The molecular landscape of cutaneous melanoma and its clinical implications

Roy Rabbie
Wolfson College
University of Cambridge

February, 2021

Dissertation submitted for the degree of
Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text. It is not substantially the same as any that I have submitted, or is being concurrently submitted, for a degree or other qualification at the University of Cambridge or any other university. It does not exceed the prescribed limit of 60,000 words.

Roy Rabbie
February 2021

Abstract

Title: The molecular landscape of cutaneous melanoma and its clinical implications

Author: Roy Rabbie

Cutaneous melanoma arises due to the uncontrolled proliferation of melanocytes. It is the deadliest form of skin cancer and its incidence, particularly in younger adults, is rapidly rising worldwide. Over the past decade, multiple large-scale sequencing studies have uncovered the molecular landscape of cutaneous melanoma. In particular, an improved understanding of the melanoma genome and regulation of the immune system have led to the development of effective targeted and immunotherapies that have revolutionised the treatment landscape. Nevertheless, only a subset of patients demonstrate durable responses to these therapies and identifying those patients most likely to benefit remains an important unmet ned.

Melanoma is an aggressive malignancy that often metastasises beyond its primary site. It has a particular propensity to metastasise to the brain and the mechanisms underlying this devastating complication remain incompletely understood. Melanoma risk stratification has traditionally relied on the examination of clinical and pathological features, however it has become clear that traditional staging may fall short in accurately assessing an individual patients' risk. There is therefore a need for robust prognostic biomarkers capable of identifying truly high-risk patients, or patients with biologically indolent tumours, which could be used as an adjunct to conventional clinicopathologic assessments, and would afford a unique insight into the underlying tumour cell biology.

In this thesis, I explore the mutational landscape of cutaneous melanoma focussing on four key clinical cohorts; adolescent cutaneous melanoma, patients with brain metastases, widespread lethal metastatic disease and patients with high-risk primary melanoma. I use a range of high-throughput sequencing technologies to identify the tumour-specific somatic alterations characterising these cohorts and apply detailed multi-site phylogenetic analyses to uncover some of the key evolutionary changes. In the final chapter, I apply deep RNA sequencing to primary melanomas embedded within a prospective phase III clinical trial, to define and validate a gene expression signature identifying primary melanoma patients at higher risk of adverse outcomes.

Together, these studies demonstrate how detailed analyses of molecular sequencing can uncover novel biological insights. It is hoped that further prospective molecular analyses coupled with high-quality experimental validation will pave the way towards novel therapeutic strategies for these patient cohorts.

Acknowledgements

It is with warm and sincere gratitude that I first wish to thank my doctoral supervisor David Adams. It was clear from the outset that Dave would be incredibly supportive in allowing me to pursue my interests, as well guide my personal and professional development. Throughout my PhD he invited me to attend meetings with world experts, supported my attendance at national and international conferences, as well as opportunities to present our work at the highest levels. These instilled in me a belief that our work is important and valued. Together, we pored over countless figures and data, and spent hundreds of hours discussing melanoma research, cancer genomics and its clinical implications. Looking back, it was through these meetings, observing and imitating his rigorous approach to data and scientific integrity, that I gained the knowledge, skills and momentum to pursue this work. There is no doubt that his time and input, as a supervisor and mentor, have left a lifelong impact on me and a determination to pursue a career in cancer research. I was fortunate to be dually supervised by Pippa Corrie and had the opportunity to work within the melanoma clinic throughout my PhD. I am hugely indebted to Pippa for her unwavering support, for always providing a clear clinical focus and for helping guide my academic progression.

I would also like to acknowledge several key members of the Adams' lab who helped support this work. Mamunur Rashid, Vivek Iyer, Kim Wong and Alastair Rust for guiding me through coding and for helping me think critically and constructively about data analyses. James Hewinson and Elizabeth Anderson for their exemplary management of all our projects. Victoria Harle for her advice and help with cloning, and Louise van der Weyden for her guidance and encouragement. I am also incredibly grateful to my senior colleagues in medical oncology; Christine Parkinson, Sarah Welsh and Leila Khoja for giving of their time, including countless patient discussions and truly valuable insights into our work.

I would like to highlight three exceptional external collaborators; Manik Garg from the European Bioinformatics Institute, Naser Ansari-Pour from the Big Data Institute and Peter Ferguson from the Melanoma Institute of Australia. It was an absolute pleasure exchanging ideas and a shared passion for research. I was also extremely fortunate to have had regular contact with their respective senior group leaders including Alvis Brazma (European Bioinformatics Institute), David Wedge (Big Data Institute) and Richard Scolyer (Melanoma Institute Australia). It has been an honour learning from such world-class researchers and I look forwards to future collaborations.

I am incredibly grateful to my dad Jack and brothers Gil & Daniel who have always been a smart, sensible and unwavering support. My wife Nora, whom I met during the start of my first PhD year and has been there every moment since, including through our respective clinical redeployments across two continents. In memory of my late mother Lili Rabbie. Her love, dedication and courage will live in us forever.

Roy Rabbie
February 2021

Preface: Thesis structure

This thesis reflects work carried out through the course of my Cancer Research UK funded Clinical Research Training Fellowship (PhD) in Biological Sciences, based at the Wellcome Trust Sanger Institute in Cambridge. I commenced my academic studies in October 2016. In late March 2020 I interrupted my PhD in response to the call for academic clinical trainees to return to the NHS in the wake of the Covid-19 pandemic. Over the first three months of the pandemic I worked on a 24/7 on-call rota and, due to the ongoing high clinical service demands within the department, since late June 2020 I have continued covering the oncology inpatient/outpatient service within our hospital. In light of my ongoing clinical responsibilities, in November 2020 my mode of PhD study was formally changed from full- to part-time.

During the course of my PhD, I successfully led a number of research projects. These resulted in 4 peer-reviewed primary research papers, including 3 as first-author (1-3) and one as senior/corresponding author (4). I have also published 2 further first-author comprehensive reviews (5, 6) and have made substantial contributions to a number of related research publications within this field (7-10). As I have had the opportunity to write a substantial body of peer-reviewed research and, given the restricted academic time within my final PhD year, I applied to the University Degree Committee for permission to submit my PhD as a collection of published research papers (as described within the University 'requirements for research degrees' (11)). This application was fully supported by my PhD supervisors, clinical training programme director and the Committee of Graduate Studies (CoGS) at the Sanger Institute (supporting letters available on request). On the 7th December 2020, this was unanimously approved by the University Degree Committee (12).

The thesis is structured according to the 'requirements for research degrees' guidelines (11). In particular, this includes a substantial introductory statement summarising the rationale, placing this work in the context of the wider field of study as well as the extent to which these works represent a consistent body of research. This is then followed by the full text and figures for the first four peer-reviewed publications, summarised below (1-4). I have further detailed my contributions to these studies, as well as the contributions of our collaborators. It is important to mention that, although I have only elected to highlight the first four publications within the text of this thesis (1-4), the other six publications were similarly conducted during the course of my PhD and relate to the same area of research (5-10).

I would like to formally thank the University for their support during this very challenging time and take great pride in presenting this independent body of research.

References 1-4, corresponding to chapters 1-4 respectively

1. **Rabbie R**, Rashid M, Arance AM, Sánchez M, Tell-Marti G, Potrony M, Conill C, van Doorn R, Dentro S, Gruis NA, Corrie P, Iyer V, Robles-Espinoza CD, Puig-Butille JA, Puig S, Adams DJ. Genomic analysis and clinical management of adolescent cutaneous melanoma. *Pigment Cell & Melanoma Research*. 2017 May;30(3):307-316. doi: https://doi.org/10.1111/pcmr.12574. Epub 2017 Apr 19. PMID: 28097802; PubMed Central PMCID: PMC5435926.

   As the primary author I critically analysed the somatic mutational data in the context of molecular data from The Cancer Genome Atlas. I further prepared the case histories, the clinical images and summarized the extended children's phenotypes. Myself, together with Mamunur Rashid, performed the bioinformatic analyses. I drafted the manuscript.

   - Mamunur Rashid helped me perform the alignment and variant calling. Stefan Dentro undertook the copy number calls. Vivek Iyer analysed the germline sequencing data.
   - Susana Puig, Remco van Doorn, Marcelo Sánchez, Ana Arance, Carles Conill, Joan Puig-Butille, Gemma Tell-Marti and Miriam Potrony were the external collaborators who provided the samples.

- Remco van Doorn, Pippa Corrie and Carla Daniela Robles-Espinoza reviewed the manuscript and provided expert external dermatological, oncological and genomic inputs respectively.
- David J. Adams provided overall support and was the study lead.

2. **Rabbie R**, Ferguson P, Wong K, Couturier DL, Moran U, Turner C, Emanuel P, Haas K, Saunus JM, Davidson MR, Lakhani SR, Shivalingam B, Long GV, Parkinson C, Osman I, Scolyer RA, Corrie P, Adams DJ. The mutational landscape of melanoma brain metastases presenting as the first visceral site of recurrence. *British Journal of Cancer*. 2020 Oct 7. doi: https://doi.org/10.1038/s41416-020-01090-2. PMID: 33024263.

As the primary author I conceived the project, designed the study, coordinated the sample collections, the clinical and molecular data extraction and the next generation sequencing. I analysed the sequencing data and drafted the paper.

- Peter Ferguson helped design and set-up the project and was an instrumental link to sample collections from Australia and New Zealand.
- Peter Ferguson, Clinton Turner, Patrick Emanuel, Una Moran, Iman Osman, Jodie M. Sanus, Morgan R. Davidson, Sunil R. Lakhani, Brindha Shivalingam, Georgina V. Long and Richard A. Scolyer were the external collaborators who provided the externally-acquired samples and the associated clinical details.
- Kim Wong helped review and perform some of the bioinformatic analyses including somatic variant and copy number calling. Kerstin Haas provided the copy number calls on SKCM-TCGA samples. Dominique-Laurent Couturier independently reproduced the statistical analyses.
- Christine Parkinson and Pippa Corrie provided clinical supervision, including critical review of the manuscript.
- David J. Adams provided overall supervision and was the study lead.

3. **Rabbie R**, Ansari-Pour N, Cast O, Lau D, Scott F, Welsh SJ, Parkinson C, Khoja L, Moore L, Tullett M, Wong K, Ferreira I, Gómez JMM, Levesque M, Gallagher FA, Jiménez-Sánchez A, Riva L, Miller ML, Allinson K, Campbell PJ, Corrie P, Wedge DC, Adams DJ. Multi-site clonality analysis uncovers pervasive heterogeneity across melanoma metastases. *Nature Communications*. 2020 Aug 27;11(1):4306. doi: https://doi.org/10.1038/s41467-020-18060-0. PMID: 32855398; PubMed Central PMCID: PMC7453196.

As the primary author I designed and set-up the study including consenting the patients. I collected the clinical and whole-genome sequencing data, analysed the sequencing data, conducted the validation analyses and drafted the paper.

- Naser Ansari-Pour analysed undertook the phylogenetic analyses and helped construct the evolutionary trees.
- Oliver Cast helped analyse the Affymetrix expression data and performed the Consensus^{TME} clustering. Kim Wong analysed the copy number profiles for WES multi-site cases. Laura Riva ran the mutational signature analyses.
- Doreen Lau, Francis Scott and Ferdia A. Gallagher performed the radiological imaging analyses.
- Kieren Allinson, Mark Tullett, Ingrid Ferreira, Luiza Moore, Julia N. Martintez Gomez and Mitchell Levesque performed the histopathological and IHC analyses.
- Martin Miller and Alejandro Jiminez-Sanchez provided overall supervision on the expression analyses and Consensus^{TME} clustering.
- Leila Khoja, Pippa Corrie, Christine Parkinson and Sarah J. Welsh provided overall clinical supervision.
- David C. Wedge and David J. Adams provided overall study supervision and were the co-study leads.

4. Garg M, Couturier D, Nsengimana J, Fonseca NA, Wongchenko M, Yan Y, Lauss M, Jönsson GB, Newton-Bishop J, Parkinson C, Middleton MR, Bishop T, Corrie P, Adams DJ, Brazma A, **Rabbie R**. Tumour gene expression signature in primary melanoma predicts long-term outcomes. *Nature Communications* 12**,** 1137 (2021). https://doi.org/10.1038/s41467-021-21207-2.

   As the senior and corresponding author, I conceived this project and (together with Alvis Brazma and David J. Adams) co-led this study. I co-ordinated the extraction of the relevant clinical and translational research data and oversaw all aspects of the data analyses. All the manuscript figures and text were co-written by myself, alongside the first author.

   - Manik Garg carried out the machine learning and reproduced the survival analyses.
   - Dominique Laurent-Couturier reviewed and re-ran the statistical analyses.
   - Matt Wongchenko and Yibling Yang co-ordinated the mRNA extraction and RNA sequencing. Nuno A. Fonseca derived the counts from raw sequencing reads.
   - Sarah McDonald, Nikki Stefanos and John Tadross annotated the histopathological slides with the tumour-infiltrating (TIL) scores.
   - Jeremie Nsengimana, Julia Newton-Bioshop, Tim Bishop, Martin Lauss, Goran B. Jonsson, Ismael A. Vergara, Serigne Lo, Felicity Newell, James S. Wilmott, John F. Thompson, Georgina V. Long and Richard A. Scolyer were the external collaborators who provided the data for external validation of the signature.
   - Mark R. Middleton, Christine Parkinson and Pippa Corrie provided senior input on the clinical scope of the project.
   - Alvis Brazma and David J. Adams helped co-led the project.

Additional research studies published during the course of my PhD. Although I have elected not to include these within the text of this thesis, they relate to the same research carried out during my PhD:

5. **Rabbie R**, Ferguson P, Molina-Aguilar C, Adams DJ, Robles-Espinoza CD. Melanoma subtypes: genomic profiles, prognostic molecular markers and therapeutic possibilities. *Journal of Pathology*. 2019 Apr;247(5):539-551. doi: 10.1002/path.5213. Epub 2019 Feb 15. Review. PubMed PMID: 30511391; PubMed Central PMCID: PMC6492003.
   *Awarded the top downloaded paper in the Journal of Pathology 2018-2019 (certification available on request).

6. **Rabbie R**, Adams DJ. Desmoplastic melanoma: C>Ts and NF-κB. *Pigment Cell & Melanoma Research*. 2016 Mar;29(2):120-1. doi: 10.1111/pcmr.12451. Epub 2016 Jan 20. PubMed PMID: 26663830; PubMed Central PMCID: PMC4864442.

7. Din S, Wong K, Mueller MF, Oniscu A, Hewinson J, Black CJ, Miller ML, Jiménez-Sánchez A, **Rabbie R**, Rashid M, Satsangi J, Adams DJ, Arends MJ. Mutational Analysis Identifies Therapeutic Biomarkers in Inflammatory Bowel Disease-Associated Colorectal Cancers. *Clinical Cancer Research*. 2018 Oct 15;24(20):5133-5142. doi: 10.1158/1078-0432.CCR-17-3713. Epub 2018 Jun 27. PubMed PMID: 29950348; PubMed Central PMCID: PMC6193541.

8. Küsters-Vandevelde HVN, Germans MR, **Rabbie R**, Rashid M, Ten Broek R, Blokx WAM, Prinsen CFM, Adams DJ, Ter Laan M. Whole-exome sequencing of a meningeal melanocytic tumour reveals activating *CYSLTR2* and *EIF1AX* hotspot mutations and similarities to uveal melanoma. *Brain Tumor Pathology*. 2018 Apr;35(2):127-130. doi: 10.1007/s10014-018-0308-1. Epub 2018 Feb 23. PubMed PMID: 29476293.

9. Campbell NR, Rao A, Zhang M, Baron M, Heilmann S, Deforet M, Kenny C, Ferretti L, Huang TH, Garg M, Nsengimana J, Montal E, Tagore M, Hunter M, Newton-Bishop J, Middleton MR, Corrie P, Adams DJ, **Rabbie R**, Levesque MP, Cornell RA, Yanai I, Xavier JB, White RM. Cell state diversity promotes metastasis through heterotypic cluster formation in melanoma. bioRxiv 2020.08.24.265140; doi: https://doi.org/10.1101/2020.08.24.265140 **[In review, Cancer Cell].**

10. van der Weyden L, Harle V, Turner G, Offord V, Iyer V, Droop A, Swiatkowska A, **Rabbie R**, Campbell AD, Sansom OJ, Calvo MP, Choudhary J, Ferreira I, Tullett M, Arends MJ, Speak AO, Adams DJ. *In vivo* CRISPR activation screen identifies novel membrane proteins enhancing pulmonary metastatic colonisation. *Communications Biology* 2021. **[*In press,* fully accepted]**.

11. https://www.cambridgestudents.cam.ac.uk/grad-code-of-practice/requirements-research-degrees [accessed November 2020]

12. See confirmation (email) below:

From: ld356 <ld356@cam.ac.uk>
Subject: RE: Dr Roy Rabbie, thesis submission as published papers
Date: 9 December 2020 at 12:13:57 GMT
To: Roy Rabbie <rr13@sanger.ac.uk>
Cc: "Rabbie, Roy" <roy.rabbie@addenbrookes.nhs.uk>, DC Biology <DCBiology@admin.cam.ac.uk>

Dear Roy

Hello – I hope you're well.

I'm just contacting you further to the Degree Committee meeting on 7 December 2020 where your request to submit a thesis by publication was considered.

I'm very pleased to tell you that it was unanimously approved.

All very best,
Lesley

Lesley Dixon
Faculty Administrator (Postgraduate School)
Office of the School of Biological Sciences
17 Mill Lane
Cambridge
CB2 1RX

Contents

Introduction

The cancer genome

Somatic mutations give rise to cancer

Cancer is characterised by uncontrolled cellular division and is caused by the accumulation of genetic alterations resulting from environmental exposures and/or endogenous mutagenesis. These processes confer a selective fitness advantage to a cell lineage that drives clonal expansion [1]. In the multistep process of tumorigenesis first proposed in 1958 [2], fixed mutations in individual cells are transmitted from one generation to another and it is the gradual accumulation of somatic mutations that eventually leads to an established cancer cell [3]. It is well recognised that tumours acquire thousands of mutations in their lifetime, however only a fraction of these mutations contribute to malignant initiation or progression [4]. In particular, it is the *driver* mutations that confer a clonal growth advantage and are positively selected during cancer evolution. The vast majority of driver mutations fall within the protein-coding region of the genome, although only around 600 of the <20,000 protein-coding genes are considered driver mutations [5, 6].

Identifying the driver mutations contributing to the development of cancer is a key step to understanding tumour biology and can include analyses of mutation frequency (comparing the observed mutation rate to the expected background rate [7]) as well as experimental approaches. Oncogenes are defined as driver genes for which the oncogenic function is activating resulting in a new function and these tend to be point mutations, amplifications or deletions and translocations. Tumour suppressor genes are inactivating driver genes, which tend to be focal deletions or nonsense, frameshift or splice-site mutations.

Next generation sequencing of cancer genomes

Over the past two decades, advances in high-throughout (next generation) sequencing (NGS) technologies have allowed for genome-wide screening of genomic alterations, leading to a rich discovery of somatic alterations [8]. Sequencing of exomes (whole exome sequencing, WES) provides access to all protein-coding regions (approximately 1-2% of the entire genome [9]) at base-pair resolution (single base substitutions), while genome sequencing (whole genome sequencing, WGS) also yields mutations in non-coding regions, copy number and structural alterations.

NGS follows a series of core steps (**Figure 1**). In the first step, genomic DNA is fragmented into numerous short segments, usually around 50-300 base-pairs in length. The segments are pulled out using probes complementary to each target region. In the library preparation stage, the DNA segments are then ligated with common adaptor sequences to each fragment end (these are primer recognition sequences - from which subsequent sequencing reactions are initiated). Adapter-ligated fragments are then amplified using PCR (in order to provide enough template copies for sequencer detection) and purified. The library is then loaded onto a flow cell, whereby fragments are hybridized onto a glass slide and captured onto surface-bound complementary oligos. Finally, the unbound library DNA is washed away, and the captured regions are read base-by-base (sequencing can occur from one end only or from both ends, single- vs pair-end reads respectively). The newly characterised raw sequencing reads are then aligned onto a reference genome using bioinformatic software. Following alignment, the differences between the reference genome and the newly sequenced reads can be called, filtered and annotated to identify pathogenic variants. This ‘massively parallel’ approach to sequence data generation (sequencing across millions of fragments (reads) of DNA simultaneously [10]), dramatically decreases the time and cost of data generation and has fundamentally advanced our understanding of cancer biology [9].

