Ancestry and somatic profile predict acral melanoma origin and prognosis

Authors:

Patricia Basurto-Lozada[1], Martha Estefania Vázquez-Cruz[1], Christian Molina-Aguilar[1], Amanda Jiang[2,3], Dekker C. Deacon[2,3], Dennis Cerrato-Izaguirre[4], Irving Simonin-Wilmer[1], Fernanda G. Arriaga-González[1,5], Kenya L. Contreras-Ramírez[1], Eric T. Dawson[6], J. Rene C. Wong-Ramirez[1,7], Johana Itzel Ramos-Galguera[1], Alethia Álvarez-Cano[8], Dorian Y. García-Ortega[9], Omar Isaac García-Salinas[1,5], Alfredo Hidalgo-Miranda[10], Mireya Cisneros-Villanueva[10], Héctor Martínez-Said[9], Mark J. Arends[11], Ingrid Ferreira[5], Mark Tullett[12], Rebeca Olvera-León[1,5], Louise van der Weyden[5], Martín del Castillo Velasco Herrera[5], Rodrigo Roldán-Marín[13], Helena Vidaurri de la Cruz[14], Luis Alberto Tavares-de-la-Paz[15], Diego Hinojosa-Ugarte[15], Rachel L. Belote[2,16], D. Timothy Bishop[17], Marcos Díaz-Gay[18-20], Ludmil B. Alexandrov[18-20], Yesennia Sánchez-Pérez[4], Gino K. In[21], Richard M. White[22,23], Patrícia A. Possik[24], Robert L. Judson-Torres[2,3], David J. Adams[5], Carla Daniela Robles-Espinoza[1,5,*]

Affiliations:

1. Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Santiago de Querétaro, Mexico, 76230
2. Huntsman Cancer Institute, University of Utah Health Sciences Center, Salt Lake City, Utah, USA
3. Department of Dermatology, University of Utah, Salt Lake City, UT, USA
4. Subdirección de Investigación Básica, Instituto Nacional de Cancerología (INCan), San Fernando No. 22, Tlalpan, Ciudad de México CP. 14080, Mexico.
5. Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK.
6. Nvidia Corporation, Santa Clara, CA, USA.
7. Research Program in Systems Oncology, University of Helsinki, Helsinki, Finland.
8. Surgical Oncology, Christus Muguerza Alta Especialidad, Monterrey, Nuevo Leon, Mexico.
9. Surgical Oncology, Skin, Soft Tissue & Bone Tumors Department, National Cancer Institute, Mexico City, Mexico.
10. Laboratorio de Genómica del Cáncer, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, Mexico.
11. Edinburgh Pathology, Cancer Research UK Scotland Centre, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK
12. Department of histopathology, University Hospitals Sussex, St Richard hospital, Spitalfield lane, Chichester
13. Dermato-Oncology Clinic, Research Division, Faculty of Medicine, Universidad Nacional Autónoma de México, Mexico City, Mexico.
14. Pediatric Dermatology Service, General Hospital of Mexico Dr. Eduardo Liceaga, Ministry of Health. Mexico City, Mexico.
15. Surgical Oncology, Bajio Regional High Specialty Hospital, Leon, Mexico.
16. The Ohio State University, Department of Molecular Genetics, Columbus, Ohio, United States
17. Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK.
18. Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA.
19. Department of Bioengineering, University of California San Diego, La Jolla, CA, USA.
20. Moores Cancer Center, University of California San Diego, La Jolla, CA, USA.
21. University of Southern California, Keck School of Medicine, Norris Comprehensive Cancer Center, Division of Oncology, Los Angeles, CA, USA.
22. Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, USA
23. Nuffield Department of Medicine, Ludwig Institute for Cancer Research, University of Oxford, Oxford, UK
24. 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, here we characterise the genome and transcriptome of 128 acral melanoma tumours from 96 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 number of structural variants. These *BRAF*-mutated tumours have a transcriptional profile 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 transcriptional profiling defined three expression clusters; these characteristics were associated with overall survival. We highlight novel low-frequency drivers, such as *SPHKAP*, which correlate with a distinct genomic profile and clinical characteristics. 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 on the nail unit (subungual location), and its genome differs significantly from other CM subtypes[5]. In contrast to UV-induced 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 grossly 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.

104 In this study, we analysed 128 AM samples from 96 Mexican patients through genotyping, exome
105 sequencing, SNV and insertion/deletion (indel) variant calling, copy number estimation, and gene
106 expression profiling, and examined the correlation of these molecular characteristics with clinical
107 variables. We found a significant correlation between genetic ancestry and *BRAF* somatic
108 mutations, as well as a distinct transcriptomic profile in these tumours compared to non-*BRAF*
109 mutated samples. We also identified significant differences in recurrence-free survival among
110 patients with driver mutations compared to patients with wild-type tumours, and in overall survival
111 among patients with distinct gene expression profiles.
112
113 **Results**
114
115 *Ancestry and clinical characteristics of Mexican AM patients*
116 A total of 128 uniformly ascertained samples from 96 patients from a large Mexican tertiary referral
117 hospital were analysed in this study (**Methods**, **Supplementary Table 1**). Ninety-three of these
118 tumours were primaries, 28 were metastases, five were recurrencies, one was a lesion in transit,
119 and one was unknown (**Supplementary Table 1**). Latin American genomes are generally a
120 mixture of European, African and Amerindian ancestry. Of note, 89% of genotyped samples in
121 this study had predominantly Amerindian ancestry (median 79%) (**Supplementary Figure 1,**
122 **Supplementary Table 2**) with European and African ancestries contributing a median of 14%
123 and 2.6%, respectively. The median age of the patients in this cohort was 61, with 61% of the
124 patients being female. Most patients were stage III (AJCC 8th edition)[17] at diagnosis, and the most
125 common primary site was the foot, most frequently the sole. The median Breslow thickness was
126 4.7mm and the majority of tumours were ulcerated (65%) (**Table 1**). It should be noted that
127 virtually no patients received immune check point inhibitors or targeted therapy, due to lack of
128 access.
129
130 *Genomic profiling of AM samples identifies correlations of ancestry and age with somatic*
131 *alterations*
132 Considering all 128 samples, AM tumours showed a SNV/indel [hereinafter referred to as tumour
133 mutational burden (TMB)] mean of 3.37 mutations per megabase (mut/Mb), a median of 2.75
134 mut/Mb (range: 1.05-11.46 mut/Mb). When including only one sample per patient, with primaries
135 being preferentially selected, the most frequently mutated genes were *NRAS* (15% of samples,
136 q-value < 4.33×10^{-8}), *KIT* (14%, q-value=4.33×10^{-8}), *BRAF* (11%, q-value=1.96×10^{-6}) and *NF1*
137 (7%, q-value=0.057) (**Figure 1a**). These genes were identified as being under positive selection

138 (**Methods**) and represent known driver genes. These genes showed the characteristic mutational
139 profile of oncogenes with a predominance of hotspot missense mutations, except for *NF1*, which
140 showed a pattern characteristic of a tumour suppressor and had frameshift insertions, deletions,
141 and nonsense mutations distributed throughout (**Figure 1b**). Notably, these genes exhibit mutual
142 exclusivity (only two patients have tumours with mutations in more than one of these genes, with
143 one patient having both a $BRAF^{G606W}$ mutation, which is a suspected loss of function[18], and a
144 $NRAS^{Q61R}$) which likely reflects their functional redundancy in activating the MAPK pathway. Other
145 genes previously reported as mutated in other melanoma subtypes, as well as other cancer types
146 are also mutated in this cohort, such as *TP53*, *HRAS* and *KRAS* (**Figure 1a**). In summary, the
147 "classic" melanoma driver genes (*N/H/KRAS*, *BRAF* and *NF1*) are mutated in fewer than 40% of
148 Mexican AM samples, with most of the samples in this cohort therefore being classified as "triple
149 wild type" melanomas. We next appraised those tumours without mutations in any of the
150 abovementioned four driver genes ("quadruple wild-type"), revealing three patients that carried
151 mutations in *SPHKAP* in their primary tumour (**Supplementary Figure 2a**). The mutations in
152 these tumours are protein-changing (one stop gained, one inframe deletion and one missense,
153 predicted deleterious and damaging) (**Supplementary Figure 2b**). *SPHKAP* codes for an
154 inhibitor of sphingosine kinase 1 (*SPHK1*), which in turn plays a key role in the activation of the
155 NFκB and TNF-α signalling pathways. Other recurrently mutated genes in quadruple-wild type
156 melanomas include *POU3F3*, *RDH5*, *MED12* and *TP53* (**Supplementary Figure 2**) which may
157 represent low frequency drivers.
158
159 When examining the relationship between ancestry and somatic profile, we identified significantly
160 higher odds (*P*-value=0.02) of carrying a *BRAF* somatic mutation with increasing European
161 ancestry in a linear model controlling for age at diagnosis, sex, self-reported socioeconomic status
162 and mutational burden (**Figure 1c**). Patients with mutations in *KIT* showed a tendency for higher
163 Amerindian ancestry (**Figure 1c**, **Supplementary Figure 3**). We also found that patients with
164 *NRAS* mutations are significantly younger at diagnosis (median and mean age of diagnosis for
165 patients with *NRAS* mutations= 50 and 51.4 vs without = 62.5 and 62.7, respectively, *P*-
166 value=0.01) (**Figure 1d**).
167
168 *Somatic copy number landscape of AM samples identifies correlations with somatic alterations*
169 Somatic copy number alteration (SCNA) analysis across all samples showed a higher burden of
170 amplifications than deletions (**Figure 2a**). Examination of 70 samples, one per patient, that
171 passed quality filtering for this type of analysis (**Methods**), showed that 24 regions were frequently

172 amplified, and 15 regions were frequently deleted. About a fifth (21%) of samples had an
173 estimated ploidy around 4, suggesting whole genome duplication (WGD). Potential driver genes
174 in frequently amplified regions include *CRKL* (47% of samples), *CCND1* (34%), *CDK4* (20%), and
175 *KIT* (18.5%) (**Supplementary Tables 3-5**). Regions that showed recurrent deletions contained
176 genes such as *CDKN2A*, *CDKN2B, ATM* and *FOXO3*. *CDKN2A* and *CDKN2B* had deletions in
177 66% of samples, while *ATM* and *FOXO3* both presented heterozygous deletions in 50% of
178 samples, respectively.
179
180 When stratifying samples by mutational status (considering *BRAF*-, *NRAS*-, *NF1*-, *KIT*-mutated
181 and multi-hit, which included two samples with mutations in more than one of these drivers), we
182 saw statistically significant differences in SCNA among groups (**Figure 2b**). Specifically, *NRAS*-
183 and *BRAF*- mutated tumours had significantly fewer SCNAs (Global copy number alteration score
184 [GCS], **Methods**) than *KIT*- and *NF1*-mutated tumours (**Figure 2c**). Samples without mutations
185 in these drivers had a range of GCS scores. Samples with *BRAF* and *NRAS* mutations had the
186 lowest median TMB as well, with *NF1*-mutated tumours having the highest median TMB
187 (**Supplementary Figure 4**). We did not see a significant correlation between GCS score and TMB
188 (Pearson's product moment correlation coefficient=0.20, *P*-value=0.09) (**Figure 2d**). Tumours
189 from the subungual region also had a higher median GCS score than those found on the hands
190 and feet (**Figure 2e**).
191
192 *Mutational signature analyses identify potential sources of mutation*
193 Single-base substitution mutational signature analysis across samples identified previously
194 reported COSMICv3.4 signatures SBS1, SBS5, SBS7a, SBS7b, SBS40a and some residual
195 SBS45. Apart from clock-like signatures SBS1 and SBS5[19], SBS40a was also prevalent across
196 the cohort, contributing 28.24% of mutations to the total. SBS40a is of unknown origin but has
197 been identified in many cancer types[20]. Indel mutational signature analysis identified two
198 contributing signatures, clock-like ID2 and ID12, also of unknown origin. Copy number signature
199 analysis identified a number of previously reported signatures across different samples[21,22]. CN1,
200 which has been associated with a diploid state and CN9, which is potentially caused by local loss
201 of heterozygosity on a diploid background, dominated the CN landscape (**Methods,**
202 **Supplementary Information, Supplementary Figure 5**). Nevertheless, this analysis is
203 precluded by small numbers of mutations and the formalin-fixed paraffin-embedded (FFPE) origin
204 of these samples.

