Inactivation of TGFβ receptors in stem cells drives cutaneous squamous cell carcinoma

Patrizia Cammareri1,13, Aidan M Rose2,13, David F Vincent1,13, Jun Wang3, Ai Nagano3, Silvana Libertini1, Rachel A Ridgway1, Dimitris Athineos1, Philip J Coates4, Angela McHugh2, Celine Pourreyron2, Jasbani H S Dayal2, Jonas Larsson5, Simone Weidlich6, Lindsay C Spender2, Gopal P Sapkota6, Karin J Purdie7, Charlotte M Proby2, Catherine A Harwood7, Irene M Leigh2,7, Hans Clevers8, Nick Barker9, Stefan Karlsson5, Catrin Pritchard10, Richard Marais11, Claude Chelala3, Andrew P South2,12, Owen J Sansom1* and Gareth J Inman2*.

1 Cancer Research UK Beatson Institute, Institute of Cancer Sciences, Glasgow University, Glasgow, G61 1BD, UK
2Division of Cancer Research, School of Medicine, University of Dundee, Dundee, DD1 9SY, UK
3Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, London, E1 4NS, UK
4Tayside Tissue Bank, School of Medicine, University of Dundee, Dundee, DD1 9SY, UK
5Molecular Medicine and Gene Therapy, Lund Strategic Center for Stem Cell Biology, Lund University, Lund, 221 00, Sweden
6MRC Protein Phosphorylation Unit, School of Life Sciences, University of Dundee, Dundee, DD1 5EH, UK
7Centre for Cutaneous Research, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, E1 2AT, London, UK
These authors contributed equally to this work.

*Correspondence: Gareth J Inman, Jacqui Wood Cancer Centre, Ninewells Hospital and Medical School, Dundee, DD1 9SY. g.j.inman@dundee.ac.uk; Owen J Sansom, Cancer Research UK Beatson Institute, Garscube Estate, Glasgow, G61 1BD. o.sansom@beatson.gla.ac.uk
Abstract

Melanoma patients treated with oncogenic BRAF inhibitors can develop cutaneous squamous cell carcinoma (cSCC) within weeks of treatment, driven by paradoxical RAS/RAF/MAPK pathway activation. Here, we identify frequent TGFBR1 and TGFBR2 mutations in human vemurafenib-induced skin lesions and in sporadic cSCC. Functional analysis reveals these mutations ablate canonical TGFβ Smad signaling which is localised to bulge stem cells in both normal human and murine skin. MAPK pathway hyperactivation (through Braf\textsuperscript{V600E} or Kras\textsuperscript{G12D} knockin) and TGFβ signaling ablation (through Tgfbr1 deletion) in LGR5\textsuperscript{+ve} stem cells enables rapid cSCC development in the mouse. Mutation of Tp53 (which is commonly mutated in sporadic cSCC) coupled with Tgfbr1 deletion in LGR5\textsuperscript{+ve} cells also results in cSCC development. These findings indicate that LGR5\textsuperscript{+ve} stem cells can act as cells of origin for cSCC and that RAS/RAF/MAPK pathway hyperactivation or Tp53 mutation, coupled with loss of TGFβ signaling, are driving events of skin tumorigenesis.

Introduction

The development of epithelial tumors is generally accepted to take place over several years, involving the accumulation of mutations which drive tumor progression\textsuperscript{1}. However, some tumors contain a relatively low mutation burden\textsuperscript{2} and develop rapidly, without progression from benign intermediary stages, suggesting a potential stem cell origin\textsuperscript{3}. Data from murine model systems illustrate a tumor’s ability to form from both stem and differentiated cells.
Within intestinal epithelium, loss of Apc in the LGR5^ve stem cell compartment leads to adenoma, whilst tumors rarely form from differentiated cells⁴. Conversely, we have shown that targeting Kras, in addition to Apc, can de-differentiate intestinal villi and permit tumor formation⁵. Thus, the tumor cell of origin remains unclear, as does the standard model of progression from benign tumor to malignant carcinoma.

Discord with the progression model is exemplified in the skin, which carries a high mutation burden⁶. Asymptomatic normal skin carries frequent mutations in TP53⁷,⁸ and NOTCH⁶,⁸.

Classic chemical carcinogenesis DMBA/TPA experiments demonstrate Hras mutations can lie dormant in the skin (without the addition of TPA), at no obvious consequence to the tissue⁹. Indeed even when Ras mutation is targeted to stem cell compartments (e.g LRIG1^ve cells or bulge stem cells¹⁰,¹¹), this does not lead to cancer unless there is a disruption of tissue homeostasis through wounding. These findings support the hypothesis that homeostasis within stem cell compartments plays an important tumor suppressive role in highly organised structures such as skin.

We reasoned that, in the absence of wounding, mutations in other oncogenic/tumour suppressor genes might facilitate rapid skin tumorigenesis. Using targeted sequence analysis and whole exome sequencing (WES), we identify frequent mutation in both TGFβ type 1 receptor (TGFBR1) and TGFβ type 2 receptor (TGFBR2) genes in human primary cSCC samples. IntOgen mutation analysis reveals TGFβ signaling as a pathway significantly altered by mutation and functional analysis of several TGFβ receptor mutants indicates that many of these mutations result in loss of function. Pathway activation studies reveal highly localised TGFβ signaling in both normal human and mouse hair follicle bulge stem cells. In murine skin, targeted activation of the RAS/RAF/MAPK pathway, coupled with deletion of
5 Tgfbr1 in LGR5+ve stem cells promotes rapid development of cSCC which, in the absence of wounding, may mimic the kinetics of tumor induction in vemurafenib-induced cSCC. Combined Tp53 mutation/inactivation coupled with Tgfbr1 loss in LGR5+ve stem cells also results in cSCC with longer latency providing a model for cSCC development without RAS activation.

RESULTS

TGFBR1 and TGFBR2 are frequently mutated in human cSCC

Cutaneous squamo-proliferative lesions (including keratoacanthomas and cSCC) arise in a significant proportion of patients treated with the type I RAF inhibitor vemurafenib. Such lesions develop within a few weeks of treatment. Targeted sequencing has revealed that these lesions contain a high frequency of activating mutations in HRAS. Cutaneous lesions isolated from patients treated with sorafenib (the “pan-RAF” inhibitor) also harbour mutations in HRAS, TP53 and TGFBR1. Employing targeted deep sequencing of 39 squamo-proliferative lesions from 7 patients (including cSCC and actinic keratosis (AK); Supplementary Table 1) treated with vemurafenib (using a percentage variance criterion of >10%) we identified frequent coding mutations in both TGFBR1 (8/39, 21% of samples) and TGFBR2 (5/39, 13% of samples) revealing mutation of TGFβ receptors in 28% of lesions (Fig. 1a, Supplementary Data 1). These mutational events were only surpassed in frequency by mutations in NOTCH1/NOTCH2 (56%) and activating mutations of HRAS (38%). TP53 mutations arose in 26% of lesions (Fig. 1a, Supplementary Data 2). In contrast to NOTCH, (using our mutational call cut off, see Methods) we did not detect any mutations in TGFβ receptors or HRAS in the normal or perilesional skin samples (n=6 from 4 patients, 3 of which had lesions containing TGFβ receptor mutations). These findings imply that a
combination of potential mutational inactivation of TGFβ signaling and activation of HRAS may be important driving events in vemurafenib-induced skin lesions and skin tumorigenesis.

We next sought to investigate whether loss of TGFβ signaling is a frequent event in sporadic cSCC. We employed targeted 454 pyrosequencing of TGFBR1 and TGFBR2 in 91 human primary cSCC samples (Supplementary Table 2) and 21 human cell lines derived from primary cSCC15, all of which were recently sequenced for common genetic alterations6. Using a percentage variance criterion of >10% we detected mutations of TGFBR1 in 22% and TGFBR2 in 30% of primary cSCC samples and 14% of cell lines (Fig. 1b,c and Supplementary Data 3). Overall, mutation of TGFβ receptors occurred in 43% of primary cSCC samples. These mutational events were only surpassed in frequency by mutations in NOTCH1/2 (86%), and this time TP53 (63%) (Fig. 1b, Supplementary Data 4, Ref 6). In sporadic cSCC oncogenic activation of RAS only occurred in 9% of samples (Fig. 1b, Supplementary Data 4, Ref 6). We then sequenced normal blood samples from 8 patients with sporadic cSCC whose lesions harboured mutations in TGFβ receptors (Supplementary Data 3) and found no TGFβ receptor mutations. Next we prospectively collected a further Dundee cohort of 7 primary cSCC samples with complementary matched normal distant and perilesional skin (Supplementary Table 3). This cohort demonstrated a comparable spectrum of mutation in our selected gene panel and in both TGFβ receptors (Fig. 1d, Supplementary Data 5). TGFβ receptor mutations were again not identified in either distant or perilesional skin. To assess the potential lesion specific, non-germline significance of TGFβ receptor mutations, we interrogated the pyrosequencing analysis in depth from all of the samples containing normal matched tissue (Supplementary Data 6). We observed only 8 variant reads out of 1348
reads in total in 4 out of 25 matched normal sample reads. Three of these samples were from peri-lesional skin and likely reflect rare contaminating tumour cells. In comparison we observed 237 variant reads out of 1340 reads in the tumour samples. Employing Fisher’s exact 2 sided tests to compare variant allele frequencies (VAFs) in matched samples, we determined that 17/25 of the TGFβ receptor mutations reached tumour specific VAF statistical significance confirming the lesion specific non-germline nature of these mutations (Supplementary Data 6).

