncifrederick.cancer.gov/Lasp/MouseRepository/MouseModels/StrainDetails.aspx?StrainNum=01XC2&g=Tr
1033 p53; 129S4-Trp53tm2Tyj/Nci mice,
1034 https://ncifrederick.cancer.gov/Lasp/MouseRepository/MouseModels/StrainDetails.aspx?StrainNum=01XM2&g=T
1035 rp53; Cancer Gene Census, http://cancer.sanger.ac.uk/census/; Mouse Genome Informatics (MGI)
1036 database, ftp://ftp.informatics.jax.org/pub/reports/index.html; The Cancer Genome Atlas (TCGA),
1037 http://cancergenome.nih.gov/; Enrichr, http://amp.pharm.mssm.edu/Enrichr/, STRING, https://string-db.org/.

1038 **Supplementary Tables.** Supplementary Table datasets may be downloaded from NIH.figshare.com at
1039 (http://dx.doi.org/10.35092/yhjc.11441130). Brief descriptions of each table are provided for reference.

1040 **Supplementary Table 1**: Tumor incidence and subgroup classifications by cohort

1041 **Supplementary Table 2**: Sequencing projects summary by cohort

1042 **Supplementary Table 3**: Specimen metafile data for projects sequenced using Splink_PCR protocol with 454 GS-
1043 FLX Titanium pyrosequencer

1044 **Supplementary Table 4**: BED file of SB insertions for cuSCC83_454

1045 **Supplementary Table 5**: Discovery and progression SB Driver Analysis for cuSCC83_454

1046 **Supplementary Table 6**: Trunk SB Driver Analysis for cuSCC83_454

1047 **Supplementary Table 7**: BED file of SB insertions for cuKA51_454

1048 **Supplementary Table 8**: Discovery and progression SB Driver Analysis for cuKA51_454

1049 **Supplementary Table 9**: Trunk SB Driver Analysis for cuKA51_454

1050 **Supplementary Table 10**: Specimen metafile data for projects sequenced using SBCapSeq protocol with Ion
1051 Torrent Proton sequencer

1052 **Supplementary Table 11**: BED file of SB insertions for cuSCC60_SBC.

1053 **Supplementary Table 12**: Discovery and progression SB Driver Analysis for cuSCC60_SBC.

1054 **Supplementary Table 13**: Trunk SB Driver Analysis for cuSCC60_SBC.

1055 **Supplementary Table 14**: BED file of SB insertions for cuKA11_SBC

1056 **Supplementary Table 15**: Discovery and progression SB Driver Analysis for cuKA11_SBC

1057 **Supplementary Table 16**: Trunk SB Driver Analysis for cuKA11_SBC

1058 **Supplementary Table 17**: BED file of SB insertions for cuSK32_SBC

1059 **Supplementary Table 18**: Discovery and progression SB Driver Analysis for cuSK32_SBC

1060 **Supplementary Table 19**: BED file of SB insertions for 4 cuSCC genomes selected for multi-region resequencing
1061 because they had intermixing of cuSCC and cuKA histologies

1062 **Supplementary Table 20**: SBCapSeq read depth and analysis for 4 cuSCC genomes selected for multi-region
1063 resequencing because they had intermixing of cuSCC and cuKA histologies

1064 **Supplementary Table 21**: BED file of SB insertions for cuSK10_SBC

1065 **Supplementary Table 22**: Discovery and progression SB Driver Analysis for cuSK10_SBC

1066 **Supplementary Table 23**: Contingency Tables comparing Cancer Gene Census (GRCh38 COSMICv86) driver
1067 enrichment with cuSCC cohorts

1068 **Supplementary Table 24**: Venn diagram for overlap of Cancer Gene Census (GRCh38 COSMICv86) driver
1069 enrichment with all SB|cuSCC Trunk Driver and SB|cuKA Trunk Driver lists