205 *BRAF-mutated acral melanomas exhibit a transcriptional signature more characteristic of non-*
206 *acral cutaneous melanomas*
207 In our study, *BRAF*/*NRAS*-mutated tumours exhibited different SCNA profiles compared to all
208 other AMs and were associated with distinct demographic and clinical features, suggestive of a
209 divergent aetiology for this genetic subset. As it has been previously postulated that *BRAF*-
210 mutated acral melanomas might be more biologically like melanomas from non-acral sites than
211 to other acral melanomas[10,23], we investigated this hypothesis. We successfully extracted and
212 sequenced RNA from 80 primary tumours from different patients in this collection
213 (**Supplementary Table 1, Methods**). We then generated a gene signature-based score for
214 identifying acral- versus cutaneous-derived melanomas. For this, we sourced a list of candidate
215 genes from AM and CM datasets (**Methods, Supplementary Table 6**) and identified twenty
216 genes with high classification accuracy in a training cohort of 10 primary AMs and 10 primary CMs
217 (**Figure 3a-b**). We then obtained scores for samples in our dataset of AMs, separating primary
218 *BRAF*-missense (n=9), *NRAS*-missense (n=12) vs *BRAF/NRAS*-wildtype (n=59) tumours. We
219 observed a difference between *BRAF*-mutated and *BRAF/NRAS*-wildtype tumours (*P*-
220 value=0.055) (**Figure 3c**). We then replicated this analysis in an independent cohort of 63 AMs
221 from Newell *et al* (2020)[7] (*BRAF*-missense n=13, wild-type n=50), which confirmed these results
222 (*P*-value=0.039) (**Figure 3d**). In these comparisons, *BRAF*-missense tumours expressed a more
223 "CM-like" transcriptional program, indicating that *BRAF*-mutated melanomas that occur at acral
224 sites are transcriptionally more similar to non-acral cutaneous melanomas, and are associated
225 with increasing European genetic ancestry.
226
227 *Transcriptional landscape of AM tumours identifies three subgroups with distinct clinical and*
228 *prognostic characteristics*
229 We then applied a more stringent quality filtering, including coverage and alignment features, to
230 primary tumours in this collection with 47 samples remaining for further analyses (**Methods**,
231 **Supplementary Table 1**). Consensus clustering of gene expression identified three sample
232 groups with distinct transcriptional profiles (**Figure 4a, Supplementary Table 7**). Cluster 1 was
233 characterised by a high expression of cytokines (*e.g.*, *CXCL12*, *CCL13, ICOSLG, IL7, IL4R, IL1R,*
234 *CD69, IL15RA, CXCL14*), immune-related (*e.g*. *CD209*, *INHBA*) and invasion-related (*e.g.*, *AXL*,
235 *ZEB1*) genes, which we termed "mixed"; Cluster 2 was characterised by a "proliferative" and
236 "pigmentation"-related signature, with high expression of genes such as *MITF*, *SOX10*, *TYR* and
237 *DCT*; and Cluster 3 showed expression mostly of keratins and epidermal-related genes ("keratin-
238 related"). Interestingly, Cluster 1 was associated with better prognostic clinical characteristics,

239 such as a small proportion of ulcerated samples, lower Breslow depth and earlier clinical stages,
240 and a tendency for lower mitotic rates (**Figure 4b**). Deconvolution of gene expression profiles
241 also indicated differences in immune cell infiltration composition, with Cluster 1 having a higher
242 proportion of endothelial cells, CD4+ T cells and cancer-associated fibroblasts (CAFs) (**Figure**
243 **4c-e**).
244
245 *Somatic and gene expression profile influence recurrence-free survival*
246 Next, we evaluated whether the genomic and transcriptomic characteristics had any impact on
247 patient overall or recurrence-free survival. We included in the analysis those participants whose
248 primary could be analysed (n=87, **Methods**). The mean time between diagnosis and recruitment
249 was 2.06 years, including 20 participants recruited within 6 months; the range was from a few
250 days to over 10 years. Among these participants, twelve primary tumours had an *NRAS* mutation,
251 eleven had a mutation in *KIT*, nine had a *BRAF* mutation, six had *NF1* mutations, two had multiple
252 hits and 47 were classified as wild-type.
253
254 For analysis of the covariates influencing time to recurrence, only participants without
255 documented recurrence prior to consent were included (n=69). Twenty-four of these participants
256 had a recurrence, occurring at an average 1.56 years after recruitment, while 45 did not have a
257 recurrence and were followed for an average of 3.66 years. Those with a driver mutation (*BRAF*,
258 *NRAS*, *KIT*, *NF1* or multihit) had a significantly higher probability of having earlier recurrences
259 (Log-rank test *P*-value < 0.05) (**Supplementary Table 8**, **Figure 5a**), with *NF1* mutations likely
260 having a stronger effect (**Supplementary Table 9**, **Figure 5b**). These analyses suggest that
261 tumours with a driver mutation have a higher risk of recurrence in any time. To examine this
262 suggestion, we analysed the time until recurrence among the participants who had a recurrence
263 prior to recruitment (n=18). Of these, seven had wild-type tumours and eleven had a driver
264 mutation. The mean time until recurrence among those with wild-type tumours is about twice as
265 long as those with a driver mutation, suggesting once again that there is a higher rate of
266 recurrence among those with a driver mutation (**Supplementary Table 10**, Log-rank test *P*-value
267 < 0.01). No significant relationship was found between the transcriptomic clusters and recurrence.
268
269 For the analysis of covariates influencing overall survival, although driver mutation carriers have
270 a higher risk of death, no significant relationship was found perhaps due to small numbers of
271 patients who carry mutations and have died (Log-rank test *P*-value = 0.21). There were significant
272 differences in overall survival among patients with different transcriptomic cluster tumours, with

Cluster 1 patients having the best overall survival and Cluster 2 patients having the worst (Log-rank test *P*-value < 0.04) (**Supplementary Table 11**, **Figure 5c**), with this statistical difference maintained when controlling for age, sex, and stage at diagnosis (Cox proportional hazards model *P*-value < 0.05) (**Supplementary Table 12**).

Discussion

In this study, we report the analysis of the somatic and transcriptomic profile of 128 acral melanoma samples from Mexican patients, one of the largest cohorts reported for this type of cancer. In our view, this study helps address several research gaps: 1) The underrepresentation of samples of Latin American ancestry in cancer sample repositories[14]: As it has been shown previously, genetic ancestry and environment influence the somatic profile of tumours, with potential impacts on patient management and treatment[11–13], 2) the relative lack of studies of acral melanoma, when compared to other types of the disease, as this type of melanoma constitutes the majority of cases in some low- and middle-income countries (LMICs)[3], and 3) the relative paucity of genomic studies performed and directed from LMICs, such as Mexico.

Most patients in this study had predominantly Amerindian genetic ancestry, which allowed us to perform an analysis of genetic ancestry correlates with somatic mutation profile. We identified a positive correlation between European ancestry and *BRAF* mutation rate (**Figure 1c**). A possible link between European ancestry and $BRAF^{V600E}$ mutation had been described previously[10], and this study provides further confirmatory evidence. Other similar correlations have recently been described for other types of cancer, such as a positive relationship between Native American ancestry and *EGFR* mutation rate in lung cancer[13], and an increased rate of somatic *FBXW7* in African patients compared to European patients[11]. In accordance with this observation, other cohorts of acral melanoma, which studied patients with predominantly European ancestry, have a higher *BRAF* mutation rate than that in this study (*e.g.*, 23% in Australian patients with predominantly European ancestry[7]). These observations should provide the basis for future studies exploring the relationships between ancestry and somatic mutation rate.

We were intrigued to discover that *BRAF*-missense acral melanomas exhibit a more 'CM-like' transcriptome than other genetic subtypes of acral melanoma. One possible explanation is that this gene signature is uniquely downstream of a *BRAF* missense mutation. However, we do not

307 favour this explanation, as the CM specimens used to derive the scoring metric were not selected
308 by genetic subtype and likely included a variety of genetic profiles. An alternative explanation
309 involves the distinct origins of *BRAF*-missense acral melanomas compared to other acral
310 melanomas. In our previous work[24], we identified distinct subclasses of human epidermal
311 melanocytes: a common type enriched in limbs (c-type) and a rare type enriched in volar regions
312 (v-type). We observed that most acral melanomas generally retained a transcriptional signature
313 like v-type melanocytes, while a significant subset appeared more akin to c-type melanocytes[24].
314 The current work indicates that these tumours are more likely to belong to the *BRAF*-missense
315 genetic subtype, suggesting that a subset of volar melanomas might be more accurately classified
316 by cell of origin and/or genetic profile as non-acral CM, rather than bona fide acral melanomas.
317 Future studies could explore the diagnosis of cutaneous melanoma as acral versus non-acral
318 based on molecular signatures rather than solely on anatomic location. The fact that *BRAF*-
319 mutated tumours occur less frequently on patients of non-European ancestry highlights the need
320 to study a diverse set of samples to maximise clinical benefit to all patients. Other observations,
321 such as a tendency for *KIT*-mutated tumours to occur in patients with a higher Amerindian
322 ancestry, are intriguing and will need to be investigated in future studies.
323
324 Additionally, Kaplan-Meier analyses identified that patients with mutations in any driver gene, and
325 especially in *NF1*, had worse recurrence-free survival than those without mutations in these
326 genes. This, to the best of our knowledge, has not previously been reported. Separately, patients
327 with Cluster 1 tumours, which we refer to as a 'mixed', also showed a better prognosis than other
328 patients, which is not surprising given their associated clinical characteristics (lower Breslow
329 thickness, earlier stages at diagnosis, and a tendency for lower mitotic indexes). However, what
330 is surprising is the gene expression profile characteristic of this cluster. More CAFs, CD4+ T cells
331 and endothelial cells were found by deconvolution to be associated to Cluster 1 than other
332 clusters, signatures commonly associated with immunosuppression. Moreover, the genes that
333 were overexpressed by this cluster in comparison with others have roles in tumour invasion, such
334 as *AXL*, *ZEB1*, and others. A possible explanation is that early-stage tumours are associated with
335 immunosuppressive microenvironments, a balance which, in later tumours, may have been tilted
336 in favour of tumour cell growth. Another potential explanation may involve the recently described
337 roles of CAFs in immunostimulation[25]. Patients with Cluster 2 tumours, with a
338 'proliferative/pigmentation' signature showed the worst survival, with an overexpression of genes
339 associated with proliferation and pigmentation. It has previously been observed in a zebrafish
340 model and in TCGA samples that a pigmentation signature also predicts worse survival[26], and, in

341 a recent report by Liu and collaborators[27], AM tumours with a proliferative signature also were
342 associated with worse survival than other tumours. This study both extends and replicates these
343 findings in acral melanoma.
344
345 Overall, we were able to identify novel associations of the germline and somatic profile in AM,
346 genomic-clinical correlates of overall and recurrence-free survival, as well as transcriptional
347 differences in *BRAF*-mutated AMs. This study shows the value of studying diverse populations,
348 allowing us to uncover previously unreported relationships and better understand tumour
349 evolution.

