# Ancestry and somatic profile predict acral melanoma origin and prognosis


## Authors:

Patricia Basurto-Lozada[1,2], Martha Estefania Vázquez-Cruz[1], Christian Molina-Aguilar[1], Amanda Jiang[3,4], Dekker C. Deacon[3,4], Dennis Cerrato-Izaguirre[5], Irving Simonin-Wilmer[1], Fernanda G. Arriaga-González[1,6], Kenya L. Contreras-Ramírez[1], Emiliano Ferro-Rodríguez[1], Jamie Billington[6], Eric T. Dawson[7], J. Rene C. Wong-Ramirez[1,8], Johana Itzel Ramos-Galguera[1], Alethia Álvarez-Cano[9], Dorian Y. García-Ortega[10], Omar Isaac García-Salinas[1,6], Alfredo Hidalgo-Miranda[11], Mireya Cisneros-Villanueva[11], Peter A. Johansson[12,13], Héctor Martínez-Said[10], Pilar Gallego-García[14], Mark J. Arends[15], Ingrid Ferreira[6], Mark Tullett[16], Rebeca Olvera-León[1,6], Louise van der Weyden[6], Martín del Castillo Velasco Herrera[6], Rodrigo Roldán-Marín[17], Helena Vidaurri de la Cruz[18], Luis Alberto Tavares-de-la-Paz[19], Diego Hinojosa-Ugarte[19], Rachel L. Belote[20], D. Timothy Bishop[21], Marcos Díaz-Gay[14,22-24], Ludmil B. Alexandrov[22-24], Yesennia Sánchez-Pérez[5], Gino K. In[25], Richard M. White[26,27], Patrícia A. Possik[28], Robert L. Judson-Torres[3,4], David J. Adams[6], Carla Daniela Robles-Espinoza[1,6,*]

## Affiliations:

1. Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Santiago de Querétaro, Mexico, 76230
2. Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Unidad de Posgrado, Edificio D, 1° Piso, Circuito de Posgrados, Ciudad Universitaria, Coyoacán, C.P. 04510, CDMX, México
3. Huntsman Cancer Institute, University of Utah Health Sciences Center, Salt Lake City, Utah, USA
4. Department of Dermatology, University of Utah, Salt Lake City, UT, USA
5. Subdirección de Investigación Básica, Instituto Nacional de Cancerología (INCan), San Fernando No. 22, Tlalpan, Ciudad de México CP. 14080, Mexico.
6. Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK.
7. Nvidia Corporation, Santa Clara, CA, USA.
8. Research Program in Systems Oncology, University of Helsinki, Helsinki, Finland.
9. Surgical Oncology, Christus Muguerza Alta Especialidad, Monterrey, Nuevo Leon, Mexico.
10. Surgical Oncology, Skin, Soft Tissue & Bone Tumors Department, National Cancer Institute, Mexico City, Mexico.
11. Laboratorio de Genómica del Cáncer, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, Mexico.
12. QIMR Berghofer, Brisbane, QLD, Australia
13. University of Queensland, Brisbane, QLD, Australia
14. Digital Genomics Group, Cancer Genomics Program, Spanish National Cancer Research Center (CNIO), Madrid, Spain
15. Edinburgh Pathology, Cancer Research UK Scotland Centre, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK
16. Department of histopathology, University Hospitals Sussex, St Richard hospital, Spitalfield lane, Chichester
17. Dermato-Oncology Clinic, Research Division, Faculty of Medicine, Universidad Nacional Autónoma de México, Mexico City, Mexico.
18. Pediatric Dermatology Service, General Hospital of Mexico Dr. Eduardo Liceaga, Ministry of Health. Mexico City, Mexico.
19. Surgical Oncology, Bajio Regional High Specialty Hospital, Leon, Mexico.
20. The Ohio State University, Department of Molecular Genetics, Columbus, Ohio, United States
21. Leeds Institute of Medical Research, University of Leeds, Leeds, UK.
22. Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA.
23. Department of Bioengineering, University of California San Diego, La Jolla, CA, USA.
24. Moores Cancer Center, University of California San Diego, La Jolla, CA, USA.
25. University of Southern California, Keck School of Medicine, Norris Comprehensive Cancer Center, Division of Oncology, Los Angeles, CA, USA.
26. Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

27. Nuffield Department of Medicine, Ludwig Institute for Cancer Research, University of Oxford, Oxford, UK
28. Division of Basic and Experimental Research, Brazilian National Cancer Institute, Rua Andre Cavalcanti 37, Rio de Janeiro, RJ, 20231-050, Brazil.

***Correspondence to:** Carla Daniela Robles-Espinoza, drobles@liigh.unam.mx

**Abstract**

Acral melanoma, which is not ultraviolet (UV)-associated, is the most common type of melanoma in several low- and middle-income countries including Mexico. Latin American samples are significantly underrepresented in global cancer genomics studies, which directly affects patients in these regions as it is known that cancer risk and incidence may be influenced by ancestry and environmental exposures. To address this, we characterise the genome and transcriptome of 123 acral melanoma tumours from 92 Mexican patients, a population notable because of its genetic admixture. Compared with other studies of melanoma, we found fewer frequent mutations in classical driver genes such as *BRAF*, *NRAS* or *NF1*. While most patients had predominantly Amerindian genetic ancestry, those with higher European ancestry had increased frequency of *BRAF* mutations and a lower median number of structural variants. The tumours with activating *BRAF* mutations have a transcriptional profile more similar to cutaneous non-volar melanocytes, suggesting that acral melanomas in these patients may arise from a distinct cell of origin compared to other tumours arising in these locations. *KIT* mutations were found in a subset of these tumours, and quadruple wild-type samples (non *BRAF/NRAS/NF1/KIT*) differed from mutated samples in their structural genomic profile and overall and recurrence-free survival patterns. Transcriptional profiling defined three expression clusters; these characteristics were associated with recurrence-free and overall survival. We highlight potential novel low-frequency drivers, such as *PTPRJ*, *NF2* and *RDH5*. Our study enhances knowledge of this understudied disease and underscores the importance of including samples from diverse ancestries in cancer genomics studies.

## Introduction

Melanoma is classified into several clinicopathological subtypes based on tumour site of presentation and histopathological features. Acral melanoma (AM) is an understudied melanoma subtype due to its low incidence globally, and because it represents a small proportion of melanoma cases in European-descent populations[1,2]; however, AM represents the vast majority of melanoma cases in some Latin American, African and Asian countries due to the lower incidences of ultraviolet (UV)-induced melanoma subtypes[3]. Additionally, the causes of this type of disease are unknown, with patients managed in a similar way to UV-associated cutaneous melanoma (CM). However, its site of presentation and genomic characteristics are vastly different[4].

AM arises on the glabrous (non-haired) skin of soles, palms and in the nail unit (subungual location), and its genome differs significantly from other CM subtypes[5]. In contrast to UV-linked subtypes like superficial spreading or lentigo maligna melanoma, AM has a lower burden of single nucleotide variants (SNVs), a higher burden of structural variants, and a low prevalence of mutational signatures SBS7a/b/c/d, which are associated with UV irradiation[6–10]. Genes that are frequently mutated in CM such as *BRAF*, the *RAS* genes and *NF1*, are reported to be altered at a significantly lower frequency in AM. This, coupled with the comparatively lower number of studies of AM when compared to other CM subtypes, has translated into limited available therapies for AM management.

It is known that cancer risk and incidence, as well as tumour genomic profiles, vary with ancestry and geographical location[11–13]. Since most genomic studies on AM have been performed on patients of European or Asian ancestry, we considered it necessary to examine the genomics of this subtype of melanoma in Latin Americans. Specifically, Latin American populations have been substantially underrepresented in cancer genomic studies, with only about 1% of all samples in cohorts such as the Pan-Cancer Analysis of Whole Genomes (PCAWG), the Cancer Genome Atlas (TCGA) and other repositories, and those contributing to cancer genome-wide association studies (GWAS), being of Latin American origin[14–16]. Identification of differences in the genomic profile among populations can potentially aid the discovery of germline/inherited or environmental factors related to AM aetiology, as well as identify optimal therapeutic strategies for all patients.

113 In this study, we analysed 123 AM samples from 92 Mexican patients through genotyping, exome
114 sequencing, SNV and insertion/deletion (indel) variant calling, copy number estimation, and gene
115 expression profiling, and examined the correlation of these molecular characteristics with clinical
116 variables. We reveal a significant correlation between genetic ancestry and *BRAF* somatic
117 mutations, as well as a distinct transcriptomic profile in these tumours compared to non-*BRAF*
118 mutated samples. We also identify significant differences in recurrence-free and overall survival
119 among patients with driver mutations compared to patients with no mutations in driver genes, as
120 well as among patients with tumours with distinct gene expression profiles.
121
122 **Results**
123
124 *Ancestry and clinical characteristics of Mexican AM patients*
125 A total of 123 uniformly ascertained samples from 92 patients from a large Mexican tertiary referral
126 hospital were analysed in this study (**Methods**, **Supplementary Table 1**). Eighty-nine of these
127 tumours were primaries, 27 were metastases, five were recurrences, one was an in transit lesion,
128 and one was unknown (**Supplementary Table 1**). Latin American genomes are generally a
129 mixture of European, African and Amerindian ancestry. Of note, 90% of genotyped samples
130 (n=80) in this study had predominantly Amerindian ancestry (median 81%) (**Supplementary**
131 **Figure 1, Supplementary Table 2**) with European and African ancestries contributing a median
132 of 13.6% and 2.5%, respectively. The median age of the patients in this cohort was 60, with 59%
133 of the patients being female. Most patients were stage III (AJCC 8th edition)[17] at diagnosis, and
134 the most common primary site was the foot, most frequently the sole. The median Breslow
135 thickness was 4.0mm and most primary tumours were ulcerated (68%) (**Table 1**). It should be
136 noted that only four patients received immune check point inhibitors or targeted therapy, due to
137 lack of access.
138
139 *Genomic profiling of AM samples identifies correlations of sex and ancestry with somatic*
140 *alterations*
141 Considering all 123 samples, AM tumours showed a SNV+indel [hereinafter referred to as tumour
142 mutational burden (TMB)] mean of 0.95 mutations per megabase (mut/Mb) and a median of 0.87
143 mut/Mb (range: 0-3.49 mut/Mb). When including only one sample per patient, with primaries being
144 preferentially selected, the most frequently mutated genes were *NRAS* (14% of samples, *q*-value
145 $< 4.97\times10^{-10}$), *KIT* (14% of samples, *q*-value = $4.97\times10^{-10}$), *BRAF* (13%, *q*-value=$3.86\times10^{-7}$) and
146 *NF1* (9%, *q*-value=0.0001) (**Figure 1a, Supplementary Table 3**). For *BRAF*, all mutations except

147 one were V600E, with one L597R (**Figure 1b**). These genes were identified as being under
148 positive selection (**Methods**) and represent known driver genes. These genes showed the
149 characteristic mutational profile of oncogenes with a predominance of hotspot missense
150 mutations, except for *NF1*, which showed a pattern characteristic of a tumour suppressor and had
151 large deletions (an exon or larger), frameshift insertions, deletions, splice donor and nonsense
152 mutations distributed throughout (**Figure 1b**). Notably, these genes exhibit mutual exclusivity
153 (only one patient has tumours with mutations in more than one of these genes) which likely reflects
154 their functional redundancy in activating the MAPK pathway. Separate capillary sequencing of the
155 *TERT* promoter in 76 samples belonging to 64 patients identified that six carry the -124 promoter
156 mutation (9.3%) and two out of 59 patients for which the -146 position was successfully amplified
157 carry a mutation in this position (3.4%) (**Supplementary Table 4**). In total, we estimate that 10.5%
158 of patients have an activating *TERT* promoter mutation, which is similar to estimates in other
159 studies[7,18]. All samples from all patients that had multiple samples sequenced and that could be
160 assessed had a concordant *TERT* genotype, in agreement with an early emergence of this
161 mutation during tumour evolution[18]. Other genes previously reported as mutated in other
162 melanoma subtypes, as well as other cancer types are also mutated in this cohort, such as *TP53*,
163 *HRAS* and *KRAS* (**Figure 1a, Supplementary Figure 2**). In summary, the “classic” melanoma
164 driver genes (*N/H/KRAS*, *BRAF* and *NF1*) are mutated in 40% of Mexican AM samples, with most
165 of the samples in this cohort therefore being classified as "triple wild type" melanomas. Apart from
166 the known *HRAS*, *SPRED1*, *TP53*, and *KRAS* driver genes, we also find mutations in *PTPRJ*,
167 *ATM*, *NF2* and *RDH5* (**Supplementary Figure 2a,b**). Specifically, in those tumours without
168 mutations in any of the abovementioned four driver genes [*BRAF*, *NRAS*, *KIT*, *NF1*, “quadruple
169 wild-type” (QWT)], we find two tumours each from different patients with deleterious mutations in
170 *ATM* and *RDH5* (**Supplementary Figure 2b**). The mutations in these genes are all protein-
171 changing and deleterious. All these genes have previously been linked to tumour suppressor
172 activities in either acral or mucosal melanomas[19–24], as well as other cancer types, and may
173 represent low-frequency drivers. We also observe a significantly higher proportion of women
174 carrying mutations in driver genes vs men (Two-tailed Fisher’s test $P$=0.003) (**Supplementary**
175 **Table 5**). After adjusting for date of diagnosis, age at diagnosis, ancestry and tumour stage, the
176 Odds Ratio (OR) of having a mutation in a driver gene in female patients (compared to men) was
177 estimated to be 3.83 (95% CI 1.32, 11.07) (multivariate logistic regression, $P$ = 0.013).
178
179 When examining the relationship between ancestry and somatic profile, we identified significantly
180 higher odds ($P$-value=0.02) of carrying a *BRAF* somatic mutation with increasing European

181 ancestry in a linear model controlling for age at diagnosis, sex, and TMB (**Figure 1c**). Patients
182 with mutations in *KIT* showed a tendency for higher Amerindian ancestry (**Figure 1c**,
183 **Supplementary Figure 3**). We also found that patients with *NRAS* mutations were younger at
184 diagnosis (median and mean age of diagnosis for patients with *NRAS* mutations = 49 and 50.84
185 vs without = 63 and 62.9, respectively), but this effect is likely mediated by ulceration status, as
186 patients with *NRAS* mutations have a significantly lower rate of ulceration (Two-tailed Fisher's
187 Exact test *P*=0.016).
188
189 Out of 22 patients for whom we sequenced at least two different samples (*e.g.*, a primary and a
190 metastasis), 13 have mutations in *NRAS/KIT/BRAF/NF1*. For *BRAF*, all four patients have
191 mutations across all samples, and for *NRAS*, three out of four patients have a *NRAS* mutation in
192 both the primary and lymph node metastasis. These data suggest that these mutations appear
193 as an early event in tumour evolution. For patients with *KIT* mutations, about half of the times the
194 mutation is found in the primary and lost in the metastasis. These data agree with Wang *et al*
195 (2024)[18], and suggest that metastases are seeded before the appearance of these mutations
196 (**Supplementary Table 6**).
197
198 Collectively, these results are similar to those reported in other acral melanoma studies[6,7,10,25–28],
199 with some important differences: First, the genetic composition of the patients in our study
200 includes a significant proportion of Native American ancestry, which is severely underrepresented
201 in already published cancer genomics studies and permits the identification of relationships of
202 specific ancestries with somatic characteristics. In addition, the fraction of activating *BRAF*
203 mutations is lower than in the studies with predominantly European-descent patients, and more
204 similar to those with Chinese patients, probably due to the positive relationship between *BRAF*
205 mutation and European ancestry (**Figure 1d**).
206
207 *Somatic copy number landscape of AM samples identifies correlations with somatic alterations*
208 Somatic copy number alteration (SCNA) analysis across all samples showed a higher burden of
209 amplifications than deletions (**Figure 2a**). Examination of 47 samples, one per patient, that
210 passed our stringent quality filtering for this type of analysis (**Methods**), showed that 18 regions
211 were significantly amplified, and six regions were frequently deleted. About a quarter (11, 23%)
212 of these 47 samples had whole genome duplication (WGD) events. Potential driver genes in
213 frequently amplified regions include *TERT* (43%), *CRKL* (36% of samples), *GAB2* (30%) and
214 *CCND1* (28%) (**Supplementary Tables 4,7-9**). Regions that showed recurrent deletions

contained genes such as *CDKN2A*, *CDKN2B, ATM* and *TP53*. Specifically, *CDKN2A* and *CDKN2B* had deletions in 57% of samples, with only two of them (out of 27) being homozygous deletions, while *ATM* and *TP53* both had deletions in 38% and 34% of samples, respectively. No association was found between ancestry and any of the significantly altered CNA regions.

SCNA profiles varied depending on whether samples had mutations in driver genes or were QWT. Specifically, samples with mutations in driver genes (n=22) showed preferential amplification of *NOTCH2* (*P*=0.021, one-sided exact Fisher test), whereas *CCND1* (*P*=0.044), *ARF6* and *SOS2* (both in same amplification peak, *P*=0.043, all one-sided exact Fisher tests) were preferentially amplified in QWT tumours (n=25). The 8p12 region, containing genes such as *FGFR1* and *TACC1*, was only found amplified in five QWT samples, whereas regions 11q13.4 and 6p24 were found amplified along several deleted regions in mutated tumours (**Supplementary Tables 10-15, Supplementary Figures 4-5**).

When stratifying samples by mutational status (considering *BRAF*-, *NRAS*-, *NF1*-, *KIT*-mutated and multi-hit, which included one sample with mutations in more than one of these drivers), we saw statistically significant differences in SCNA among groups, specifically, *NRAS*- and *BRAF*-mutated tumours had significantly fewer SCNAs (Global copy number alteration score [GCS], **Methods**) than *NF1*-mutated tumours (**Figure 2b-c**). Samples without mutations in these drivers had a range of GCS scores. Considering all samples, those with *NRAS* and *BRAF* mutations had the lowest median TMB, while *KIT*-mutated and multi-hit tumours had the highest median TMB (**Supplementary Figure 6**). We found a significant correlation between GCS score and TMB (Pearson's product moment correlation coefficient=0.72, *P*-value<0.0001) (**Figure 2d**). Tumours from the subungual region also had a higher median GCS score than those found on the feet (**Figure 2e**).

