RESOURCE

nature medicine

Genomic landscape of carcinogen-induced and genetically induced mouse skin squamous cell carcinoma

Dany Nassar[1,5], Mathilde Latil[1,5], Bram Boeckx[2,3,5], Diether Lambrechts[2,3,6] & Cédric Blanpain[1,4,6]

Mouse models of cancers are routinely used to study cancer biology. However, it remains unclear whether carcinogenesis in mice is driven by the same spectrum of genomic alterations found in humans. Here we conducted a comprehensive genomic analysis of 9,10-dimethyl-1,2-benzanthracene (DMBA)-induced skin cancer, the most commonly used skin cancer model, which appears as benign papillomas that progress into squamous cell carcinomas (SCCs). We also studied genetically induced SCCs that expressed G12D mutant Kras (Kras G12D) but were deficient for p53. Using whole-exome sequencing, we uncovered a characteristic mutational signature of DMBA-induced SCCs. We found that the vast majority of DMBA-induced SCCs presented recurrent mutations in *Hras*, *Kras* or *Rras2* and mutations in several additional putative oncogenes and tumor-suppressor genes. Similar genes were recurrently mutated in mouse and human SCCs that were from different organs or had been exposed to different carcinogens. Invasive SCCs, but not papillomas, presented substantial chromosomal aberrations, especially in DMBA-induced and genetically induced *Trp53*-mutated SCCs. Metastasis occurred through sequential spreading, with relatively few additional genetic events. This study provides a framework for future functional cancer genomic studies in mice.

Epidemiological studies indicate that the vast majority of human cancers are associated with environmental and occupational exposures[1]. SCCs in various locations and tissues such as head and neck, lung, esophagus and skin can be induced by carcinogens such as tobacco, alcohol and ultraviolet radiation[2]. Mouse models of carcinogen-induced SCCs have been used since the early 20th century and have become the most extensively used model for studying cancer *in vivo*[1]. The procedure most frequently used to induce carcinogenesis is topical application of DMBA, a polycyclic aromatic hydrocarbon that induces DNA alterations, followed by tetradecanoyl-phorbol acetate (TPA), which stimulates inflammation and epidermal proliferation[3]. In this model, during the course of TPA administration, benign tumors (papillomas) arise, some of which progress to invasive SCCs. It is generally thought that during the course of TPA administration, additional genomic alterations occur that lead to malignant progression[3].

Balmain and colleagues discovered that the majority of DMBA-induced papillomas and mouse skin SCCs (msSCCs) contained the same point mutation in *Hras*, $Hras^{Q61L}$, which is frequently associated with *Hras* gene amplification. This suggested that *Hras* was an early and key driver in the transformation of msSCCs[4–6]. In addition, *Trp53* mutations are found in 25%–50% of malignant SCCs, but not in papillomas[7,8]. Cytogenetic analysis and low-resolution comparative genomic hybridization array revealed trisomy of chromosomes 6 and 7 in most msSCCs[9] and focal amplification of the distal region in chromosome 7 containing *Hras* and *Ccnd1* (ref. 10). However, it remains unknown whether additional somatic mutations besides those in *Hras* and *Trp53* or additional gene copy-number alterations (CNAs) are responsible for the initiation and progression of DMBA-induced msSCCs. Also, the extent of the similarity between mouse models of chemically induced cancers and their human counterparts in terms of the mutational landscape remains unclear.

In addition to mouse models of carcinogen-induced cancer, genetically engineered mouse models of cancer are widely used to study skin SCC. The combination of $Kras^{G12D}$ expression and *Trp53* deletion is required for malignant progression of msSCC[11,12]. However, the additional genomic alterations that promote tumorigenesis in this genetic model remain unknown.

Here we defined the genomic landscape of both genetically induced and DMBA-induced msSCCs at different stages of tumor evolution ranging from benign papillomas to fully malignant SCCs and their metastases. We uncovered a characteristic mutational signature associated with DMBA-induced skin tumorigenesis and identified several recurrently mutated genes that are likely drivers of tumorigenesis, and our findings demonstrate the importance of chromosomal alterations during malignant progression. The identification of the key genetic events during mouse skin tumorigenesis will be useful for rationalizing future functional mouse genetic studies and determining which genes are most likely to present conserved functions in mouse and human cancers.

RESULTS

Whole-exome sequencing of DMBA-induced skin msSCC

To determine the somatic changes underlying msSCC development, we used FACS to isolate tumor epithelial cells (TECs) with

[1]Institut de recherche interdisciplinaire en biologie humaine et moléculaire (IRIBHM), Université libre de Buxelles (ULB), Brussels, Belgium. [2]Vesalius Research Center, Vlaams Instituut voor Biotechnologie VIB, Leuven, Belgium. [3]Laboratory for Translational Genetics, Department of Oncology, Katholieke Universiteit Leuven (KUL), Leuven, Belgium. [4]WELBIO, Brussels, Belgium. [5]These authors contributed equally to this work. [6]These authors jointly directed this work. Correspondence should be addressed to C.B. (Cedric.Blanpain@ulb.ac.be) or D.L. (Diether.Lambrechts@vib-kuleuven.be).

Received 22 October 2014; accepted 18 May 2015; published online 13 July 2015; doi:10.1038/nm.3878; corrected after print 6 August 2015

high purity[13] from DMBA-induced msSCCs in the tumor-sensitive FVB/N inbred mouse strain and performed whole-exome sequencing (WES) and low-depth whole-genome sequencing (shallow WGS) (**Supplementary Fig. 1a**). WES was done on 27 primary DMBA-induced msSCCs, 20 of which were induced by a single DMBA application and 7 by multiple DMBA applications (**Fig. 1a**,**b**). We also sequenced six matched bone marrow samples. After completing mapping and variant calling relative to the GRCm38/mm10 mouse reference genome, we selected somatic mutations by removing all variants reported in the FVBN/NJ variant database or detected in one of the six FVB/N bone marrow samples. Overall, we identified 17,395 nonsynonymous somatic mutations in the exomes of 27 DMBA msSCCs. A random selection of these was validated independently with Sequenom MassARRAY, which showed a true positive rate of 94% (**Supplementary Data**). Notably, the number of nonsynonymous somatic events per tumor in msSCCs that arose from a single DMBA application (mean, 431 mutations per tumor; $n = 20$) was less than the number in tumors arising from multiple DMBA applications (mean, 1,253 mutations per tumor; $n = 7$, $P < 1 \times 10^{-4}$). The cancer cell fraction (CCF), which is the percentage of cancer cells carrying a given mutation, was ≥80% for ≥80% of mutations as determined by ABSOLUTE; this indicated that the vast majority of mutations were present in the majority of cancer cells (**Supplementary Fig. 2a**).

msSCCs with a unique mutational signature

Different cancers present with different mutational signatures depending on the types of carcinogens to which they have been exposed[14]. Here we assessed the mutational signature of DMBA-induced msSCCs. DMBA has been reported to induce substitutions affecting adenine (A) and guanine (G) residues[15], inducing the A-to-T (A>T) hotspot transversion in *Hras*[5]. Our WES analysis showed that A>T transversions were by far the most common type of event, contributing to >60% of all detected mutations (**Fig. 1c**,**d**). This signature was different from any other signature observed in human cancer[14]. G>T transversions accounted for another 15% of the mutations detected in DMBA-induced msSCC (**Fig. 1c**,**d**). Together, A>T and G>T substitutions represented nearly 80% of all somatic mutations. We found that A>T transversions occurred more frequently when the A (or T) was preceded by a C (or G) and followed by a G (or C) (**Fig. 1e–g**). Notably, this context-dependent effect on the mutation signature was identical in mice receiving one DMBA application and mice receiving more than one (**Fig. 1e–g**). This effect can possibly be explained by the fact that CAG (or GTC) trinucleotides are more accessible to mutagens or by the lower efficiency of the base excision-repair machinery that repairs lesions in this specific nucleotide context[16]. The characteristic and rather unusual mutation signature of DMBA-induced tumors further indicated that most nonsynonymous mutations occurred at the time of DMBA application, even when DMBA was applied only once.

Mutations in three *Ras* family members are the most common drivers in msSCC

To identify putative driver mutations in msSCC, we searched for genes mutated more frequently than expected given the average observed mutation frequency and the length of a given gene (χ^2 analysis). We also used the Genome Mutational Significance in Cancer (Genome MuSiC) suite of tools, which uses the observed mutation signature and background mutation rates to identify significantly mutated genes. To further select driver genes, we eliminated genes that were not expressed in DMBA msSCC ($n = 6$) or in normal epidermis ($n = 5$) and considered only genes that were expressed in at least six samples (msSCCs or normal epidermis). Overall, χ^2 analysis identified 279 significantly mutated genes at a false discovery rate (FDR) of <0.1, 78 of which were expressed in either msSCC or normal epidermis