350 **Methods**

351

352 *Patient recruitment and sample collection*

353 The protocol for sample collection was approved by the Mexican National Cancer Institute's
354 (Instituto Nacional de Cancerología, INCan, México) Ethics and Research committees
355 (017/041/PBI;CEI/1209/17) and the United Kingdom's National Health Services (NHS, UK)
356 (18/EE/00076).

357

358 Recruitment of patients and sample collection took place from 2017 to 2019. Patients attending
359 follow up appointments at INCan that had previously been diagnosed with AM were offered to
360 participate in this study, and upon signing a written consent form, were asked to provide access
361 to a formalin-fixed paraffin-embedded (FFPE) sample of their tumour tissue that had been kept at
362 the INCan tumour bank, as well as a saliva or normal adjacent tissue sample. Note that, to help
363 anonymise patient data, in tables and figures patient ages are shown rounded down to the nearest
364 5-year tier and dates are shown in the month/year format. However, all analyses in this work used
365 exact ages and dates. FFPE samples underwent inspection by a medical pathologist to establish
366 whether sufficient tumour tissue was available for exome sequencing. Saliva samples were
367 collected using the oragenDNA kit (DNAGenotek, # OG-500).

368

369 *DNA and RNA extraction*

370 DNA extraction from all saliva samples was performed at the International Laboratory for Human
371 Genome Research from the National Autonomous University of México (LIIGH-UNAM) using the
372 reagent prepITL2P (DNAGenotek, # PT-L2P) and the AllPrep DNA/RNA/miRNA Universal Kit
373 (Qiagen, #80224). DNA and RNA extraction from FFPE samples was performed at the Wellcome
374 Sanger Institute (UK) using the All-prep DNA/RNA FFPE Qiagen kit. Samples with >0 and
375 <0.1ng/μl of total DNA were sequenced using the Sanger Institute's low-DNA pipeline, whereas
376 all samples with DNA >0.1ng/μl were sequenced using the standard pipeline.

377

378 *Genotyping*

379 Genotyping was performed using Illumina's Infinium Multi-Ethnic AMR/AFR-8 v1.0 array at King's
380 College London and Infinium Global Screening Array v3.0 at University College London. Sufficient
381 germline DNA was available for genotyping for 84 out of 96 samples (87.5%). Ancestry estimation
382 was performed using the ADMIXTURE[28] unsupervised analysis together with the
383 superpopulations of the 1000 Genomes dataset. Five superpopulations were identified,

384 corresponding to AFR (Q1), AMR (Q2), SAS (Q3), EAS (Q4), and EUR (Q5) (**Supplementary**
385 **Table 2**, **Supplementary Figure 1**).
386
387 *Exome sequencing and data quality control*
388 FFPE samples, saliva and normal adjacent tissue underwent whole exome sequencing as follows:
389 Exome capture was performed using Agilent SureSelect AllExon v5 probes and paired-end
390 sequencing was performed at the Wellcome Sanger Institute (UK) in Illumina HiSeq4000
391 machines. Control and tumour samples were sequenced to a mean depth of 101x. Alignment was
392 done using BWA-mem[29], using the GRCh38 reference genome. Sequencing quality filters were
393 performed using samtools stats[30] and fastqc[31]. Sample contamination was estimated using the
394 GATK tool CalculateContamination[32]. Concordance between sample pairs was estimated using
395 the Conpair tool[33]. Samples that had <90% similarity with their pair (tumour-normal) or showed a
396 level of contamination above 5% were excluded from the study. After this step, 128 samples
397 remained for further analysis.
398
399 *Somatic SNV calling and identification of driver genes and mutations*
400 Somatic variant calling was done using three different tools (MuTect[34], Mutect2[35] and Varscan2[36]),
401 keeping only the variants identified by a minimum of two out of the three tools. When available
402 within the variant calling tool, strand bias filters were applied. A minimum base quality score of 30
403 on the Phred scale was used. Indel calling was performed using Strelka2[37] using indel candidates
404 identified by the structural variant caller manta[38]. When selecting one sample per patient,
405 preference was given to primaries, and metastases or recurrences were chosen only when a
406 primary had not been collected.
407
408 Significantly mutated genes were identified using the tool dNdScv[39] with default parameters using
409 SNVs identified by two of the three tools used for variant calling and indels identified by Strelka2
410 as input data. Positive selection was considered for genes that had global q-values below 0.1
411 according to the dNdScv tool recommendations.
412
413 *Analysis of correlation between driver mutations and clinical covariates and ancestry*
414 Statistical tests were performed to identify potential clinical and ancestry covariates that correlated
415 with driver mutational status. For tumour stage, sex, ulceration status and tumour site, which are
416 discrete variables, association was tested with contingency Chi-squared tests. No association
417 was found between these variables and driver mutational status. For age at diagnosis, a

418 continuous variable, the model was age ~ *BRAF* status + *NRAS* status + *NF1* status + *KIT* status
419 + Multihit status. Only samples with mutations in these four genes were included for this test.
420 For each of the four driver genes, a logistic regression model was fitted to predict the presence
421 or absence of a mutation on the AM samples using the inferred ADMIXTURE[28] cluster related to
422 the European ancestry component from the 1000 Genomes Project, correcting for age, sex, self-
423 reported socioeconomic level (SE level), and tumour mutational burden (TMB, SNPs + indels), as
424 such: Driver gene status ~ EUR related cluster proportion + age + sex + SE level + TMB. Then
425 the log odds related to the EUR cluster were plotted with their respective confidence intervals.
426 The models were constructed using 81 samples out of the 96, which were those with available
427 genotyping information and with all tested covariables available.
428
429 *Somatic DNA copy number calling*
430 Copy number alterations (CNAs), cellularity and ploidy of the samples were estimated using the
431 tool Sequenza[40]. Samples underwent manual quality filtering, eliminating samples with estimated
432 cellularity below 0.2, and samples with oversegmentation. Whole genome duplication events were
433 considered when samples had an estimated ploidy above 3.6. Significantly affected regions by
434 CNAs were identified using GISTIC2[41]. Amplifications were classified as low-level amplifications
435 when regions had a copy number gain above 0.1 and below 0.9, and as high-level amplifications
436 when regions had a copy number gain above 0.9 according to GISTIC2 values; partial deletions
437 were those with copy number 1, and homozygous deletions as copy number 0. Only peaks with
438 residual q-values < 0.1 were considered as significantly altered. For the analyses of differences
439 in CNA burden by sample group (*e.g.*, mutational status or site of presentation), we used the
440 CNApp tool[42] to generate copy number alteration scores for global (GCS), focal (FCS) and broad
441 (BCS) CNA burden with default parameters. All paired comparisons between groups were
442 evaluated with a Mann-Whitney test.
443
444 *Mutational signature analysis*
445 Mutational matrices were generated using SigProfilerMatrixGenerator[43]. These matrices, with
446 single nucleotide mutations found by at least two of the three variant callers and all insertions and
447 deletions identified by Strelka2, were used as input for mutational signature extraction using
448 SigProfilerExtractor[44] and decomposition to COSMICv3.4[45] and assignment using
449 SigProfilerAssignment[46]. For single base substitutions, the SBS-96 mutational context was
450 selected, and default parameters were used, with a minimum and maximum number of output
451 signatures being set as 1 and 5, respectively. After a first round, samples that had more than 50%

452 of mutations assigned to artifactual signatures were removed, and a second run with the
453 remaining 107 samples and the same parameters was performed. For indel mutational signature
454 analysis, the same 107 samples were used, and the ID-83 mutational context was selected. For
455 copy number signature analysis, all 85 samples with available copy number data were used with
456 default parameters, and selecting the CN-48 context.
457
458 *RNA sequencing and data quality control*
459 Total RNA library preparation followed by exome capture using Agilent SureSelect AllExon v5
460 was performed on Illumina HiSeq 4000 machines on 146 samples. Reads were aligned to the
461 GRCh38 reference genome using the splice-aware aligner STAR[47]. Of these, we focused on the
462 80 samples that came from different patients, that had matching DNA and were primaries for the
463 score analysis (Methods below). We then applied further quality control filters for the consensus
464 clustering analysis: samples were excluded if total read counts were fewer than 25 million, or if
465 the sum of ambiguous reads and no feature counts was greater than the sum of all gene read
466 pair counts. Forty-seven samples remained for downstream analysis. Counts were generated with
467 HTSeq[48]. Transcripts per million (TPM) normalisation was performed and values were
468 $\log_2$(TPM+1) transformed.
469
470 *Acral vs. non-acral cutaneous tumour score*
471 Invasive acral and non-acral cutaneous melanomas were identified and collected as part of the
472 University of Utah IRB umbrella protocol #76927, Project #60, and RNA was extracted and
473 quantified as previously described[49]. A custom NanoString nCounter XT CodeSet (NanoString
474 Technologies) was designed to include genes differentially expressed between glabrous and non-
475 glabrous melanocytes[24,50]. Sample hybridization and processing were performed in the Molecular
476 Diagnostics core facility at Huntsman Cancer Institute. Data were collected using the nCounter
477 Digital Analyzer. Raw NanoString counts were normalised using the nSolver Analysis Software
478 (NanoString Technologies). Normalisation was carried out using the geometric mean of
479 housekeeping genes included in the panel (**Supplementary Table 6**). Background thresholding
480 was performed using a threshold count value of 20. Fold change estimation was calculated by
481 partitioning by acral vs. cutaneous melanoma. The log2 normalised gene expression data were
482 subjected to Principal Component Analysis (PCA) using the PCA function in Prism version 10.2.1
483 (GraphPad Software). PCA was performed to identify the main sources of variability in the data
484 and to distinguish between acral and cutaneous samples.
485

To determine the top differentially expressed genes contributing to the variance between acral melanomas and cutaneous melanomas, the loadings of the second principal component (PC2) were examined. Genes with the highest positive and negative loadings on PC2 were selected as the top 10 and bottom 10 genes, respectively. Log2 expression values of these genes were used to generate a multiplicative score, producing the ratio of acral to cutaneous melanocyte genes. Statistical analyses were performed using Prism version 10.2.1 (GraphPad Software). Differences in acral to cutaneous ratios were assessed using the Mann-Whitney U test.