TGFB1 and TGFB2 mutations are driver events in human cSCC

Next we examined a further cohort of 30 primary cSCC samples with matched normal tissue (Supplementary Table 4) employing next generation WES (see Methods) and interrogated in detail NOTCH1, NOTCH2, TP53, CDKN2A, HRAS, KRAS, NRAS, TGFB1 and TGFB2 genes for mutational and copy number changes (Fig. 2a and Supplementary Data 7). We observed alterations in all of these genes with a similar frequency to that of our previous 454 pyrosequencing analysis. None of the mutational events were found in the matched normal samples and all except two of these were statistically significant (Fisher’s exact t-test) (Supplementary Data 8). Importantly we observed changes in TGFB1 in 30% of the samples and changes in TGFB2 in 40% of the samples with a combined alteration in 53% of samples, confirming a frequent alteration of TGFβ receptor genes in cSCC. Copy number analysis also revealed that loss of heterozygosity (LOH) occurred in both TGFB1 and TGFB2 genes including in tumors with missense mutations in TGFB2 (Fig. 2a, Supplementary Data 8). Somatic single nucleotide variants (SNVs) of TGFβ receptors were detected in 30% of our samples consistent with our 454 pyrosequencing analysis and the recent sequencing analyses of two North American cSCC cohorts which, when combined,
detected TGFβ receptors protein altering SNVs in 15.7% of samples\textsuperscript{16,17}. Given the high mutational burden of cSCC, it is probable that many mutations identified will be passenger mutations with no functional consequence for tumorigenesis. We investigated the potential functional consequence of the mutations detected by WES employing MutsigCV\textsuperscript{18} and IntOgen analysis\textsuperscript{19}. MutsigCV detected TP53, CDKN2A, NOTCH1 and NOTCH2 as significant drivers but no RAS genes and IntOgen detected TP53, CDKN2A NOTCH1 and HRAS as significant drivers but did not identify NOTCH2, KRAS or NRAS (Supplementary Data 9). Neither analysis detected TGFBR1 or TGFBR2 individually as significant drivers (Supplementary Data 9), but IntOgen pathway analysis revealed TGFβ signaling as a significantly altered signaling pathway (Oncodrive-fm functional impact bias, FM bias\textsuperscript{19}, p=0.0019, Supplementary Data 10). We assessed the clonality of our candidate driver genes using the ABSOLUTE algorithm\textsuperscript{20}. WES data were of sufficient quality for 24/30 exomes and ABSOLUTE analysis revealed purity and ploidy estimates ranging from 0.2-0.73 and 1.78-5.79 respectively (Supplementary Data 11). ABSOLUTE clonality analysis indicated that all NOTCH1, CDKN2A, and RAS mutations were clonal as were all bar one TP53, three NOTCH2 and one TGFBR1 mutation which were subclonal (Fig. 2b,c, Supplementary Data 12). Mutations present in nearly all tumor cells (clonal) would suggest early events and therefore represent initiating “driver” genes as appears to be the case here for NOTCH1, NOTCH2, CDKN2A, HRAS, KRAS, TP53 and importantly both TGFBR1 and TGBFR2.

Having established the likely driver event of mutation of TGFBR1 and TGFBR2 in our WES data set we extended this analysis to include our samples assessed by targeted sequencing. We first calculated average percentage VAFs for our candidate drivers and these ranged from 48.7% for CDKN2A to 20% for TGFBR1 (Fig. 3a). TGFBR1 VAF was significantly lower
than that of CDKN2A, TP53, HRAS, NOTCH1, TGFBR2 and NOTCH2 but not KRAS and NRAS (Fig. 3a, Supplementary Data 13). TGFBR2 VAF was only statistically significantly lower than CDKN2A and TP53 but equivalent to KRAS, NOTCH2 and NRAS (Fig. 3a, Supplementary Data 13). The VAFs of the TGFβ receptors are of a similar range to those observed in other cSCC driver genes. UV light is the major oncogenic stimulus of cSCC and the % of mutations conforming to a UV signature (C-T or G-A transitions) of our candidate drivers ranged from 79.7% in CDKN2A to 30.4% in HRAS (Fig. 3b) with mutations in both TGFβ receptor genes lying within this range. VAFs were statistically significantly higher for UV signature mutations for NOTCH2, CDKN2A and TGFBR2 (Supplementary Fig. 1a, Supplementary Data 14). If these candidate genes represent potential driver genes then the mutational consequence should be predicted to change protein function. We classified these mutations as potentially damaging if they were predicted to be so by at least two of the four mutation function prediction programmes SIFT\textsuperscript{21}, PolyPhen-2\textsuperscript{22}, Provean\textsuperscript{23} and Mutation Assessor\textsuperscript{24} or were a splice site or PTC mutation (Fig. 3c). Damaging mutation rates ranged from 89% for TP53 to 53.5% for TGFBR1 (Fig. 3c, Supplementary Data 15-23) were statistically significantly higher for those with a UV signature for NOTCH2, TGFBR2 and TP53 (Fig. 3d, Supplementary Data 24) and damaging mutations had higher VAFs for NOTCH2, CDKN2A, TGFBR2 and NOTCH1 (Supplementary Fig. 1b, Supplementary Data 25). Together our data suggest that approximately 70% of TGFBR2 and 50% of TGFBR1 mutations will alter protein function with the potential to drive cSCC development. In its entirety our analysis conservatively estimates functionally relevant TGFBR1 and TGFBR2 mutations in ~10% and ~16% of samples respectively and therefore 20% of cSCC samples could harbour damaging TGFβ receptor mutations.
TGFβ receptor mutation inactivates canonical Smad signaling

Identified missense and nonsense mutations were found throughout the coding exons of both TGFBR1 and TGFBR2, occurring in the extracellular and kinase domains of each protein (Fig. 3e). Structural analysis of the extracellular domains of TGFBR1 (Supplementary Fig. 2) and TGFBR2 (Supplementary Fig. 3) indicated mutations occur in, or in close proximity to, highly conserved disulphide bonds, ligand interaction motifs and/or receptor interaction motifs. These findings suggest significant potential for loss of function.

TGFβ signals via activation of a heterotetrameric complex of TGFBR2:TGFBR1, resulting in TGFBR1-kinase driven c-terminal phosphorylation of SMAD2 and SMAD3. Once phosphorylated (PO4), SMAD2 and SMAD3 form hetero-oligomeric complexes with the co-Smad SMAD4, accumulate in the nucleus and regulate gene expression of hundreds of target genes. Activity of SMAD-dependent reporter gene constructs and steady state levels of SMAD2/3 c-terminal phosphorylation can be used as measures of canonical TGFβ signaling. To assess the functional consequence of these TGFβ receptor mutations we generated a panel of four TGFBR1 and five TGFBR2 mutant expression plasmids from mutations identified in our original targeted sequencing series. We assayed each mutant receptor for functional activity in transient transfection reporter gene assays. TGFBR1 expression plasmids were co-transfected into TGFBR1 null MEFS and TGFBR2 expression plasmids were co-transfected into TGFBR2 null T47D breast cancer cells, in addition to the TGFβ responsive reporter construct SMAD7 Promoter-luciferase (Fig. 4a and Fig. 4b respectively). Wild type TGFβ receptor expression elevated reporter activity over empty vector controls, which was further elevated by TGFβ treatment (Fig. 4a,b). We confirmed this activity was dependent on intact SMAD binding elements in the SMAD7 promoter.
(Supplementary Fig. 4a,b). The TGFBR1 mutants H331R and W277C and all of the TGFBR2 mutants (S474F, C486R, C96R, R2323W, A556T) failed to efficiently activate the reporter gene, despite similar levels of expression of the receptors, as assayed by western blotting (Fig. 4a,b). These findings indicate that mutation of TGFBR1 and TGFBR2 in cSCC frequently results in a loss of ability to activate canonical SMAD signaling. To demonstrate corollary of these findings in primary human tissue we then established conditions to monitor c-terminal PO4-SMAD3 levels using a c-terminal Ser433/Ser435 PO4-SMAD3 specific antibody in cSCC by immunohistochemistry (IHC) (Supplementary Fig. 5). We measured PO4-SMAD3 activity in 8 primary tumors harbouring wild type receptors and 8 primary tumors harbouring mutant TGFβ receptors with a combined VAF of >20% (Supplementary Data 26). Wild type tumors exhibited readily detectable PO4-SMAD3 activity whereas mutant tumors showed significantly reduced PO4-SMAD3 activity (Fig. 4c, Supplementary Fig. 6), consistent with our observation that mutation of TGFβ receptors results in loss of canonical SMAD signaling activity. Both wild type and mutant tumors exhibited heterogeneity of staining consistent with our previous observations that cSCC is heterogenous in nature\(^6\) and with the VAFs observed in mutant tumors.

