The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature

Gareth J. Inman1, Jun Wang2, Ai Nagano2, Ludmil B. Alexandrov3, Karin J. Purdie4, Richard G. Taylor1, Victoria Sherwood1, Jason Thomson4, Sarah Hogan4, Lindsay C. Spender1, Andrew P. South5, Michael Stratton6, Claude Chelala2, Catherine A. Harwood4, Charlotte M. Proby1 & Irene M. Leigh1

Cutaneous squamous cell carcinoma (cSCC) has a high tumour mutational burden (50 mutations per megabase DNA pair). Here, we combine whole-exome analyses from 40 primary cSCC tumours, comprising 20 well-differentiated and 20 moderately/poorly differentiated tumours, with accompanying clinical data from a longitudinal study of immunosuppressed and immunocompetent patients and integrate this analysis with independent gene expression studies. We identify commonly mutated genes, copy number changes and altered pathways and processes. Comparisons with tumour differentiation status suggest events which may drive disease progression. Mutational signature analysis reveals the presence of a novel signature (signature 32), whose incidence correlates with chronic exposure to the immunosuppressive drug azathioprine. Characterisation of a panel of 15 cSCC tumour-derived cell lines reveals that they accurately reflect the mutational signatures and genomic alterations of primary tumours and provide a valuable resource for the validation of tumour drivers and therapeutic targets.

1Division of Cancer Research, Jacqui Wood Cancer Centre, School of Medicine, University of Dundee, Dundee DD1 9SY, UK. 2Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, London EC1M 6BQ, UK. 3Department of Cellular and Molecular Medicine and Department of Bioengineering and Moores Cancer Center, University of California, San Diego, La Jolla, CA 92093, USA. 4Centre for Cell Biology and Cutaneous Research, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, UK. 5Department of Dermatology and Cutaneous Biology, Thomas Jefferson University, Philadelphia, PA 19107, USA. 6 Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. These authors contributed equally: Gareth J. Inman, Jun Wang, Ai Nagano, Catherine A. Harwood, Charlotte M. Proby, Irene M. Leigh. Correspondence and requests for materials should be addressed to G.J.I. (email: g.j.inman@dundee.ac.uk) or to J.W. (email: j.a.wang@qmul.ac.uk) or to I.M.L. (email: i.m.leigh@dundee.ac.uk)
The incidence of keratinocyte skin cancers in fair skinned populations currently exceeds that of all other cancers combined and is increasing year on year in our ageing population. More than one million new cases of cutaneous squamous cell carcinoma (cSCC) are diagnosed annually in the US, with significant health economic implications. In contrast to most other epithelial malignancies, more than a third of patients develop multiple primary cSCC. This is especially true in immunosuppressed individuals with evidence in organ transplant recipients (OTR) of a more than 100-fold increased risk of developing cSCC. Although many cSCC are effectively treated with surgery or radiotherapy, morbidity is significant, particularly as tumours frequently occur on cosmetically sensitive sites. Metastasis occurs in ~5% of cases and cSCC is responsible for approximately one-quarter of skin cancer-related deaths and estimates for annual mortality are similar to those for melanoma in parts of the USA. There are few effective treatments for advanced cSCC, with five-year survival of less than 30% reported for metastatic disease.

Cutaneous SCC is poorly understood at a molecular level and the drivers of disease progression have not been fully elucidated. Whole-exome sequencing (WES) has revealed a high mutation rate with on average 50 mutations per megabase pair DNA. Additionally, gross chromosomal aberrations and numerous copy number alterations (CNA) indicative of genetic instability create a complex molecular landscape. Epidemiological and clinical evidence supports cumulative life time exposure to ultraviolet radiation (UVR) as the principle environmental carcinogen responsible for cSCC and the majority of mutations found in cSCC contain a ‘UV signature’. Our understanding of the genomic landscape of cSCC is further complicated by keratinocyte clones in apparently normal sun-exposed skin bearing mutations in cSCC tumour suppressor genes (TSGs) such as NOTCH1/2 and TP53. There is evidence that the mutational profile of cSCC harbours similarities to those of other epithelial tumours, including head and neck, lung and oesophageal SCCs (reviewed in ref.11) and the squamous subtype of pancreatic SCC, but comparisons are limited by small cSCC data sets.

Analyses of the patterns of mutations in cancer genomes has revealed the presence of distinct ‘mutational signatures’. They are described using a simple classification based on the six classes of single base mutations in combination with the base 5’ and 3’ to each mutation and characteristic UVR-induced C > T mutations at pyrimidines appear to predominate in all cSCC studies to date (COSMIC signature 7).

Here, we perform WES of 40 primary cSCC from both immunosuppressed (IS) and immunocompetent (IC) patients classified according to histological differentiation status (well-differentiated vs. moderately/poorly differentiated categories). We have taken an integrated bioinformatics approach with the intention of identifying a limited number of actionable pathways onto which this multiplicity of genetic events will converge. In addition to mutational, chromosomal and clonality analyses, a mutational signature analysis has been performed, revealing a novel mutational signature associated with patients who have been exposed to the immunosuppressive agent, azathioprine.

Furthermore, we describe the molecular characterisation of 15 primary cSCC cell lines. These reflect the mutational signatures, mutational and genomic alteration profiles and gene expression changes of primary tumours and provide a resource for investigating the contribution of genes, pathways, and processes for cSCC progression.

Results

Mutational burden and signatures. Forty cSCC samples from 37 patients (Supplementary Data 1) were analysed/reanalysed by WES (including 30 previously described). Twenty-nine patients were organ transplant recipients receiving immunosuppressive drugs and one was immunosuppressed for Crohn’s disease. This sample set contained equal numbers of well-differentiated (WD) vs. poorly/moderately differentiated (MD/PD) tumours (20 samples each). WES coverage averaged 54.3 × (range 25.8–79.1x) with 91% of bases covered > 10 × (range 73.5–97.8%) and 81.4% > 20 × (range 47.2–93%) (Supplementary Data 2). We identified 71,242 somatic mutations in the tumours (Supplementary Data 3). The number of nonsynonymous and synonymous mutations varied across each group averaging 1795 somatic variants in WD (range 23–6867) and 1793 somatic variants in MD/PD (range 13–5188) (Fig. 1a and Supplementary Data 4). Computational analysis of patterns of somatic mutations has enabled the identification of 30 distinct mutational signatures from more than 12,000 samples from a wide range of cancer types (for a complete set of currently known signatures see http://cancer.sanger.ac.uk/cosmic/signatures). Many of these signatures have been attributed to environmental exposure to mutagens. Mutational signature analysis was performed on 37/40 of the cSCC whole-exome sequences (Fig. 1b and Supplementary Data 4). Signature 7 was observed in 33 samples (Fig. 1b), which reflected the C > T mutations of pyrimidines and has previously been described as a ‘UVR signature’ (Supplementary Fig. 1B). Twenty-one of these UVR signature positive tumours were from confirmed sun exposed sites (Fig. 1b; head and neck, wrist and hand, Supplementary Data 1) and the remaining 12 were from sites which included the leg, chest, shoulder and back, which may also have been sun exposed (Supplementary Data 1). MD07, which displays no evidence of UVR exposure, was isolated from the foot of the patient and was, therefore, unlikely to be sun exposed (Fig. 1b and Supplementary Data 1, 4).