The acral:cutaneous (A:C) ratio was calculated for each of the 80 primary acral tumours using the method described above. Differences in the A:C gene expression ratio scores between *BRAF* missense mutation-positive and *BRAF/NRAS* wildtype acral melanoma samples were assessed using a Mann-Whitney U test. The same normalisation, scoring method and statistical testing was applied to the 63 transcriptomes from acral melanoma tumours considering *BRAF*-missense (n=13) and wild-type (n=50) in Newell *et al* (2020)[7]. All available samples in this cohort were used, as only one primary had a *BRAF* mutation.

Consensus clustering and deconvolution based on gene expression

To identify molecular subgroups based on transcriptome data, we performed consensus clustering using the Cola R package[51]. Standard preprocessing of the input matrix was performed, including removal of rows in which >25% of the samples had NA values, imputation of missing values, replacement of values higher than the 95th percentile or less than 5th by corresponding percentiles, removal of rows with zero variance, and removal of rows with variance less than the 5th percentile of all row variances. Subsequently, standard statistical metrics were used to assess the number of clusters and the stability of the partitions, including 1-PAC score, concordance and jaccard index, and visual inspection of the consensus matrix through heatmaps visualisations. Afterwards, signature analysis and functional enrichment on the identified clusters were performed.

The EPIC algorithm[52] was used in the R programming environment to perform deconvolution in order to infer immune and stromal cell fractions within AM tumours. We used the TRef signature method with default parameters, which includes gene expression reference profiles from tumour-infiltrating cells. The algorithm generated an absolute score that could be interpreted as a cell fraction.

520 *Survival analyses*

521 Consenting and recruitment of patients started in December 2017 and ended in October 2019.
522 Because of the challenges of recruiting significant numbers of participants with AM, patients
523 diagnosed in earlier years who were still attending follow-up clinics were recruited. To ensue
524 comparability of data, only participants with a primary available for analysis were the subject of
525 focus in analyses of time to recurrence and/or death. In total, 89 participants were recruited whose
526 primary was available for analysis. For two of these participants (PD51948 and PD51972), the
527 date of recruitment was not available and so these are excluded in the following analyses.
528
529 Lifetable analysis and Cox proportion hazards were applied to both recurrence and death. For
530 recurrence, all participants with a recurrence prior to consent were excluded from the analysis
531 and treated as a prospective cohort starting at recruitment. In these analyses, where relevant, the
532 date of last note was changed to the date of death.
533
534 Because of the large number of different driver mutations and the small number with each
535 mutation, we combined the data into two groups including those participants with a mutation in
536 any of the driver genes and those without a mutation in one of these genes ("WildType" tumours).
537
538 For the Cox Proportional Hazards analysis, time to event was measured in days since recruitment
539 and we adjusted by age (in years), sex (F vs M) and stage at diagnosis; within these analyses,
540 tumours with stage 1 or stage 2 were considered "lower stage" while those with stage 3 or stage
541 4 tumours were regarded as "higher stage". Comparisons were on higher vs lower stage.
542

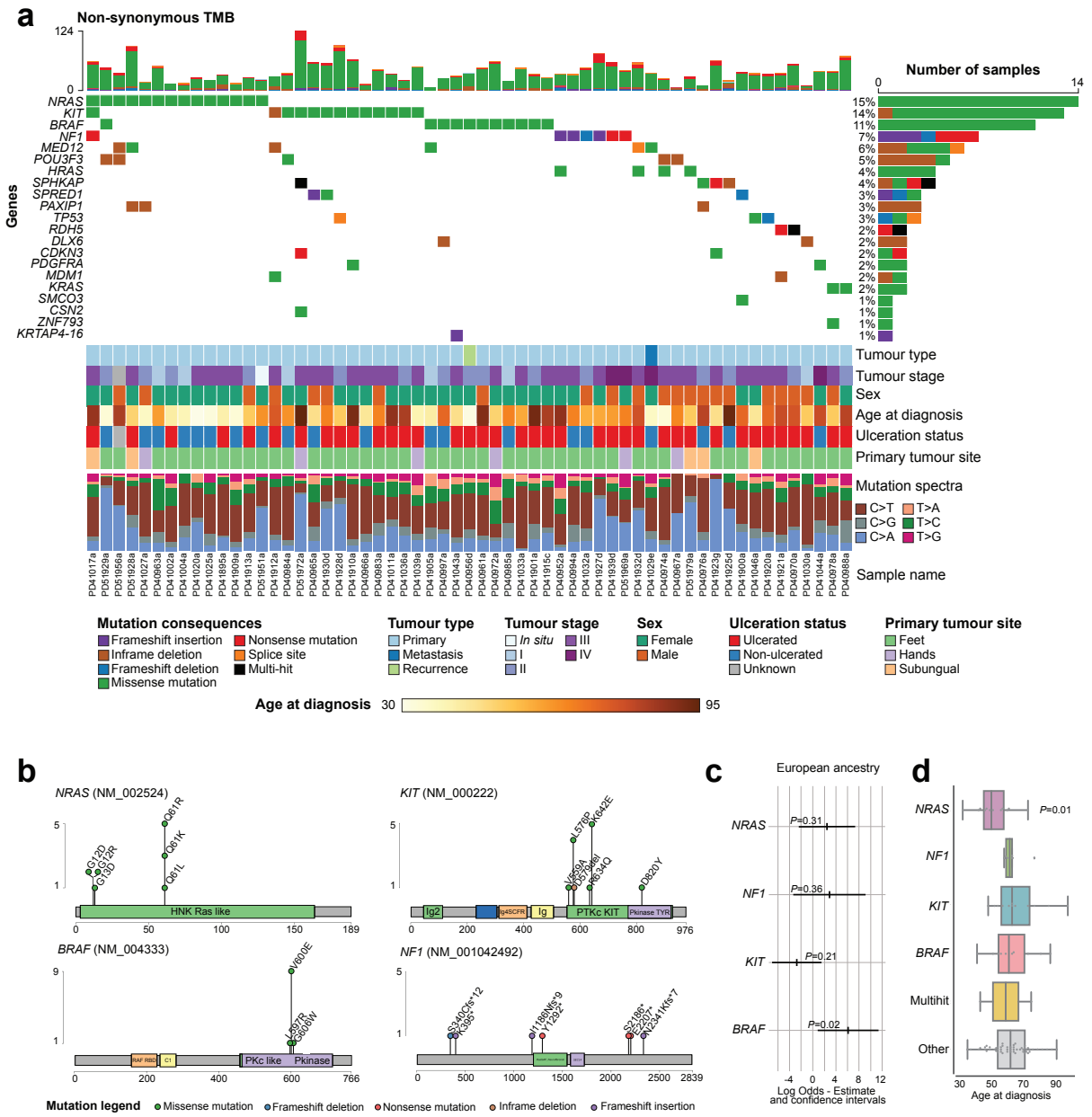


Figure 1. Somatic landscape of acral melanoma in Mexican patients. A) Oncoplot depicting the 21 most mutated genes and their status in the samples with mutations in these (59 samples out of 96, one per patient). Tumour type, tumour stage, sex, age at diagnosis, ulceration status, tumour site and mutational spectra are shown by sample. B) Mutations found in *NRAS*, *BRAF*, *KIT* and *NF1*, which are the most significantly mutated genes. C) A logistic regression model controlling for age, sex, self-reported socioeconomic level 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) Boxplot showing the age of diagnosis of patients classified into genomic subgroups. Statistical significance was assessed by the linear model age ~ *BRAF* status + *NRAS* status + *KIT* status + *NF1* status + Multihit status. 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.