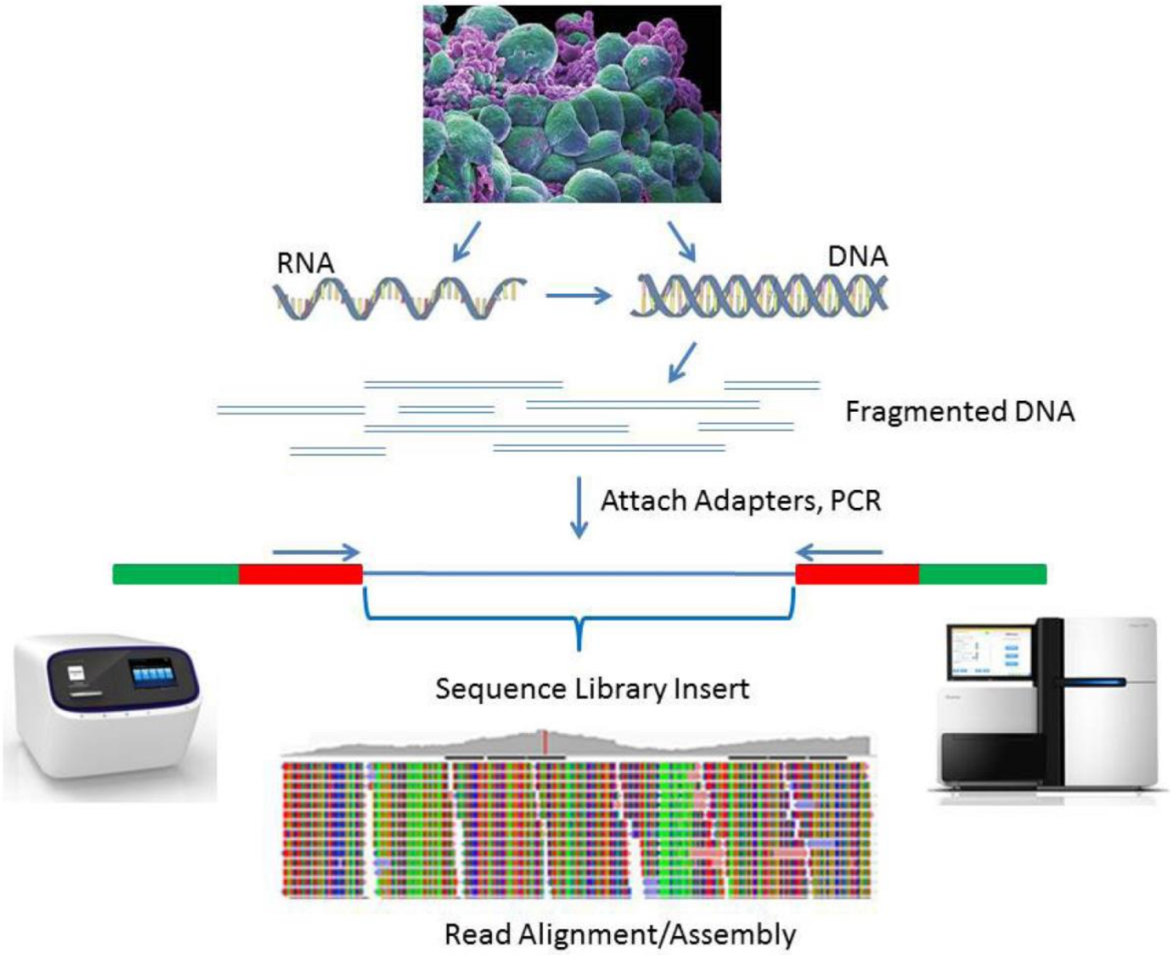


Figure 1. Next generation sequencing library preparation. Extracted DNA (or RNA) is fragmented. DNA fragments are then ligated by sequencing adaptors (which initiate the sequencing steps). The library is then amplified and sequenced. Figure taken from [11].

Distinguishing somatic alterations from inherited variants is of critical importance. To do this, the same sequencing experiment should be performed on a matched non-malignant (germline) DNA sample from the same patient. The germline sample is typically obtained from peripheral blood, but could also be from skin, buccal swabs or adjacent normal tissue. The germline serves as the reference or control, from which the large number variants within the tumour can be subtracted to distinguish somatic mutations [12]. The analysis of the germline DNA can also reveal potentially important cancer susceptibility alterations.

The ability to undertake NGS from a multitude of cancer patients at high resolution has had a profound impact on cancer biology. Importantly, one of the very first large-scale screening analyses across all genes in the mitogen-activated protein kinase (MAPK) pathway in the early

2000s first highlighted the immense potential of this approach, identifying *BRAF* as the mostly commonly mutated cancer driver gene in melanoma [13]. Whole-genome sequencing efforts have further lead to the discovery of recurrent non-exonic mutations in the regulatory region of the telomerase reverse transcriptase (*TERT*) promotor gene [14]. This gene, mutated in 71% of melanomas [14], encodes the catalytic subunit of reverse transcriptase and is critical for the development of melanoma [15] and a number of other cancers [16].

The availability of cost-effective NGS technologies have led to the emergence of large cancer sequencing consortia including The Cancer Genome Atlas (TCGA) [17], International Cancer Genome Consortium (ICGC) [18] and the Pan-Cancer Analysis of Whole Genome Consortium (PCAWG) [19]. These open-access publicly available databases have generated a massive amount of genomic sequencing data, of >50,000 tumours across >30 cancer types, as well as data on the transcriptome, methylome and proteome. There is no doubt that these data have massively propelled research efforts in cancer biology.

Mutational signatures

With the ability to interrogate the full spectrum of somatic mutations, it is possible to unravel some of the processes generating these mutations. Each mutational process can leave a DNA damage imprint or *mutational signature* on the genome. Pan-cancer analyses of WES data demonstrate that melanoma and lung cancer, both carcinogen-exposed malignancies, have higher numbers of somatic mutations than many other common cancers [20]. The overwhelming majority of melanoma mutations are C>T/G>A substitutions, a direct consequence of ultraviolet (UV) mutagenesis [12]. Lung cancer harbours more mutations in unmethylated CpG dinucleotides, consistent with a smoking signature [12]. Similar mutational signatures may uncover deficient DNA damage repair processes or the exposure to certain carcinogens [20, 21]. Studies by Alexandrov *et al* introduced a novel computational framework to consider all 96 trinucleotide mutation contexts (6 classes of single base substitutions and 16 possible sequence contexts), thus far identifying > 30 distinctive mutational signatures (https://cancer.sanger.ac.uk/cosmic/signatures) (**Figure 2**) [20, 22].

The biological mechanisms underpinning many of the mutational signatures have shed light on some of the exogenous and endogenous processes attributed to different cancers. However, obtaining evidence of the proposed aetiology of mutational signatures is not straightforward and many signatures remain unexplained [22]. There have recently been promising data showing that mutational signatures could aid in therapeutic decision-making. Prominent examples of this include the mismatch repair (MMR) signature and profound responses to immunotherapy [23, 24] and homologous recombination deficiency/dependency on other DNA repair pathways and sensitivity to PARP inhibition [25]. In summary, somatic mutational signatures in WGS data can help unravel the aetiological factors for some cancers and may be implicated in cancer therapeutics. Further studies from well-annotated datasets alongside functional perturbations in experimental models will further unravel their clinical and biologic significance [26].

Figure 2. Mutational signatures and their associated causes. The profile of each signature is indicated using the six single base substitution subtypes (C>A, C>G, C>T, T>A, T>C and T>G) and 16 possible sequence contexts, generating 96 possible trinucleotide mutation contexts. Figure taken from [27], under the terms of the Creative Commons Attribution-NonCommercial License.

Cancer heterogeneity

An important outcome of large-scale genomic analyses has been the realisation that inter- (between individuals with the same tumour type) and intra-tumour (within a single tumour) heterogeneity characterises many cancers. Genomic heterogeneity results from Darwinian selection, whereby tumour cells better suited to their local microenvironment proliferate quicker in a process called *clonal expansion* [7]. This expansion of fitter tumour subclones acts as a substrate for tumour evolution.

Somatic alterations detectable in every clone and tumour region are thought to arise early in the natural history of a tumour, so-called *trunk* mutations [28]. As the tumour develops it acquires more alterations leading to the expansion of clones only present in a subset of cells. These *subclonal* alterations make up the branches of the evolutionary tree and represent later stages of tumour evolution [29]. The assignment of variants to their respective clones is based on the cancer cell fraction (CCF). The CCF is estimated by correcting the variant allele frequency (the fraction of sequencing reads carrying an alteration) with tumour purity and copy number alteration (which also impact the allele frequency). The CCF gives a numerical estimation of the cellular prevalence of a mutation. Cells from the clonal cluster have CCF of 1 indicating they are carried in 100% of the tumour cells and cells from the subclonal cluster have CCF <1 [30]. This information can then be used to identify the clonal makeup of a tumour and reconstruct its evolutionary trajectory [31-36]. These approaches have been formalised into algorithms which, when applied to a range of NGS data, can identify clonal lineages and infer the phylogenetic relationships between tumours [34, 36].

Following the pioneering discoveries of Allison and colleagues [37], antibodies that target immune checkpoint proteins (including CTLA-4, PD-1 and others), have shown remarkable activity across a range of malignancies [38]. These observations suggest that the endogenous T-cell repertoire can recognise epitopes within the tumour microenvironment resulting in the control of tumour growth. Research over the years have shown that these T cell epitopes arise as a consequence of somatic DNA alterations, generating new immunogenic peptides called *neoantigens* [39]. As a tumour grows, somatic mutations lead to new neoantigens that could trigger an immune response [39]. As would be expected, the neoantigen load is higher in tumours with a high mutational burden and there is some evidence suggesting that tumours with an elevated mutational load respond better to immunotherapy [40-42]. However, the correlation between neoantigen load and response to immunotherapy is imperfect and it cannot explain all responses [43]. By considering both neoantigen clonality and burden, McGranahan and colleagues have shown that non-small cell lung cancers with a high clonal neoantigen burden and low intra-tumoural heterogeneity have an improved response to immune checkpoint inhibitors [44] and longer disease-free survival [45]. However, the mechanisms linking intratumour heterogeneity with an anti-tumour immune response remain elusive.

Dissecting the tumour microenvironment

It is increasingly clear that understanding alterations within tumour cells only represents part of the picture and that oncogenesis depends on the interactions between tumours and the (non-malignant) multicellular tumour microenvironment (TME) [46]. As such NGS technologies, originally focussing on tumour-centric analyses, are now increasingly being used to unravel the intercellular crosstalk between the tumour and its surrounding multicellular ecosystem.

The characterization of the transcriptome using RNA sequencing (RNA-seq) can be particularly informative. Changes in transcriptional activity and regulation generally underlie cellular diversity and provide important insights into the organisation of immune cells in the TME. Specifically, RNA-seq data can be used to predict patient-specific neoantigens which might elicit an anti-tumour immune response [47]. It can also be used to profile different immune cell subtypes within the microenvironment, for example using either gene-set enrichment analyses [48] or deconvolution tools [49].

Gene expression quantification from bulk-sequencing approaches however only represent average expression profiles of the constituent cells and are influenced by the particular transcriptional profiles, as well as the abundance, of a multitude of different cell types and states within the sample. This becomes particularly relevant when considering the detection limits that might preclude the identification of low-level subclones. Recent progress in single cell sequencing and spatial transcriptomics have allowed more detailed analyses of subpopulations of cells and their relationships within the tumour microenvironment [50], although are beyond the scope of this thesis.

Clinical implications of genome sequencing technologies

Because driver mutations are responsible for oncogenic addiction, targeted therapies directed towards the resulting protein have remarkable therapeutic activity [51]. The era of precision oncology (tailoring anti-cancer treatments to genomic alterations [52]) was born two decades ago, with the approval of imatinib to treat *BCR-ABL-fusion* positive CML [53] and trastuzumab to treat *HER2*-amplified metastatic breast cancer [54]. Because specific genetic variations are only typically present in small patient populations, basket trials (enrolling patients with different histologies harbouring the same genetic alteration), have allowed for broadened access to targeted therapies and molecular screening [55], although the results of many of these studies have thus far been disappointing [56]. Interestingly, NGS testing has also found value in characterising ‘exceptional responders’ [57] and programmes such as the National Cancer Institute exceptional responders initiative are helping to systematically characterise these cases [58].

However, despite the initial successful application of targeted therapies, it is well recognised that only a minority of patients have tumours harbouring therapeutically actionable mutations. Many driver mutations remain incompletely characterised or may currently be considered ‘undruggable’. Critically, most patients commencing targeted therapies eventually develop acquired resistance [52]. Next generation sequencing technologies have helped uncover some of the molecular mechanisms rendering a tumour insensitive to inhibition. These can include; pre-existing or de novo secondary mutations [59, 60], the emergence of resistant subclones [61], the activation of downstream effectors leading to pathway reactivation or switching to alternative pathways [62] amongst others. Efforts to develop rational combination- or serial-therapies are underway, aiming to circumvent some of these resistance mechanisms. In addition to the known biomarkers predicting response to targeted therapies, more complex molecular predictors are emerging. Recently, the FDA have approved TRK fusions and microsatellite instability as histology-agnostic biomarkers for the use of Larotrectinib [63]/entrectinib [64] and pembrolizumab [23, 24] respectively. In conclusion, genomics-based assays are increasingly being used in patient stratification and treatment.

Ongoing challenges and limitations in NGS studies

Despite the enormous progress in NGS technologies, a number of important challenges remain, particularly in the application of these technologies to clinical practice. Firstly, there are important technical challenges relating to the quantity and quality of tumour specimens. This is particularly pertinent to primary melanomas, which are often thin/small and may contain scarce tumour tissue for nucleic acid extraction [65]. Primary melanomas can also be infiltrated with a mixture of non-malignant cells (including keratinocytes, stromal cells, immune cells and others) and a reduced tumour purity could alter the sensitivity of variant detection [66]. The high number of (UV-induced) mutations with low-allelic frequency can also lead to decreased levels of detection sensitivity. This may account for the relative paucity of molecular sequencing data on primary relative to metastatic melanomas in TCGA and other consortia [67, 68] (the opposite picture to most cancers, where many more primaries have been sequenced relative to metastases). In addition, most primary and metastatic tumours are processed and stored as formalin-fixed paraffin-embedded (FFPE) blocks [65], which preserves the cellular architecture of the tumour. However formalin-fixing can lead to cross-linking of the phosphodiester backbone of DNA resulting in fragmentation and the detection of false positive genomic alterations [69, 70]. Improvements in the preparation of nucleic acids for sequencing, including the ability to create NGS libraries from very-low DNA input [71], have helped broaden the NGS application.

An important consideration when designing an NGS experiment is the balance between pursuing comprehensive approaches such as whole-genome sequencing versus more targeted hybrid-capture approaches (including whole-exome and targeted amplicon sequencing). Whole-genome approaches clearly have higher computational requirements, longer turnaround time and higher costs. On the other hand, targeted sequencing might afford deeper coverage (the number of DNA template strands at each given position), which could be particularly beneficial in polyclonal or low-purity samples. In light of these practical considerations, molecular analyses in primary melanomas have historically focussed on targeted gene panels of mutational hotspots [72-76], which only capture a tiny fraction of the genomic complexity. Similar trade-offs in breadth versus depth are also relevant for RNA sequencing [77]. In addition, the inclusion of patient-matched germline DNA is critical to distinguish somatic mutations from inherited germline variants. However this is not always clinically available and in the absence of germline DNA, tumoural variants must be filtered through large databases of common recurrent single nucleotide polymorphisms (SNPs) [78] which can sometimes be problematic, particularly in incurring additional false positive alterations [79].

With the increasing expansions in sequence outputs and the precipitous reduction in sequencing costs, the challenge increasingly lies in the bioinformatic analyses and their meaningful biological interpretation. Challenges also remain in harmonising genomic data attained from different sequencing platforms. This can be particularly problematic in low-depth sequencing or across more complicated genomic regions [80]. The continued open exchange and sharing of data, expertise and technology will help drive further innovation.

Melanoma biology

The origins of melanoma

Melanocytes are neural crest-derived cells found primarily at the basal level of the epidermis where they produce the UV absorbing pigment melanin [81]. Melanoma is marked by the uncontrolled division of melanocytes and the main causative agent is ultraviolet (UV) radiation through its direct damage to DNA [82]. This thesis is on focussed cutaneous melanoma (CM) arising from the skin, much rarer melanoma subtypes include mucosal, acral lentiginous, uveal and conjunctival. We have reviewed the genomics of these subtypes elsewhere [83].

Epidemiology

Melanoma comprises less than 5% of all cutaneous malignancies, yet it accounts for 75% of all skin cancer-related deaths [84]. The incidence of CM has been steadily rising since the 1960's and it is one of the few cancers that continues to rise in incidence worldwide (**Figure 3**) [84]. This is of particular concern given the unusual age demographics. Melanoma disproportionately affects younger patients and is the commonest cause of cancer-related death in 15-30 year-olds [85]. The combination of a rapidly rising incidence and lethality highlights that CM still represents a serious public health concern.

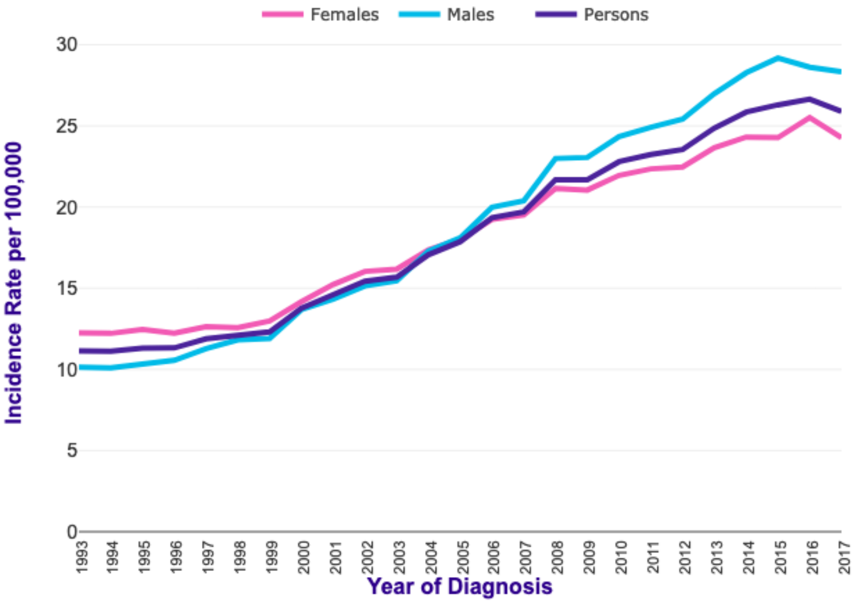


Figure 3. European age-standardised incidence rates in the UK, 1993-2007. Image from [86].

The melanoma genome

Genomic sequencing studies have illuminated the causative role of UV radiation in the pathogenesis of melanoma, identifying UV-induced mutations, including C>T and CC>TT substitutions, in nearly all CM [68].

One of most commonly mutated pathways in melanoma is the mitogen-activated protein kinase (MAPK) pathway [87]. This regulates cell proliferation, apoptosis and survival, and amplification leads to cell cycle dysregulation and uninhibited cellular growth (**Figure 4**) [88]. *BRAF* is one of the key signalling kinases within the MAPK pathway [89]. Somatic mutations in *BRAF* are the commonest genetic alteration in CM, found in 34–41% of cases [13, 90]. The most common and well-characterised *BRAF* mutation (representing >85% of alterations [13]) results from a missense mutation (transversion of T>A) at nucleotide 1799, which causes the substitution of valine (V) for glutamic acid (E) at codon position 600 [91]. Over 90% of activating *BRAF* mutations in melanomas are located in codon 600, less common mutations include V600K (8-20%), V600R (1%), V600M (0.3%) and V600D (0.1%) [92]. Epidemiological analyses have shown that V600E mutations are associated with younger patients and with melanomas on skin sites exposed to intermittent (rather than chronically) sun-induced damage, such as the extremities [93]. In contrast, melanomas lacking *BRAF*V600E mutations generally occur in older patients and are associated with skin sites exposed to chronic sun-damage, such as the head and neck [94].