Finally, we used primary human cSCC cell lines to assess whether TGFβ receptor mutation results in a loss of TGFβ signaling. Exogenous treatment of normal human keratinocytes (NHK) with TGFβ1 resulted in a dose dependent decrease in cell proliferation (Fig. 4d). The TGFBR2 mutant harbouring cell lines SCCIC8 and SCCIC12 (Supplementary Data 3) failed to respond to exogenous TGFβ stimulation by either PO4-SMAD activation (Supplementary Fig. 7) or by any effect on cell proliferation (Fig. 4d). Co-transfecting these TGFBR2 mutant cells with either empty vector, or wild type TGFBR2 expression plasmids in addition to a GFP expression plasmid, we measured cell proliferation in real-time using Incucyte-Zoom\(^{TM}\)
imager over 6 days. Cell proliferation of the GFP\textsuperscript{+ve} cells indicated that cells expressing wild type TGFBR2 proliferated at a slower rate in the presence of exogenous TGF\textbeta\textsuperscript{(Fig. 4e). The degree of inhibition was commensurate to the degree of restoration of SMAD activity as measured using the multimerised SMAD binding element reporter gene CAG\textalpha\textsubscript{12}-Luciferase\textsuperscript{32} (Supplementary Fig. 7c,d). These findings indicate that re-expression of wild-type TGFBR2 restores canonical TGF\textbeta\textsuperscript{signaling and proliferative inhibition, confirming mutational loss of TGF\textbeta tumor suppressive activity.

Matrix cells exhibit active TGF\textbeta\textsuperscript{signaling}

Given this potential aetiological loss of TGF\textbeta\textsuperscript{signaling, we sought to identify sites of active TGF\textbeta\textsuperscript{signaling in normal skin to gain insight into the cellular origin of cSCC RAF inhibitor induced lesions. PO\textsubscript{4}-SMAD3 activity was barely detectable by IHC analysis in normal human epidermis (Supplementary Fig. 8) but showed strong immunoreactivity in the hair matrix zone of anagen hair follicles (Fig. 5a, Supplementary Fig. 8). PO\textsubscript{4}-SMAD3 positivity was also detected in the hair matrix of anagen hair follicles in mouse back skin (Fig. 5a). In anagen, the hair follicle transit-amplifying (TA) cells are localized in the matrix and are positive for Sonic hedgehog (SHH)\textsuperscript{11}. Elegant studies by the Blanpain group have demonstrated these cells are unable to act as a cell of origin for papilloma formation, even when both oncogenic Kras and Tp53 were targeted\textsuperscript{11}. This suggests that these PO\textsubscript{4}-SMAD3\textsuperscript{+ve} hair matrix cells are unlikely to be the cell of origin for the rapid cSCC observed in humans following RAF inhibitor treatment. To investigate this in the mouse, we tested if Tgfbr1 deletion could permit the transformation of TA cells. RAF inhibitors stimulate paradoxical activation of the MAPK pathway in cells with wild-type BRAF harbouring upstream pathway activation, via mechanisms such as: up-regulated receptor tyrosine kinases, oncogenic RAS via RAF dimer
formation\textsuperscript{33-35}, or relief of inhibitory auto-phosphorylation\textsuperscript{36}. Circumventing pharmacological enhancement of MAPK signaling in the presence of mutated RAS, we modelled hyperactivation of the MAPK pathway in the SHH\textsuperscript{+ve} compartment by targeting downstream oncogenic \textit{Braf}\textsuperscript{V600E} and oncogenic activation of \textit{Kras}\textsuperscript{G12D}. We crossed our previously described \textit{LSL-Braf}\textsuperscript{V600E} mice\textsuperscript{37}, which allow inducible expression of \textit{Braf}\textsuperscript{V600E} from the endogenous \textit{Braf} gene, with the \textit{ShhCRE\textsuperscript{ER}} strain\textsuperscript{38}. This permits tamoxifen inducible activation of the Cre recombinase in SHH\textsuperscript{+ve} cells. To assess the role of TGF\textbeta signaling in the SHH\textsuperscript{+ve} cells, we then crossed these animals with \textit{Tgfbr1}\textsuperscript{fl} mice\textsuperscript{30} (Supplementary Fig. 9a,b). No tumors formed in the skin of \textit{ShhCRE\textsuperscript{ER} Braf}\textsuperscript{V600E} and \textit{ShhCRE\textsuperscript{ER} Braf}\textsuperscript{V600E} \textit{Tgfbr1}\textsuperscript{fl/+} mice (Fig. 5b, Supplementary Fig. 10a,b). A small percentage of \textit{ShhCRE\textsuperscript{ER} Braf}\textsuperscript{V600E} \textit{Tgfbr1}\textsuperscript{fl/fl} mice developed minimally proliferative papillomatous lesions (as evidenced by low level BrdU staining) mainly in the lips, but only at long latency (Fig. 5b, Supplementary Fig. 10c). No mice developed cSCC. Mice failed to develop any skin lesions following oncogenic activation of \textit{Kras}\textsuperscript{G12D} with or without deletion of \textit{Tgfbr1} in this cell compartment (Fig. 5c). Together these studies indicate that the SHH\textsuperscript{+ve} cells are unlikely to be the cell of origin for either rapid onset vemurafenib-induced cSCC, or sporadic cSCC.

**TGF\textbeta signaling is active in telogen bulge stem cells**

Approximately 90\% of human hair follicles are present in the anagen phase of the hair cycle with the remaining 10\% existing in catagen or the resting telogen phase. Analysis of human telogen hair follicles revealed highly localised PO\textsubscript{4}-SMAD3 staining in the bulge stem cells, characterised in part by KERATIN 15 staining (Fig. 6a). This pattern was recapitulated in mouse telogen hair follicles (Fig. 6b), characterised by the expression of the stem cell marker LGR5\textsuperscript{39}. To investigate further, we used the \textit{Lgr5-EGFP-Ires-CREERT2} knockin mouse.
(hereafter termed Lgr5CREER), where the endogenous Lgr5 promoter controls expression of enhanced green fluorescent protein (EGFP) and the CREERT2 fusion protein. IHC analysis for GFP revealed a staining pattern similar of that observed for PO4-SMAD3 (Fig. 6c). Furthermore, co-immunofluorescence revealed LGR5+ve cells (stained for EGFP) are highly enriched for both PO4-SMAD3 and TGFBR1 (Fig. 6d and Supplementary Fig. 11).

Recent studies indicate that the dermal papilla may provide a source of TGFβ2, activating SMAD signaling in overlying hair germ stem cells. We sorted epithelial EGFP positive LGR5+ve stem cells from murine back skin. Quantitative reverse-transcriptase PCR (Q-RTPCR) analysis revealed LGR5+ve cells express enhanced levels of Tgfbr1, Tgfb1 and Tgfb3 mRNA when compared to LGR5-ve cells, with negligible amounts of Tgfb2 (Fig. 6e). Expression of Tgfbr2 was readily detected in GFP+ve and GFP-ve compartments (Fig. 6e). This indicates enriched autocrine TGFβ signaling in the LGR5+ve compartment. We observed high levels of the TGFβ target gene Smad7 in LGR5+ve cells (Fig. 6e). Together these findings indicate that autocrine TGFβ signaling is highly localised to the LGR5+ve hair follicle bulge stem cells in the mouse, and the KERATIN 15+ve hair follicle bulge stem cells in humans, and that this cell compartment may give rise to both vemurafenib-induced and sporadic cSCC.