*Mutational signature analyses identify potential sources of mutation*

Single-base substitution mutational signature analysis across 116 samples that carried at least one SNV revealed previously reported COSMICv3.4 signatures SBS1, SBS5, and SBS40a (**Supplementary Figure 7**). Apart from clock-like signatures SBS1 and SBS5[29], SBS40a was also prevalent across the cohort, contributing 53.38% of mutations to the total (**Supplementary Figure 8**). SBS40a is of unknown origin but was originally identified in kidney cancer and is present in many cancer types[30]. Copy number signature analysis identified a number of previously reported signatures across different samples (**Methods**, n=60 samples)[31,32]. CN1, which has been

249 associated with a diploid state and CN9, which is potentially caused by local loss of heterozygosity
250 on a diploid background, dominated the CN landscape (**Methods, Supplementary Figure 7**). As
251 expected, signatures related to chromothripsis (CN7, CN8) were also found in several samples
252 across the cohort. The number of indels in the samples was too low to add meaningful information
253 (average: 2.52 indels per sample), so signature analysis for indels was not performed. We
254 similarly did not find any significant associations between ancestry and any mutational signature.
255 Nevertheless, this analysis is limited by the small numbers of mutations and the formalin-fixed
256 paraffin-embedded (FFPE) origin of these samples.
257
258 *Acral melanomas with activating* BRAF *mutations exhibit a transcriptional signature more*
259 *characteristic of non-acral cutaneous melanomas*
260 It has been previously postulated that *BRAF*-mutated acral melanomas might be more biologically
261 similar to melanomas from non-acral sites than to other acral melanomas[10,33], and, due to the
262 observed correlation of European ancestry with *BRAF* mutation rate, we decided to investigate
263 this hypothesis. We successfully extracted and sequenced RNA from 77 primary tumours from
264 different patients (**Supplementary Table 1, Methods**). We then generated a gene signature-
265 based score for identifying acral- versus cutaneous-derived melanomas. For this, we sourced a
266 list of candidate genes from AM and CM datasets (**Methods, Supplementary Table 16**) and
267 identified twenty genes with high classification accuracy in a training cohort of 10 primary AMs
268 (used to derive a v-mel score, or "A" for acral) and 10 primary CMs (used to derive a c-mel score,
269 or "C" for cutaneous) recruited at the University of Utah (**Figure 3a-b**). We then obtained scores
270 (v-mel/c-mel, or A:C) for samples in our dataset of AMs, separating primary *BRAF*-activating
271 (n=10) vs *BRAF*-wildtype (n=67) tumours. We observed a difference between *BRAF*-activating
272 and *BRAF*-wildtype tumours (*P*-value=0.045), with *BRAF*-activating tumours having a score
273 closer to CMs (**Figure 3c**). We then replicated this analysis in an independent cohort of 63 AMs
274 from Newell *et al* (2020)[7] (*BRAF*-activating n=10, wild-type n=53), which further confirmed these
275 results (*P*-value=0.039) (**Figure 3d**). Therefore, we explored the possibility that this difference
276 could be due to downstream mutated *BRAF* signalling. First, we replicated this analysis in the
277 TCGA cutaneous melanoma data, finding no significant differences among *BRAF* and non-*BRAF*
278 mutated samples (**Supplementary Figure 9a**). We then examined datasets in which mutant
279 *BRAF* was introduced into primary melanocytes in a doxycycline-inducible manner[34]. We found
280 that the c-mel signature genes were not activated downstream of mutant *BRAF*, further
281 suggesting that the classifier does not simply reflect *BRAF*-driven transcriptional changes (**Figure**
282 **3d**). Using a recently developed method for assessing gene signature similarity[35], we compared

283 the c-mel gene signature from our classifier to a previously published set of genes directly
284 activated by mutant *BRAF* in melanoma cells[36]. We found no significant correlation between these
285 signatures (**Supplementary Figure 9b**).
286
287 In conclusion, in these comparisons, *BRAF*-activating tumours expressed a more "CM-like"
288 transcriptional program, indicating that *BRAF*-mutated melanomas that occur at acral sites are
289 transcriptionally closer to non-acral cutaneous melanomas, a transcriptional program that is not
290 explained by *BRAF* downstream signalling and are associated with increasing European genetic
291 ancestry.
292
293 *Transcriptional landscape of AM tumours identifies three subgroups with distinct clinical and*
294 *prognostic characteristics*
295 We then applied a more stringent quality filter, including coverage and alignment features, to
296 primary tumours in this collection with 44 samples remaining for further analyses (**Methods**,
297 **Supplementary Table 1**). Consensus clustering of gene expression identified three sample
298 groups with distinct transcriptional profiles (**Methods**, **Figure 4a-b, Supplementary Table 17**).
299 Cluster 1 was characterised by an epidermal/immune profile, with high expression of keratin
300 genes (*e.g*. *KRT1*, *KRT9*), cytokines (*e.g*., *CXCL16*, *IL7, IL4R, IL1R, IL15RA, CXCL14*), and
301 processes such as epidermis development, cell-cell junction organisation and wound healing
302 (**Supplementary Figure 10**, **Supplementary Table 18**); Cluster 2 was characterised by a
303 mitotic/proliferative-related signature, with high expression of genes such as *MITF* and *TYR,* and
304 processes such as chromosome segregation, nuclear division and mitochondrial translation
305 (**Supplementary Figure 11**, **Supplementary Table 19**), and Cluster 3 showed expression mostly
306 of respiration and oxidative phosphorylation-related genes and some pigmentation-related genes
307 (e.g. *SOX10*, *DCT*) (**Supplementary Figure 12**, **Supplementary Table 20**). Interestingly, Cluster
308 1 was associated with better prognostic clinical characteristics, such as a smaller proportion of
309 ulcerated samples (53%, vs 93% in Cluster 2 and 57% in Cluster 3), a tendency for earlier clinical
310 stages, and lower mitotic rates (Mann-Whitney U C1 vs C2 *P*-value: 0.041, C1 vs C3 *P*-value:
311 0.052) (**Figure 4c**). Deconvolution of gene expression profiles also indicated differences in
312 immune cell infiltration composition, with Cluster 1 having a higher proportion of CD4+ T cells and
313 cancer-associated fibroblasts (CAFs), and Cluster 2 having a higher proportion of B-cell infiltration
314 (**Figure 4d-f, Supplementary Figure 13, Supplementary Table 21**).
315
316 *Somatic and gene expression profile influence overall and recurrence-free survival*

317 Next, we evaluated whether the genomic and transcriptomic characteristics had any impact on
318 patient overall or recurrence-free survival. We included in the analysis those participants whose
319 primary could be analysed (n=85, **Methods**). The mean time between diagnosis and recruitment
320 was 2.06 years, including 21 participants recruited within 6 months; the range was from a few
321 days to over 10 years. Among these participants, twelve primary tumours had an *NRAS* mutation,
322 eleven had a mutation in *KIT*, eleven had a *BRAF* mutation, seven had *NF1* mutations, one had
323 multiple hits and 43 were classified as wild-type.
324
325 Carrying any driver mutation was not associated with age at diagnosis or tumour stage (data not
326 shown). Having a tumour with a driver mutation was, however, associated with a reported
327 recurrence, with 66.7% of mutated tumours having a recurrence as compared to 37.2% of QWT
328 tumours (Pearson two-tailed Chi-squared test $P$ =0.007). After adjusting for date of diagnosis,
329 sex, age at diagnosis, ancestry and tumour stage (n=73, primaries with ancestry information
330 available), the OR for a mutated tumour having a recurrence compared to QWT tumours was OR
331 = 5.31 (95% CI 1.56, 18.12), (multivariate logistic regression, $P$ = 0.008) (**Figure 5a**). Notably,
332 among the mutated tumours, for each different gene, tumour recurrence was increased over QWT
333 tumours (**Figure 5a**, **Supplementary Table 22**), most notably for *NF1* where all seven of the
334 mutated tumours recurred.
335
336 Overall, 44 of the tumours could be classified transcriptomically into one of the three clusters.
337 There was no association between tumour driver mutation and transcriptomic cluster (data not
338 shown). There was, however, evidence of differences in recurrence frequency by cluster, with
339 35.7% of cluster 1 tumours, 81.2% of cluster 2 tumours and 57.1% of cluster 3 tumours having a
340 recurrence (Fisher's Exact Test, $P$ = 0.04 for homogeneity). Logistic regression adjusting for age
341 at diagnosis, sex, diagnosis date and stage at presentation showed that those tumours from
342 cluster 2 had a higher rate of recurrence as compared to cluster 1 (OR = 6.68, 95%CI 0.97, 46.3),
343 multivariate logistic regression, $P$ = 0.054, **Supplementary Table 23**) while cluster 3 had
344 intermediate rates of recurrence (**Figure 5b**).
345
346 Fourteen participants (16.5%) died during the study period. Seven percent of participants with
347 QWT tumours and 26.2% of participants with tumours with driver mutations died ($P$ = 0.017 for
348 homogeneity). Log rank analysis of time to death from diagnosis showed increased risk of death
349 among those with any mutation (n=73, primaries with ancestry information available) ($P$ = 0.036)
350 (**Supplementary Table 24**) while similar analysis by specific mutation showed more extreme

351 significance ($P < 0.0001$) (**Supplementary Table 25**). Cox proportional hazards analysis
352 adjusting for age, sex, tumour stage and ancestry indicated participants whose tumour had any
353 mutation in a driver gene had a significant increase in death in keeping with the analysis of
354 recurrence (HR =5.07, 95% CI 1.02, 25.17, $P = 0.047$) (**Supplementary Table 26**) (**Figure 5c)**.
355
356 Finally, survival analysis based on the 44 tumours with transcriptomic classification showed
357 significant variation between the clusters with 0% of cluster 1, 37.5% of cluster 2 and 21.4% of
358 cluster 3 having died (Pearson two-tailed Chi-squared $P$=0.04), again in keeping with the analysis
359 of recurrence, a known major risk factor for survival (**Supplementary Table 27**, **Figure 5d**). Cox
360 proportional hazards analysis adjusting for age at diagnosis, sex, stage at presentation and
361 ancestry did not provide significant evidence of differences between the clusters in terms of
362 mortality rates in keeping with the limited sample size.
363
364
365 **Discussion**
366
367 In this study, we report the analysis of the somatic and transcriptomic profile of 123 acral
368 melanoma samples from Mexican patients, one of the largest cohorts reported for this type of
369 cancer. In our view, this study helps address several research gaps: 1) The underrepresentation
370 of samples of Latin American ancestry in cancer sample repositories[14]: As it has been shown
371 previously, genetic ancestry and environment influence the somatic profile of tumours, with
372 potential impacts on patient management and treatment[11–13], 2) the relative lack of studies of acral
373 melanoma, when compared to other types of the disease, as this type of melanoma constitutes
374 the majority of cases in some low- and middle-income countries (LMICs)[3], and 3) the relative
375 paucity of genomic studies performed and directed from LMICs, such as Mexico.
376
377 Most patients in this study had predominantly Amerindian genetic ancestry, which allowed us to
378 perform an analysis of genetic ancestry correlates with somatic mutation profile. We identified a
379 positive correlation between European ancestry and *BRAF* mutation rate (**Figure 1c**). A possible
380 link between European ancestry and $BRAF^{V600E}$ mutation had been described previously[10], and
381 this study provides further confirmatory evidence. Other similar correlations have recently been
382 described for other types of cancer, such as a positive relationship between Native American
383 ancestry and *EGFR* mutation rate in lung cancer[13], and an increased rate of somatic *FBXW7* in
384 African patients compared to European patients[11]. In accordance with this observation, other

385 cohorts of acral melanoma, which studied patients with predominantly European ancestry, have
386 a higher *BRAF* mutation rate than that in this study (*e.g.*, 23% in Australian patients with
387 predominantly European ancestry[7]). These observations should provide the basis for future
388 studies exploring the relationships between ancestry and somatic mutation rate.
389
390 We were intrigued to discover that *BRAF*-missense acral melanomas exhibit a more 'CM-like'
391 transcriptome than other genetic subtypes of acral melanoma. One possible explanation is that
392 this gene signature is uniquely downstream of a *BRAF* missense mutation. However, in further
393 analyses, we do not see evidence for this explanation (**Figures 3c,d** and **Supplementary Figure**
394 **9**). An alternative explanation involves the distinct origins of *BRAF*-missense acral melanomas
395 compared to other acral melanomas. In our previous work[37], we identified distinct subclasses of
396 human epidermal melanocytes: a common type enriched in limbs (c-type) and a rare type
397 enriched in volar regions (v-type). We observed that most acral melanomas generally retained a
398 transcriptional signature like v-type melanocytes, while a significant subset appeared more akin
399 to c-type melanocytes[37]. The current work indicates that these tumours are more likely to belong
400 to the *BRAF*-activating genetic subtype, suggesting that a subset of volar melanomas might be
401 more accurately classified by cell of origin and/or genetic profile as non-acral CM, rather than
402 bona fide acral melanomas. It is important to clarify that the hypothesis that acral melanomas may
403 arise from different cells of origin is not solely based on this study but is supported by prior work.
404 Our previous research has demonstrated transcriptional diversity among melanocytes in different
405 anatomic locations, including distinct populations of epidermal melanocytes in the palms and
406 soles[37]. Additionally, our zebrafish model studies have shown that AM-associated drivers
407 preferentially (though not exclusively) induce tumours in the limbs (fins), whereas CM-associated
408 *BRAF* mutations preferentially (but not exclusively) lead to tumours in the trunk[38]. Furthermore,
409 we have demonstrated that *BRAF* mutations selectively drive hyperproliferation in only a subset
410 of less-pigmented primary human melanocytes[34]. Therefore, while our additional analyses do not
411 strongly support an oncogene signature as the explanation for the differences in transcriptional
412 scores, thus favouring the cell-of-origin hypothesis, it is possible that in some cases these two
413 phenomena could be intertwined. For example, recent data has shown that some acral
414 melanomas harbouring amplifications of the *CRKL* oncogene depend upon *HOX13* positional
415 identity programs already present in the cell of origin, suggesting that oncogenes and cell-of-origin
416 programs can synergize[38]. Future studies could shed light on this ongoing research area, and
417 could explore the diagnosis of cutaneous melanoma as acral versus non-acral based on
418 molecular signatures rather than solely on anatomic location. The fact that *BRAF*-mutated

419 tumours occur less frequently on patients of non-European ancestry highlights the need to study
420 a diverse set of samples to maximise clinical benefit to all patients. Other observations, such as
421 a tendency for *KIT*-mutated tumours to occur in patients with a higher Amerindian ancestry, are
422 intriguing and will need to be investigated in future studies.
423
424 The observation that any mutation in a driver gene (*BRAF*/*NRAS*/*KIT*/*NF1*) leads to worse
425 prognosis is intriguing. Similar observations, namely that mutation in a MAPK pathway gene leads
426 to worse prognosis, have been made in cutaneous melanoma. For example, Long and
427 collaborators reported in 2011 that untreated patients with $BRAF^{V600E}$ mutations have worse
428 overall survival than patients with wild-type *BRAF* tumours[39], an observation that has been
429 replicated in other studies[40–42], with these observations supporting the use of *BRAF* inhibitors in
430 the clinic[43]. Other mutations in the MAPK pathway also led to shorter overall survival, sometimes
431 in combination with other prognostic factors[44,45]. A previous study also found that mutations in the
432 MAPK pathway predict worse survival in acral melanoma[46]. Regarding *NF1* specifically, previous
433 reports in cutaneous melanoma have suggested that *NF1*-mutated tumours have differing
434 characteristics, sometimes also associated with worse survival and differential response to
435 immune checkpoint inhibition[47,48]. Our results, where we find that patients carrying *NF1* mutations
436 have worse recurrence-free survival than patients with no mutations in this gene, support and
437 expand these previous observations.
438
439 Patients with Cluster 1 tumours also showed a better prognosis than other patients, which is not
440 surprising given their associated clinical characteristics (lower Breslow thickness, a tendency for
441 earlier stages at diagnosis, and lower mitotic indexes). However, what is surprising is the gene
442 expression profile characteristic of this cluster. More CAFs and CD4+ T cells were found by
443 deconvolution to be associated to Cluster 1 than other clusters, signatures commonly associated
444 with immunosuppression. A possible explanation is that early-stage tumours are associated with
445 immunosuppressive microenvironments, a balance which, in later tumours, may have been tilted
446 in favour of tumour cell growth. Another potential explanation may involve the recently described
447 roles of CAFs in immunostimulation[49]. Patients with Cluster 2 tumours, with a
448 'proliferative/pigmentation' signature showed the worst survival, with an overexpression of genes
449 associated with proliferation and pigmentation. It has previously been observed in a zebrafish
450 model and in TCGA samples that a pigmentation signature also predicts worse survival[50], and, in
451 a recent report by Liu and collaborators[25], AM tumours with a proliferative signature also were

452 associated with worse survival than other tumours. This study both extends and replicates these
453 findings in acral melanoma.
454
455 This study has several limitations. Firstly, the nature of our samples (FFPE) may introduce
456 artifacts that affect our ability to identify SNVs and indels accurately. Therefore, to mitigate this
457 risk and increase specificity, we used three different variant calling tools, albeit at the cost of
458 reduced sensitivity. As formalin fixation can generate DNA fragmentation, this may also affect
459 copy number estimation analyses and consequently copy number mutational signature analysis.
460 To mitigate this, we stringently filtered the samples used for this analysis, which affected our
461 statistical power due to a reduced sample size. This study was also done using whole exome
462 data, which limits our ability to call mutations in non-coding regions of the genome.
463
464 The characteristics of the analysed cases reflect in part the challenges of setting up such a study.
465 For instance, the fact that the year of diagnosis preceded the date of recruitment by up to 10 years
466 means that somatic mutations associated with higher mortality rates would be under-represented
467 among those recruited, while survival is likely extended over a similarly sized cohort of
468 prospectively recruited cases. To assess the impact of any biases on our interpretation of the
469 impact of mutations, we restricted attention to the 40 tumours diagnosed after mid-2016, *i.e.,*
470 those closest to the time of recruitment. Analysis of survival gave qualitatively and quantitatively
471 similar results to those reported above; in total, 7 of 23 (30.4%) of cases with mutated tumours
472 had died, as opposed to 0 of 17 among cases with QWT tumours. In the analysis of recurrence,
473 again results matched with 70.0% of cases with mutated tumours recurring (16 of 23 cases)
474 compared to 17.7% of QWT tumours (3 of 17, $P = 0.001$). Results for the RNA clusters were
475 similar to the results quoted above for both survival and recurrence.
476
477 Overall, we were able to identify novel associations of the germline and somatic profile in AM,
478 genomic-clinical correlates of overall and recurrence-free survival, as well as transcriptional
479 differences in *BRAF*-mutated AMs. This study shows the value of studying diverse populations,
480 allowing us to uncover previously unreported relationships and better understand tumour
481 evolution.