NRAS is involved in the transduction of extracellular growth signals through the MAPK and the PI3K/AKT pathways, and is the second most frequently mutated gene in CM [67]. *NRAS* mutations occur in 5-15% of cases, usually missense mutations in codons 12, 13 and 61 [67]. Mutations in *NRAS* are generally independent of *BRAF*, though dual expression has been reported [95]. *NRAS*-mutant melanomas are associated with a nodular morphology and tend to occur on chronically sun-damaged skin and are correlate with a higher tumour stage as well as a more aggressive clinical behaviour [96].

NF1 is a tumour suppressor gene mutated in 10-15% of CM [67, 68] and also more commonly found in melanomas associated with chronically sun-exposed skin [97]. Loss of function in *NF1* leads to hyperactivation of the NRAS protein and up-regulated MAPK and PI3K signalling

[98]. Other commonly mutated driver mutations include; *PTEN*, *KIT*, *TP53*, *TERT* promoter, *ARID2* and *CDKN2A* (**Figure 4**) [67, 68].

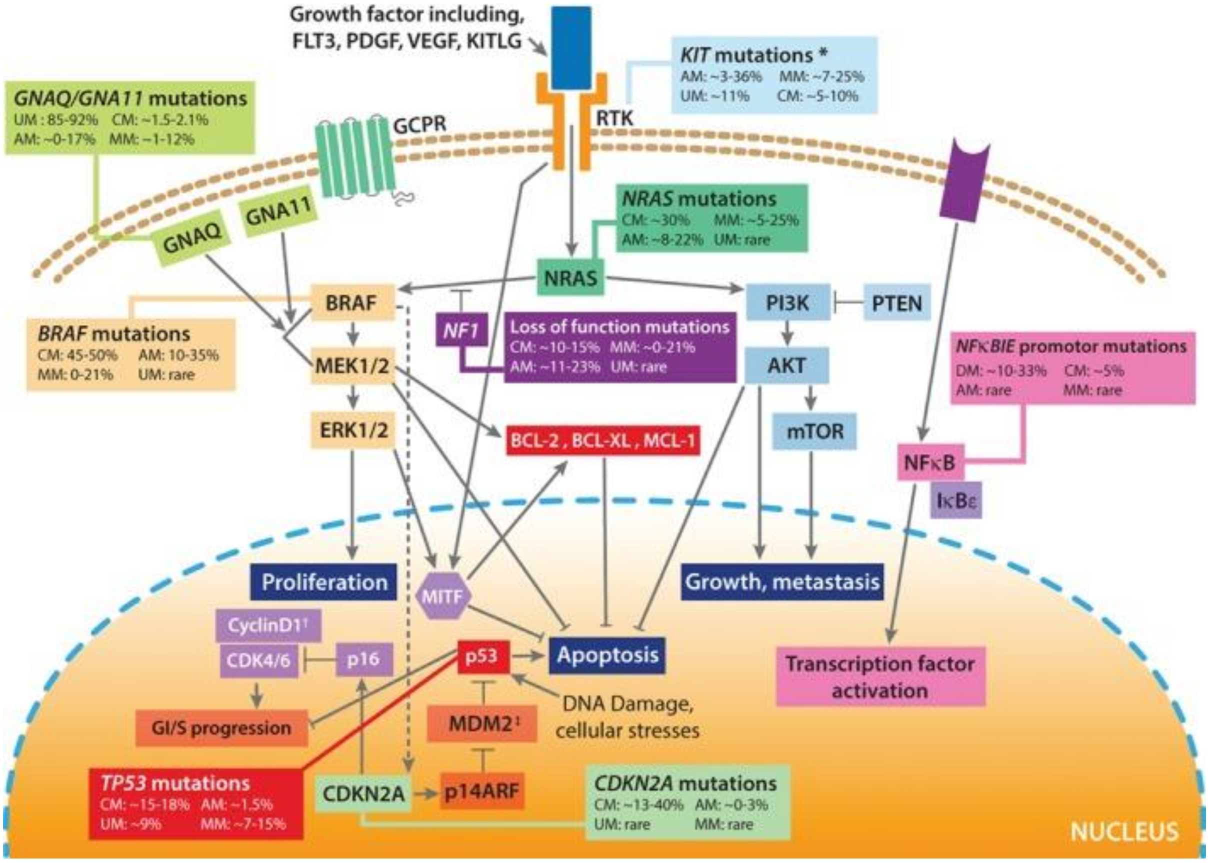


Figure 4. Molecular representation of the mutations associated with the RAS/RAF/MEK/ERK pathways in melanoma, including the MITF signalling cascade. GPCR, G-protein coupled receptor; RTK, receptor tyrosine kinase. Figure reproduced with permission from [83].

From melanocytes to melanoma

Melanocytic naevi are benign proliferations of melanocytes and are both markers of melanoma risk as well as potential precursors of melanoma. In a landmark study, Shain *et al* performed targeted panel sequencing on melanomas and their adjacent precursor melanocytic lesions [76]. They demonstrated that (in these rare cases) *BRAF*V600E mutations were found in the majority of naevi. On the other hand, *BRAF* non-V600E and *NRAS* mutations were found more frequently in intermediate or 'dysplastic' naevi and loss of *CDKN2A* (encoding for p16INK4A) was only identified in established malignant lesions [76]. Based on this evidence, Shain *et al* have proposed a stepwise pathway to transform a naevus to melanoma, with the UV signature identified across all stages of progression (**Figure 5**) [99]. However, the majority of benign melanocytes never actually progress to melanoma and oncogenic *BRAF* mutations alone are not sufficient for melanoma development [100] and more recent data suggest that this may be an over-simplification [87].

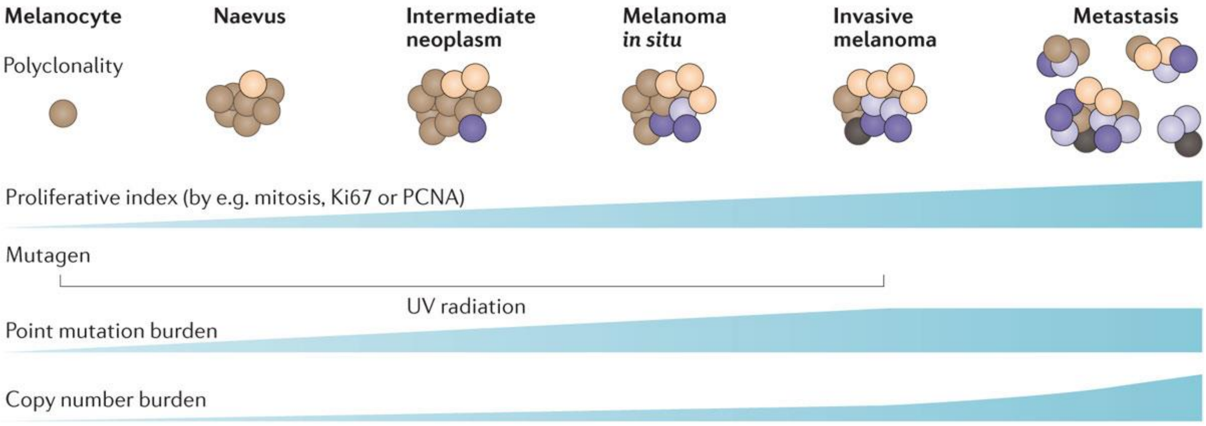


Figure 5. The evolution of melanocytic neoplasms. Showing the accumulation of mutations and copy number alterations. UV-induced mutagenesis was identified across all stages of tumour evolution. Image taken from [99].

Principles of management

Following the initial confirmatory biopsy, surgical excision comprising wide-local excision of the tumour and surrounding healthy tissue is the definitive treatment for the majority of early-stage tumours. Patients with higher-risk primary tumours also undergo a sentinel lymph node biopsy. If melanoma cells are found in a sentinel lymph node, the remaining lymph nodes are also often excised (although there remains considerable debate regarding the optimal management of patients with positive sentinel nodes) [101, 102]. However once metastases develop, surgical interventions alone are no longer curative and systemic therapies are indicated. For the majority of the past 50 years, the prognosis of metastatic melanoma has been poor with very limited effective systemic therapies. However, the treatment landscape fundamentally changed in 2011 with the approval of immune checkpoint inhibitors and *BRAF*-targeted therapies, and these two branches of therapy remain the mainstay of systemic therapy.

Mitogen-activated protein kinase (MAPK) signalling pathway inhibitors targeting either *BRAF* or *MEK* kinases result in response rates exceeding 50% and have revolutionised the treatment of melanoma [103, 104]. However, the majority of patients inevitably develop acquired resistance during therapy [105]. Immune checkpoint inhibitors target the inactivation signals of the immune system leading to an uninhibited endogenous cytotoxic effector T-cell response [106]. There has been a rapid development of immunotherapeutics for the treatment of advanced melanoma and the main agents currently in widespread use target CTLA-4 or PD-1. In particular, anti-PD-1 monotherapy is currently considered the standard first-line therapy in advanced melanoma due to the durable responses, with a well-established efficacy and toxicity profile [107]. Over the past decade, the approval of newer immune and targeted therapies have further improved outcomes for patients with advanced melanoma. These include single or combination immunotherapy agents as well as combinations with existing anti-cancer therapies [108].

With this shift in outcome for advanced melanoma, similar results are seen when these therapies are applied in the adjuvant setting, and adjuvant therapy is now established in stage III and stage IV after complete tumour resection [109]. Adjuvant dabrafenib and trametinib is available for patients with *BRAF*-V600E/K mutations [110] and immunotherapy with pembrolizumab [111] is available for stage III and with nivolumab [112] for stage III and resected stage IV disease. Studies investigating the role of immunotherapy in stage II disease have already been initiated [113].

Despite these important successes, it is clear from both trial and real-world clinical data that durable long-term responses to both targeted and immune therapies are limited to only a small fraction of patients [114, 115]. A better understanding of the resistance mechanisms and the reasons why some patients don’t respond are needed. In particular, there remains an important need for prognostic and predictive biomarkers to better stratify patient risk and select those patients who will most likely achieve a favourable outcome. This is particularly relevant in the current era of effective adjuvant therapies.

Melanoma brain metastases

Adjusting for cancer incidence, melanoma has the highest propensity to metastasise to the brain amongst all cancer types [116]. However the mechanism regulating this organotropism remains poorly understood. Although the true prevalence of brain metastases may be difficult to estimate, studies have shown that approximately 10% of patients with a history of invasive melanoma (across any stage) will ultimately develop clinically apparent brain metastases [117]. However melanoma brain metastases (MBM) complicate 44-50% of patients with stage IV disease [118] and up to 75% of melanoma patients have brain metastasis detected in autopsy series [119].

The initial clinical presentation of MBM comprises headaches, seizures and neurological impairment or can be asymptomatic. They often present as multiple lesions and have a high tendency for haemorrhage. Left untreated, MBM progress rapidly with an average survival of approximately 3 months [120]. Although recent therapeutic developments with immune checkpoint inhibitors have dramatically improved the outcomes for MBM patients [121-123], the prognosis remains dismal. Treatment failure is common and the brain is often the major site of disease progression, even when extracranial disease is well controlled [116].

Early detection of MBM is critical as smaller tumours may be more amenable to surgery/radiosurgery [124] and immunotherapies have demonstrated greatest efficacy in patients with small, asymptomatic brain metastases [123]. It therefore remains crucial to identify patients at an increased risk of brain metastasis. Previous studies examining the clinical factors predictive of MBM have shown an association with depth, ulceration and mitotic rate of the primary tumour as well as location on the head and neck [117, 125, 126]. However, the identification of genomic alterations specific to brain metastases represent an important area of unmet need. An important study undertook whole-exome sequencing of 86 matched primary tumours and brain metastases [127]. The study, which predominantly included patients with breast, lung and renal cancer, reported that 53% of brain metastases had clinically actionable alterations that were not detected in the primary tumours. The identification of brain-specific mutations suggests that clones of metastatic cells growing within the brain may have evolved separately to those within the primary tumour. Further evolutionary analyses across multiple anatomically and temporally distinct brain metastases proposed a divergent evolution at metastatic sites (branched evolution) [127]. However metastatic dissemination to the brain is a complex multistep process that includes the escape of malignant cells from the primary tumour, haematogenous dissemination through the vasculature, passage through the blood brain barrier and successful colonisation of the brain [128]. Once in the brain, in addition to tumoural factors, experimental models are increasingly highlighting the importance of molecular crosstalk between metastatic cells and the surrounding microenvironment [129]. Dissecting the interaction between melanoma brain metastasis and the brain microenvironment is beyond the scope of this thesis, but will be critical to furthering our understanding of the pathogenesis of this devastating complication.

Melanoma prognostic factors

Clinical classification and staging

For the past 40 years, risk stratification in cutaneous melanoma has been based on traditional clinical and pathological features [130]. This is based on the analysis of thousands of melanoma patients over many years and is formalised by the American Joint Committee on Cancer (AJCC) staging, classifying CM into four ordinal disease stages [131]. Staging combines clinical assessment with histological characteristics of the primary tumour. The most powerful of these predictors is Breslow thickness, which stratifies risk based on the invasive depth of melanoma cells in the primary tumour (**Figure 6**). This can be further stratified by the presence of ulceration and the tumour mitotic rate. The most recent (8th) edition of the AJCC staging manual was published in 2017 and included some important refinements supported by the prognostic importance of a number of parameters only available after the 7th edition of the manual was published [132].

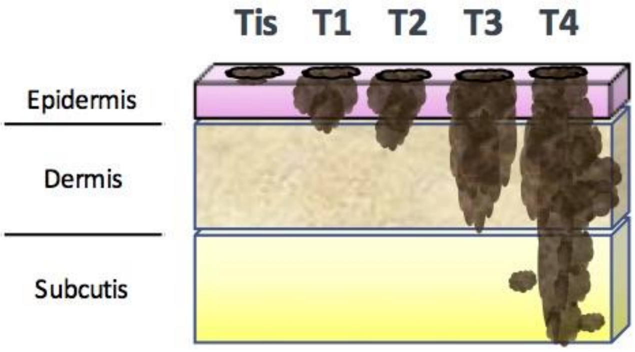


Figure 6. Primary tumour thickness (Breslow depth). This represents melanoma *in-situ*. T1 <0.8mm thickness, T2 >1.0-2.0mm, T3 >2.0-4.0mm and T4 >4.0mm.

Staging allows patients to be stratified into prognostic groups which inform the most appropriate management plans (**Figure 7**). However, due to the subjectivity of some of these visual observations, there remains high inter- and intra-observer variability [133]. In addition, due to the phenotypic and genetic heterogeneity of melanoma, conventional clinical characteristics are also somewhat limited in their ability to accurately predict individual outcomes. As an example, most patients present with early-stage disease and have an excellent prognosis. However, it is well recognised that around 20-30% of early-stage patients will develop a distant recurrence within 5 years (**Figure 7**) [132]. This indicates that there are tumours with a biologic propensity to metastasise that may-not be currently identified through traditional staging.

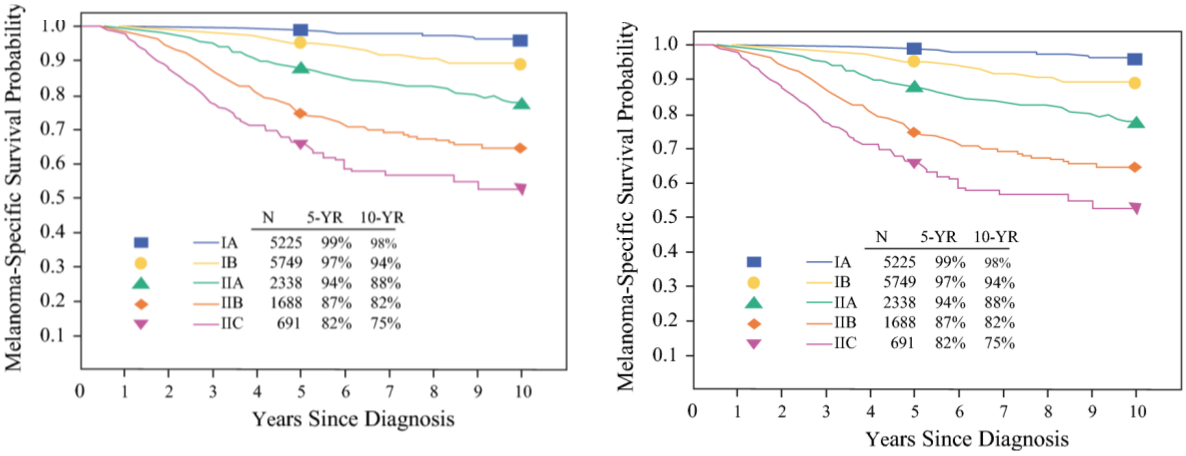


Figure 7. Melanoma-specific survival according to the 8th edition of AJCC. Left panel shows stages I-II and right panel stage III. Figure taken from [132].

Prognostic biomarkers in early-stage disease

Given that the majority of melanomas are diagnosed at an early-stage, it is now understood that around one-fifth of melanoma deaths are recorded in patients initially diagnosed with early-stage disease [134]. In fact, due in part to the high incidence (>70%) of thin primary melanomas (≤1.00 mm Breslow thickness), more patients die of thin melanoma than thick melanoma (>4.00 mm thickness) [134]. The early detection of recurrence or distant metastasis also has therapeutic relevance. Depending on the degree of tumour burden, some patients with oligometastatic disease can be treated with curative-intent surgery. In addition, several studies in stage IV disease have shown that contemporary melanoma therapies (including targeted and immunotherapies) elicit improved responses in patients with a lower tumour burden [121, 135, 136]. Following the approval of adjuvant therapy for high-risk primary melanoma, most patients with stage III (lymph node positive) disease are currently offered adjuvant systemic therapy. However, many of these patients have a favourable risk profile and will likely remain progression-free without therapy. Treatment with immune checkpoint inhibitors can be complicated by immune-mediated toxicity and may necessitate permanent discontinuation or lead to lifelong secondary conditions [137]. A further important area to consider is the wider financial impact associated with offering adjuvant therapy to all melanoma patients [138]. There is therefore a pressing need to identify and validate better prognostic biomarkers that could be used as an adjunct to conventional clinicopathologic assessments. Intensive surveillance can then be focussed on those patients with biologically aggressive tumours and those truly low-risk patients could potentially be spared the toxicities associated with adjuvant therapies. Such robust biomarkers might also offer an opportunity to better understand the relevant intrinsic tumour cell biology.

There have been exciting developments in the study of predictive biomarkers guiding the administration of modern systemic therapies. Identifying biomarkers for immunotherapy in particular requires a comprehensive approach encompassing the complexity of the host immune system (age, gender, gut microbiome, genetic susceptibility to autoimmunity and possibly ethnicity), tumour biology (mutational load) and the tumour microenvironment (PD-L1 expression, T cell infiltrate, dendritic cell activation) amongst other emerging factors [139]. This will require deep analyses of multi-omics datasets and we may ultimately find that no single test has sufficient predictive value. This thesis is focussed primarily on prognostic biomarkers.

Types of prognostic biomarkers

Early research efforts initially focussed on the discovery of protein-based prognostic biomarkers. In particular, serum lactate dehydrogenase (LDH) is well recognised as an important biomarker [140] and is incorporated in the AJCC staging for categorising metastasis [132]. However, its clinical utility is limited to advanced disease stages. Melanomas are among the most immunoreactive of malignancies and are typically infiltrated by a variety of immune cells [141]. A number of studies have shown that a high level of immune cell infiltration is associated with a favourable prognosis in both primary [142] and metastatic melanoma [143]. This has been further assessed using a variety of immunohistochemical markers to define and detect tumour-infiltrating lymphocytes including CD3, CD4, CD8 and FoxP3 [144]. Due to its central importance in predicting response to targeted therapy, *BRAF* mutation status remains the most important molecular biomarker. Although immunohistochemical analyses using the commercially available VE1 monoclonal antibody is highly sensitive and specific for the detection of *BRAF*-V600E melanoma [145], it is the only V600E mutation that can be detected using clinically-approved immunohistochemical methods and DNA-based methods are currently considered the gold standard approach [146].