Rapid cSCC formation from Lgr5+ve stem cells

To investigate the consequence of hyperactivation of the MAPK pathway coupled with ablation of TGFβ signaling in LGR5+ve stem cells we crossed the Lgr5CREER mice with the LSL-BrafV600E mice (Supplementary Fig. 9c), or LSL-KrasG12D mice and Tgfbr1fl mice (Supplementary Fig. 9d). Loss of TGFβ signaling alone was not sufficient to initiate tumorigenesis (Fig. 7a). Targeted activation of BRAF to LGR5+ve cells resulted in decreased
survival, with all mice euthanized 300-days post induction of the transgene by injection of tamoxifen (Median survival 276 days). Whilst 6 out of 14 mice succumbed to adrenal tumors, 50% of these mice presented with papillomas consistent with LGR5 expression in murine skin (Fig. 7a,c). However, the long latency period suggests Braf mutation requires additional events to facilitate papilloma development. The combined targeted inactivation of one allele of Tgfbr1 reduced survival (median survival 231 days) and enhanced both the number of mice with skin lesions, and the number of lesions per mouse (Fig. 7a,c and Supplementary Fig. 12a). Inactivation of both Tgfbr1 alleles significantly increased the numbers of tumor per mouse and dramatically shortened both skin tumor free survival (all mice developing skin lesions within 63 days of induction) and overall survival (median survival 51 days) (Fig. 7a,c and Supplementary Fig. 12a). Phenotypically these lesions appeared as differentiated papillomas in Tgfbr1 wild-type and heterozygous mice (Fig. 7c-e and Supplementary Fig. 12b). Remarkably, in the homozygous Tgfbr1fl/fl mice tumors appeared as ulcerative cSCC (Fig. 7c and Supplementary Fig. 12b). Elegant work by the Blanpain and Jensen laboratories10,11 have shown when Kras is targeted to skin stem cells, there is long latency to papilloma formation (similar to the BrafV600E allele described here) and most of these lesions form around areas associated with wounding. Targeted activation of Kras alone mainly failed to produce skin lesions, however, when we targeted inactivation of both alleles of Tgfbr1 and the KrasG12D mutation to the LGR5+ve compartment, mice developed rapid cSCC with kinetics comparable to BrafV600E mice (Fig. 7b,c). Additionally, KERATIN1 (Fig. 7d) and KERATIN 5 staining (Fig. 7e) revealed that cSCC lesions in both the Braf and Kras mice are poorly differentiated cSCC. Importantly these lesions were highly proliferative (Supplementary Fig. 12c) and never progressed via a papillomatous stage, recapitulating the rapid cSCC onset observed in humans12,13. PO4-SMAD3 activity exhibited a
dose dependent reduction in tumors isolated from these mice, indicating loss of TGFβ signaling (Supplementary Fig. 12d,e). Q-RTPCR analysis of these tumors revealed loss of Tgfbr1 expression (Supplementary Fig. 12f) without any significant change in ligand mRNA expression (Supplementary Fig. 13).

Skin tissue compartmentalization has been recently proposed as a mechanism involved in tissue maintenance. To test if Tgfbr1 deletion perturbed such compartmentalization, we lineage traced LGR5+ve cells by intercrossing Lgr5CreER with the Rosa^LSL-RFP reporter mice (Supplementary Fig. 9e). We observed that RFP positive cells were confined to the hair follicle and were never detected in the sebaceous gland, or interfollicular epidermis regions of Lgr5Cre ER Braf^V600E, or Lgr5Cre ER Braf^V600E Tgfbr1^fl/+ mice, at early time points post induction (Supplementary Fig. 14). The cSCC arising within Lgr5Cre ER Braf^V600E Tgfbr1^fl/fl mice were fully recombined and RFP positive. In the normal skin comparator for these tumours, but also at earlier time points, the LGR5+ve cells and their progeny were localised in their normal compartment (Supplementary Fig. 14). These results indicate that perturbation of TGFβ signaling is insufficient to disrupt compartmentalization, but acts as a tumour suppressor in LGR5+ve stem cells.

Given the infrequent coincident activation of RAS genes and mutational inactivation of TGFβ receptors in sporadic cSCC we finally sought to model this disease by inactivating TP53 function coupled with deletion of Tgfbr1 in LGR5+ve cells (Supplementary Fig. 9f). Knockin of mutant Tp53 (R172H) coupled with deletion of the wild type allele had no discernible phenotype (Fig. 8a). Heterozygous knockin or deletion of Tp53 coupled with homozygous deletion of Tgfbr1 resulted in the emergence of skin tumours in a few mice (30% and 20%
respectively) with long latency. Combined knockin of mutant Tp53 with deletion of the wild
type allele of Tp53 coupled with deletion of Tgfbr1 resulted in skin tumor development in
81% of mice with increased tumor number at a shorter latency (Fig. 8a,b). These tumors
exhibited loss of differentiation expressing low levels of KERATIN 1 and higher levels of
KERATIN 5 (Fig. 8c).

Discussion
Recent studies have revealed an exceptionally high mutation burden (50 mutations per
megabase of DNA) in cSCC. This rate is second only to that of the commonest skin
malignancy basal cell carcinoma. This translates to potentially thousands of mutations per
tumor, providing a particular challenge in identifying driver mutations. This challenge is
further compounded by varying efficiencies in deep sequencing technologies and profound
tumor heterogeneity. Our studies here reveal that targeted deep sequencing using
fluidigm PCR amplification and Roche 454 pyrosequencing can provide a robust platform to
identify mutations in NOTCH1, NOTCH2, TGFBR1 and TGFBR2 genes. This approach has also
implicated alterations of NOTCH, TP53 and RAS in cSCC tumor development. We further
these studies by revealing mutation of TGFβ receptors in 43% of sporadic human cSCC and
28% of vemurafenib-induced skin lesions (Fig. 1). The prevalent tumor initiating event in
cSCC is UV-induced damage, which manifests as C-T and G-A transitions. Approximately
68% of all nucleotide changes observed in our cSCC samples present with this signature.
Analysis of mutational signatures in TGFβ receptors reveals that 42% conform to a UV
signature (Fig. 3b, Supplementary Data 13). This figure increases to 56.1% when scored as
possibly damaging events via protein function prediction programmes (Fig. 3d,
Supplementary Data 24). This indicates that UV damage may also be responsible for inactivation of TGFβ receptors. Mutation prediction programmes scored 53.5% of TGFBR1 and 71.1% of TGFBR2 receptor mutants as damaging indicating that approximately 20% of cSCC harbour TGFβ receptor inactivation (Supplementary Data 15-16, Fig. 3c). Subsequent functional analysis of 4 TGFBR1 mutants and 5 TGFBR2 mutants indicated that half of the TGFBR1 mutants and all five TGFBR2 mutants were loss of function for canonical Smad signaling and that tumors harbouring TGFβ receptor mutations had reduced PO4-SMAD3 activity (Fig. 4). Restoration of TGFBR2 expression to TGFBR2 null cell lines restricted cell proliferation (Fig. 4). Taken together these findings indicate that loss of TGFβ tumour suppressor function is a common event in cSCC.