482 **Methods**

483 A flowchart describing the analyses, steps, and number of samples used in each individual
484 section can be found in **Supplementary Figure 14**.

485

486 *Patient recruitment and sample collection*

487 The protocol for sample collection was approved by the Mexican National Cancer Institute's
488 (Instituto Nacional de Cancerología, INCan, México) Ethics and Research committees
489 (017/041/PBI;CEI/1209/17) and the United Kingdom's National Health Services (NHS, UK)
490 (18/EE/00076). Patient samples collected for the Utah cohort analysis were derived as previously
491 described[51].

492

493 Recruitment of patients and sample collection took place from 2017 to 2019. Patients attending
494 follow up appointments at INCan that had previously been diagnosed with AM were offered to
495 participate in this study, and upon signing a written consent form, were asked to provide access
496 to a formalin-fixed paraffin-embedded (FFPE) sample of their tumour tissue that had been kept at
497 the INCan tumour bank, as well as a saliva or normal adjacent tissue sample. FFPE samples
498 underwent inspection by a medical pathologist to establish whether sufficient tumour tissue was
499 available for exome sequencing. Saliva samples were collected using the oragene DNA kit
500 (DNAGenotek, # OG-500).

501

502 *DNA and RNA extraction*

503 DNA extraction from all saliva samples was performed at the International Laboratory for Human
504 Genome Research from the National Autonomous University of México (LIIGH-UNAM) using the
505 reagent prepITL2P (DNAGenotek, # PT-L2P) and the AllPrep DNA/RNA/miRNA Universal Kit
506 (Qiagen, #80224). DNA and RNA extraction from FFPE samples was performed at the Wellcome
507 Sanger Institute (UK) using the All-prep DNA/RNA FFPE Qiagen kit. Samples with >0.1ng/μl were
508 sequenced through the Sanger Institute's standard pipeline. Saliva and adjacent tissue samples
509 were used for whole exome sequencing, and only saliva samples were used for genotyping.

510

511 *Genotyping*

512 Genotyping was performed using Illumina's Infinium Multi-Ethnic AMR/AFR-8 v1.0 array at King's
513 College London and Infinium Global Screening Array v3.0 at University College London. Sufficient
514 germline DNA was available for genotyping for 80 out of 92 samples (86.9%). Ancestry estimation
515 was performed using the ADMIXTURE[52] v1.3.0 unsupervised analysis together with the

516 superpopulations of the 1000 Genomes dataset. Five superpopulations were identified,
517 corresponding to AFR (Q1), AMR (Q2), SAS (Q3), EAS (Q4), and EUR (Q5) (**Supplementary**
518 **Table 2**, **Supplementary Figure 1**).
519
520 *Exome sequencing and data quality control*
521 FFPE samples, saliva and normal adjacent tissue underwent whole exome sequencing as follows:
522 Exome capture was performed using Agilent SureSelect AllExon v5 probes and paired-end
523 sequencing was performed at the Wellcome Sanger Institute (UK) in Illumina HiSeq4000
524 machines. Control and tumour samples were sequenced to a mean coverage of 43.72x.
525 Alignment was done using BWA-mem[53] v0.7.17, using the GRCh38 reference genome.
526 Sequencing quality filters were performed using samtools stats[54] and fastqc[55]. Sample
527 contamination was estimated using the GATK v4.2.3.0 tool CalculateContamination[56].
528 Concordance between sample pairs was estimated using the Conpair v0.2 tool[57]. Samples that
529 had <90% similarity with their pair (tumour-normal) or showed a level of contamination above 5%
530 were excluded from the study. After this step, 123 samples remained for further analysis.
531
532 *Somatic SNV calling and identification of driver genes and mutations*
533 Somatic variant calling was done using three different tools (CaVEMan[58] v1.15.2, Mutect2[59]
534 v4.1.0.0 and Varscan2[60] v2.3.9), keeping only the variants identified by a minimum of two out of
535 the three tools. For *BRAF*$^{V600E}$ mutations, we kept these variants even if it was only identified by
536 one of the tools as its oncogenicity and relevance in melanoma is well known. When available
537 within the variant calling tool, strand bias filters were applied. A minimum base quality score of 30
538 on the Phred scale was used. Indel calling was performed using pindel[61,62] v1.15.2 using indel
539 candidates identified by the structural variant caller manta[63]. When selecting one sample per
540 patient, preference was given to primaries, and metastases or recurrences were chosen only
541 when a primary had not been collected.
542
543 Significantly mutated genes were identified using the tool dNdScv[64] v0.0.1.0 with default
544 parameters using SNVs identified by two of the three tools used for variant calling and indels
545 identified by pindel as input data. Positive selection was considered for genes that had global q-
546 values below 0.1 according to the dNdScv tool recommendations.
547

548 Two lymph node metastasis samples (one from a patient that had a *BRAF*$^{V600E}$ mutation and
549 another one with an *NRAS* mutation) and their primaries were annotated as having the same
550 mutation for follow up analysis after manual inspection using IGV[65].
551
552 *Analysis of correlation between driver mutations and clinical covariates and ancestry*
553 Statistical tests were performed to identify potential clinical and ancestry covariates that correlated
554 with driver mutational status. For tumour stage, sex, ulceration status and tumour site, which are
555 discrete variables, association was tested with contingency Chi-squared tests.
556 For each of the four driver genes, a logistic regression model was fitted to predict the presence
557 or absence of a mutation on the AM samples using the inferred ADMIXTURE[52] cluster related to
558 the European ancestry component from the 1000 Genomes Project, correcting for age, sex, and
559 tumour mutational burden (TMB, SNVs + indels), as such: Driver gene status ~ EUR related
560 cluster proportion + age + sex + TMB. Then the log odds related to the EUR cluster were plotted
561 with their respective confidence intervals. The models were constructed using 80 samples out of
562 the 92, which were those with available genotyping information and with all tested covariables
563 available.
564
565 *Somatic DNA copy number calling*
566 Sequenza[66] v3.0.0 and ASCAT[67] v3.1.2 were used to estimate ploidy and purity values for each
567 sample. These values per sample were compared between the two tools, and samples that had
568 a high discrepancy in their purity estimates (<0.15 vs 1, respectively) were filtered out
569 (**Supplementary Table 28**). Samples with an estimated goodness of fit below 95 were also
570 filtered out. Subsequently, copy number, cellularity and ploidy values estimated by ASCAT were
571 used in follow-up analyses. Whole genome duplication events were assigned as reported by
572 ASCAT. Significantly affected regions by CNAs were identified using GISTIC2[68] v2.0.23.
573 Amplifications were classified as low-level amplifications when regions had a copy number gain
574 above 0.25 and below 0.9, and as high-level amplifications when regions had a copy number gain
575 above 0.9 according to GISTIC2 values; deletions were classified as low-level deletions when
576 regions had a copy number change between -0.25 and -1.3 and as high-level deletions when
577 regions had a copy number change lower than -1.3. Only peaks with residual q-values < 0.1 were
578 considered as significantly altered. For the analyses of differences in CNA burden by sample
579 group (*e.g.*, mutational status or site of presentation), we used the CNApp tool[69] to generate copy
580 number alteration scores for global (GCS), focal (FCS) and broad (BCS) CNA burden, calculating
581 segment means (seg.mean) as log2(cn/ploidy) and using default parameters. GCS (Global copy

582 number alteration score) is a number quantifying the copy number aberration level in each sample
583 provided by the CNApp tool[69]. Higher GCS scores indicate a higher burden of copy number
584 aberrations compared with all other samples in the cohort. GCS is the sum of the normalised BCS
585 (Broad copy number alteration score) and FCS (focal copy number alteration score), which are
586 calculated considering broad (chromosome and arm-level) and focal (weighted focal CNAs
587 corrected by the amplitude and length of the segment) aberrations per sample. These values are
588 calculated using as input the number of DNA copies normalised by sample ploidy. A more detailed
589 explanation can be found in the original publication[69]. GCS values were used for comparisons
590 between sample groups. All paired comparisons between groups were evaluated with a Mann-
591 Whitney test.
592
593 *Mutational signature analysis*
594 Mutational matrices were generated using SigProfilerMatrixGenerator[70] v1.2.20. These matrices,
595 with single nucleotide mutations found by at least two of the three variant callers and all insertions
596 and deletions identified by pindel, were used as input for mutational signature extraction using
597 SigProfilerExtractor[71] v1.1.23 and decomposition to COSMICv3.4 reference mutational
598 signatures[72]. For single base substitutions, the standard SBS-96 mutational context was used,
599 with default parameters and a minimum and maximum number of output signatures being set as
600 1 and 5, respectively. A total of 116 samples with an SNV count > 0, were used for this analysis.
601 For copy number signature analysis, all 60 samples with available copy number data were used
602 with default parameters, and using the standard CN-48 context from COSMICv3.4.
603
604 *RNA sequencing and data quality control*
605 Total RNA library preparation followed by exome capture using Agilent SureSelect AllExon v5
606 was performed on Illumina HiSeq 4000 machines on 146 samples. Reads were aligned to the
607 GRCh38 reference genome using the splice-aware aligner STAR[73] v2.5.0. Of these, we focused
608 on the 77 samples that came from different patients, that had matching DNA and were primaries
609 for the score analysis (Methods below). We then applied further quality control filters for the
610 consensus clustering analysis: samples were excluded if total read counts were fewer than 25
611 million, or if the sum of ambiguous reads and no feature counts was greater than the sum of all
612 gene read pair counts. Forty-four samples remained for downstream analysis. Counts were
613 generated with HTSeq[74] v0.7.2. Transcripts per million (TPM) normalisation was performed and
614 values were $\log_2$(TPM+1) transformed.
615

616 *Acral vs. non-acral cutaneous tumour score*

617 Patient samples collected for the Utah cohort analysis were derived as previously described[51].
618 Invasive acral and non-acral cutaneous melanomas were identified and collected as part of the
619 University of Utah IRB umbrella protocol #76927, Project #60, and RNA was extracted and
620 quantified as previously described[51]. A custom NanoString nCounter XT CodeSet (NanoString
621 Technologies) was designed to include genes differentially expressed between glabrous and non-
622 glabrous melanocytes[37,38]. Sample hybridization and processing were performed in the Molecular
623 Diagnostics core facility at Huntsman Cancer Institute. Data were collected using the nCounter
624 Digital Analyzer. Raw NanoString counts were normalised using the nSolver Analysis Software
625 (NanoString Technologies). Normalisation was carried out using the geometric mean of
626 housekeeping genes included in the panel (**Supplementary Table 16**). Background thresholding
627 was performed using a threshold count value of 20. Fold change estimation was calculated by
628 partitioning by acral vs. cutaneous melanoma. The log2 normalised gene expression data were
629 subjected to Principal Component Analysis (PCA) using the PCA function in Prism version 10.2.1
630 (GraphPad Software). PCA was performed to identify the main sources of variability in the data
631 and to distinguish between acral and cutaneous samples.
632
633 To determine the top differentially expressed genes contributing to the variance between acral
634 melanomas and cutaneous melanomas, the loadings of the second principal component (PC2)
635 were examined. Genes with the highest positive and negative loadings on PC2 were selected as
636 the top 10 and bottom 10 genes, respectively. Log2 expression values of these genes were used
637 to generate a multiplicative score, producing the ratio of acral to cutaneous melanocyte genes.
638 Statistical analyses were performed using Prism version 10.2.1 (GraphPad Software). Differences
639 in acral to cutaneous ratios were assessed using the Mann-Whitney U test.
640
641 The acral:cutaneous (A:C) ratio was calculated for each of the 77 primary acral tumours using the
642 method described above after batch correction (limma v3.64.1) on normalised and transformed
643 expression data processed by the R package DESeq2 v1.48.1. Differences in the A:C gene
644 expression ratio scores between *BRAF* activating mutation-positive and *BRAF*-wildtype acral
645 melanoma samples were assessed using a Mann-Whitney U test. The same normalisation,
646 scoring method and statistical testing was applied to the 63 transcriptomes from acral melanoma
647 tumours considering *BRAF*-activating (n=10) and wild-type (n=53) in Newell *et al* (2020)[7]. All
648 available samples in this cohort were used, as only one primary had a *BRAF* mutation. Only

649 samples with *BRAF*-activating mutations (V600E and L597R for the Mexican acral melanoma set)
650 were included in the *BRAF* group.
651
652 To determine whether the CM classifier genes are induced by *BRAF*$^{V600E}$ signalling in
653 melanocytes, we analysed RNA-seq data from McNeal *et al*[34], which consisted on bulk-RNA
654 sequencing of primary human melanocytes transduced with *BRAF*$^{V600E}$ and cultured under two
655 conditions: PMA (phorbol 12-myristate 13-acetate) and ET1 (endothelin-1). We extracted
656 normalised expression values for CM classifier genes across four conditions: PMA, PMA +
657 *BRAF*$^{V600E}$, ET1, and ET1 + *BRAF*$^{V600E}$. Normalised expression levels were compared using the
658 Mann-Whitney U test in Prism version 10.2.1 (GraphPad Software).
659
660 We evaluated the A:C classifier in clinical melanoma samples using RNA-seq data from The
661 Cancer Genome Atlas (TCGA) Skin Cutaneous Melanoma Firehose Legacy cohort. Normalised
662 gene expression data were downloaded from cBioPortal. Samples were classified as *BRAF*-
663 activating or *BRAF*-wildtype in the same way as for Figure 3C. We calculated the product of the
664 expression of CM classifier genes for each category. Differences were assessed using the Mann-
665 Whitney U test.
666
667 We used an interactive Shiny application, What Is My Melanocytic Signature (WIMMS;
668 https://wimms.tanlab.org) to compare transcriptional programs associated with distinct
669 melanocytic cell states. WIMMS classifies melanocytic gene expression profiles by aggregating
670 previously published gene expression signatures and clusters them into seven principal cell state
671 categories. We input our gene signature into WIMMS to assess correlation with these reference
672 states. The resulting dendrogram represents the relationship between our classifier-derived CM
673 (cutaneous melanoma) genes and known signatures.
674
675 *Consensus clustering and deconvolution based on gene expression*
676 To identify molecular subgroups based on transcriptome data, we performed consensus
677 clustering using the Cola R v2.10.1 package[75]. Standard preprocessing of the input matrix was
678 performed, including removal of rows in which >25% of the samples had NA values, imputation
679 of missing values, replacement of values higher than the 95th percentile or less than 5th by
680 corresponding percentiles, removal of rows with zero variance, and removal of rows with variance
681 less than the 5th percentile of all row variances. Subsequently, standard statistical metrics were
682 used to assess the number of clusters and the stability of the partitions, including 1-PAC score,

683 concordance and jaccard index, and visual inspection of the consensus matrix through heatmaps
684 visualisations. Afterwards, signature analysis and functional enrichment on the identified clusters
685 were performed.
686
687 The EPIC algorithm[76] v1.1.7 was used in the R programming environment to perform
688 deconvolution in order to infer immune and stromal cell fractions within AM tumours. We used the
689 TRef signature method with default parameters, which includes gene expression reference
690 profiles from tumour-infiltrating cells. The algorithm generated an absolute score that could be
691 interpreted as a cell fraction.
692
693
694 *Survival analyses*
695 Consenting and recruitment of patients started in 2017 and ended in 2019. Because of the
696 challenges of recruiting significant numbers of participants with AM, patients diagnosed in earlier
697 years who were still attending follow-up clinics for primary or recurrent disease were recruited.
698 To ensue data consistency, only participants with a primary available for analysis were the subject
699 of focus on analyses of recurrence and/or death (n=85 patients). In total, 73 participants were
700 recruited whose primary and ancestry data were available for analysis on driver mutations. A total
701 of 44 patients with primary and RNA cluster data available were used for analysis on clusters
702 regardless of their ancestry data availability.
703
704 To compare the effect of distinct driver mutations and the RNA clusters, we examined 2 measures
705 of disease severity (i) the recurrence of the primary and (ii) overall survival. For recurrence, we
706 examined time to recurrence using a life-table approach from date of primary diagnosis onwards
707 as a descriptor, and based analyses of differences between mutations (and clusters) on logistical
708 regression adjusting for date of diagnosis, age at diagnosis, sex, stage at diagnosis
709 (advanced/early), ancestry (only for mutations) and time from diagnosis through either death or
710 last known to be alive. Primary tumours were either classified as QWT or mutated for known
711 drivers; tumours with multiple driver mutations were classified as “multihit”. We also conducted
712 analyses based on a binary exposure of “QWT” or “Mutated” tumour based on the existence of
713 one or more mutations in a known driver gene.
714
715 For survival analysis, we also included a life-table approach, again as a descriptor, based on time
716 from diagnosis through to date of death or date last known to be alive. Statistical assessment of

717 the effect of each mutation and/or cluster were based on Cox proportional hazard analysis with
718 followup starting from date of recruitment through to date of death (or last date alive) adjusting for
719 date of diagnosis, age at diagnosis, sex and ancestry (as a sensitivity analysis). Analysis of
720 combined mutations were also conducted as for the recurrence analysis.
721
722