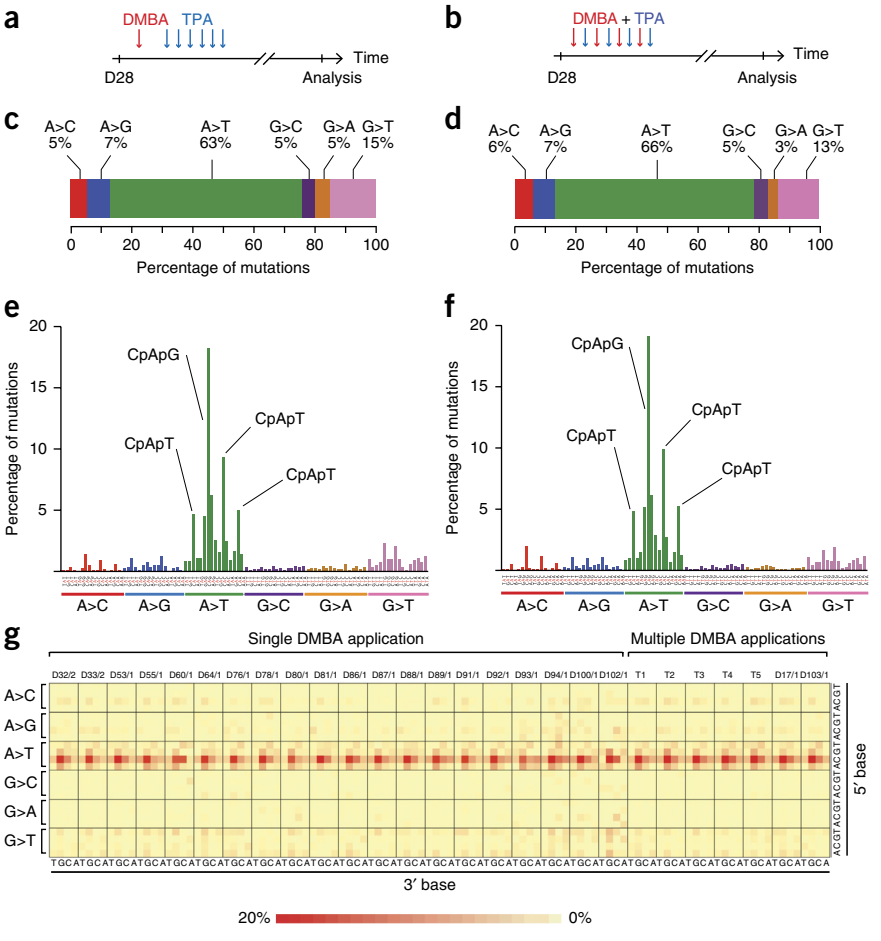

Figure 1 Mutational signature of DMBA-TPA–induced msSCCs. (**a**,**b**) msSCCs were induced in FVB/N mice by single ($n = 7$; **a**) and multiple ($n = 20$; **b**) DMBA applications. (**c**,**d**) Frequency of the different nonsynonymous single-nucleotide substitutions found in msSCCs arising from single (**c**) and multiple (**d**) DMBA applications. (**e**,**f**) The mutation signature retrieved for msSCCs arising from single (**e**) or multiple (**f**) DMBA applications. Frequencies for each of the six substitutions are displayed in different colors and broken into 16 different permutations according to the neighboring 5′ and 3′ bases. The predominant nucleotide context for the A>T transversion is highlighted on the histograms. (**g**) Heat map representing the frequency of each substitution stratified for the preceding (adjacent 5′) and following (adjacent 3′) bases in DMBA-induced msSCCs.

(**Supplementary Data**). Genome MuSiC identified 26 genes with FDR < 0.1 (**Supplementary Data**); 16 of these were also identified in the χ^2 analysis (FDR < 0.1). Like in previous studies involving mouse tumor WES data[17], MutSigCV was too stringent, identifying *Hras* as the only significantly mutated gene (FDR = 1.5×10^{-5}) (**Supplementary Data**). *Kras* and *Trp53* were among the ten highest-ranking genes in MutSigCV analysis but did not remain in the list after correction for multiple testing. We therefore considered 81 expressed genes, containing 477 mutations in total, that were identified by either GenomeMuSiC or χ^2 analysis (FDR < 0.1) as significantly mutated. Functional prediction with PROVEAN showed that more than 60%–80% of the mutations in expressed genes with FDR < 0.1 according to χ^2 analysis or Genome MuSiC were predicted to be functionally disruptive, as compared to 45% of mutations in a gene with FDR > 0.2 (**Supplementary Fig. 2b–e**). Moreover, the ratio of synonymous mutations to nonsynonymous mutations in genes identified with FDR < 0.2 was strongly enriched for nonsynonymous mutations, further supporting the idea that these mutations are more likely to act as drivers than as passenger mutations (**Supplementary Fig. 2f,g**).

The highest-ranked gene was *Hras*, which was mutated in 67% of DMBA msSCCs, with all tumors presenting the activating mutation Q61L[5] (**Fig. 2a**). *Kras* mutations were found in 19% of the DMBA msSCCs, representing >50% of *Hras* mutation–negative tumors (five out of nine tumors). All detected *Kras* mutations (i.e., G13R, Q61L and Q61H; **Fig. 2a**) have been previously identified in DMBA msSCCs and shown to confer constitutive activation of the Ras pathway[18]. In addition, we identified a substitution in another Ras member (*Rras2*) that induced a similar mutation (*Rras2*^{Q72L}). That mutation had not previously been found in msSCC but exhibited oncogenic potential *in vitro* similar to that of *Hras*^{Q61L} (ref. 19) (**Fig. 2a,b**). Although mutations in *Kras* and *Rras2* were not identified as significantly enriched by Genome MuSiC (FDR = 0.30 and 1.0, respectively) and those in *Rras2* were not enriched according to χ^2 analysis (FDR = 1.0) (**Fig. 2c**) because of the high frequency of mutations in DMBA-induced msSCCs, these mutations are clearly biologically relevant, as previously demonstrated by their oncogenic properties *in vitro* and *in vivo*[18,19]. Furthermore, mutual-exclusivity analysis identified *Hras* and *Kras* mutations as the most significant pair ($P = 6 \times 10^{-4}$) (**Fig. 2c** and **Supplementary Data**). Overall, these data indicate that activating mutations are found in *Hras*, *Kras* or *Rras2* in a mutually exclusive manner in 90% of DMBA-induced msSCCs.

Novel putative drivers of msSCC

In addition to *Hras* mutations, mutations in the tumor-suppressor gene (TSG) *Trp53* have been reported in msSCC[7,8]. In our study, *Trp53* was the second most significantly mutated gene (**Fig. 2c**). Point mutations in *Trp53* were found in 30% of msSCCs, with an additional deletion inducing *Trp53* loss of heterozygosity in 25% of those tumors. All *Trp53* mutations occurred in conserved amino acids located in the gene's DNA-binding domain, including truncating mutations and mutations located at previously identified hotspots such as R273H (R270H in humans). The majority (70%) of *Trp53* mutations did not have the DMBA signature, which suggested that *Trp53* mutations occurred later in tumor progression (**Supplementary Fig. 3a**).

We identified recurrent mutations with FDR < 0.1 by χ^2 analysis and FDR < 0.25 by Genome MuSiC in several functionally demonstrated TSGs, including *Myh9* (ref. 20) (22%), *Blm*[21] (22%), *Notch1* (ref. 22) (30%), *Ptprm*[23] (22%) and *Rpl36* (ref. 24) (7%). These mutations often affected functional domains previously noted in human SCC (**Fig. 2c**, **Supplementary Fig. 3a–e** and **Supplementary Data**). Histone modifiers and epigenetic regulators including *Gltscr1* (26%), *Nsd1* (ref. 25) (7%) and *Vprbp* (22%) were recurrently mutated in msSCCs in similar functional domains as in human cancers (**Fig. 2c**, **Supplementary Fig. 3f–h** and **Supplementary Data**). Genes regulating cell adhesion, polarity and cytoskeleton and implicated in cancer, such as *Nav3* (ref. 26) (26%), *Slit3* (ref. 27) (26%), *Ap1m2* (ref. 28) (26%), *Wwc1* (ref. 29) (26%) and *Palld*[30] (26%), were also recurrently mutated in msSCCs, as were genes regulating drug resistance (*Abcb1a*, *Sult1a1* and *Gsto2*) and RNA metabolism (*Sf3a3* and *Rexo1*) (**Fig. 2c**, **Supplementary Fig. 3i** and **Supplementary Data**).

The mouse SCC mutational landscape is similar to that of human SCC

We then assessed whether the somatic mutations identified in DMBA-induced msSCC were similar to the mutations found in human SCCs from different body locations such as skin[31–34], head and neck[25,35,36], lung[37], esophagus[38,39], mouth[40], cervix[41] and nasopharynx[42]. Many mouse genes with human counterparts reported to be recurrently mutated in human SCCs from different organs and exposed to different carcinogens, such as *Hras*, *Kras*, *Trp53*, *Fat1*, *Notch1*, *Notch2*, *Unc13c*, *Ash1l*, *Kmt2d*, *Kmt2c*, *Ep300*, *Setd2*, *Usp9x*, *Arid1a*, *Arid2*, *Smg1*, *Nsd1*, *Nfe2l2*, *Syne1-2* and *Tgfbr2* (refs. 25,31–42), were also frequently mutated in DMBA msSCCs (**Fig. 2d**).

Next we assessed more systematically whether genes that are mutated in human SCC are indeed significantly enriched in DMBA msSCC. Specifically, we calculated the overlap between genes recurrently (>10%) mutated in msSCC and human SCCs such as head and neck SCC[25], esophageal SCC[39], lung SCC[37] and cervical SCC (http://www.cbioportal.org/study.do?cancer_study_id=cesc_tcga) while considering three unrelated cancers (acute myeloid leukemia (http://www.cbioportal.org/study.do?cancer_study_id=laml_tcga), melanoma (http://www.cbioportal.org/study.do?cancer_study_id=skcm_tcga) and lung adenocarcinoma[43]) (**Supplementary Data**). We observed significantly more overlapping genes between msSCCs and human SCCs than between msSCCs and unrelated cancers; specifically, 35%, 44%, 25% and 18% of genes overlapped between msSCC and human SCC from head and neck ($P = 5.8 \times 10^{-5}$), cervix ($P = 2.6 \times 10^{-5}$), esophagus ($P = 0.026$) and lung ($P = 5.9 \times 10^{-5}$), respectively, as compared to 0% overlap with acute myeloid leukemia ($P = 1$), 14% overlap with melanoma ($P = 0.015$) and 17% overlap with lung adenocarcinoma ($P = 3.1 \times 10^{-5}$), with the last two cancers induced by similar carcinogens. These data suggest significant similarity between recurrently mutated genes in DMBA-induced msSCC and human SCCs of different origins and exposed to different carcinogens.

Recurrent CNAs in msSCC

To define the role of chromosomal aberrations in DMBA msSCCs, we assessed somatic DNA CNAs via shallow WGS of nine premalignant papillomas and 24 out of 27 msSCCs (comparative genomic hybridization array data were available for the remaining 3 msSCCs). Although 80% of the benign papillomas did not contain any CNAs, all msSCCs were aneuploid with frequent whole-arm or whole-chromosome gains or losses, as well as focal CNAs (**Fig. 3a,b**).

