Whole genome landscapes of major melanoma subtypes

Nicholas K. Hayward[1,2*#], James S. Wilmott[1,3*], Nicola Waddell[2,4*], Peter A. Johansson[2*], Matthew A. Field[5], Katia Nones[2,4], Ann-Marie Patch[2,4], Hojabr Kakavand[3], Ludmil B. Alexandrov[6], Hazel Burke[1], Valerie Jakrot[1], Stephen Kazakoff[2,4], Oliver Holmes[2,4], Conrad Leonard[2,4], Radhakrishnan Sabarinathan[7,8], Loris Mularoni[7,8], Scott Wood[2,4], Qinying Xu[2,4], Nick Waddell[4], Varsha Tembe[9], Gulietta M. Pupo[9], Ricardo De Paoli-Iseppi[3], Ricardo E. Vilain[3], Ping Shang[3], Loretta M.S. Lau[10], Rebecca A. Dagg[11], Sarah-Jane Schramm[9], Antonia Pritchard[2], Ken Dutton-Regester[2], Felicity Newell[2], Anna Fitzgerald[12], Catherine A. Shang[12], Sean M. Grimmond[13], Hilda A. Pickett[10], Jean Y. Yang[14], Jonathan R. Stretch[1], Andreas Behren[15], Richard F. Kefford[1,16], Peter Hersey[1,17], Georgina V. Long[1,18], Jonathan Cebon[15], Mark Shackleton[19], Andrew J. Spillane[1], Robyn P. M. Saw[1], Núria López-Bigas[7,8,20], John V. Pearson[2,4#], John F. Thompson[1#], Richard A. Scolyer[1,3,21#], Graham J. Mann[1,9#§]

[1] Melanoma Institute Australia, The University of Sydney, NSW, Australia

[2] QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

[3] Discipline of Pathology, Sydney Medical School, The University of Sydney, Sydney, NSW, Australia

[4] Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD, Australia

[5] Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, QLD, Australia

[6] Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

[7] Research Program on Biomedical Informatics, IMIM Hospital del Mar Medical Research Institute, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain

[8] Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Barcelona, Spain

[9]Centre for Cancer Research, Westmead Institute for Medical Research, The University of Sydney, Westmead, NSW, Australia

[10] Children's Medical Research Institute, The University of Sydney, Westmead, NSW, Australia

[11] Children's Hospital at Westmead, The University of Sydney, Westmead, NSW, Australia

[12] Bioplatforms Australia, North Ryde, NSW, Australia

[13] University of Melbourne Centre for Cancer Research, University of Melbourne, Parkville, VIC, Australia

[14] School of Mathematics and Statistics, The University of Sydney, NSW, Australia

[15] Olivia Newton-John Cancer Research Institute, Latrobe University, Austin Health, Heidelberg, Vic, Australia

[16] Macquarie University, North Ryde, NSW, Australia

[17] Centenary Institute, The University of Sydney, NSW, Australia

[18] Department of Medical Oncology, Royal North Shore Hospital, New South Wales, Australia

[19] Peter MacCallum Cancer Centre and University of Melbourne, Melbourne, VIC, Australia

[20] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

[21] Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, New South Wales, Australia

*These authors contributed equally

#These authors jointly directed the work

§ Correspondence should be addressed to: Graham J. Mann

Summary

Cutaneous, acral and mucosal subtypes of melanoma were evaluated by whole-genome sequencing. The heavily mutated landscape of coding and non-coding mutations in cutaneous melanoma resolved novel signatures of ultraviolet radiation mutagenesis. However, acral and mucosal melanomas were dominated by structural changes and mutation signatures of unknown aetiology, not previously identified in melanoma. The number of genes affected by recurrent mutations disrupting non-coding sequences was similar to that affected by recurrent mutations to coding sequences. Significantly mutated genes included *BRAF*, *CDKN2A*, *NRAS* and *TP53* in cutaneous melanoma, *BRAF*, *NRAS* and *NF1* in acral melanoma and *SF3B1* in mucosal melanoma. Mutations affecting the *TERT* promoter were the most frequent of all, however, neither they nor *ATRX* mutations, which correlate with alternative telomere lengthening, were associated with greater telomere length. Most melanomas had potentially actionable mutations, the majority in components of the mitogen-activated protein kinase and phosphoinositol kinase pathways.

Melanoma genomes have the highest mutation load of any cancer[1,2] and a predominant C>T nucleotide transition signature attributable to ultraviolet radiation (UVR)[1]. However, some melanomas lack this signature and have fewer point mutations[3,4]. Non-UVR driven melanomas are uncommon in cutaneous sites, but dominate those occurring in eyes, mucosal surfaces, and acral (hands and feet) sites; the latter account for a much higher proportion of melanomas in Asians[5] than Europeans. Elucidating mutational processes and genomic drivers in all subtypes is key to further progress in melanoma epidemiology, prevention and targeted treatment globally.

The Cancer Genome Atlas (TCGA) study of 333 melanomas[2] excluded acral, mucosal and ocular subtypes and used exome and low-pass whole genome sequencing (WGS). Examination of mutational process signatures, and of structural changes that drive melanoma was restricted, and mutations outside the exome were not surveyed. Other sequencing studies have included acral and mucosal melanomas[6,7], but have been limited by sample size and/or restriction to exome sequencing.

We present the first large, high-coverage WGS study of melanomas from cutaneous, acral and mucosal sites, revealing distinct mutation processes, profiles and drivers across subtypes, the landscape of non-coding mutations, and paradoxical relationships between telomere maintenance gene mutations and telomere length. Mutational signatures of UVR exposure dominated cutaneous melanomas, while structural variants formed the majority of

aberrations in acral and mucosal melanomas. These latter subtypes fundamentally differ in pathogenesis and therapeutic targets from cutaneous melanoma.

Study sample and mutation burden

The 183 melanoma samples comprised 75 primary tumours, 93 metastases and 15 cell lines derived from metastases; 35 were acral melanomas (arising from palms, soles or nail beds of the hands or feet), eight were mucosal and the remainder cutaneous (Supplementary Table 1). WGS identified 20,894,255 substitutions (range 1,610 to 777,275 per tumour) and 96,467 small insertions/deletions (indels) (range 0 to 10,114 per tumour) at an average rate of 38.23 mutations per megabase (Mb) (range 0.54 to 260 mutations/Mb) (Supplementary Table 1 and Fig. 1), the 156,770 coding mutations (range 19 to 5,536 per tumour) comprising 0.8% (Supplementary Table 2). Single nucleotide variant (SNV) and indel frequencies varied by melanoma subtype (t-test, P <1 X 10^{-7}), with >18 times more of these mutations in cutaneous (mean 49.17 mutations/Mb, range 0.71 to 259.69 mutations/Mb) than in acral and mucosal melanomas (mean 2.64 mutations/Mb, range 0.54 to 15.25 mutations/Mb) (Fig. 1b). Somatic structural variants (SV) were more frequent in acral and mucosal (mean 342.40, range 6 to 1,148) than in cutaneous melanomas (mean 101, range 3 to 1,123) (t-test, P <1 X 10^{-6}) (Fig. 1c,d). Greater proportions of the acral and mucosal melanoma genomes showed copy number variation (CNV), as in previous exome sequencing[3]; more regions were amplified (Mann-Whitney test, P < 1.46 X 10^{-4}) and more genomes were aneuploid (chi-square test, P = 0.049) than in cutaneous melanomas (Extended Data 1).

UVR and non-UVR mutation spectra

An algorithm using the sequence context of each mutation[1] identified 12 signatures, which varied considerably by melanoma subtype (Fig. 2a-c). Most melanomas (total n = 139; 136 cutaneous) were dominated by three novel signatures of UVR exposure (signatures 7a, 7b, and 7c), recapitulating UVR signature 7 previously identified by exome sequencing[1] (Extended Data 2). The predominantly C>T substitutions at TpC dinucleotides (mutated base underlined) of signature 7a are likely due to repair of 6,4-photoproducts[8]. Signature 7b involves C>T substitutions at CpC and CpC, characteristic of cyclobutane pyrimidine dimers[8]. Signature 7c has high levels of T>C and T>A mutations, potentially due to indirect DNA damage after UVR. Forty-four tumours, comprising 40 acral or mucosal melanomas and four cutaneous, all from infrequently sun-exposed sites, showed striking differences from the UVR-dominant group (Fig. 2). Six of nine signatures observed in this group have been described in other cancers[9] but never previously in melanoma and are presumably unrelated to UVR exposure. The most frequent were signature 1, the result of endogenous

spontaneous deamination of 5-methylcytosine, and signature 5, ubiquitous in cancers of many types and of unknown aetiology[10] (Extended Data 2). Signature 13, attributed to activity of AID/APOBEC cytidine deaminases, occurred in 52 samples (28%) and all melanoma subtypes. Signatures 8, 17 and 18, of unknown aetiology, occurred in all melanoma subtypes at a combined frequency of 7-28%. Signature 6, associated with defective mismatch repair, was seen in the single cutaneous melanoma with a high MSIsensor[11] score.

Structural variants, telomere length

The combined frequency of all SV (deletions, duplications, tandem duplications, and foldback inversions), was higher in acral and mucosal melanomas than in cutaneous melanomas (t-test, $P < 1 \times 10^{-6}$) (Fig. 1c and Supplementary Table 3). Acral and mucosal melanomas exhibited substantially more clusters of breakpoints, indicating more frequent complex rearrangements such as breakage-fusion-bridge and chromothripsis (Extended Data 3a). We detected complex rearrangements including breakage-fusion-bridge-like in 18% (*n* = 32) of tumours, most of which (*n* = 20, 63%) were acral or mucosal melanomas (Extended Data 4, 5).

WGS provided the first opportunity to assess effects of mutagenic processes and driver mutation profiles of melanoma on immortalisation via telomere length[12,13] (Methods and Extended Data 6b). Telomere length varied widely among the 183 samples, ranging from 6-fold shorter to 10-fold longer than the matched normal for each patient (Extended Data 6a). Telomere length was not correlated with melanoma subtype (Extended Data 6c), chromothripsis, or breakage-fusion-bridge events (Extended Data 3).

Non-coding mutations

All mutations were annotated for genomic context including regulatory, untranslated and protein coding regions, the first such comprehensive survey in melanoma and one of the first in any cancer. *TERT* promoter mutations[14,15] were most common (Sanger sequencing), with 115/167 (69%) overall and 86% of cutaneous melanomas mutated at one or more of four positions upstream of the initiation codon (Extended Data 6d). These mutations created new binding sites for the ETS family transcription factor GABP[16], and were mutually exclusive with the exception of two samples with extremely high mutation burdens. In keeping with previous findings for acral melanoma[17], these mutations were less prevalent in acral and mucosal melanomas (4/38: 11%), accounting for their association with lower SV and higher SNV burden (Fisher's exact test, $P < 1 \times 10^{-17}$). There was no association between *TERT* promoter and *BRAF* mutation (chi-square test, $P > 0.09$), in contrast to previous reports[18].

These putatively activating *TERT* promoter mutations were associated with modestly reduced telomere length (Mann-Whitney, $P = 0.0022$) (Extended Data 6e-g). Telomere elongation in cancer depends either on *TERT* or a process of alternative lengthening of telomeres (ALT)[19], often driven by mutations in *ATRX*[20]. The latter occur mutually exclusively with *TERT* promoter mutations in glioblastoma[21]. Ten melanomas (all cutaneous) had predicted loss-of-function mutations in *ATRX,* including two nonsense mutations, but paradoxically nine of these tumours also carried mutations in the *TERT* promoter and there was no association between *ATRX* mutation and telomere length (data not shown). Therefore, mutational activation of *TERT* does not simply facilitate telomere length extension, and the relationships between alternative telomere maintenance mechanisms, telomere length and immortalisation in melanoma must be complex.

All genes with recurrent mutations in the promoter, 5'-UTR or 3'-UTR are listed (Supplementary Tables 4-6). To detect non-coding elements with an accumulation of functional mutations we used OncodriveFML[22], which detects signals of positive selection in mutations of coding and non-coding elements. Because of the high mutation rate at active transcription factor binding sites (TFBS) in melanoma[22], we used a version of OncodriveFML (version 1.1) that permutes mutations locally to assess if the mutations observed disrupt or create TFBS more often than expected by chance. Promoter regions of nine genes were significantly (q-value < 0.25) biased towards high impact mutations: creating new TFBS in *BLCAP*, *KBTBD8*, *TERT* and *ZNF778*, disrupting them in *NSUN6* and *RNF185*, and with other impacts in *RALY*, *RPL29* and *RPS27* (Supplementary Table 4). Only in the case of *TERT* (as previously reported[15]), *RNF185* and *RPS27* was an effect of these mutations on gene expression supported. In a further six genes with five or fewer occurrences of a promoter mutation (*NUDT10*, *PNPLA2*, *CCP110*, *UAP1*, *UBAC2* and *ZWINT*) OncodriveFML predicted creation of novel TFBS, however too few tumours had such mutations for expression comparisons with wild type tumours to be informative.

Recurrent mutations in the promoter of *NFKBIE* were present in 10/183 (5%) tumours including 3/9 desmoplastic melanomas[23]. The commonest (chr6: 44233400) occurred in five melanomas, all cutaneous but lacking a desmoplastic component (Extended Data 7b). Mutations at this and one other position (chr6: 44233417) have been reported previously, but the other four sites are novel (Extended Data 7b) and presumably lie outside the captured exome in the previous study. *NFKBIE* promoter mutations are thus enriched in desmoplastic melanomas (Fisher's exact test, $P < 0.0084$), but are also potential drivers of other cutaneous melanomas.

The most frequent 5'-UTR hotspot mutation, detected in 19/183 samples, was at the transcription start site of *RPS27* (chr1: 153963239 C>T)[2,24]and a dinucleotide mutation was detected in one patient (chr1: 153963239-153963240 CT>TC). *RPS27* also had several recurrent promoter mutations (Extended Data 7c), none predicted to cause gain or loss of TFBS. Recurrent mutations also occurred in the 5'-UTR of *PES1* (Extended Data 7d), *RPS14, CHCHD2* and *MRPS31* (Supplementary Table 5), the latter identified as a common hotspot coding mutation by TCGA[2]. The 5'-UTR regions of *ERGIC3*, *PSMD8* and *TGS1* were significantly biased towards high functional impacting mutations and were expressed in most tumours, however none showed expression differences associated with mutation status so their significance as drivers remains uncertain.

TCGA did not report 3'-UTR mutations; we observed recurrent (n≥5) 3'-UTR mutations in 18 genes (Supplementary Table 6). Only 11 of these genes were expressed in the tumour cohort and none has been associated with melanoma, except *PTPRT* which was recently reported to have frequent coding mutations[25].