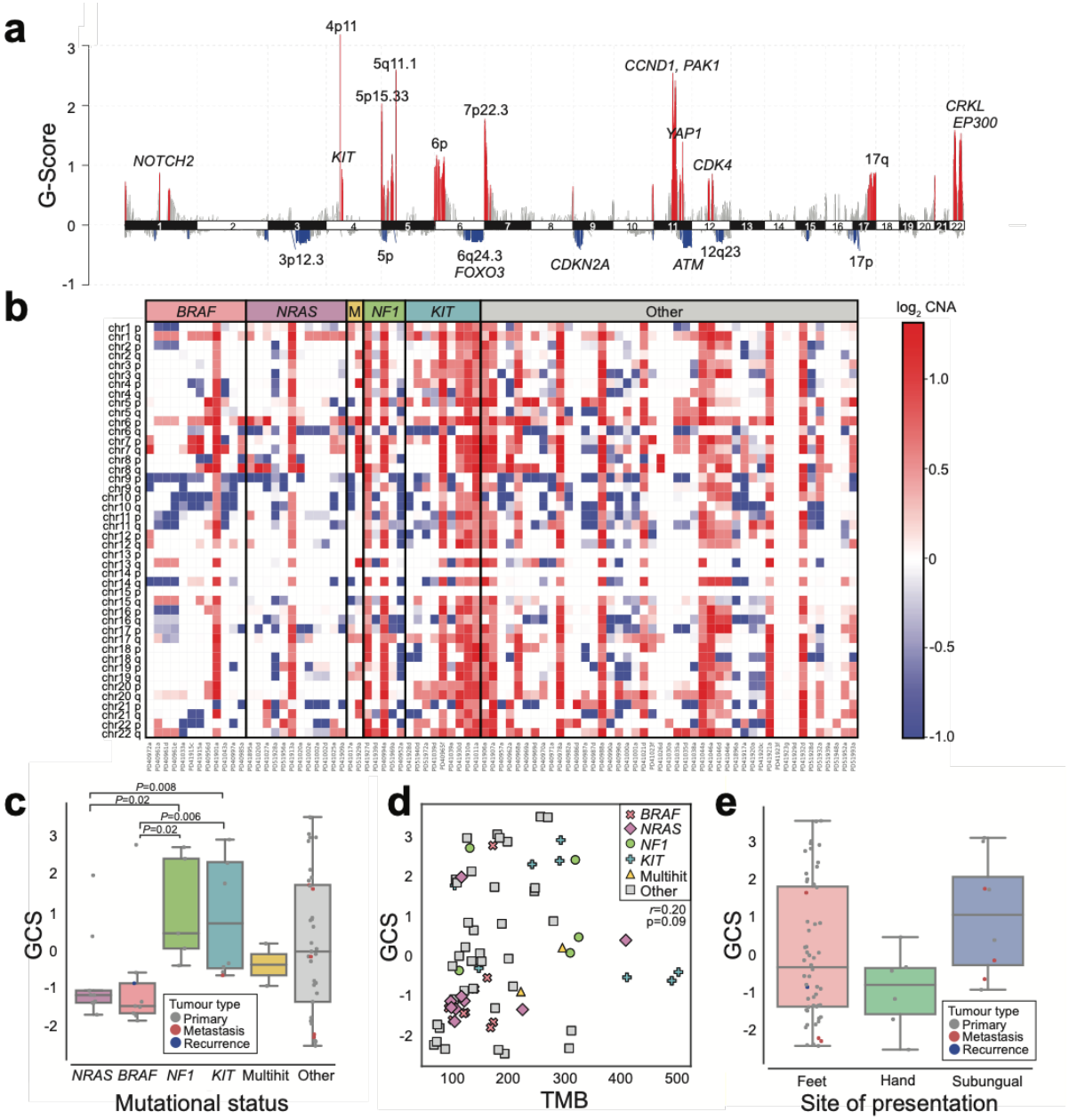
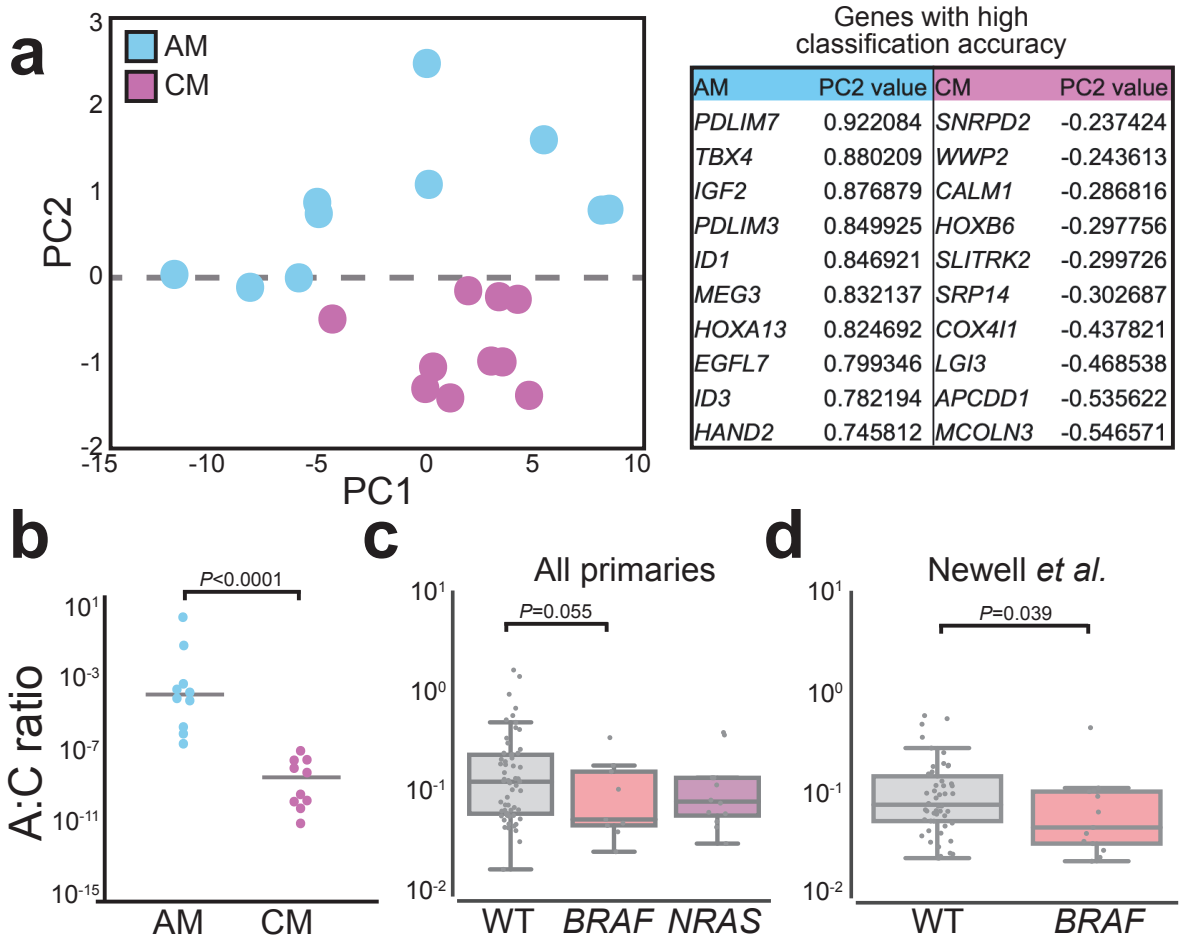


Figure 2. DNA copy number landscape of acral melanoma and molecular and clinical correlates in Mexican patients. A) Regions of amplification (red) and deletion (blue) in 70 acral melanoma samples, one per patient, as identified by GISTIC2. Known drivers, or the chromosomal regions, are shown. B) Heatmap showing regions of amplification (red) and deletion (blue) by sample and chromosomal arm in all samples classified into genomic subgroups. C) Boxplot of global copy-number scores (GCS) of 70 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 70 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 70 samples, one per patient, classified by tumour site. For figures 2c, 2d, 2e one sample (PD40965f) was not plotted as it did not have annotation of sample type, but was included in statistical tests for the paired comparisons. 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.



Genes with high classification accuracy

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*-, *NRAS*-mutated and *BRAF*/*NRAS* 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 < 0.0001). c) Comparison of A:C gene expression ratio in AM samples with different mutation status. Box and whiskers plot comparing three groups: non-*BRAF*/*NRAS* mutated tumours (WT), *BRAF*-missense mutated tumours, and *NRAS*-missense mutated tumours. d) Comparison of A:C gene expression ratio in AM samples with *BRAF* mutations and *BRAF*-wild type 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.

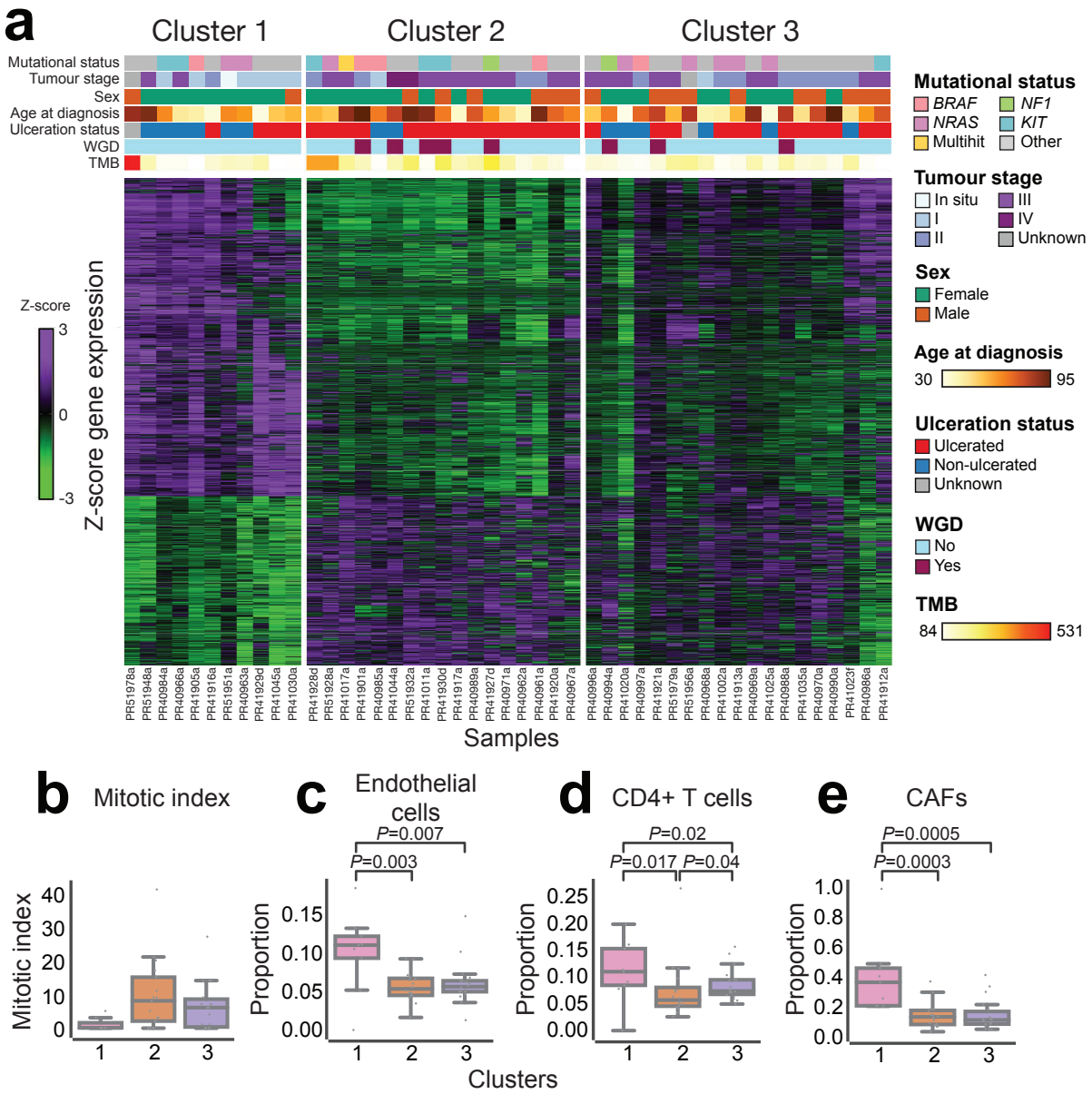


Figure 4. Unsupervised gene expression clustering of primary acral melanoma samples from Mexican patients identifies three main groups. A) Gene expression heatmap showing the 5,439 genes identified as differentially expressed among sample clusters. Samples are in the X axis and genes are in the Y axis. Mutational status and clinical covariates by sample are shown above the heatmap. B) Box plot of mitotic index (Y axis) per sample classified by transcriptional cluster. C) Box plot of endothelial 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.