At the arm level, GISTIC analysis showed that chromosomes 6 (harboring *Kras*) and 15 (harboring *Myc*) were the most frequently amplified chromosomes in 48% and 44% of tumors, respectively (**Fig. 3c** and **Supplementary Data**). Chromosome 9, which contains *Atm*, *Atr* and *Cdkn2d* (*p19*), was significantly deleted in 26% of msSCCs (**Fig. 3c** and **Supplementary Data**). GISTIC analysis of focal CNAs identified amplified regions on 7qB5, containing *Herc2*, a gene

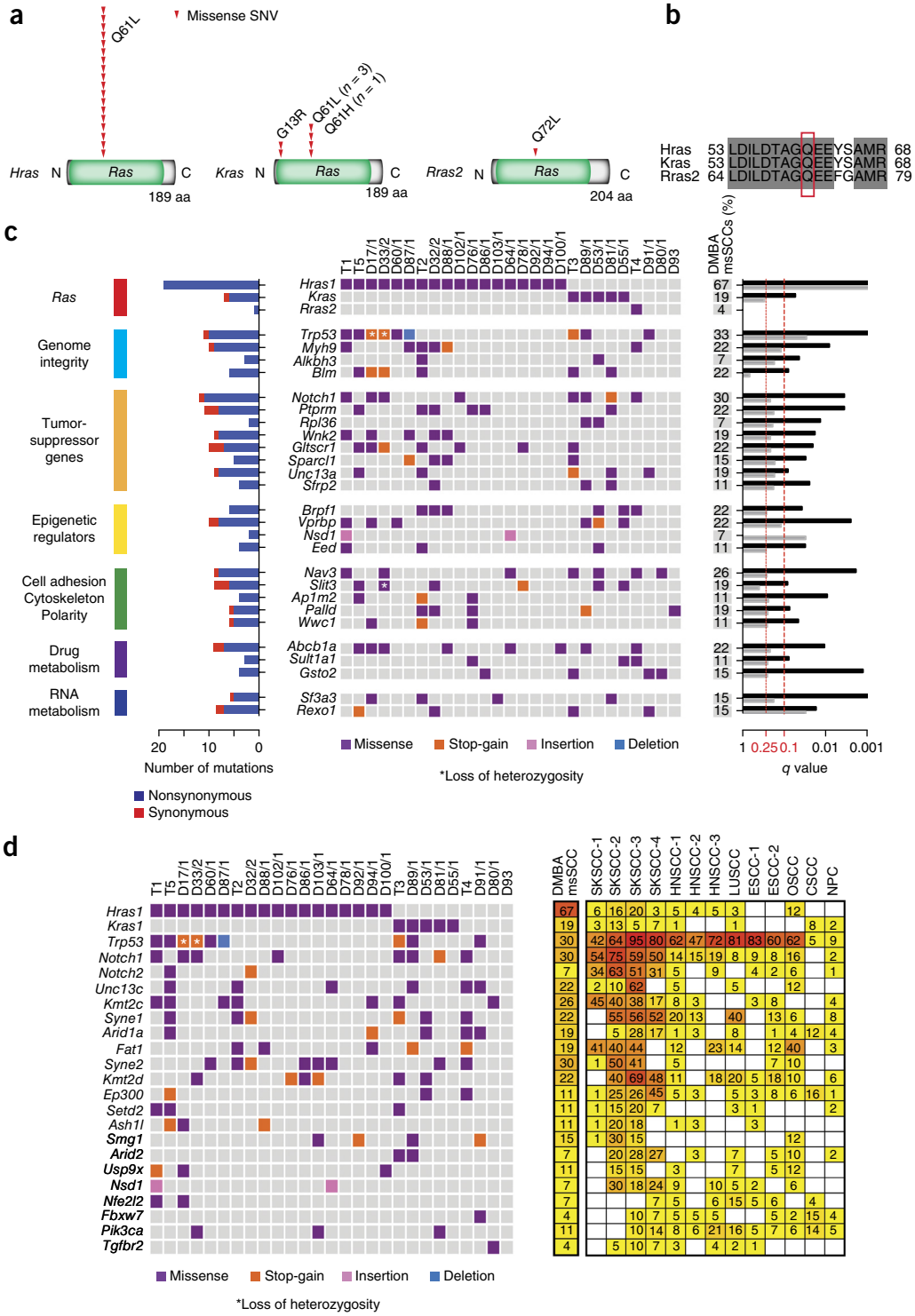

Figure 2 Recurrent somatic mutations in DMBA-induced msSCCs. (**a**,**b**) Prevalence and localization of somatic mutations in *Hras*, *Kras* and *Rras2*. (**a**) Previously described $Hras^{Q61L}$, $Kras^{Q61L}$, $Kras^{Q61H}$ and $Kras^{G13R}$ activating mutations were identified in 23 of 27 msSCCs. A newly identified $Rras2^{Q72L}$ mutation was found that corresponded to amino acid position 61 in the Ras domain of Hras and Kras, as shown by protein sequence alignment (**b**). (**c**) Data matrix showing *Ras* mutations occurring in msSCCs in a mutually exclusive manner and significantly mutated genes in msSCCs classified according to their function. Somatic mutations are presented according to the type of mutation (missense, stop-gain, insertion or deletion; colors defined in key). The proportion of msSCC tumors exhibiting mutations in the corresponding gene is shown on the right, along with FDR *q* values as determined by Genome MuSiC (gray bars) and χ^2 (black bars) statistical tests. Genes with an FDR $q < 0.1$ for χ^2 and $q < 0.25$ for Genome MuSiC were considered significant. The number of synonymous and nonsynonymous mutations for each gene is shown to the left of the histogram. (**d**) Data matrix showing the nonsynonymous substitutions found in DMBA-induced msSCCs and recurrently mutated genes in human SCCs. The proportion of msSCCs showing mutations of the corresponding gene is plotted on the right, along with the mutation frequency (%) in human SCCs from different organs. SKSCC, skin SCCs 1 (refs. 31–34), 2 (ref. 32), 3 (ref. 33) and 4 (ref. 34); HNSCC, head and neck SCCs 1 (ref. 25), 2 (ref. 35) and 3 (ref. 36); LUSCC, lung SCC[37]; ESSC, esophageal SCCs 1 (ref. 38) and 2 (ref. 39); OSCC, oral SCC[40]; CSCC, cervical SCC 1 (ref. 41); NPC, nasopharyngeal carcinoma[42].