Over the past decade, the advent of high-throughput sequencing has enabled discovery of numerous prognostic and predictive molecular biomarkers. In an important landmark study, Jonsson used gene expression data from a cohort of stage IV melanoma patients to identify four key molecular subtypes, and validated this in another cohort of stage III and IV patients [147, 148]. In this study, the immune-high subtype, characterised by the increased expression of immune related genes, associated with the best prognosis which was independent of AJCC stage.

Gene expression technologies have also been used to define prognostic biomarkers in early-stage disease. The most widely reported test in this setting is a 31-gene expression signature first developed by Gerami and colleagues [72]. This has since been formalised into a commercially available test (DecisionDx-Melanoma, Castle Biosciences). The test analyses 28 prognostic genes and 3 control genes using RT-PCR technology. Thereby classifying patients with primary melanoma into tumours that are high- or low-risk for metastases, independent of traditional AJCC staging [72]. This test has been studied in a number of retrospective and prospective cohort studies and has been proposed as a risk-stratification tool guiding follow-up frequency and imaging intensity [149-151]. However, these studies have often suffered from small and biased gene panels, the lack of consideration to tumoural and host microenvironment heterogeneity, limited patient numbers and often a lack of long-term follow-up. At the present time, additional data are needed before molecular predictors can be advocated as part of formalised clinical guidelines [152, 153].

The aims of this project

In this thesis we explore the molecular landscape of CM and evaluate whether these data can improve patient stratification. Specifically, we apply a range of next-generation sequencing technologies to primary and metastatic melanomas, focussing our analyses on four key areas of clinical need, for which previous studies may have been confounded by smaller sample sizes and less robust sequencing technologies.

Although melanoma accounts for >10% of all cancers in adolescents and young adults [85], there is little known about the molecular landscape in this population. In Chapter 1 we present a detailed analysis of germline and somatic sequencing in adolescent cutaneous melanomas. Brain metastases pose a particular challenge for genome sequencing studies, not least as obtaining tissue from the central nervous system is invasive and often difficult, particularly in patients who are poor surgical candidates or have tumours within inaccessible sites. In Chapter 2, we present one of the first studies analysing the molecular landscape of melanoma brain metastases. In these first two chapters, we make use of pooled sample collections coordinated through large international consortia, which vastly increased our detection power. Melanomas are known to carry a high mutational load, however they have previously been considered to lack subclonal heterogeneity [87, 154]. In Chapter 3, we apply genome-wide sequencing analyses across multiple temporally and spatially-separated treatment-resistant metastases, uncovering a previously unrecognised layer of genomic heterogeneity. Complementary to the analyses in DNA, tumoural RNA can reveal key important insights into the function of genomic alterations. However, RNA is more labile than DNA, requiring careful handling/quality control and additional computational complexities in the downstream analyses [155]. In Chapter 4 we analyse deep RNA sequencing data from primary and regional lymph node melanoma to study gene expression profiles in early-stage disease. Integrating these findings with clinical outcome data from a well-conducted prospective phase III clinical trial [156, 157], we were able to assess how the molecular architecture of the primary tumour could inform prognostic stratification.

The availability of next-generation sequencing has revolutionised the understanding of cancer biology [158]. It is anticipated that similar studies integrating multi-omic analyses will ultimately establish a central role for molecular-based technologies in understanding melanoma biology and in ultimately translating these findings for the benefit of patients.

References

1. Stratton MR, Campbell PJ, Futreal PA: **The cancer genome**. *Nature* 2009, **458**(7239):719-724.
2. Foulds L: **The natural history of cancer**. *J Chronic Dis* 1958, **8**(1):2-37.
3. Hanahan D, Weinberg RA: **The hallmarks of cancer**. *Cell* 2000, **100**(1):57-70.
4. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C *et al*: **Patterns of somatic mutation in human cancer genomes**. *Nature* 2007, **446**(7132):153-158.
5. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B *et al*: **Comprehensive Characterization of Cancer Driver Genes and Mutations**. *Cell* 2018, **173**(2):371-385.e318.
6. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E *et al*: **COSMIC: the Catalogue Of Somatic Mutations In Cancer**. *Nucleic Acids Res* 2019, **47**(D1):D941-d947.
7. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW: **Cancer genome landscapes**. *Science* 2013, **339**(6127):1546-1558.
8. Meyerson M, Gabriel S, Getz G: **Advances in understanding cancer genomes through second-generation sequencing**. *Nat Rev Genet* 2010, **11**(10):685-696.
9. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C: **Ten years of next-generation sequencing technology**. *Trends Genet* 2014, **30**(9):418-426.
10. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M: **Comparison of next-generation sequencing systems**. *J Biomed Biotechnol* 2012, **2012**:251364.
11. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P: **Library construction for next-generation sequencing: overviews and challenges**. *Biotechniques* 2014, **56**(2):61-64, 66, 68, passim.
12. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordóñez GR, Bignell GR *et al*: **A comprehensive catalogue of somatic mutations from a human cancer genome**. *Nature* 2010, **463**(7278):191-196.
13. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W *et al*: **Mutations of the BRAF gene in human cancer**. *Nature* 2002, **417**(6892):949-954.

14. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA: **Highly recurrent TERT promoter mutations in human melanoma**. *Science* 2013, **339**(6122):957-959.
15. Heidenreich B, Nagore E, Rachakonda PS, Garcia-Casado Z, Requena C, Traves V, Becker J, Soufir N, Hemminki K, Kumar R: **Telomerase reverse transcriptase promoter mutations in primary cutaneous melanoma**. *Nature communications* 2014, **5**:3401.
16. Wu XQ, Huang C, He X, Tian YY, Zhou DX, He Y, Liu XH, Li J: **Feedback regulation of telomerase reverse transcriptase: new insight into the evolving field of telomerase in cancer**. *Cell Signal* 2013, **25**(12):2462-2468.
17. **Comprehensive genomic characterization defines human glioblastoma genes and core pathways**. *Nature* 2008, **455**(7216):1061-1068.
18. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabé RR, Bhan MK, Calvo F, Eerola I, Gerhard DS *et al*: **International network of cancer genome projects**. *Nature* 2010, **464**(7291):993-998.
19. **Pan-cancer analysis of whole genomes**. *Nature* 2020, **578**(7793):82-93.
20. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL *et al*: **Signatures of mutational processes in human cancer**. *Nature* 2013, **500**(7463):415-421.
21. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC *et al*: **Landscape of somatic mutations in 560 breast cancer whole-genome sequences**. *Nature* 2016, **534**(7605):47-54.
22. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN *et al*: **The repertoire of mutational signatures in human cancer**. *Nature* 2020, **578**(7793):94-101.
23. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D *et al*: **PD-1 Blockade in Tumors with Mismatch-Repair Deficiency**. *The New England journal of medicine* 2015, **372**(26):2509-2520.
24. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, Lu S, Kemberling H, Wilt C, Luber BS *et al*: **Mismatch-repair deficiency predicts response of solid tumors to PD-1 blockade**. *Science* 2017, **357**(6349):409-413.
25. Golan T, Hammel P, Reni M, Van Cutsem E, Macarulla T, Hall MJ, Park JO, Hochhauser D, Arnold D, Oh DY *et al*: **Maintenance Olaparib for Germline BRCA-**

Mutated Metastatic Pancreatic Cancer. *The New England journal of medicine* 2019, **381**(4):317-327.

26. Riva L, Pandiri AR, Li YR, Droop A, Hewinson J, Quail MA, Iyer V, Shepherd R, Herbert RA, Campbell PJ *et al*: **The mutational signature profile of known and suspected human carcinogens in mice**. *Nat Genet* 2020, **52**(11):1189-1197.
27. Nakagawa H, Fujita M: **Whole genome sequencing analysis for cancer genomics and precision medicine**. *Cancer Sci* 2018, **109**(3):513-522.
28. Dentro SC, Wedge DC, Van Loo P: **Principles of Reconstructing the Subclonal Architecture of Cancers**. *Cold Spring Harbor perspectives in medicine* 2017, **7**(8).
29. Yap TA, Gerlinger M, Futreal PA, Pusztai L, Swanton C: **Intratumor heterogeneity: seeing the wood for the trees**. *Sci Transl Med* 2012, **4**(127):127ps110.
30. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, Raine K, Jones D, Marshall J, Ramakrishna M *et al*: **The life history of 21 breast cancers**. *Cell* 2012, **149**(5):994-1007.
31. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A *et al*: **Intratumor heterogeneity and branched evolution revealed by multiregion sequencing**. *The New England journal of medicine* 2012, **366**(10):883-892.
32. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, Shafi S, Johnson DH, Mitter R, Rosenthal R *et al*: **Tracking the Evolution of Non-Small-Cell Lung Cancer**. *The New England journal of medicine* 2017, **376**(22):2109-2121.
33. Mitchell TJ, Turajlic S, Rowan A, Nicol D, Farmery JHR, O'Brien T, Martincorena I, Tarpey P, Angelopoulos N, Yates LR *et al*: **Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer: TRACERx Renal**. *Cell* 2018, **173**(3):611-623.e617.
34. Wedge DC, Gundem G, Mitchell T, Woodcock DJ, Martincorena I, Ghori M, Zamora J, Butler A, Whitaker H, Kote-Jarai Z *et al*: **Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets**. *Nat Genet* 2018, **50**(5):682-692.

35. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I, Alexandrov LB, Van Loo P, Haugland HK, Lilleng PK *et al*: **Genomic Evolution of Breast Cancer Metastasis and Relapse**. *Cancer Cell* 2017, **32**(2):169-184.e167.

36. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, Brewer DS, Kallio HML, Hognas G, Annala M *et al*: **The evolutionary history of lethal metastatic prostate cancer**. *Nature* 2015, **520**(7547):353-357.

37. Leach DR, Krummel MF, Allison JP: **Enhancement of antitumor immunity by CTLA-4 blockade**. *Science* 1996, **271**(5256):1734-1736.

38. Ribas A, Wolchok JD: **Cancer immunotherapy using checkpoint blockade**. *Science* 2018, **359**(6382):1350-1355.

39. Schumacher TN, Schreiber RD: **Neoantigens in cancer immunotherapy**. *Science* 2015, **348**(6230):69-74.

40. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, Walsh LA, Postow MA, Wong P, Ho TS *et al*: **Genetic basis for clinical response to CTLA-4 blockade in melanoma**. *The New England journal of medicine* 2014, **371**(23):2189-2199.

41. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS *et al*: **Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer**. *Science* 2015, **348**(6230):124-128.

42. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Foppen MHG, Goldinger SM *et al*: **Genomic correlates of response to CTLA-4 blockade in metastatic melanoma**. *Science* 2015, **350**(6257):207-211.

43. Jardim DL, Goodman A, de Melo Gagliato D, Kurzrock R: **The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker**. *Cancer Cell* 2020.

44. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, Jamal-Hanjani M, Wilson GA, Birkbak NJ, Hiley CT *et al*: **Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade**. *Science* 2016, **351**(6280):1463-1469.

45. Rosenthal R, Cadieux EL, Salgado R, Bakir MA, Moore DA, Hiley CT, Lund T, Tanić M, Reading JL, Joshi K *et al*: **Neoantigen-directed immune escape in lung cancer evolution**. *Nature* 2019, **567**(7749):479-485.

46. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation**. *Cell* 2011, **144**(5):646-674.

47. De Mattos-Arruda L, Vazquez M, Finotello F, Lepore R, Porta E, Hundal J, Amengual-Rigo P, Ng CKY, Valencia A, Carrillo J *et al*: **Neoantigen prediction and computational perspectives towards clinical benefit: recommendations from the ESMO Precision Medicine Working Group**. *Ann Oncol* 2020, **31**(8):978-990.

48. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles**. *Proc Natl Acad Sci U S A* 2005, **102**(43):15545-15550.

49. Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, Waldner MJ, Bindea G, Mlecnik B, Galon J *et al*: **Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy**. *Genome Biol* 2015, **16**(1):64.

50. Suvà ML, Tirosh I: **Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging Challenges**. *Mol Cell* 2019, **75**(1):7-12.

51. Mateo J, Chakravarty D, Dienstmann R, Jezdic S, Gonzalez-Perez A, Lopez-Bigas N, Ng CKY, Bedard PL, Tortora G, Douillard JY *et al*: **A framework to rank genomic alterations as targets for cancer precision medicine: the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT)**. *Ann Oncol* 2018, **29**(9):1895-1902.

52. Shin SH, Bode AM, Dong Z: **Precision medicine: the foundation of future cancer therapeutics**. *NPJ Precis Oncol* 2017, **1**(1):12.

53. Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, Lydon NB, Kantarjian H, Capdeville R, Ohno-Jones S *et al*: **Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia**. *The New England journal of medicine* 2001, **344**(14):1031-1037.

54. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M *et al*: **Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2**. *The New England journal of medicine* 2001, **344**(11):783-792.

55. West HJ: **Novel Precision Medicine Trial Designs: Umbrellas and Baskets**. *JAMA oncology* 2017, **3**(3):423.

56. Chen AP, Eljanne M, Harris L, Malik S, Seibel NL: **National Cancer Institute Basket/Umbrella Clinical Trials: MATCH, LungMAP, and Beyond**. *Cancer J* 2019, **25**(4):272-281.

57. Cummings M, Lehrer EJ, Drabick JJ, Gusani NJ, Trifiletti DM, Zaorsky NG: **Exceptional Responders in Oncology: A Systematic Review and Meta-Analysis of Patient Level Data**. *Am J Clin Oncol* 2019, **42**(8):624-635.

58. Conley BA, Staudt L, Takebe N, Wheeler DA, Wang L, Cardenas MF, Korchina V, Zenklusen JC, McShane LM, Tricoli JV *et al*: **The Exceptional Responders Initiative: Feasibility of a National Cancer Institute Pilot Study**. *J Natl Cancer Inst* 2021, **113**(1):27-37.

59. Gainor JF, Dardaei L, Yoda S, Friboulet L, Leshchiner I, Katayama R, Dagogo-Jack I, Gadgeel S, Schultz K, Singh M *et al*: **Molecular Mechanisms of Resistance to First- and Second-Generation ALK Inhibitors in ALK-Rearranged Lung Cancer**. *Cancer discovery* 2016, **6**(10):1118-1133.

60. Kwak EL, Ahronian LG, Siravegna G, Mussolin B, Borger DR, Godfrey JT, Jessop NA, Clark JW, Blaszkowsky LS, Ryan DP *et al*: **Molecular Heterogeneity and Receptor Coamplification Drive Resistance to Targeted Therapy in MET-Amplified Esophagogastric Cancer**. *Cancer discovery* 2015, **5**(12):1271-1281.

61. McGranahan N, Swanton C: **Biological and therapeutic impact of intratumor heterogeneity in cancer evolution**. *Cancer Cell* 2015, **27**(1):15-26.

62. Van Allen EM, Wagle N, Sucker A, Treacy DJ, Johannessen CM, Goetz EM, Place CS, Taylor-Weiner A, Whittaker S, Kryukov GV *et al*: **The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma**. *Cancer discovery* 2014, **4**(1):94-109.

63. Drilon A, Laetsch TW, Kummar S, DuBois SG, Lassen UN, Demetri GD, Nathenson M, Doebele RC, Farago AF, Pappo AS *et al*: **Efficacy of Larotrectinib in TRK Fusion-Positive Cancers in Adults and Children**. *The New England journal of medicine* 2018, **378**(8):731-739.

64. Doebele RC, Drilon A, Paz-Ares L, Siena S, Shaw AT, Farago AF, Blakely CM, Seto T, Cho BC, Tosi D *et al*: **Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1-2 trials**. *Lancet Oncol* 2020, **21**(2):271-282.

65. Millán-Esteban D, Reyes-García D, García-Casado Z, Bañuls J, López-Guerrero JA, Requena C, Rodríguez-Hernández A, Traves V, Nagore E: **Suitability of melanoma FFPE samples for NGS libraries: time and quality thresholds for downstream molecular tests**. *Biotechniques* 2018, **65**(2):79-85.
66. Wilmott JS, Field MA, Johansson PA, Kakavand H, Shang P, De Paoli-Iseppi R, Vilain RE, Pupo GM, Tembe V, Jakrot V *et al*: **Tumour procurement, DNA extraction, coverage analysis and optimisation of mutation-detection algorithms for human melanoma genomes**. *Pathology (Phila)* 2015, **47**(7):683-693.
67. **Genomic Classification of Cutaneous Melanoma**. *Cell* 2015, **161**(7):1681-1696.
68. Hayward NK, Wilmott JS, Waddell N, Johansson PA, Field MA, Nones K, Patch AM, Kakavand H, Alexandrov LB, Burke H *et al*: **Whole-genome landscapes of major melanoma subtypes**. *Nature* 2017, **545**(7653):175-180.
69. Do H, Dobrovic A: **Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization**. *Clin Chem* 2015, **61**(1):64-71.
70. Williams C, Pontén F, Moberg C, Söderkvist P, Uhlén M, Pontén J, Sitbon G, Lundeberg J: **A high frequency of sequence alterations is due to formalin fixation of archival specimens**. *Am J Pathol* 1999, **155**(5):1467-1471.
71. Van Allen EM, Wagle N, Stojanov P, Perrin DL, Cibulskis K, Marlow S, Jane-Valbuena J, Friedrich DC, Kryukov G, Carter SL *et al*: **Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine**. *Nat Med* 2014, **20**(6):682-688.
72. Gerami P, Cook RW, Russell MC, Wilkinson J, Amaria RN, Gonzalez R, Lyle S, Jackson GL, Greisinger AJ, Johnson CE *et al*: **Gene expression profiling for molecular staging of cutaneous melanoma in patients undergoing sentinel lymph node biopsy**. *J Am Acad Dermatol* 2015, **72**(5):780-785.e783.
73. Haqq C, Nosrati M, Sudilovsky D, Crothers J, Khodabakhsh D, Pulliam BL, Federman S, Miller JR, 3rd, Allen RE, Singer MI *et al*: **The gene expression signatures of melanoma progression**. *Proc Natl Acad Sci U S A* 2005, **102**(17):6092-6097.
74. Winnepenninckx V, Lazar V, Michiels S, Dessen P, Stas M, Alonso SR, Avril MF, Ortiz Romero PL, Robert T, Balacescu O *et al*: **Gene expression profiling of primary cutaneous melanoma and clinical outcome**. *J Natl Cancer Inst* 2006, **98**(7):472-482.

75. Ellerhorst JA, Greene VR, Ekmekcioglu S, Warneke CL, Johnson MM, Cooke CP, Wang L-E, Prieto VG, Gershenwald JE, Wei Q *et al*: **Clinical Correlates of NRAS and BRAF Mutations in Primary Human Melanoma**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2011, **17**(2):229-235.

76. Shain AH, Yeh I, Kovalyshyn I, Sriharan A, Talevich E, Gagnon A, Dummer R, North J, Pincus L, Ruben B *et al*: **The Genetic Evolution of Melanoma from Precursor Lesions**. *The New England journal of medicine* 2015, **373**(20):1926-1936.

77. Levin JZ, Berger MF, Adiconis X, Rogov P, Melnikov A, Fennell T, Nusbaum C, Garraway LA, Gnirke A: **Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts**. *Genome Biol* 2009, **10**(10):R115.

78. Das R, Ghosh SK: **Genetic variants of the DNA repair genes from Exome Aggregation Consortium (EXAC) database: significance in cancer**. *DNA Repair (Amst)* 2017, **52**:92-102.

79. Garofalo A, Sholl L, Reardon B, Taylor-Weiner A, Amin-Mansour A, Miao D, Liu D, Oliver N, MacConaill L, Ducar M *et al*: **The impact of tumor profiling approaches and genomic data strategies for cancer precision medicine**. *Genome Med* 2016, **8**(1):79.