Key coding mutations and TCGA genomic subgroups

Two tools were used to identify significantly mutated genes (SMG) in each of the three melanoma subtypes (MutSigCV and OncodriveFML, q<0.1; Supplementary Table 7). Four cutaneous melanoma SMG were supported by both methods (*BRAF*, *CDKN2A*, *NRAS* and *TP53*), and five (*ARID2*, *CWH43, NF1*, *PTEN* and *RB1*) were supported by one method. *SF3B1* was identified as a SMG, for the first time in mucosal melanoma, by OncodriveFML. In acral melanoma, OncodriveFML identified *BRAF*, *KIT, MAP2K2, NF1* and *NRAS* as SMG.

Mutation frequencies of previously reported melanoma driver genes[2,3,25-29], varied considerably between melanoma subtypes. For example, no acral or mucosal melanomas had mutations in *TP53, PTEN, DDX3X, RASA2, PPP6C, RAC1* or *RB1,* indicating that the molecular pathways driving these melanoma subtypes differ markedly from those driving cutaneous melanoma. In line with previous studies[30], *KIT* mutations occurred more frequently in acral (3/35) and mucosal melanomas (2/8) than in cutaneous melanomas (6/140) (chi-square test, $P < 0.044$). The distribution of mutations in selected melanoma genes (Supplementary Table 8) is shown in Fig. 3. SV and CNV contributed heavily to the overall mutation frequency of some melanoma driver genes, especially *NF1*, *TP53*, *PTEN* and *KIT* (Extended Data 1b,c; Fig. 3). *BRAF* hotspot mutations were usually amplified, and these SV tended to involve *MET*, adjacent on chromosome 7. Similarly, *KIT* and *PDGFRA,* co-located within 0.35 Mb on chromosome 4, were frequently co-amplified. *MITF*

amplifications were present in 10% of samples, and in all three melanoma subtypes. Many acral and mucosal melanomas had SV and high-level amplifications on the long arm of chromosome 11, often targeting *CCND1* (Extended Data 3c).

We analysed the genomic subgroups of melanoma proposed by TCGA: *BRAF* mutated, *RAS* mutated, *NF1* mutated or triple wild-type (Extended Data 8 and 9, Fig. 3a)[2]. *NF1* aberrations occurred in 32/183 (17%) samples, and in 14 they were bi-allelic; 23 had point mutations/small indels only (15 nonsense, four missense, three splice site and one frame-shift). SV contributed heavily (Fig. 3b) to mutation burden in *NF1,* so previous studies in which SV were not assessed will have underestimated the *NF1*-mutated group and overestimated the proportion of triple wild-type melanomas. *NF1* mutation was found to co-occur with *RASA2* mutation (Fisher's exact test, $P < 0.003$), as recently described[28,31].

Most mucosal and acral (n = 22/43, 51%) melanomas, but only 15/140 (11%) cutaneous melanomas lacked *BRAF*, *NRAS* or *NF1* mutations; mutations in other cancer driver genes were frequent in this triple wild-type group. The latter included loss-of-function mutations in *CDKN2A, TP53* and *ARID2*, and activating hotspot mutations in *GNAQ* and *SF3B1.* Notably, one *GNAQ* and three *SF3B1* mutations, commonly mutated in uveal[32,33] but not cutaneous melanoma[2], were found exclusively in four of six triple wild-type mucosal melanomas, marking an unappreciated set of driver genes and pathways shared by mucosal and uveal melanomas. Seven triple wild-type melanomas carried mutations in *KIT*, five of which had high-level amplifications. Triple wild-type melanomas had more focal amplifications of *KIT*, *CCND1*, *MDM2* and *KRAS* than did non-triple wild-type melanomas (Fisher's exact test, $P < 5.4 \times 10^{-5}$, 1.8×10^{-5}, 0.018, 0.003, respectively).

Gene fusions

Gene fusions targeted several genes previously associated with melanoma (*RAF1*[34], *RAC1*[3], *MAP2K2*[35], *GRM3*[36], *MAP3K9*[29], *TRRAP*[37] and *PLCB4*[37]) (Supplementary Table 3) and known melanoma tumour suppressor genes: *NF1* (n = 3), *PARK2* (n = 2), *TP53*, *PTEN*, and *PPP6C*. *RAF1* gene fusions, shown to increase MEK phosphorylation compared to wild-type RAF1, were observed in one triple wild-type and one *NF1*-mutated tumour[38]. The same *RAF1* exons were retained in both fusions, albeit fused to different genes (*CDH3* and *GOLGA4*) (Extended Data 9). In RNAseq of 44 samples 17 of 136 potential fusion genes were expressed (Supplementary Table 3).

Fusions involved several kinase genes (Supplementary Table 3), the most frequently affected being *PAK1* (n = 5 samples: four acral melanomas, two of which were triple wild-

type, and one cutaneous carrying a *BRAF* V600E mutation). These *PAK1* fusions are unlikely to be activating: the kinase domain was retained in four of five samples, in each case there was a different gene fusion partner (*AQP11*, *FCHSD2*, *TENM4*, *OR9Q1*) and two to four additional breakpoints within the *PAK1* gene. The next most frequently fused kinase gene was *DGKB*, with two fusions in acral triple wild-type melanomas and one in a cutaneous *BRAF* V600E mutated sample, each with multiple breakpoints within the *DGKB* gene with likely loss of function.

Signalling activation and clinical implications

All subtypes were dominated by mutations to the MAPK, PI3K pathways and RTK (Supplementary Table 9, Fig. 4b). The MAPK signalling pathway was enriched 1.34 fold (Fisher's exact test, $q < 1 \text{ X } 10^{-51}$) and bore mutations in 179 of 183 (98 %) samples. Samples that harboured a mutated gene in both the MAPK and PI3K pathways had the highest proportion of p-ERK positive tumour nuclei, whereas samples that were wild-type for all genes in the PI3K and MAPK pathways had the lowest proportion (Fig. 4c, 4d). The highest p-AKT activation levels were seen in cases with only PI3K altered, whereas MAPK only and MAPK/PI3K wild-type cases had the lowest (differences not significant, one-way ANOVA and post hoc Tukey's test).

Nearly all cutaneous (138/140, 99 %) and most non-cutaneous (39/43, 91 %) melanomas harbored a SNV/indel or SV breakpoint in one or more genes that can confer sensitivity to an agent that is FDA-approved or currently in a cancer clinical trial (Extended data 10). The frequency of mutation and CNV of some potentially actionable genes (*CCND1*, *CCND2*, *KIT*, *KRAS*, *MDM2* and *PIP5K1A*) was significantly higher in non-cutaneous compared with cutaneous melanomas (two-proportion z-test, all $P < 0.05$).

Discussion

This is the largest WGS analysis of melanoma to date and the largest to compare cutaneous, acral and mucosal subtypes. Acral and mucosal melanomas showed a dramatically different genomic landscape from cutaneous melanoma, with a far lower mutation burden dominated by large-scale structural variants. Novel mutagen signatures in cutaneous melanomas likely reflect uncharacterised mechanisms of UVR damage. The principal mutation mechanisms driving mucosal and most acral melanomas were, in contrast, not attributable to UVR and imply novel carcinogenic exposures that are shared with cancers other than skin cancer. Notably, some mucosal melanomas exhibited SMG (*GNAQ*, *SF3B1*) previously characteristic of uveal melanoma. In further contrast to cutaneous melanomas, acral and mucosal melanomas were infrequently driven by *TP53*,

PTEN or *RB1* pathway lesions, and exhibited a variety of "triple wild-type" mechanisms such as *KIT* mutations and focal amplifications of *KIT*, *CCND1*, *MDM2* and *KRAS*. Novel recurrently fused genes were also identified (*PAK1*, *DGKB*). Most cutaneous and non-cutaneous melanomas exhibited mutations in genes for which targeted therapies exist or are being evaluated in clinical trials.

Non-coding mutations were surveyed for the first time at scale in melanoma, and their significance assessed taking into account the high mutation load at TFBS in melanoma[24]. The number of genes affected by recurrent non-coding mutations predicted to have a strong functional impact was equivalent to the number of genes affected by recurrent significant coding mutations. Eight genes in addition to *TERT* showed potential driver promoter mutations (*BLCAP*, *KBTBD8*, *NSUN6*, *RALY*, *RNF185*, *RPL29, RPS27* and *ZNF778*). Recurrent, likely functional 5'-UTR mutations were observed in *CHCHD2*, *PES1* and *RPS14*, as well as in genes previously reported as mutated (*RPS27*, *MRPS31*). Recurrent 3'-UTR mutations were common, and of uncertain significance, although *PTPRT* also carries frequent coding mutations[27].

Finally, we show the highly prevalent *TERT* promoter mutations in melanoma must confer selective advantage via dysregulation of telomere protection, whether via telomerase or the ALT mechanism, since they result in reduced, not increased telomere length.

References

1 Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421, doi:10.1038/nature12477 (2013).

2 Cancer Genome Atlas, N. Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681-1696, doi:10.1016/j.cell.2015.05.044 (2015).

3 Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nature genetics* **44**, 1006-1014, doi:10.1038/ng.2359 (2012).

4 Berger, M. F. *et al.* Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature* **485**, 502-506, doi:10.1038/nature11071 (2012).

5 Chi, Z. *et al.* Clinical presentation, histology, and prognoses of malignant melanoma in ethnic Chinese: a study of 522 consecutive cases. *BMC cancer* **11**, 85, doi:10.1186/1471-2407-11-85 (2011).

6 Furney, S. J. *et al.* Genome sequencing of mucosal melanomas reveals that they are driven by distinct mechanisms from cutaneous melanoma. *The Journal of pathology* **230**, 261-269, doi:10.1002/path.4204 (2013).

7 Furney, S. J. *et al.* The mutational burden of acral melanoma revealed by whole-genome sequencing and comparative analysis. *Pigment cell & melanoma research* **27**, 835-838, doi:10.1111/pcmr.12279 (2014).

8 Ravanat, J. L., Douki, T. & Cadet, J. Direct and indirect effects of UV radiation on DNA and its components. *Journal of photochemistry and photobiology. B, Biology* **63**, 88-102 (2001).

9 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 reports* **3**, 246-259, doi:10.1016/j.celrep.2012.12.008 (2013).
10 Alexandrov, L. B. & Stratton, M. R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Current opinion in genetics & development* **24**, 52-60, doi:10.1016/j.gde.2013.11.014 (2014).
11 Niu, B. *et al.* MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* **30**, 1015-1016, doi:10.1093/bioinformatics/btt755 (2014).
12 Mardin, B. R. *et al.* A cell-based model system links chromothripsis with hyperploidy. *Molecular systems biology* **11**, 828, doi:10.15252/msb.20156505 (2015).
13 Nones, K. *et al.* Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nature communications* **5**, 5224, doi:10.1038/ncomms6224 (2014).
14 Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959-961, doi:10.1126/science.1230062 (2013).
15 Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957-959, doi:10.1126/science.1229259 (2013).
16 Bell, R. J. *et al.* Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science* **348**, 1036-1039, doi:10.1126/science.aab0015 (2015).
17 Liau, J. Y. *et al.* TERT promoter mutation is uncommon in acral lentiginous melanoma. *Journal of cutaneous pathology* **41**, 504-508, doi:10.1111/cup.12323 (2014).
18 Fredriksson, N. J., Ny, L., Nilsson, J. A. & Larsson, E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nature genetics* **46**, 1258-1263, doi:10.1038/ng.3141 (2014).
19 Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A. & Reddel, R. R. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nature medicine* **3**, 1271-1274 (1997).
20 Ramamoorthy, M. & Smith, S. Loss of ATRX Suppresses Resolution of Telomere Cohesion to Control Recombination in ALT Cancer Cells. *Cancer cell* **28**, 357-369, doi:10.1016/j.ccell.2015.08.003 (2015).
21 Wiestler, B. *et al.* ATRX loss refines the classification of anaplastic gliomas and identifies a subgroup of IDH mutant astrocytic tumors with better prognosis. *Acta neuropathologica* **126**, 443-451, doi:10.1007/s00401-013-1156-z (2013).
22 Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264-267, doi: 10.1038/nature17661 (2016).
23 Shain, A. H. *et al.* Exome sequencing of desmoplastic melanoma identifies recurrent NFKBIE promoter mutations and diverse activating mutations in the MAPK pathway. *Nature genetics* **47**, 1194-1199, doi:10.1038/ng.3382 (2015).
24 Dutton-Regester, K. *et al.* A highly recurrent RPS27 5'UTR mutation in melanoma. *Oncotarget* **5**, 2912-2917 (2014).
25 Ding, L. *et al.* Clonal architectures and driver mutations in metastatic melanomas. *PloS one* **9**, e111153, doi:10.1371/journal.pone.0111153 (2014).
26 Andersen, L. B. *et al.* Mutations in the neurofibromatosis 1 gene in sporadic malignant melanoma cell lines. *Nature genetics* **3**, 118-121, doi:10.1038/ng0293-118 (1993).
27 Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251-263, doi:10.1016/j.cell.2012.06.024 (2012).
28 Krauthammer, M. *et al.* Exome sequencing identifies recurrent mutations in NF1 and RASopathy genes in sun-exposed melanomas. *Nature genetics* **47**, 996-1002, doi:10.1038/ng.3361 (2015).

29 Stark, M. S. *et al.* Frequent somatic mutations in MAP3K5 and MAP3K9 in metastatic melanoma identified by exome sequencing. *Nature genetics* **44**, 165-169, doi:10.1038/ng.1041 (2012).
30 Curtin, J. A., Busam, K., Pinkel, D. & Bastian, B. C. Somatic activation of KIT in distinct subtypes of melanoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24**, 4340-4346, doi:10.1200/JCO.2006.06.2984 (2006).
31 Arafeh, R. *et al.* Recurrent inactivating RASA2 mutations in melanoma. *Nature genetics*, doi:10.1038/ng.3427 (2015).
32 Harbour, J. W. *et al.* Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nature genetics* **45**, 133-135, doi:10.1038/ng.2523 (2013).
33 Van Raamsdonk, C. D. *et al.* Frequent somatic mutations of GNAQ in uveal melanoma and blue naevi. *Nature* **457**, 599-602, doi:10.1038/nature07586 (2009).
34 Palanisamy, N. *et al.* Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nature medicine* **16**, 793-798, doi:10.1038/nm.2166 (2010).
35 Nikolaev, S. I. *et al.* Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nature genetics* **44**, 133-139, doi:10.1038/ng.1026 (2012).
36 Prickett, T. D. *et al.* Exon capture analysis of G protein-coupled receptors identifies activating mutations in GRM3 in melanoma. *Nature genetics* **43**, 1119-1126, doi:10.1038/ng.950 (2011).
37 Wei, X. *et al.* Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nature genetics* **43**, 442-446, doi:10.1038/ng.810 (2011).
38 Jones, D. T. *et al.* Oncogenic RAF1 rearrangement and a novel BRAF mutation as alternatives to KIAA1549:BRAF fusion in activating the MAPK pathway in pilocytic astrocytoma. *Oncogene* **28**, 2119-2123, doi:10.1038/onc.2009.73 (2009).