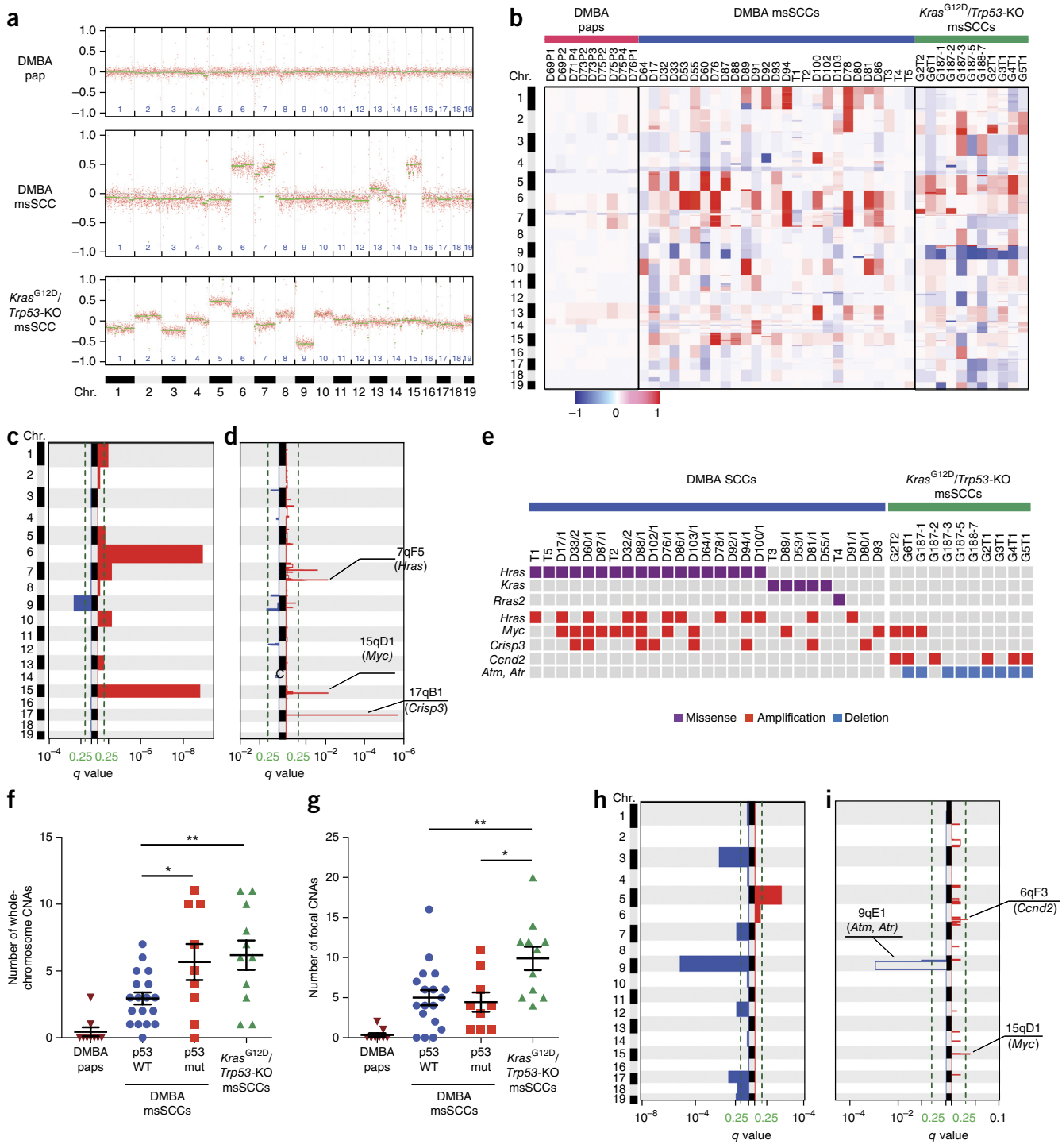

Figure 3 Recurrent CNAs in DMBA-induced and genetically induced msSCC. (**a**) Shallow WGS plots for a representative papilloma (pap; $n = 9$), DMBA msSCC ($n = 27$) and genetic *Kras*^{G12D} expression with *Trp53* knockout (*Kras*^{G12D}/*Trp53*-KO) msSCC ($n = 11$). (**b**) Heat map of copy-number gains and losses identified in DMBA-induced papillomas and msSCCs and in genetically induced *Kras*^{G12D}-expressing *Trp53*-knockout msSCCs. Red denotes a copy-number gain, and blue denotes a copy-number loss. Chromosome numbers are shown on the left. (**c**) GISTIC analysis for recurrent whole-chromosome amplification or deletion in DMBA msSCCs identified statistically significant recurrent amplification of chromosomes 6 (*Kras*) and 15 (*Myc*) and deletion of chromosome 9 (*Atm*, *Atr*, *Cdkn2d*). (**d**) GISTIC analysis of recurrent focal CNAs identified statistically significant recurrent copy-number amplifications on 7qB5, 7qF5 containing *Hras*, 15qD1 containing *c-Myc* and 17qB1 containing *Crisp3*. (**e**) Matrix showing focal CNAs in DMBA-induced and genetically induced msSCC tumors. (**f**,**g**) Number of whole-chromosome (**f**) and focal (**g**) CNAs per tumor found in DMBA-induced papillomas ($n = 9$), DMBA-induced msSCCs expressing wild-type (WT) p53 ($n = 18$), DMBA-induced msSCCs expressing mutant (mut) p53 ($n = 9$) and genetic *Kras*^{G12D} *Trp53*-knockout msSCCs ($n = 11$) (black bars denote mean ± s.e.m.; $^*P < 0.05$, $^{**}P < 0.01$, Student's *t*-test). (**h**) GISTIC analysis for recurrent whole-chromosome CNAs in genetic *Kras*^{G12D} *Trp53*-knockout msSCCs identified statistically significant deletions of chromosomes 3 and 9 and amplification of chromosome 5. (**i**) GISTIC analysis of focal CNAs in genetic *Kras*^{G12D} *Trp53*-knockout msSCCs identified statistically significant amplification of 15qD1 (*Myc*) and 6qF3 (*Ccnd2*), as well as deletion in 9qE1 (*Atm*, *Atr*).

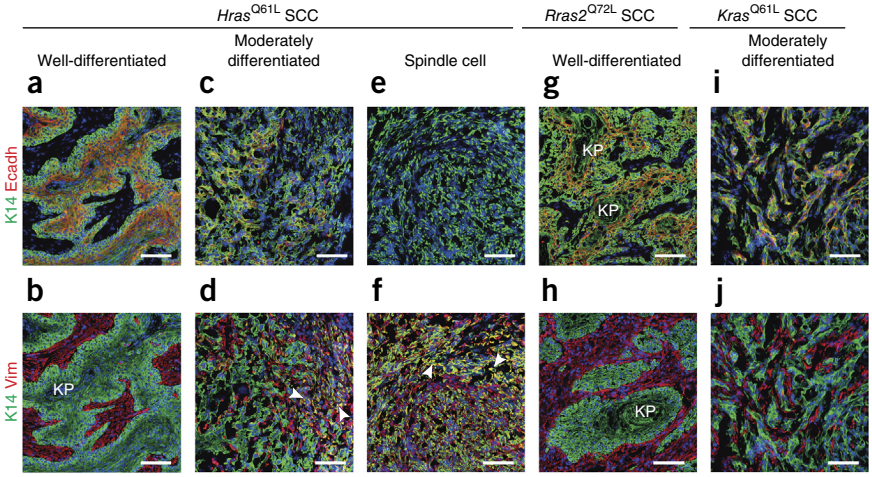

Figure 4 Histological heterogeneity of DMBA-induced msSCCs. (**a**–**j**) Immunostaining of keratin 14 (K14), E-cadherin (Ecadh) and vimentin (Vim) on *Hras*-mutated (**a**–**f**), *Rras2*-mutated (**g**,**h**) and *Kras*-mutated (**i**,**j**) DMBA msSCC. This analysis showed well-differentiated *Hras*-mutated ($n = 3$) (**a**,**b**) and *Rras2*-mutated ($n = 1$) (**g**,**h**) msSCC with keratin pearls (KP) and well-organized keratin nests expressing K14 and E-cadherin; moderately differentiated *Hras*-mutated msSCC ($n = 8$) (**c**,**d**) with more invasive and less cohesive tumor cells, some expressing the mesenchymal marker vimentin; spindle cell msSCC ($n = 1$) (**e**,**f**) that did not express E-cadherin but did express vimentin; and moderately differentiated *Kras*-mutated msSCCs ($n = 3$) (**i**,**j**) with smaller keratin nests and more invasive architecture. Arrowheads indicate double-labeled cells. Scale bars, 50 µm.

encoding a ubiquitin ligase that degrades Brca1 (ref. 44) (60%); 7qF5, containing *Hras* (44%); and 15qD1, containing the *Myc* oncogene (41%) (**Fig. 3d**,**e** and **Supplementary Data**). Amplifications of *Hras* affected 10 out of 19 msSCCs with $Hras^{Q61L}$ mutations. In two tumors the amplification specifically involved the mutated allele, supporting the idea that focal *Hras* amplifications occur after DMBA administration and control tumor progression[5]. *c-Myc* amplification has been reported in human skin SCC[45] and is associated with poorly differentiated tumors[45]. In addition, we identified a recurrent amplification of region 17qB1 (30%) containing *Crisp3* (**Fig. 3d**,**e**), a gene expressed in a subset of prostate cancers with poor prognosis[46], and *Pgk2*, which expresses an enzyme of the glycolytic pathway. Overall, these data indicate that whole-chromosome amplifications and focal CNAs are likely to influence the progression of DMBA-induced tumors by affecting regions harboring well-established and putative oncogenes.

Interestingly, the number of whole-chromosome events, but not of focal CNAs, was higher in *Trp53*-mutated tumors than in *Trp53* wild-type tumors (**Fig. 3f**,**g** and **Supplementary Data**). To validate this observation, we performed WES and shallow WGS of genetically induced msSCCs expressing $Kras^{G12D}$ and deficient for p53 (refs. 11,12). WES showed that these tumors had very few additional somatic mutations besides the initial $Kras^{G12D}$ mutation (four mutations per tumor on average) (**Supplementary Data**). However, the number of whole-chromosome events in genetically p53-deficient msSCCs was significantly higher than that in DMBA msSCCs expressing wild-type *Trp53*, but it was comparable to that in *Trp53*-mutated DMBA msSCCs (**Fig. 3f–h** and **Supplementary Data**), supporting the notion that p53 deficiency promotes aneuploidy in skin SCCs.

Correlation between mutational landscape and tumor histology

To assess whether the mutational landscape influences the tumor phenotype, we conducted detailed histological analysis of the sequenced tumors. We stained 18 tumors with various markers to assess their differentiation, invasion and proliferation status (**Fig. 4a–j** and **Supplementary Fig. 4**). All tumors were invasive carcinomas containing zones where tumor cells invaded the underlying mesenchyme. All but one tumor were histologically classified as moderately differentiated to well-differentiated SCCs with keratin pearls (**Fig. 4a–d**,**g–j** and **Supplementary Fig. 4a**,**b**). The remaining tumor was a spindle cell carcinoma showing loss of epithelial markers such as E-cadherin and strong expression of vimentin (**Fig. 4e**,**f** and **Supplementary Fig. 4a**).

Hras- and *Rras2*-mutated tumors presented with very similar histology and consisted of well-differentiated to moderately differentiated SCC (**Fig. 4a**,**b**,**g**,**h** and **Supplementary Fig. 4a**,**b**). *Kras*-mutated tumors were composed of smaller epithelial sheets that were more intermingled with mesenchymal cells (**Fig. 4i–j**), suggesting that *Kras* tumors are usually more invasive. However, the only spindle cell carcinoma was mutated for $Hras^{Q61L}$ (**Fig. 4e–f**), which suggested that *Hras* mutation can lead to very invasive tumors as well. The other tumors that did not have mutated *Ras* genes consisted of moderately differentiated SCCs expressing K14 (**Supplementary Fig. 4a**,**c**).