80. Alioto TS, Buchhalter I, Derdak S, Hutter B, Eldridge MD, Hovig E, Heisler LE, Beck TA, Simpson JT, Tonon L *et al*: **A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing**. *Nature communications* 2015, **6**:10001.

81. Kanitakis J: **Anatomy, histology and immunohistochemistry of normal human skin**. *Eur J Dermatol* 2002, **12**(4):390-399; quiz 400-391.

82. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA *et al*: **Mutational heterogeneity in cancer and the search for new cancer-associated genes**. *Nature* 2013, **499**(7457):214-218.

83. Rabbie R, Ferguson P, Molina-Aguilar C, Adams DJ, Robles-Espinoza CD: **Melanoma subtypes: genomic profiles, prognostic molecular markers and therapeutic possibilities**. *J Pathol* 2019, **247**(5):539-551.

84. Schadendorf D, van Akkooi ACJ, Berking C, Griewank KG, Gutzmer R, Hauschild A, Stang A, Roesch A, Ugurel S: **Melanoma**. *Lancet* 2018, **392**(10151):971-984.

85. Weir HK, Marrett LD, Cokkinides V, Barnholtz-Sloan J, Patel P, Tai E, Jemal A, Li J, Kim J, Ekwueme DU: **Melanoma in adolescents and young adults (ages 15-39 years): United States, 1999-2006**. *J Am Acad Dermatol* 2011, **65**(5 Suppl 1):S38-49.

86. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/melanoma-skin-cancer/incidence **[accessed November 2020].**

87. Shain AH, Joseph NM, Yu R, Benhamida J, Liu S, Prow T, Ruben B, North J, Pincus L, Yeh I *et al*: **Genomic and Transcriptomic Analysis Reveals Incremental Disruption of Key Signaling Pathways during Melanoma Evolution**. *Cancer Cell* 2018, **34**(1):45-55.e44.

88. Fecher LA, Amaravadi RK, Flaherty KT: **The MAPK pathway in melanoma**. *Curr Opin Oncol* 2008, **20**(2):183-189.

89. Gray-Schopfer VC, da Rocha Dias S, Marais R: **The role of B-RAF in melanoma**. *Cancer Metastasis Rev* 2005, **24**(1):165-183.

90. Lee JH, Choi JW, Kim YS: **Frequencies of BRAF and NRAS mutations are different in histological types and sites of origin of cutaneous melanoma: a meta-analysis**. *Br J Dermatol* 2011, **164**(4):776-784.

91. Ribas A, Flaherty KT: **BRAF targeted therapy changes the treatment paradigm in melanoma**. *Nat Rev Clin Oncol* 2011, **8**(7):426-433.

92. Long GV, Menzies AM, Nagrial AM, Haydu LE, Hamilton AL, Mann GJ, Hughes TM, Thompson JF, Scolyer RA, Kefford RF: **Prognostic and clinicopathologic associations of oncogenic BRAF in metastatic melanoma**. *J Clin Oncol* 2011, **29**(10):1239-1246.

93. Pracht M, Mogha A, Lespagnol A, Fautrel A, Mouchet N, Le Gall F, Paumier V, Lefeuvre-Plesse C, Rioux-Leclerc N, Mosser J *et al*: **Prognostic and predictive values of oncogenic BRAF, NRAS, c-KIT and MITF in cutaneous and mucous melanoma**. *J Eur Acad Dermatol Venereol* 2015, **29**(8):1530-1538.

94. Menzies AM, Haydu LE, Visintin L, Carlino MS, Howle JR, Thompson JF, Kefford RF, Scolyer RA, Long GV: **Distinguishing clinicopathologic features of patients with V600E and V600K BRAF-mutant metastatic melanoma**. *Clin Cancer Res* 2012, **18**(12):3242-3249.

95. Chiappetta C, Proietti I, Soccodato V, Puggioni C, Zaralli R, Pacini L, Porta N, Skroza N, Petrozza V, Potenza C *et al*: **BRAF and NRAS mutations are heterogeneous and not

mutually exclusive in nodular melanoma. *Appl Immunohistochem Mol Morphol* 2015, **23**(3):172-177.

96. Jakob JA, Bassett RL, Jr., Ng CS, Curry JL, Joseph RW, Alvarado GC, Rohlfs ML, Richard J, Gershenwald JE, Kim KB *et al*: **NRAS mutation status is an independent prognostic factor in metastatic melanoma**. *Cancer* 2012, **118**(16):4014-4023.

97. Wiesner T, Kiuru M, Scott SN, Arcila M, Halpern AC, Hollmann T, Berger MF, Busam KJ: **NF1 Mutations are Common in Desmoplastic Melanoma**. *The American journal of surgical pathology* 2015, **39**(10):1357-1362.

98. Krauthammer M, Kong Y, Bacchiocchi A, Evans P, Pornputtapong N, Wu C, McCusker JP, Ma S, Cheng E, Straub R *et al*: **Exome sequencing identifies recurrent mutations in NF1 and RASopathy genes in sun-exposed melanomas**. *Nat Genet* 2015, **47**(9):996-1002.

99. Shain AH, Bastian BC: **From melanocytes to melanomas**. *Nat Rev Cancer* 2016, **16**(6):345-358.

100. Gray-Schopfer V, Wellbrock C, Marais R: **Melanoma biology and new targeted therapy**. *Nature* 2007, **445**(7130):851-857.

101. Faries MB, Thompson JF, Cochran AJ, Andtbacka RH, Mozzillo N, Zager JS, Jahkola T, Bowles TL, Testori A, Beitsch PD *et al*: **Completion Dissection or Observation for Sentinel-Node Metastasis in Melanoma**. *The New England journal of medicine* 2017, **376**(23):2211-2222.

102. Leiter U, Stadler R, Mauch C, Hohenberger W, Brockmeyer N, Berking C, Sunderkötter C, Kaatz M, Schulte KW, Lehmann P *et al*: **Complete lymph node dissection versus no dissection in patients with sentinel lymph node biopsy positive melanoma (DeCOG-SLT): a multicentre, randomised, phase 3 trial**. *Lancet Oncol* 2016, **17**(6):757-767.

103. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M *et al*: **Improved survival with vemurafenib in melanoma with BRAF V600E mutation**. *The New England journal of medicine* 2011, **364**(26):2507-2516.

104. Robert C, Karaszewska B, Schachter J, Rutkowski P, Mackiewicz A, Stroiakovski D, Lichinitser M, Dummer R, Grange F, Mortier L *et al*: **Improved overall survival in**

melanoma with combined dabrafenib and trametinib. *The New England journal of medicine* 2015, **372**(1):30-39.

105. McArthur GA, Chapman PB, Robert C, Larkin J, Haanen JB, Dummer R, Ribas A, Hogg D, Hamid O, Ascierto PA *et al*: **Safety and efficacy of vemurafenib in BRAF(V600E) and BRAF(V600K) mutation-positive melanoma (BRIM-3): extended follow-up of a phase 3, randomised, open-label study**. *Lancet Oncol* 2014, **15**(3):323-332.

106. Lee L, Gupta M, Sahasranaman S: **Immune Checkpoint inhibitors: An introduction to the next-generation cancer immunotherapy**. *J Clin Pharmacol* 2016, **56**(2):157-169.

107. Coit DG, Thompson JA, Albertini MR, Barker C, Carson WE, Contreras C, Daniels GA, DiMaio D, Fields RC, Fleming MD *et al*: **Cutaneous Melanoma, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology**. *J Natl Compr Canc Netw* 2019, **17**(4):367-402.

108. Gellrich FF, Schmitz M, Beissert S, Meier F: **Anti-PD-1 and Novel Combinations in the Treatment of Melanoma-An Update**. *J Clin Med* 2020, **9**(1).

109. Eggermont AMM, Robert C, Ribas A: **The new era of adjuvant therapies for melanoma**. *Nat Rev Clin Oncol* 2018, **15**(9):535-536.

110. Long GV, Hauschild A, Santinami M, Atkinson V, Mandala M, Chiarion-Sileni V, Larkin J, Nyakas M, Dutriaux C, Haydon A *et al*: **Adjuvant Dabrafenib plus Trametinib in Stage III BRAF-Mutated Melanoma**. *The New England journal of medicine* 2017, **377**(19):1813-1823.

111. Eggermont AMM, Blank CU, Mandala M, Long GV, Atkinson V, Dalle S, Haydon A, Lichinitser M, Khattak A, Carlino MS *et al*: **Adjuvant Pembrolizumab versus Placebo in Resected Stage III Melanoma**. *The New England journal of medicine* 2018, **378**(19):1789-1801.

112. Weber J, Mandala M, Del Vecchio M, Gogas HJ, Arance AM, Cowey CL, Dalle S, Schenker M, Chiarion-Sileni V, Marquez-Rodas I *et al*: **Adjuvant Nivolumab versus Ipilimumab in Resected Stage III or IV Melanoma**. *The New England journal of medicine* 2017, **377**(19):1824-1835.

113. Luke JJ, Ascierto PA, Carlino MS, Gershenwald JE, Grob JJ, Hauschild A, Kirkwood JM, Long GV, Mohr P, Robert C *et al*: **KEYNOTE-716: Phase III study of adjuvant pembrolizumab versus placebo in resected high-risk stage II melanoma**. *Future Oncol* 2020, **16**(3):4429-4438.

114. Hamid O, Robert C, Daud A, Hodi FS, Hwu WJ, Kefford R, Wolchok JD, Hersey P, Joseph R, Weber JS *et al*: **Five-year survival outcomes for patients with advanced melanoma treated with pembrolizumab in KEYNOTE-001**. *Ann Oncol* 2019, **30**(4):582-588.

115. Asher N, Ben-Betzalel G, Lev-Ari S, Shapira-Frommer R, Steinberg-Silman Y, Gochman N, Schachter J, Meirson T, Markel G: **Real World Outcomes of Ipilimumab and Nivolumab in Patients with Metastatic Melanoma**. *Cancers* 2020, **12**(8).

116. Cohen JV, Tawbi H, Margolin KA, Amravadi R, Bosenberg M, Brastianos PK, Chiang VL, de Groot J, Glitza IC, Herlyn M *et al*: **Melanoma central nervous system metastases: current approaches, challenges, and opportunities**. *Pigment cell & melanoma research* 2016, **29**(6):627-642.

117. Qian M, Ma MW, Fleming NH, Lackaye DJ, Hernando E, Osman I, Shao Y: **Clinicopathological characteristics at primary melanoma diagnosis as risk factors for brain metastasis**. *Melanoma Res* 2013, **23**(6):461-467.

118. Barnholtz-Sloan JS, Sloan AE, Davis FG, Vigneau FD, Lai P, Sawaya RE: **Incidence proportions of brain metastases in patients diagnosed (1973 to 2001) in the Metropolitan Detroit Cancer Surveillance System**. *J Clin Oncol* 2004, **22**(14):2865-2872.

119. Fife KM, Colman MH, Stevens GN, Firth IC, Moon D, Shannon KF, Harman R, Petersen-Schaefer K, Zacest AC, Besser M *et al*: **Determinants of outcome in melanoma patients with cerebral metastases**. *J Clin Oncol* 2004, **22**(7):1293-1300.

120. Davies MA, Liu P, McIntyre S, Kim KB, Papadopoulos N, Hwu WJ, Hwu P, Bedikian A: **Prognostic factors for survival in melanoma patients with brain metastases**. *Cancer* 2011, **117**(8):1687-1696.

121. Kluger HM, Chiang V, Mahajan A, Zito CR, Sznol M, Tran T, Weiss SA, Cohen JV, Yu J, Hegde U *et al*: **Long-Term Survival of Patients With Melanoma With Active Brain Metastases Treated With Pembrolizumab on a Phase II Trial**. *J Clin Oncol* 2019, **37**(1):52-60.

122. Tawbi HA, Forsyth PA, Algazi A, Hamid O, Hodi FS, Moschos SJ, Khushalani NI, Lewis K, Lao CD, Postow MA *et al*: **Combined Nivolumab and Ipilimumab in Melanoma Metastatic to the Brain**. *The New England journal of medicine* 2018, **379**(8):722-730.

123. Long GV, Atkinson V, Lo S, Sandhu S, Guminski AD, Brown MP, Wilmott JS, Edwards J, Gonzalez M, Scolyer RA *et al*: **Combination nivolumab and ipilimumab or nivolumab alone in melanoma brain metastases: a multicentre randomised phase 2 study**. *Lancet Oncol* 2018, **19**(5):672-681.

124. Shaw E, Scott C, Souhami L, Dinapoli R, Kline R, Loeffler J, Farnan N: **Single dose radiosurgical treatment of recurrent previously irradiated primary brain tumors and brain metastases: final report of RTOG protocol 90-05**. *Int J Radiat Oncol Biol Phys* 2000, **47**(2):291-298.

125. Frankel TL, Bamboat ZM, Ariyan C, Coit D, Sabel MS, Brady MS: **Predicting the development of brain metastases in patients with local/regional melanoma**. *J Surg Oncol* 2014, **109**(8):770-774.

126. Haydu LE, Lo SN, McQuade JL, Amaria RN, Wargo J, Ross MI, Cormier JN, Lucci A, Lee JE, Ferguson SD *et al*: **Cumulative Incidence and Predictors of CNS Metastasis for Patients With American Joint Committee on Cancer 8th Edition Stage III Melanoma**. *J Clin Oncol* 2020:Jco1901508.

127. Brastianos PK, Carter SL, Santagata S, Cahill DP, Taylor-Weiner A, Jones RT, Van Allen EM, Lawrence MS, Horowitz PM, Cibulskis K *et al*: **Genomic Characterization of Brain Metastases Reveals Branched Evolution and Potential Therapeutic Targets**. *Cancer discovery* 2015, **5**(11):1164-1177.

128. Kircher DA, Silvis MR, Cho JH, Holmen SL: **Melanoma Brain Metastasis: Mechanisms, Models, and Medicine**. *International journal of molecular sciences* 2016, **17**(9).

129. Quail DF, Joyce JA: **The Microenvironmental Landscape of Brain Tumors**. *Cancer Cell* 2017, **31**(3):326-341.

130. Rebecca VW, Sondak VK, Smalley KS: **A brief history of melanoma: from mummies to mutations**. *Melanoma Res* 2012, **22**(2):114-122.

131. Balch CM, Gershenwald JE, Soong SJ, Thompson JF, Atkins MB, Byrd DR, Buzaid AC, Cochran AJ, Coit DG, Ding S *et al*: **Final version of 2009 AJCC melanoma staging and classification**. *J Clin Oncol* 2009, **27**(36):6199-6206.

132. Gershenwald JE, Scolyer RA, Hess KR, Sondak VK, Long GV, Ross MI, Lazar AJ, Faries MB, Kirkwood JM, McArthur GA *et al*: **Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual**. *CA Cancer J Clin* 2017, **67**(6):472-492.

133. Dickson PV, Gershenwald JE: **Staging and prognosis of cutaneous melanoma**. *Surg Oncol Clin N Am* 2011, **20**(1):1-17.

134. Lo SN, Scolyer RA, Thompson JF: **Long-Term Survival of Patients with Thin (T1) Cutaneous Melanomas: A Breslow Thickness Cut Point of 0.8 mm Separates Higher-Risk and Lower-Risk Tumors**. *Ann Surg Oncol* 2018, **25**(4):894-902.

135. Ribas A, Hamid O, Daud A, Hodi FS, Wolchok JD, Kefford R, Joshua AM, Patnaik A, Hwu WJ, Weber JS *et al*: **Association of Pembrolizumab With Tumor Response and Survival Among Patients With Advanced Melanoma**. *JAMA* 2016, **315**(15):1600-1609.

136. Robert C, Ribas A, Hamid O, Daud A, Wolchok JD, Joshua AM, Hwu WJ, Weber JS, Gangadhar TC, Joseph RW *et al*: **Durable Complete Response After Discontinuation of Pembrolizumab in Patients With Metastatic Melanoma**. *J Clin Oncol* 2018, **36**(17):1668-1674.

137. Michot JM, Bigenwald C, Champiat S, Collins M, Carbonnel F, Postel-Vinay S, Berdelou A, Varga A, Bahleda R, Hollebecque A *et al*: **Immune-related adverse events with immune checkpoint blockade: a comprehensive review**. *Eur J Cancer* 2016, **54**:139-148.

138. Franken MG, Kanters TA, Coenen JL, de Jong P, Koene HR, Lugtenburg PJ, Jager A, Uyl-de Groot CA: **Potential cost savings owing to the route of administration of oncology drugs: a microcosting study of intravenous and subcutaneous administration of trastuzumab and rituximab in the Netherlands**. *Anticancer Drugs* 2018, **29**(8):791-801.

139. Schumacher TN, Kesmir C, van Buuren MM: **Biomarkers in cancer immunotherapy**. *Cancer Cell* 2015, **27**(1):12-14.

140. Diem S, Kasenda B, Spain L, Martin-Liberal J, Marconcini R, Gore M, Larkin J: **Serum lactate dehydrogenase as an early marker for outcome in patients treated with anti-PD-1 therapy in metastatic melanoma**. *Br J Cancer* 2016, **114**(3):256-261.

141. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, Jiang P, Shen H, Aster JC, Rodig S *et al*: **Comprehensive analyses of tumor immunity: implications for cancer immunotherapy**. *Genome Biol* 2016, **17**(1):174.

142. Thomas NE, Busam KJ, From L, Kricker A, Armstrong BK, Anton-Culver H, Gruber SB, Gallagher RP, Zanetti R, Rosso S *et al*: **Tumor-infiltrating lymphocyte grade in**

primary melanomas is independently associated with melanoma-specific survival in the population-based genes, environment and melanoma study. *J Clin Oncol* 2013, **31**(33):4252-4259.

143. Ladanyi A: **Prognostic and predictive significance of immune cells infiltrating cutaneous melanoma**. *Pigment cell & melanoma research* 2015, **28**(5):490-500.

144. Anichini A, Molla A, Vegetti C, Bersani I, Zappasodi R, Arienti F, Ravagnani F, Maurichi A, Patuzzo R, Santinami M *et al*: **Tumor-reactive CD8+ early effector T cells identified at tumor site in primary and metastatic melanoma**. *Cancer Res* 2010, **70**(21):8378-8387.

145. Anwar MA, Murad F, Dawson E, Abd Elmageed ZY, Tsumagari K, Kandil E: **Immunohistochemistry as a reliable method for detection of BRAF-V600E mutation in melanoma: a systematic review and meta-analysis of current published literature**. *J Surg Res* 2016, **203**(2):407-415.

146. Vanni I, Tanda ET, Spagnolo F, Andreotti V, Bruno W, Ghiorzo P: **The Current State of Molecular Testing in the BRAF-Mutated Melanoma Landscape**. *Front Mol Biosci* 2020, **7**:113-113.

147. Jonsson G, Busch C, Knappskog S, Geisler J, Miletic H, Ringner M, Lillehaug JR, Borg A, Lonning PE: **Gene expression profiling-based identification of molecular subtypes in stage IV melanomas with different clinical outcome**. *Clin Cancer Res* 2010, **16**(13):3356-3367.

148. Harbst K, Staaf J, Lauss M, Karlsson A, Masback A, Johansson I, Bendahl PO, Vallon-Christersson J, Torngren T, Ekedahl H *et al*: **Molecular profiling reveals low- and high-grade forms of primary melanoma**. *Clin Cancer Res* 2012, **18**(15):4026-4036.

149. Gastman BR, Gerami P, Kurley SJ, Cook RW, Leachman S, Vetto JT: **Identification of patients at risk for metastasis using a prognostic 31-gene expression profile in subpopulations of melanoma patients with favorable outcomes by standard criteria**. *J Am Acad Dermatol* 2018.

150. Zager JS, Gastman BR, Leachman S, Gonzalez RC, Fleming MD, Ferris LK, Ho J, Miller AR, Cook RW, Covington KR *et al*: **Performance of a prognostic 31-gene expression profile in an independent cohort of 523 cutaneous melanoma patients**. *BMC Cancer* 2018, **18**(1):130.

151. Vetto JT, Hsueh EC, Gastman BR, Dillon LD, Monzon FA, Cook RW, Keller J, Huang X, Fleming A, Hewgley P *et al*: **Guidance of sentinel lymph node biopsy decisions in patients with T1-T2 melanoma using gene expression profiling**. *Future Oncol* 2019, **15**(11):1207-1217.