1070 **Supplementary Table 25**: Venn diagram for overlap of Cancer Gene Census (GRCh38 COSMICv86) driver
1071 enrichment with all SB|cuSCC Progression Driver and SB|cuKA Progression Driver lists

1072 **Supplementary Table 26**: Venn diagram for overlap of Cancer Gene Census (GRCh38 COSMICv86) driver
1073 enrichment with cuSCC60_SBC Trunk Driver, cuSCC83_454 Trunk Driver, and cuKA51_454 Trunk Driver lists

1074 **Supplementary Table 27**: Venn diagram for overlap of Cancer Gene Census (GRCh38 COSMICv86) driver
1075 enrichment with cuSCC60_SBC Progression Driver, cuSCC83_454 Progression Driver, and cuKA51_454 Progression
1076 Driver lists

1077 **Supplementary Table 28**: Enrichr gene set pathway enrichment analysis of cuSCC drivers

1078 **Supplementary Table 29**: Normalized microarray values per gene from RNA isolated from cuSCC genomes with
1079 and without Zmiz1 insertions

1080 **Supplementary Table 30**: Normalized microarray values per probe from RNA isolated from cuSCC genomes with
1081 and without Zmiz1 insertions

1082 **Supplementary Table 31**: Top 96 genes with differential expression analysis from microarray data from RNA
1083 isolated from cuSCC genomes with and without Zmiz1 insertions

1084 **Supplementary Table 32**: Enrichr gene set pathway enrichment analysis of the top 96 differentially expressed
1085 genes from microarray data from RNA isolated from cuSCC genomes with and without Zmiz1 insertions

1086 **Supplementary Table 33**: Enrichr gene set gene ongology term enrichment analysis of the top 96 differentially
1087 expressed genes from microarray data from RNA isolated from cuSCC genomes with and without Zmiz1 insertions

1088 **Supplementary Table 34**: All 289 genes with differential expression analysis from microarray data from RNA
1089 isolated from cuSCC genomes with and without Zmiz1 insertions with FDR adjusted-$P<0.0001$

1090 **Supplementary Table 35**: Summary of 7 cuSCC transcriptomes selected for whole transcriptome RNA-seq analysis

1091 **Supplementary Table 36**: BED file of SBfusion insertions in 7 cuSCC genomes by whole transcriptome RNA-seq
1092 analysis

1093 **Supplementary Table 37**: Venn diagram for overlap of genes with SBfusion reads detected by whole transcriptome
1094 RNA-seq analysis and cuSCC60_SBC discovery drivers

1095 **Supplementary Table 38**: Venn diagram for overlap of genes with SBfusion reads detected by whole transcriptome
1096 RNA-seq analysis and all cuSCC drivers

1097 **Supplementary Table 39**: Transcripts per million (TPM) normalized whole transcriptome RNA-seq values per gene
1098 from RNA isolated from cuSCC genomes with and without Zmiz1 insertions

1099 **Supplementary Table 40**: Fragments Per Kilobase of Transcripts per Million (FPKM) normalized whole
1100 transcriptome RNA-seq values per gene transcript from RNA isolated from cuSCC genomes with and without Zmiz1
1101 insertions

1102 **Supplementary Table 41**: siRNA and TaqMan probes used in this study

1103 **Supplementary Table 42**: siRNA primary screen results

1104 **Supplementary Table 43**: Lentiviral vectors containing shRNAs used in this study

1105 **References**

1106 Andricovich, J., Perkail, S., Kai, Y., Casasanta, N., Peng, W., and Tzatsos, A. (2018). Loss of KDM6A Activates Super-
1107 Enhancers to Induce Gender-Specific Squamous-like Pancreatic Cancer and Confers Sensitivity to BET
1108 Inhibitors. Cancer Cell *33*, 512-526 e518.

1109 Beliakoff, J., and Sun, Z. (2006). Zimp7 and Zimp10, two novel PIAS-like proteins, function as androgen receptor
1110 coregulators. Nucl Recept Signal *4*, e017.