Clonal evolution during metastasis

It is still unclear whether the metastatic process requires the acquisition of additional genetic events or whether metastasis is mainly controlled by epigenetic and transcriptional changes. It is also not known at which stage metastasis occurs and how lymph node (LN) and lung metastasis are connected in the DMBA model. To address these questions, we compared the mutational landscape of five primary tumors and their metastases (four LN metastases and two lung metastases) sequenced at high depth. Most mutations were common between the primary tumor and the metastasis, and only very few nonsynonymous mutations or CNA changes were unique to the metastasis (**Fig. 5a** and **Supplementary Data**), which suggested that metastasis is a relatively late event and that few additional genetic changes are associated with it. Mutations unique to metastases indeed had a different mutational signature, confirming the DMBA-independent appearance of the metastatic clone (**Fig. 5b** and **Supplementary Data**). Assessment of CCFs showed that primary tumors and paired metastases consisted of a predominant clone in which shared mutations had a CCF > 95% (**Fig. 5c**). In contrast, mutations unique to the primary SCC or metastasis had a much lower CCF (~50%) (**Fig. 5c**,**d**), which enabled us to define the clonal-lineage relationship between the primary tumor and paired metastases and confirmed the late occurrence of the metastatic clone. One mouse presented with a primary SCC, a positive LN and a lung metastasis (**Fig. 5e**). The primary tumor and its metastases contained 397 shared mutations, demonstrating their common origin. The lung metastasis contained 12 mutations that were shared with the primary tumor but not with the LN. The LN metastasis did not present mutations shared exclusively with the primary tumor (i.e., that were not found in the lung metastasis) and contained only one additional mutation relative to the primary tumor (**Fig. 5f**). Overall, this suggested an early metastatic seeding event from the

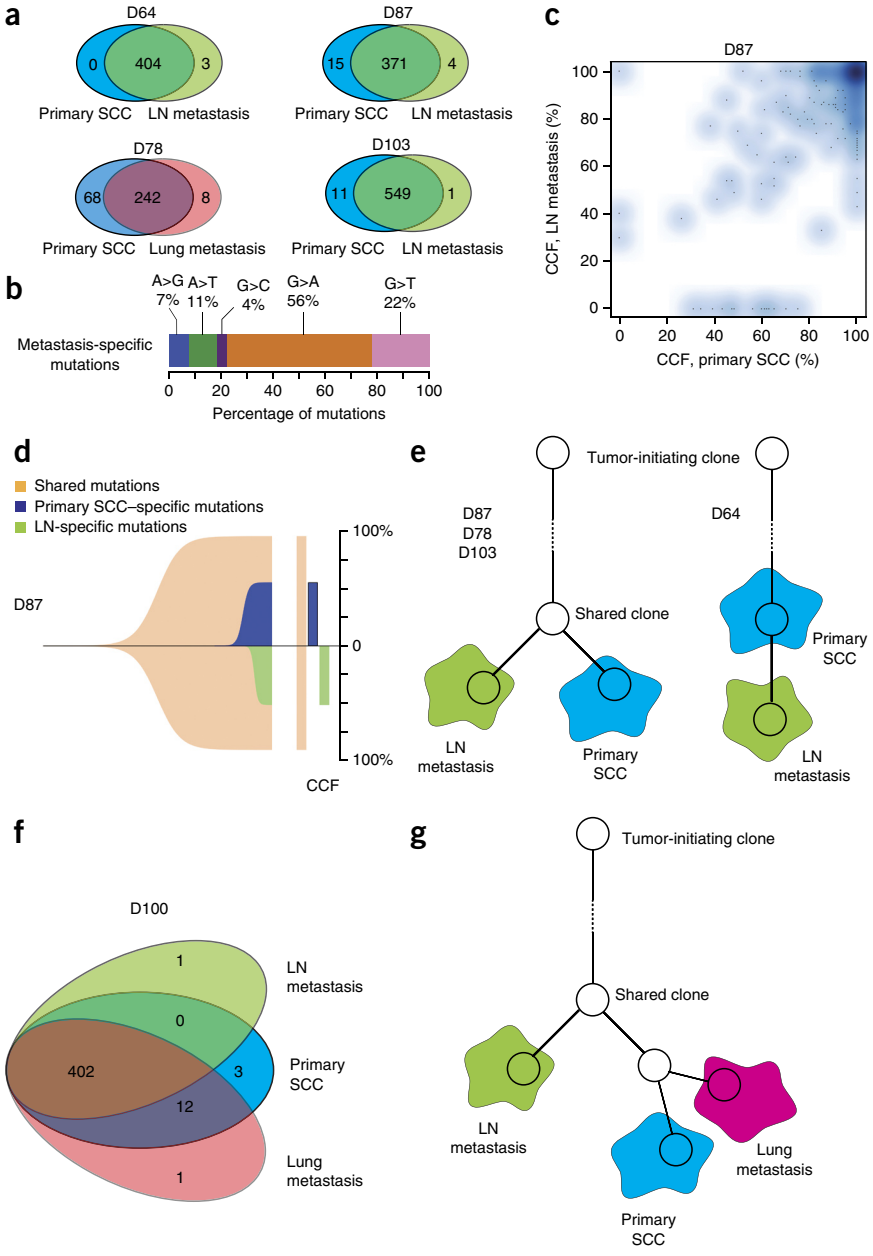

Figure 5 Clonal evolution in primary and metastatic DMBA-induced msSCC. WES was performed on primary msSCCs ($n = 5$) and paired lung or LN metastases ($n = 6$). ABSOLUTE was used to assess the mutation CCF and to model clonal evolution. (**a**) Venn diagrams representing numbers of shared and tumor-specific mutations in four primary tumors and paired LN or lung metastases (**Supplementary Data**). (**b**) Frequencies of the different metastasis-specific nonsynonymous mutations showing a different distribution than found in primary msSCCs. (**c**) CCFs of all mutations detected in a representative msSCC-LN pair (mouse D87) showing a predominant clone containing mutations in >95% of the cancer cells from the primary SCC (*x*-axis) and LN metastasis (*y*-axis). (**d**) Lineage evolution of a representative msSCC-LN pair (mouse D87) showing late appearance of the metastatic clone. (**e**) Phylogenetic trees depicting the branched evolution pattern of three msSCC-metastasis (LN or lung) pairs (mice D87, D78 and D103). (**f**) Venn diagrams representing numbers of shared and tumor-specific mutations in a primary tumor and its lung and LN metastases (**Supplementary Data**). (**g**) Phylogenetic tree depicting a parallel-branched evolution pattern of a LN metastasis and a later independent seeding event from the primary tumor to the lung.

primary SCC tumor to the LN and a later independent seeding event to the lung from a subclone of the primary tumor during lung metastasis (**Fig. 5g**).

DISCUSSION

Our study provides the first comprehensive analysis of the mutational landscape in the frequently used DMBA-induced skin SCC mouse model. DMBA induced a distinctive mutational signature suggesting that most mutations occurred at the time of DMBA administration. The mutational signature of DMBA-induced skin SCCs is very different from the mutational signature of other human cancers[14] and of mouse lung tumors induced by carcinogens such as *N*-methyl-*N*-nitrosourea (MNU)[17]. Strikingly, despite this difference in DNA assault, the mutational heterogeneity induced by these different carcinogens results in the selection of different activating variants of the same gene. For instance, MNU induces $Kras^{G12D}$ mutations, urethane leads to $Kras^{Q61R}$ mutations and DMBA induces $Kras^{Q61L}$ mutations[17].

Our analysis showed that 90% of msSCCs were driven by activating mutations in three different Ras genes; the most frequently mutated was *Hras*, followed by *Kras* and *Rras2*. In addition to activating *Ras* mutations, these tumors contained inactivating mutations in well-known TSGs such as *Trp53* and *Myh9*, or in other DNA-repair genes. A sizable proportion of tumors presented mutations in epigenetic and chromatin regulators, which could be important for reprogramming of the targeted oncogene and potentially could explain the relatively long delay between DMBA application (about 16 weeks) or genetically induced $Kras^{G12D}$ expression with *Trp53* knockout (2 months) and the appearance of skin SCC. Many genes uncovered in functional screens of skin tumorigenesis, by either whole-genome shRNA silencing (e.g., *Myh9*)[20] or transposon-induced mutagenesis (e.g., *Notch1*, *Nsd1* and *Fat1*)[47], were also frequently mutated in the DMBA-induced msSCCs, which served as functional validation of some of the genes found to be recurrently mutated in DMBA msSCC and showed that these genes are indeed *bona fide* TSGs in DMBA-induced SCC carcinogenesis. Interestingly, we found considerable similarity in the genes frequently mutated in mouse and human SCCs, suggesting that chemical-induced skin carcinogenesis, despite its completely different mutational landscape, selects for mutations in the same genes that are mutated in human SCCs. The relative frequency of these mutated genes differed between mouse and human SCCs; for example, *Trp53* mutations were more frequently found in human SCCs, whereas *Ras* mutations were more frequent in msSCCs.

In addition to somatic mutations, msSCCs also contained frequent and recurrent focal amplifications of *Hras* and *c-Myc* oncogenes, as

well as whole-chromosome duplications or deletions. Most premalignant papillomas were devoid of such CNAs, suggesting that the acquisition of chromosomal aberrations is associated with malignant progression. The whole-chromosome CNA events were strongly associated with the presence of *Trp53* mutations in the DMBA-induced msSCCs. Likewise, in *Kras*^{G12D}-expressing *Trp53*-knockout SCCs, we identified very few additional somatic mutations but noted substantial chromosomal rearrangements, as observed in p53-deficient lung tumors[48] and lymphoma[49], and this seems to be a hallmark of genetically induced p53-deficient mouse tumors.

Genomic analysis of LN and lung metastases from msSCCs showed that SCC metastasis involved very few additional mutations and/or chromosomal rearrangements compared to the primary tumors; this suggested that epigenetic, transcriptional and environmental changes could be the major determinants of the metastatic process. Future studies will be required to determine whether metastasis-associated mutations have a functional role and which mechanisms aside from genetic tumor heterogeneity control the metastatic process.

In summary, this study delineated the mutational landscape of the most common carcinogen-induced mouse tumors and the genetic changes that control tumor heterogeneity and accompany tumor initiation, progression and metastasis. This study can serve as an important resource for establishing the framework for future cancer genomics studies in mice that better mimic the genetic heterogeneity found in human cancers and that will be essential to define the mechanisms by which the newly identified recurrently mutated genes in human cancer promote tumor initiation and progression.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The raw data from whole-exome sequencing, whole-genome sequencing and RNA sequencing are available in the ArrayExpress database under accession numbers E-MTAB-2887, E-MTAB-2888 and E-MTAB-2889, respectively.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

C.B. is an investigator at Walloon Excellence in Life Sciences and Biotechnology (WELBIO). D.N. and M.L. are supported by Télévie and an EMBO long-term fellowship, respectively. The computational resources and services used in this work were provided by the Flemish Supercomputer Center (VSC), funded by the Hercules Foundation and the Flemish Government, Department of Economy, Science and Innovation (EWI). This work was supported by Fonds de la Recherche Scientifique (FNRS), Télévie, the Interuniversity Attraction Program (PAI), a research grant and an infrastructure grant from Fondation Contre le Cancer/Stichting tegen Kanker, Fondation ULB, Fond Yvonne Boël, Fond Gaston Ithier, the the Baillet Latour Foundation and the European Research Council (ERC).