A previously undescribed mutational signature, termed signature 32 here-after, was found in 27 samples from both the WD and MD/PD tumours (Fig. 1b, orange bars). It is predominately C > T mutations (75%) in combination with C A T, A C T, and T C mutations (Fig. 1c). There seems to be a clear presence for C > X at ApCpN and T > X at GpTpN. This novel mutational signature exhibits a very strong transcriptional strand bias potentially indicating an interplay with transcription coupled nucleotide excision repair (TC-NER) due to adducts on guanine and adenine. The pattern of signature 32 is distinct from any of the previously known COSMIC mutational signatures, including signatures that have a pattern of mutations with predominately C > T transitions such as signature 7. Only two of the COSMIC mutational signatures exhibit a correlation > 0.5 between their respective patterns and that of signature 32. The pattern of signature 30 (attributed to mutations in the base excision repair gene NTHL1) has a correlation of 0.58 with signature 32 and that of signature 11 (attributed to exposure to alkylating agents) has a correlation of 0.65 with signature 32. These correlations are lower than those of other distinct COSMIC signatures, for example COSMIC signatures 1 and 6 have a correlation of 0.81. Further, signatures 11 and 30 do not exhibit the same transcriptional strand bias as signature 32. Thus, signature 32 is an entirely novel mutational signature.

Signature 32 was prominent in many of the samples from immunosuppressed patients (Fig. 1b) indicating a potential association with exposure to immunosuppressant drugs. Most immunosuppressed patients received azathioprine (Fig. 1b) although this was usually part of a complex regimen involving at least one other immunosuppressive drug (Supplementary Data 1). However, three of the tumours with signature 32 were from patients who had received azathioprine alone (MD02,
PD01, WD03) and both tumours analysed from immunosuppressed patients whom had not received azathioprine (MD08, WD15) did not exhibit mutational signature 32 (Fig. 1b and Supplementary Data 1, 4). There is a significant association only between a confirmed history of azathioprine exposure and the presence of signature 32 (Fisher’s exact test, \( p < 0.0001 \); Supplementary Fig. 2a) and not exposure to the other immunosuppressant drugs (Supplementary Fig. 2a). We did not identify any other significant associations between any of the mutational signatures and exposure to any of the immunosuppressant drugs (Supplementary Fig. 2b). Furthermore, analysis of treatment times revealed a strong positive correlation with the estimated time of azathioprine exposure and the prevalence of signature 32 (Spearman’s rank order correlation \( r_s(26) = 0.679, \ p < 0.0001 \); Supplementary Fig. 3). Taken together these findings indicate that chronic exposure to azathioprine correlates with the presence of mutational signature 32. All tumours contained lower levels of signatures 1 and 5, which are found in most cancers and are attributed to so called ‘clock-like’ mutational processes20 (Fig. 1b and Supplementary Data 4). Four tumour samples also contained mutational signature contributions from Signatures 2 and 13 (MD06, MD07, WD02 and WD18; Fig. 1b).
and Supplementary Data 4), which are attributed to the activity of the AID/APOBEC family of cytidine deaminases converting cytosine to uracil\textsuperscript{14}.

**Somatic copy number aberrations.** Somatic copy number aberrations (CNA) and loss of heterozygosity (LOH) events were identified across WD and MD/PD samples (Supplementary Data 5), with MD/PD samples having more large-scale genomic events than WD samples (Fig. 2a, b). 62.5% of the samples (25/40) had detectable copy number variations (CNVs) in at least 2 regions, which is consistent with previous observations\textsuperscript{21–24} with a significant positive correlation with previous studies for top frequent CNAs (Supplementary Fig. 4, Pearson’s correlation \( r = 0.37, p = 0.015 \)). GISTIC\textsuperscript{25} analysis further identified significantly amplified and deleted regions (\( q < 0.1 \), Supplementary Data 6), including amplifications at 3q26, 5q22, 7q21, 11q22 and 15q11, and deletions at 3p12, 5q13 and 9q21. Amplifications at 3q\textsuperscript{22–24}, 5q\textsuperscript{21, 24}, and copy losses at 3p\textsuperscript{21–24} and 5q\textsuperscript{24} have also previously reported. Frequent gain of 9q and loss of 9p are also observed in this and previous studies of cSCC (Supplementary Data 7).

**Significantly mutated genes (SMGs) in cSCC.** We employed three statistical mutational significance methods; MutsigCV\textsuperscript{26} (significance cutoff \( p < 0.05 \), Supplementary Data 8), Oncodrive-FM\textsuperscript{27} (significance cutoff \( q < 0.05 \), Supplementary Data 9) and Oncodrive-CLUST\textsuperscript{28} (significance cutoff \( q < 0.05 \), Supplementary Data 10). These identified 109, 97 and 171 SMGs respectively totalling 351 unique genes (Fig. 3a and Supplementary Data 8–10). We identified 22 SMGs by at least two methods (Fig. 3a, b). This revealed recurrent mutations in a number of genes previously implicated in human cSCC including NOTCH1/2, TP53 and CDKN2A (Fig. 3b and Supplementary Data 11, 12, Supplementary Fig. 5). The frequency and statistical significance of these potential driver genes are consistent with previous exome\textsuperscript{10} and targeted sequencing studies\textsuperscript{9} (Supplementary Data 13).

Additional SMGs included HRAS, MAP3K9, PTEN, SF3B1, VPS41 and WHSC1 (Fig. 3b, Supplementary Fig. 6 and Supplementary Data 14) all of which have known genetic alterations in human malignancies\textsuperscript{29–38}. We identified oncogenic activating mutations in HRAS in three tumours (Fig. 3b and Supplementary Fig. 6), which have previously been identified in 3–20% of cSCC\textsuperscript{9,10} (Supplementary Data 13) and keratoacanthomas (reviewed in ref. 4). Of note, 10% of the samples exhibit copy number loss of HRAS (Fig. 3b), which has been observed by others\textsuperscript{8}, warranting a need for better understanding of the role of HRAS in cSCC. PTEN alterations have been previously described in cSCC\textsuperscript{9} (Supplementary Data 13), but the other SMGs including FLNB, GLIS3, CACNA1C, HERC6, TRAPPC9, MAPK1P1L (Fig. 3b, Supplementary Fig. 7 and Supplementary Data 15), GRHL2, CLEC3, TMEM51, ATP1A1, LCLAT1 and CRY1 (Fig. 3b, Supplementary Fig. 8 and Supplementary Data 16) are of unknown significance in cancer and present with no obvious hotspot mutations. Non-silent mutations in a few of these SMGs have also been observed in cSCC\textsuperscript{10} (Supplementary Data 13). Comparison of the frequency of alterations in the 22 SMGs between tumours isolated from immunosuppressed and immunocompetent patients revealed no significant difference in mutation frequencies indicating that tumours that arise in these patient groups have common drivers (Fig. 3b and Supplementary Data 17). Next, we assessed the probability for each mutational somatic substitution in the 22 SMGs to be caused by each of the operative mutational signatures. Overall the majority of the SMG gene mutations are caused by azathioprine signature 32 (66.2%): 24.4% are due to UVR driven signature 7; 7.8% are caused by