1111 Boukamp, P., Petrussevska, R.T., Breitkreutz, D., Hornung, J., Markham, A., and Fusenig, N.E. (1988). Normal
1112 keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. J Cell Biol *106*, 761-771.

1113 Cancer Genome Atlas, N. (2015). Comprehensive genomic characterization of head and neck squamous cell
1114 carcinomas. Nature *517*, 576-582.

1115 Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr:
1116 interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

1117 Chitsazzadeh, V., Coarfa, C., Drummond, J.A., Nguyen, T., Joseph, A., Chilukuri, S., Charpiot, E., Adelmann, C.H., Ching,
1118 G., Nguyen, T.N*., et al.* (2016). Cross-species identification of genomic drivers of squamous cell carcinoma
1119 development across preneoplastic intermediates. Nat Commun *7*, 12601.

1120 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.*
1121 (2018). KMT2C Mutations in Diffuse-Type Gastric Adenocarcinoma Promote Epithelial-to-Mesenchymal
1122 Transition. Clin Cancer Res *24*, 6556-6569.

1123 Collier, L.S., and Largaespada, D.A. (2007). Transposons for cancer gene discovery: Sleeping Beauty and beyond.
1124 Genome Biol *8 Suppl 1*, S15.

1125 Copeland, N.G., and Jenkins, N.A. (2010). Harnessing transposons for cancer gene discovery. Nat Rev Cancer *10*, 696-
1126 706.

1127 de la Rosa, J., Weber, J., Friedrich, M.J., Li, Y., Rad, L., Ponstingl, H., Liang, Q., de Quiros, S.B., Noorani, I., Metzakopian,
1128 E*., et al.* (2017). A single-copy Sleeping Beauty transposon mutagenesis screen identifies new PTEN-cooperating
1129 tumor suppressor genes. Nat Genet *49*, 730-741.

1130 de Ridder, J., Uren, A., Kool, J., Reinders, M., and Wessels, L. (2006). Detecting statistically significant common
1131 insertion sites in retroviral insertional mutagenesis screens. PLoS Comput Biol *2*, e166.

1132 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
1133 Required to Maintain Broad H3K4me3 Peaks and Super-Enhancers at Tumor Suppressor Genes. Mol Cell *70*,
1134 825-841 e826.

1135 Doyle, B., Morton, J.P., Delaney, D.W., Ridgway, R.A., Wilkins, J.A., and Sansom, O.J. (2010). p53 mutation and loss
1136 have different effects on tumourigenesis in a novel mouse model of pleomorphic rhabdomyosarcoma. J Pathol
1137 *222*, 129-137.

1138 Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G., and Jenkins, N.A. (2005). Mammalian mutagenesis using a
1139 highly mobile somatic Sleeping Beauty transposon system. Nature *436*, 221-226.

1140 Dupuy, A.J., Rogers, L.M., Kim, J., Nannapaneni, K., Starr, T.K., Liu, P., Largaespada, D.A., Scheetz, T.E., Jenkins, N.A.,
1141 and Copeland, N.G. (2009). A modified sleeping beauty transposon system that can be used to model a wide
1142 variety of human cancers in mice. Cancer Res *69*, 8150-8156.

1143 Friedel, R.H., Friedel, C.C., Bonfert, T., Shi, R., Rad, R., and Soriano, P. (2013). Clonal expansion analysis of transposon
1144 insertions by high-throughput sequencing identifies candidate cancer genes in a PiggyBac mutagenesis screen.
1145 PLoS One *8*, e72338.

1146 Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A
1147 census of human cancer genes. Nat Rev Cancer *4*, 177-183.

1148 Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe
1149 level. Bioinformatics *20*, 307-315.

1150 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.,
1151 *et al.* (2013). Sleeping Beauty mutagenesis in a mouse medulloblastoma model defines networks that
1152 discriminate between human molecular subgroups. Proc Natl Acad Sci U S A *110*, E4325-4334.