AUTHOR CONTRIBUTIONS

C.B., D.N. and D.L. designed the experiments. B.B. and D.L. performed bioinformatic analysis. D.N. and M.L. performed biological experiments. All authors performed data analysis. C.B., D.L. and D.N. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Luch, A. Nature and nurture —lessons from chemical carcinogenesis. *Nat. Rev. Cancer* **5**, 113–125 (2005).
2. Poirier, M.C. Chemical-induced DNA damage and human cancer risk. *Nat. Rev. Cancer* **4**, 630–637 (2004).
3. Abel, E.L., Angel, J.M., Kiguchi, K. & DiGiovanni, J. Multi-stage chemical carcinogenesis in mouse skin: fundamentals and applications. *Nat. Protoc.* **4**, 1350–1362 (2009).
4. Balmain, A., Ramsden, M., Bowden, G.T. & Smith, J. Activation of the mouse cellular Harvey-ras gene in chemically induced benign skin papillomas. *Nature* **307**, 658–660 (1984).
5. Quintanilla, M., Brown, K., Ramsden, M. & Balmain, A. Carcinogen-specific mutation and amplification of Ha-ras during mouse skin carcinogenesis. *Nature* **322**, 78–80 (1986).
6. Bizub, D., Wood, A.W. & Skalka, A.M. Mutagenesis of the Ha-ras oncogene in mouse skin tumors induced by polycyclic aromatic hydrocarbons. *Proc. Natl. Acad. Sci. USA* **83**, 6048–6052 (1986).
7. Ruggeri, B. *et al.* Alterations of the p53 tumor suppressor gene during mouse skin tumor progression. *Cancer Res.* **51**, 6615–6621 (1991).
8. Kemp, C.J., Donehower, L.A., Bradley, A. & Balmain, A. Reduction of p53 gene dosage does not increase initiation or promotion but enhances malignant progression of chemically induced skin tumors. *Cell* **74**, 813–822 (1993).
9. Aldaz, C.M., Trono, D., Larcher, F., Slaga, T.J. & Conti, C.J. Sequential trisomization of chromosomes 6 and 7 in mouse skin premalignant lesions. *Mol. Carcinog.* **2**, 22–26 (1989).
10. Quigley, D.A. *et al.* Network analysis of skin tumor progression identifies a rewired genetic architecture affecting inflammation and tumor susceptibility. *Genome Biol.* **12**, R5 (2011).
11. Lapouge, G. *et al.* Identifying the cellular origin of squamous skin tumors. *Proc. Natl. Acad. Sci. USA* **108**, 7431–7436 (2011).
12. White, A.C. *et al.* Defining the origins of Ras/p53-mediated squamous cell carcinoma. *Proc. Natl. Acad. Sci. USA* **108**, 7425–7430 (2011).
13. Beck, B. *et al.* A vascular niche and a VEGF-Nrp1 loop regulate the initiation and stemness of skin tumours. *Nature* **478**, 399–403 (2011).
14. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
15. Dipple, A., Pigott, M., Moschel, R.C. & Costantino, N. Evidence that binding of 7,12-dimethylbenz(a)anthracene to DNA in mouse embryo cell cultures results in extensive substitution of both adenine and guanine residues. *Cancer Res.* **43**, 4132–4135 (1983).
16. Cai, Y., Patel, D.J., Broyde, S. & Geacintov, N.E. Base sequence context effects on nucleotide excision repair. *J. Nucleic Acids* **2010**, 174252 (2010).
17. Westcott, P.M. *et al.* The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature* **517**, 489–492 (2015).
18. Bos, J.L. *ras* oncogenes in human cancer: a review. *Cancer Res.* **49**, 4682–4689 (1989).
19. Saez, R., Chan, A.M., Miki, T. & Aaronson, S.A. Oncogenic activation of human R-ras by point mutations analogous to those of prototype H-ras oncogenes. *Oncogene* **9**, 2977–2982 (1994).
20. Schramek, D. *et al.* Direct *in vivo* RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas. *Science* **343**, 309–313 (2014).
21. Chu, W.K. & Hickson, I.D. RecQ helicases: multifunctional genome caretakers. *Nat. Rev. Cancer* **9**, 644–654 (2009).
22. Nicolas, M. *et al.* Notch1 functions as a tumor suppressor in mouse skin. *Nat. Genet.* **33**, 416–421 (2003).
23. Laczmanska, I. *et al.* Copy number alterations of chromosomal regions enclosing protein tyrosine phosphatase receptor-like genes in colorectal cancer. *Pathol. Res. Pract.* **210**, 893–896 (2014).
24. Provost, E. *et al.* The tumor suppressor rpl36 restrains KRAS(G12V)-induced pancreatic cancer. *Zebrafish* **11**, 551–559 (2014).
25. The Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).
26. Cohen-Dvashi, H. *et al.* Navigator-3, a modulator of cell migration, may act as a suppressor of breast cancer progression. *EMBO Mol. Med.* **7**, 299–314 (2015).
27. Dickinson, R.E. *et al.* Epigenetic inactivation of SLIT3 and SLIT1 genes in human cancers. *Br. J. Cancer* **91**, 2071–2078 (2004).
28. Hase, K. *et al.* AP-1B-mediated protein sorting regulates polarity and proliferation of intestinal epithelial cells in mice. *Gastroenterology* **145**, 625–635 (2013).
29. Yu, J. *et al.* Kibra functions as a tumor suppressor protein that regulates Hippo signaling in conjunction with Merlin and Expanded. *Dev. Cell* **18**, 288–299 (2010).
30. Pogue-Geile, K.L. *et al.* Palladin mutation causes familial pancreatic cancer and suggests a new cancer mechanism. *PLoS Med.* **3**, e516 (2006).
31. Lee, C.S. *et al.* Recurrent point mutations in the kinetochore gene *KNSTRN* in cutaneous squamous cell carcinoma. *Nat. Genet.* **46**, 1060–1062 (2014).
32. South, A.P. *et al.* NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. *J. Invest. Dermatol.* **134**, 2630–2638 (2014).
33. Pickering, C.R. *et al.* Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin. Cancer Res.* **20**, 6582–6592 (2014).
34. Li, Y.Y. *et al.* Genomic analysis of metastatic cutaneous squamous cell carcinoma. *Clin. Cancer Res.* **21**, 1447–1456 (2015).
35. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157–1160 (2011).

36. Agrawal, N. *et al.* Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in *NOTCH1*. *Science* **333**, 1154–1157 (2011).
37. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
38. Song, Y. *et al.* Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* **509**, 91–95 (2014).
39. Lin, D.C. *et al.* Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nat. Genet.* **46**, 467–473 (2014).
40. India Project Team of the International Cancer Genome Consortium. Mutational landscape of gingivo-buccal oral squamous cell carcinoma reveals new recurrently-mutated genes and molecular subgroups. *Nat. Commun.* **4**, 2873 (2013).
41. Ojesina, A.I. *et al.* Landscape of genomic alterations in cervical carcinomas. *Nature* **506**, 371–375 (2014).
42. Lin, D.C. *et al.* The genomic landscape of nasopharyngeal carcinoma. *Nat. Genet.* **46**, 866–871 (2014).
43. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107–1120 (2012).
44. Wu, W. *et al.* HERC2 is an E3 ligase that targets BRCA1 for degradation. *Cancer Res.* **70**, 6384–6392 (2010).
45. Toll, A. *et al.* MYC gene numerical aberrations in actinic keratosis and cutaneous squamous cell carcinoma. *Br. J. Dermatol.* **161**, 1112–1118 (2009).
46. Al Bashir, S. *et al.* Cysteine-rich secretory protein 3 (CRISP3), ERG and PTEN define a molecular subtype of prostate cancer with implication to patients' prognosis. *J. Hematol. Oncol.* **7**, 21 (2014).
47. Quintana, R.M. *et al.* A transposon-based analysis of gene mutations related to skin cancer development. *J. Invest. Dermatol.* **133**, 239–248 (2013).
48. McFadden, D.G. *et al.* Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* **156**, 1298–1311 (2014).
49. Dudgeon, C. *et al.* The evolution of thymic lymphomas in p53 knockout mice. *Genes Dev.* **28**, 2613–2620 (2014).

ONLINE METHODS

Mouse strains. FVB/N mice, an inbred strain sensitive to DMBA-induced skin SCCs[50], were obtained from Charles River Laboratories. Lgr5CREER/*Kras*$^{LSL\text{-}G12D}$/*Trp53*$^{fl/fl}$/*Rosa-YFP* mice[51] were maintained on a mixed genetic background. Both male and female mice were used in the experiments. Mouse colonies were maintained in a certified animal facility in accordance with European guidelines. The experimental protocol (483N) was approved by the ethical committee of the Université Libre de Bruxelles.

DMBA-TPA–induced skin tumors. FVB/N mice were treated with DMBA and TPA (12-O-tetradecanoyl phorbol-13-acetate) as previously described[3]. For single DMBA-dose administration (n = 20 mice), DMBA (200 µl of 0.25 mg/ml solution in acetone) was applied to shaved skin on the backs of 28-d-old mice, when hair follicles are in their regenerative stage (anagen), as administration of DMBA during this stage promotes tumor development[52]. TPA (200 µl of 0.02 mg/ml solution in acetone) was then applied once weekly to these mice for 16 weeks. For the group of mice receiving multiple DMBA administrations (n = 7), DMBA was applied once weekly for 6 weeks, starting on postnatal day 28, and TPA was applied once weekly for 16 weeks. In total, 27 msSCCs, six paired lung or lymph node metastases, and nine premalignant tumors (papillomas) on 30 different DMBA-treated mice were included in this study for statistical analysis of gene mutation enrichment and copy-number alterations. All collected SCCs were included, and no mice were excluded from analysis or experimentation. As all mice used for DMBA-induced carcinogenesis experiments were from the FVB/N inbred strain, WES was done on six bone marrow (BM) samples from six different FVB/N mice to provide germline DNA for use in the elimination of eventual single-nucleotide polymorphisms (SNPs).