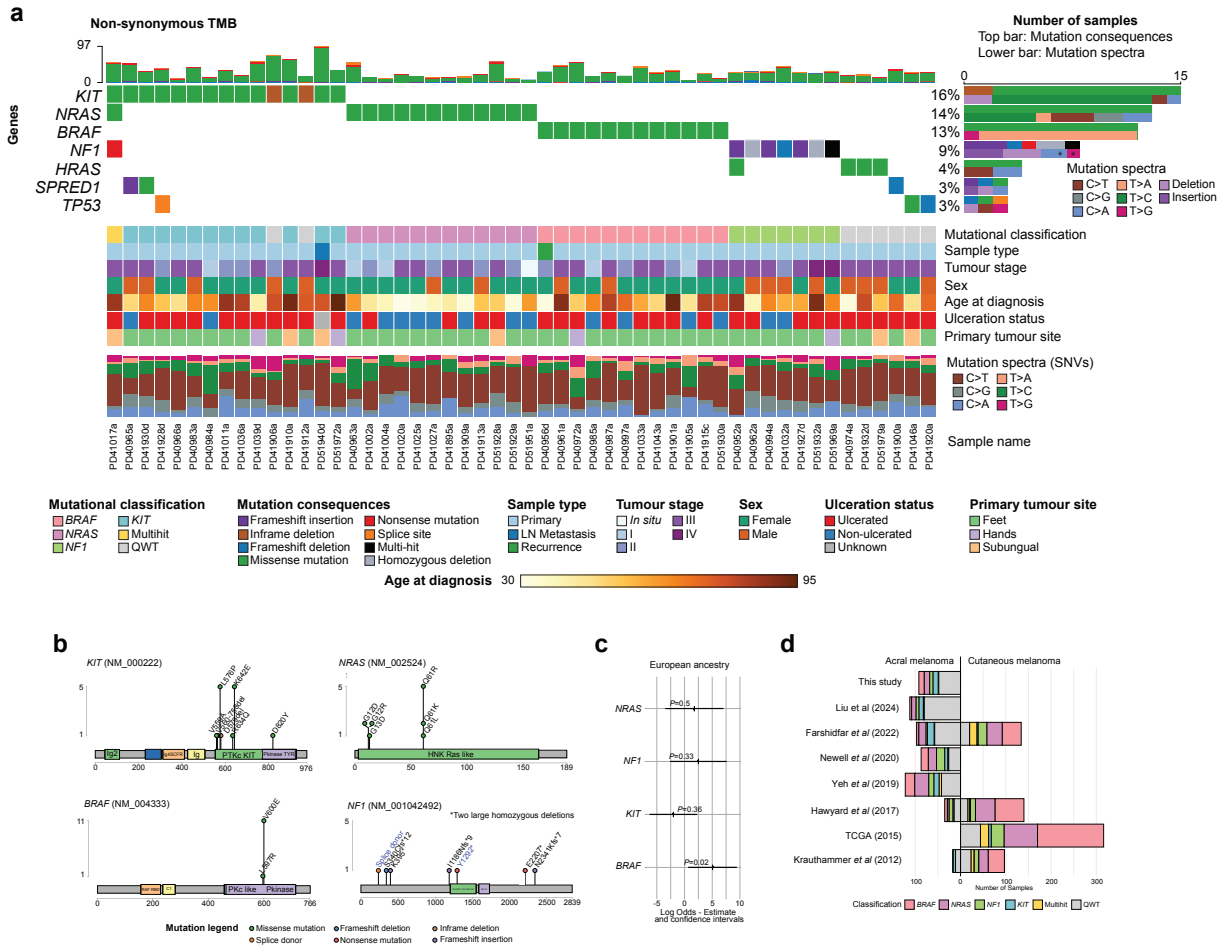



**Figure 1. Somatic landscape of acral melanoma in Mexican patients.** A) Oncoplot depicting the 7 most mutated genes according to dNdScv and their status in the samples with mutations in these (52 samples out of 92, one per patient). Mutational classification, sample type, tumour stage, sex, age at diagnosis, ulceration status, tumour site and mutational spectra are shown by sample. B) Mutations found in *KIT*, *NRAS*, *BRAF* and *NF1*, which are the most significantly mutated genes. C) A logistic regression model controlling for age, sex, and TMB was fitted to predict the presence or absence of a mutation on the AM samples using the inferred ADMIXTURE cluster related to the European ancestry component. Log odds estimate and confidence intervals are depicted for the four driver genes. D) Barplot depicting the number and mutational classification of samples in different acral and cutaneous melanoma studies.

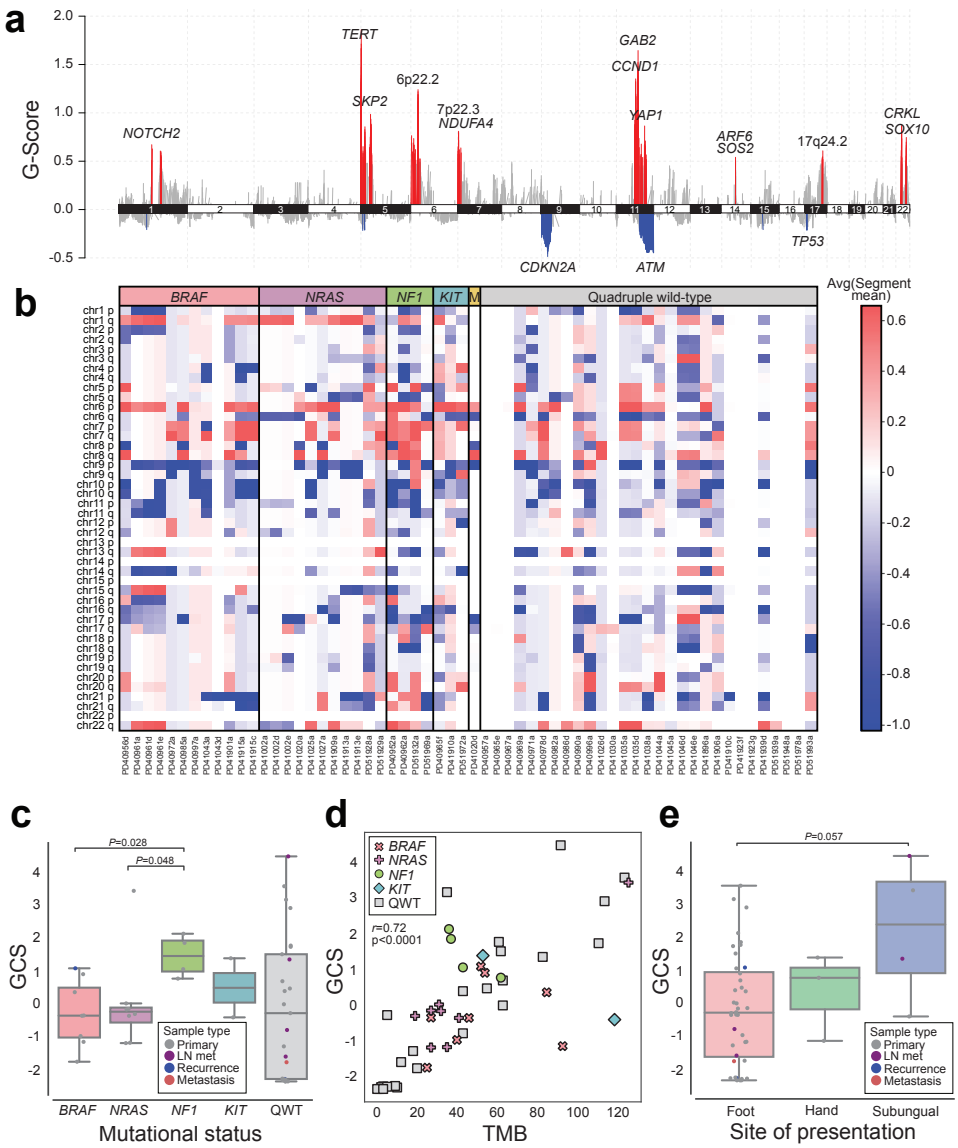
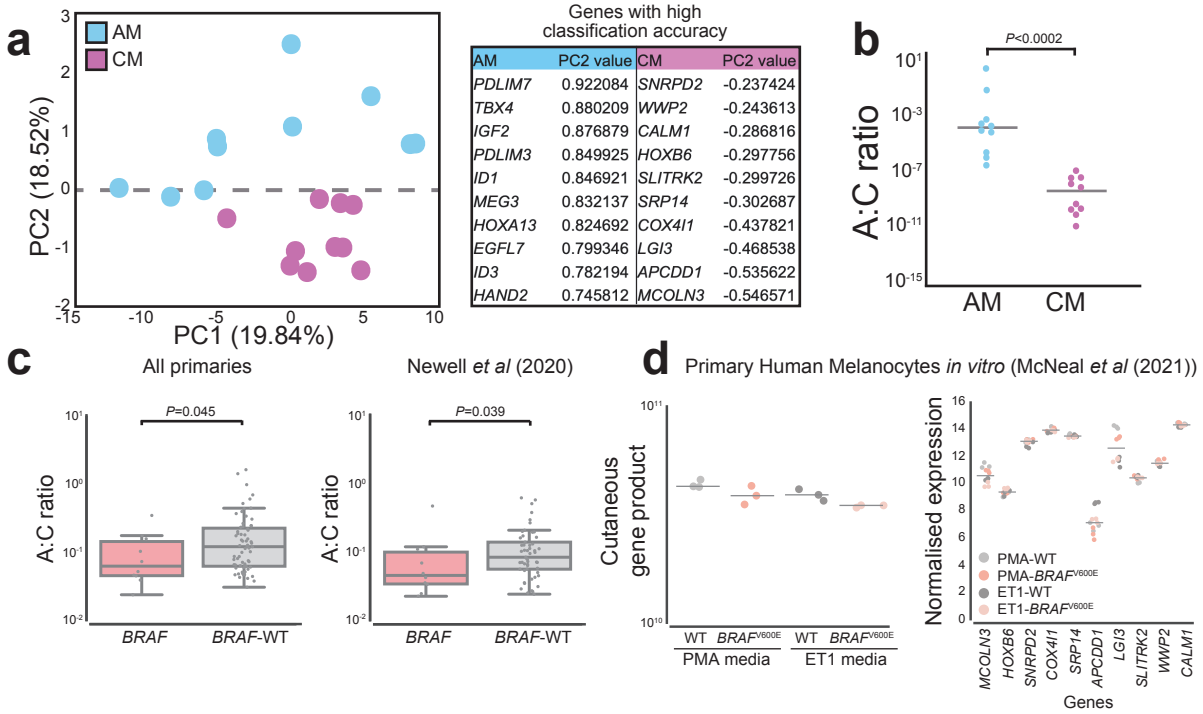



**Figure 2. DNA copy number landscape of acral melanoma and molecular and clinical correlates in Mexican patients.** A) Regions of amplification (red) and deletion (blue) in 47 acral melanoma samples, one per patient, as identified by GISTIC2. Known drivers, or the chromosomal regions, are shown. This analysis shows alteration with respect to the normal sample, *i.e.*, with respect to a ploidy of 2. B) Heatmap showing regions of amplification (red) and deletion (blue) by sample and chromosomal arm in all 60 samples classified into genomic subgroups. This analysis shows alterations with respect to the estimated tumour sample ploidy. C) Boxplot of global copy-number scores (GCS) of 47 samples, one per patient, classified by genomic subgroup. *P*-values are from Wilcoxon-Mann-Whitney paired tests. D) Scatter plot of TMB (X axis) and GCS (Y axis) for 47 samples, one per patient. Dots represent samples, coloured by genomic subtype. Pearson's product-moment correlation coefficient and associated *P*-value is shown. E) Boxplot of GCS of 47 samples, one per patient, classified by tumour site. GCS scores are calculated with respect to tumour sample ploidy. For box plots, the central line within each box represents the median value, the box boundaries represent the interquartile range (IQR), and the whiskers extend to the lowest or highest data point still within 1.5xIQR.




| AM | PC2 value | CM | PC2 value |
|---|---|---|---|
| *PDLIM7* | 0.922084 | *SNRPD2* | -0.237424 |
| *TBX4* | 0.880209 | *WWP2* | -0.243613 |
| *IGF2* | 0.876879 | *CALM1* | -0.286816 |
| *PDLIM3* | 0.849925 | *HOXB6* | -0.297756 |
| *ID1* | 0.846921 | *SLITRK2* | -0.299726 |
| *MEG3* | 0.832137 | *SRP14* | -0.302687 |
| *HOXA13* | 0.824692 | *COX4I1* | -0.437821 |
| *EGFL7* | 0.799346 | *LGI3* | -0.468538 |
| *ID3* | 0.782194 | *APCDD1* | -0.535622 |
| *HAND2* | 0.745812 | *MCOLN3* | -0.546571 |

**Figure 3. Comparisons of the transcriptional profile of *BRAF*-activating and *BRAF*-wildtype AM tumours.** A) Elucidation of genes used to classify acral vs cutaneous melanoma samples. PCA of acral melanoma (blue) and cutaneous melanoma (purple) samples (left panel). Loadings on PC2 were used to identify the top differentially expressed genes contributing to the variance between acral melanomas and cutaneous melanomas (right panel). B) Scatter plot showing the distribution of the acral:cutaneous (A:C) gene expression ratios between test acral and cutaneous melanoma samples. AM samples are represented by blue dots, and CM samples are represented by purple dots ($P$-value < 0.0002). c) Left panel. Comparison of A:C gene expression ratio in AM samples with different mutation status. Box and whiskers plot comparing two groups: *BRAF*-WT and *BRAF*-activating mutated tumours. Right panel. Comparison of A:C gene expression ratio in AM samples with *BRAF*-activating mutations and *BRAF*-WT tumours from Newell *et al* (2020)[7]. The central line within each box represents the median value, the box boundaries represent the interquartile range (IQR), and the whiskers extend to the lowest or highest data point still within 1.5xIQR. Individual data points are plotted as dots. Statistical significance was assessed using individual Wilcoxon Mann-Whitney tests. D) Left panel. Comparison of the product of the cutaneous genes in normal human melanocytes. Melanocytes were cultured in PMA or ETA containing media with or without doxycycline-induced *BRAF*$^{V600E}$ expression. Each dot is an individual biological replicate with horizontal lines indicating median values. Right panel. Relative expression levels of cutaneous genes across individual normal human melanocytes. Melanocytes were cultured in PMA or ETA containing media with or without doxycycline-induced *BRAF*$^{V600E}$ expression. Each point represents a biological replicate with horizontal lines indicating median values.

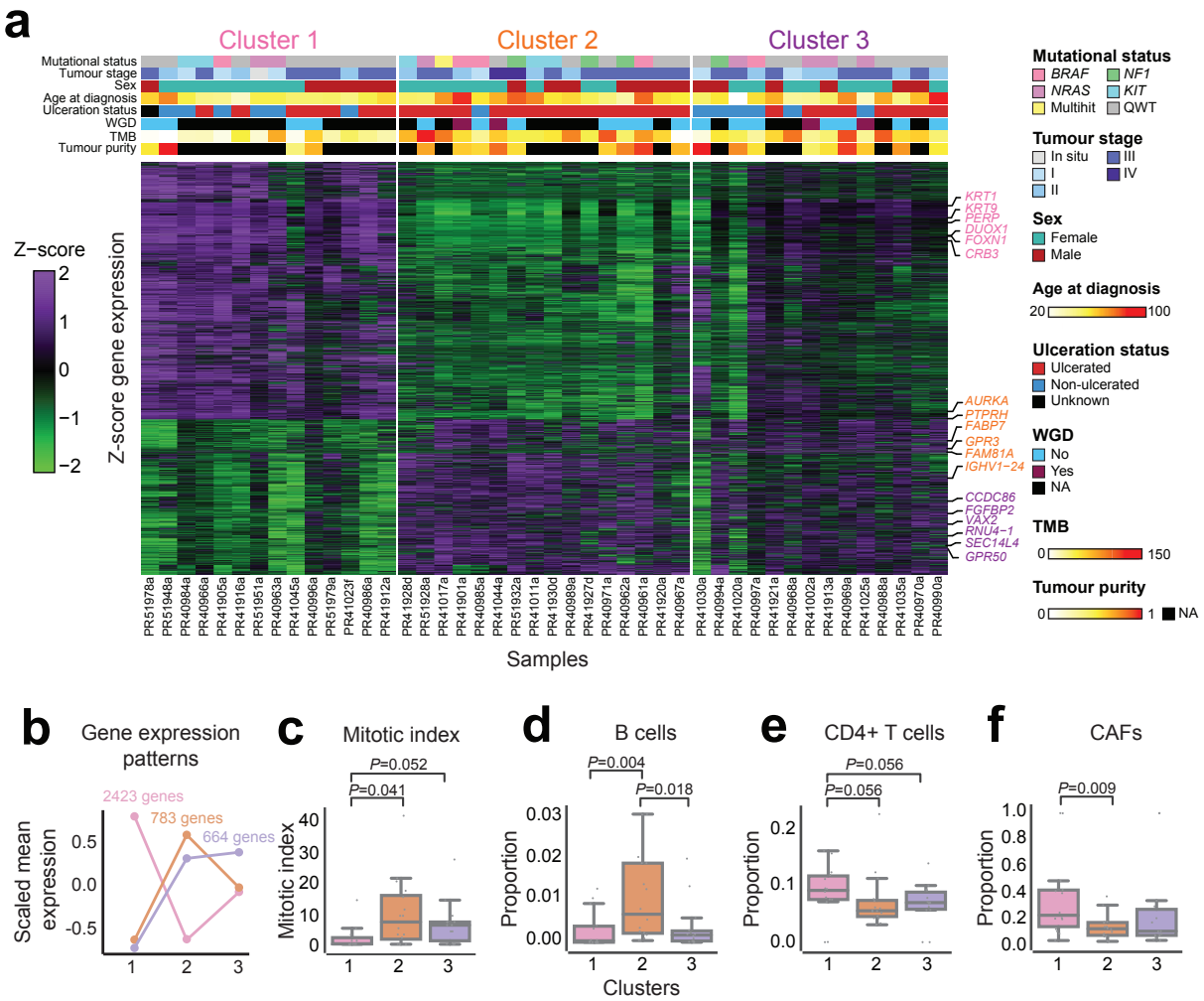



**Figure 4. Unsupervised gene expression clustering of primary acral melanoma samples from Mexican patients identifies three main groups.** A) Gene expression heatmap showing the 3,870 genes identified as differentially expressed among sample clusters. Samples are in the X axis and genes are in the Y axis. Mutational status and clinical covariates by sample are shown above the heatmap. B) Scaled mean expression patterns per cluster for the three 'gene programs' defining each cluster. C) Box plot of mitotic index (Y axis) per sample classified by transcriptional cluster. C) Box plot of B cell proportion (Y axis), as calculated by deconvolution, per sample classified by transcriptional cluster. D) Box plot of CD4+ T cell proportion (Y axis), as calculated by deconvolution, per sample classified by transcriptional cluster. E) Box plot of cancer-associated fibroblasts (CAFs, Y axis), as calculated by deconvolution, per sample classified by transcriptional cluster. The central line within each box represents the median value, the box boundaries represent the interquartile range (IQR), and the whiskers extend to the lowest or highest data point still within 1.5xIQR. Individual data points are plotted as dots. Wilcoxon-Mann-Whitney paired tests were performed.