1153 Gozdecka, M., Meduri, E., Mazan, M., Tzelepis, K., Dudek, M., Knights, A.J., Pardo, M., Yu, L., Choudhary, J.S.,
1154 Metzakopian, E., *et al.* (2018). UTX-mediated enhancer and chromatin remodeling suppresses myeloid
1155 leukemogenesis through noncatalytic inverse regulation of ETS and GATA programs. Nat Genet *50*, 883-894.

1156 Hanel, W., Marchenko, N., Xu, S., Yu, S.X., Weng, W., and Moll, U. (2013). Two hot spot mutant p53 mouse models
1157 display differential gain of function in tumorigenesis. Cell Death Differ *20*, 898-909.

1158 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.
1159 (2012). Dissection of stromal and cancer cell-derived signals in melanoma xenografts before and after treatment
1160 with DMXAA. Br J Cancer *106*, 1134-1147.

1161 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,
1162 enhances androgen receptor-mediated transcription and interacts with SWI/SNF-like BAF complexes. Mol
1163 Endocrinol *19*, 2915-2929.

1164 Huguet, E.L., McMahon, J.A., McMahon, A.P., Bicknell, R., and Harris, A.L. (1994). Differential expression of human
1165 Wnt genes 2, 3, 4, and 7B in human breast cell lines and normal and disease states of human breast tissue.
1166 Cancer Res *54*, 2615-2621.

1167 Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvak, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like
1168 transposon from fish, and its transposition in human cells. Cell *91*, 501-510.

1169 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.*
1170 (2018). Crebbp Loss Drives Small Cell Lung Cancer and Increases Sensitivity to HDAC Inhibition. Cancer Discov
1171 *8*, 1422-1437.

1172 Jonkers, J., Meuwissen, R., van der Gulden, H., Peterse, H., van der Valk, M., and Berns, A. (2001). Synergistic tumor
1173 suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. Nat Genet *29*, 418-425.

1174 Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and
1175 bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat
1176 Genet *45*, 592-601.

1177 Kahyo, T., Nishida, T., and Yasuda, H. (2001). Involvement of PIAS1 in the sumoylation of tumor suppressor p53. Mol
1178 Cell *8*, 713-718.

1179 Knatko, E.V., Praslicka, B., Higgins, M., Evans, A., Purdie, K.J., Harwood, C.A., Proby, C.M., Ooi, A., and Dinkova-Kostova,
1180 A.T. (2017). Whole-Exome Sequencing Validates a Preclinical Mouse Model for the Prevention and Treatment of
1181 Cutaneous Squamous Cell Carcinoma. Cancer Prev Res (Phila) *10*, 67-75.

1182 Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M.,
1183 Lachmann, A., *et al.* (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.
1184 Nucleic Acids Res *44*, W90-97.

1185 Lambert, S.R., Mladkova, N., Gulati, A., Hamoudi, R., Purdie, K., Cerio, R., Leigh, I., Proby, C., and Harwood, C.A. (2014).
1186 Key differences identified between actinic keratosis and cutaneous squamous cell carcinoma by transcriptome
1187 profiling. Br J Cancer *110*, 520-529.

1188 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,
1189 L.C., *et al.* (2004). Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. Cell
1190 *119*, 861-872.

1191 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

1192 Lee, C.S., Bhaduri, A., Mah, A., Johnson, W.L., Ungewickell, A., Aros, C.J., Nguyen, C.B., Rios, E.J., Siprashvili, Z., Straight,
1193 A., *et al.* (2014). Recurrent point mutations in the kinetochore gene KNSTRN in cutaneous squamous cell
1194 carcinoma. Nat Genet *46*, 1060-1062.