---

**Fig. 2** Somatic copy number aberrations (SCNA) and LOH events in 40 cSCC samples. a OncoPrint of copy gain, loss and copy-neutral (CN)-LOH segments across moderate/poor and well-differentiated groups. IS immunosuppressed, IC immunocompetent. b Box-and-whisker plot indicating the percentage of the genome altered by CNA/LOH across the two groups. Individual values were shown by the blue dots. Moderate/poor differentiated group had significantly higher percentage of altered genome than well-differentiated group (Student’s \( t \)-test, two sided, \( p = 0.018 \)).
clock-like mutational signatures\(^\text{20}\) (signatures 1 and 5) and the remaining 1.7\% by APOBEC related signatures 2 and 13, suggesting the driving mutational processes responsible for the initiation and progression of the tumours in our sample set are due to a combination of UVR and azathioprine exposure (Fig. 3c).

**Inference of clonal evolution and order of genetic changes.** Next, we performed clonal analysis, integrating somatic mutations (SNVs and indels) and CNAs using EXPANDs\(^\text{39}\), and SciClone\(^\text{40}\) which have been used to infer patterns of clonal evolution\(^\text{41,42}\). In 35 of 40 cSCC, WES data were sufficient for the identification of dominant clone and nested subpopulations in each sample, thus allowing for the differentiation between clonal and subclonal mutations. EXPANDs and SciClone had concordant estimates of tumour purity (Supplementary Fig. 9, Pearson’s correlation \(r = 0.9, p = 3.42e-13\), Supplementary Data 19). There was no correlation between the number of somatic mutations identified in our samples and tumour purity (Pearson’s correlation \(r = 0.10, p = 0.57\)). Clonality analysis showed tumour heterogeneity across our samples, varying from 1 to 11 clones as estimated by EXPANDs (Fig. 4a) and 1 to 9 clones as estimated by SciClone (Fig. 4b) with some correlation in the number of clusters estimated by the two methods (Supplementary Fig. 10). 26/35 (74\%) had at least two clones, 10/35 (29\%) had at least three clones and 7/35 (20\%) had at least five clones identified by both programmes (Fig. 4a, b and Supplementary Data 20) indicating a high degree of heterogeneity in cSCC.

For the 22 SMGs, we assessed the order of driver mutation acquisition inferred from the aggregate frequencies at which they were found to be clonal or subclonal\(^\text{43}\). We characterised the clonal and subclonal features of the nonsynonymous mutations for these 22 genes with 74\% of all clonal mutations called by EXPANDs also identified as clonal by SciClone (Supplementary Fig. 10). Over 70\% of mutations in CDKN2A, ATP1A1, HERC6, MAPK11P1L, GRHL2 and TP53 were clonal using both methods implying alterations in these genes tended to arise earlier during cSCC development (Fig. 4c, d). Rank order of clonality similarly indicated that mutations in these genes and also those in NOTCH1, TRAPPC9, FLNB and MAP3K9 may also be early events, whereas mutations in HRAS, VPS41, GLIS3, CACNA1C, NOTCH2, CLCN3, LCLAT1, TMEM51 and SF3B1 appeared to accumulate a higher proportion of subclonal nonsynonymous mutations (Fig. 4c, d and Supplementary Data 21), indicating that lesions in these genes are more likely to occur later.

**Integration of genomic drivers and gene expression profiles.** To investigate potential tumour suppressor or tumour promoter roles of our 22 SMGs and genes encoded within areas of significant CNAs, we analysed gene expression profiles (GEP) in five independent data sets comprising samples from normal skin, actinic keratoses (AK) and SCC. These included GSE45216\(^\text{44}\), coupled with additional samples described here as Data set 1 (Supplementary Data 22), GSE42677\(^\text{45}\), GSE25034\(^\text{6}\), GSE84293\(^\text{7}\) and GSE32628\(^\text{21}\). GLIS3 and LCLAT1 were determined to have a low level of expression in normal and/or lesional skin across samples and were excluded from this analysis (see Methods). The remaining 20 SMGs were clearly separated into two clusters in GSE42677 (Fig. 5a). One cluster was highly expressed in normal skin and became downregulated in AK and SCC, and the other cluster demonstrated the opposite trend (Fig. 5a). The down- and upregulated patterns for these genes were largely consistent between the five data sets (Fig. 5b). Among them, TP53 and NOTCH2 were downregulated in SCC compared to normal skin, while seven genes HERC6, CDKN2A, FLNB, MAP3K9, TMEM51, MAPK11P1L and WHSC1 were upregulated in SCC relative to

---

**Fig. 3** Twenty-two SMGs identified in 40 cSCC samples. **a** Venn diagram of overlap of significantly mutated genes as assessed by MutSigCV, OncodriveFM and OncodriveClust. **b** Mutation OncoPrint of the 22 SMGs identified by at least two of the three methods, with their aberrational frequencies also indicated with percentage bars. IS immunosuppressed, IC immunocompetent. **c** Pie chart of the mutational signature contribution to SMG mutations.
regions identified to play tumour promoting roles in cSCC. Nine genes within deleted regions in the mRNA level in cSCC compared to normal skin in GSE42677 (Fig. 6a), 38 in Data Set 1, 78 in GSE32628, 7 in GSE2503 and 43 in GSE84293 (Fig. 6b, top Venn diagram). Overlap analysis revealed that 19 genes were encoded in deleted regions and upregulated at the mRNA level in at least three of these data sets (Fig. 6b, c) indicating that these genes may play tumour promoting roles in cSCC.

Within the amplified regions identified by GISTIC analysis, the mRNA levels of 23 genes were upregulated in cSCC compared to normal skin in GSE42677 (Fig. 6a), 38 in Data Set 1, 78 in GSE32628, 7 in GSE2503 and 43 in GSE84293 (Fig. 6b, top Venn diagram). Overlap analysis revealed that 19 genes were encoded in amplified regions and upregulated at the mRNA level in at least three of these data sets (Fig. 6b, c) indicating that these genes may play tumour promoting roles in cSCC. Nine genes within deleted regions identified by GISTIC analysis were also downregulated at the mRNA level in cSCC compared to normal skin in GSE42677 (Fig. 6a), 11 in Data set 1, 24 in GSE32628, 2 in GSE2503 and 5 in GSE84293 (Fig. 6b, bottom Venn diagram). Overlap analysis revealed that five genes were encoded in deleted regions and downregulated at the mRNA level in at least three of these data sets (Fig. 6b, c) indicating that these genes may play tumour suppressor roles in cSCC.