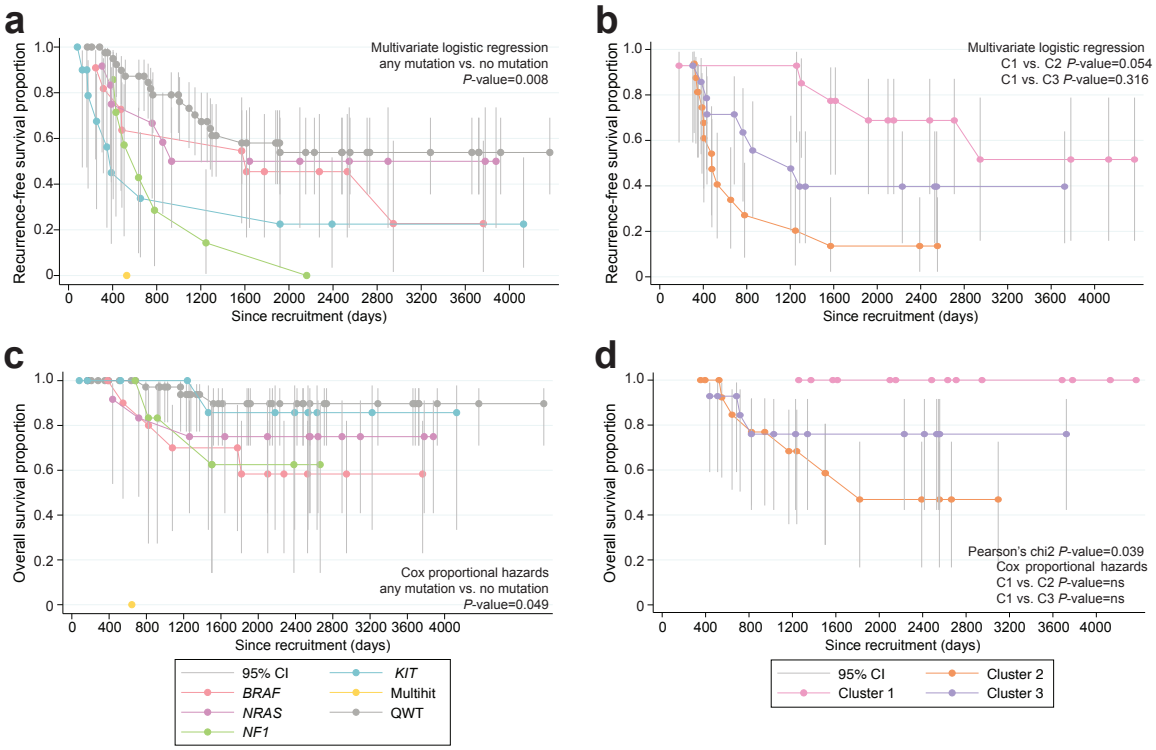
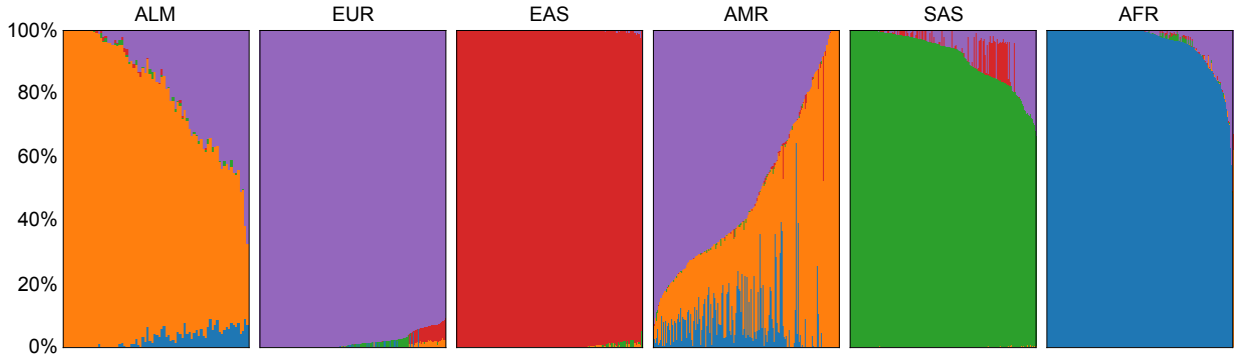
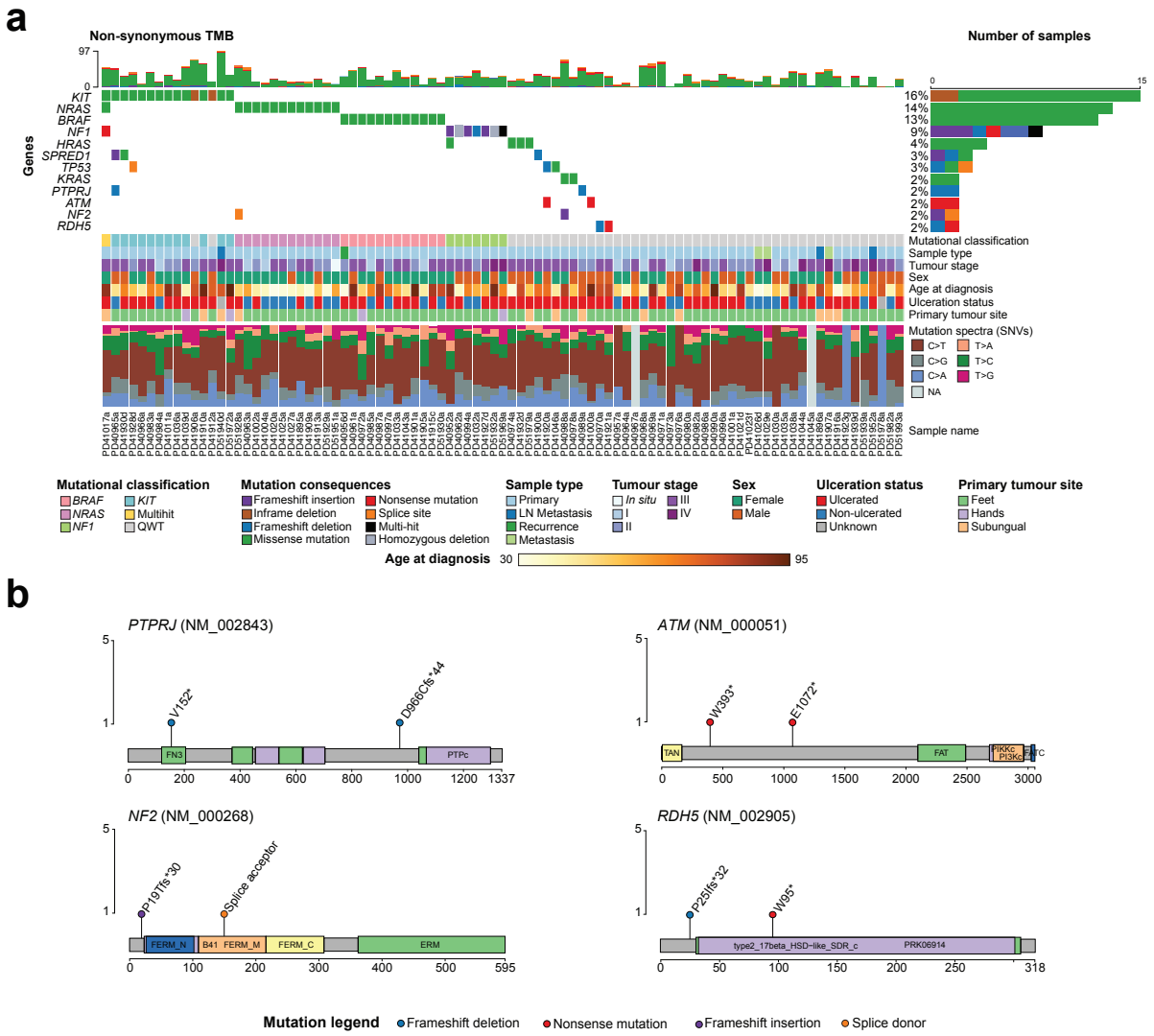


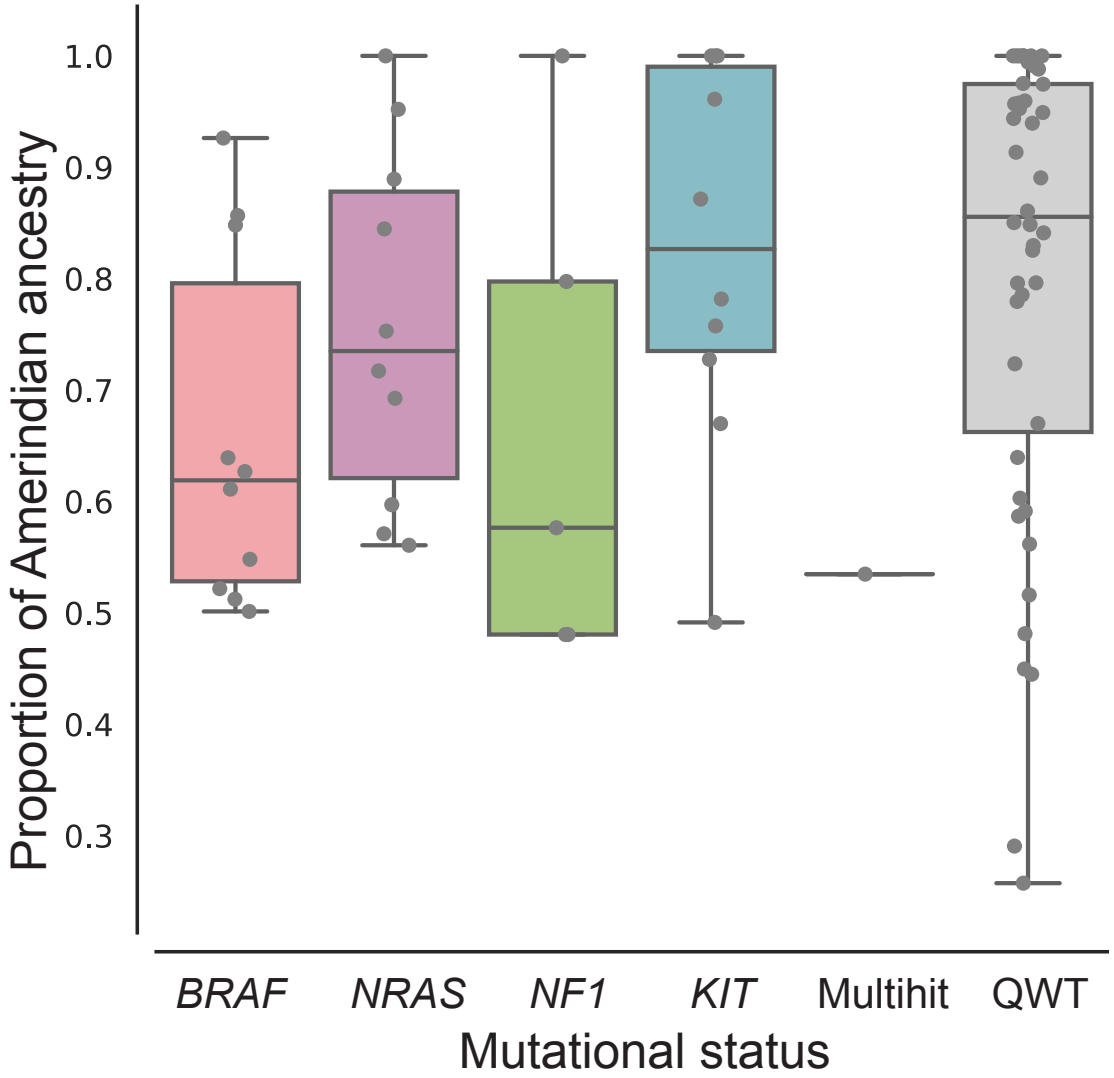

**Figure 5. Kaplan-Meier plots of overall and recurrence-free survival for patients by tumour mutational and transcriptional status**. A) Recurrence-free survival of patients with and without driver mutations, depicted by each category of the mutational classification. B) Recurrence-free survival for patients with tumours in each of the three transcriptional clusters. C) Overall survival of patients with and without driver mutations, depicted by each category of the mutational classification. D) Overall survival for patients with tumours in each of the three transcriptional clusters. Each panel indicates the test that the depicted *P*-values comes from.



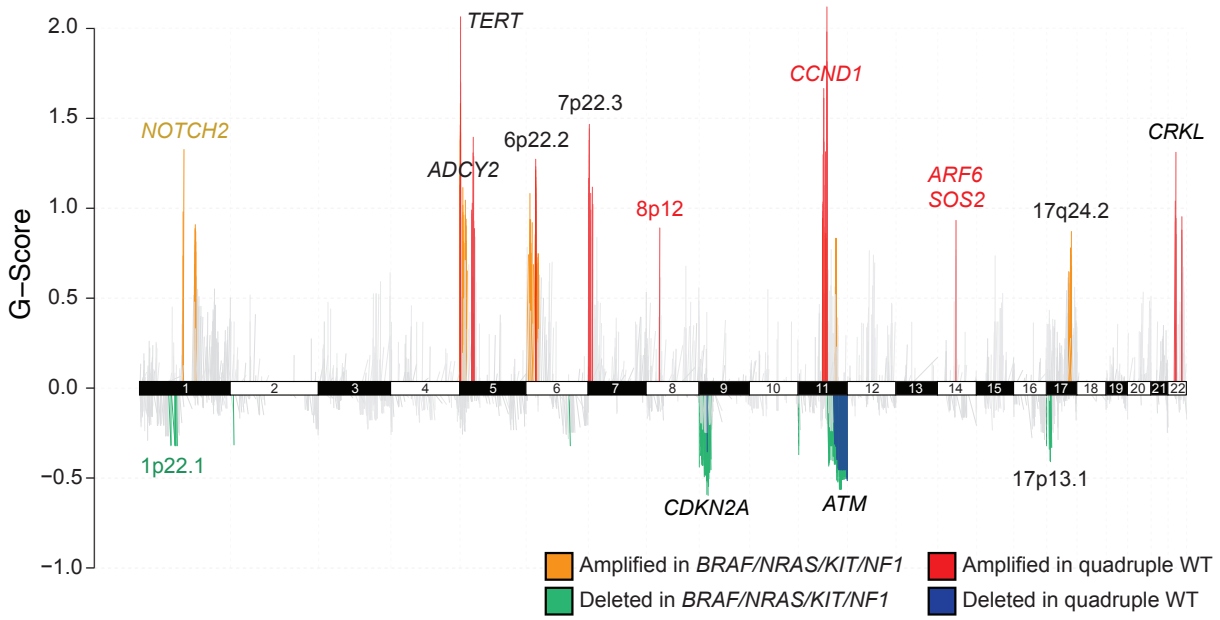

**Supplementary Figure 1. Estimation of ancestry proportions per sample together with the superpopulations of the 1000 Genomes dataset.** The leftmost panel corresponds to the samples genotyped in this study (n=80). The following panels correspond to the superpopulations in the 1000 Genomes Project. Five superpopulations are plotted, corresponding to African (AFR, blue), American (AMR, orange), South Asian (SAS, green), East Asian (EAS, red), and European (EUR, purple).



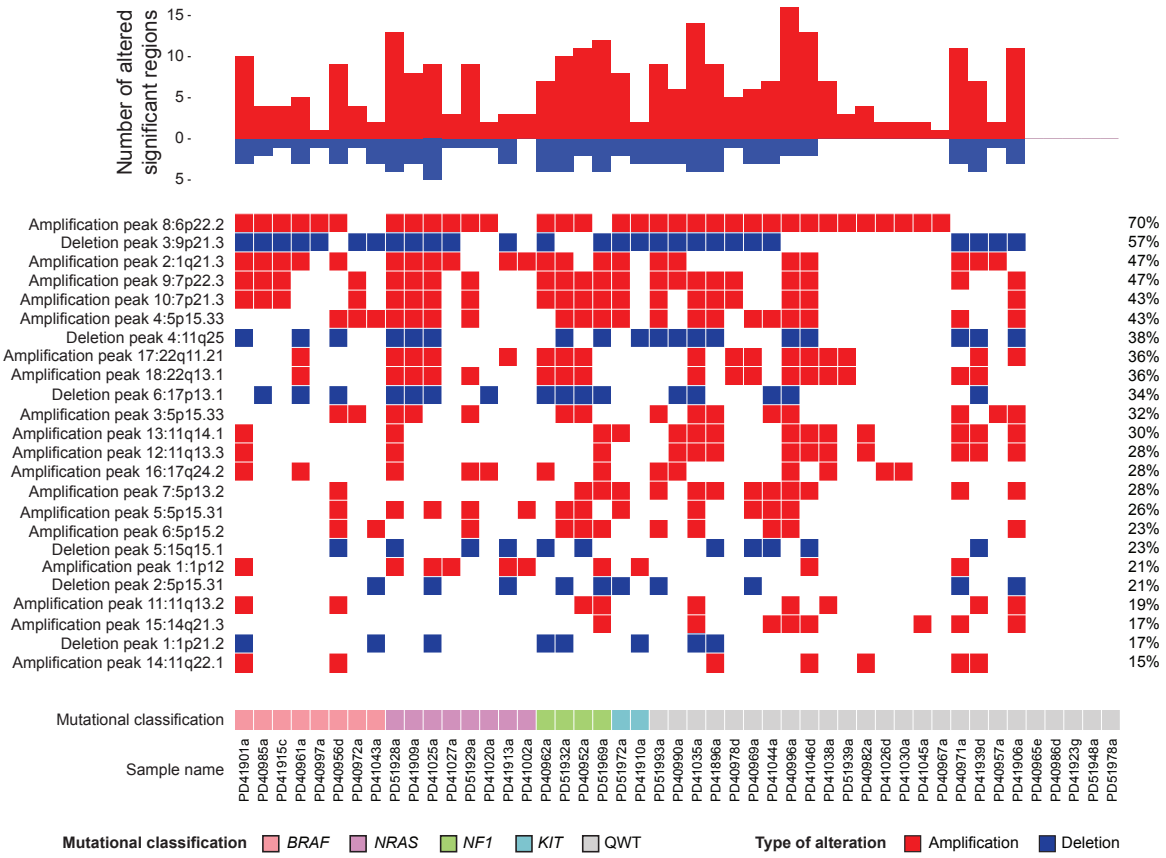

**Supplementary Figure 2. Somatic landscape of all acral melanoma samples.** a) Oncoplot depicting the seven most mutated genes according to dNdScv and five selected genes based on mutational frequency and biological function. Mutational classification, sample type, tumour stage, sex, age at diagnosis, ulceration status, tumour site and mutational spectra are shown by sample. b) Mutations found in *PTPRJ*, *ATM, NF2* and *RDH5*, for which all mutations are deleterious and are found altered each in two samples.