152. Marchetti MA, Bartlett EK, Dusza SW, Bichakjian CK: **Use of a prognostic gene expression profile test for T1 cutaneous melanoma: Will it help or harm patients?** *J Am Acad Dermatol* 2019, **80**(6):e161-e162.

153. Kovarik CL, Chu EY, Adamson AS: **Gene Expression Profile Testing for Thin Melanoma: Evidence to Support Clinical Use Remains Thin**. *JAMA Dermatol* 2020, **156**(8):837-838.

154. Birkeland E, Zhang S, Poduval D, Geisler J, Nakken S, Vodak D, Meza-Zepeda LA, Hovig E, Myklebost O, Knappskog S *et al*: **Patterns of genomic evolution in advanced melanoma**. *Nature communications* 2018, **9**(1):2665.

155. Koch CM, Chiu SF, Akbarpour M, Bharat A, Ridge KM, Bartom ET, Winter DR: **A Beginner's Guide to Analysis of RNA Sequencing Data**. *Am J Respir Cell Mol Biol* 2018, **59**(2):145-157.

156. Corrie PG, Marshall A, Dunn JA, Middleton MR, Nathan PD, Gore M, Davidson N, Nicholson S, Kelly CG, Marples M *et al*: **Adjuvant bevacizumab in patients with melanoma at high risk of recurrence (AVAST-M): preplanned interim results from a multicentre, open-label, randomised controlled phase 3 study**. *Lancet Oncol* 2014, **15**(6):620-630.

157. Corrie PG, Marshall A, Nathan PD, Lorigan P, Gore M, Tahir S, Faust G, Kelly CG, Marples M, Danson SJ *et al*: **Adjuvant bevacizumab for melanoma patients at high risk of recurrence: survival analysis of the AVAST-M trial**. *Ann Oncol* 2018, **29**(8):1843-1852.

158. Malone ER, Oliva M, Sabatini PJB, Stockley TL, Siu LL: **Molecular profiling for precision cancer therapies**. *Genome Medicine* 2020, **12**(1):8.

Genomic analysis and clinical management of adolescent cutaneous melanoma

Roy Rabbie[1,2,*], Mamunur Rashid[1,*], Ana M. Arance[3], Marcelo Sánchez[4], Gemma Tell-Marti[5,6], Miriam Potrony[7], Carles Conill[8], Remco van Doorn[9], Stefan Dentro[1], Nelleke A. Gruis[9], Pippa Corrie[2], Vivek Iyer[1], Carla Daniela Robles-Espinoza[1,10], Joan A. Puig-Butille[6,11], Susana Puig[5,6,*] and David J. Adams[1,*]

1 Experimental Cancer Genetics, The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK **2** Department of Oncology, Cambridge University Hospitals National Health Service Foundation Trust, Cambridge, UK **3** Department of Medical Oncology and Targeted Therapeutics in Solid Tumors Group (IDIBAPS), Hospital Clınic de Barcelona, Barcelona, Spain **4** Melanoma Unit, Radiology Service, Hospital Clınic, IDIBAPS, University of Barcelona, Barcelona, Spain **5** Melanoma Unit, Department of Dermatology, Hospital Clınic de Barcelona, Barcelona, Spain **6** Centre of Biomedical Research on Rare Diseases (CIBERER), ISCIII, Barcelona, Spain **7** Melanoma Unit, Department of Dermatology, Hospital Clınic de Barcelona, IDIBAPS, Barcelona University, Barcelona, Spain **8** Melanoma Unit, Radiotherapy Oncology, Hospital Clınic, IDIBAPS, Barcelona University, Barcelona, Spain **9** Leiden University Medical Centre, Leiden, The Netherlands **10** Laboratorio Internacional de Investigacion sobre el Genoma Humano, Universidad Nacional Autonoma de Mexico, Santiago de Queretaro, Mexico **11** Biochemistry and Molecular Genetics Department, Melanoma Unit, Hospital Clinic de Barcelona, IDIBAPS, Barcelona, Spain

CORRESPONDENCE David J. Adams, e-mail: da1@sanger.ac.uk

*These authors contributed equally.

KEYWORDS adolescent melanoma/germline mutation/ultraviolet radiation/BRAF mutation/immunotherapy

PUBLICATION DATA Received 17 October 2016, revised and accepted for publication 11 January 2017, published online 17 January 2017

doi: 10.1111/pcmr.12574

Summary

Melanoma in young children is rare; however, its incidence in adolescents and young adults is rising. We describe the clinical course of a 15-year-old female diagnosed with AJCC stage IB non-ulcerated primary melanoma, who died from metastatic disease 4 years after diagnosis despite three lines of modern systemic therapy. We also present the complete genomic profile of her tumour and compare this to a further series of 13 adolescent melanomas and 275 adult cutaneous melanomas. A somatic *BRAF*^{V600E} mutation and a high mutational load equivalent to that found in adult melanoma and composed primarily of C>T mutations were observed. A germline genomic analysis alongside a series of 23 children and adolescents with melanoma revealed no mutations in known germline melanoma-predisposing genes. Adolescent melanomas appear to have genomes that are as complex as those arising in adulthood and their clinical course can, as with adults, be unpredictable.

Significance

The survival from advanced melanoma in adults has been revolutionized by the introduction of immune checkpoint inhibitors and molecular targeted therapies. However, children and adolescents younger than 18 years have had limited access to the registration clinical trials. We present a detailed genomic analysis of a series of adolescent melanomas and the clinical course of one such patient who died from metastatic disease. A high mutational load was observed and suggests that immune-based therapies may be relevant, but response cannot be guaranteed. Germline mutations in established adult melanoma-predisposing genes were not evident in an extended childhood and adolescent series. Given the complexities around diagnosis and the paucity of prospective clinical studies for younger individuals, melanoma in this age group represents a particular clinical challenge requiring specialist management by a dedicated multidisciplinary team.

© 2017 The Authors. *Pigment Cell & Melanoma Research* published by John Wiley & Sons Ltd.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Introduction

Melanoma in children is rare accounting for only 2% of all malignancies in patients younger than 20 years (Howlader et al., 2016). Melanoma in infancy and early childhood (1–10 years) comprises around 8% of newly diagnosed cases in young people, whereas adolescents (11–20 years) account for the majority (92%) of melanoma cases (Lorimer et al., 2016). Importantly, the incidence of melanoma in the adolescent population is rising at a rate of 2% per year (Austin et al., 2013). Melanocytic lesions in children comprise a heterogeneous group of neoplasms that can be broadly divided based on histology and age onset, and three major subtypes have been described (Barnhill and Kerl, 2006). Firstly, melanoma can arise in association with a pre-existing, usually large, congenital melanocytic naevus (CMN) (Guegan et al., 2016; Trozak et al., 1975). The lifetime risk of malignant transformation from a CMN is 5–10% and 50% of these transformations are said to occur in the first decade of life (Bett, 2005; Krengel et al., 2006). The second type are termed spitzoid melanocytic tumours, which comprise a wider spectrum of histological variants including spitzoid melanoma and atypical Spitz tumours. The vast majority of Spitz naevi occur in individuals younger than 20 years and often arise on the extremities and face (Reed et al., 2013). The third subtype, generally occurring in adolescents, has been termed 'conventional' melanoma, owing to its shared clinical and histological features typical of adult melanoma. In contrast to infantile and childhood cases, post-pubertal melanoma is most often sporadic, occurring as a de-novo lesion in patients with fair-coloured skin and substantial sun exposure (Wood, 2016).

Cutaneous melanoma in adults is characterised by a high prevalence of somatic mutations and the mutational pattern depicts a characteristic ultraviolet-light (UV)-induced signature associated with frequent transitions at dipyrimidine sites (Cancer Genome Atlas Network, 2015). A recent comprehensive genomic analysis found that melanomas from adolescents bear a remarkably similar mutational rate and spectrum to tumours from adults, suggesting that sun protection practices are important in early life (Anderson et al., 2009; Lu et al., 2015). In addition to its rarity and the low clinical suspicion for malignancy, there is recognition that melanomas in young people are commonly amelanotic and the clinico-pathologic features may overlap with proliferative nodules and other benign skin lesions that are generally more common in children than adults (Cordoro et al., 2013; Moscarella et al., 2012). This can lead to delays both in diagnosis and treatment (Neier et al., 2012).

Several high-risk mutations have been identified in melanoma-dense families, including mutations in the cyclin-dependent kinase inhibitor 2A (*CDKN2A*) gene (Cannon-Albright et al., 1992), the cyclin-dependent kinase 4 (*CDK4*) gene (Zuo et al., 1996) and more recently in the Breast cancer 1 (*BRCA1*)-associated protein 1 (*BAP1*) (Aoude et al., 2013; Wiesner et al., 2011) and protection of telomeres 1 (*POT1*) genes (Robles-Espinoza et al., 2014; Shi et al., 2014). However, the prevalence of these predisposing mutations amongst younger patients is largely unknown. A deeper understanding of the molecular drivers of childhood and adolescent melanoma would advance our understanding of its pathogenesis, particularly the role of gene–environment interactions in susceptible cases and could help define particular high-risk subgroups that might benefit from specialist screening and surveillance.

Results

In this study, we present a detailed clinical history of one patient and an extensive genomic analysis of their germline and tumour and that of a wider series of adolescent and childhood melanomas.

Patient presentation

The 15-year-old female described had blonde hair, blue eyes, skin phototype II on the Fitzpatrick Classification Scale (Fitzpatrick, 1988) and a history of multiple (>50) benign skin naevi. Her mother had a history of uveal melanoma, and her maternal grandmother had pancreatic adenocarcinoma (Figure 1A). She presented in February 2011 with an enlarging symmetric raised light brown papule on the right lower posterior chest wall at the level of the costal margin, which measured less than 1 cm in diameter. The lesion was removed by shave excision at her local hospital and was found to be a non-ulcerated cutaneous malignant melanoma, Clark's level IV, Breslow thickness 0.9 mm and 6 mitoses/mm^2 (Figure 1B). As the lesion extended to the excision margins, wide local excision was undertaken with subsequent clear margins. Ultrasonography revealed no pathological regional lymph nodes and she underwent active multimodality 6-monthly surveillance. Two and a half years later in October 2013, a 0.8 mm pigmented lesion appeared in the centre of the existing wide local excision scar (Figure 1C). Dermoscopic examination revealed a homogeneous pattern and reflectance confocal microscopy showed atypical dendritic cells in the dermo-epidermal junction (Figure 1D). This lesion was diagnosed as melanoma in situ, which was completely excised. In March 2014, no abnormalities were detected on surveillance clinical examination or ultrasonography and serum s-100 levels were recorded as normal at 0.13 μg/l (normal <0.15 μg/l). However, 2 months later, during a separate clinic consultation for acne treatment, an enlarged lymph node was detected in the right axilla and serum s-100 levels were now elevated to 0.7 μg/l. A PET/CT scan revealed avid FDG uptake in multiple liver and bone metastases as well as right axillary lymph nodes (Figure 2A). A single asymptomatic parietal lobe brain metastasis was also identified on imaging

© 2017 The Authors. *Pigment Cell & Melanoma Research* published by John Wiley & Sons Ltd.

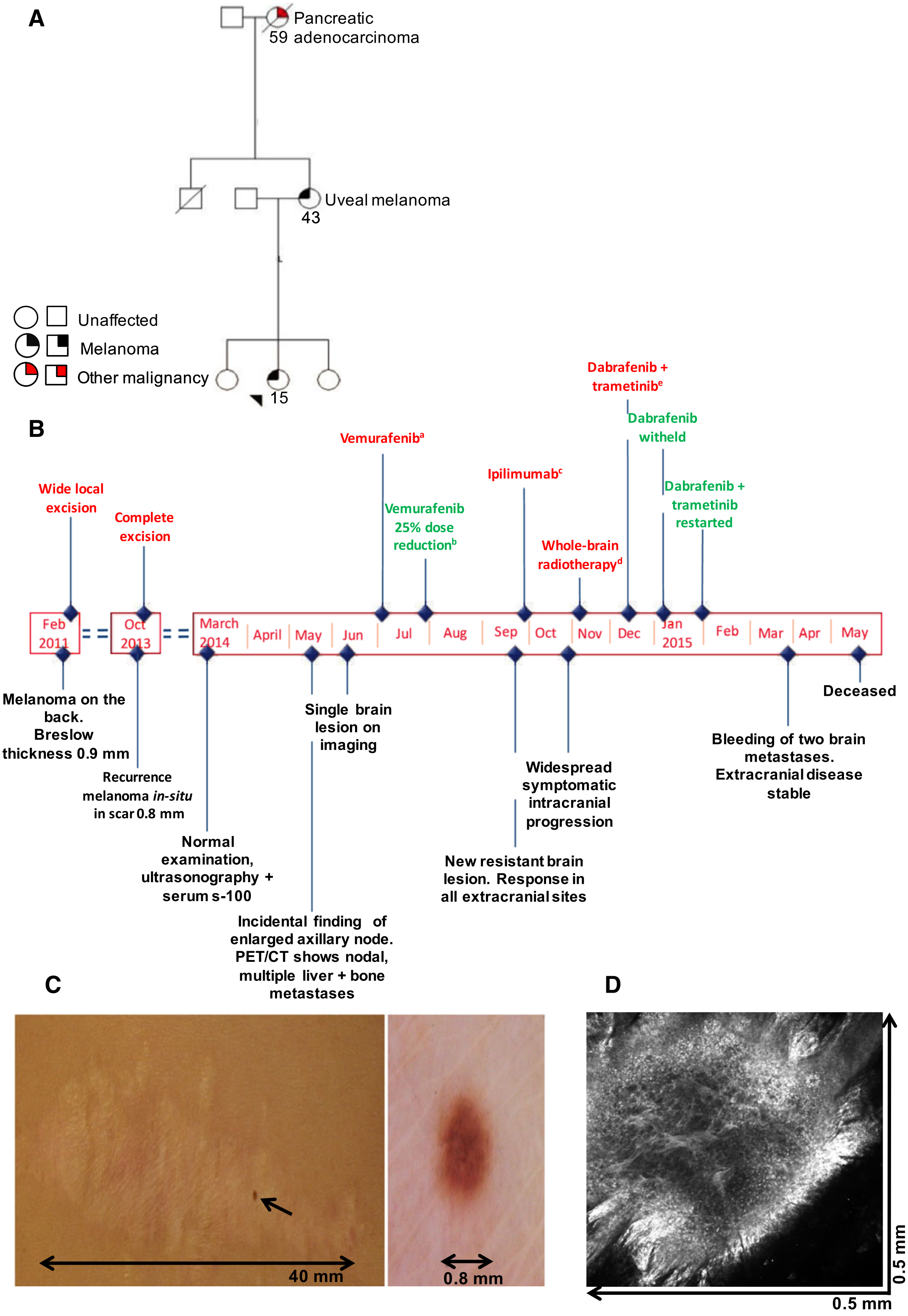


Figure 1. Clinical timeline of the 15-year-old index patient (M_4180). (A) Family pedigree. The proband is indicated with an arrow, ages at diagnosis are shown. (B) Timeline of diagnosis and treatment. (C) New pigmented melanoma in situ appearing in the centre of previous melanoma wide local excision scar. Accompanied dermoscopic image of the in-situ melanoma prior to further wide local excision (beside). (D) Reflectance confocal microscopy at the dermo-epidermal level, showing proliferation of dendritic atypical melanocytes. [a]Vemurafenib starting dose 960 mg twice a day. [b]Dose reduction vemurafenib to 720 mg twice a day. [c]Ipilimumab 3 mg/kg every 3 weeks. [d]Whole-brain radiotherapy 10 Gy in 10 fractions. [e]Dabrafenib 150 mg twice a day, trametinib 2 mg once a day.

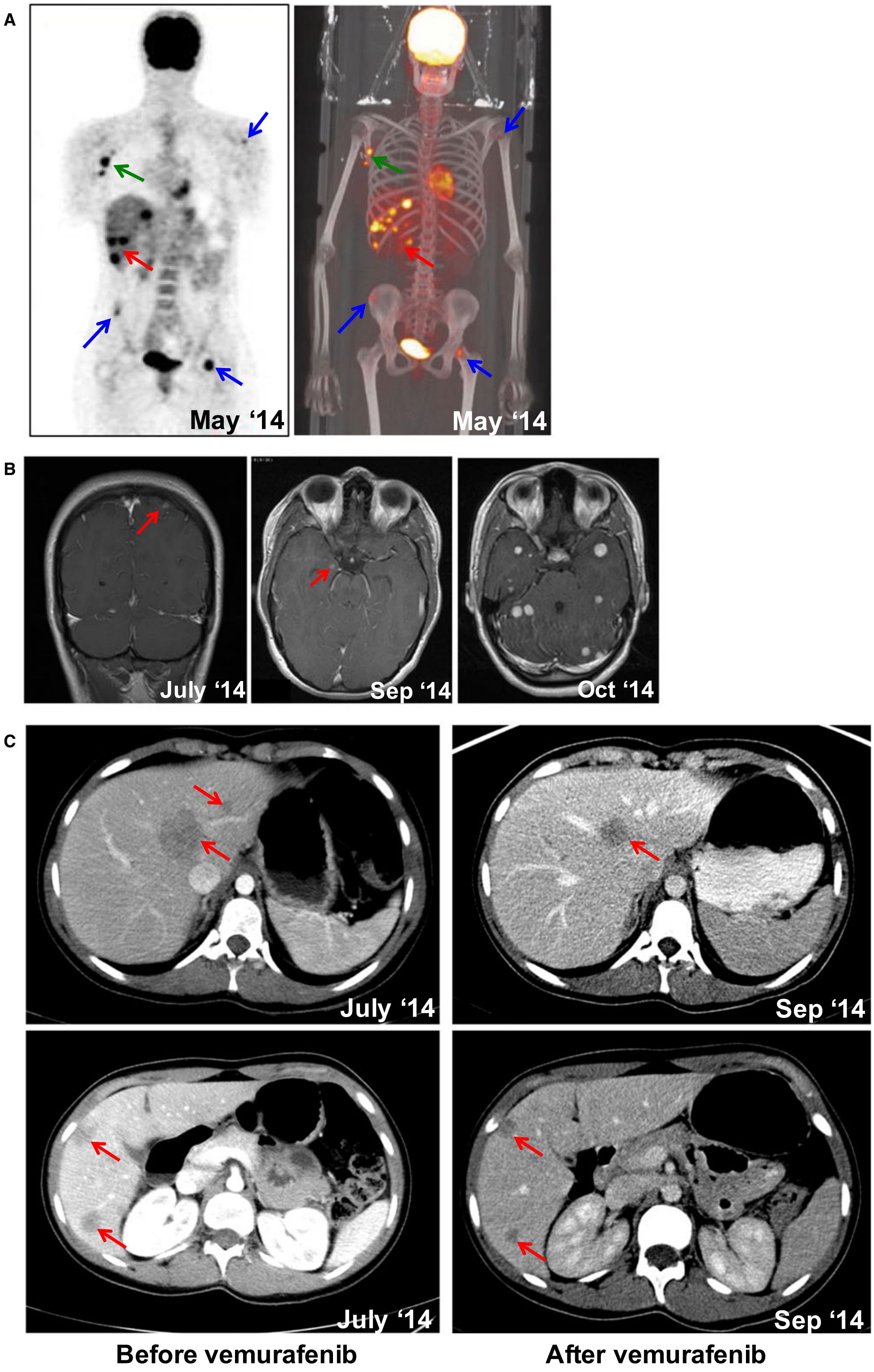


Figure 2. Radiological evaluation through treatment. (A) 18F-FDG PET/CT alongside 3D colour reconstruction. Arrows indicate avid FDG tracer uptake in the right axilla, left humeral head, left femoral neck and right iliac crest (blue) as well as widespread liver uptake (red). (B) Post-contrast T1-weighted MR images showing tiny enhancing lesions in the left parietal lobe (July 2014) and right amygdala (September 2014). Axial post-contrast MR images prior to whole-brain radiotherapy showing multiple and supra- and infratentorial lesions with no significant mass effect (October 2014). (C) Cross-sectional CT images of the liver post-IV-contrast in the portal phase. Baseline images show hypodense focal lesions corresponding to segment 1 in the left hepatic lobe (left upper) and the caudal segments of the right hepatic lobe (left lower). On the right, post-treatment images indicate a partial response in all liver lesions (arrows).