Supplementary Information

Supplementary Table 1 | Clinical and mutation Data

Supplementary Table 2 | Coding mutations (SNV and indel) in melanoma (MAF)

Supplementary Table 3 | Structural rearrangements in melanoma

Supplementary Table 4 | Gene promoters frequently mutated in melanoma

Supplementary Table 5 | Recurrent 5' UTR mutations in melanoma

Supplementary Table 6 | Recurrent 3' UTR mutations in melanoma

Supplementary Table 7 | Significantly mutated genes

Supplementary Table 8 | Key to Fig 3b: Significantly mutated genes and selected published melanoma driver genes

Supplementary Table 9 | Perturbed pathways in melanoma

Acknowledgments

This work was supported by Melanoma Institute Australia (MIA), Bioplatforms Australia, NSW Ministry of Health, Cancer Council NSW, National Health and Medical Research Council of Australia (NHMRC), Cancer Institute NSW, Australian Cancer Research Foundation and the National Collaborative Research Infrastructure Strategy. NKH, NW, RAS, JSW and KD-R were supported by NHMRC Fellowships, KN by a Keith Boden Fellowship, GVL by University of Sydney Medical Foundation, MS by Pfizer Australia, the Victorian Endowment for Knowledge, Science and Innovation and NHMRC, LBA by a J. Robert Oppenheimer Fellowship at Los Alamos National Laboratory, and NLB by the European Research Council (Consolidator Grant 682398). Biobanking was supported by MIA, Victorian Cancer Agency, Victorian Cancer Biobank, Victorian State Government Operational Infrastructure Support Program, Melanoma Research Alliance and the Melbourne Melanoma Project, and the efforts of patients, clinicians and other staff at health services across Australia. Cell lines were provided via the ABN-Oncology group, supported by NHMRC. Research at Los Alamos National Laboratory was under the auspices of the National Nuclear Security Administration of the US Department of Energy; the Los Alamos National Laboratory Institutional Computing Program was supported by contract DE-AC52-06NA25396. The authors gratefully acknowledge the support of colleagues at MIA, Royal Prince Alfred Hospital, NSW Health Pathology, Westmead Institute for Medical Research, Peter MacCallum Cancer Centre and Olivia Newton-John Cancer Research Institute.

Author Contributions

PAJ, MAF, KN, AMP, LBA, AP, SK, OH, CL, SW, QX, NMW, KDR and FN analysed genomic data; VJ, PS, HK, RdP, VT, GMP and HB collected, prepared and analysed samples and data; LMSL, RAD and HP validated telomere length; LM, RS and NLB analysed selection on coding and non-coding mutations; AF, CAS, JYY and SJS supported design and planning; JVP, NW and SMG developed and directed the analysis pipeline; JFT, MS, AB, JC, JRS, RFK, PH, GVL, AJS, RPMS and REV collected samples and data; NKH, JSW, PAJ, NW, RAS and GJM designed and directed the study, analysed data and wrote the manuscript, which all authors reviewed.

Author Statement

The authors declare no competing financial interests. BAM files are available at EGA (accession: EGAS00001001552). Direct correspondence and requests for materials to GJM.

FIGURES

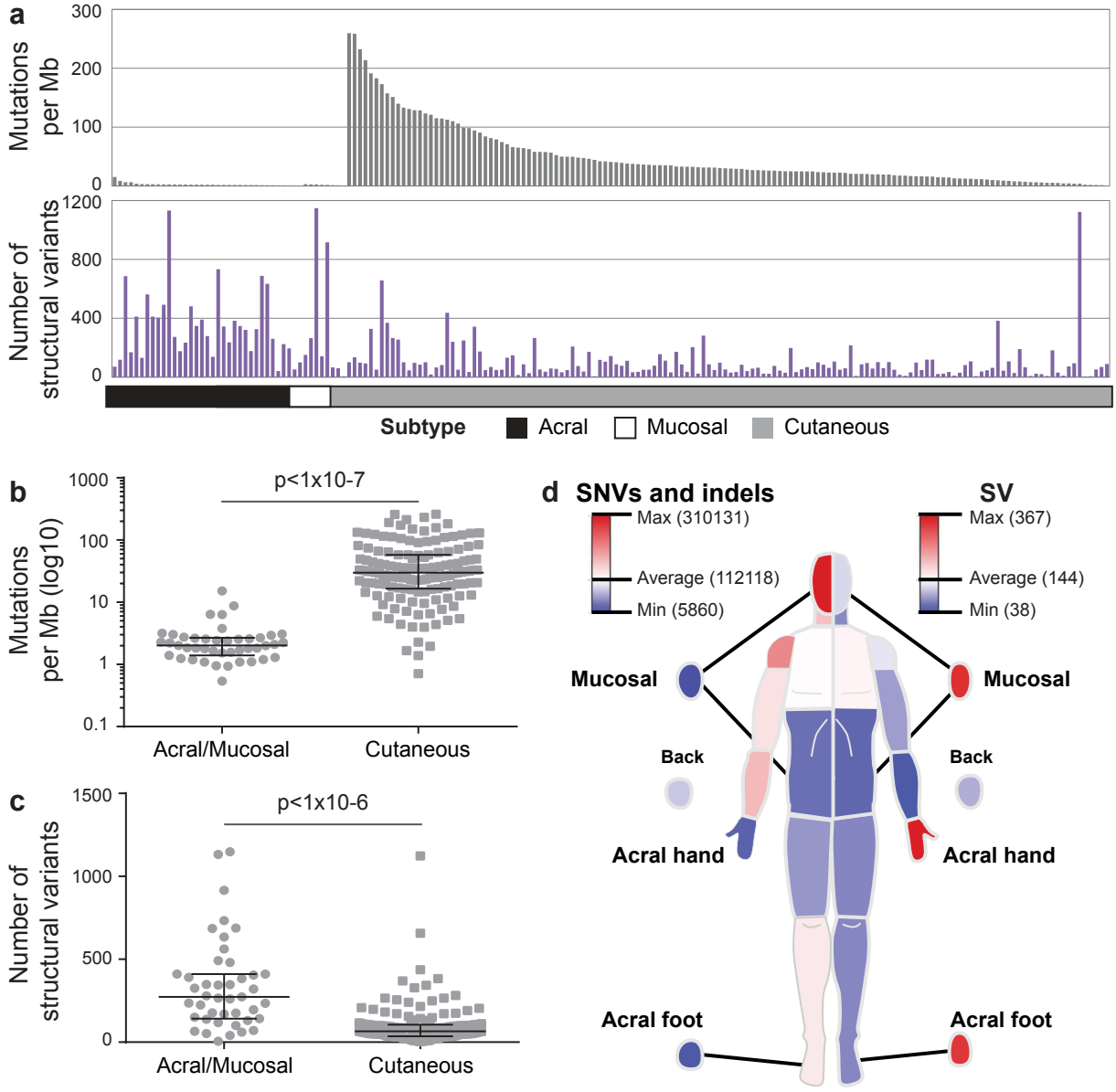


Figure 1 | Mutation burden. a, Rate of substitution/indel (SNV, upper panel) and structural (SV, lower panel) mutations. **b,** Differences in SNV frequency by subtype (t-test $P < 1 \times 10^{-7}$): acral and mucosal melanomas carried >18-fold fewer mutations than cutaneous melanomas. **c,** Differences in SV frequency by subtype (t-test $P < 1 \times 10^{-6}$): more SV in acral and mucosal than cutaneous melanomas. **d,** Relative abundance of aberrations by body site of the antecedent primary tumour. Highest (red) average SNV counts occurred in sun-exposed areas; highest SV counts occurred in sun-shielded body sites of mucosal and acral origin.

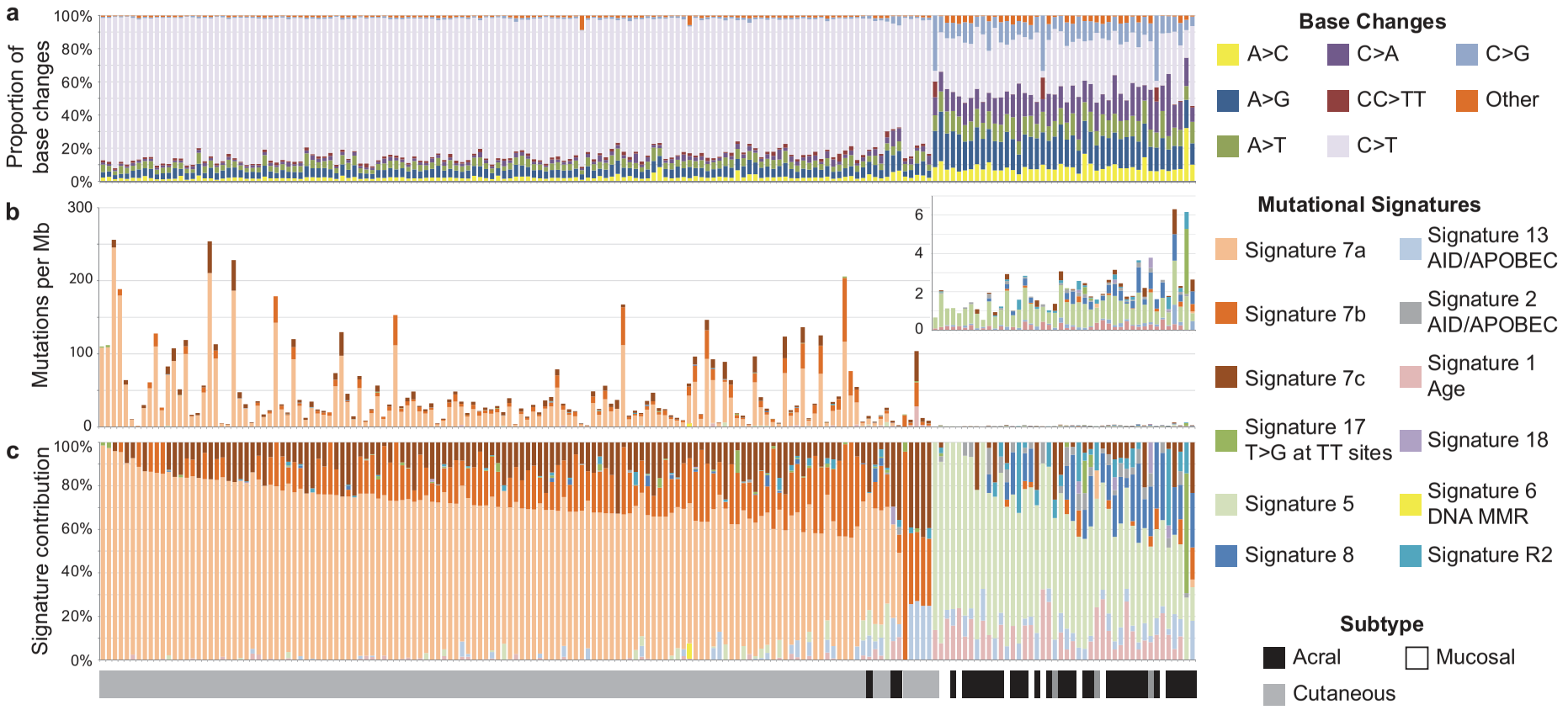

a
Proportion of base changes
100%
80%
60%
40%
20%
0%
b
Mutations per Mb
300
200
100
0
6
4
2
0
c
Signature contribution
100%
80%
60%
40%
20%
0%
Base Changes
A>C
C>A
C>G
A>G
CC>TT
Other
A>T
C>T
Mutational Signatures
Signature 7a
Signature 13 AID/APOBEC
Signature 7b
Signature 2 AID/APOBEC
Signature 7c
Signature 1 Age
Signature 17 T>G at TT sites
Signature 18
Signature 5
Signature 6 DNA MMR
Signature 8
Signature R2
Subtype
Acral
Mucosal
Cutaneous

Figure 2 | Mutational processes. a, Proportion of somatic base changes: melanomas segregate into two groups with nearly all cutaneous melanomas showing a high proportion of C>T transitions. Twelve mutational signatures shown as **b**, mutations/Mb and **c**, percentage. Cutaneous melanomas were dominated by three novel UVR signatures, whereas melanomas with low mutation burden were dominated by signature 1 (associated with age), signature 5 (aetiology unknown and ubiquitous in diverse cancers), and signature 8 (seen in medulloblastoma and breast cancer).