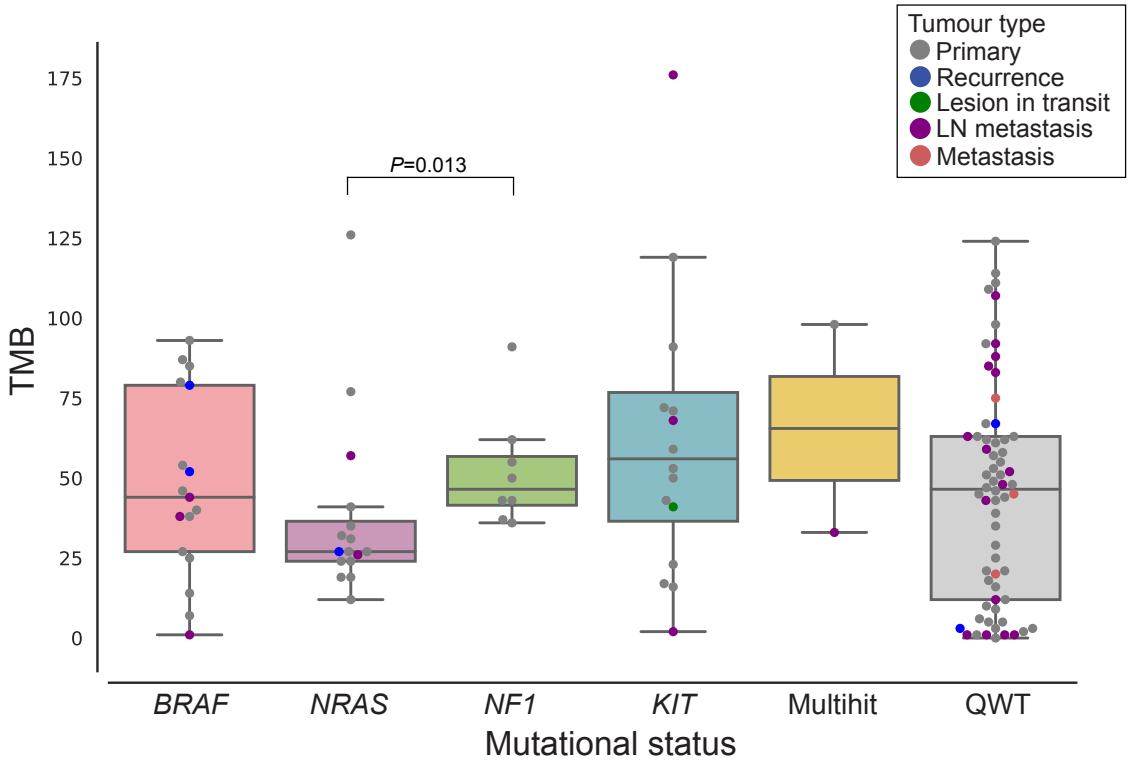

**Supplementary Figure 3. Boxplot of the proportion of Amerindian ancestry among patients classified by genomic subtype.** Each dot corresponds to a sample. The central line within each box represents the median value, the box boundaries represent the interquartile range (IQR), and the whiskers extend to the lowest or highest data point still within 1.5xIQR.



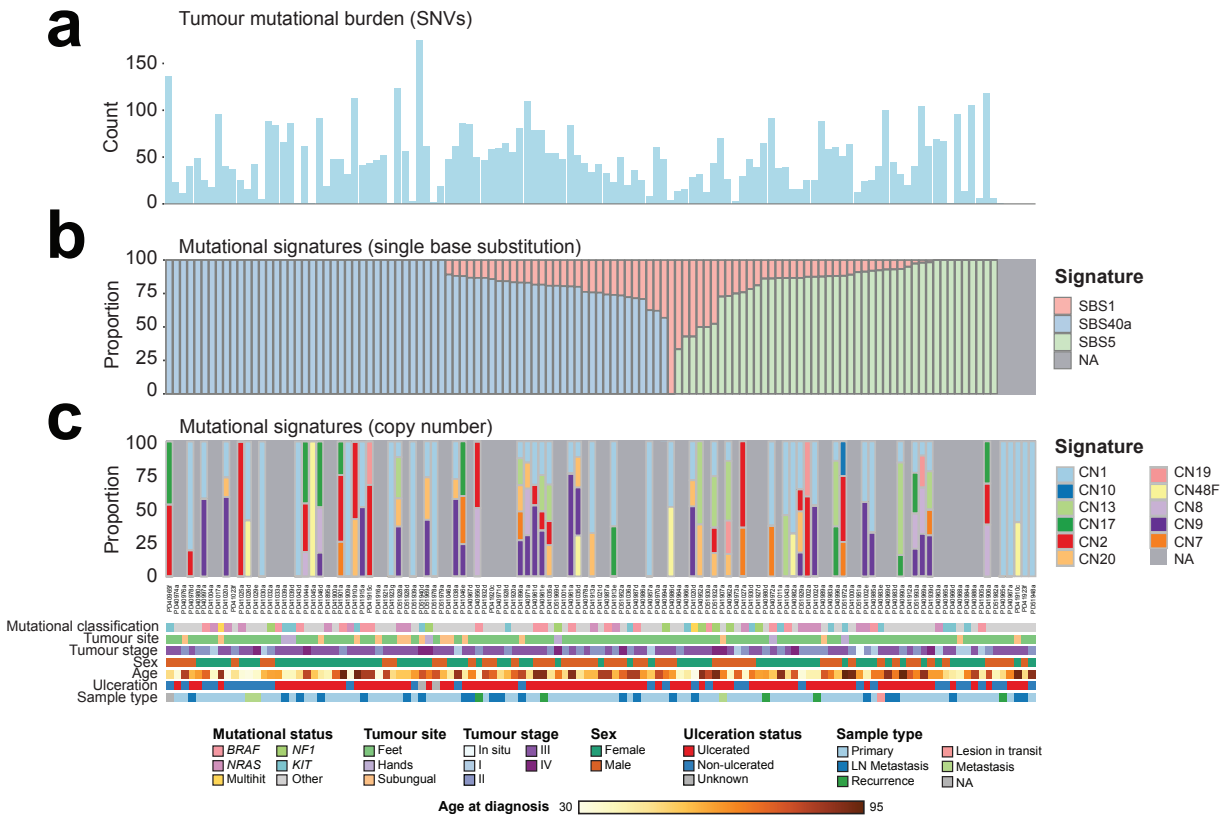

**Supplementary Figure 4. Copy number profile of *BRAF/NRAS/KIT/NF1*-mutated tumours vs. quadruple WT.** All depicted regions have been identified by GISTIC2 analysis per group of samples. Statistically significant differences are marked in colour, yellow = amplified in mutated tumours, red = amplified in QWT, green = deleted in mutated tumours, blue = deleted in QWT tumours. Differences are determined first by assessing the global GISTIC2 output and determining differences between groups by one-sided Fisher's exact test ($P < 0.05$). If a region is not found in the global GISTIC2 output, but it is found only in the analysis per group, we have indicated it as statistically different. Number of mutated tumours = 22, number of QWT tumours = 25.



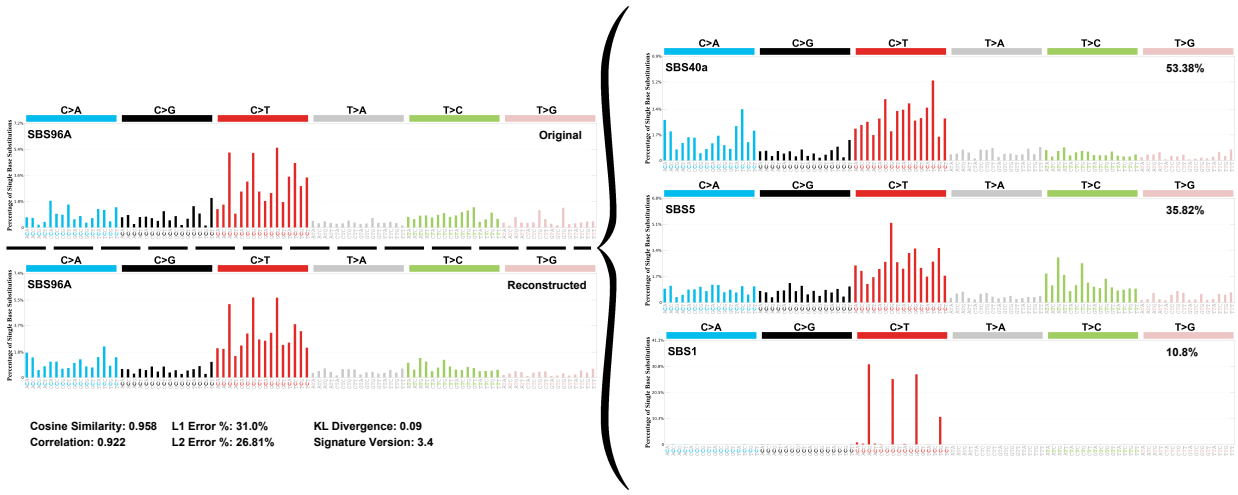

**Supplementary Figure 5**. **Significantly altered regions per sample**. Top panel. Number of significant regions altered by sample. Bottom panel. Binary heatmap showing the significantly altered regions identified by GISTIC2 per sample. One sample per patient is shown. Heatmap is ordered on the X axis by mutational classification and on the Y axis by frequency of alterations per region.



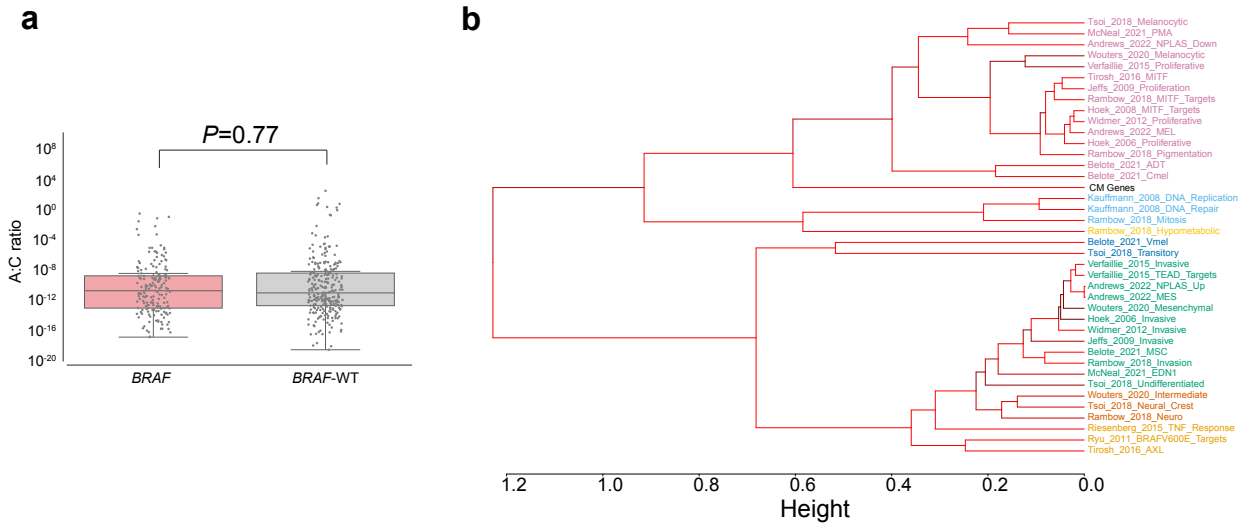

**Supplementary Figure 6. Boxplot of TMB for all samples classified by genomic subtype.** Each dot corresponds to a sample, and colours represent tumour type. The central line within each box represents the median value, the box boundaries represent the interquartile range (IQR), and the whiskers extend to the lowest or highest data point still within 1.5xIQR. *P*-values shown are from Wilcoxon Mann-Whitney tests.



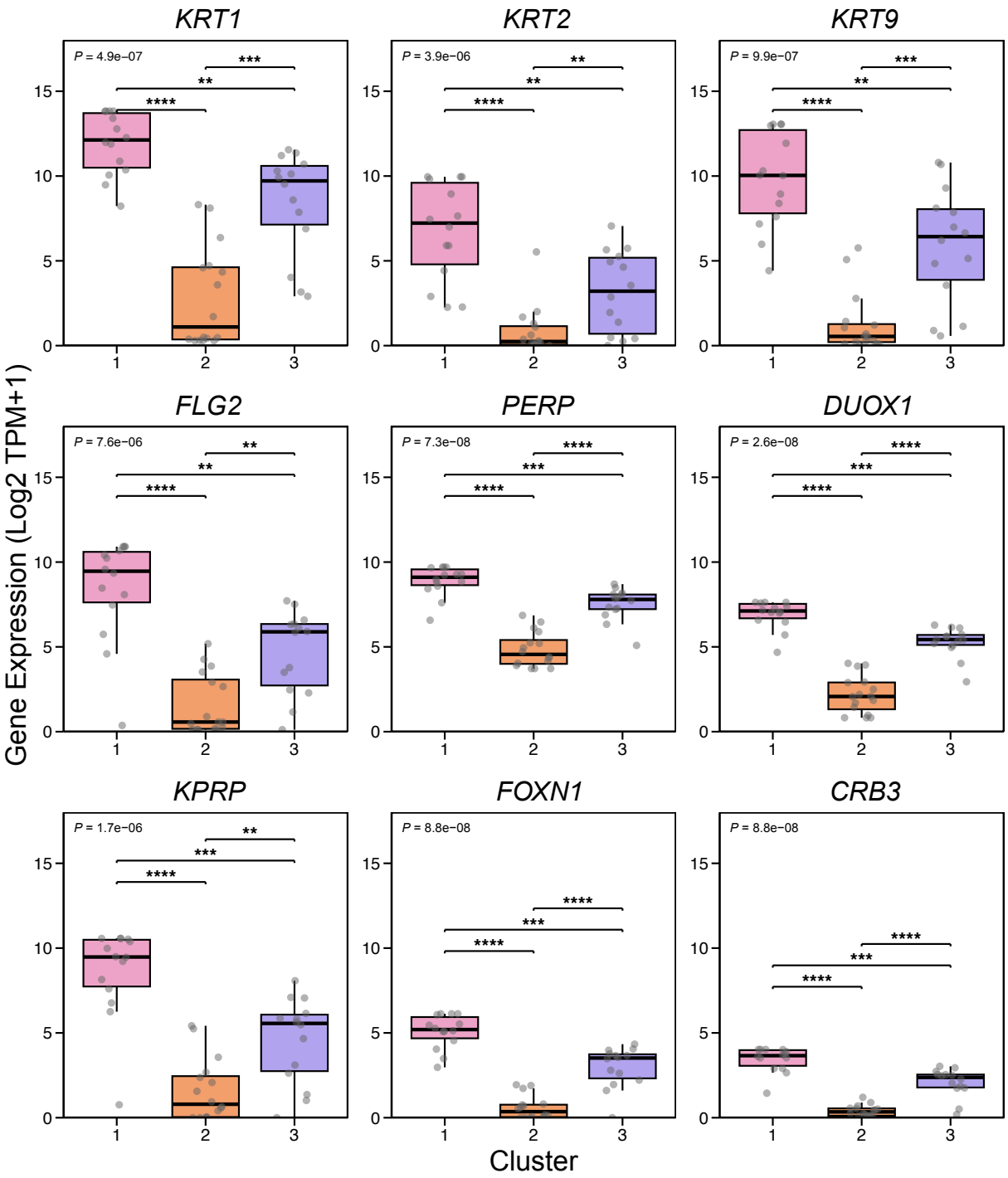

**Supplementary Figure 7. Mutational signatures found in acral melanoma samples from Mexican patients**. a) The SNV component of tumour mutational burden per sample. b-c) Proportions of mutational signatures per sample are shown in stacked bars for single base substitutions (b), and copy-number aberrations (c). In b) and c), samples with a light gray background did not have data available. Genomic subtypes and clinical characteristics are plotted at the bottom.