(Figure 2B). Three cutaneous metastases were evident, one of which was excised. Molecular analysis of the excised metastasis using PCR revealed a *BRAF*^{V600E} mutation. In July 2014, she was commenced on systemic therapy with the BRAF kinase inhibitor, vemurafenib. Ten days into therapy, she experienced arthralgia, blepharitis, meibomian gland inflammation (presenting with suppuration from the sebaceous gland at the rim of the eyelids and treated with topical and oral antibiotics), as well as a widespread cutaneous rash necessitating interruption of treatment (Erfan et al., 2017; Figure S1). Treatment was reintroduced 2 weeks later at a 25% dose reduction. Repeat cross-sectional imaging 2 months later showed a response in all the nodal and liver lesions (Figure 2C). There was also response in the parietal lobe lesion, but a new brain metastasis within the amygdala was now evident (Figure 2B). Vemurafenib was therefore stopped and, following a 3-week washout, immune checkpoint inhibitor therapy with ipilimumab was commenced. Following the second cycle, she was admitted to hospital with migraine and unsteadiness of gait and neuroimaging revealed widespread multiple brain metastases (Figure 2B). Her symptoms improved with corticosteroids and whole-brain radiotherapy. In December 2014, combination MAP kinase inhibitor therapy with dabrafenib and trametinib was commenced. Treatment was associated with pyrexia necessitating brief interruption of dabrafenib, but subsequent resumption of the combination regimen. At the end of March 2015, she was readmitted with a sudden-onset severe headache. Imaging revealed bleeding and perilesional oedema into two existing brain metastasis and the appearance of a further new brain metastasis. She died from progressive metastatic melanoma 2 months later.

Tumour genomic analyses

Whole-genome sequencing of a cutaneous metastasis and matched germline DNA from the patient described above revealed somatic mutations in melanoma driver genes including a *BRAF*^{V600E} mutation, and a truncating *CDKN2A* mutation (Figure 3A). In total, we identified 133 mutations in the protein-coding region of the genome, of which 89 were protein-altering and 44 were silent (non-synonymous to silent mutation ratio = 2.022; Tables S4 and S5). 15,853 somatic mutations were identified genome wide with a mutation frequency of 5.12 mutations per megabase (Figure 3A, C). The tumour displayed a disproportionately high level of cytidine to thymidine (C>T) transitions accounting for >80% of all nucleotide changes. The mutational spectrum bore closest resemblance to the UV-exposure signature (signature 7) described by Alexandrov et al. (2013) (cosine similarity test 0.63; Figure S2). We validated 42 randomly selected loci via Sanger sequencing of tumour and germline DNA and found 36 (86%) to be true somatic variants (Table S7). A further 13 ‘conventional’ melanomas (so-called due to their shared clinical and histological features typical of adult cutaneous melanoma) were identified from Lu et al. (2015). These patients had a median age of 16 years (13–20) and ranged from stage IB to IV disease at initial diagnosis. The primary tumours were generally from sun exposed sites (six from the head and neck, three from the trunk, three from the extremities and one unknown) and were mainly of common histological subtypes (six nodular, five superficial spreading, one acral and one unknown; Table S1). Pooling variants from our patient with somatic variant calls from these 13 conventional melanomas revealed a median of 10.23 mutations per megabase (3.21–52.65; Figure 3A; Table S6). We obtained further mutation data from 275 adult cutaneous melanomas from The Cancer Genome Atlas (mean age 56.62 years). A Wilcoxon test comparing these to the adolescent melanoma series did not reveal any significant difference between the mutation rates of adolescent vs adult cutaneous melanoma (P value = 0.2721).

Germline genomic analyses

We investigated this 15-year-old patient’s germline genome for known melanoma-predisposing genes including *CDKN2A, CDK4* and *BAP1* but failed to find any rearrangements, copy number neutral changes, point mutations or other alternations that may explain her presentation. A wider analysis of 23 additional children and adolescents, including five new cases with resected primary melanoma that we whole-genome-sequenced for this study and 18 children described in Lu et al. (2015), also failed to identify variants in established melanoma-predisposing genes. These five new cases had a median age of 10 years (6–16), were all of the superficial spreading histological subtype and had AJCC stage I disease at first presentation, while the remaining 18 cases identified from Lu et al. had a median age of 15 years (9 months–20 years) and included a wider spectrum of both stages and histological subtypes (Table S1). We noted that our patient carried R142H and V60L alleles in the melanocortin 1 receptor (*MC1R*) gene, contributing to her pale complexion (Garcia-Borron et al., 2014) (Figure S1). Other *MC1R* variants were also discovered in the children and adolescents analysed in our study (Table S3). In view of the emerging evidence implicating telomere dysregulation in familial melanoma (Robles-Espinoza et al., 2014; Shi et al., 2014), we further searched for alterations in genes encoding the shelterin protein complex that protect the ends of telomeres. In particular, the protection of telomeres 1 (*POT1*) gene, adrenocortical dysplasia homolog (*ACD*) gene and telomeric repeat binding factor 2 interacting protein (*TERF2IP*) genes have been shown to be important in some melanoma families (Aoude et al., 2015). We found 1 of 24 patients carried a missense mutation in *TERF2IP* (allele frequency 0.00378 in The Exome Aggregation Consortium (ExAC) (Table S3), although the pathogenicity of this mutation is unknown.

© 2017 The Authors. *Pigment Cell & Melanoma Research* published by John Wiley & Sons Ltd.

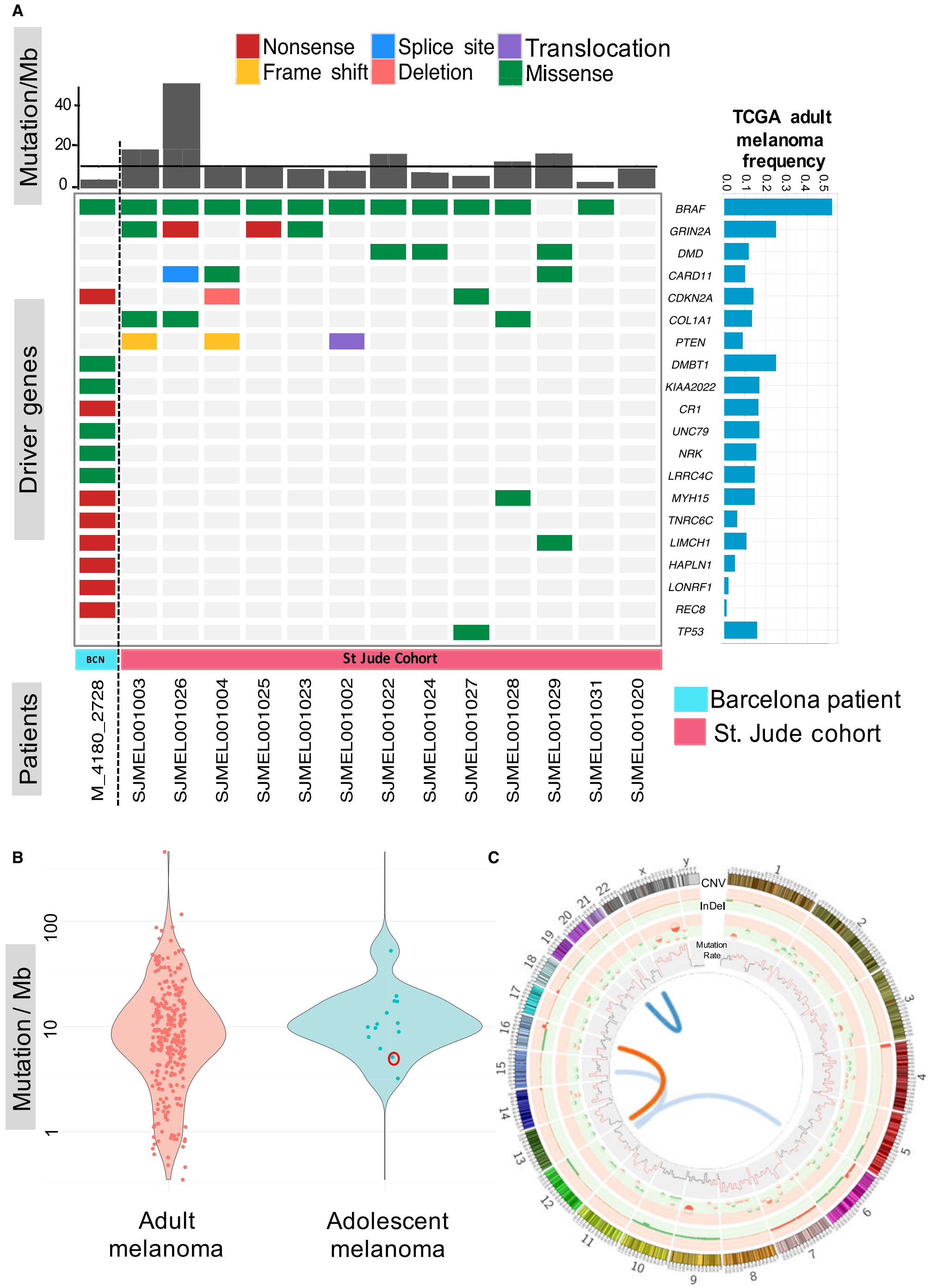


Figure 3. Somatic genomic analyses of adolescent melanoma. (A) Mutational landscape of adolescent melanoma. Driver mutations from the patient described are shown in the first column on the left-hand side. Remaining cases are from Lu et al. (2015) and indicate the 13 conventional adolescent melanoma patients described within this cohort and for whom genome sequencing data was available. Bar chart across the top panel shows the mutation rate per megabase (Mb) while the right panel shows the mutational frequency in adult cutaneous melanoma found in The Cancer Genome Altas (TCGA; Cancer Genome Atlas Network, 2015), straight line indicates the median number of mutations across all patients. Genes were selected based on those most frequently mutated in The Cancer Genome Atlas (adult) and in Lu et al. (childhood and adolescent; Lu et al., 2015), as well as the loss-of-function mutations detected in this 15-year-old patient. A number of commonly mutated genes identified in the TCGA melanoma cohort are omitted owing to the absence of mutations of these genes in our adolescent data set (including *NRAS*, *NF1*, *MAP2K1* and *RB1*). (B) Cluster plot of mutational frequency of adolescent versus adult cutaneous melanoma. The index patient described is circled in red. (C) Circos plot of somatic changes in the 15-year-old patient described. The outermost track shows large copy number gains (red) and losses (green) (Table S8). Middle track shows small insertions and deletions (Table S9). The inner most track shows mutations per Mb (regions marked in red have mutation rates higher than 15 mutations/Mb). Interchromosomal translocations are shown in the centre and were seen in; t(12;6)(q21; q2), t(12;15)(q14;q1), t(16;12)(q23;q2) and t(20;22)(q13;q32) (Table S10).

Discussion

Metastatic spread of melanoma is relatively rare amongst children; however, there are data that suggests that when this occurs the prognosis is particularly poor (Strouse et al., 2005). The adolescent described in this study presented with a AJCC stage IB primary melanoma, which is associated with a 95% 5-year survival (Balch et al., 2009). Despite this, she developed extensive metastases 3 years after diagnosis and died of metastatic disease within 12 months despite three lines of modern systemic therapies known to offer potential for survival gain.

Notably, and as reported previously (Lu et al., 2015), we identified a preponderance of UV-induced mutations across 'conventional' adolescent melanomas, which was unexpected given the relatively limited exposure to UV light compared to an adult population. This 15-year-old patient had intermittent sun exposure amounting to approximately 120 h/yr, yet was always appropriately sun protected. We were unable to find germline predisposing alleles in an extended series of children and adolescents, suggesting that established high-penetrance predisposition genes do not explain most cases. However, many of these patients carried *R* variants of *MC1R* associated with red hair, freckles and pale skin (Valverde et al., 1995).

Given the variability in clinical behaviour, wide histological variation and the rarity of melanoma in infancy and early childhood, studies in this population are scarce. Consequently, our understanding of the pathogenesis in this younger cohort is more limited. Analysis of a recent large national data set of over 350 000 melanoma patients showed that children (1–10 years) and adolescents (11–20 years) had differing survivals, suggesting inherent differences in the biology of the disease (Lorimer et al., 2016). The distinct clinical and histopathological features of melanomas arising in a CMN and Spitzoid tumours suggest that their molecular features are likely to be very different from the 'conventional' adolescent tumours described herein (Kinsler et al., 2013; Lu et al., 2015; Shakhova et al., 2012). Additional studies on the genomic evolution of these rarer subtypes could help facilitate improved diagnostics and tailored therapies.

The reason for the rise in incidence of melanoma during adolescence remains unclear; however, the finding of a high mutational load driven by UV exposure supports the need for education and behavioural modification as an important preventative strategy starting in early life (Green et al., 2011). The strong therapeutic effect of immune checkpoint blockade in some patients has been linked to the expression of neoantigens, mutant peptides presented by MHC Class I. A higher overall mutational burden would be expected to lead to the expression of more neoantigens, with mutation number being associated with improved efficacy of immunotherapy (Snyder et al., 2014; Van Allen et al., 2015). This adolescent developed metastatic disease at 18 years and accessed a range of modern treatments through clinical trials. It is imperative that adolescents are given the opportunity to participate in relevant clinical trials that include novel therapies (Pappo, 2014).

Methods

Patient enrolment

We whole-genome-sequenced six patients as part of our study. Our first patient (M_4180), whose treatment we detail, was a 15-year-old female who attended the Department of Dermatology at the University Hospital Clínic of Barcelona, Spain. Five additional children with resected primary melanoma were also identified from the University Hospital Clínic of Barcelona and from Leiden University Medical Center, the Netherlands. The remaining cases were selected from a cohort of paediatric melanomas identified and sequenced at St Jude Children's Hospital, Memphis, TN (Lu et al., 2015), as part of the Paediatric Cancer Genome Project (Downing et al., 2012) study accession through the European Genome-phenome Archive; EGAS00001000901 (Table S1). Written informed consent was obtained from the patients' parents.

Dermoscopy, histopathology and imaging

Total body photography and digital dermoscopy were performed by SP using MoleMax™ HD (Derma Medical Systems, Vienna, Austria) and DermLite® FOTO (Dermlite®, San Juan Capistrano, CA, USA). Histopathological analyses were performed by an expert dermatopathologist.

Sample processing

Tumour DNA extraction from the index 15-year-old patient (M_4180) was performed using the Qiagen DNA Micro Kit. Germline DNA was extracted from peripheral blood mononuclear cells using the salting out method.

Tumour genomic analyses

DNA from a metastatic cutaneous deposit and whole blood DNA from the index 15-year-old patient were genome sequenced on the Illumina X10 platform (Table S2). Whole-genome-sequenced reads were aligned against the human reference genome (GRCh37) using the Burrows–Wheeler Aligner (Li and Durbin, 2009; Table S2). We used a somatic caller merging approach to identify somatic variants selecting only those detected using four or more algorithms for further analysis (Rashid et al., 2013). These calls were further filtered for germline polymorphic variants using the 1000 Genomes Project (Auton et al., 2015), and other standard quality filters were also applied (e.g. depth of coverage ≥10, read mapping quality ≥15). Small insertions and deletions were identified using Scalpel (Narzisi et al., 2014). Randomly selected candidate variants were validated by capillary sequencing. Large somatic copy number aberrations were detected using the Batternberg algorithm. Somatic variants from a series of 13 'conventional' paediatric melanomas described by Lu et al. (2015) (so-called due to their shared clinical and histological features typical of adult cutaneous melanoma) and for whom genome sequencing data were available were used for a comparative analysis. Exome sequencing data from a further 275 adult cutaneous melanomas were downloaded from The Cancer Genome Atlas and used for comparison with adult-onset disease (Table S6).

Germline genomic analyses

Germline DNA from the peripheral blood of five children with resected primary melanoma was whole-genome-sequenced on the

Illumina HiSeq2500 platform (Tables S1 and S2). These sequences, and that of the index case, were combined with whole-genome and whole-exome sequences from a collection of 18 children sequenced at St Jude Children's Hospital (Lu et al., 2015) comprising 13 children from the 'conventional' melanoma cohort described above and five from the other histological subgroups described therein (Table S3). Germline variants were called using samtools mpileup (Li et al., 2009) and bcftools (Narasimhan et al., 2016). These variants were annotated for consequence using Ensembl Variant Effect Predictor (McLaren et al., 2016) and filtered for non-synonymous variants. They were then further restricted to those variants known to be rare (allele frequency < 10^{-3}) by comparison with the Exome Aggregation Consortium (ExAc) data set (Lek et al., 2016) or that were private to a single child.

Data accession IDs

M_4180_2728	EGAN00001232866	Tumour of patient M_4180
M_4180	EGAN00001195811	Germline of patient M_4180
M_509	EGAN00001197185	Germline of patient M_509
M_1064	EGAN00001197186	Germline of patient M_1064
M_3629	EGAN00001197187	Germline of patient M_3629
M_4117	EGAN00001197188	Germline of patient M_4117
D1_10_02707	EGAN00001197189	Germline of patient D1_10_02707

For details, see Table S1.

Acknowledgements

We would like to sincerely thank this patient and her family for giving us the opportunity to undertake this research and for allowing us to share these data. We extend this immense personal appreciation to all the patients and families involved. We would also like to thank Armita Bahrami and Alberto Pappo from St Jude Children's Hospital for their helpful discussions and for sharing their experience and data.

Author contributions

Susana Puig identified the patient, provided the clinical history and undertook the dermoscopy as well as identifying four further suitable patients for this analysis. Susana Puig also secured informed consent. Remco van Doorn identified the patient from the University Hospital Leiden and secured informed consent. Marcelo Sánchez reviewed the radiology. Ana Arance and Carles Conill were the treating medical and radiation oncologists respectively. Joan Puig-Butille was the clinical geneticist. Gemma Tell-Marti extracted the tumour DNA for the patient described and Miriam Potrony collated the germline data from University Hospital Barcelona. Mamunur Rashid performed the bioinformatics analyses. Stefan Dentro undertook the copy number calls for this patient. Vivek Iyer analysed the germline sequencing data from the five further paediatric cases described, together with the germline data from the 18 children identified from Lu et al. (2015). Mamunur Rashid, Roy Rabbie and Carla Daniela Robles-Espinoza critically analysed the somatic mutational data in the context of molecular data from The Cancer Genome Atlas. Roy Rabbie wrote the case history, prepared the clinical images and summarized the extended children's phenotypes. Roy Rabbie and Mamunur Rashid drafted the manuscript. Remco van Doorn, Pippa Corrie and Carla Daniela Robles-Espinoza reviewed the manuscript and provided expert external dermatological, oncological and genomic input respectively. David Adams provided overall support and supervision including concept, analyses and critical manuscript revisions. All authors approved the final manuscript. The funders did not play a role in the design of this study or in the interpretation of the results.

Funding

This work was supported by Cancer Research UK and the Wellcome Trust. The research at the Melanoma Unit in Barcelona is partially funded by Spanish Fondo de Investigaciones Sanitarias grants PI12/00840, PI15/00716 and PI15/00956; CIBER de Enfermedades Raras of the Instituto de Salud Carlos III, Spain, co-financed by European Development Regional Fund 'A way to achieve Europe' ERDF; AGAUR 2014_SGR_603 of the Catalan Government, Spain; European Commission under the 6th Framework Programme, Contract No. LSHC-CT-2006-018702 (GenoMEL) and by the European Commission under the 7th Framework Programme, Diagnoptics; a grant from 'Fundació La Marató de TV3, 201331-30', Catalonia, Spain; a grant from Telemaraton of Spain 'Todos somos raros'; and a grant from 'Asociación Española Contra el Cáncer (AECC)'. The work was carried out at the Esther Koplowitz Center, Barcelona. MP is the recipient of a PhD fellowship FI14/00231 (PFIS) from Instituto de Salud Carlos III, Spain. RvD and NAG were supported by a grant from the Dutch Cancer Society (UL 2012–5489).