1195 Lee, S.H., Zhu, C., Peng, Y., Johnson, D.T., Lehmann, L., and Sun, Z. (2013). Identification of a novel role of ZMIZ2
1196 protein in regulating the activity of the Wnt/beta-catenin signaling pathway. J Biol Chem *288*, 35913-35924.

1197 Lewandoski, M., Meyers, E.N., and Martin, G.R. (1997). Analysis of Fgf8 gene function in vertebrate development.
1198 Cold Spring Harb Symp Quant Biol *62*, 159-168.

1199 Liu, B., Xia, X., Zhu, F., Park, E., Carbajal, S., Kiguchi, K., DiGiovanni, J., Fischer, S.M., and Hu, Y. (2008). IKKalpha is
1200 required to maintain skin homeostasis and prevent skin cancer. Cancer Cell *14*, 212-225.

1201 Mann, K.M., Jenkins, N.A., Copeland, N.G., and Mann, M.B. (2013). Transposon Insertional Mutagenesis Models of
1202 Cancer. Cold Spring Harb Protoc.

1203 Mann, K.M., Jenkins, N.A., Copeland, N.G., and Mann, M.B. (2014a). Transposon insertional mutagenesis models of
1204 cancer. Cold Spring Harb Protoc *2014*, 235-247.

1205 Mann, K.M., Newberg, J.Y., Black, M.A., Jones, D.J., Amaya-Manzanares, F., Guzman-Rojas, L., Kodama, T., Ward, J.M.,
1206 Rust, A.G., van der Weyden, L., *et al.* (2016). Analyzing tumor heterogeneity and driver genes in single myeloid
1207 leukemia cells with SBCapSeq. Nat Biotechnol *34*, 962-972.

1208 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,
1209 D.K., *et al.* (2012). Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic
1210 adenocarcinoma. Proc Natl Acad Sci U S A *109*, 5934-5941.

1211 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.,
1212 McMahon, M., *et al.* (2015). Transposon mutagenesis identifies genetic drivers of Braf(V600E) melanoma. Nat
1213 Genet *47*, 486-495.

1214 Mann, M.B., Jenkins, N.A., Copeland, N.G., and Mann, K.M. (2014b). Sleeping Beauty mutagenesis: exploiting forward
1215 genetic screens for cancer gene discovery. Curr Opin Genet Dev *24*, 16-22.

1216 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.,
1217 Uren, A., *et al.* (2011). Insertional mutagenesis identifies multiple networks of cooperating genes driving
1218 intestinal tumorigenesis. Nature genetics *43*, 1202-1209.

1219 Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B.,
1220 Tubio, J.M., *et al.* (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in
1221 normal human skin. Science *348*, 880-886.

1222 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).
1223 Genome-wide investigation of intragenic DNA methylation identifies ZMIZ1 gene as a prognostic marker in
1224 glioblastoma and multiple cancer types. Int J Cancer.

1225 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,
1226 C.W. (2017). ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. Nat
1227 Genet *49*, 296-302.

1228 Megidish, T., Xu, J.H., and Xu, C.W. (2002). Activation of p53 by protein inhibitor of activated Stat1 (PIAS1). J Biol
1229 Chem *277*, 8255-8259.

1230 Morton, J.P., Timpson, P., Karim, S.A., Ridgway, R.A., Athineos, D., Doyle, B., Jamieson, N.B., Oien, K.A., Lowy, A.M.,
1231 Brunton, V.G., *et al.* (2010). Mutant p53 drives metastasis and overcomes growth arrest/senescence in
1232 pancreatic cancer. Proc Natl Acad Sci U S A *107*, 246-251.

1233 Nadeau, J.H., and Dudley, A.M. (2011). Genetics. Systems genetics. Science *331*, 1015-1016.

1234 Nassar, D., Latil, M., Boeckx, B., Lambrechts, D., and Blanpain, C. (2015). Genomic landscape of carcinogen-induced
1235 and genetically induced mouse skin squamous cell carcinoma. Nat Med *21*, 946-954.