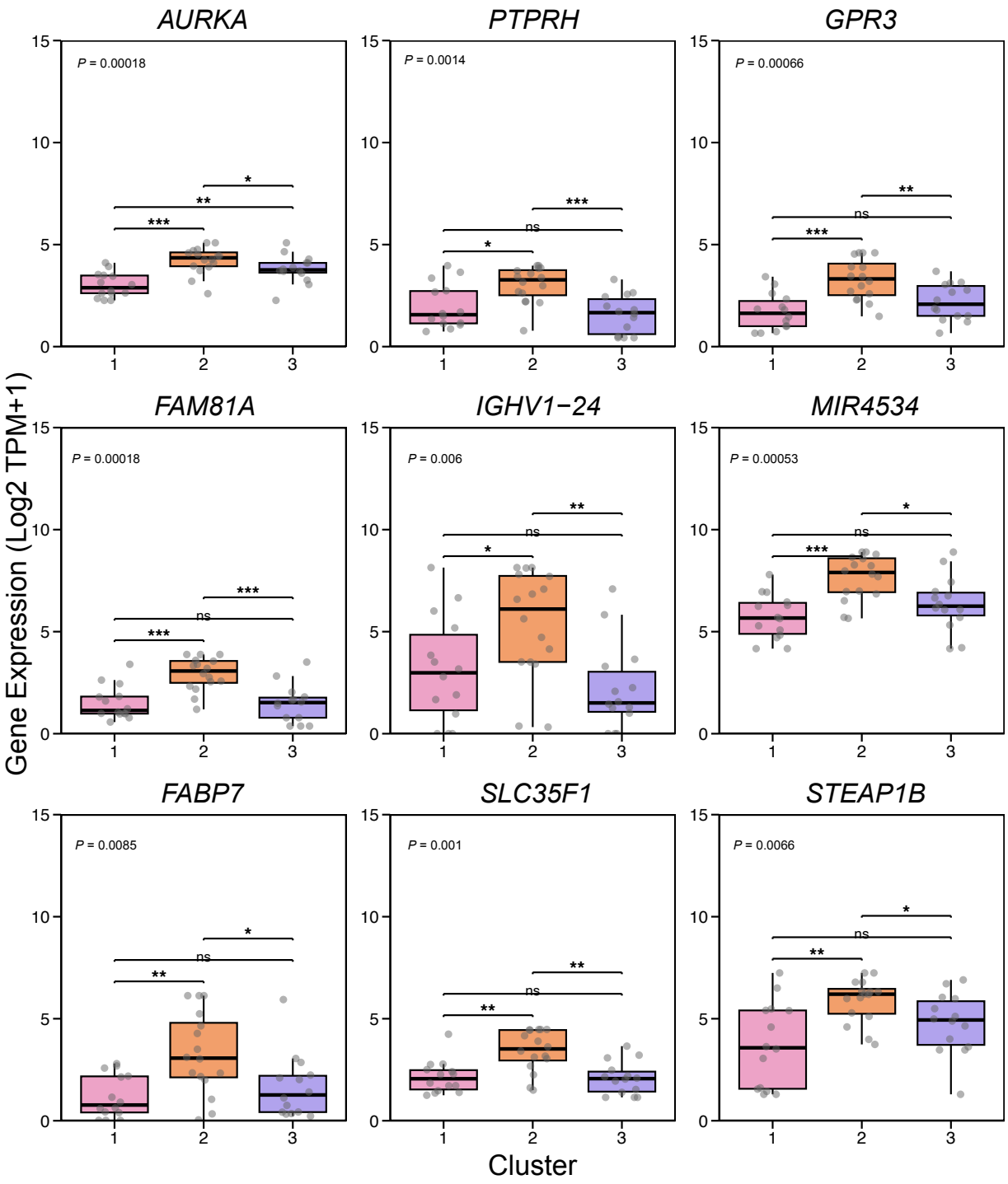

**Supplementary Figure 8. Decomposition plot of single-base substitution mutational signatures for all acral melanoma samples.** Samples that had mutation data (116) were included in this analysis.



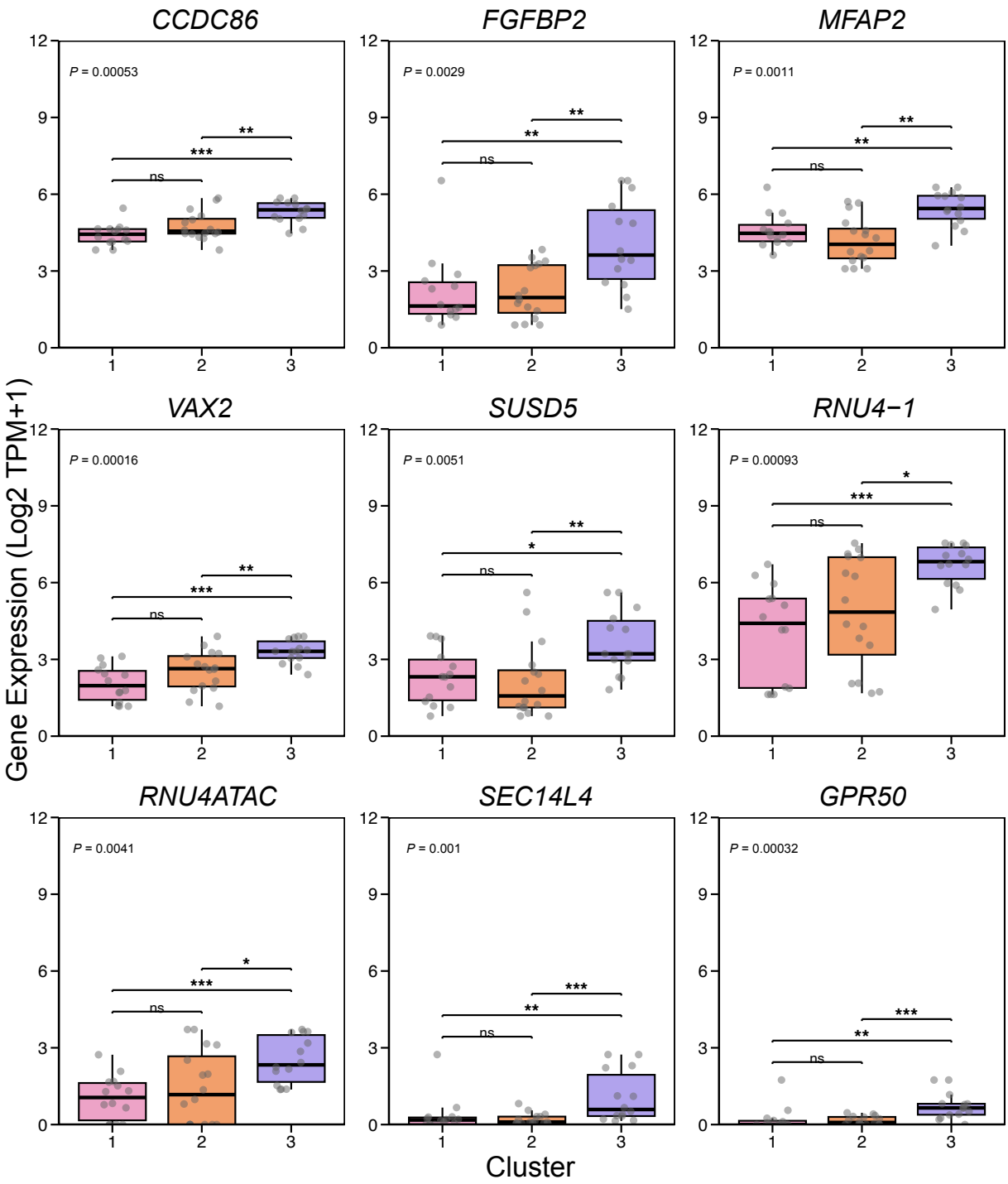

**Supplementary Figure 9. Analyses testing the association of the transcriptional signature found in acral *BRAF*-mutated tumours with downstream *BRAF* signalling**. A) Scatter plot comparison of Acral:Cutaneous gene expression ratio in cutaneous melanoma samples from The Cancer Genome Atlas (TCGA) stratified by activating mutation status. Samples were compared in two groups: non-*BRAF* activated tumours and *BRAF*-activating tumours. Statistical significance was assessed using individual Mann-Whitney U test. B) Hierarchical clustering dendrogram generated using the WIMMS platform to compare the cutaneous melanoma classifier genes to other published molecular signatures, including a signature of genes activated by mutant *BRAF*$^{V600E}$ in melanoma cells (Ryu_2011_BRAFV600E_Targets)[36].



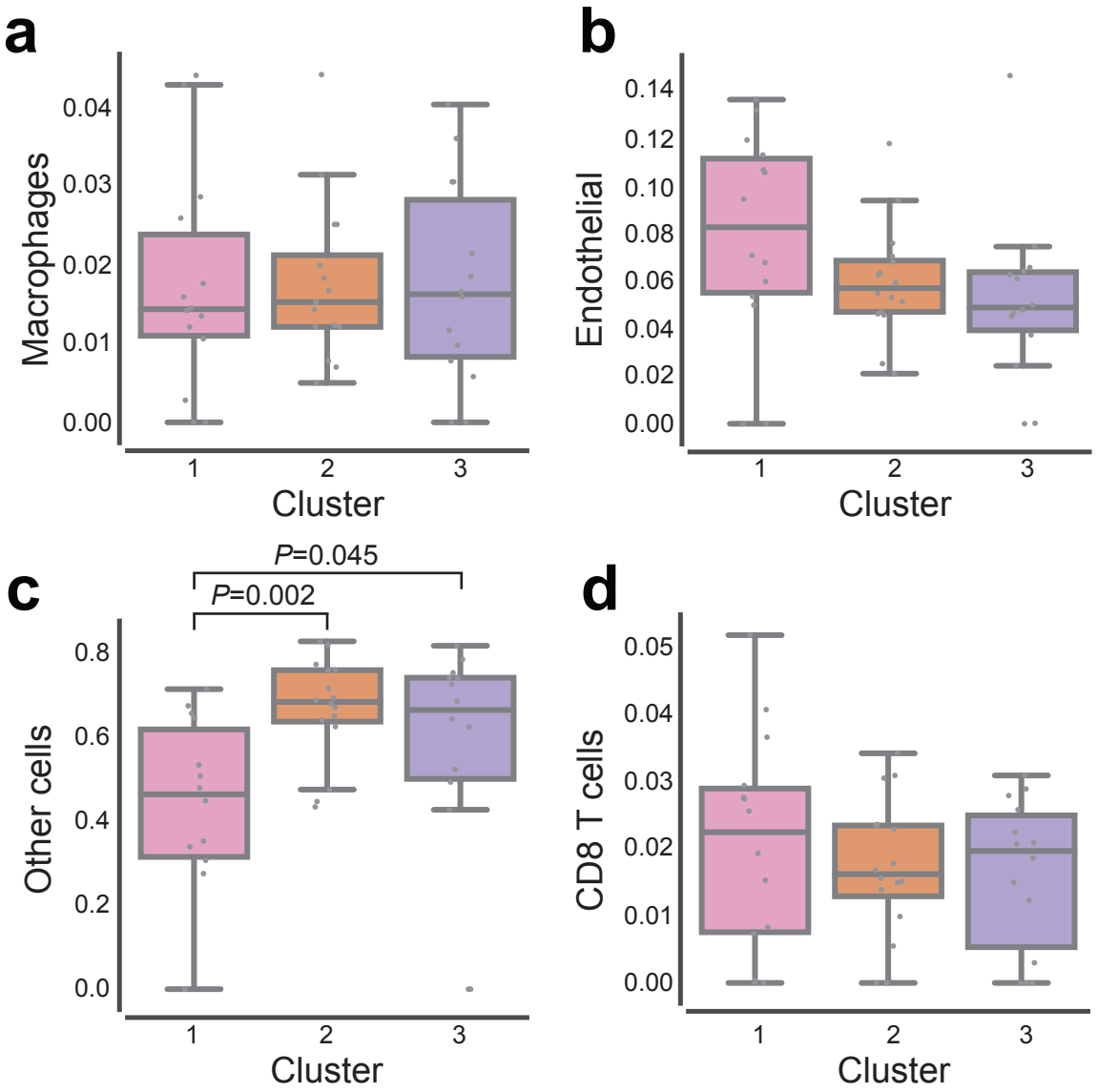

**Supplementary Figure 10. Gene expression for a selection of genes found associated to Cluster 1.** *P*-values were estimated with Kruskal-Wallis tests.



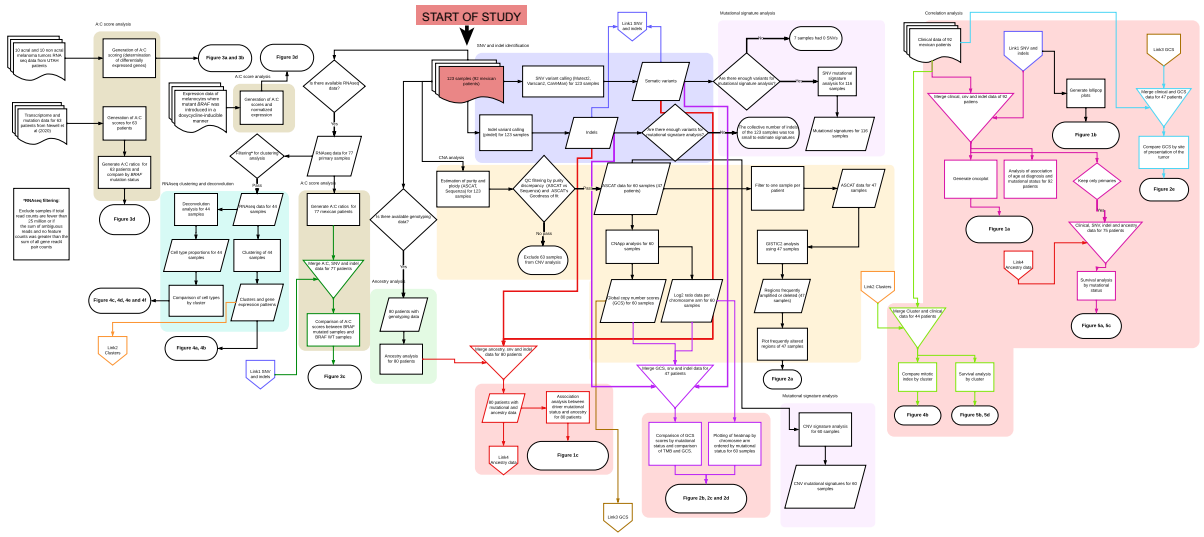

**Supplementary Figure 11. Gene expression for a selection of genes found associated to Cluster 2.** *P*-values were estimated with Kruskal-Wallis tests.




**Supplementary Figure 12. Gene expression for a selection of genes found associated to Cluster 3.** *P*-values were estimated with Kruskal-Wallis tests.




**Supplementary Figure 13. Deconvolution results for macrophages, endothelial cells, CD8 T cells and other cells.** A) Box plot of macrophage proportion (Y axis), as calculated by deconvolution, per sample classified by transcriptional cluster. B) Box plot of endothelial cell proportion (Y axis), as calculated by deconvolution, per sample classified by transcriptional cluster. C) Box plot of other cells (non-immune), as calculated by deconvolution, per sample classified by transcriptional cluster. D) Box plot of CD8 T cell proportion (Y axis), as calculated by deconvolution, per sample classified by transcriptional cluster. The central line within each box represents the median value, the box boundaries represent the interquartile range (IQR), and the whiskers extend to the lowest or highest data point still within 1.5xIQR. Individual data points are plotted as dots. Wilcoxon-Mann-Whitney paired tests were performed.




**Supplementary Figure 14. Flowchart describing all analyses and steps in this work.** Different colours represent different sections of the study, the resulting main figure from each analysis is also indicated in the text.

## Table legends

**Table 1.** Clinical information for patients included in this study.

| **Number of patients** | 92 | |
|---|---|---|
| **Age** | | |
| | *Median* | 60 |
| | *Mean* | 61.22 |
| | *Min* | 32 |
| | *Max* | 98 |
| **Sex** | | |
| | *Female* | 54 |
| | *Male* | 38 |
| **Stage** | | |
| | *IV* | 9 |
| | *III* | 47 |
| | *II* | 25 |
| | *I* | 10 |
| | *In situ* | 1 |
| **Ulceration** | | |
| | *Yes* | 61 |
| | *No* | 29 |
| | *NA* | 2 |
| **Site** | | |
| | *Feet* | 76 |
| | *Hands* | 4 |
| | *Subungual* | 12 |
| **Breslow Thickness** | | |
| | *Median* | 4 |
| | *Mean* | 6.61 |
| | *Min* | 0.45 |
| | *Max* | 50 |

906 **Supplementary Table legends**

907

908 **Supplementary Table 1.** Clinical and molecular information for patients and samples analysed
909 in this study. Clinical information, along with genomic subtype and classification, TMB count, copy
910 number alteration counts, transcriptomic cluster and socioeconomic status are included. For self-
911 reported socioeconomic status, the range is 1-7, with 1 being the lowest and 7 being the highest.
912 The value of Breslow thickness and ulceration for metastatic samples is derived from its
913 corresponding primary.

914

915 **Supplementary Table 2**. Ancestry proportions for five superpopulations for samples in this study
916 and 1000 Genomes Project samples. The header is labelled with the inferred population from
917 comparison with the 1000 Genomes projects.

918

919 **Supplementary Table 3. dNdScv results of neutrality tests at the gene level.** This table
920 includes information on the number of substitutions of each class observed in each gene,
921 maximum-likelihood estimates of dN/dS ratios for each gene, *P*-values obtained by Likelihood-
922 Ratio Tests as described in (Martincorena et al, 2017)[64] and q-values obtained by Benjamini-
923 Hodgberg's multiple testing correction.

924

925 **Supplementary Table 4. *TERT* promoter sequencing results**. In the "Amplification column", 0
926 = not amplified, 1 = low amplification, 2 = high amplification, LQ = low quality. TERT status
927 classification: 1=mutated, 0=wild type.

928

929 **Supplementary Table 5. Analysis of clinical, molecular and transcriptional characteristics**
930 **by sex.** Two percentages are shown in each cell, the first one is by column and second one by
931 row.

932

933 **Supplementary Table 6. Patients with multiple samples and their driver mutational status.**

934

935 **Supplementary Table 7. Amplification and deletion peaks found in acral melanoma**
936 **samples.**

937

938 **Supplementary Table 8. Cytobands, q values, location and genes contained within**
939 **amplification peaks.**

940
941 **Supplementary Table 9. Cytobands, q values, location and genes contained within deletion**
942 **peaks.**
943
944 **Supplementary Table 10. Amplification and deletion peaks found in QWT acral melanoma**
945 **samples.**
946
947 **Supplementary Table 11. Cytobands, q values, location and genes contained within**
948 **amplification peaks in QWT samples.**
949
950 **Supplementary Table 12. Cytobands, q values, location and genes contained within**
951 **deletion peaks in QWT samples.**
952
953 **Supplementary Table 13. Amplification and deletion peaks found in driver-mutated acral**
954 **melanoma samples.**
955
956 **Supplementary Table 14. Cytobands, q values, location and genes contained within**
957 **amplification peaks in driver-mutated samples.**
958
959 **Supplementary Table 15. Cytobands, q values, location and genes contained within**
960 **deletion peaks in driver-mutated samples.**
961
962 **Supplementary Table 16. List of candidate genes from acral and cutaneous melanoma**
963 **datasets.**
964
965 **Supplementary Table 17. Mean gene expression value per gene per RNA cluster.**
966
967 **Supplementary Table 18. Gene enrichment analysis for genes expressed in Cluster 1.**
968
969 **Supplementary Table 19. Gene enrichment analysis for genes expressed in Cluster 2.**
970
971 **Supplementary Table 20. Gene enrichment analysis for genes expressed in Cluster 3.**
972
973 **Supplementary Table 21. RNA deconvolution results per sample.**

974

975 **Supplementary Table 22. Multivariate logistic regression to test the association of any**
976 **driver mutation on recurrence-free survival.**

977

978 **Supplementary Table 23. Multivariate logistic regression to test the association of cluster**
979 **assignmnent on recurrence-free survival.**

980

981 **Supplementary Table 24. Log-rank test for homogeneity, any mutation on overall survival.**

982

983 **Supplementary Table 25. Log rank test of homogeneity, mutation classification on overall**
984 **survival.**

985

986 **Supplementary Table 26. Cox proportional hazards analysis testing the association of any**
987 **mutation on overall survival.**

988

989 **Supplementary Table 27. Chi squared test of independence, cluster assignment and**
990 **survival,**

991

992 **Supplementary Table 28. Comparison between ASCAT and Sequenza purity and ploidy**
993 **estimates.** Samples in black font passed all QC.