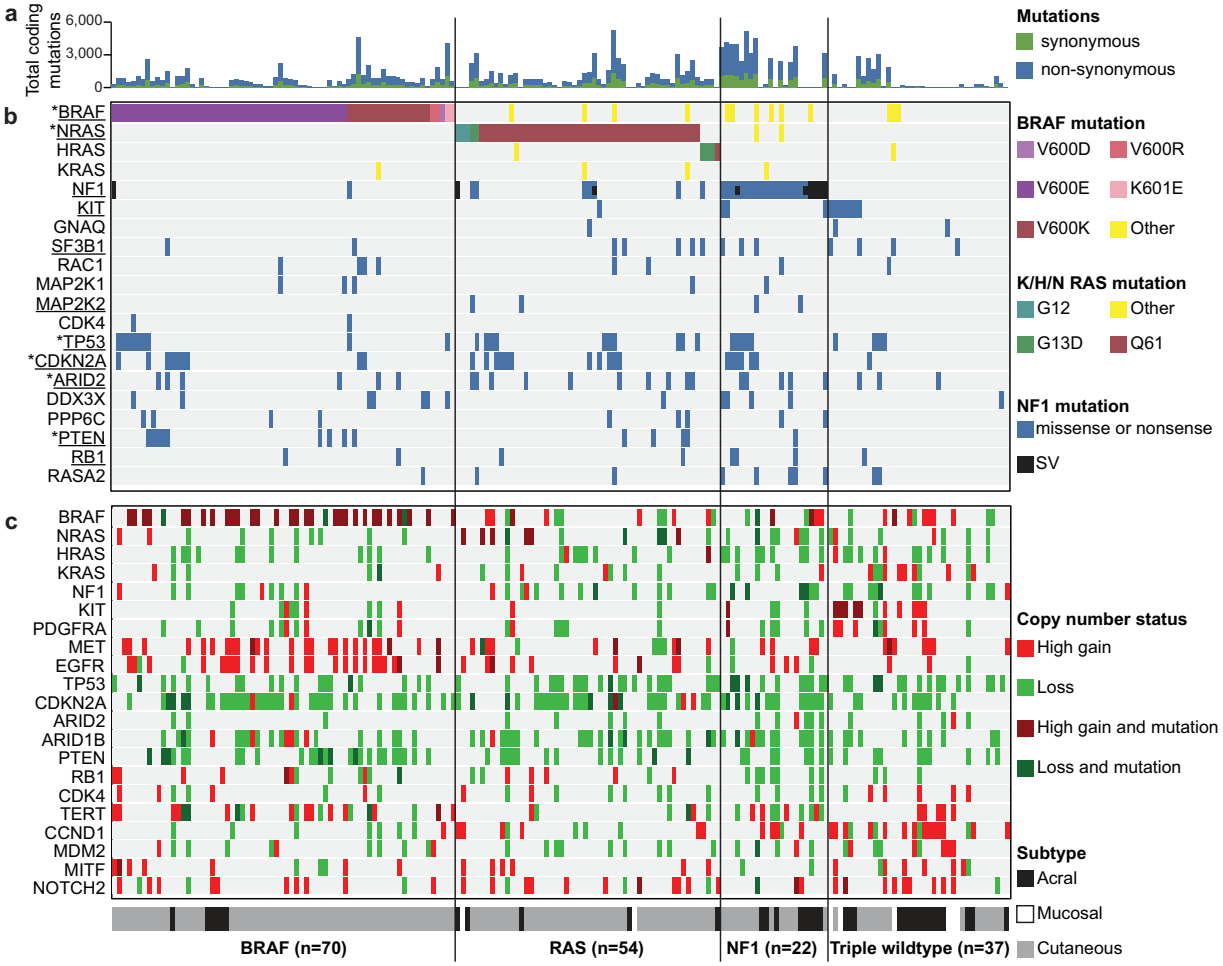


Figure 3 | Mutations and copy number changes in selected published melanoma driver genes. a, Somatic coding mutation rate. **b,** Mutations in significantly mutated genes (underlined) and selected published melanoma driver genes (Supplementary Table 8); *BRAF* and *RAS* aberrations coloured by mutation type; genes marked (*) differentially mutated between subtypes (Fisher's exact test, $P < 0.05$; missense, nonsense or SV, or hotspot for *BRAF* and *RAS*), each in cutaneous more than in acral/mucosal melanoma. **c,** Copy number changes in selected melanoma-associated genes: loss (green), high gain (≥6 copies, red). Melanoma subtype shown below.

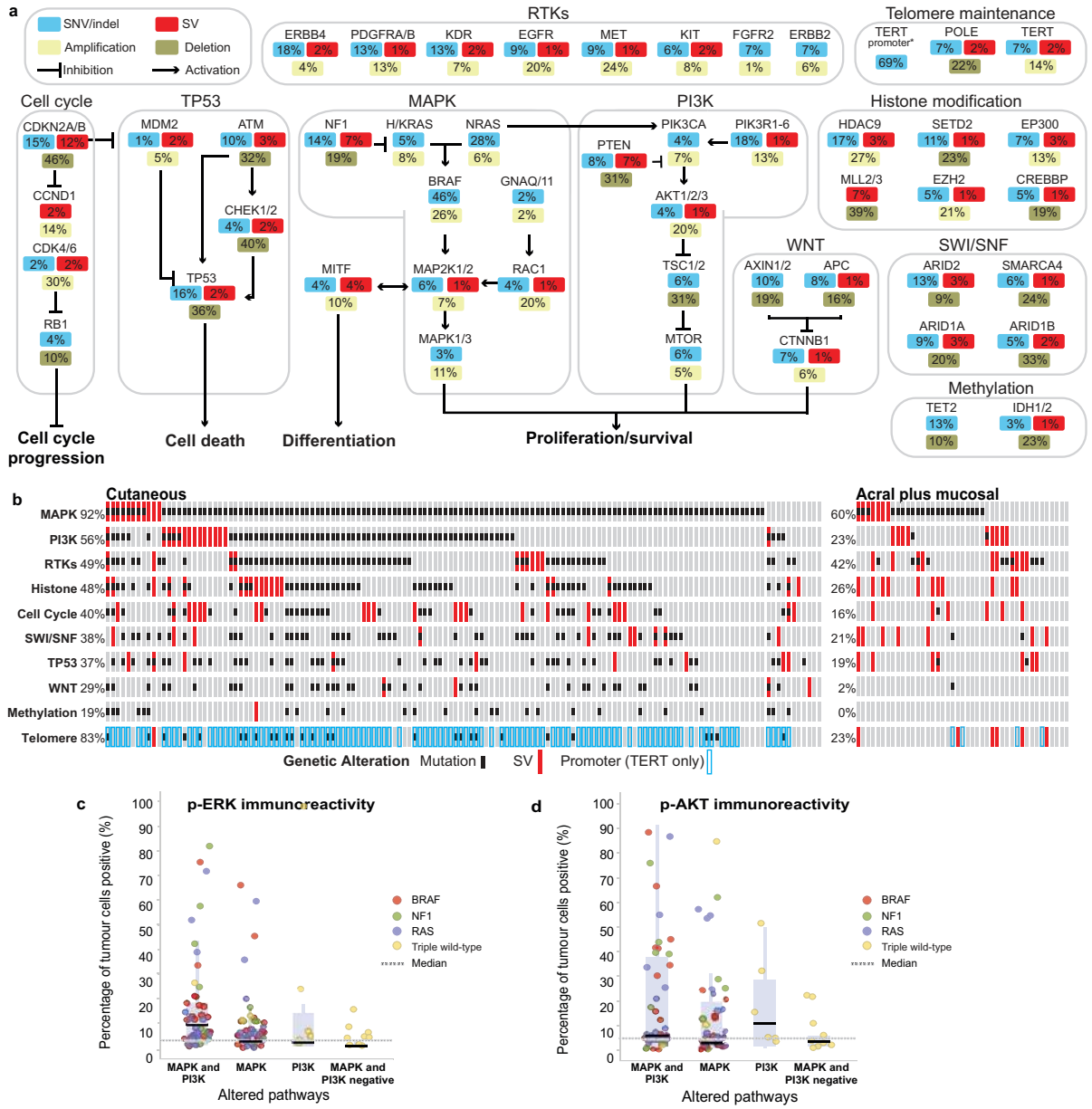


Figure 4 | Genes and signaling pathways recurrently altered in melanoma.

a, Percentage of samples with protein-affecting aberrations in candidate driver genes, grouped by pathway: substitution/indels (blue), structural variants (red), copy number amplification (copy number >5, yellow), homozygous deletion (green). **b,** Frequency of aberrations in pathways as percentage of cutaneous (n = 140) or non-cutaneous melanomas (n = 43). **c** and **d**, Percentage of tumour nuclei immunoreactive to p-ERK or p-AKT by multiplex-immunofluorescent staining; cases grouped according to mutated genes from the MAPK and PI3K pathways.

Methods

Human melanoma samples

The fresh-frozen tissue and blood samples analysed in the current study were obtained from Australian melanoma biospecimen banks, including the Melanoma Institute Australia (MIA) (n=160), Australasian Biospecimen Network-Oncology Cell Line Bank QIMR-Berghofer Institute of Medical Research (QIMR) (n=15; all lines authenticated by DNA profiling and tested for Mycoplasma contamination), Ludwig Institute for Cancer Research (n=4), Peter MacCallum Cancer Centre/Victorian (n=4) biospecimen banks. All tissues and bloods form part of prospective collection of fresh-frozen samples accrued with written informed patient consent and institutional review board approval. Fresh surgical specimens were macro-dissected and tumour tissues were procured (with as little contaminating normal tissue as possible) and snap frozen in liquid nitrogen within 1 hour of surgery. All samples were pathologically assessed prior to inclusion into the study, with samples requiring greater than 80 % tumour content and less than 30 % necrosis to be included. All samples were independently reviewed by expert melanoma pathologists (RAS, REV) to confirm the presence of melanoma and qualification of the above criteria. Samples requiring tumour enrichment underwent macrodissection or frozen tissue coring (Cryoxtract, Woburn MA USA) using a marked H&E slide as a reference.

The histopathology of all mucosal and acral samples was reviewed by RAS to confirm the diagnosis. Acral melanomas were classified as occurring within acral skin of the palm of the hand, sole of the foot and under nail beds. The lack of hair follicles, thickened stratum corneum and clinical site was confirmed in all cases. Mucosal melanomas were defined as occurring in the mucosal membranes lining oral, respiratory, gastrointestinal and urogenital tracts. The H&E slides of the primary melanomas were reviewed for all mucosal and acral samples and any tumour that arose in the junction of the acral/mucosal and cutaneous skin was excluded. Occult/unknown primary melanomas were considered cutaneous, since their genome landscape is indistinguishable from that of melanomas arising in the skin[39].

DNA extractions

Tumour DNA was extracted using DNeasy Blood and Tissue Kits (69506, Qiagen Ltd), according to the manufacturer's instructions. Blood DNA was extracted from whole blood using Flexigene DNA Kits (51206, Qiagen Ltd). All samples were quantified using the NanoDrop (ND1000, Thermoscientific) and Qubit® dsDNA HS Assay (Q32851, Lifetechnologies) and DNA size and quality were tested using gel electrophoresis. Samples with a concentration of less than 50 ng/µl, or absence of a high molecular weight band in electrophoresis gels, were excluded from further analyses.

Whole genome sequencing (WGS)

WGS was performed on Illumina Hiseq 2000 instruments (Illumina, San Diego, CA, USA) at three Australian sequencing facilities (Australian Genomic Research Facility, Ramaciotti Centre for Genomics, John Curtin School of Medical Research) and Macrogen (Geumcheon-gu, Seoul, South Korea). All facilities performed library construction using TruSeq DNA Sample Preparation kits (Illumina) as per Illumina instructions. The subsequent 100 bp pair-end libraries were sequenced using Truseq SBS V3-HS kits to average depth 85x (range 43-219x) in the tumour sample and 64x (range 30-214x) in the matched normal.

Whole genome sequence processing and quality control

Sequence data was aligned to the GRCh37 assembly using multi-threaded BWA 0.6.2-mt resulting in sorted lane level files in sequence alignment/mapping (SAM) format which were compressed and converted to binary file (BAM) created by samtools 0.1.19. Sample-level merged BAMs, one each for matched germline and tumour samples were produced by in-house tools and duplicate reads marked with Picard MarkDuplicates 1.97 (http://picard.sourceforge.net). Quality assessment and coverage estimation was carried out by qProfiler and qCoverage (http://sourceforge.net/projects/adamajava). To test for the presence of sample or data swaps all sequence data were assessed for concordance at approximately 1.4 million polymorphic genomic positions including the genotyping array data by qSignature.

Somatic mutation analysis

Somatic mutations and germline variants were detected using a dual calling strategy using qSNP[40] and GATK[41] and indels of 1–50 bp in length were called with Pindel[42] and GATK. All mutations were submitted to the International Cancer Genome Consortium (ICGC)[43] Data Coordination Centre (DCC). Mutations were annotated with gene consequence using Ensembl gene annotation with SnpEff. Somatic genes which were significantly mutated were identified using two approaches: MutSigCV and OncodriveFML 1.1[22] using a threshold of $q<0.1$. Significant non-coding elements were detected using OncodriveFML 1.1[22].

Somatic copy number and ploidy were determined using the TITAN tool[44]. SV were identified using the qSV tool and chromosomes containing highly significant non-random distributions of breakpoints were identified as previously described[45]. Chromosomes identified to have clustering of breakpoints were inspected against criteria for chromothripsis[46] and the breakage-fusion-bridge cumulative rearrangement model[47]. Chromosomes with high

numbers of translocations were identified with a minimum threshold of 10 translocation breakpoints per chromosome following manual review.

Mutational signatures were predicted in each sample using a published framework[1]. Essentially the substitution mutations across the whole genome in all cases were analysed in context of the flanking nucleotides (96 possible trinucleotide combinations). Identified signatures were compared to other validated signatures and the frequency of each signature per Mb was determined.

Recurrent non-coding mutations

Statistical significance of recurrent non-coding mutations was estimated using a permutation test. A null distribution of recurrence was estimated by randomly shuffling all mutations within each sample and record number of recurrent mutations within the regions of interest. To take into account not only the varying mutation burden but also the different mutation signatures, we restricted the random shuffling such that the mutation in the middle of a trinucleotide, ABC, was only shuffled to the same trinucleotide.

Functional impact bias of mutations in promoters, 3' UTRs and 5' UTRS

Promoters and UTRs that are likely to play a role in tumorigenesis were identified with OncodriveFML[22]: a framework able to detect signals of positive selection in both the coding and the non-coding regions of the genome by measuring the bias toward the accumulation of functional mutations. The functional impact of mutations in gene promoters was assessed using the CADD (Combined Annotation Dependent Depletion)[48], TFBS creation, and TFBS disruption scores. The CADD scores measure the deleteriousness of mutations, and are calculated by integrating multiple annotations into a single metric by contrasting variants that survived natural selection with simulated mutations. The scores for TFBS creation (motif gain) and disruption (motif break) were computed by following the steps described by Fu et al.[49]. The score value indicates the difference between position weight matrix matching scores of the germline and mutant alleles. 5' UTRs were analyzed using the TFBS disruption scores while 3' UTRs were analyzed using the CADD scores. The statistical significance of promoters and UTRs was derived by comparing the average functional impact score of the mutations in the element with the functional impact scores obtained by permuting 100,000-fold the observed mutations, maintaining their trinucleotide context. In addition, since the rate of somatic mutations in melanoma is highly increased at active transcription factor binding sites (TFBS)[24], OncodriveFML was adapted (version 1.1) to perform a strictly local permutation in windows of 51 bp (25 nucleotides at each side of the mutation). This variation in the background model of OncodriveFML allowed us to better estimate whether the

mutations observed in tumours disrupted or created TFBS more than expected by chance. The statistical significance of promoters and UTRs mutated in at least two samples was adjusted with the Benjamini-Hochberg correction for multiple testing.

We also used miRanda v3.3a to predict whether the recurrent 3'UTR mutations alter (disrupt or create) miRNA binding sites. The 50 base sequence surrounding each 3'UTR was used as input to miRanda. miRNAs that were predicted to hit either the wild-type or mutant sequence (but not both) were considered potential targets and further filtered as follows. We required a hit to perfectly align against the seed region of the miRNA (nt 2-8), that the mutation lay within the seed, and that the predicted binding energy was higher (lower dG) in the non-hit than in the hit.