1236 Newberg, J.Y., Black, M.A., Jenkins, N.A., Copeland, N.G., Mann, K.M., and Mann, M.B. (2018a). SB Driver Analysis: a
1237 Sleeping Beauty cancer driver analysis framework for identifying and prioritizing experimentally actionable
1238 oncogenes and tumor suppressors. Nucleic Acids Res *46*, e94.

1239 Newberg, J.Y., Black, M.A., Jenkins, N.A., Copeland, N.G., Mann, K.M., and Mann, M.B. (2018b). SB Driver Analysis: a
1240 Sleeping Beauty cancer driver analysis framework for identifying and prioritizing experimentally actionable
1241 oncogenes and tumor suppressors. Nucleic Acids Res.

1242 Newberg, J.Y., Mann, K.M., Mann, M.B., Jenkins, N.A., and Copeland, N.G. (2017). SBCDDB: Sleeping Beauty Cancer
1243 Driver Database for gene discovery in mouse models of human cancers. Nucleic Acids Res.

1244 Newberg, J.Y., Mann, K.M., Mann, M.B., Jenkins, N.A., and Copeland, N.G. (2018c). SBCDDB: Sleeping Beauty Cancer
1245 Driver Database for gene discovery in mouse models of human cancers. Nucleic Acids Res *46*, D1011-D1017.

1246 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
1247 gain of function in two mouse models of Li-Fraumeni syndrome. Cell *119*, 847-860.

1248 Parman, C., Halling, C., and Gentleman, R. (2005). affyQCReport: QC Report Generation for affyBatch objects. R
1249 package version *1*, 1.

1250 Perez-Mancera, P.A., Rust, A.G., van der Weyden, L., Kristiansen, G., Li, A., Sarver, A.L., Silverstein, K.A., Grutzmann, R.,
1251 Aust, D., Rummele, P., *et al.* (2012). The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma.
1252 Nature *486*, 266-270.

1253 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.,
1254 *et al.* (2014). Mutational landscape of aggressive cutaneous squamous cell carcinoma. Clin Cancer Res *20*, 6582-
1255 6592.

1256 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
1257 PIAS-like Coactivator Zmiz1 Is a Direct and Selective Cofactor of Notch1 in T Cell Development and Leukemia.
1258 Immunity *43*, 870-883.

1259 Potla, L., Boghaert, E.R., Armellino, D., Frost, P., and Damle, N.K. (2002). Reduced expression of EphrinA1 (EFNA1)
1260 inhibits three-dimensional growth of HT29 colon carcinoma cells. Cancer Lett *175*, 187-195.

1261 R_Core_Team (2008; 2013). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation
1262 for Statistical Computing).

1263 Ramirez, A., Page, A., Gandarillas, A., Zanet, J., Pibre, S., Vidal, M., Tusell, L., Genesca, A., Whitaker, D.A., Melton, D.W.,
1264 *et al.* (2004). A keratin K5Cre transgenic line appropriate for tissue-specific or generalized Cre-mediated
1265 recombination. Genesis *39*, 52-57.

1266 Rangel, R., Lee, S.C., Hon-Kim Ban, K., Guzman-Rojas, L., Mann, M.B., Newberg, J.Y., Kodama, T., McNoe, L.A.,
1267 Selvanesan, L., Ward, J.M., *et al.* (2016). Transposon mutagenesis identifies genes that cooperate with mutant
1268 Pten in breast cancer progression. Proc Natl Acad Sci U S A *113*, E7749-E7758.

1269 Rheinwald, J.G., and Beckett, M.A. (1981). Tumorigenic keratinocyte lines requiring anchorage and fibroblast
1270 support cultured from human squamous cell carcinomas. Cancer Res *41*, 1657-1663.

1271 Rheinwald, J.G., and Green, H. (1977). Epidermal growth factor and the multiplication of cultured human epidermal
1272 keratinocytes. Nature *265*, 421-424.