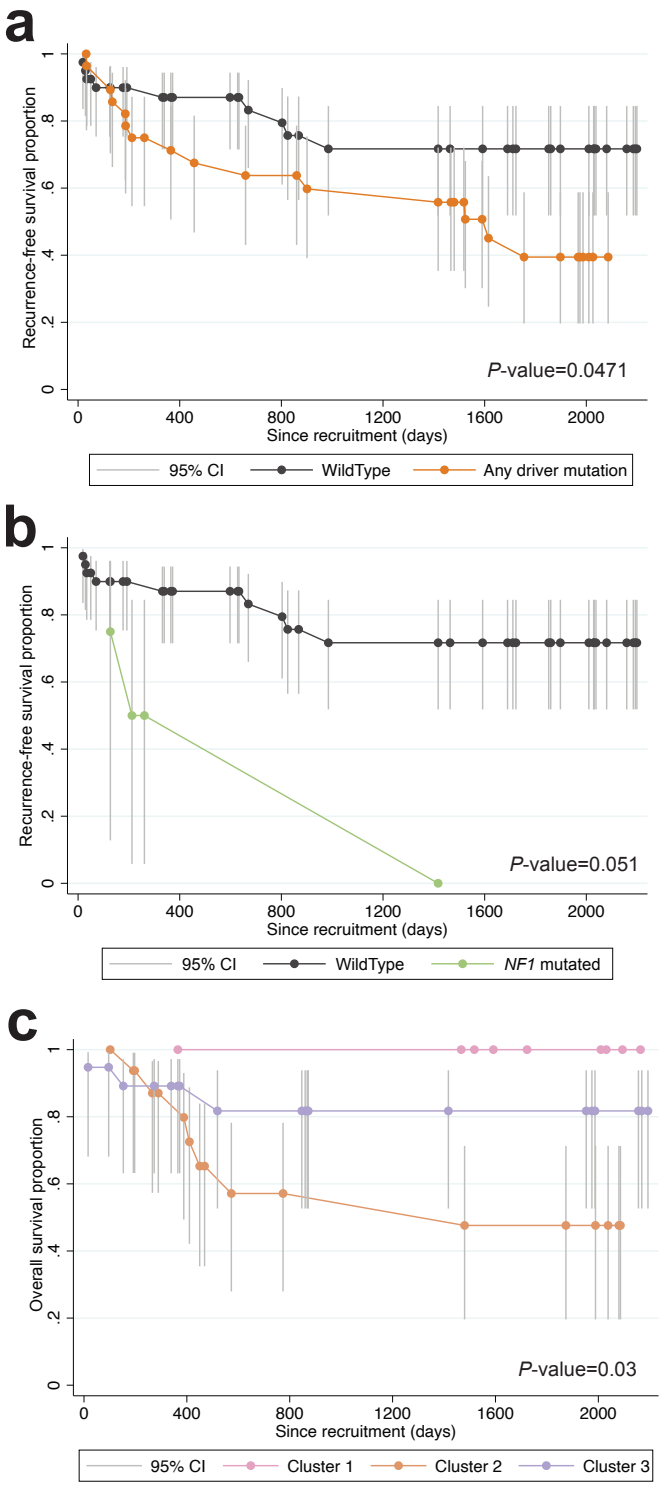
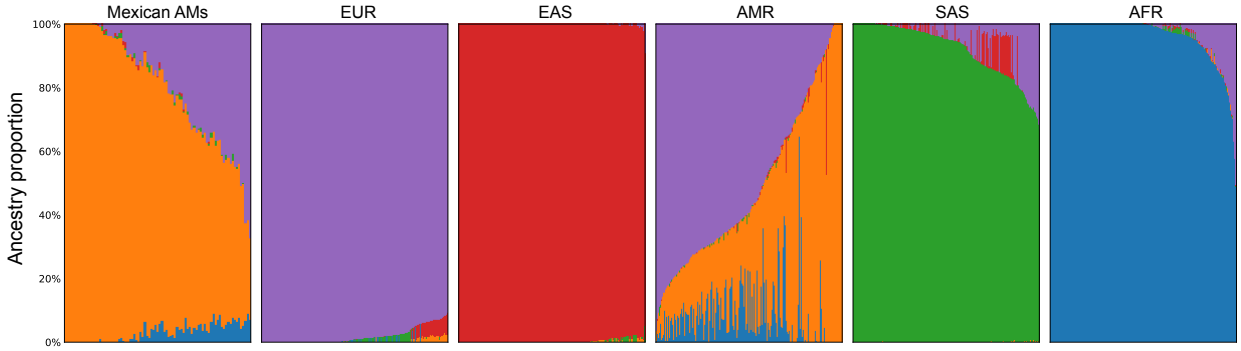
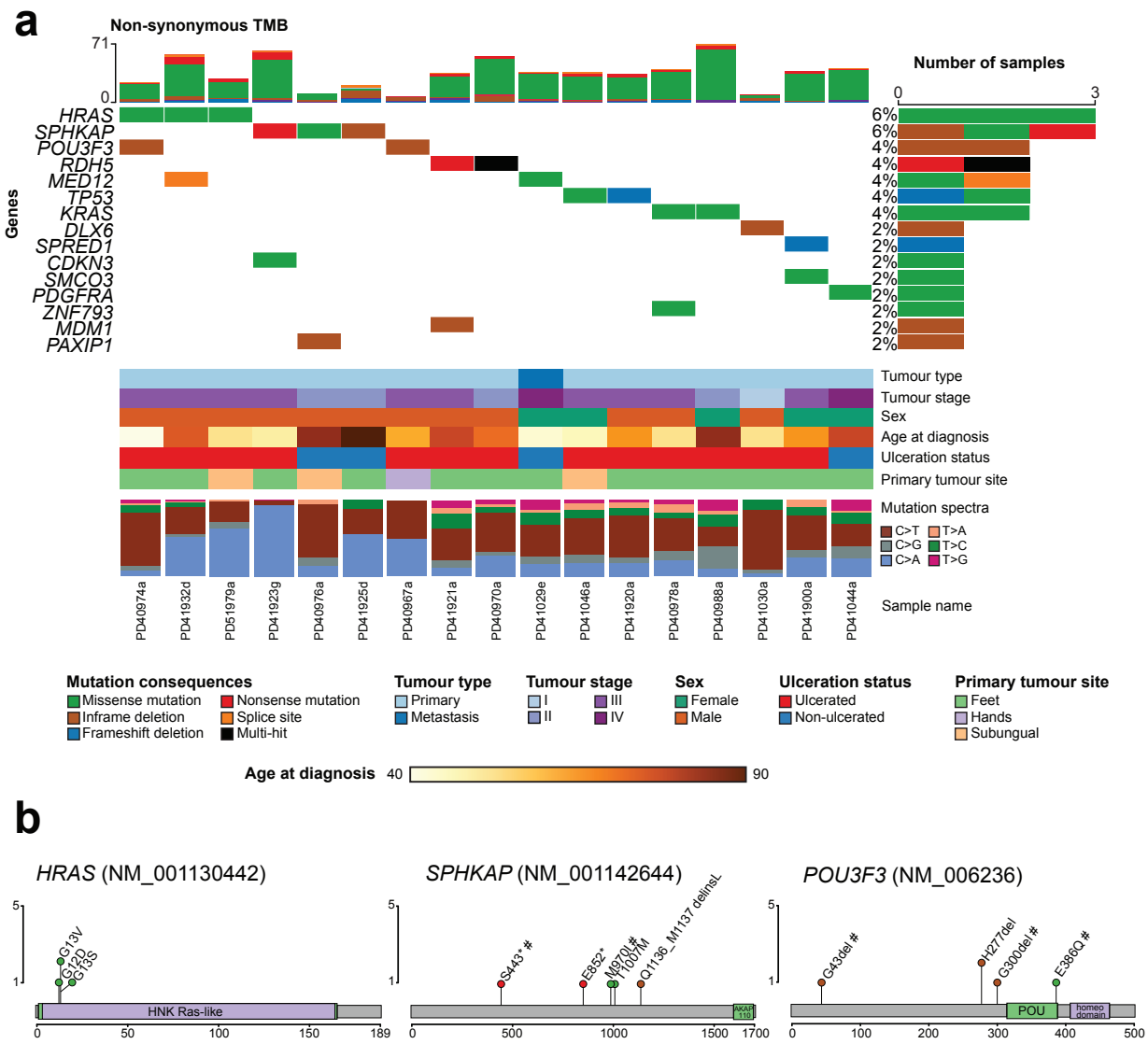


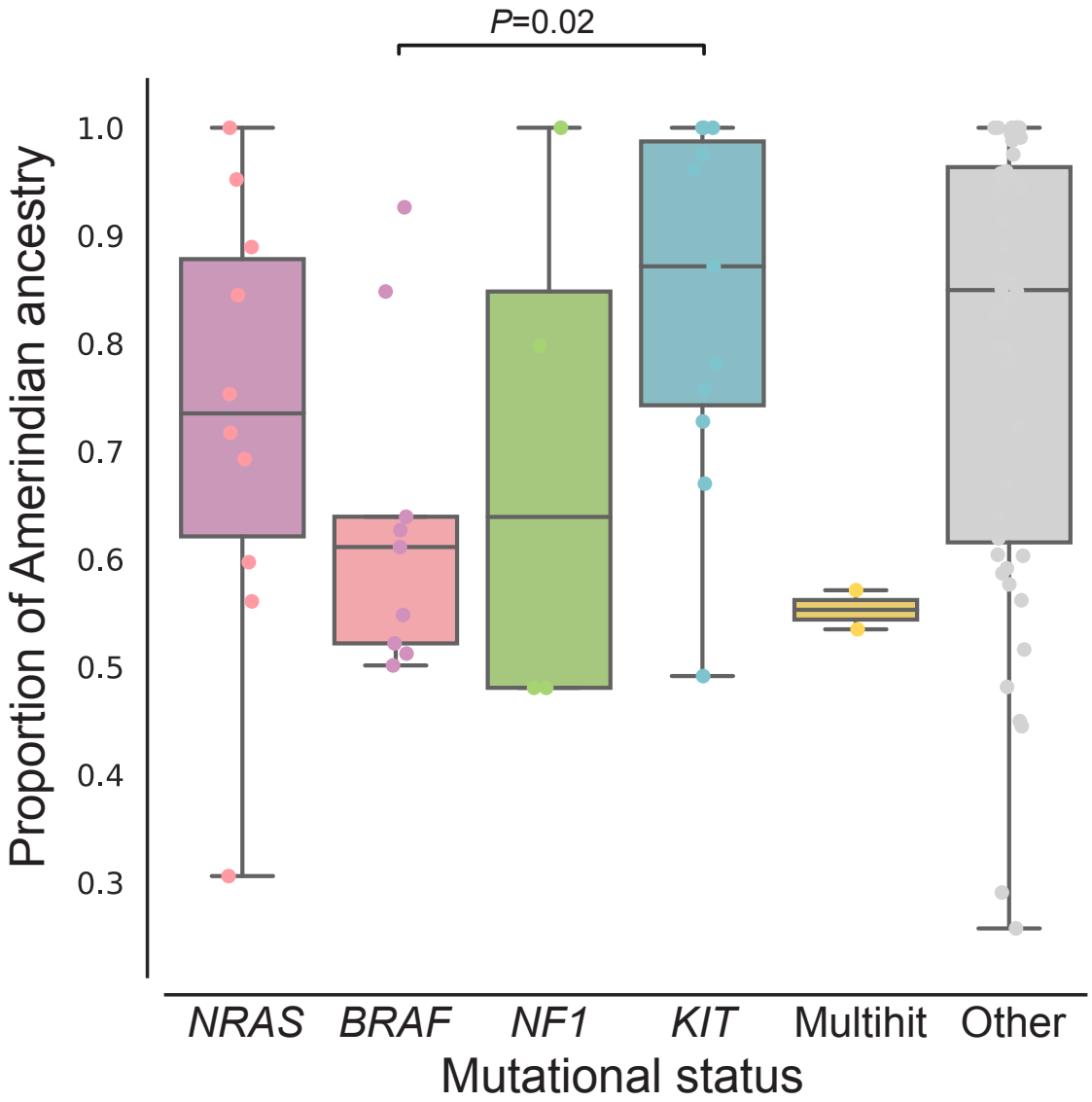
Figure 5. Kaplan-Meier plots of overall and recurrence-free survival for patients by tumour mutational and transcriptional status. A) Recurrence-free survival of patients with and without driver mutations. B) Recurrence-free survival for patients with tumours with and without mutations in *NF1*. C) Overall survival for patients with tumours in each of the three identified transcriptional clusters. *P*-values shown are from Log-rank tests.