The assessment of VAFs provides an indication of the clonality of tumors and aids the potential identification of early driver mutationss during tumor development2. We ranked mutational events by potential order of occurrence in the 7 genes we have previously implicated in cSCC development and TGFβ receptors by measurement of VAF (Supplementary Data 15-16). These analyses indicate that potentially damaging mutations in TGFBR1 occur early in 25% of tumors harbouring these mutations and in 42% of tumors harbouring potentially damaging TGFBR2 mutations. Although this analysis is limited to the 9 genes studied in-depth here (but importantly including NOTCH genes previously identified as gatekeeper mutations in cSCC5), 11 samples exhibited TGFβ receptor mutations with the highest VAF indicating that this could be an initiating event in the development of cSCC. Strong support for this hypothesis comes from ABSOLUTE clonality analysis of our WES samples which revealed that 7/8 TGFβ receptor mutations were clonal and represent probable driver events in these lesions.
Several of our samples display VAFs of 10-20% for the TGFβ receptors, which is not too dissimilar to the VAFs of the other known cSCC tumor suppressors studied here and likely reflects the heterogeneous nature of cSCC. It is however an intriguing possibility that in some cases low TGFβ receptor VAF may reflect spontaneous regression of TGFβ receptors mutant clones as observed in multiple self-healing squamous epithelioma (MSSE) patients who harbour germline mutations in **TGFBR1**. TGFβ signaling has been demonstrated to play both positive and negative roles in cSCC development in various mouse models acting to limit tumor cell proliferation but also to promote tumor initiating capacity and drug resistance. Effects are dependent on the timing of aberrant TGFβ signaling and the cooperating oncogenic driving events (reviewed in 48,50-52). We provide evidence that a dose-dependent loss of TGFβ signaling drives tumor progression - emphasising its role as a major tumor suppressor in the skin. Although the cell of origin in cSCC in humans remains poorly defined, our observations indicate that tumors can initiate efficiently and rapidly from LGR5^+ve stem cells compared to TA cells, and mutational modulation of two signaling pathways within this cellular compartment is sufficient to drive rapid progression directly to carcinoma, without the need for protracted tumor evolution. The kinetics of this event mimic exactly that observed during development of cutaneous lesions in RAF inhibitor treated patients. Intriguingly this rapid process requires MAPK pathway activation as targeted interference with TP53 function coupled with Tgfbr1 loss results in the development of skin tumours with long latency. Importantly our studies revealed highly localised TGFβ signaling in KERATIN 15^+ve bulge stem cells in human telogen hair follicles, mirrored exquisitely the location of specific autocrine TGFβ signaling activity identified in LGR5^+ve bulge stem cells of murine telogen hair follicles. It has been proposed that stem cell
quiescence acts as a tumor suppressive mechanism in murine skin and that LGR5^+ve stem cells are refractory to oncogenic transformation\textsuperscript{53}. Our data clearly indicate that oncogenic activation of the RAS/RAF/MAPK pathway, or TP53 modulation, coupled with loss of TGFβ signaling, is capable of leading to tumor development from this compartment. As we demonstrate that mutational inactivation of TGFβ receptors is a frequent event in human cSCC, and that TGFβ signaling is highly localised to stem cells in normal skin, we propose that these cells represent a cell of origin for human cSCC. It remains possible that loss of TGFβ signaling may also contribute to cSCC development from other cell compartments in the skin and this warrants further investigation.

Our data, both in human and mice, indicate that TGFβ signaling inactivation can be an initiating event in sporadic cSCC. This is clearly the case in MSSE where germline loss of function mutations in \textit{TGFBR1} have been identified as the underlying genetic lesion\textsuperscript{47}. We speculate that activation of the RAS/RAF/MAPK pathway, or p53 modulation, may be a cooperating event in the development of this disease and that these tumors may originate from the bulge stem cell compartment. TGFβ signaling inactivation may also occur following the acquisition of other driving mutational events and act as a limiting factor for tumor development. Intriguingly, initial clinical trials targeting systemic TGFβ inhibition with GC1008 (a pan-TGFβ neutralising antibody) have also reported the occurrence of spontaneous cSCC as a side effect\textsuperscript{54}. This provides further compelling support for the tumor suppressive role of TGFβ in skin carcinogenesis.

\textbf{Methods}
**Samples.** Ethical approval for this investigation was obtained from the East London and City Health Authority and the Tayside Tissue bank local ethics committee and the study was conducted according to the Declaration of Helsinki Principles. All patients participating in this study were from dermatology and plastic surgery units in the UK and all provided written, informed consent. Punch biopsies of cSCC tissue were collected and processed as previously reported. Normal human keratinocytes were isolated from normal skin samples according to previously published protocols. Human tumor cell lines SCCIC1, SCCIC4, SCCIC8, SCCIC15, SCCIC12, SCCIC18, SCCIC19, SCCIC21, SCCT1, SCCT2, SCCT6, SCCT8, PM1, MET1, MET4, SCCT9, SCCT10, SCCT11, RDEBSCC2, RDEBSCC3, RDEBSCC4 and NTERT cells were established by our laboratories and were cultured as described. TGFBR1 null MEFs and T47D cells (ATCC) were maintained in DMEM and RPMI supplemented with 10% FCS respectively. Mycoplasma contamination checks were carried out on all cultures as routine and all lines were confirmed mycoplasma negative.

**454 sequencing.** TGFBR1 and TGFBR2 primers were designed and validated by Fluidigm (Fluidigm Corporation, San Francisco, CA) as per recommended guidelines for Roche Titanium sequencing (Roche, Mannheim, Germany). Primers for NOTCH1, NOTCH2, TP53, CDKN2A, HRAS, KRAS and NRAS were previously described and all primer sequences are listed in Supplementary Data 27. Each primer included sample-specific Fluidigm 454 barcode primer and adapter sequences. Sequencing was performed in the same manner as our previous study. Briefly for thermal cycling a Fluidigm FC1 Cycler was used. The libraries were normalized and pooled prior to purification using Agencourt AMPure XP system (Beckman, UK). Library components were clonally amplified utilising the GSJunior emPCR
Lib-A Kit (Roche) by inputting 1 molecule of library DNA per capture bead. Pyrosequencing was done using the GS Junior system (Roche/454 Life Sciences).

**454 variant analyses.** Variant analysis was performed as previously described\(^6\). Briefly, reads were mapped to the hg19 build of the human genome using LASTZ via the public GALAXY instance and filtered to exclude those mapping to <100 loci using tools available through GALAXY.

**Coding Variants and splice site detection.** Pileup files were generated and filtered using SAMTools\(^57\). Variants present in a single read or less than 10% of the total reads were excluded using a custom java program available from [https://github.com/mattsouth/laszt-variant-filter](https://github.com/mattsouth/laszt-variant-filter) (last accessed 6th May 2013). Coding variants were called against the RefSeq gene list using the amino acid tool via the public GALAXY instance. Variants present in <3 reads were excluded. Variants present in >1 independent sample and adjacent to a homopolymer >3 bases were excluded unless present in COSMIC\(^58\). Variants present in >30 samples were excluded unless present in COSMIC. All variants present in the exome variant server database ([http://evs.gs.washington.edu/EVS/](http://evs.gs.washington.edu/EVS/)) were excluded unless present in COSMIC. Splice sites were called from the pileup variant list if present in >4 reads and within 2 bases of Refseq coding sequence using Excel (Microsoft Inc., CA).

**Whole exome sequencing data analysis.** 20 previously published cSCC whole exomes\(^6\) were re-analysed with the addition of 10 new cSCC whole exomes with the overall mean coverage of 63x (Supplementary Table 4), using a previous pipeline\(^59\). SNVs and short indels were identified using the Strelka pipeline\(^60\) with a minimum coverage of 10 reads at the targeted
sites. Annotation of somatic variants was performed using the Oncotator tool. Mutations in our targeted genes were further identified across the 30 cSCC WES samples.

**Copy number analysis using WES data.** Two independent approaches were applied. First, to generate SNP and indel variant genotyping information, the tumor-normal pair was processed together against the reference genome using the VarScan2 germline variants calling method mpileup2cns. The minimum coverage for identified sites was 10 reads for both tumor and normal. Next the logR and BAF (B-allele frequency) files were created based on the tumor-normal pair genotyping information, with the depth information normalized by dividing the depth of each variant by the median depth across all variants. The ASCAT R packages were then used to perform allele-specific copy number analysis to identify copy number aberrations (CNA) and loss-of-heterozygosity (LOH) regions. The second approach was based on numbers of reads aligned to each exon between the tumor and normal pair. VarScan2 copy number calling method was firstly applied. Raw copy number calls were adjusted as previously reported. Finally results from the two approaches were cross-compared to produce the final CNA and acquired uniparental disomy (aUPD) calls for targeted genes.

**Identification of potential cancer drivers and significantly mutated pathways.** Based on all mutations identified from the 30 cSCC WES dataset, we used the IntOGen platform to identify significantly mutated genes and pathways, based on the significance (p-value) of the FM bias (i.e., the bias toward the accumulation of mutations with high functional impact). The significantly mutated signaling pathways (based on the IntOGen Oncodrive-fm functional impact bias, FM bias $p < 0.05$) were further selected (Supplementary Data 10).
MutsigCV\textsuperscript{18} was also used to detect significant genes with point mutations above the background mutation rate.

**Estimating the clonality of mutations.** For the somatic mutations of *TGFBR1/2, TP53, CDKN2A, NOTCH1/2* and *RAS* genes identified by WES (Supplementary Data 7), we further classified them as clonal or subclonal on the basis of the posterior probability that the cancer cell fraction (CCF) exceeded 0.95 using ABSOLUTE\textsuperscript{20}. Numbers of reads supporting the reference and alternative alleles were extracted, and the copy number segmentation files were generated based on the DNAcopy CBS segments using WES data. Mutations with the somatic clonal probability > 0.5 were classified as clonal with high confidence. Those mutations with clonal probability > 0.25 but with very small subclonal probability scores were also called clonal (Supplementary Data 12). Tumor purity and ploidy were also estimated (Supplementary Data 11). For samples with *TGFBR1/2* mutations, CCFs for *TGFBR1/2* were further compared to those for other genes to determine the clonality orders.