Tamoxifen administration for genetic *Kras*^{G12D} p53-deficient model. Lgr5CREER/*Kras*$^{LSL\text{-}G12D}$/*Trp53*$^{fl/fl}$/*Rosa-YFP* mice[51] were treated with four 2.5-mg doses of tamoxifen on days 28–31 after birth. Carcinomas were harvested 2 months after tamoxifen administration. In total, 11 SCCs from seven different mice were included in WES and shallow WGS. Because these were outbred strains, WES and shallow WGS were done on the BM of each mouse (n = 7) to provide germline DNA for the elimination of SNPs.

Histology, immunostaining and imaging. DMBA tumors were embedded in OCT (optimal-cutting-temperature compound), and 5-µm cryosections were fixed in 4% paraformaldehyde for 10 min at room temperature. Samples were sectioned into 4–6 mm sections with a CM3050S Leica cryostat (Leica Microsystems). The following primary antibodies were used: anti-Ki67 (polyclonal rabbit, 1:200; ab15580, Abcam), anti-vimentin (rabbit, 1:2,000; clone EPR3776, Abcam), anti-K14 (polyclonal chicken, 1:1,000; SIG-3476-0100, Covance), anti-K8 (monoclonal rat, 1:1,000, TROMA-I, Developmental Studies Hybridoma Bank, University of Iowa) and anti–E-cadherin (rat, 1:1,000; clone ECCD-2, Invitrogen). Sections were incubated in blocking buffer (5% normal donkey serum, 1% BSA, 0.2% Triton in PBS) for 1 h at room temperature. Primary antibodies were incubated overnight at 4 °C. Sections were rinsed three times in PBS and incubated with secondary antibodies diluted at 1:400 for 1 h at room temperature. The following secondary antibodies were used: anti-rabbit rhodamine Red-X (Jackson ImmunoResearch), anti-rat Cy5 (Jackson ImmunoResearch) and anti-chicken Alexa Fluor 488 (Molecular Probes). Nuclei were stained in Hoechst solution (4 mM), and slides were mounted in DAKO mounting medium supplemented with 2.5% Dabco (Sigma). Pictures were acquired with an Axio Imager M1 Microscope and an AxioCamMR3 camera using Axiovision software (Carl Zeiss).

FACS isolation of tumor epithelial cells. For FACS isolation of tumor epithelial cells, DMBA-induced and genetically induced msSCCs were minced and digested in collagenase I (Sigma) for 1 h at 37 °C on a rocking plate. EDTA (5 mM) in PBS was used to neutralize the collagenase for 30 min. The cell suspension was filtered through 60-µm cell strainers (BD Biosciences) and then rinsed in PBS supplemented with 2% FCS. For antibody labeling, cells were blocked for 20 min at room temperature in PBS supplemented with 30% FCS. Immunostaining was done with phycoerythrin (PE)-conjugated anti-CD31 (clone MEC13.3, 1:100, BD Pharmingen), PE-conjugated anti-CD140a (clone APA5, 1:100, eBiosciences), PE-conjugated anti-CD45 (clone 30F11, 1:100, eBiosciences) and allophycocyanin-Cy7–conjugated anti-Epcam (clone G8.8, 1:100, Biolegend) for 30 min on ice. Cell isolation was done with FACSAria and FACSDiva software (BD Biosciences). Living tumor cells were selected by forward scatter, side scatter and Hoechst dye exclusion (Life Technologies). For DMBA-induced msSCCs, tumor epithelial cells were isolated on the basis of expression of the epithelial marker Epcam after exclusion of fibroblasts, leukocytes and endothelial cells labeled with CD140a, CD45 and CD31, respectively (**Supplementary Fig. 1a**). For genetic *Kras*^{G12D}-expressing *Trp53*-knockout Rosa-YFP msSCCs, tumor cells were isolated on the basis of YFP expression (**Supplementary Fig. 1b**). Sorted cells were collected in PBS and processed for nucleic acid extraction.

Isolation of normal epidermal cells. Back skin from 28-d-old FVB/N female mice was dissected and incubated overnight in trypsin-EDTA at 4 °C after the removal of subcutaneous adipose tissue as previously described[53]. Then the trypsin was neutralized and the epidermis was detached, minced, and filtered through 70-µM and then 40-µM pores (BD Biosciences). Single-cell suspensions in 2% FCS in PBS were then counted and used for RNA extraction.

DNA and RNA extraction. Genomic DNA was purified from approximately 2 million cells isolated from tumors by FACS or from normal epidermis (see above) using the Qiagen DNeasy blood and tissue kit (Qiagen) according to the manufacturer's recommendations. mRNA extraction (approximately 2×10^5 cells isolated from sorted tumors of normal epidermis) and DNase treatment were done with the RNeasy micro kit (Qiagen) according to the manufacturer's recommendations. Genomic DNA from DMBA-induced premalignant papillomas and from lymph node and lung metastases was extracted from total tissues because of the small size of the tumors. For this, a tissue sample of approximately 2 mm^3 was dissected, minced and directly digested for genomic DNA extraction with the Qiagen DNeasy blood and tissue kit (Qiagen) according to the manufacturer's recommendations.

WES, somatic mutation calling and annotation. Whole-genome DNA libraries were created with the Illumina TruSeq DNA sample preparation kit V2 according to the manufacturer's instructions, and whole-exome capture was done with the Nimblegen SeqCap EZ Developer Library kit (110624_MM9_Exome L2R_D02_EZ_HX1_9999042611). The resulting whole-exome libraries were then sequenced on a HiSeq2000 (Illumina) using a V3 or V4 flow cell generating two 100- or 125-bp paired-end reads. Raw sequencing reads were mapped to the mouse reference genome (GRCm38/mm10) using the Burrows-Wheeler Aligner[54] and were processed further with SAMtools[55]. The sequencing metrics of the individual samples are summarized in the **Supplementary Data**. On average, samples were sequenced with a sequencing depth of 68× with 94% of the exome covered by >10×. Genome Analysis Toolkit 2 (GATK)[56] was used to perform base recalibration and local realignment around insertion-deletions (indels). Substitutions were called by the Unified Genotyper of GATK, whereas small indels were detected using Dindel (v1.01)[57]. In the 27 DMBA-induced tumor samples, on average 72,330 such variants were detected, whereas in the six BM samples an average of 71,231 variants were detected. Similarly, on average 154,367 and 183,145 indels were called in tumor and BM samples, respectively. There were fewer detected substitutions in the genetic msSCCs (60,335 variants) and a similar number for the indels (189,912). We selected somatic variants by removing variants present in any of the six FVB/N BM samples (for DMBA msSCCs) or in the corresponding BM sample from each mouse (for genetic msSCCs)[58], as well as by removing SNPs and indels reported by the Mouse Genome Project of the Sanger Institute in the 129P2, 129S1, 129S5, C57BL/6NJ and/or FVB/NJ backgrounds. Remaining variants were annotated by ANNOVAR (v2013Jun21)[59] and the corresponding RefGene annotation track. Only nonsynonymous substitutions and frameshift indels were selected. Finally, the number of reference and alternative reads for each remaining mutation was counted. Mutations with >5% alternative reads in any of the BM samples and mutations with <10% alternative reads in the tumor were discarded. As a result, 22,839 somatic mutations (846 per tumor on average) and 255 somatic indels (8 per tumor on average) were identified in the DMBA msSCCs, and 45 somatic mutations (4 per tumor on average) and 2 somatic indels were identified in the genetic msSCCs.

Validation of somatic mutations in DMBA msSCC. All somatic mutations that were present in at least two independent tumors were considered as potential recurrent somatic mutations. First, all potential recurrent mutations were manually curated in IGV to exclude the possibility that they represented germline SNPs (**Supplementary Data**). Next, we subjected these 98 substitutions to an orthogonal validation experiment using Sequenom MassARRAY. Five mutations were detected in at least one germline sample, suggesting that they were SNPs; 45 were detected in only one tumor, suggesting that they were unique substitutions; and 16 were false positive substitutions in both samples. The design or validation of 17 recurrent mutations failed, often because of issues with the primer design, as mutations were often located in highly repetitive DNA sequences, suggesting that they probably did not represent true positives. For the few mutations for which the design failed, a very careful manual inspection in IGV was considered as validation. Overall, 14 substitutions were confirmed as recurrent substitutions.

The total set of unique substitutions (i.e., substitutions detected in only one tumor sample) was too large to be independently validated with an orthogonal technology. Nevertheless, validation of a random set of 80 unique substitutions showed that 73 substitutions were true positives (**Supplementary Data**). Inspection in IGV showed that the false positives were identified as somatic substitutions most likely because mapping issues had precluded their identification in the FVB/N reference genome. Indeed, most false positive variants were located on clusters of several somatic substitutions. To resolve this, we applied an additional filtering strategy. We calculated the average number of mutations divided by the length of the exon. For small exons the minimum length was set at 100 bp. When assessing the 80 validated unique substitutions, we observed that substitutions in an exon with a mutation frequency of <0.001 and in the range of 0.001–0.005 represented likely true positives, whereas substitutions in exons with a mutation rate higher than 0.005 were less likely to be true positives. On the basis of these results, we removed somatic substitutions that were clustered in regions exhibiting a mutation rate exceeding 0.005. Consequently, 17,247 somatic mutations (639 per tumor on average) were considered further as high-confidence true positive mutations.

Our analysis revealed 255 potential somatic indels. Manual inspection in IGV showed that 74 indels were also detectable in the matched germline, albeit in a small fraction of sequence reads, and therefore most likely represented SNPs. The remaining 192 potential somatic indels were subjected to orthogonal validation on Sequenom. Overall, 192 indels were successfully genotyped (**Supplementary Data**): 148 indels as true positives, 43 indels as false positives and one in a germline sample.

The final lists of high-confidence mutations and indels in DMBA msSCC, metastases and genetic msSCC are provided in the **Supplementary Data**. The numbers of mutations observed in mice receiving single and multiple DMBA treatments were both normally distributed. On average, 1,253 ± 333 and 431 ± 230 true positive mutations were detected in mice receiving multiple and single DMBA administrations, respectively. The raw data from whole-exome sequencing, whole-genome sequencing and RNA sequencing are available in the ArrayExpress database under accession numbers E-MTAB-2887, E-MTAB-2888 and E-MTAB-2889, respectively.