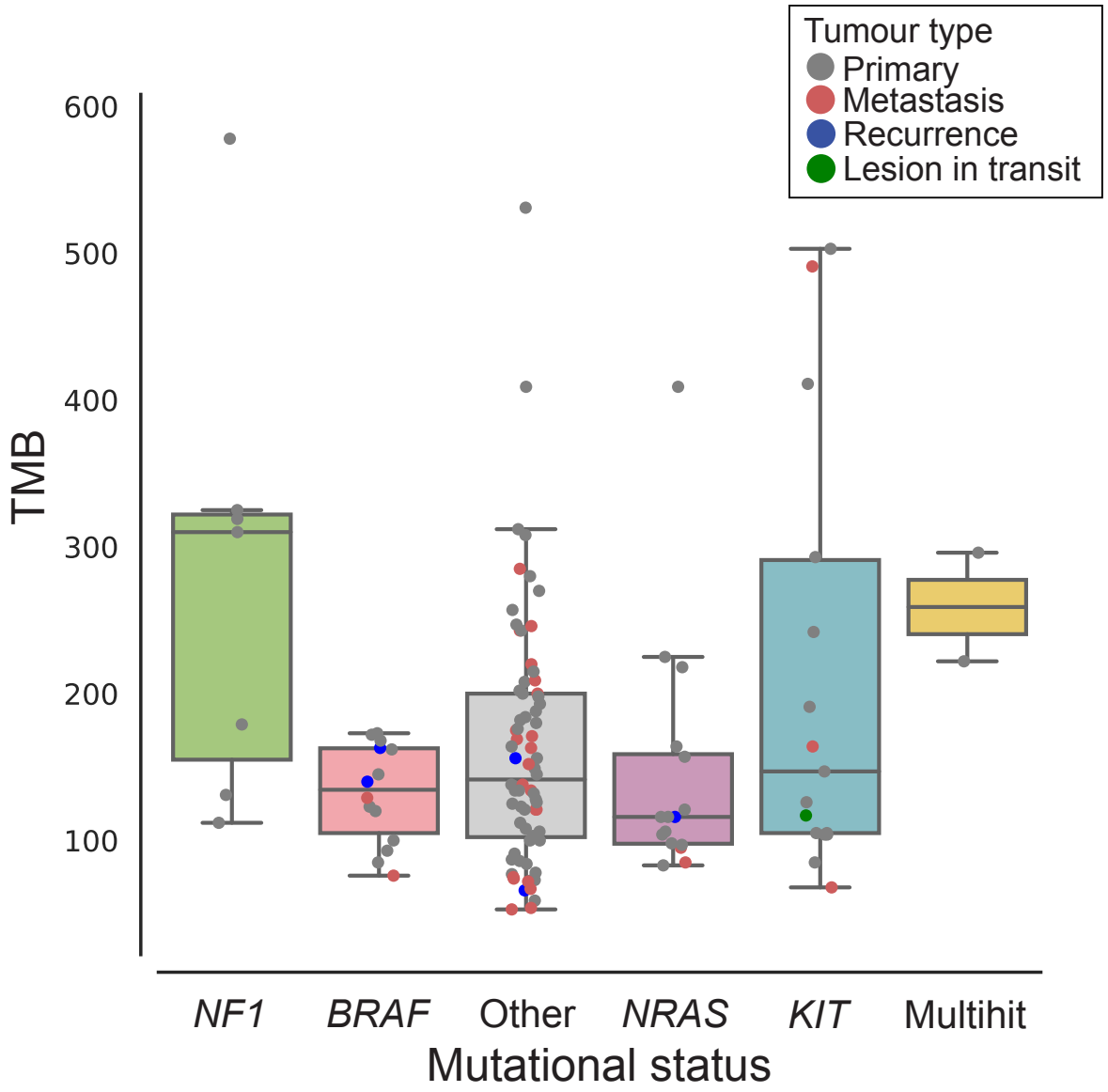
Supplementary Figure 1. Estimation of ancestry proportions per sample together with the superpopulations of the 1000 Genomes dataset. The leftmost panel corresponds to the samples genotyped in this study (n=84). The following panels correspond to the superpopulations in the 1000 Genomes Project. Five superpopulations are plotted, corresponding to African (AFR, blue), Admixed American (AMR, orange), South Asian (SAS, green), East Asian (EAS, red), and European (EUR, purple).



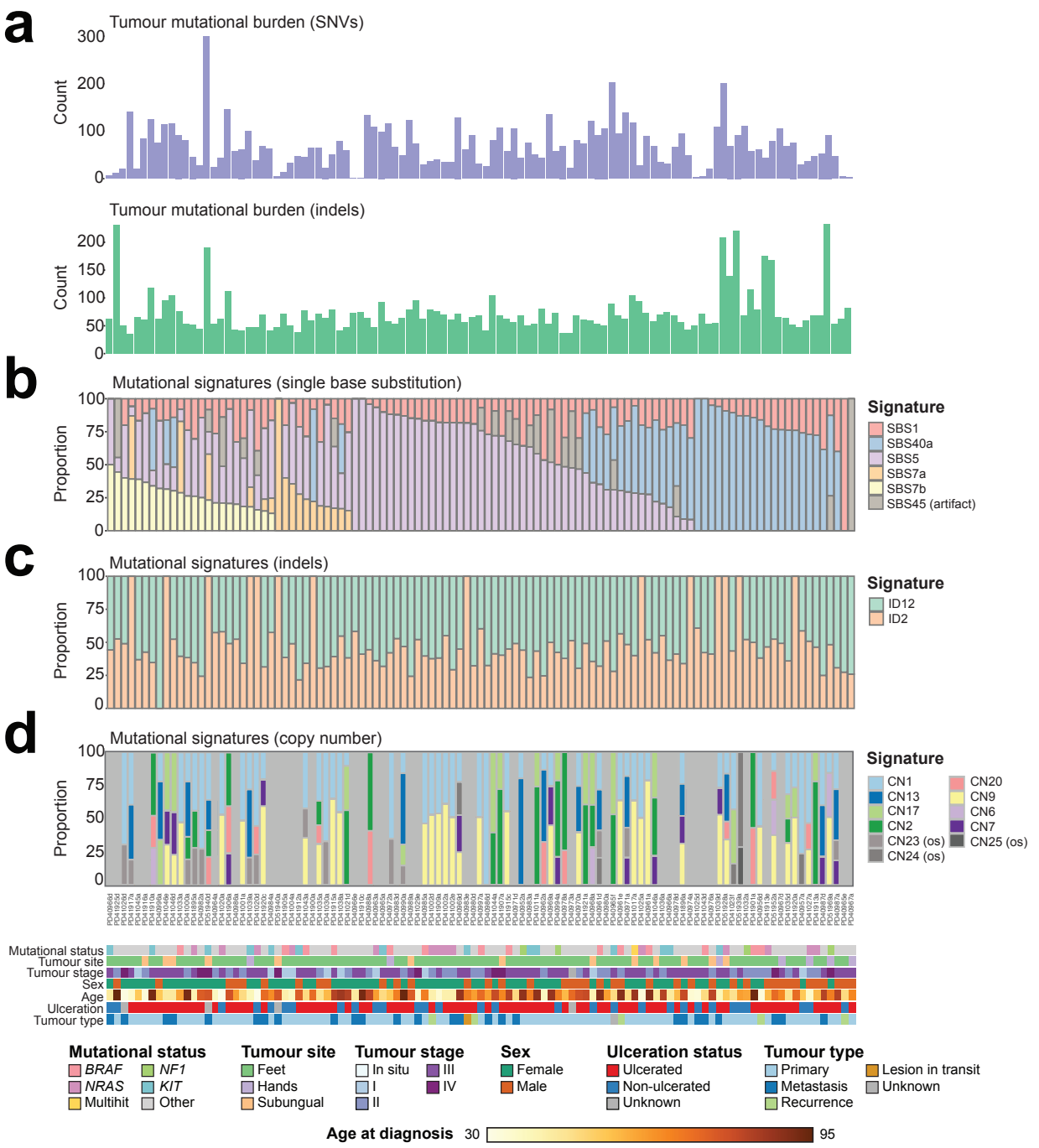
Supplementary Figure 2. Somatic landscape of acral melanoma samples without mutations in *BRAF*, *NRAS*, *NF1* and *KIT* in Mexican patients. a) Oncoplot depicting the 15 most mutated genes and their status in the samples with mutations in these and without mutations in established driver genes (17 samples out of 96, one per patient). Tumour type, tumour stage, sex, age at diagnosis, ulceration status, tumour site and mutational spectra are shown by sample. b) Mutations found in *HRAS*, *SPHKAP and POU3F3*, which are the top mutated genes after the established drivers, including the whole cohort (not only the *BRAF*/*NRAS*/*NF1*/*KIT* wildtype tumours). For *SPHKAP*, mutations with hash symbols are found in the same sample, and are in a *KIT*-mutated sample. For *POU3F3*, mutations with a hash symbol are found in samples with driver mutations.



Supplementary Figure 3. Boxplot of the proportion of Amerindian ancestry among patients classified by genomic subtype. Each dot corresponds to a sample. *P*-value is from a one-tailed Mann-Whitney-Wilcoxon test. 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. 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.



Supplementary Figure 5. Mutational signatures found in acral melanoma samples from Mexican patients. a) Tumour mutational burden per sample, plotted separately for SNVs (top panel) and indels (bottom panel). b-d) Proportions of mutational signatures per sample are shown in stacked bars for single base substitutions (b), indels (c), and copy-number aberrations (d). Known artifacts are shown in darker gray shades. In d), samples with a light gray background did not have data available. Genomic subtypes and clinical characteristics are plotted at the bottom. os: Oversegmentation artifact.

652 **Table legends**

653

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

655 **Supplementary Table legends**

656

657 **Supplementary Table 1.** Clinical and molecular information for patients and samples analysed
658 in this study. Clinical information, along with genomic subtype and classification, TMB, copy
659 number alteration counts, transcriptomic cluster and socioeconomic status are included. For self-
660 reported socioeconomic status, the range is 1-7, with 1 being the lowest and 7 being the highest.
661 To help anonymise the patients, dates are shown as month/year and ages were rounded down
662 to the nearest 5-year tier.

663

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

667

668 **Supplementary Table 3.** Amplification and deletion peaks found in acral melanoma samples.

669

670 **Supplementary Table 4.** Cytobands, q values, location and genes contained within amplification
671 peaks.

672

673 **Supplementary Table 5.** Cytobands, q values, location and genes contained within deletion
674 peaks.

675

676 **Supplementary Table 6.** List of candidate genes from acral and cutaneous melanoma datasets.

677

678 **Supplementary Table 7.** Genes, pathways and biological processes associated to each
679 transcriptional cluster.

680

681 **Supplementary Table 8.** Log-rank test for equality of survivors for recurrence-free survival by
682 mutation status.

683

684 **Supplementary Table 9.** Log-rank test for equality of survivors for recurrence-free survival by
685 genomic subtype.