Telomere length estimation

To estimate telomere length we counted telomere motifs in the whole gene data using the qPCR-validated qmotif tool (https://sourceforge.net/p/adamajava/wiki/qMotif) implemented in JAVA using the Picard library (version 1.110). qmotif is driven by a single plain-text configuration file in the "Windows INI-file" style and the input is a WGS BAM file that has been duplicate-marked and coordinate-sorted. Essentially, qmotif uses a two-stage matching system where the first stage is a quick-but-strict string match and the second stage is a slower but more flexible regular expression match; only reads that pass stage 1 go on to the much slower match in stage 2. For telomere quantification, a string that represents 3 concurrent repeats of the canonical telomere motif (TTAGGGTTAGGGTTAGGG) was used as the stage 1 match and a simple pattern match for stage 2 which captures any read with 2 or more concurrent occurrences of the telomeric repeat with variation allowed in the first 3 bases. The relative tumour telomere length was then estimated by comparing the tumour read counts to the matched normal sample. Telomere length was validated by qPCR[50].

Sanger sequencing of TERT promoter

Direct PCR amplification and Sanger sequencing was performed using primers: hTERT_F ACGAACGTGGCCAGCGGCAG and hTERT_R CTGGCGTCCCTGCACCCTGG which amplify a 474 bp region of the *TERT* promoter[51]. PCR was carried out in a 13 µl volume containing 1 µl of 20 ng/µl gDNA, 1.25 µl 10X $MgCL_2$, 2.5 µl Betaine, 1.25 µl dNTPs (2.5 mM), 1 µl of 10 µM primers and 0.25 µl of PFU Turbo (#600250, Agilent). PCR reactions were performed under the following conditions: 95 °C for 5 mins, followed by four cycles of 95 °C for 30 sec, annealing at 66 °C for 1 min and polymerization at 72 °C for 1 min. This was followed by 4 more cycles with a lowered annealing temperature of 64 °C for 1 min, followed by 28 cycles with annealing temperatures of 62 °C. Subsequent PCR products

were sequenced on a (AbiPrism 3130xl Genetic Analyzer, Applied Biosystems) and data analysed using Sequence Scanner Software 2 (Applied Biosystems) with reference to the sequences from the NCBI gene database, *TERT* (chr5: 1,295,071 - 1,295,521).

Ultra-deep custom amplicon sequencing of non-coding mutations

The Illumina TruSeq Custom Amplicon V1.5 was used to validate 20 recurrent non-coding point mutations in the promoter (n=11), 3'-UTR (n=3) and 5'-UTR (n=6) regions of genes with frequent non-coding mutations in 164 of the 183 samples. The Illumina DesignStudio (Illumina, Inc., San Diego, CA, USA) was used to design 250 bp sequences of the target regions. Sequencing libraries were prepared using the TruSeq Custom Amplicon Library Preparation Guide and the TruSeq Custom Amplicon Index Kit and sequenced on a MiSeq Illumina sequencer V2 (Illumina, Inc.). Sequences were aligned to the reference genome (GRCh37/hg19) using BWA 0.6.2-mt. A pileup approach was used to determine the base count at each position of interest. A mutation was considered verified if the mutant allele frequency was >10 %.

Exome sequencing

Exome capture was performed on 1 µg of DNA extracted from tumour and normal blood using the Illumina TruSeq Exome Library Prep Kit. Libraries were sequenced (2 x 100 bp paired-end sequencing) on the Illumina HiSeq2000 platform with a minimum coverage of 75X. Exome sequence data was produced for 53 patients in the cohort and used to validate coding mutations detected by whole-genome sequencing.

RNA extraction and sequencing

Total RNA was extracted from fresh frozen tissue using the mirVana™ miRNA Isolation Kit (Applied Biosystems, Cat# AM1560). RNA quality and presence of a small RNA fraction were measured using the Agilent 2100 RNA 6000 Nano and small RNA kits. RNA-sequencing was performed using 1 µg of total RNA, which was converted into mRNA libraries using the Illumina mRNA TruSeq kit. RNA sequencing was performed using 2x75 bp paired end reads on the Illumina Hiseq2000. Small RNA-Sequencing was performed using 1 µg of total RNA which was converted into a small RNA libraries, size selection range 145-160 bp (RNA of 18-33 nt) using Illumina's TruSeq Small RNA Library Preparation Kit and sequenced on an Illumina Hiseq2000 using 50 bp single read sequencing with 1 % control spiked in.

RNA sequence reads were aligned to transcripts corresponding to ensemble 70 annotations using RSEM[52]. RSEM data were normalized using TMM (weighted trimmed mean of M-

values) as implemented in the R package 'edgeR'. For downstream analyses, normalized RSEM data were converted to counts per million (c.p.m.). Genes with at least 5 c.p.m. in at least 2 samples were considered expressed.

Body site distribution of SNV/indel and SV

Total numbers of SNV/Indel and SV were compared according to primary melanoma body site, categorized into abdomen, acral hand, acral foot, back, lower arm, lower leg, mucosal, neck, shoulder, thorax, upper arm, upper leg and face and scalp. Any samples with an unknown primary site or occult classification were excluded from analysis. Heat maps were produced in Spotfire-Tibco (version 6.0, spotfire.tibco.com) based on the combined total number of SNV and indels, or by SV. A two-colour heat map (red high, blue low) was produced and colours were overlaid using Illustrator CS6 (Adobe) onto an adapted SVG human body diagram that was created using Illustrator CS6 (Adobe). We thank Doug Stetner for computing assistance.

Clinically actionable genes

The frequency of clinically actionable mutations was assessed by annotating genomic variants using the IntOGen Cancer Drivers Actionability (ICDA) database (2014), which identifies mutations that may confer sensitivity to therapeutic agents[53]. The ICDA database was used to assign an activating or loss of function role to mutated genes" . Loss of heterozygosity, silent mutations, deletions to activating genes or amplifications to loss of function genes were not included in the analysis. Additionally, visual inspection using the Integrative Genome Viewer (IGV, Broad) was used to identify only high confidence structural rearrangement breakpoints with clustered supporting reads with both discordant read pair and soft clipping evidence. SVs with a high incidence of random non-clustered background signal surrounding the breakpoints along with low numbers of supporting non-duplicate reads were excluded from analysis for this figure (Extended Data 10). The proportion of tumours with a mutation to a particular actionable gene was calculated and classified based on mutation type into: 1) SNV/indel, 2) SNV/indel and SV, 3) SV or 4) copy number variation only.

Commonly mutated genes and pathways

A hand-curated list of commonly mutated tumour suppressor genes and oncogenes was created and analysed for the frequency of mutation (Fig. 4a). Mutations were defined as SNV/indel, SV, copy number amplifications and copy number deletions. Loss of heterozygosity, silent mutations, RNA mutations, deletions to oncogenes or amplifications to tumour suppressor genes were not included in the figure. SV breakpoints were subjected to

manual inspection using the Integrative Genome Viewer (IGV, Broad) and only events confirmed as somatic and predicted to alter transcription processing were considered further. We then overlaid the alterations from Fig. 4a onto pathways defined by KEGG and GO gene sets from MSigDB v5.0. A pathway was considered altered in a given sample if at least one gene in the pathway contained an SNV/indel or SV. The pathways were then stratified according to cutaneous or non-cutaneous subtypes. A mutation file with samples IDs and their mutated pathways was entered for analysis into the OncoPrinter tool (http://cbioportal.org).

Multiplex immunofluorescent staining

MAPK and PI3K pathway status was also assessed by multiplex immunofluorescent staining for phosphorylated ERK and AKT (106/183). All immunohistochemical staining was performed on a Dako Autostainer Plus (Dako, Glostrup, Denmark) using the Dako Envision Flex detection kit (K8000, Dako) and OPAL 7-color IHC Kit for visualisation (NEL797B001KT, PerkinElmer). Consecutive staining rounds included p-AKT (1:100, NCL-L-Akt-Phos, Leica), p-ERK (1:1600, CS4370, Cell Signalling) and SOX10 (1:800, ACI 3099A, Biocare). Multispectral quantitative image analysis was performed on a Vectra 3 slide scanner (PerkinElmer, USA). The captured multispectral images were analysed using the quantitative InForm image analysis software (PerkinElmer, USA).

Data deposition

All somatic variants for this study have been submitted to the ICGC Data Coordination Centre (DCC) and are publicly available through the ICGC DCC Data Portal (https://dcc.icgc.org). The BAM files are in the EGA (accession EGAS00001001552).

Code availability

Tools used in this publication that were developed in-house are available from the SourceForge public code repository under the AdamaJava project (http://sourceforge.net/projects/adamajava/).

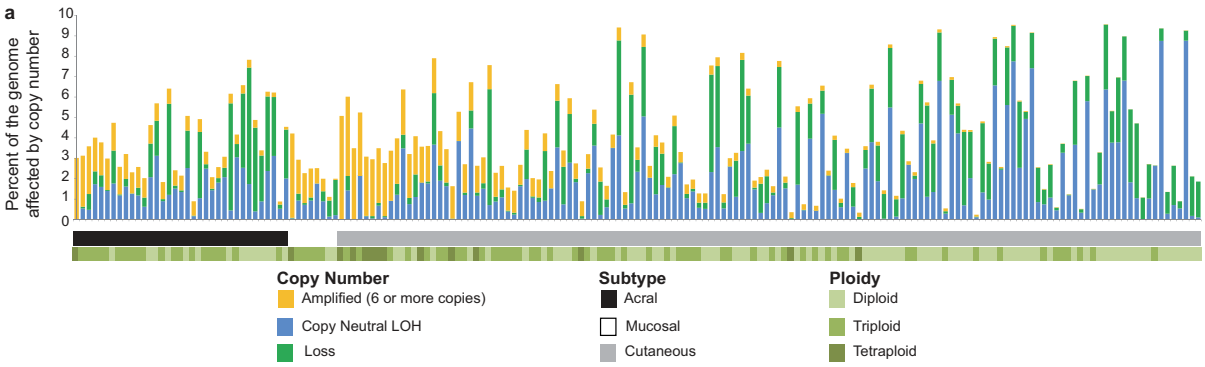
39 Dutton-Regester, K. *et al.* Melanomas of unknown primary have a mutation profile consistent with cutaneous sun-exposed melanoma. *Pigment cell & melanoma research* **26**, 852-860, doi:10.1111/pcmr.12153 (2013).

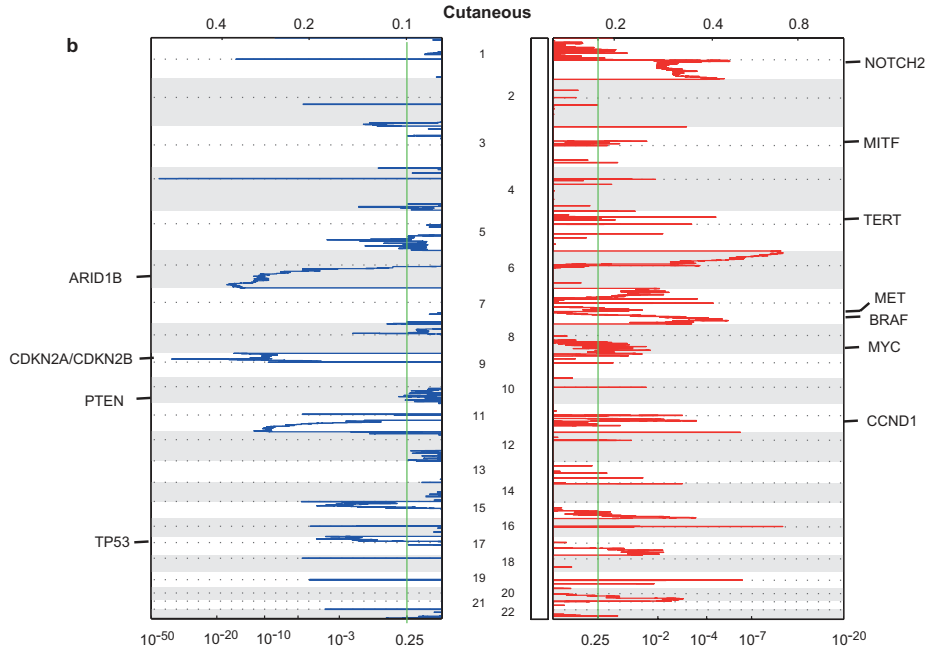
40 Kassahn, K. S. *et al.* Somatic point mutation calling in low cellularity tumors. *PloS one* **8**, e74380, doi:10.1371/journal.pone.0074380 (2013).

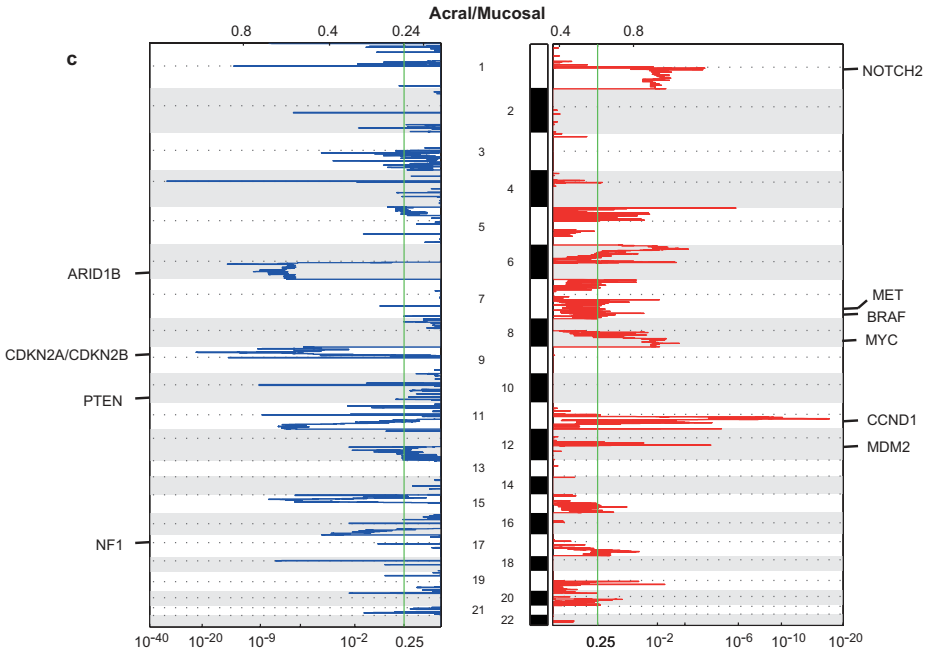
41 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

42 Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865-2871, doi:10.1093/bioinformatics/btp394 (2009).
43 Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464**, 993-998 (2010).
44 Ha, G. *et al.* TITAN: inference of copy number architectures in clonal cell populations from tumor whole-genome sequence data. *Genome research* **24**, 1881-1893, doi:10.1101/gr.180281.114 (2014).
45 Patch, A. M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489-494, doi:10.1038/nature14410 (2015).
46 Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226-1236, doi:10.1016/j.cell.2013.02.023 (2013).
47 Kinsella, M. & Bafna, V. Combinatorics of the breakage-fusion-bridge mechanism. *Journal of computational biology : a journal of computational molecular cell biology* **19**, 662-678, doi:10.1089/cmb.2012.0020 (2012).
48 Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* **46**, 310-315, doi:10.1038/ng.2892 (2014).
49 Fu, Y. *et al.* FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol* **15**, 480, doi:10.1186/s13059-014-0480-5 (2014).
50 Lau, L. M. *et al.* Detection of alternative lengthening of telomeres by telomere quantitative PCR. *Nucleic acids research* **41**, e34, doi:10.1093/nar/gks781 (2013).
51 Griewank, K. G. *et al.* TERT promoter mutations are frequent in atypical fibroxanthomas and pleomorphic dermal sarcomas. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **27**, 502-508, doi:10.1038/modpathol.2013.168 (2014).
52 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).
53 Rubio-Perez, C. *et al.* In Silico Prescription of Anticancer Drugs to Cohorts of 28 Tumor Types Reveals Targeting Opportunities. *Cancer cell* **27**, 382-396, doi:10.1016/j.ccell.2015.02.007 (2015).