**Functional prediction of mutations.** A combination of four approaches were used to predict the functional impact of identified mutations by targeted sequencing, (i) SIFT\textsuperscript{21}, which uses sequence homology and protein conservation to predict the effects of all possible substitutions at each position in the protein sequence; (ii) PolyPhen-2\textsuperscript{22}, which predicts possible functional impact of an amino acid substitution on the structure and function level using physical and comparative considerations; (iii) Provean\textsuperscript{23}, which predicts the damaging effects of SNVs and indels using a versatile alignment-based score; and (iv) Mutation Assessor\textsuperscript{24}, which measures the functional impact scores for amino acid residue changes.
using evolutionary conservation patterns derived from aligned families and sub-families of
sequence homologs within and between species. Mutations predicted as functional
damaging by at least two of the four approaches were classified being potentially
damaging/deleterious.

**In Vivo analyses.** All experiments were performed under the UK Home Office guidelines.
Mice were segregating for C57BL6J and S129 background. Alleles used throughout this study
were: Lgr5-cre-ER\textsuperscript{T240}, ShhCre\textsuperscript{ER38}, Braf\textsuperscript{V600E37}, Kras\textsuperscript{G12D65}, Tgfbr1\textsuperscript{f30} and Rosa\textsuperscript{LSL-RFP66}. A mix of
males and females were used. Recombination in the Lgr5-cre-ER\textsuperscript{T2} mouse model was
induced with 1 intraperitoneal (IP) injection of 3 mg Tamoxifen (Sigma) followed by 1
injection of 2 mg Tamoxifen for 3 days. Mice were induced post 7 weeks of age.
Recombination in the ShhCre\textsuperscript{ER} mouse model was induced with 1 IP injection of 2.5 mg
Tamoxifen. Mice were induced post 28 days of age. For proliferation analysis mice were
injected with 250 µl of BrdU (Amersham Biosciences) 2 hours before being sacrificed.

**FACS analysis.** Epidermis was prepared as previously described\textsuperscript{67}. Briefly, fat was scraped
from the mouse back and left at 37°C in a dish (dermis down) in 0.25% of Trypsin/EDTA
(Invitrogen) for 90 min. Epidermis was removed using a scalpel and dissociated by pipetting.
Cells were filtered through a 40 µm strainer, centrifuged at 250g for 5 min and washed with
PE (PBS/EDTA). Cells were washed with 0.1% BSA/PE, centrifuged at 250g for 5 min and
used for Lgr5-GFP sorting.
**Immunohistochemistry.** Immunohistochemistry (IHC) was performed on formalin-fixed skin sections. Standard IHC techniques were used throughout this study. Primary antibodies were as follows: TGFBR1 (Santa Cruz, V22, 1:100), PO$_4$-SMAD3 (Abcam, EP823Y, (52903), 1:50), GFP (Abgent, 168AT1211, 1:100), KERATIN1 (Covance, AF109, 1:1000), KERATIN 5 (Covance, AF138, 1:4000), KERATIN 15 (Abcam, 80522 (LHK15), 1:1000), KI67 (Thermo, RM-9106-S), BrdU (BD Biosciences, 347580, 1:200;). Mouse PO$_4$-SMAD3 score was performed in a blinded fashion. For each antibody, staining was performed on at least 3 mice of each genotype and at least six sections of normal human skin. Representative images are shown for each staining. PO$_4$-SMAD3 antibody was optimised for IHC use using FFPE embedded SCCIC4 cells treated with and without recombinant TGFβ1 or the TGFBR1 kinase inhibitor SB-431542$^{68}$ (Supplementary Figure 5). PO$_4$-SMAD3 IHC scoring was performed in a blinded manner using the histoscore method.

**RNA isolation and quantitative PCR** RNA was isolated using a Qiagen RNeasy Mini Kit (Qiagen, Crawly, West Sussex, UK) according to the manufacturer’s instructions. DNA-free (Ambion/Applied Biosystems, Warrington, UK) was used to remove genomic DNA contamination according to the manufacturer’s instructions. 1µg of RNA was reverse transcribed to cDNA using a DyNAmo SYBR Green 2-step qPCR kit (Finnzymes, Espoo, Finland) in a reaction volume of 20 µl. GAPDH was used to normalize for differences in RNA input.

**qRT-PCR primers.** qRT–PCR primers were as follows. mTgfbr1 F-TGCCATAACCGCACTGTCA, mTgfbr1 R-AATGAAAGGGCGATCTAGTGATG, mTgfbr2 F-CCGGAA GTTCTAGAATCCAG,
Western Blotting. Cells were lysed directly in 4xSDS sample buffer at 60-80% confluence. Lysates were subjected to standard SDS-PAGE. Bands were detected using enhanced chemiluminescence solution (ECL, Amersham). Secondary antibodies used throughout: HRP-conjugated polyclonal goat anti-mouse Ig (Dako, P0448, 1:2000), HRP-conjugated polyclonal goat anti-rabbit Ig (Dako, P0260, 1:2000). Primary antibodies were PO4-SMAD3 (Abcam, 52903, 1:1000), SMAD3 (Cell Signaling, 9523, 1:1000), TGFBR1 (Santa Cruz, 398 (V22), 1:500), TGFBR2 (Santa Cruz, 17792, (E6), 1:500). For TGFBR2 western blots, lysates were prepared directly from transfected cells using the Dual-luciferase cell lysis buffer (Promega). For TGFBR1 western blots, parallel transfections to the luciferase assays were performed and samples were lysed directly in 4xSDS sample buffer. Original uncropped western blot scans are also provided (Supplementary Figure 15).

Plasmids. The full-length wild type human TGFBR1 and pathogenic mutants, amplified with BgIII/NotI restriction sites, were shuttled into pCMV5 mammalian cell expression vectors onto the BamHI/NotI sites. The full-length wild type human TGFBR2 and pathogenic mutants were sub-cloned into pCMV5 using the BamHI/NotI restriction sites. Site-directed
mutagenesis was carried out using the QuickChange method (Stratagene) but substituting the Taq with KOD Hot Start DNA polymerase (Novagen). All DNA constructs were verified by DNA sequencing (by the DNA Sequencing Service at University of Dundee; www.dnaseq.co.uk). GFP expression plasmid was from Amaxa.

**Transient transfection analysis.** All transfections were performed in 24-well format in biological triplicate using LipofectAMINE 2000 (Invitrogen, Carlsbad, CA) according to manufacturer’s instructions. Cells were transfected overnight with 100ng of reporter gene (SMAD7 Promoter-Luciferase or CAGA_{12}-Luciferase) and 10ng of internal Renilla-Luciferase control (pRL-TK, E2241, Promega) with empty vector (pCMV5, 211175, Stratagene), wild type or mutant TGFβ receptor plasmids (range 150-300ng). Recombinant human TGFβ1 (Peprotech) was dissolved in 4mM HCL/1mg/ml BSA and used at final concentration of 5ng/ml and cells were treated for 4 hours prior to harvest. Luciferase activities were measured using the Dual Luciferase assay (Promega) and firefly luciferase activity was normalised to renilla luciferase activity.

**Cell proliferation assays.** Cells were seeded at a density of 500-1000 cells/well of 96-well plates in keratinocyte media (RM⁺) without growth factors and incubated overnight. Cells were fed 50μl of medium supplemented with treatment and controls every 2 days until harvest. All cultures were performed in sextuplet (n=6). Cells were assayed for proliferation using the CellTitreGlo™ Luminescent Cell Viability assay (NHKs) as per the manufacturer’s instructions (Promega, UK - Luminescence was measured on a Berthold Orion II microplate luminometer) or IncucyteZoom™ Live cell imager.
Data availability

The whole exome sequencing data for the 30 samples have been deposited in the European Genome-phenome Archive (EGA) under accession code EGAS00001001892. The authors declare that all other relevant data supporting the findings of this study are available within the article and its supplementary information files. Additional information can be obtained from the corresponding authors (GJI and OJS).

References


Acknowledgments

OS is supported by a Cancer Research UK core grant (A21139) and an ERC starting grant (311301). PC is supported by FP7 Health CP-IP - Large-scale integrating project grant (278568), DFV is supported by ERC Starting grant (311301). AMR is supported by Cancer Research UK centre grant (A12976). LCS was supported by WWCR grant 11-0788. GJI was supported by WWCR fellowship (03-0900). GJI, IML, CMP, CAH, KJP, AM, CP and APS were supported by a Cancer Research UK programme grant (A21358) and an ERC grant (250170).

Thanks to the research and scientific services at the CRUK Beatson Institute in general.