1273 Rogers, L.M., Riordan, J.D., Swick, B.L., Meyerholz, D.K., and Dupuy, A.J. (2013). Ectopic expression of Zmiz1 induces
1274 cutaneous squamous cell malignancies in a mouse model of cancer. J Invest Dermatol *133*, 1863-1869.

1275 Schmidt, D., and Muller, S. (2002). Members of the PIAS family act as SUMO ligases for c-Jun and p53 and repress p53
1276 activity. Proc Natl Acad Sci U S A *99*, 2872-2877.

1277 Schmidt, D., and Muller, S. (2003). PIAS/SUMO: new partners in transcriptional regulation. Cell Mol Life Sci *60*, 2561-
1278 2574.

1279 Sharma, M., Li, X., Wang, Y., Zarnegar, M., Huang, C.Y., Palvimo, J.J., Lim, B., and Sun, Z. (2003). hZimp10 is an
1280 androgen receptor co-activator and forms a complex with SUMO-1 at replication foci. EMBO J *22*, 6101-6114.

1281 Soler, G., Radford-Weiss, I., Ben-Abdelali, R., Mahlaoui, N., Ponceau, J.F., Macintyre, E.A., Vekemans, M., Bernard, O.A.,
1282 and Romana, S.P. (2008). Fusion of ZMIZ1 to ABL1 in a B-cell acute lymphoblastic leukaemia with a
1283 t(9;10)(q34;q22.3) translocation. Leukemia *22*, 1278-1280.

1284 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.,
1285 McHugh, A., *et al.* (2014). NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. J
1286 Invest Dermatol *134*, 2630-2638.

1287 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.,
1288 Dupuy, A.J., *et al.* (2009). A transposon-based genetic screen in mice identifies genes altered in colorectal cancer.
1289 Science *323*, 1747-1750.

1290 Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G.V., Lawrence, M.S.,
1291 Sougnez, C., McKenna, A., *et al.* (2011). The mutational landscape of head and neck squamous cell carcinoma.
1292 Science *333*, 1157-1160.

1293 Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H.,
1294 Bork, P., *et al.* (2019). STRING v11: protein-protein association networks with increased coverage, supporting
1295 functional discovery in genome-wide experimental datasets. Nucleic Acids Res *47*, D607-D613.

1296 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.
1297 (2015). Transposon mutagenesis identifies genes and evolutionary forces driving gastrointestinal tract tumor
1298 progression. Nat Genet *47*, 142-150.

1299 Team, R.C. (2013). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for
1300 Statistical Computing).

1301 Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics
1302 *25*, 1105-1111.

1303 Vin, H., Ojeda, S.S., Ching, G., Leung, M.L., Chitsazzadeh, V., Dwyer, D.W., Adelmann, C.H., Restrepo, M., Richards, K.N.,
1304 Stewart, L.R.*, et al.* (2013). BRAF inhibitors suppress apoptosis through off-target inhibition of JNK signaling.
1305 Elife *2*, e00969.

1306 Wang, L., Zhao, Z., Ozark, P.A., Fantini, D., Marshall, S.A., Rendleman, E.J., Cozzolino, K.A., Louis, N., He, X., Morgan,
1307 M.A.*, et al.* (2018). Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer
1308 therapy. Nat Med *24*, 758-769.

1309 Wickham, H. (2009). Ggplot2 : elegant graphics for data analysis (New York: Springer).

1310 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
1311 NCOA2 by recurrent fusion with LACTB2 in colorectal cancer. Oncogene *35*, 187-195.

1312 Zhang, C., Spevak, W., Zhang, Y., Burton, E.A., Ma, Y., Habets, G., Zhang, J., Lin, J., Ewing, T., Matusow, B.*, et al.* (2015).
1313 RAF inhibitors that evade paradoxical MAPK pathway activation. Nature *526*, 583-586.

1314