686

687 **Supplementary Table 10.** Two-sample t-test with equal variance for time to recurrence among
688 patients with recurrences prior to recruitment.

689

690 **Supplementary Table 11.** Log-rank test for equality of survivor functions for overall survival by
691 transcriptomic cluster.

692

693 **Supplementary Table 12.** Cox proportional hazards model evaluating the relationship of
694 transcriptional clusters to overall survival.

695 **Supplementary Information**

696

697 *Mutational signature analyses identify potential sources of mutation and chromosomal*
698 *aberrations*

699 TMB ranges were 1-393 SNVs and 36-292 indels across the whole cohort of samples. Mutational
700 signature analysis was performed on all 128 samples, and after a filtering step where samples
701 with more than 50% of mutations assigned to artifactual signatures were removed, 107 samples
702 remained, with TMB ranges 1-300 SNVs and 36-232 indels (**Supplementary Figure 4a**). A
703 second mutational signature analysis was performed (**Supplementary Figures 4b-d**). Single-
704 base substitution mutational signature analysis across these 107 samples identified signatures
705 SBS1, SBS5, SBS7a, SBS7b, SBS40a and some residual SBS45 (**Supplementary Figure 4b**).
706 The first two of these have been previously classified as clock-like signatures, while SBS1 is
707 related to spontaneous deamination of 5-methylcytosine[19]. SBS7a and SBS7b are related to the
708 UV mutagenic process[6]. SBS40a, which contributes 28.24% of mutations (2005) to the total, is of
709 unknown origin, but has been identified in many cancer types[20]. SBS45 has been recognised as
710 a sequencing artefact[53], which potentially relates to the FFPE origin of these samples.
711 Nevertheless, this analysis is precluded by small numbers of mutations and the FFPE origin of
712 these samples. Indel mutational signature analysis identified two contributing mutagenic patterns,
713 which have been catalogued as ID2 and ID12 (**Supplementary Figure 4c**). ID2 has been
714 proposed to be caused by slippage during DNA replication of the template strand and has been
715 found across many types of cancer[54]. ID12 is of unknown aetiology.

716

717 Copy number signature analysis identified a number of patterns across many samples
718 (**Supplementary Figure 4d**). CN1, which has been associated with a diploid state and CN9,
719 which is potentially caused by local loss of heterozygosity (LOH) on a diploid background,
720 dominated the CN landscape. Nearly a quarter (22%) of samples with a diploid background also
721 showed CN13, which has been associated to chromosomal LOH. Samples with WGD, illustrated
722 by the dominance of CN2, usually also showed CN17, a signature of homologous recombination
723 (HRD) deficiency. No somatic mutations in these samples were found in *BRCA1*, *BRCA2*, *CDK12*,
724 *PALB2* or *FBXW7,* and no other signatures of HRD were found in these samples (possibly due to
725 the small number of point mutations). These analyses illustrate the complexity of these samples'
726 genomes and the heterogeneity of genome compositions across distinct samples.

727

728

729 **References**

730

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

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

736 3. Basurto-Lozada, P. *et al.* Acral lentiginous melanoma: Basic facts, biological characteristics
737 and research perspectives of an understudied disease. *Pigment Cell & Melanoma Research*
738 **34**, 59–71 (2021).

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

741 5. Curtin, J. A. *et al.* Distinct sets of genetic alterations in melanoma. *N Engl J Med* **353**, 2135–
742 2147 (2005).

743 6. Hayward, N. K. *et al.* Whole-genome landscapes of major melanoma subtypes. *Nature* **545**,
744 175–180 (2017).

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

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

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

751 10. Yeh, I. *et al.* Targeted Genomic Profiling of Acral Melanoma. *J. Natl. Cancer Inst.* **111**,
752 1068–1077 (2019).

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

755 12. Nassar, A. H. *et al.* Ancestry-driven recalibration of tumor mutational burden and disparate
756 clinical outcomes in response to immune checkpoint inhibitors. *Cancer Cell* **40**, 1161-
757 1172.e5 (2022).
758 13. Carrot-Zhang, J. *et al.* Genetic Ancestry Contributes to Somatic Mutations in Lung Cancers
759 from Admixed Latin American Populations. *Cancer Discovery* **11**, 591–598 (2021).
760 14. Molina-Aguilar, C. & Robles-Espinoza, C. D. Tackling the lack of diversity in cancer
761 research. *Disease Models & Mechanisms* **16**, dmm050275 (2023).
762 15. Park, S. L., Cheng, I. & Haiman, C. A. Genome-Wide Association Studies of Cancer in
763 Diverse Populations. *Cancer Epidemiol Biomarkers Prev* **27**, 405–417 (2018).
764 16. Campbell, P. J. *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
765 17. Amin, M. B. *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a
766 bridge from a population-based to a more 'personalized' approach to cancer staging. *CA*
767 *Cancer J Clin* **67**, 93–99 (2017).
768 18. Ng, P. K.-S. *et al.* Systematic Functional Annotation of Somatic Mutations in Cancer. *Cancer*
769 *Cell* **33**, 450-462.e10 (2018).
770 19. Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat Genet*
771 **47**, 1402–1407 (2015).
772 20. Senkin, S. *et al.* Geographic variation of mutagenic exposures in kidney cancer genomes.
773 *Nature* **629**, 910–918 (2024).
774 21. Steele, C. D. *et al.* Signatures of copy number alterations in human cancer. *Nature* **606**,
775 984–991 (2022).
776 22. Everall, A. *et al.* Comprehensive repertoire of the chromosomal alteration and mutational
777 signatures across 16 cancer types from 10,983 cancer patients. 2023.06.07.23290970
778 Preprint at https://doi.org/10.1101/2023.06.07.23290970 (2023).
779 23. Yeh, I. & Bastian, B. C. Melanoma Pathology 2.0 – New Approaches and Classification. *Br J*
780 *Dermatol* **185**, 282–293 (2021).

24. Belote, R. L. *et al.* Human melanocyte development and melanoma dedifferentiation at single-cell resolution. *Nat Cell Biol* **23**, 1035–1047 (2021).
25. Tsoumakidou, M. The advent of immune stimulating CAFs in cancer. *Nat Rev Cancer* **23**, 258–269 (2023).
26. Kim, I. S. *et al.* Microenvironment-derived factors driving metastatic plasticity in melanoma. *Nat Commun* **8**, 14343 (2017).
27. Liu, H. *et al.* Integrative molecular and spatial analysis reveals evolutionary dynamics and tumor-immune interplay of in situ and invasive acral melanoma. *Cancer Cell* **42**, 1067-1085.e11 (2024).
28. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–1664 (2009).
29. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio]* (2013).
30. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
31. Andrews, S. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).
32. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
33. Bergmann, E. A., Chen, B.-J., Arora, K., Vacic, V. & Zody, M. C. Conpair: concordance and contamination estimator for matched tumor–normal pairs. *Bioinformatics* **32**, 3196–3198 (2016).
34. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213–219 (2013).
35. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. 861054 Preprint at https://doi.org/10.1101/861054 (2019).

807 36. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in
808 cancer by exome sequencing. *Genome Res* **22**, 568–576 (2012).
809 37. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat*
810 *Methods* **15**, 591–594 (2018).
811 38. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and
812 cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
813 39. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*
814 **171**, 1029-1041.e21 (2017).
815 40. Favero, F. *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor
816 sequencing data. *Ann Oncol* **26**, 64–70 (2015).
817 41. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets
818 of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41 (2011).
819 42. Franch-Expósito, S. *et al.* CNApp, a tool for the quantification of copy number alterations
820 and integrative analysis revealing clinical implications. *eLife* **9**, e50267 (2020).
821 43. Bergstrom, E. N. *et al.* SigProfilerMatrixGenerator: a tool for visualizing and exploring
822 patterns of small mutational events. *BMC Genomics* **20**, 685 (2019).
823 44. Islam, S. M. A. *et al.* Uncovering novel mutational signatures by *de novo* extraction with
824 SigProfilerExtractor. *Cell Genomics* **2**, 100179 (2022).
825 45. Sondka, Z. *et al.* COSMIC: a curated database of somatic variants and clinical data for
826 cancer. *Nucleic Acids Research* **52**, D1210–D1217 (2024).
827 46. Díaz-Gay, M. *et al.* Assigning mutational signatures to individual samples and individual
828 somatic mutations with SigProfilerAssignment. *Bioinformatics* **39**, btad756 (2023).
829 47. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
830 48. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-
831 throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

832 49. Deacon, D. C. *et al.* Classification of Cutaneous Melanoma and Melanocytic Nevi with
833 microRNA Ratios is Preserved in the Acral Melanoma Subtype. *Journal of Investigative*
834 *Dermatology*.
835 50. Weiss, J. M. *et al.* Anatomic position determines oncogenic specificity in melanoma. *Nature*
836 **604**, 354–361 (2022).
837 51. Gu, Z., Schlesner, M. & Hübschmann, D. cola: an R/Bioconductor package for consensus
838 partitioning through a general framework. *Nucleic Acids Res* **49**, e15 (2020).
839 52. Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E. & Gfeller, D. Simultaneous
840 enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife*
841 **6**, e26476 (2017).
842 53. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature*
843 **578**, 94–101 (2020).
844 54. Thatikonda, V. *et al.* Comprehensive analysis of mutational signatures reveals distinct
845 patterns and molecular processes across 27 pediatric cancers. *Nat Cancer* **4**, 276–289
846 (2023).
847
848

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 Carlos S. Flores from the Laboratorio Nacional de Visualización Científica Avanzada and Jair S. García Sotelo, Abigayl Hernández, Eglee Lomelín, Iliana Martínez, Rebeca Muciño, María A. Ávila, Alejandra Castillo and Carina Díaz from the International Laboratory for Human Genome Research, National Autonomous University of Mexico. 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. P.B. is a PhD student from Programa de Doctorado en Ciencias Biológicas, Universidad Nacional Autónoma de México (UNAM), and is supported by Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCyT) (holder no. 562546, scholarship no.762536). This work represents a substantial part of her dissertation.

875 **Author Contributions**

876 P.B.-L., M.E.V.-C., D.C.-I., 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
877 M.C.V.H. performed bioinformatic and statistical analyses. C.M.-A., F.G.A.-G., M.C.-V., R.O.-L.
878 and L. v.d.W. did sample cataloguing and nucleic acid extraction. E.T.D. provided computational
879 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.-
880 P. and D.H.-U. assessed patients and provided access to biological samples. A.H.-M. provided
881 facilities for sample processing and supervised that part of the work. M.J.A., I.F. and M.T.
882 performed sample histopathology. M.D.-G. and L.B.A. supervised the mutational signatures
883 analysis. Y.S.-P. provided access to patient clinical information and supervised that part of the
884 work. G.K.I., R.L.B. and R.M.W. provided data and information that crucially helped the
885 interpretation of the results in this manuscript. D.T.B. performed survival statistical analyses.
886 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
887 manuscript with assistance from P.B., P.A.P., R.L.J.-T. and D.J.A.

888

889 **Code and data availability statement**

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

894

895 **Competing interests**

896 The authors declare no competing interests.

897