Author Contributions

OJS, GJI, PC, AMR, DV, JW and APS contributed to study design; PC, AMR, DFV, SL, DA contributed to the data acquisition; PC, AMR, DFV, JW, AN, SL, RAR, DA, PJCV, AM, CP, JHSD, JL, SW, LCS, GPS, KJP, CMP, CAH, IRL, HC, NB, SK, CP, RM, CC, APS, OJS, GJI contributed to the data analysis and interpretation of the data; OJS, GJI, PC, AMR, APS, JW to drafting the manuscript.
Competing financial interests: the authors declare that they have no competing financial interest.

Figure Legends

Figure 1. TGFβ receptors are frequently mutated in vemurafenib-induced skin lesions and sporadic cSCC tumors. Mutation frequency, distribution and relationship with pathological features of skin lesions isolated from vemurafenib treated patients (n=39) (a), sporadic cSCC (n=91) (b), 21 cSCC cell lines (RDEB = recessive dystrophic epidermolysis bullosa) (c), (a-c, adapted from J Invest Dermatol. (2014) with permission from Elsevier (ref 6) and 7 sporadic cSCC tumors (T) and normal distant/perilesional skin samples (N) collected in Dundee (d). Mutation status for nine genes is indicated and the overall percentage mutation is shown on the left. Each column represents a single case. Colors correspond to specific mutations as shown. Details of clinical parameters are included in Supplementary Tables 1-3 and mutations are included in Supplementary Data 1-5.

Figure 2. Target gene mutation and copy number and clonality analysis in 30 cSCC primary tumors analysed by whole exome sequencing. (a) Mutation frequency, distribution and relationship with pathological features from 30 cSCC primary tumors. Mutation and copy number status (gain, loss and aUPD) for nine genes is indicated and the overall percentage alteration is shown on the left. Each column represents a single case. Colors correspond to specific mutations and copy number changes as shown. Split columns indicate where more than one mutation type is present in a single case. Details of clinical parameters and mutations are included in Supplementary Table 4 and Supplementary Data 7. (b) ABSOLUTE clonality analysis of potential driver genes of cSCC indicates that all nine genes are frequently clonal. (c) Cancer cell fraction clonality analysis indicates clonal and subclonal mutations in the indicated tumors.
Figure 3. Mutational frequencies and spectrum of driver mutations. Ranked driver gene mutations by (a) % VAF (p-values represent student’s t-test (2-tailed) are shown above the figure, (b) % UV spectrum (c) predicted mutational consequence of Damaging/Non-damaging and (d) combined UV and damaging analysis (p-values are shown above the figure and represent Chi-squared Fishers exact-test). For all statistics; * defines statistical significance (*** p<0.001, ** p<0.01 and * P<0.05). Numbers of samples are contained in Supplementary Data 13. (e) Domain structures of TGFBR1 and TGFBR2 are shown. Exons are numbered and functional domains colour coded (see key). SNPs identified in TGFBR1 and TGFBR2 are labelled in AA sequence [using UniProtKB codes: P-36897-1, TGFBR1, HUMAN and P37173, P37173-2, TGFBR2_HUMAN] for sporadic cSCC (above) and Vemurafenib associated cSCC (below). Asterisked SNPs are those found in cSCC cell lines. Amino acid numbers for TGFBR2 refer to isoform 2.

Figure 4. Mutation of TGFβ receptors results in loss of function. (a) Indicated TGFBR1 plasmids were co-transfected into TGFBR1 null MEFs and assayed for SMAD7-Promoter Luciferase (SMAD7-Luc) reporter gene activity and receptor expression levels by western blot (lower panels) with and without TGFβ stimulation for 4 hours. β-ACTIN is used as a loading control. EV is empty vector control, WT is wild type. Data are mean +/- s.d., n=3. (b) Indicated TGFBR2 plasmids were co-transfected into TGFBR2 null T47D cells and assayed for SMAD7-Promoter Luciferase (SMAD7-Luc) reporter gene activity and receptor expression levels by western blot (lower panels) with and without TGFβ stimulation for 4 hours. β-ACTIN is used as a loading control. EV is empty vector control, WT is wild type. Data are mean +/- s.d., n=3. (c) PO₄-SMAD3 activity was assessed by IHC in wild type and mutant...
tumors (n=8, *** p=0.001, Mann-Whitney U test) Representative images are shown. Scale bar, 100 μM. (d) Effects of TGFβ stimulation on growth of NHK and cSCC cell lines. Data represent Cell Titre Glo measurement of cell proliferation over the indicated time course of cells treated with the indicated dose of TGFβ1. Normal human keratinocytes (NHK) and cell lines harbouring mutant TGFBR2 (SCCIC8, SCCIC12) are shown. Data represent the mean +/- s.d. n=6. (e) Restoration of wild type TGFBR2 restores growth inhibition. SCC1C8 and SCCIC12 cells were co-transfected with empty vector control (EV) or wild type TGFBR2 expression plasmids (TGFBR2) and a GFP expression plasmid. Proliferation of GFP+ve cells was assessed using real-time Incucyte Zoom imaging over 6 days. Data represent the mean +/- s.d. n=6. *, ** and *** = p< 0.05, p<0.01 and p<0.001 respectively (Student’s t-test).

Figure 5. Transit-amplifying SHH positive cells do not allow cSCC development. (a) PO4-SMAD3 IHC in anagen human (left panel) and mouse (right panel) hair follicles reveals immunoreactivity in the hair matrix. CTL connective tissue layer, ORS outer root sheath, IRS inner root sheath, DP dermal papilla. Scale bar, 100 μm (b) Kaplan-Meier survival curve of \( \text{ShhCre}^{\text{ER}} \text{Braf}^{\text{V600E}} \) (n=13), \( \text{ShhCre}^{\text{ER}} \text{Braf}^{\text{V600E}} \text{Tgfbr1}^{\text{fl/+}} \) (n=28) and \( \text{ShhCre}^{\text{ER}} \text{Braf}^{\text{V600E}} \text{Tgfbr1}^{\text{fl/fl}} \) (n=39) mice. (c) Kaplan-Meier survival curve of \( \text{ShhCre}^{\text{ER}} \text{Kras}^{\text{G12D}} \) (n=17) and \( \text{ShhCre}^{\text{ER}} \text{Kras}^{\text{G12D}} \text{Tgfbr1}^{\text{fl/fl}} \) (n=21) mice.

Figure 6. TGFβ signaling is active in LGRS+ve stem cells. (a) IHC analysis of PO4-SMAD3 (left panels) and KERATIN 15 (right panels) in human normal skin. Insert shows strong PO4-SMAD3 staining in the telogen hair follicle KERATIN 15+ve bulge stem cells. Scale bar, 100 μm (b) IHC analysis of PO4-SMAD3 (IFE inter-follicular epidermis, SG sebaceous gland, BG Bulge, DP dermal papilla) and (c) LGR5-GFP in murine skin in the telogen phase of the hair cycle.
Scale bar, 100 µm (d) Immunofluorescence (IF) analysis of LGR5-GFP and PO4-SMAD3 in murine telogen skin. Nuclei are counterstained with DAPI. (e) Q-RTPCR analysis of Tgfbr1, Tgfbr2, Tgfbr1, Tgfbr2, Tgfbr3 and Smad7 in LGR5\textsuperscript{+ve} (n=3 biological replicates) and LGR5\textsuperscript{-ve} cells (n=3 biological replicates) freshly isolated from back skin in the telogen phase. Data are shown as ratios to the internal Gapdh control with error bars representing s.e.m.. Statistical significance *p=0.04 (Mann-Whitney U-test, one-tailed test).

**Figure 7.** Deletion of Tgfbr1 coupled with BRAF/KRAS activation leads to skin tumorigenesis. (a) Kaplan-Meier survival curve (left panel) of Lgr5Cre\textsuperscript{ER} Tgfbr1\textsuperscript{fl/+} (n=12), Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E} (n=14), Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E} Tgfbr1\textsuperscript{fl/+} (n=23) and Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E} Tgfbr1\textsuperscript{fl/fl} (n=26) mice (p≤0.0001 by Log-Rank (Mantel-Cox)). Skin tumor free survival curve (right panel) of Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E} (n=5), Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E} Tgfbr1\textsuperscript{fl/+} (n=10) and Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E} Tgfbr1\textsuperscript{fl/fl} (n=19) mice. (b) Kaplan-Meier survival curve of Lgr5Cre\textsuperscript{ER} Kras\textsuperscript{G12D} (n=9) and Lgr5Cre\textsuperscript{ER} Kras\textsuperscript{G12D} Tgfbr1\textsuperscript{fl/fl} (n=14) mice (p<0.0001 by Log-Rank (Mantel-Cox)). (c) Macroscopic pictures of skin tumours from Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E}, Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E} Tgfbr1\textsuperscript{fl/+}, Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E} Tgfbr1\textsuperscript{fl/fl} and Lgr5Cre\textsuperscript{ER} Kras\textsuperscript{G12D} Tgfbr1\textsuperscript{fl/fl} mice. (d) Representative staining of KERATIN 1 (K1) and (e) KERATIN 5 (K5) on Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E}, Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E} Tgfbr1\textsuperscript{fl/+}, Tgfbr1\textsuperscript{fl/fl}, Lgr5Cre\textsuperscript{ER} Braf\textsuperscript{V600E} Tgfbr1\textsuperscript{fl/+} and Lgr5Cre\textsuperscript{ER} Kras\textsuperscript{G12D} Tgfbr1\textsuperscript{fl/fl} mice. Scale bar, 100 µm.