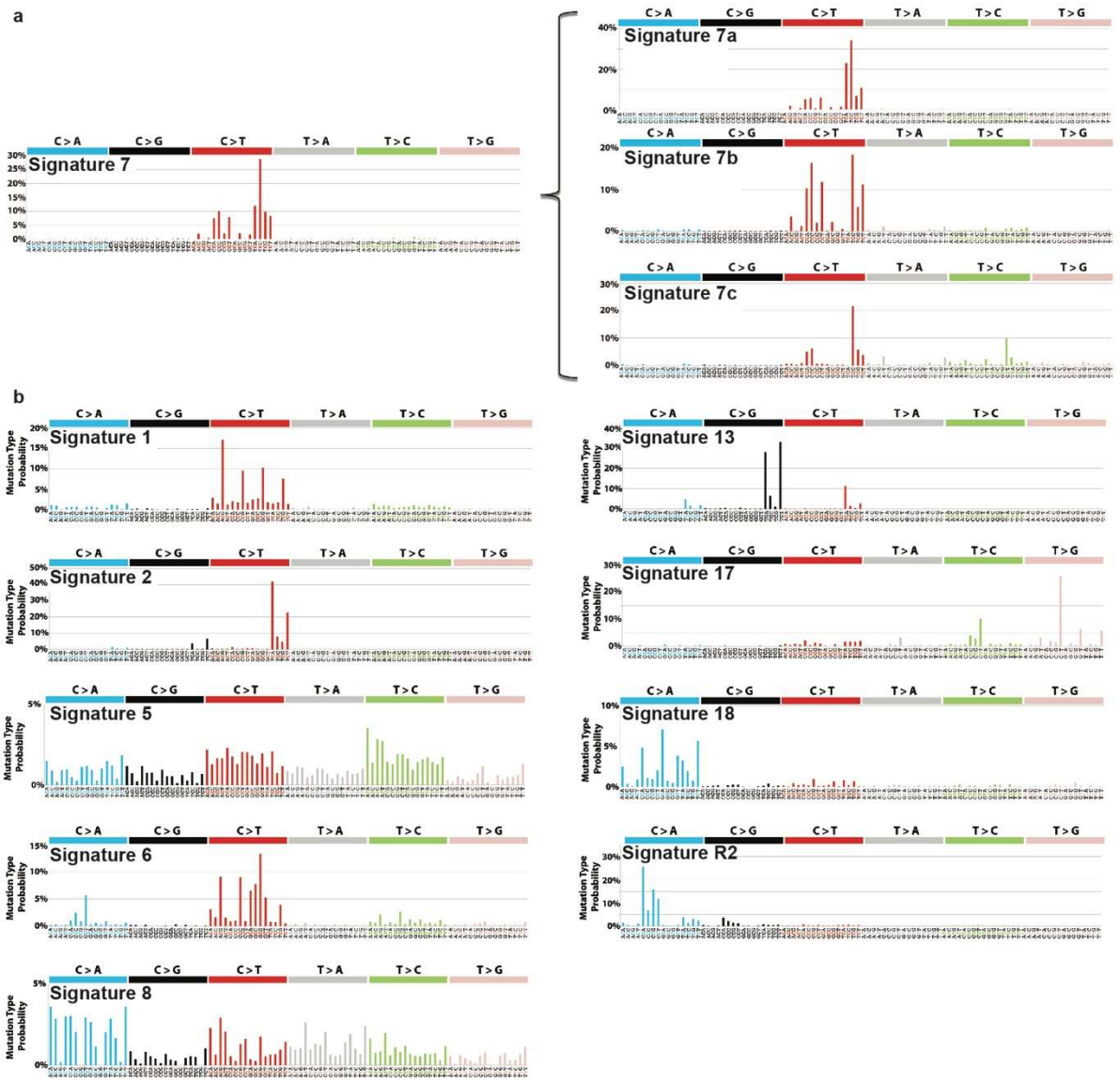
EXTENDED DATA





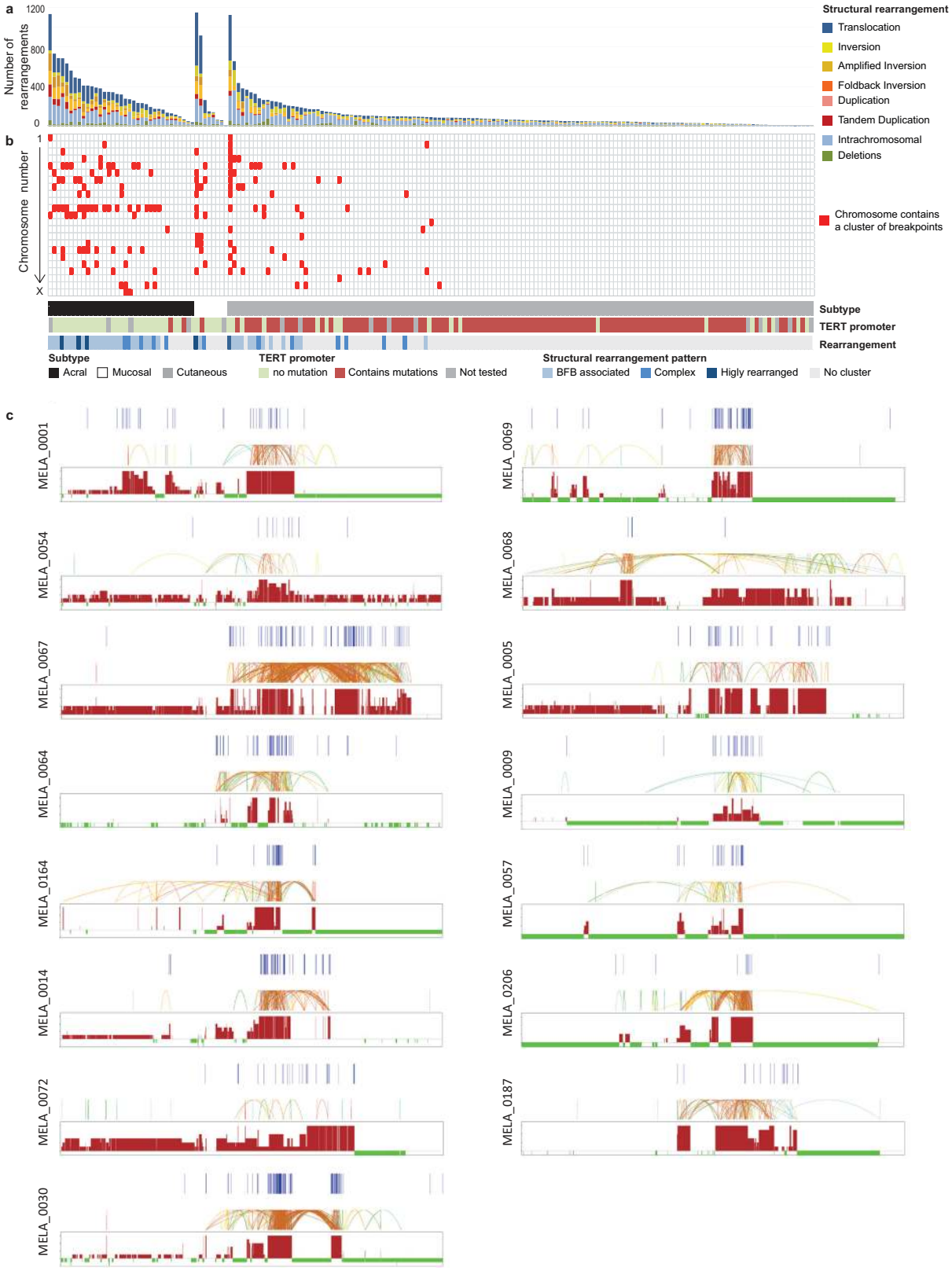


Extended Data 1 | Copy number and ploidy in melanoma. a. The proportion of each of the melanoma genomes that were affected by loss (copy number 1 or 0), copy neutral loss of heterozygosity and high gain (≥6 copies) are shown in the histogram. The melanoma subtype and degree of ploidy are illustrated in the colour bar beneath the histogram. GISTIC analysis was performed to determine significant regions of recurrent copy number change in the **b** cutaneous and **c** acral and mucosal tumours.



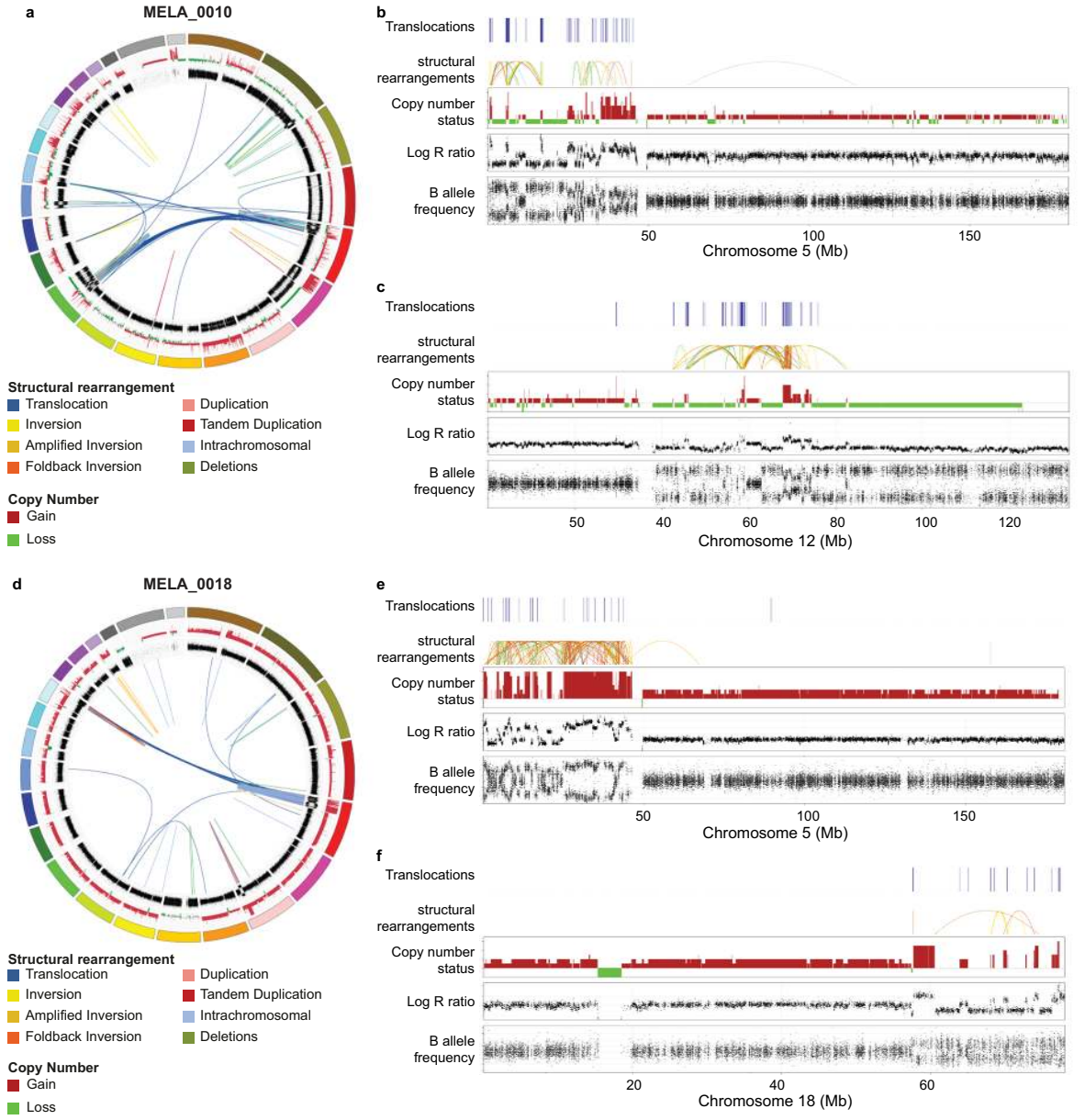
Extended Data 2 | Mutation signatures in melanoma. Twelve mutation signatures in melanoma were identified taking into account the sequence context immediately before and after the mutation. Each signature is displayed showing the probability for each of the 96 mutation types. **a,** Three novel signatures, associated with ultraviolet radiation (UVR), were identified. These signatures perfectly recapitulate the UVR mutation signature previously extracted from exome sequencing data. Signature 7a is predominately C>T transitions occurring at TpC dinucleotides and, based on similarity with the sequence context, it is most likely due to the formation of 6,4-photoproducts. Signature 7b is described by C>T transitions at CpC dinucleotides. This sequence context is characteristic for the formation of cyclobutane pyrimidine dimers due to UVR exposure. Signature 7c has a high proportion of C>T substitutions, and high levels of T>C and T>A mutations, demonstrating the ability of UVR to generate both transitions and transversions. The underlying processes damaging DNA and resulting in signature 7c are currently unknown but they could potentially be due to

indirect DNA damage. **b.** Of the nine non-UVR associated signatures, three (signature 1, 5 and 17) have previously been observed in melanoma and six have been seen in other cancer types but not previously noted in melanoma. Signature 1 is the result of an endogenous mutational process initiated by spontaneous deamination of 5-methylcytosine and is associated with age of the patient. Signature 5 (aetiology unknown) was the dominant signature in tumours with no UVR signature; it is together with signature 1 the most common signature, observed in diverse cancer types. Signature 17 (aetiology unknown) has been seen in melanoma previously and was here observed in 52 tumours (exclusively cutaneous and acrals). Signature 8 was one of the dominant signatures in tumours with no UVR signature (together with signature 5 and 1); it has previously been observed in breast cancer and medulloblastoma and was evident in 39 melanomas. Signatures 2 and 13 have been attributed to activity of the AID/APOBEC family of cytidine deaminases and are often observed in the same samples; here most non-UV samples showed both signatures, but the highest level of signature 2 was observed in a cluster of four samples in which signature 2, 7a, and 7c were observed. Signature 6 is associated with defective DNA mismatch repair and typically found in microsatellite-unstable tumours; it was observed in one melanoma in this cohort. Signature 18 has been observed frequently in neuroblastoma; it was identified in 13 melanomas. Signature R2 reflects a sequencing artefact.