Sequenom validation. The selected indels and substitutions were validated with Sequenom MassARRAY genotyping according to the manufacturer's instructions. Primers were designed with MassARRAY Assay Design software v3.1. Automated genotyping calls were generated with MassARRAY RTTM software v4.0 and were validated by manual review of the raw mass spectra. Recurrent substitutions and indels that initially could not be successfully genotyped were redesigned for a second attempt using a new set of Sequenom primers (for example, by designing new extension primers that annealed on the opposite DNA strand relative to the extension primer from the first validation round). If they could not be genotyped in the second round, they were considered as false positive findings.

Significantly mutated genes in DMBA msSCCs. We used three approaches to identify significantly mutated genes. First we used the Genome MuSiC smg option[60]. This analysis takes the sequencing coverage, the number of variants detected, the regions of interest (exons) and the reference genome into account. Three statistical tests are performed by the tool: FCPT (Fisher's combined *P* value test), LRT (likelihood ratio test) and CT (convolution test). All the resulting *P* values are FDR corrected. Genes with a *q* value less than 0.1 in at least one of the three performed statistical tests were considered significantly mutated in this study (**Supplementary Data**). Second, we performed a χ^2 statistical test to identify genes with an enriched mutation load compared to the average mutation load per gene length and calculated corresponding *q* values (**Supplementary Data**). Third, we used MutSigCV[61], with a modified code, as the original is designed for human data. The covariates we used (i.e., replication timing and expression) were based on mouse data; the replication time used is described in more detail in ref. 62. For the expression we performed RNA-Seq on msSCC, with the genomic properties adapted according to the mouse genome. MutSigCV noted only *Hras* as a significantly mutated gene and thus was not appropriate for this mouse model because of the high proportion of false negatives (**Supplementary Data**). To select drivers, we considered only genes that were expressed in at least half of the samples analyzed by RNA sequencing (six msSCC and five normal epidermis). In total, χ^2 testing reported 279 statistically significant genes, 78 of which were expressed in msSCCs or normal epidermis, and Genome MuSiC reported 26 significantly mutated genes, 6 of which were expressed in msSCCs or normal skin (**Supplementary Data**). For every mutated gene, mouse and human homologs were aligned using UniProt, and somatic mutations reported in human cancer on homologous positions were searched in the COSMIC database.

To demonstrate that our strategy for identifying significantly mutated genes selects true cancer driver genes, we predicted the functional effect of each mutation on its respective proteins using Protein Variant Effect Analyzer (PROVEAN)[63]. PROVEAN predicted that up to 81.4%, 79.2% or 69.7% of mutations in an expressed gene with a χ^2-detemined FDR of <0.0001, <0.001 or <0.01, respectively, would be functionally disruptive, compared to 49.7% of mutations in a gene with an FDR of >0.1. Similarly, PROVEAN predicted that up to 100% or 81.6% of mutations with an FDR of <0.01 or <0.1 as determined by Genome MuSiC, respectively, would be functionally disruptive, compared to 49.9% of mutations with an FDR of >0.1. The same trend was seen when we stratified genes on the basis of other FDR cutoffs (**Supplementary Fig. 2b–e**), which verified that our selection methods enrich for genes with functional mutations.

As driver genes were shown to be enriched in nonsynonymous mutations[64], we also calculated the ratio of nonsynonymous to synonymous mutations in the DMBA msSCC mutated genes. The ratio of nonsynonymous to synonymous mutations was 6.24 in our list of significantly mutated genes and 3.42 in non-significantly mutated genes, further showing that our selection method enriches for driver genes. When calculating these ratios for different sets of genes on the basis of their FDR values, we noticed that genes with a low FDR value were enriched for a high ratio of nonsynonymous to synonymous mutations (**Supplementary Fig. 2f,g**).

RNA sequencing, analysis and annotation. RNA sequencing was performed on FACS-sorted cells from six DMBA msSCCs and five normal epidermal samples. RNA libraries were created using the Illumina TruSeq RNA sample preparation kit V2 according to the manufacturer's instructions, and resulting whole-exome libraries were sequenced on a HiSeq2000 (Illumina) using a V3 flow cell generating 50-bp reads. Raw sequencing reads were mapped to the transcriptome and the mouse reference genome (GRCm38/mm10) using TopHat 2.0 and Bowtie2.0 (ref. 59). Reads were assigned to ensemble gene I.D.s with the HTSeq software package. On average, 13,076,919 reads were assigned to genes. These reads were normalized with EDASeq[65]. A gene with less than one read per million counts in a sample was considered as not expressed in that sample. Only genes that showed expression in at least 6 of the 11 analyzed samples (msSCCs and normal skin) were considered further for mutation selection (**Supplementary Data**).

Similarity between msSCC and human SCC data. Genomics data from human data sets on squamous cell carcinomas from head and neck[25], lung[37], esophagus[39] and cervix (http://www.cbioportal.org/study.do?cancer_study_id=cesc_tcga) were explored to investigate similarities to the genes affected in our data set of msSCCs. Genomics data for melanoma

doi:10.1038/nm.3878

(http://www.cbioportal.org/study.do?cancer_study_id=skcm_tcga), acute myeloid leukemia (http://www.cbioportal.org/study.do?cancer_study_id=laml_tcga) and lung adenocarcinoma[43] were also compared to our data set as controls. The human data sets were either published or assessed using cBioportal[66,67]. We compared the recurrently mutated genes (>10%) in human and DMBA msSCCs. Only mouse genes with corresponding human genes and vice versa were taken into account. The following approach was applied to statistically assess similarities between the lists of genes identified in mSSCs and in human data sets. First the number of genes shared between the mouse and human lists was calculated; then 100,000 randomly generated gene lists were created with the same length as the human list, and the number of genes shared between the msSCCs and each randomly generated list was calculated. The number of overlapping genes obtained for each simulation was then ranked, and the percentile of the overlap between the mouse and human gene lists was determined, yielding an accurate *P* value. This procedure was repeated three times, and the average *P* value indicative of the similarity between msSCCs and human SCCs and controls was then calculated (**Supplementary Data**).

Mutational co-occurrence and exclusivity. To identify relationships of mutation co-occurrence or mutual exclusivity in genes across the msSCC cases, we performed a Genome MuSiC mutation-relation analysis using 10,000 permutations on a list of detected significantly mutated genes (**Supplementary Data**).

CNA detection by shallow WGS. Shallow WGS was performed on 9 DMBA premalignant papillomas, 24 DMBA msSCCs, 8 paired metastases and 11 genetic msSCCs. Whole-genome DNA libraries were created using the Illumina TruSeq DNA sample preparation kit V2 according to the manufacturer's instructions, and resulting whole-genome libraries were sequenced at low coverage on a HiSeq2000 (Illumina) using a V3 flow cell generating 50-bp reads. Raw sequencing reads were mapped to the mouse reference genome (GRCm38/mm10) using Burrows-Wheeler Aligner and showed 6,259,633 mapped reads on average. We removed PCR duplicates (8%) with Picard (v1.32 and v1.43) and obtained an average of 5,810,209 unique mapped reads per sample (**Supplementary Data**). Subsequently, we identified CNAs by binning the reads in 30-kb windows, correcting for genomic waves using the PennCNV software package[68] and transforming the number of reads per 30-kb window into log R values. Resulting data were then segmented using the Ascat algorithm. After some modifications to the source code of GISTIC 2.0 (Genomic Identification of Significant Targets in Cancer)[69], this software package was able to handle mouse data and was used to identify recurrent CNAs in our set of msSCCs (**Supplementary Data**). The BM samples extended with our in-house database of >25 germline samples were used to eliminate non-somatic copy-number events.

Modeling clonal evolution. An ABSOLUTE[70] analysis was performed on the WGS data, in combination with the detected mutations when available. The detected CCFs, in combination with the shared and private mutations in tumors and metastases, were used to derive the clonal evolution. To be sure only true positive mutations were used to estimate the CCFs, all unique mutations were additionally validated in mice D64 and D87, with a true positive rate of >95%.

50. Hennings, H. *et al.* FVB/N mice: an inbred strain sensitive to the chemical induction of squamous cell carcinomas in the skin. *Carcinogenesis* **14**, 2353–2358 (1993).
51. Lapouge, G. *et al.* Skin squamous cell carcinoma propagating cells increase with tumour progression and invasiveness. *EMBO J.* **31**, 4563–4575 (2012).
52. Miller, S.J. *et al.* Mouse skin is particularly susceptible to tumor initiation during early anagen of the hair cycle: possible involvement of hair follicle stem cells. *J. Invest. Dermatol.* **101**, 591–594 (1993).
53. Blanpain, C., Lowry, W.E., Geoghegan, A., Polak, L. & Fuchs, E. Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell* **118**, 635–648 (2004).
54. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
55. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
56. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
57. Albers, C.A. *et al.* Dindel: accurate indel calls from short-read data. *Genome Res.* **21**, 961–973 (2011).
58. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
59. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
60. Dees, N.D. *et al.* MuSiC: identifying mutational significance in cancer genomes. *Genome Res.* **22**, 1589–1598 (2012).
61. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
62. Farkash-Amar, S. *et al.* Global organization of replication time zones of the mouse genome. *Genome Res.* **18**, 1562–1570 (2008).
63. Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. & Chan, A.P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7**, e46688 (2012).
64. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
65. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-Seq data. *BMC Bioinformatics* **12**, 480 (2011).
66. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
67. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
68. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
69. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl. Acad. Sci. USA* **104**, 20007–20012 (2007).
70. Carter, S.L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).

Erratum: Genomic landscape of carcinogen-induced and genetically induced mouse skin squamous cell carcinoma

Dany Nassar, Mathilde Latil, Bram Boeckx, Diether Lambrechts & Cédric Blanpain

***Nat. Med.* 21, 946–954 (2015); published online 13 July 2015, corrected after print 6 August 2015**

In the published article, the format was listed as Article, but this is a Resource. The error has been corrected in the HTML and PDF versions of the article.