**Figure 8.** Deletion of Tgfbr1 coupled with Tp53 mutation/deletion leads to skin tumorigenesis. (a) Kaplan-Meier survival curve of Lgr5Cre\textsuperscript{ER} Tp53\textsuperscript{R172H/fl} (n=31), Lgr5Cre\textsuperscript{ER} Tp53\textsuperscript{R172H/fl} Tgfbr1\textsuperscript{fl/fl} (n=10) and Lgr5Cre\textsuperscript{ER} Tp53\textsuperscript{R172H/fl} Tgfbr1\textsuperscript{fl/+} (n=5) and Lgr5Cre\textsuperscript{ER} Tp53\textsuperscript{R172H/fl} Tgfbr1\textsuperscript{fl/fl} (n=11) mice (p≤0.0001 by Log-Rank (Mantel-Cox)). (b) Tumor number of indicated
genotypes. (c) Macroscopic picture of a Lgr5Cre\textsuperscript{ER} Tp53\textsuperscript{R172H/f} Tgfbr1\textsuperscript{f/f} mouse with a skin tumour. Representative staining of H&E, KERATIN 1 (K1) and KERATIN 5 (K5) is shown. Scale bar, 100 µm.
### Vemurafenib-induced skin lesions (n=39)

<table>
<thead>
<tr>
<th></th>
<th>Squamoproliferative / AK (n=27)</th>
<th>SCC (n=12)</th>
<th>Normal Skin (n=6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOTCH1</td>
<td>49%</td>
<td>29%</td>
<td>30%</td>
</tr>
<tr>
<td>NOTCH2</td>
<td>21%</td>
<td>14%</td>
<td>13%</td>
</tr>
<tr>
<td>TPS3</td>
<td>13%</td>
<td>29%</td>
<td>30%</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>5%</td>
<td>29%</td>
<td>23%</td>
</tr>
<tr>
<td>HRAS</td>
<td>16%</td>
<td>8%</td>
<td>13%</td>
</tr>
<tr>
<td>NRAS</td>
<td>13%</td>
<td>21%</td>
<td>14%</td>
</tr>
<tr>
<td>KRAS</td>
<td>22%</td>
<td>8%</td>
<td>5%</td>
</tr>
<tr>
<td>TGFBR1</td>
<td>5%</td>
<td>13%</td>
<td>8%</td>
</tr>
<tr>
<td>TGFBR2</td>
<td>30%</td>
<td>5%</td>
<td>5%</td>
</tr>
</tbody>
</table>

### Sporadic cSCC (n=91)

<table>
<thead>
<tr>
<th></th>
<th>Well Differentiated cSCC (n=31)</th>
<th>Moderately Differentiated cSCC (n=31)</th>
<th>Poorly Differentiated cSCC (n=29)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOTCH1</td>
<td>75%</td>
<td>63%</td>
<td>49%</td>
</tr>
<tr>
<td>NOTCH2</td>
<td>63%</td>
<td>63%</td>
<td>63%</td>
</tr>
<tr>
<td>TPS3</td>
<td>23%</td>
<td>23%</td>
<td>23%</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>16%</td>
<td>16%</td>
<td>16%</td>
</tr>
<tr>
<td>HRAS</td>
<td>13%</td>
<td>13%</td>
<td>13%</td>
</tr>
<tr>
<td>NRAS</td>
<td>5%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>KRAS</td>
<td>22%</td>
<td>22%</td>
<td>22%</td>
</tr>
<tr>
<td>TGFBR1</td>
<td>30%</td>
<td>30%</td>
<td>30%</td>
</tr>
<tr>
<td>TGFBR2</td>
<td>30%</td>
<td>30%</td>
<td>30%</td>
</tr>
</tbody>
</table>

### cSCC Cell Lines (n=21)

<table>
<thead>
<tr>
<th></th>
<th>UV cSCC (n=18)</th>
<th>RDER Cell Lines (n=21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOTCH1</td>
<td>49%</td>
<td>43%</td>
</tr>
<tr>
<td>NOTCH2</td>
<td>21%</td>
<td>43%</td>
</tr>
<tr>
<td>TPS3</td>
<td>13%</td>
<td>43%</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>8%</td>
<td>43%</td>
</tr>
<tr>
<td>HRAS</td>
<td>16%</td>
<td>14%</td>
</tr>
<tr>
<td>KRAS</td>
<td>14%</td>
<td>5%</td>
</tr>
<tr>
<td>NRAS</td>
<td>14%</td>
<td>5%</td>
</tr>
<tr>
<td>TGFBR1</td>
<td>14%</td>
<td>14%</td>
</tr>
<tr>
<td>TGFBR2</td>
<td>14%</td>
<td>5%</td>
</tr>
</tbody>
</table>

### Dundee cSCC (n=7)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NOTCH1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOTCH2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPS3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDKN2A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRAS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KRAS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRAS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGFBR1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGFBR2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1**

- **Activating RAS**: Insertion/Deletion, Splice Site, Missense
- Primary cSCC
- Matched normal skin
Figure 2

Panel a: Heatmap showing the percentage of mutations across different stages of differentiation for genes NOTCH1, NOTCH2, TP53, CDKN2A, HRAS, KRAS, NRAS, TGFBR1, and TGFBR2. The bars represent the percentage of mutations in Well Diff, Mod Diff, and Poor Diff stages.

Panel b: Bar graph illustrating the percentage of clonal and subclonal mutations for genes NOTCH1, CDKN2A, HRAS, TP53, NOTCH2, and TGFBR1/2. The number of samples for each gene is provided as N = 12, 5, 4, 15, 23, and 7, respectively.

Panel c: Scatter plot showing the cancer cell fraction (CCF) for TGFBR1/2, NOTCH2, and other genes. The x-axis represents TGFBR1/2, and the y-axis represents other genes. The plot includes points for TGFBR1/2 Others, NOTCH2(3) MD04, NOTCH2(2) MD04, NOTCH2(1) MD04, and TP53 WD12.
Figure 3

(a) 

(b) 

(c) 

(d) 

(e)
Figure 5

**a**
Human

Mouse

- PO_2 SMAD3
- ORS
- CTL
- IRS
- Hair Matrix

**b**

- ShhCRE\textsuperscript{ER} Bra\textsuperscript{V600E}
- ShhCRE\textsuperscript{ER} Bra\textsuperscript{V600E} Tgfbr\textsuperscript{fl/fl}
- ShhCRE\textsuperscript{ER} Kras\textsuperscript{G12D}
- ShhCRE\textsuperscript{ER} Kras\textsuperscript{G12D} Tgfbr\textsuperscript{fl/fl}

**c**

Percent survival

Days

Percent survival

Days

Figure 5
Figure 6

(a) PO₄-SMAD3 and KERATIN 15

(b) IFE, BG, SG, DP

(c) Lgr5-GFP

(d) DAPI, Lgr5-GFP, PO₄-SMAD3, Merge

(e) Tgfbr1, Tgfbr2, Tgfb1, Tgfb2, Tgfb3, Smad7

* indicates statistical significance.
Figure 8

(a) Percent survival over days for different genotypes:

- Lgr5CreER TP53\textsuperscript{R172H/fl}
- Lgr5CreER TP53\textsuperscript{R172H/+} Tgfr1\textsuperscript{fl/fl}
- Lgr5CreER TP53\textsuperscript{fl/+} Tgfr1\textsuperscript{fl/fl}
- Lgr5CreER TP53\textsuperscript{R172H/fl} Tgfr1\textsuperscript{fl/fl}

(b) Number of tumors per mouse for different genotypes:

(c) Histological images:

- TP53\textsuperscript{R172H/fl} Tgfr1\textsuperscript{fl/fl}
- H&E
- K1
- K5