**Fig. 4** Clonality analysis of cSCC tumour samples. Clonality analysis of 35 cSCC exomes using EXPANDS (a) and SciClone (b). Clonality analysis of 22 SMGs identifying clonal and subclonal nonsynonymous mutations using EXPANDS (c) and SciClone (d). The number of nonsynonymous mutations in each gene was also shown.

Identification of WD and MD/PD group-specific SMGs. Differentiation status is a prognostic marker in cSCC (reviewed in ref. 4). We developed a randomisation test strategy to identify WD and MD/PD group-specific SMGs based on the MutSigCV and OncodriveFM significance p-values (see Methods, Fig. 7a), and further selected those group-specific genes that were also regarded as expressed in the GEP data sets. This led to 16 MD/PD-specific and 6 WD-specific SMGs, respectively (Fig. 7b). The MD/PD-specific genes included TMEM51 and GRHL2 previously identified in the analysis of the whole 40 exomes but also identified additional MD/PD-specific genes ZZZF1, GMDS, NEDD4L, LP1, PRB1, HECTD4, SOS2, ICAM1, VWF, ACVR2A, POLH, CNDP2, RPLP1 and PRMT3. Their significance has yet to be described in cSCC although non-silent mutations in some of these genes have been observed in cSCC10 (Supplementary Data 23). Interestingly, the expression of ACVR2A appeared to decline gradually with the disease progression, with the highest level of expression in the normal skin and the lowest in SCCs in all five GEP data sets (Supplementary Fig. 11), suggesting its...
potential tumour suppressing role in disease progression. The WD-specific genes included the previously identified ATP1A1 and additionally SULF1, ZNF528, NRCAM, FAT1 and SEMA5A. FAT1 has been identified as a significantly mutated driver gene in cSCC and in normal sun exposed skin. None of these other WD-specific genes have been associated with cSCC although non-silent mutations in some of these genes have been observed in cSCC (Supplementary Data 23).

Mutational pathway analysis. Mutational pathway analysis revealed several KEGG signalling pathways that were more mutated in the MD/PD group, such as NOTCH, TGF-β, WNT, PI3K-AKT and neurotrophin signalling (Fig. 7c and Supplementary Data 24, 25). The MD/PD-specific mutated pathways also included those related to immune responses such as T-cell receptor, Fc epsilon RI, RIG-I-like receptor and chemokine signalling pathways (Fig. 7c). Furthermore, our mutational pathway analysis revealed that biological processes of ECM-receptor interaction, cell cycle, tight junction and glycolysis/gluconeogenesis were also more significantly mutated in the MD/PD group compared to WD group, whereas lysine degradation, regulation of the actin cytoskeleton, carbohydrate digestion and adsortion and basal transcription factors were among those specific to the WD group (Fig. 7d and Supplementary Data 25). Computational pathway analysis of mutated genes in cSCC has not previously been performed. Manual pathway assessment previously suggested involvement of the RAS/RTK/PI3K and cell cycle pathways in aggressive cSCC consistent with our observations (Fig. 7c, d). The findings that PI3K signalling is frequently altered in more aggressive SCCs such as HNSCC, lung and oesophageal SCC further supports our MD/PD pathway analysis implicating PI3K-AKT signalling in disease progression.

Patient-derived cell lines reflect primary tumours. To facilitate future functional studies, tumour keratinocytes were cultured from multiple tumours and a panel of 15 cell lines was selected for full characterisation (Supplementary Data 26). Nine were derived from five OTR patients on immunosuppressive drugs including five from a single patient (PM1, MET1, 2, 4, T9). Six cell lines were derived from five immunocompetent patients (IC1, IC1MET, IC8, IC12, IC18 and IC19) with IC1/IC1MET being paired primary and metastatic lymph node derived lines, respectively.

In general, mutational profiles of patient-derived cell lines were comparable with those of the tumour sample set (Supplementary Data 27). Mutation signature analysis showed that signature 7 was present in all cell lines, and eight lines from four patients also contained signature 32, all of which were derived from tumours from immunosuppressed patients receiving azathioprine (Fig. 8a and Supplementary Data 28). Interestingly, PM1, MET1, 2, 4, and T9 from the same patient show very similar mutation signature compositions, except that MET4 also exhibited signature 26, which correlates with the presence of signature 26. Analysis of somatic CNA/LOH revealed many recurrently altered regions that were also present in patient tumours including amplification at 3q26, 7q21 and 11q22, and deletions at 5q13 and 3p12 (Fig. 8b). The cell lines exhibited a high degree of chromosome instability similar to the most genomically altered primary tumours. Mutations in the 22 SMGs from the primary tumour set also frequently occurred in cell lines, with TP53, NOTCH1/2 and CDKN2A the most recurrently altered with similar mutational
profiles in primary tumour samples and cell lines (Fig. 8c, Supplementary Data 29 and Supplementary Fig. 13).

We also performed gene expression analysis of our cell lines and three separate isolates of primary normal human keratinocytes (NHKs) using Illumina Beadchip arrays (Supplementary Data 30). NOTCH2, a commonly downregulated SMG in the primary tumour data set, was also downregulated in cSCC cell lines in comparison to NHKs (Supplementary Fig. 14A). 2/7 commonly upregulated SMGs (CDKN2A and WHSC1) were also upregulated in cSCC cell lines compared to NHKs (Supplementary Fig. 14A). 3/19 commonly amplified and upregulated genes POLR2H, ALG3 and AP2M1, were also upregulated in the cell lines relative to NHKs and 2/5 of the commonly deleted and downregulated genes, SSBP2 and F2R, were also downregulated in cSCC relative to NHKs (Supplementary Fig. 14B). We further identified the mutational profiles for MD/WD and PD group-specific SMGs in cell lines, with MD/WD-specific SMGs ZZZEFL, NEDD4L, LRP1, VWF, HECTD4, ICAM1 and ACVR2A being the most frequently altered, and WD-specific genes FAT1, NRCAM, SULF1 and ATP1A1 recurrently altered in SCC lines (Fig. 8d).

Our cell lines thus represent a well characterised resource, which may drive tumour formation leaves a mutational signature and deciphering these signatures is a unique opportunity to reveal the mutational processes operative in different cancer types. We applied this approach to our data set and found five previously identified signatures in our tumours including signature 7, which has been associated with exposure to UV a known potent skin cancer carcinogen. The UV signature was found in 33/37 exomes examined and in all 15 cell lines. A novel signature, termed signature 32 was found in 27 of the tumour samples and 8 of the cell lines (Figs. 1, 8). Interrogation of the clinical history for each patient found that the presence of signature 32 correlated with immunosuppressive regimens including use of azathioprine. Although immunosuppressants such as mycophenolate mofetil, cyclophosphamide and cyclosporine A affect DNA damage mechanisms, including UV induced DNA excision repair, three patients received azathioprine alone and analysis of treatment times revealed a strong positive correlation with exposure time to azathioprine and signature 32 prevalence (Supplementary Fig. 3). All of the signature 32 positive cell lines were also from immunosuppressed patients who received azathioprine (Fig. 8). Furthermore, only exposure to azathioprine and not the other immunosuppressive drugs correlated with the presence of signature 32 and the only significant correlation of immunosuppressive drug use and the presence of any of the mutational signatures was with azathioprine exposure and signature 32 (Supplementary Fig. 2). This signature is likely to be of biological relevance as mutational signature analysis of mutations in the putative driver genes identified in our study, revealed that over 65% of these were signature 32 (Fig. 3c).