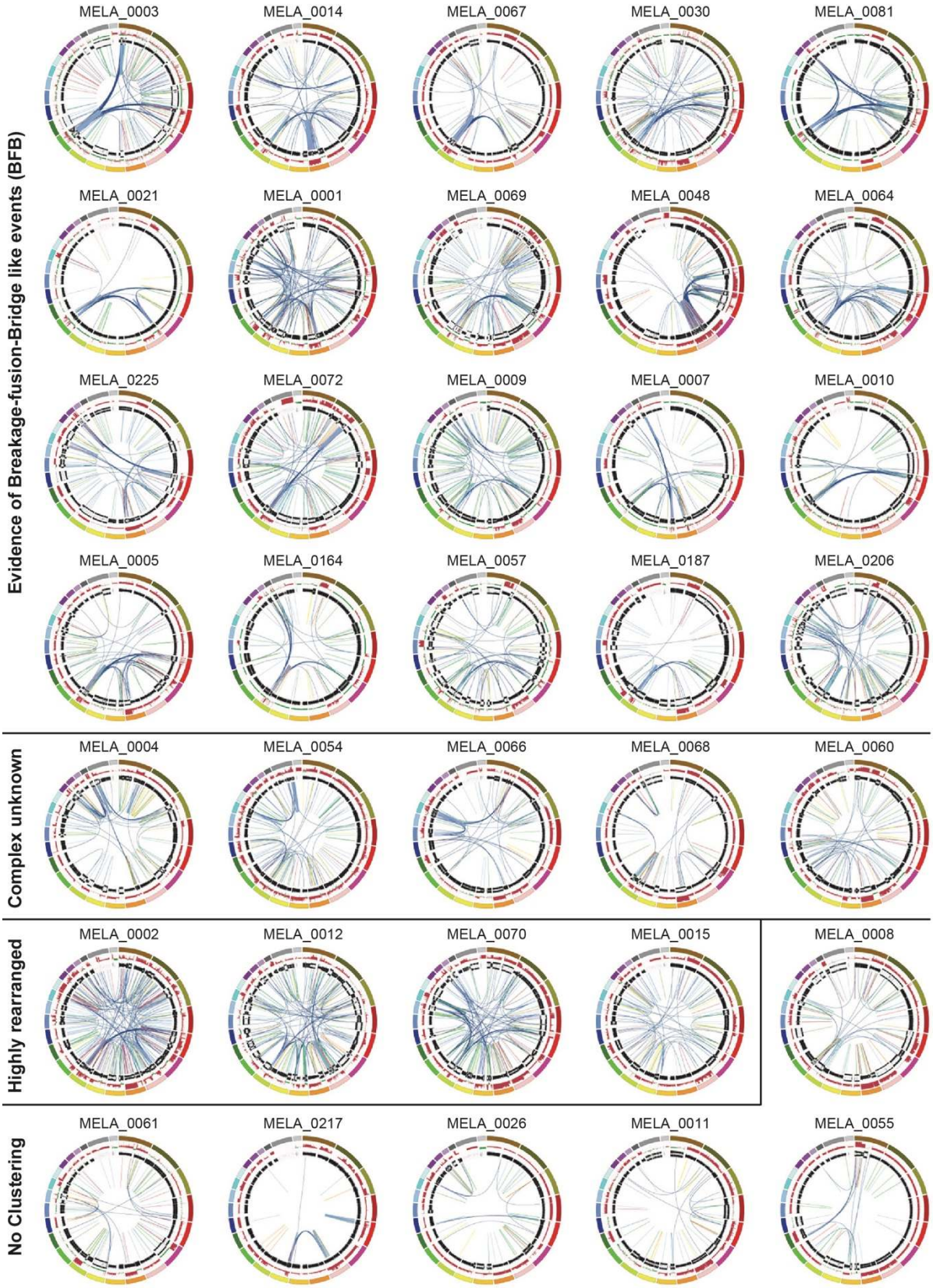
Extended Data 3 | Structural variants clustering on chromosome 11 in acral melanoma. a. The number and type of SV is shown for each tumour (n=183). **b.** Chromosomes containing complex rearrangements are shown in red defined as

chromosomes with a clustered arrangement of breakpoints (Kolmogorov-Smirnov goodness of fit test $P < 0.0001$) and a high density of breakpoints (number of breakpoints exceeds the extreme threshold value of the upper quartile plus 5* inter-quartile range for the chromosomal distribution of breakpoints). The observed pattern of breakpoint clusters and copy number events was used to classify tumours as breakage-fusion-bridge associated, complex, highly rearranged or no clusters of events. *TERT* promoter mutations are also shown. **c.** Patterns of SV and copy number in all acral melanomas that contained a cluster of SV on chromosome 11. Graphs show the most frequent translocations and other SV (copy number red=gain, green=loss).



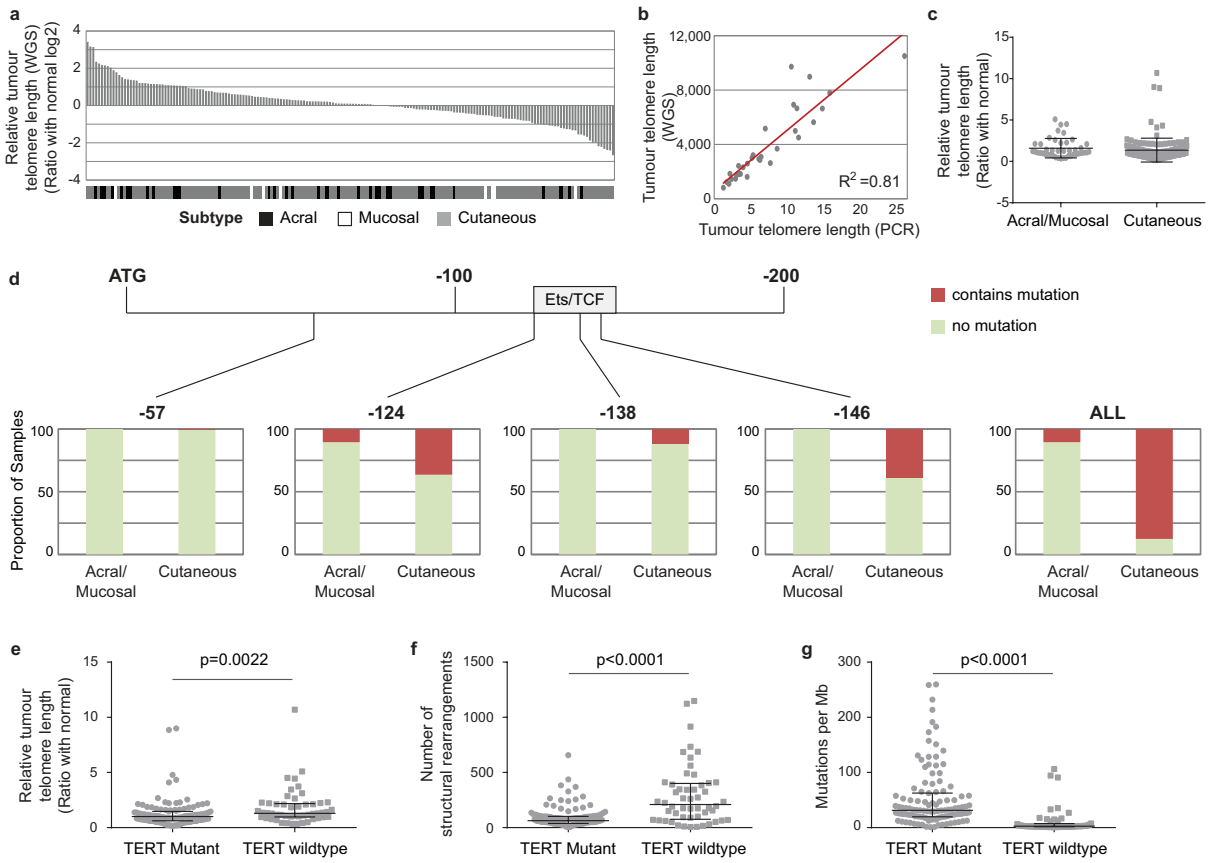
Extended Data 4 | Fine structure of two structural variants. a. Circos plot of MELA_0010 shows a cluster of SV on chromosomes 5 and 12. The outer coloured ring on the Circos plot represents each chromosome, the next ring is the copy number (red = gain and green = loss), the next ring is the B-allele frequency, and the SV are shown by the lines in the centre of the plot. **b.** The cluster of rearrangements on chromosome 5 in MELA_0010 are shown from top to bottom: translocations (blue), all other rearrangements, copy number status, logR ratio and B allele frequency. **c.** The cluster of rearrangements on chromosome 12 in MELA_0010 shows evidence of breakage-fusion-bridge (BFB) with loss of telomeric region and inversions with increased copy number. **d.** Circos plot of MELA_0018 shows a cluster of structural rearrangements on chromosomes 5 and 18. The cluster of rearrangements in **e**, MELA_0018 on chromosome 5 and **f**, chromosome 18 are shown from top to bottom:

translocations, all other rearrangements, copy number status, logR ratio and B allele frequency.

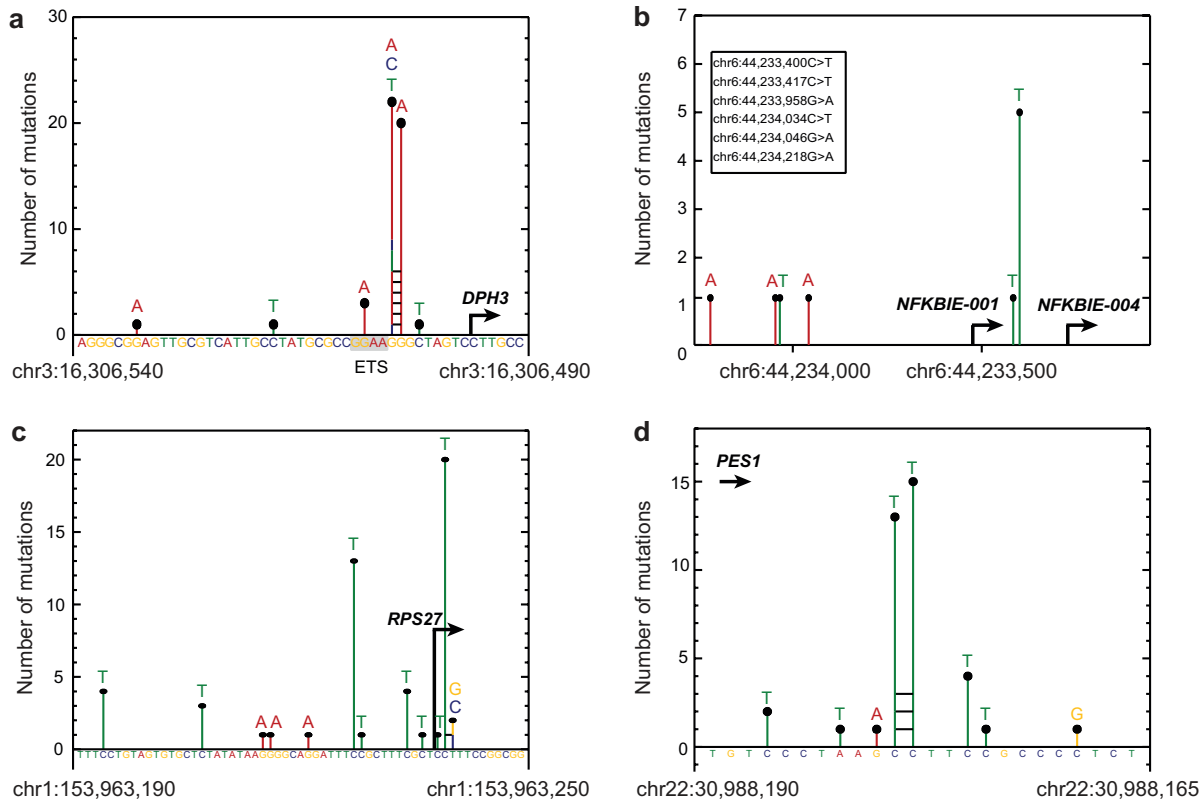


Extended Data 5 | Complex structural rearrangements are frequent in acral melanomas. Circos plots are shown for all 35 acral melanomas in the study. The outer coloured ring on the Circos plot represents each chromosome, the next ring is the copy