994

995 **References**

996


997 1. Rabbie, R., Ferguson, P., Molina-Aguilar, C., Adams, D. J. & Robles-Espinoza, C. D.
998 Melanoma subtypes: genomic profiles, prognostic molecular markers and therapeutic
999 possibilities. *J Pathol* **247**, 539–551 (2019).

1000 2. Ossio, R., Roldán-Marín, R., Martínez-Said, H., Adams, D. J. & Robles-Espinoza, C. D.
1001 Melanoma: a global perspective. *Nat. Rev. Cancer* **17**, 393–394 (2017).

1002 3. Basurto-Lozada, P. *et al.* Acral lentiginous melanoma: Basic facts, biological characteristics
1003 and research perspectives of an understudied disease. *Pigment Cell & Melanoma Research*
1004 **34**, 59–71 (2021).

1005 4. Mao, L., Qi, Z., Zhang, L., Guo, J. & Si, L. Immunotherapy in Acral and Mucosal Melanoma:
1006 Current Status and Future Directions. *Front. Immunol.* **0**, (2021).

1007 5. Curtin, J. A. *et al.* Distinct sets of genetic alterations in melanoma. *N Engl J Med* **353**, 2135–
1008 2147 (2005).

1009 6. Hayward, N. K. *et al.* Whole-genome landscapes of major melanoma subtypes. *Nature* **545**,
1010 175–180 (2017).

1011 7. Newell, F. *et al.* Whole-genome sequencing of acral melanoma reveals genomic complexity
1012 and diversity. *Nature Communications* **11**, 5259 (2020).

1013 8. Shi, K. *et al.* Distinct Genomic Features in a Retrospective Cohort of Mucosal, Acral and
1014 Vulvovaginal Melanomas. *J. Am. Acad. Dermatol.* (2019) doi:10.1016/j.jaad.2019.07.017.

1015 9. Turajlic, S. *et al.* Whole genome sequencing of matched primary and metastatic acral
1016 melanomas. *Genome Res* **22**, 196–207 (2012).

1017 10. Yeh, I. *et al.* Targeted Genomic Profiling of Acral Melanoma. *J. Natl. Cancer Inst.* **111**,
1018 1068–1077 (2019).

1019 11. Carrot-Zhang, J. *et al.* Comprehensive Analysis of Genetic Ancestry and Its Molecular
1020 Correlates in Cancer. *Cancer Cell* **37**, 639-654.e6 (2020).

1021 12. Nassar, A. H. *et al.* Ancestry-driven recalibration of tumor mutational burden and disparate
1022 clinical outcomes in response to immune checkpoint inhibitors. *Cancer Cell* **40**, 1161-
1023 1172.e5 (2022).
1024 13. Carrot-Zhang, J. *et al.* Genetic Ancestry Contributes to Somatic Mutations in Lung Cancers
1025 from Admixed Latin American Populations. *Cancer Discovery* **11**, 591–598 (2021).
1026 14. Molina-Aguilar, C. & Robles-Espinoza, C. D. Tackling the lack of diversity in cancer
1027 research. *Disease Models & Mechanisms* **16**, dmm050275 (2023).
1028 15. Park, S. L., Cheng, I. & Haiman, C. A. Genome-Wide Association Studies of Cancer in
1029 Diverse Populations. *Cancer Epidemiol Biomarkers Prev* **27**, 405–417 (2018).
1030 16. Campbell, P. J. *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
1031 17. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a
1032 bridge from a population-based to a more 'personalized' approach to cancer staging. *CA*
1033 *Cancer J Clin* **67**, 93–99 (2017).
1034 18. Wang, M. *et al.* The genetic evolution of acral melanoma. *Nat Commun* **15**, 6146 (2024).
1035 19. Wang, M., Banik, I., Shain, A. H., Yeh, I. & Bastian, B. C. Integrated genomic analyses of
1036 acral and mucosal melanomas nominate novel driver genes. *Genome Medicine* **14**, 65
1037 (2022).
1038 20. Wong, K. *et al.* Cross-species genomic landscape comparison of human mucosal
1039 melanoma with canine oral and equine melanoma. *Nature Communications* **10**, 1–14
1040 (2019).
1041 21. Hendricks, W. P. D. *et al.* Somatic inactivating PTPRJ mutations and dysregulated
1042 pathways identified in canine malignant melanoma by integrated comparative genomic
1043 analysis. *PLoS Genet* **14**, e1007589 (2018).
1044 22. Bailey, M. H. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations.
1045 *Cell* **173**, 371-385.e18 (2018).

1046 23. Turner, J. A. *et al.* Expanding the landscape of oncogenic drivers and treatment options in
1047 acral and mucosal melanomas by targeted genomic profiling. *International Journal of*
1048 *Cancer* **155**, 1792–1807 (2024).
1049 24. Jette, C. *et al.* The Tumor Suppressor Adenomatous Polyposis Coli and Caudal Related
1050 Homeodomain Protein Regulate Expression of Retinol Dehydrogenase L *. *Journal of*
1051 *Biological Chemistry* **279**, 34397–34405 (2004).
1052 25. Liu, H. *et al.* Integrative molecular and spatial analysis reveals evolutionary dynamics and
1053 tumor-immune interplay of in situ and invasive acral melanoma. *Cancer Cell* **42**, 1067-
1054 1085.e11 (2024).
1055 26. Farshidfar, F. *et al.* Integrative molecular and clinical profiling of acral melanoma links focal
1056 amplification of 22q11.21 to metastasis. *Nat Commun* **13**, 898 (2022).
1057 27. Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic RAC1 mutations in
1058 melanoma. *Nat Genet* **44**, 1006–1014 (2012).
1059 28. Cancer Genome Atlas Network. Genomic Classification of Cutaneous Melanoma. *Cell* **161**,
1060 1681–1696 (2015).
1061 29. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat Genet*
1062 **47**, 1402–1407 (2015).
1063 30. Senkin, S. *et al.* Geographic variation of mutagenic exposures in kidney cancer genomes.
1064 *Nature* **629**, 910–918 (2024).
1065 31. Steele, C. D. *et al.* Signatures of copy number alterations in human cancer. *Nature* **606**,
1066 984–991 (2022).
1067 32. Everall, A. *et al.* Comprehensive repertoire of the chromosomal alteration and mutational
1068 signatures across 16 cancer types from 10,983 cancer patients. 2023.06.07.23290970
1069 Preprint at https://doi.org/10.1101/2023.06.07.23290970 (2023).
1070 33. Yeh, I. & Bastian, B. C. Melanoma Pathology 2.0 – New Approaches and Classification. *Br J*
1071 *Dermatol* **185**, 282–293 (2021).

1072 34. McNeal, A. S. *et al.* BRAFV600E induces reversible mitotic arrest in human melanocytes via
1073 microrna-mediated suppression of AURKB. *Elife* **10**, e70385 (2021).
1074 35. Hu, M., Coleman, S., Judson-Torres, R. L. & Tan, A. C. The classification of melanocytic
1075 gene signatures. *Pigment Cell Melanoma Res* **37**, 854–863 (2024).
1076 36. Ryu, B. *et al.* Global analysis of BRAFV600E target genes in human melanocytes identifies
1077 matrix metalloproteinase-1 as a critical mediator of melanoma growth. *J Invest Dermatol*
1078 **131**, 1579–1583 (2011).
1079 37. Belote, R. L. *et al.* Human melanocyte development and melanoma dedifferentiation at
1080 single-cell resolution. *Nat Cell Biol* **23**, 1035–1047 (2021).
1081 38. Weiss, J. M. *et al.* Anatomic position determines oncogenic specificity in melanoma. *Nature*
1082 **604**, 354–361 (2022).
1083 39. Long, G. V. *et al.* Prognostic and clinicopathologic associations of oncogenic BRAF in
1084 metastatic melanoma. *J Clin Oncol* **29**, 1239–1246 (2011).
1085 40. Wu, X. *et al.* Mutations in BRAF codons 594 and 596 predict good prognosis in melanoma.
1086 *Oncol Lett* **14**, 3601–3605 (2017).
1087 41. Houben, R. *et al.* Constitutive activation of the Ras-Raf signaling pathway in metastatic
1088 melanoma is associated with poor prognosis. *J Carcinog* **3**, 6 (2004).
1089 42. Moreau, S. *et al.* Prognostic Value of BRAFV600 Mutations in Melanoma Patients After
1090 Resection of Metastatic Lymph Nodes. *Ann Surg Oncol* **19**, 4314–4321 (2012).
1091 43. Ribas, A. & Flaherty, K. T. BRAF targeted therapy changes the treatment paradigm in
1092 melanoma. *Nat Rev Clin Oncol* **8**, 426–433 (2011).
1093 44. Jakob, J. A. *et al.* NRAS Mutation Status is an Independent Prognostic Factor in Metastatic
1094 Melanoma. *Cancer* **118**, 4014–4023 (2012).
1095 45. Thomas, N. E. *et al.* Association Between NRAS and BRAF Mutational Status and
1096 Melanoma-Specific Survival Among Patients With Higher-Risk Primary Melanoma. *JAMA*
1097 *Oncol* **1**, 359–368 (2015).

1098 46. Bai, X. *et al.* MAPK Pathway and TERT Promoter Gene Mutation Pattern and Its Prognostic
1099 Value in Melanoma Patients: A Retrospective Study of 2,793 Cases. *Clin Cancer Res* **23**,
1100 6120–6127 (2017).
1101 47. Thielmann, C. M. *et al.* NF1-mutated melanomas reveal distinct clinical characteristics
1102 depending on tumor origin and respond favorably to immune-checkpoint inhibitors. *Eur J*
1103 *Cancer* **159**, 113–124 (2021).
1104 48. Cirenajwis, H. *et al.* 1-mutated melanoma tumors harbor distinct clinical and biological
1105 characteristics. *Molecular Oncology* **11**, 438–451 (2017).
1106 49. Tsoumakidou, M. The advent of immune stimulating CAFs in cancer. *Nat Rev Cancer* **23**,
1107 258–269 (2023).
1108 50. Kim, I. S. *et al.* Microenvironment-derived factors driving metastatic plasticity in melanoma.
1109 *Nat Commun* **8**, 14343 (2017).
1110 51. Deacon, D. C. *et al.* Classification of Cutaneous Melanoma and Melanocytic Nevi with
1111 microRNA Ratios is Preserved in the Acral Melanoma Subtype. *Journal of Investigative*
1112 *Dermatology*.
1113 52. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in
1114 unrelated individuals. *Genome Res* **19**, 1655–1664 (2009).
1115 53. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1116 *arXiv:1303.3997 [q-bio]* (2013).
1117 54. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–
1118 2079 (2009).
1119 55. Andrews, S. FastQC: a quality control tool for high throughput sequence data.
1120 http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).
1121 56. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing
1122 next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

1123 57. Bergmann, E. A., Chen, B.-J., Arora, K., Vacic, V. & Zody, M. C. Conpair: concordance and
1124 contamination estimator for matched tumor–normal pairs. *Bioinformatics* **32**, 3196–3198
1125 (2016).
1126 58. Jones, D. *et al.* cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect
1127 Somatic Single Nucleotide Variants in NGS Data. *Current Protocols in Bioinformatics* **56**,
1128 15.10.1-15.10.18 (2016).
1129 59. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. 861054 Preprint at
1130 https://doi.org/10.1101/861054 (2019).
1131 60. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in
1132 cancer by exome sequencing. *Genome Res* **22**, 568–576 (2012).
1133 61. Raine, K. M. *et al.* cgpPindel: Identifying Somatically Acquired Insertion and Deletion Events
1134 from Paired End Sequencing. *Current Protocols in Bioinformatics* **52**, 15.7.1-15.7.12 (2015).
1135 62. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to
1136 detect break points of large deletions and medium sized insertions from paired-end short
1137 reads. *Bioinformatics* **25**, 2865–2871 (2009).
1138 63. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and
1139 cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
1140 64. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*
1141 **171**, 1029-1041.e21 (2017).
1142 65. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).
1143 66. Favero, F. *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor
1144 sequencing data. *Ann Oncol* **26**, 64–70 (2015).
1145 67. Ross, E. M., Haase, K., Van Loo, P. & Markowetz, F. Allele-specific multi-sample copy
1146 number segmentation in ASCAT. *Bioinformatics* **37**, 1909–1911 (2021).
1147 68. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets
1148 of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).

1149 69. Franch-Expósito, S. *et al.* CNApp, a tool for the quantification of copy number alterations
1150 and integrative analysis revealing clinical implications. *eLife* **9**, e50267 (2020).
1151 70. Bergstrom, E. N. *et al.* SigProfilerMatrixGenerator: a tool for visualizing and exploring
1152 patterns of small mutational events. *BMC Genomics* **20**, 685 (2019).
1153 71. Islam, S. M. A. *et al.* Uncovering novel mutational signatures by *de novo* extraction with
1154 SigProfilerExtractor. *Cell Genomics* **2**, 100179 (2022).
1155 72. Sondka, Z. *et al.* COSMIC: a curated database of somatic variants and clinical data for
1156 cancer. *Nucleic Acids Research* **52**, D1210–D1217 (2024).
1157 73. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
1158 74. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-
1159 throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
1160 75. Gu, Z., Schlesner, M. & Hübschmann, D. cola: an R/Bioconductor package for consensus
1161 partitioning through a general framework. *Nucleic Acids Res* **49**, e15 (2020).
1162 76. Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E. & Gfeller, D. Simultaneous
1163 enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife*
1164 **6**, e26476 (2017).
1165
1166

**Acknowledgments**

We are deeply grateful to patients and their families for agreeing to form part of this study and providing access to their samples. We are also thankful to members of the CGBio lab team at LIIGH-UNAM for valuable discussions regarding the findings in this article. The authors wish to thank Luis A. Aguilar, Alejandro de León and Alejandro Avalos from the Laboratorio Nacional de Visualización Científica Avanzada and Jair S. García Sotelo, Abigayl Hernández, Eglee Lomelín, Iliana Martínez, Rebeca Muciño, María A. Ávila, Christian Molina-Aguilar, Alejandra Castillo and Carina Uribe Díaz from the International Laboratory for Human Genome Research, National Autonomous University of Mexico. We are grateful to the International Cancer Genome Consortium Data Access Committee for granting access to ICGC controlled data. We are also thankful to Kim Wong and Daniel Desposorio for useful discussions. Work included in this paper has been funded by Wellcome Trust (204562/Z/16/Z and 227228/Z/23/Z to C.D.R.-E.), the Melanoma Research Alliance (Pilot Award #825924, to C.D.R.-E.), the Mexican National Council of Humanities, Science and Technology (CONAHCYT, FOSISS A3-S-31603, to C.D.R.-E.), Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT UNAM) (IN209422 to C.D.R.-E.), and the Wellcome Sanger Institute through an International Fellowship. This project was also supported by the MRC Dermatlas project; MR/V000292/1. A.J., D.C.D., and R.L.J.-T. are supported by the Department of Dermatology and the Huntsman Cancer Foundation. This work was funded in part by the Melanoma Research Alliance Dermatology Fellows award to D.C.D., the Harry J Lloyd Charitable Trust Melanoma Research Grant to R.L.J.-T., a National Cancer Institute R01 (R01CA229896) to R.L.J.-T., and pilot funds from the Huntsman Cancer Institute Melanoma Center. We utilised the Shared Resources for Research Informatics and High-Throughput Genomics and Bioinformatics Analysis, each supported by the National Cancer Institute of the National Institutes of Health under Award Number P30CA042014. M.D.-G. and P.G.-G. were awarded fellowships within the “Generación D” initiative, Red.es, Ministerio para la Transformación Digital y de la Función Pública, for talent attraction (C005/24-ED CV1), funded by the European Union NextGenerationEU funds, through PRTR. P.B. is a PhD student from Programa de Doctorado en Ciencias Biológicas, Universidad Nacional Autónoma de México (UNAM), and was supported by Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCyT, now known as SECIHTI) (holder no. 562546, scholarship no.762536). This paper is part of P.B.’s requirements for obtaining a Doctoral degree at the Posgrado en Ciencias Biológicas, UNAM.

## Author Contributions

P.B.-L., M.E.V.-C., D.C.-I., E.F.R., J.B., P.A.J., I.S.-W., J.R.C.W.-R., K.L.C.-R., A.J., D.C.D., J.I.R.-G., O.I.G.-S. and M.C.V.H. performed bioinformatic and statistical analyses. C.M.-A., F.G.A.-G., M.C.-V., R.O.-L. and L. v.d.W. did sample cataloguing and nucleic acid extraction. E.T.D. provided computational resources and advice on statistical analyses. A.A.-C., D.Y.G.-O., H.M.-S., R.R.-M., H.V.C., L.A.T.-P. and D.H.-U. assessed patients and provided access to biological samples. A.H.-M. provided facilities for sample processing and supervised that part of the work. M.J.A., I.F. and M.T. performed sample histopathology. P.G.G., M.D.-G. and L.B.A. supervised the mutational signatures analysis. Y.S.-P. provided access to patient clinical information and supervised that part of the work. G.K.I., R.L.B. and R.M.W. provided data and information that crucially helped the interpretation of the results in this manuscript. D.T.B. performed survival statistical analyses. P.A.P., R.L.J.-T., D.J.A. and C.D.R.-E. jointly supervised this work. C.D.R.-E. wrote the manuscript with assistance from P.B., P.A.P., R.L.J.-T. and D.J.A.

## Code and data availability statement

Sequencing data are available at the European Genome-Phenome Archive (EGA). DNA sequencing data are available under ENA accession number EGAS00001003740 and RNA sequencing data under ENA accession number EGAS00001003758. Code is available at https://github.com/CGBio-Lab/Mex-acral-exomes-transcriptomes.

## Competing interests

The authors declare no competing interests.