Each biological process causing mutations in somatic cells that may drive tumour formation leaves a mutational signature and deciphering these signatures is a unique opportunity to reveal the mutational processes operative in different cancer types. We applied this approach to our data set and found five previously identified signatures in our tumours including signature 7, which has been associated with exposure to UV a known potent skin cancer carcinogen. The UV signature was found in 33/37 exomes examined and in all 15 cell lines. A novel signature, termed signature 32 was found in 27 of the tumour samples and 8 of the cell lines (Figs. 1, 8). Interrogation of the clinical history for each patient found that the presence of signature 32 correlated with immunosuppressive regimens including use of azathioprine. Although immunosuppressants such as mycophenolate mofetil, cyclophosphamide and cyclosporine A affect DNA damage mechanisms, including UV induced DNA excision repair, three patients received azathioprine alone and analysis of treatment times revealed a strong positive correlation with exposure time to azathioprine and signature 32 prevalence (Supplementary Fig. 3). All of the signature 32 positive cell lines were also from immunosuppressed patients who received azathioprine (Fig. 8). Furthermore, only exposure to azathioprine and not the other immunosuppressive drugs correlated with the presence of signature 32 and the only significant correlation of immunosuppressive drug use and the presence of any of the mutational signatures was with azathioprine exposure and signature 32 (Supplementary Fig. 2). This signature is likely to be of biological relevance as mutational signature analysis of mutations in the putative driver genes identified in our study, revealed that over 65% of these were signature 32 (Fig. 3c).

Azathioprine, an inhibitor of de novo purine synthesis, is associated with selective ultraviolet A (UVA) photosensitivity and mutagenic effects in the skin. The azathioprine metabolite 6-thioguanine (6-TG) replaces a small proportion of DNA guanine

Fig. 6 Expression profiles of genes significantly amplified or deleted in cSCC. a Expression heatmap of significantly amplified and upregulated genes (n = 23), and deleted and downregulated genes (n = 9) in the in situ vs. normal comparison that were also expressed across normal, AK, in situ and invasive SCC samples from GSE42677. b Venn diagrams of overlap across five gene expression data sets of significantly amplified genes and overexpressed (upper panel) and significantly deleted and downregulated genes (lower panel). c Log2 fold changes (FC) of pairwise comparisons of AK vs. normal and SCC vs. normal across the five data sets for 19 amplified/upregulated and 5 deleted/downregulated expressed genes shared across at least three data sets. Within the heatmap, red colour indicates the upregulation in AK or SCC compared to normal control and blue colour indicates the downregulation in AK or SCC in relation to normal skin.
and becomes a strong UVA chromophore, interacting with UVA to generate reactive oxygen species, which cause widespread DNA damage and protein oxidation; the latter damages the DNA repair proteome increasing UVB mutagenicity.\(^{33,34}\) Azathioprine photosensitivity is clinically measurable, may be associated with a specific genetic signature in basal cell carcinomas\(^{35}\) and can be reduced by switching from azathioprine to mycophenolate mofetil (MMF), although azathioprine metabolites may persist in the skin for several years after withdrawal.\(^{36}\) Given the photosensitising properties of azathioprine we postulate that the combined action of UVR and incorporation of azathioprine metabolites into DNA provide the biochemical basis for this novel signature. We propose that these azathioprine effects are associated with this new genetic signature (signature 32) that contributes to tumour progression. This has clinical significance for the many patients currently receiving azathioprine who should be counselled about their skin cancer risk and UVR photosensitivity.

We confirm the high mutational burden present in cSCC with an average of over 1700 mutations present in the primary tumour exomes (Fig. 1). WES data enabled us to assess clonality and revealed that most tumours are highly heterogenous containing 2–11 clones (Fig. 4). GISTIC analysis revealed common areas of copy number loss and gain in our samples with MD/PD tumours exhibiting higher levels of chromosomal instability compared to WD tumours (Fig. 2) further illustrating the genetic complexity of cSCC.

We employed three computational methods (MutSig, OncodriveClust, OncodriveFM) and identified 351 SMGs at the cutoffs we used. Twenty-two genes were found to be significantly mutated in two or more of these programmes enabling a further ranking of the potential driver genes (Fig. 3). Combination of these analyses with CNA/LOH analysis indicate that alteration of *NOTCH1* and *NOTCH2* genes are among the most common events in cSCC (Fig. 3) with only 3/20 MD/PD tumours and 5/20 WD tumours escaping alteration in either of these genes indicating that the
NOTCH pathway is an almost universal tumour suppressor pathway in cSCC probably due to its role in promoting cell cycle exit and differentiation. We also identify TP53 alterations in 70% of our tumours and CDKN2A changes in nearly 50% of cases. Clonality analysis revealed that CDKN2A mutations are clonal (Fig. 4c, d) indicating a likely early event in formation of the tumours in which they were identified. Strikingly, despite deep sequencing efforts, CDKN2A mutations in contrast to TP53 or FAT1 mutations have yet to be identified in sun-exposed ‘normal skin’ suggesting a key gatekeeper role of p16NK4B in cSCC. Our findings also suggest that mutations in ATP1A1, HERC6, MAPK1IP1, GRHL2, TRAPPC9, FLNB, NOTCH1 and MAP3K9 represent early driving events in cSCC (Fig. 4c, d). Within these, alteration in ATP1A1 may pre-dispose to well-differentiated tumours, whereas alterations in GRHL2 may pre-dispose to more poorly differentiated and potentially poorer prognosis tumours (Fig. 7). GRHL2 mutations are associated with an autosomal-recessive ectodermal dysplasia and this transcription factor may drive a cancer stemness phenotype in oral SCC. This gene clearly warrants further investigation in cSCC.