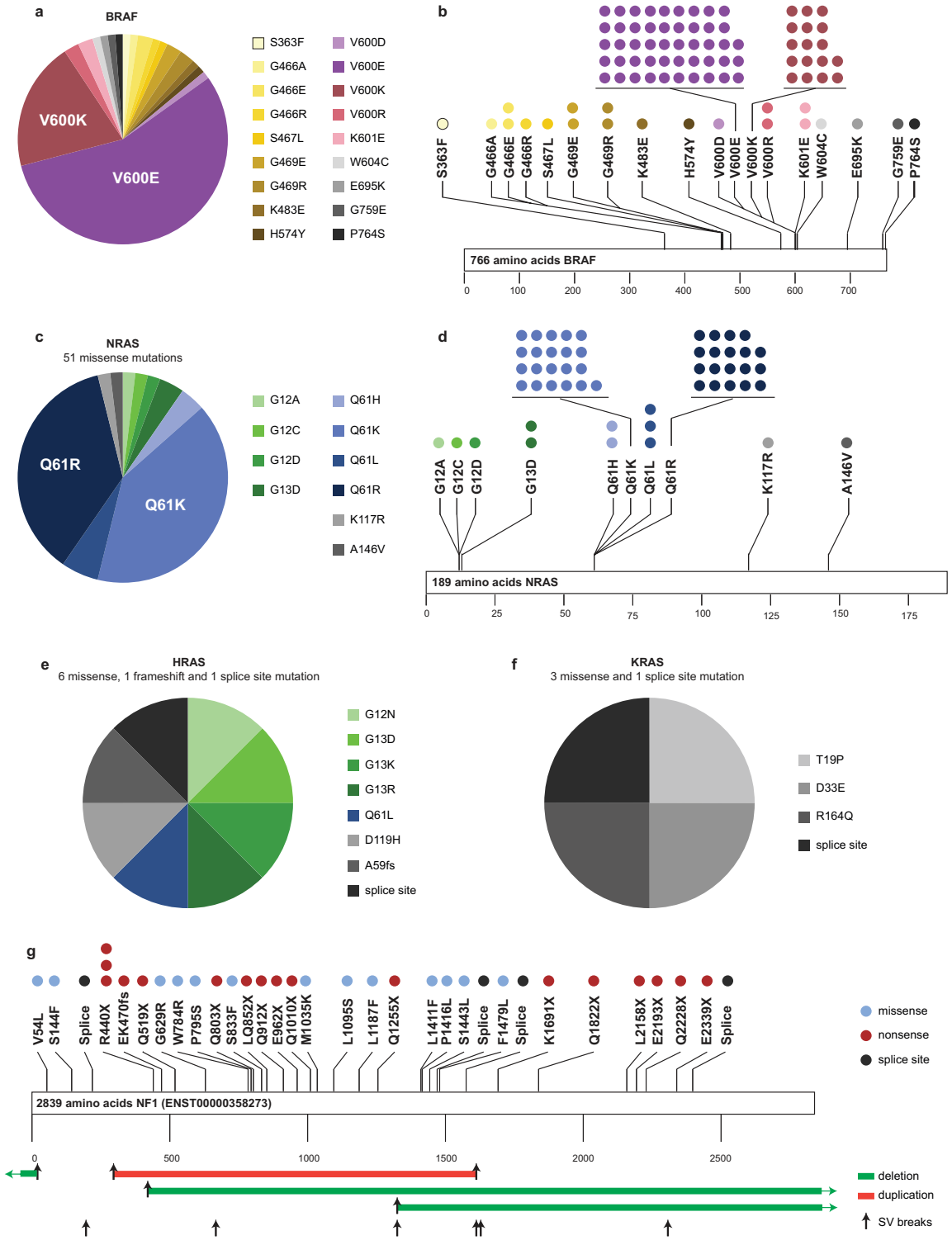
number (red = gain and green = loss), the next ring is the B-allele frequency, and the SV are shown by the lines in the centre of the plot. Tumours that contain breakage-fusion-bridge events, complex rearrangements, highly rearranged genomes or those that contain no clusters of SV are indicated.



Extended Data 6 | Telomere length and *TERT* promoter mutations. A. The relative melanoma telomere length in each tumour compared to the matched normal. **B.** Verification of telomere length in 31 samples by PCR. **c,** The telomere length was not associated with melanoma subtype. **d.** Somatic mutations in the promoter of the *TERT* gene were detected by whole-genome sequencing and/or capillary sequencing. Mutations were found at -57, -124, -138 and -146 from the ATG site. Tumours with *TERT* promoter mutations were associated with (Mann-Whitney test) **e,** telomere length; **f,** the number of structural rearrangements; and **g,** the number of mutations per Mb.

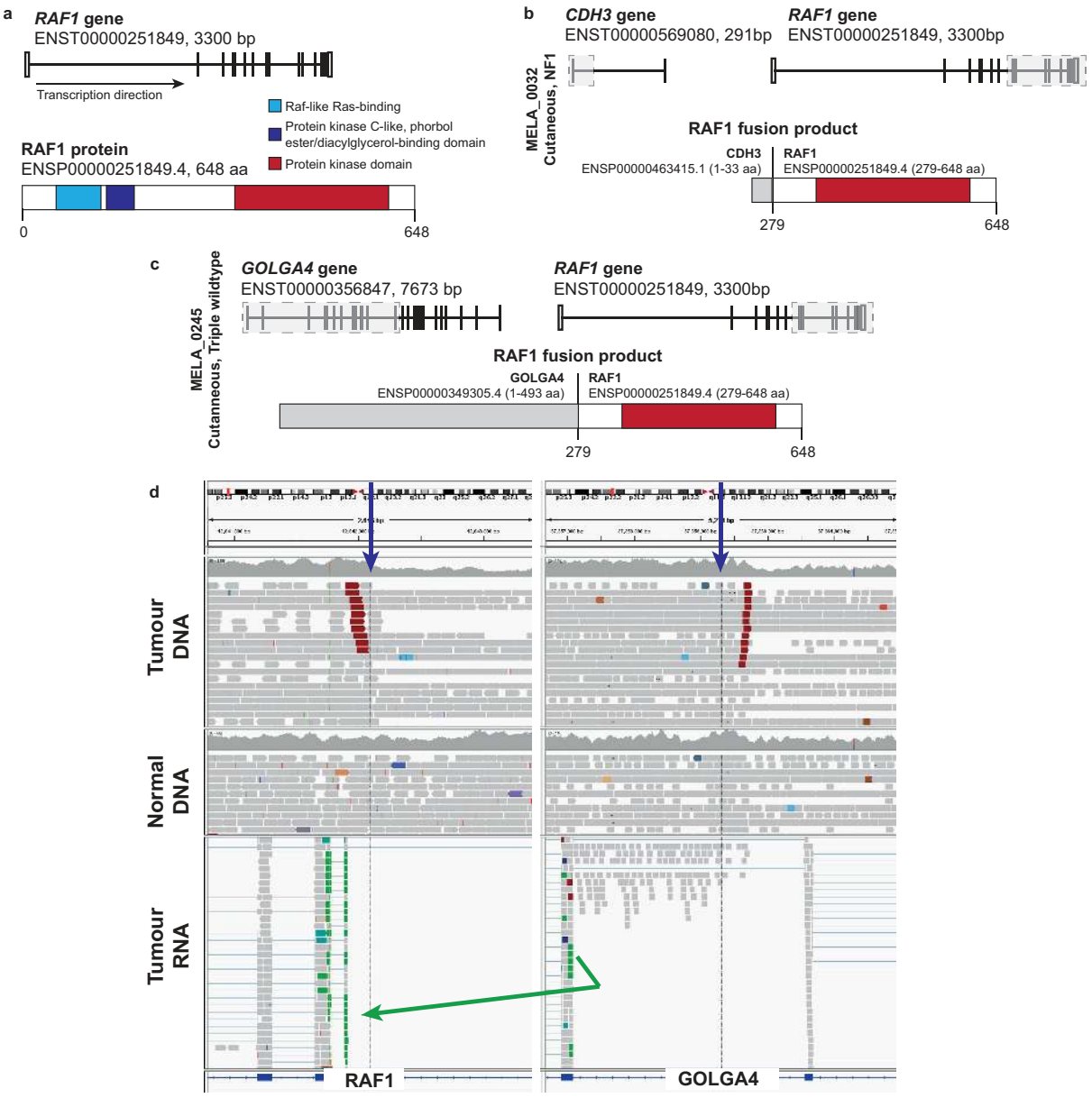


Extended Data 7 | Recurrent *DPH3*, *NFKBIE, RPS27* and *PES1* promoter mutations. Height and colour of sticks indicate recurrence and mutant base, respectively. Tandem mutations are indicated with black horizontal bars. **a.** Recurrent mutations in promoter of *DPH3* at loci 8 (n=20), 9 (n=22), and 12 (n=3) basepairs upstream from the transcription start site (TSS). The ETS transcription factor family core motif is shaded in gray. **b,** Recurrent mutation (n=5) at chr6:44,233,400 in first exon of *NFKBIE-001* isoform and 129 bases upstream from TSS of *NFKBIE-004*, which typically is the expressed transcript in most tissue including melanoma. Four novel mutations were identified further 500kb upstream. **c,** Recurrent mutations on and upstream from the TSS of *RPS27*. **d,** Recurrent mutations upstream of *PES1*.



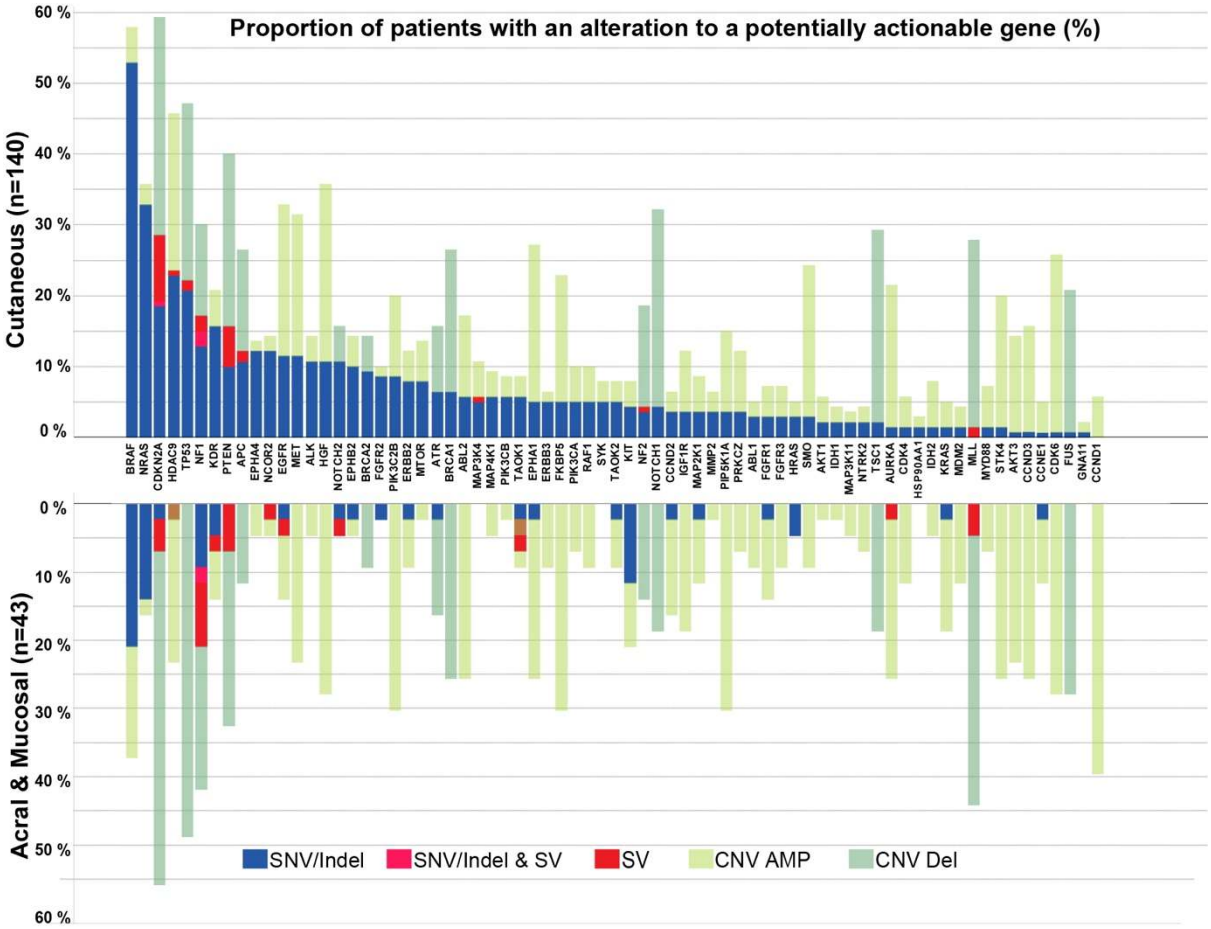
Extended Data 8 | *BRAF*, *RAS* and *NF1* mutations. a and b, *BRAF* somatic mutations were identified in 86 of the 183 samples (47 %). V600E substitutions accounted for 48 (56 %) of these mutations. Other activating mutations at codons 600 and 601 were less frequent (V600K, n=17; V600R, n=2; V600D, n=1; K601E, n=2). **c and d**, *NRAS* was mutated in 51 (28 %) tumours, 49 of these mutations occurred at 'hotspot' codons 12, 13 or 61. **e,** *HRAS* mutations occurred in 8 tumours, 5 of which were in 'hotspot' codons 12, 13 or 61. **f,** *KRAS*

mutations occurred in 4 tumours, but none occurred at the commonly activating hotspot codons. **g,** *NF1* aberrations occurred throughout the gene in 32/183 (17 %) tumours and included point mutations/small indels only (15 nonsense, 1 frameshift, 3 splice site and 4 missense). SV predicted to cause loss of function were frequent within *NF1* (SV breakpoints indicated by reverse arrows). Some rearrangements were associated with a copy number change (green = loss; red = gain).



Extended Data 9 | Candidate *RAF1* gene fusions. Whole genome sequencing data was used to identify genomic rearrangements which could result in a gene fusion product. **a,** The RAF1 gene and protein are shown. **b,** A *RAF1* gene fusion was detected with a small transcript of *CDH3*. The gene fusion is expected to produce a protein that retains the ref1 kinase domain. **c,** A RAF1-GOLGA4 gene fusion was detected which is expected to produce a protein that retains the ref1 kinase domain. **d**, IGV view of *RAF1-GOLGA4* gene fusion. DNA breakpoints (red arrows) marked by discordantly aligned red coloured read pairs that are in an incorrect orientation and too far apart. There is no matching evidence observed in the normal control sample. Supportive evidence for an expressed gene fusion is visible in RNAseq data (bottom horizontal panel) where discordantly aligned reads are highlighted in

turquoise and the orientation of a junction fusion between exon 11 of *GOLGA4* and exon 8 of *RAF1* is indicated by the green arrow.



Extended data 10 | Alterations to potentially actionable genes in cutaneous and non-cutaneous melanomas. Proportion of samples with a somatic variant containing gene that confers sensitivity to an FDA-approved or therapeutic agent being utilized in a clinical trial (in any cancer): non-silent substitution or indel (SNV/indel) blue), both SNV/indel and SV (light red), SV only (dark red) or copy number variation (CNV, high-level amplification green and homozygous deletion yellow). Genes are ordered by the difference between the frequencies of mutated genes (excluding CNV) in cutaneous versus non-cutaneous melanomas, with genes with a higher mutation frequency in cutaneous melanomas on the left.