To further sifting our list of somatically mutated genes and those that are encoded within regions of common amplification or loss we integrated our WES data with five independent gene expression data sets (Figs. 5, 6). Interestingly, despite the proposed tumour suppressor role of inactivating mutations in CDKN2A we consistently observed upregulation of CDKN2A gene expression in the GEP data sets and in our cell lines (Fig. 5 and Supplementary Fig. 14). Transcription factor motif analysis indicates that ERK signalling through ETS transcription factors may be operative during SCC progression and this may act to upregulate CDKN2A expression as part of a stress induced senescence programme possibly explaining this discrepancy. This analysis

Fig. 8 Mutation signatures, somatic CNA/LOH and mutation profiles for identified SMGs in cSCC tumour samples in SCC cell lines. a Mutation signatures across all SCC cell lines. Again, signature 7 and 32 were the most dominant signatures. IC immunocompetent, IS immunosuppressed, Aza confirmed azathioprine exposure, NC no confirmed azathioprine exposure, SE sun exposed site, U unknown if sun exposed site. b OncoPrint for copy gain, loss and CN-LOH segments. c Mutation OncoPrint for 22 SMGs detected in cSCC tumour samples across cell lines, ordered by aberration frequency. d Mutation OncoPrint for WD/MD/PD driver genes detected in cSCC tumour samples across cell lines, ordered by frequency. MD/PD-specific genes found in cSCC samples are highlighted in orange.
highlights potential oncogenic tumour promoting roles for the actin binding protein encoded by FLNB and the histone methyltransferase encoded by WHSC1, as expression of both of these genes are elevated during cSCC progression and they do not contain premature termination codon or splice site mutations (Fig. 5 and Supplementary Fig. 6, 7). We also identified genetic alteration of the arginase methyltransferase, PRMT3, specifically in the MD/PD tumours (Fig. 7B), further indicating that epigenetic modifiers may regulate cSCC progression. Genes that are amplified and upregulated during cSCC progression may act as oncogenic drivers of disease and we revealed 19 further genes with this potential role including the SKI like proto-oncogene SnoN encoded by the SKIL gene (Fig. 6). This transcription factor can modulate canonical TGF-β signalling and pathway analysis revealed that genomic alterations in TGF-β signalling are enriched in the MD/PD group of tumours (Fig. 7C). Consistent with this we also identify alterations in the ACVR2A gene that encodes for an activin/BMP receptor enriched in this tumour group, independent pathologists. Details of prescribed medications including drug, dose notes and transplant databases for all immunosuppressed patients. The estimated genetic modifiability of SNV alterations. This analysis implicates RAP2B as a well described oncogenic activator and AP2M1 may act in conjunction with the Qubit 2.0 fluorometer (both Life Technologies, Paisley, UK). Short tandem repeat (STR) genotyping was performed to authenticate the unique identity of each cell line. Sixteen loci distributed across the human genome were amplified using the AmpFLSTR Identifier PCR Amplification Kit (Perkin Elmer, Crawley, UK) and passed upon at least 80% concurrence. All cell lines were routinely screened for mycoplasma and were negative throughout the study.

DNA extraction and genetic analysis. DNA was extracted from snap-frozen tissue, cultured cells and blood using the Qiagen DNA Mini kit (Qiagen, Crawley, UK) according to the manufacturer’s instructions. Quantitation was performed using the double-stranded DNA-specific Qubit dsDNA BR assay kit in conjunction with the Qubit 2.0 fluorometer (both Life Technologies, Paisley, UK). Randomisation test to identify MD/PD and WD-specific driver genes. We used three methods to identify different positive selection signals that occurred in driver mutations: OncodriveFM, OncodriveCLUST, both implemented in the IntGen software, and MutSigCV. For OncodriveFM and OncodriveCLUST, we used corrected p-values (q < 0.05) for significance. For MutSigCV, due to the small sample size, p-values could not be adjusted, so raw p < 0.05 was applied. Overlapped genes that were identified as being significant by at least two of the three methods were further selected. Significantly mutated pathways were also identified using OncodriveFM (q < 0.05).

Identification of mutational driver genes and pathways. We used three methods to identify different positive selection signals that occurred in driver mutations: OncodriveFM, OncodriveCLUST, both implemented in the IntGen software, and MutSigCV. For OncodriveFM and OncodriveCLUST, we used corrected p-values (q < 0.05) for significance. For MutSigCV, due to the small sample size, p-values could not be adjusted, so raw p < 0.05 was applied. Overlapped genes that were identified as being significant by at least two of the three methods were further selected. Significantly mutated pathways were also identified using OncodriveFM (q < 0.05).
the two groups and recorded the significance values for our candidate genes. We repeated this randomisation procedure 100 times and the transformed p-values from Mason-McClure or One-sided Wilcoxon generally followed a normal distribution for each gene. We next calculated the probability that the observed significance in moderate/poor or well-differentiated group was stronger than that expected by chance for each candidate. Significant genes from the randomisation test (two-tail p < 0.05) were further noted and regarded as the true moderate/poor or well-differentiated specific genes.

Identification of CNA and LOH using WES data. Analyses for CNA and LOH events from WES data were based on a combinational approach previously described36,37. We adopted two independent approaches (ASCAT3 and VarScan2/DNAcopy R packages) to analyse matched tumour and normal samples for copy number analyses. Using ASCAT, a gain (amplification) is called if (predicted total copy number – predicted ploidy) > 0.6 and a loss if (predicted total copy number – predicted ploidy) < –0.6. Using DNAcopy circular binary segmentation (CBS) segments, the logR ratio was used to discriminate copy number gains and losses (logR ratio > 0.15 = gain, logR < −0.15 = loss). We then selected only the consensus calls across the two methods. In addition, we visually inspected all CNA/LOH events using the logR and B-allele frequency (BAF) plots, and further included copy number gain and loss events when they were clearly supported by one method, but just below the cutoff threshold in the other. Genes targeted by copy gain, loss and copy-neutral LOH (cn-LOH) in each sample were further identified.

Identification of significantly amplified/deleted regions. To further identify significantly amplified and deleted regions in cSCC genomes, we applied GISTIC2.038 (q < 0.1) using the CNA segments and markers generated by VarScan variant calling and ASCAT R package39. Thirty-one of the 40 cSCC exomes that passed the ASCAT analysis were used. Regions that overlap with centromere and telomere were excluded. We further removed regions that were identified as both significantly amplified and deleted.

Tumour subpopulation identification and clonality analysis. EXPANS3 was used to estimate clonal expansions and cellular frequency of each clonal and subclonal population. Somatic SNVs, indels and CNA segments derived from DNAcopy CBS algorithm were used as input for each tumour. Tumour samples with less than 200 amplified mutations were excluded from this analysis, since stable results may not be solved for these samples, resulting in processing of 35 out of the 40 WES samples. Within each tumour sample, the largest clone was identified as the dominant clone and clones with lower cellular frequencies were regarded as sub-clones. Tumour clonality was also estimated based on the cellular frequency of the dominant clone. All somatic variants were assigned to their nested clones. We also used SciClone40 to estimate mutational clusters and Infer clonal architecture for 35/40 cSCC exomes. Similar to EXPANS, Somatic SNVs, indels with their VAFs along with CNA segments were used as input. The cluster with the largest number of mutations was regarded as the dominant clone for each sample, and subpopulations and tumour purity were subsequently derived. The clonality results and tumour purity from both methods were further compared and assessed.

Gene expression microarray data analysis and integration. Five sets of expression microarray data of patient samples were selected and downloaded from Gene Expression Omnibus (GEO), GSE42677,45, GSE45216,46, GSE50354, GSE84297 and GSE36288. The GSE42677 data set included 10 normal skin, 5 AK, 5 BCC and 5 invasive SCC samples, the GSE50354 data set included 6 normal skin, 4 AK and 5 SCC samples, the GSE84293 data set included 8 normal skin, 9 AK and 9 SCC samples and the GSE36288 data set included 13 normal skin, 14 AK and 15 SCC samples. The available normalised expression values in log scale were used. The GSE45216 data set had 10 AK and 30 SCC. For this set, we further included 16 non-sun-exposed (NSE), 20 sun-exposed (SE) normal skin and 18 AK samples profiled using the same platform, Affymetrix Human Genome U133 Plus 2.0 Array. All raw.CEL files of 94 samples were assembled and processed together. Quality control (QC) and normalisation were performed on our O-Miner transcriptomics analysis platform36, where the RMA (robust multi-array average) approach was used. For all five data sets, differential expression (DE) analyses between different groups were performed using limma R package37. The DE genes were identified using the adjusted p-value < 0.05. For SMGs, as well as significantly amplified and deleted genes identified above, we further investigated their expression profiles across normal skin, AK and SCC samples. Based on the normalised non-long, transformed expression values, we regarded probes with the expression values <100 as not being expressed or expressed at a low level. Thus, we further identified genes that were expressed across normal skin, AK and SCC samples in the five data sets. We then categorised them as genes with an activating role in the tumour development if they were significantly upregulated, or genes with a suppressing role if they were significantly downregulated in SCC/AK samples compared to normal skin, in at least three out of the five data sets.

For patient-derived cell lines, gene expression data were obtained from Illumina HumanHT-12 v4 Expression BeadChip. Raw image files were first processed using Illumina GenomeStudio, and then normalised using the O-miner platform where lumi R bioconductor package38 was applied. The limma R package was used for the DE analysis. The raw and normalised data have been deposited at GEO under the accession number of GSE98780.

Data availability
All data is provided in Supplementary Data 1-30, in GSE98780 and EGA500001002612.

Received: 31 July 2017 Accepted: 7 August 2018
Published online: 10 September 2018

References
1. Guy, G. P. Jr., Machlin, S. R., Ekwueme, D. U. & Yabroff, K. R. Prevalence and costs of skin cancer treatment in the U.S., 2002-2006 and 2007-2011. Am. J. Prev. Med. 48, 183–187 (2015).
2. Harwood, C. A. et al. A surveillance model for skin cancer in organ transplant recipients: a 22-year prospective study in an ethnically diverse population. Am. J. Transplant. 13, 119–125 (2013).
3. Karia, P. S., Han, J. & Schmults, C. D. Cutaneous squamous cell carcinoma: estimated incidence of disease, nodal metastasis, and deaths from disease in the United States, 2012. J. Am. Acad. Dermatol. 68, 957–966 (2013).
4. Harwood, C. A., Proby, C. M., Inman, G. J. & Leigh, I. M. The promise of genomics and the development of targeted therapies for cutaneous squamous cell carcinoma. Acta Derm. Venerol. 96, 3–16 (2016).
5. South, A. P. et al. NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. J. Invest. Dermatol. 134, 2630–2634 (2014).
6. Martincorena, I. et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science 348, 880–886 (2015).
7. Chitsazzadeh, V. et al. Cross-species identification of genomic drivers of squamous cell carcinoma development across preneoplastic interstates. Nat. Commun. 7, 12601 (2016).
8. Lee, C. S. et al. Recurrent point mutations in the kinetochore gene KNSTRN in cutaneous squamous cell carcinoma. Nat. Genet. 46, 1060–1062 (2014).
9. Li, Y. Y. et al. Genomic analysis of metastatic cutaneous squamous cell carcinoma. Clin. Cancer Res.: Off. J. Am. Assoc. Cancer Res. 21, 1447–1456 (2015).
10. Pickering, C. R. et al. Mutational landscape of aggressive cutaneous squamous cell carcinoma. Clin. Cancer Res.: Off. J. Am. Assoc. Cancer Res. 20, 6582–6592 (2014).
11. Dotto, G. P. & Rustgi, A. K. Squamous cell cancers: A unified perspective on biology and genetics. Cancer Cell 29, 622–637 (2016).
12. Bailey, P. et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 531, 47–52 (2016).
13. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 3, 246–259 (2013).
14. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. Nature 500, 415–421 (2013).
15. Cammareri, P. et al. Inactivation of TGFbeta receptors in stem cells drives cutaneous squamous cell carcinoma. Nat. Commun. 7, 12493 (2016).
16. Petljak, M. & Alexandrov, L. R. Understanding mutagenesis through delineation of mutational signatures in human cancer. Cancer Res. 77, 531–540 (2016).
17. Hollstein, M., Alexandrov, L. B., Wild, C. P., Ardin, M. & Zavadil, J. Base changes in tumour DNA have the power to reveal the causes and evolution of cancer. Oncogene 36, 158–167 (2017).
18. Alexandrov, L. B. et al. Mutational signatures associated with tobacco smoking in human cancer. Science 354, 618–622 (2016).
19. Droit, I. et al. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science 358, 234–238 (2017).
20. Alexandrov, L. B. et al. Clock-like mutational processes in human somatic cells. Nat. Genet. 47, 1402–1407 (2015).
21. Hameetman, L. et al. Molecular profiling of cutaneous squamous cell carcinomas and actinic keratoses from organ transplant recipients. BMC Cancer 13, 58 (2013).
22. Sekulic, A. et al. Loss of inositol polyphosphate 5-phosphatase is an early event in development of cutaneous squamous cell carcinoma. Cancer Prev. Res. 3, 1277–1283 (2010).
23. Salgado, R. et al. C60SB1 amplification is a frequent event in cutaneous squamous cell carcinoma with aggressive clinical behaviour. Genes, Chromosomes Cancer 49, 1054–1061 (2010).
24. Purdie, K. J. et al. Single nucleotide polymorphism array analysis defines a specific genetic fingerprint for well-differentiated cutaneous SCCs. J. Invest. Dermatol. 129, 1562–1568 (2009).

25. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 12, R41 (2011).

26. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499, 214–218 (2013).

27. Gonzalez-Perez, A., Deu-Pons, J. & Lopez-Bigas, N. Improving the prediction of the functional impact of cancer mutations by baseline tolerance transformation. Genome Med. 4, 89 (2012).

28. Tamborero, D., Gonzalez-Perez, A. & Lopez-Bigas, N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. Bioinformatics 29, 2238–2244 (2013).

29. Slattery, M. L. et al. Diet and lifestyle factors interact with MAPK genes to influence survival: the Breast Cancer Health Disparities Study. Cancer Causes Control.: CCC 25, 1211–1225 (2014).

30. Oyer, J. A. et al. Point mutation E1099K in MMSET/NSD2 enhances its hyperactivation in diffuse large B cell lymphoma. Blood 118, e32–e39 (2011).

31. An, C. H. et al. Frameshift mutations of vacuolar protein sorting genes in gastric and colorectal cancers with microsatellite instability. Hum. Pathol. 43, 40–47 (2012).

32. Alsafadi, S. et al. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat. Commun. 7, 10615 (2016).

33. Hollander, M. C., Blumenthal, G. M. & Dennis, P. A. PTEN loss in the leading edge of cutaneous squamous cell carcinoma by microarray expression profiling. J. Invest. Dermatol. 136, 2109–2119 (2015).

34. Purdie, K. J., Pourreyron, C. & South, A. P. Isolation and culture of squamous cell carcinoma cell lines. Methods Mol. Biol. 731, 151–159 (2011).

35. Schubbert, S., Shannon, K. & Bollag, G. Hyperactive Ras in developmental disorders and cancer. Nat. Rev. Cancer 8, 714–726 (2008).

36. Popp, S. et al. Genetic characterization of a human skin carcinoma progression model: from primary tumor to metastasis. J. Invest. Dermatol. 115, 1095–1103 (2000).

37. Garrett, G. L. et al. Incidence of and risk factors for skin cancer in organ transplant recipients in the United States. JAMA Dermatol. 153, 296–303 (2017).

38. Tembe, W. D. et al. Gene expression profiling of the leading edge of cutaneous squamous cell carcinoma: II. 24-driven MPM-7. J. Invest. Dermatol. 134, 1418–1427 (2014).

39. Mittus, H. et al. Gene expression profiling of the leading edge of cutaneous squamous cell carcinoma by microarray expression profiling. Mol. Cancer 5, 30 (2006).

40. Niid, I. et al. Identification of differentially expressed genes in cutaneous squamous cell carcinoma by microarray expression profiling. Mol. Cancer Res. 24, 52–60 (2014).

41. Popp, S. et al. Genetic characterization of a human skin carcinoma progression model: from primary tumor to metastasis. J. Invest. Dermatol. 118, 1095–1103 (2000).

42. Garrett, G. L. et al. Incidence of and risk factors for skin cancer in organ transplant recipients in the United States. JAMA Dermatol. 153, 296–303 (2017).

43. Alexandrova, L. B. & Stratton, M. R. Mutation signatures: the patterns of somatic mutations hidden in cancer genomes. Curr. Opin. Genet. Dev. 24, 52–60 (2014).

44. Ming, M., Zhao, B., Qiang, L. & He, Y. Y. Effect of immunosuppressants tacrolimus and mycophenolate mofetil on the keratinocyte UVB response. Pharmacogenomics 9, 242–247 (2008).

45. Herman, M. et al. Effect of cycloporsin A on DNA repair and cancer incidence in kidney transplant recipients. J. Lab. Clin. Med. 137, 14–20 (2001).

46. McCarroll, N. et al. An evaluation of the mode of action framework for mutagenic carcinogen case study: Cyclophosphamide. Environ. Mol. Mutagen. 49, 117–131 (2008).

47. O'Donovan, P. et al. Azathioprine and UVA light generate mutagenic oxidative DNA damage. Science 309, 1871–1874 (2005).

48. Karran, P. & Brem, R. Protein oxidation, UVA human and DNA repair. DNA Repair. (Amst.) 4, 178–185 (2016).

49. Harwood, C. A. et al. PTEN mutations in basal cell carcinomas from azathioprine-treated organ transplant recipients. Br. J. Cancer 99, 1276–1284 (2008).

50. Hoffbauer, G. F. et al. Reversal of UVA skin photosensitivity and DNA damage in kidney transplant recipients by replacing azathioprine. Am. J. Transplant. 12, 218–225 (2012).

51. Peacock, J. W. et al. SEMA3C drives cancer growth by transactivating multiple receptor tyrosine kinases via Plexin B1. EMBO Mol. Med. 10, 219–238 (2018).

52. Chung, W. et al. Quantitative proteomics identifies STEAP4 as a critical regulator of mitochondrial dysfunction linking inflammation and colon cancer. Proc. Natl Acad. Sci. USA 114, E9608–E9617 (2017).

53. Buis, E. M. et al. Inflammation-mediated skin tumorigenesis induced by epidermal c-Fos. Genes Dev. 27, 1959–1973 (2013).

54. Zhu, Z., Di, J., Lu, Z., Gao, K. & Zheng, J. Rap2B GTPase: structure, functions, and regulation. Tumor Biol. J. Int. Soc. Oncodev. Biol. Med. 37, 7085–7093 (2016).

55. Le Duff, M. et al. Regulation of senescence escape by the cdk4-EZH2-AP2M1 pathway in response to chemotherapy. Cell Death Dis. 9, 199 (2018).

56. Purdie, K. J., Pourreyron, C. & South, A. P. Isolation and culture of squamous cell carcinoma cell lines. Methods Mol. Biol. 731, 151–159 (2011).

57. Okosun, J. et al. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. Nat. Genet. 46, 176–181 (2014).

58. Okosun, J. et al. Recurrent mTORC1-activating RRAGC mutations in follicular lymphomas. Nat. Genet. 48, 183–188 (2016).

59. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–576 (2012).

60. Ramos, A. H. et al. Oncotator: cancer variant annotation tool. Hum. Mutat. 36, E2423–E2429 (2015).

61. Dayemullah, A. Z., Lemoine, N. R. & Chelala, C. SNpexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). Nucl. Acids Res. 40, W65–W70 (2012).

62. Gonzalez-Perez, A. et al. IntOGen-mutations identifies cancer drivers across tumor types. Nat. Methods 10, 1081–1082 (2013).

63. Tawana, K. et al. Disease evolution and outcomes in familial AML with germline CBPA mutations. Blood 126, 1214–1223 (2015).

64. Van Loo, P. et al. Allele-specific copy number analysis of tumors. Proc. Natl Acad. Sci. USA 107, 16910–16915 (2010).

65. Cutts, R. J. et al. O-miner: an integrative platform for automated analysis and mining of -omics data. Nucl. Acids Res. 40, W560–W568 (2012).

66. Ritchie, M. E. et al. Inflammation-mediated skin tumorigenesis induced by epidermal c-Fos. Genes Dev. 27, 1959–1973 (2013).

67. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. Bioinformatics 24, 1547–1548 (2008).
C.M.P. Wrote or contributed to the writing of the manuscript: G.J.I, J.W, A.N, R.G.T, V.S., L.C.S., C.A.H., C.M.P., I.M.L.

**Additional information**

*Supplementary Information* accompanies this paper at [https://doi.org/10.1038/s41467-018-06027-1](https://doi.org/10.1038/s41467-018-06027-1).

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

**Reprints and permission** information is available online at [http://npg.nature.com/reprintsandpermissions/](http://npg.nature.com/reprintsandpermissions/)

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