Conflict of interest

No conflict of interests to declare.

References

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C. et al. (2013). Signatures of mutational processes in human cancer. Nature *500*, 415–421.

Anderson, W.F., Pfeiffer, R.M., Tucker, M.A., and Rosenberg, P.S. (2009). Divergent cancer pathways for early-onset and late-onset cutaneous malignant melanoma. Cancer *115*, 4176–4185.

Aoude, L.G., Wadt, K., Bojesen, A., Cruger, D., Borg, A., Trent, J.M., Brown, K.M., Gerdes, A.M., Jonsson, G., and Hayward, N.K. (2013). A BAP1 mutation in a Danish family predisposes to uveal melanoma and other cancers. PLoS ONE *8*, e72144.

Aoude, L.G., Pritchard, A.L., Robles-Espinoza, C.D. et al. (2015). Nonsense mutations in the shelterin complex genes ACD and TERF2IP in familial melanoma. J. Natl Cancer Inst. *107*, 1–7.

Austin, M.T., Xing, Y., Hayes-Jordan, A.A., Lally, K.P., and Cormier, J.N. (2013). Melanoma incidence rises for children and adolescents: an epidemiologic review of pediatric melanoma in the United States. J. Pediatr. Surg. *48*, 2207–2213.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., Mccarthy, S., Mcvean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Balch, C.M., Gershenwald, J.E., Soong, S.J. et al. (2009). Final version of 2009 AJCC melanoma staging and classification. J. Clin. Oncol. *27*, 6199–6206.

Barnhill, R.L., and Kerl, H. (2006). Childhood melanoma and melanoma arising in giant congenital naevi. World Health Organization Classification of Tumours: Pathology and Genetics of Skin Tumours (Lyon: International Agency for Research on Cancer (IARC)), 83-85.

Bett, B.J. (2005). Large or multiple congenital melanocytic nevi: occurrence of cutaneous melanoma in 1008 persons. J. Am. Acad. Dermatol. *52*, 793–797.

Cancer Genome Atlas Network, T. (2015). Genomic classification of cutaneous melanoma. Cell *161*, 1681–1696.

Cannon-Albright, L.A., Goldgar, D.E., Meyer, L.J. et al. (1992). Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. Science *258*, 1148–1152.

Cordoro, K.M., Gupta, D., Frieden, I.J., Mccalmont, T., and Kashani-Sabet, M. (2013). Pediatric melanoma: results of a large cohort study and proposal for modified ABCD detection criteria for children. J. Am. Acad. Dermatol. *68*, 913–925.

Downing, J.R., Wilson, R.K., Zhang, J., Mardis, E.R., Pui, C.H., Ding, L., Ley, T.J., and Evans, W.E. (2012). The pediatric cancer genome project. Nat. Genet. *44*, 619–622.

Erfan, G., Puig, S., Carrera, C., Arance, A., Gaba, L., Victoria, I., Garcia-Herrera, A., Alos, L., and Malvehy, J. (2017). Development of cutaneous toxicities during selective anti-BRAF therapies: preventive role of combination with MEK inhibitors. Acta Derm. Venereol. *97*, 258–260.

Fitzpatrick, T.B. (1988). The validity and practicality of sun-reactive skin types I through VI. Arch. Dermatol. *124*, 869–871.

Garcia-Borron, J.C., Abdel-Malek, Z., and Jimenez-Cervantes, C. (2014). MC1R, the cAMP pathway, and the response to solar UV: extending the horizon beyond pigmentation. Pigment Cell Melanoma Res. *27*, 699–720.

Green, A.C., Wallingford, S.C., and Mcbride, P. (2011). Childhood exposure to ultraviolet radiation and harmful skin effects: epidemiological evidence. Prog. Biophys. Mol. Biol. *107*, 349–355.

Guegan, S., Kadlub, N., Picard, A., Rouille, T., Charbel, C., Coulomb-L'hermine, A., How-Kit, A., Fraitag, S., Aractingi, S., and Fontaine, R.H. (2016). Varying proliferative and clonogenic potential in NRAS-mutated congenital melanocytic nevi according to size. Exp. Dermatol. *25*, 789–796.

Howlader, N., Am, N., Krapcho, M. et al. (2016). SEER Cancer Statistics Review, 1975–2013 (Bethesda, MD: National Cancer Institute) http://seer.cancer.gov/csr/1975_2013/, based on November 2015 SEER data submission, posted to the SEER web site, April 2016.

Kinsler, V.A., Thomas, A.C., Ishida, M. et al. (2013). Multiple congenital melanocytic nevi and neurocutaneous melanosis are caused by postzygotic mutations in codon 61 of NRAS. J. Invest. Dermatol. *133*, 2229–2236.

Krengel, S., Hauschild, A., and Schafer, T. (2006). Melanoma risk in congenital melanocytic naevi: a systematic review. Br. J. Dermatol. *155*, 1–8.

Lek, M., Karczewski, K.J., Minikel, E.V. et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing S. (2009). The sequence alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lorimer, P.D., White, R.L., Walsh, K., Han, Y., Kirks, R.C., Symanowski, J., Forster, M.R., Sarantou, T., Salo, J.C., and Hill, J.S. (2016). Pediatric and adolescent melanoma: a national cancer data base update. Ann. Surg. Oncol. *23*, 4058–4066.

Lu, C., Zhang, J., Nagahawatte, P. et al. (2015). The genomic landscape of childhood and adolescent melanoma. J. Invest. Dermatol. *135*, 816–823.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The ensembl variant effect predictor. Genome Biol. *17*, 122.

Moscarella, E., Zalaudek, I., Cerroni, L., Sperduti, I., Catricala, C., Smolle, J., Hofmann-Wellenhof, R., Sgambato, A., Pellacani, G., and Argenziano, G. (2012). Excised melanocytic lesions in children and adolescents – a 10-year survey. Br. J. Dermatol. *167*, 368–373.

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics *32*, 1749–1751.

Narzisi, G., O'rawe, J.A., Iossifov, I., Fang, H., Lee, Y.H., Wang, Z., Wu, Y., Lyon, G.J., Wigler, M., and Schatz, M.C. (2014). Accurate de novo and transmitted indel detection in exome-capture data using microassembly. Nat. Methods *11*, 1033–1036.

Neier, M., Pappo, A., and Navid, F. (2012). Management of melanomas in children and young adults. J. Pediatr. Hematol. Oncol. *34*(Suppl. 2), S51–S54.

Pappo, A. (2014). Pediatric melanoma: the whole (genome) story. Am. Soc. Clin. Oncol. Educ. Book e432–e435.

Rashid, M., Robles-Espinoza, C.D., Rust, A.G., and Adams, D.J. (2013). Cake: a bioinformatics pipeline for the integrated analysis of somatic variants in cancer genomes. Bioinformatics *29*, 2208–2210.

Reed, D., Kudchadkar, R., Zager, J.S., Sondak, V.K., and Messina, J.L. (2013). Controversies in the evaluation and management of atypical melanocytic proliferations in children, adolescents, and young adults. J. Natl Comp. Cancer Netw. *11*, 679–686.

Robles-Espinoza, C.D., Harland, M., Ramsay, A.J. et al. (2014). POT1 loss-of-function variants predispose to familial melanoma. Nat. Genet. *46*, 478–481.

Shakhova, O., Zingg, D., Schaefer, S.M. et al. (2012). Sox10 promotes the formation and maintenance of giant congenital naevi and melanoma. Nat. Cell Biol. *14*, 882–890.

Shi, J., Yang, X.R., Ballew, B. et al. (2014). Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma. Nat. Genet. *46*, 482–486.

Snyder, A., Makarov, V., Merghoub, T. et al. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. N. Engl. J. Med. *371*, 2189–2199.

Strouse, J.J., Fears, T.R., Tucker, M.A., and Wayne, A.S. (2005). Pediatric melanoma: risk factor and survival analysis of the surveillance, epidemiology and end results database. J. Clin. Oncol. *23*, 4735–4741.

Trozak, D.J., Rowland, W.D., and Hu, F. (1975). Metastatic malignant melanoma in prepubertal children. Pediatrics *55*, 191–204.

Valverde, P., Healy, E., Jackson, I., Rees, J.L., and Thody, A.J. (1995). Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat. Genet. *11*, 328–330.

Van Allen, E.M., Miao, D., Schilling, B. et al. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science *350*, 207–211.

Wiesner, T., Obenauf, A.C., Murali, R. et al. (2011). Germline mutations in BAP1 predispose to melanocytic tumors. Nat. Genet. *43*, 1018–1021.

Wood, B.A. (2016). Paediatric melanoma. Pathology *48*, 155–165.

Zuo, L., Weger, J., Yang, Q., Goldstein, A.M., Tucker, M.A., Walker, G.J., Hayward, N., and Dracopoli, N.C. (1996). Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. Nat. Genet. *12*, 97–99.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1. Cutaneous toxicities associated with vemurafenib in this patient.

Figure S2. Genome-wide mutational landscape of the 15-year-old patient described displayed according to the 96 substitution classification described by Alexandrov et al. (2013).

Table S1. Clinical summary.

Table S2. QC metric of NGS data.

Table S3. Germline profile.

Table S4. All somatic mutations [M_4180].

Table S5. Exonic somatic mutations [M_4180].

Table S6. Mutation rate per megabase.

Table S7. All SNV validations.

Table S8. All CNVs from battenberg.

Table S9. All InDels from scalpel.

Table S10. Translocations from lumpy.

© 2017 The Authors. *Pigment Cell & Melanoma Research* published by John Wiley & Sons Ltd.

Supplementary Figure 1

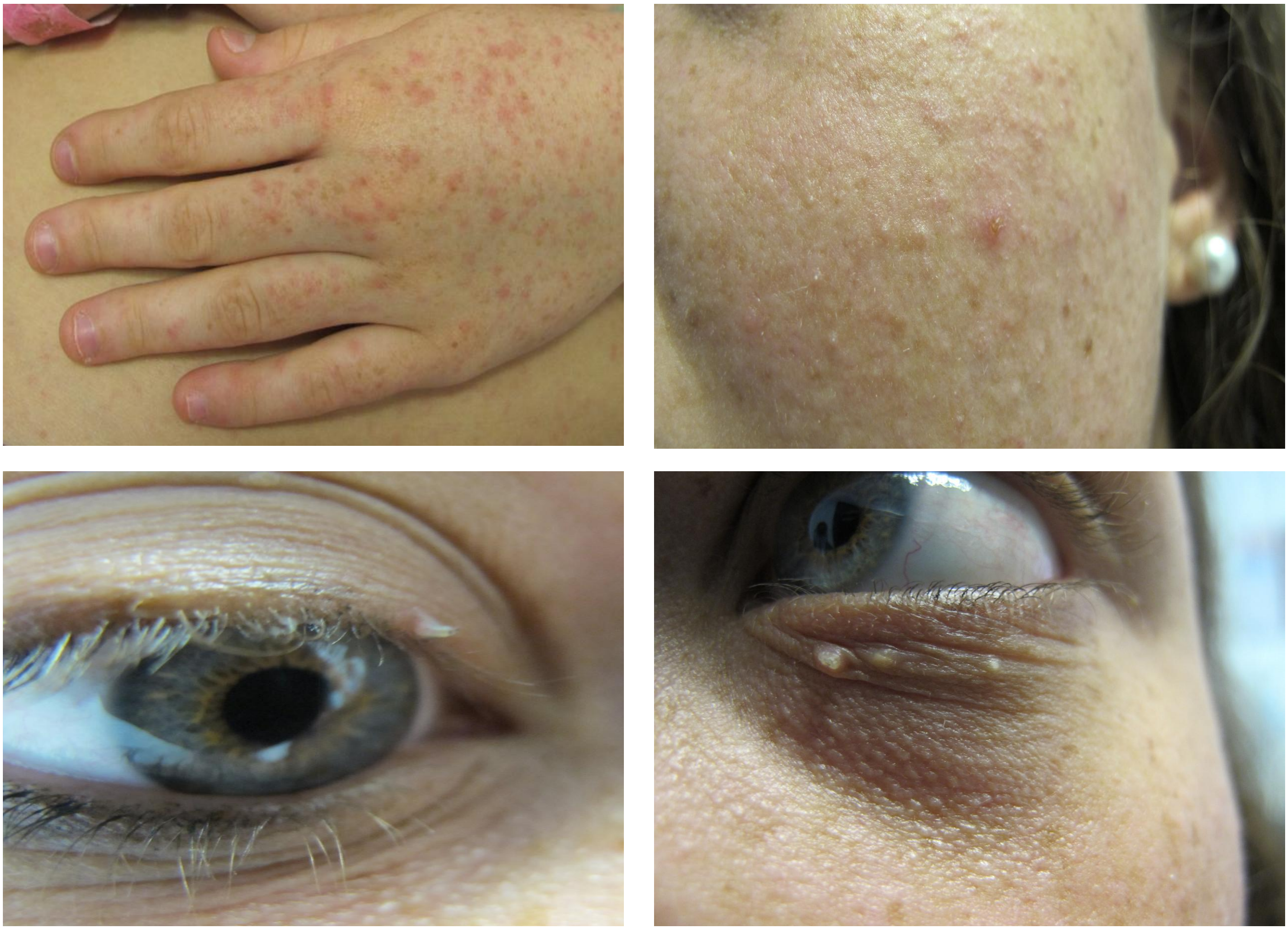


Figure S1. Cutaneous toxicities associated with vemurafenib in this patient. From left upper; photosensitive rash of the hands, follicular hyperkeratosis of cheek, milia cysts of the upper eyelid and papillomatous tumours of the lower eyelid.

Supplementary Figure 2

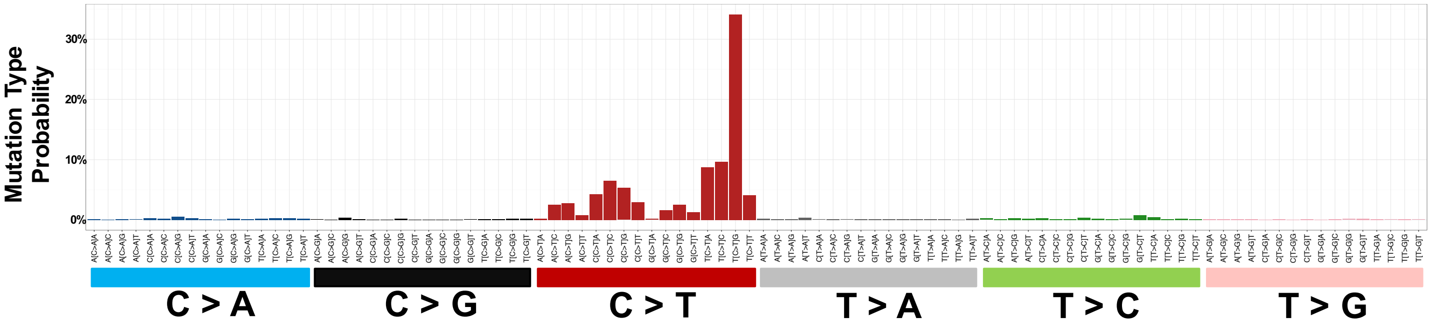


Figure S2. Genome-wide mutational landscape of the 15-year old patient described, displayed according to the 96 substitution classification described by Alexandrov *et al* (2013).

BJC
British Journal of Cancer
www.nature.com/bjc

BRIEF COMMUNICATION
Metastasis

The mutational landscape of melanoma brain metastases presenting as the first visceral site of recurrence

Roy Rabbie[1,2], Peter Ferguson[3,4,5], Kim Wong[1], Dominique-Laurent Couturier[6], Una Moran[7], Clinton Turner[8], Patrick Emanuel[9], Kerstin Haas[10], Jodi M. Saunus[11,12], Morgan R. Davidson[11,12], Sunil R. Lakhani[11,12], Brindha Shivalingam[3], Georgina V. Long[3,4,13], Christine Parkinson[2], Iman Osman[7], Richard A. Scolyer[3,4,5], Pippa Corrie[2] and David J. Adams[1]

Brain metastases are a major cause of melanoma-related mortality and morbidity. We undertook whole-exome sequencing of 50 tumours from patients undergoing surgical resection of brain metastases presenting as the first site of visceral disease spread and validated our findings in an independent dataset of 18 patients. Brain metastases had a similar driver mutational landscape to cutaneous melanomas in TCGA. However, *KRAS* was the most significantly enriched driver gene, with 4/50 (8%) of brain metastases harbouring non-synonymous mutations. Hotspot *KRAS* mutations were mutually exclusive from *BRAF*^{V600}, *NRAS* and *HRAS* mutations and were associated with a reduced overall survival from the resection of brain metastases (HR 10.01, $p = 0.001$). Mutations in *KRAS* were clonal and concordant with extracranial disease, suggesting that these mutations are likely present within the primary. Our analyses suggest that *KRAS* mutations could help identify patients with primary melanoma at higher risk of brain metastases who may benefit from more intensive, protracted surveillance.

British Journal of Cancer (2021) 124:156–160; https://doi.org/10.1038/s41416-020-01090-2

BACKGROUND

Metastases to the central nervous system (CNS) are observed in ~60% of cutaneous melanoma patients developing disseminated disease and up to 90% at autopsy.[1] Early detection of intracerebral recurrence remains critical, as isolated or oligometastatic brain metastases may be more amenable to potentially curative locoregional therapies and immunotherapies have demonstrated greatest efficacy in patients with small, asymptomatic metastases.[1,2] Early predictors of brain metastases could therefore help identify those patients most likely to benefit from closer surveillance of the brain as well as inform early use of adjuvant therapies. Importantly, epidemiological data suggest that patterns of metastatic dissemination may be partially determined by the clinical characteristics of the primary tumour.[3]

Interestingly, 15–20% of brain metastases present as the isolated first visceral site of disease spread.[4] Primary tumours in these 'early brain metastasis' cases were reported as thinner and of lower American Joint Committee of Cancer Stage when compared to other visceral metastases, challenging the current understanding of brain metastases as the final stage of tumour progression, and suggesting that these tumours could harbour distinct biological properties favouring early haematogenous dissemination to the brain.[4] Our analyses of the mutational landscape of early brain metastasis highlights key molecular features that could inform future prognostic, surveillance and intervention strategies.

METHODS

Study population

Patients with available archival paraffin-embedded melanoma brain metastases (in the absence of other sites of visceral disease, confirmed by computed tomography or magnetic resonance imaging prior to neurosurgery) were selected from prospectively maintained databases at The Melanoma Institute of Australia ($n = 34$), The Wellington School of Medicine ($n = 8$), New York University School of Medicine ($n = 4$) and Cambridge University Hospitals ($n = 4$) (discovery cohort). Samples from patients selected from The University of Queensland Australia and the Auckland region New Zealand ($n = 18$ total) made up the external validation cohort. All neuro-resections were undertaken between 2008 and 2018 at the respective academic neurosurgical centres

---

[1]Experimental Cancer Genetics, The Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK; [2]Cambridge Cancer Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK; [3]Melanoma Institute Australia, The University of Sydney, Sydney, NSW 2065, Australia; [4]Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia; [5]Royal Prince Alfred Hospital, Camperdown, NSW 2050, Australia; [6]Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, UK; [7]Interdisciplinary Melanoma Program, New York University School of Medicine, New York, NY, USA; [8]Anatomical Pathology, LabPLUS Auckland City Hospital, Auckland, New Zealand; [9]Anatomic Pathology, The University of Auckland, Auckland, New Zealand; [10]Cancer Genomics Laboratory, Francis Crick Institute, London, UK; [11]UQ Centre for Clinical Research, The University of Queensland Faculty of Medicine, Herston, QLD, Australia; [12]Pathology Queensland, Royal Brisbane Women's Hospital, Herston, QLD, Australia and [13]Royal North Shore and Mater Hospitals, Sydney, NSW 2065, Australia
Correspondence: David J. Adams (da1@sanger.ac.uk)
These authors contributed equally: Roy Rabbie, Peter Ferguson.

Received: 23 February 2020 Revised: 10 June 2020 Accepted: 24 August 2020
Published online: 7 October 2020

© The Author(s) 2020

Published by Springer Nature on behalf of Cancer Research UK

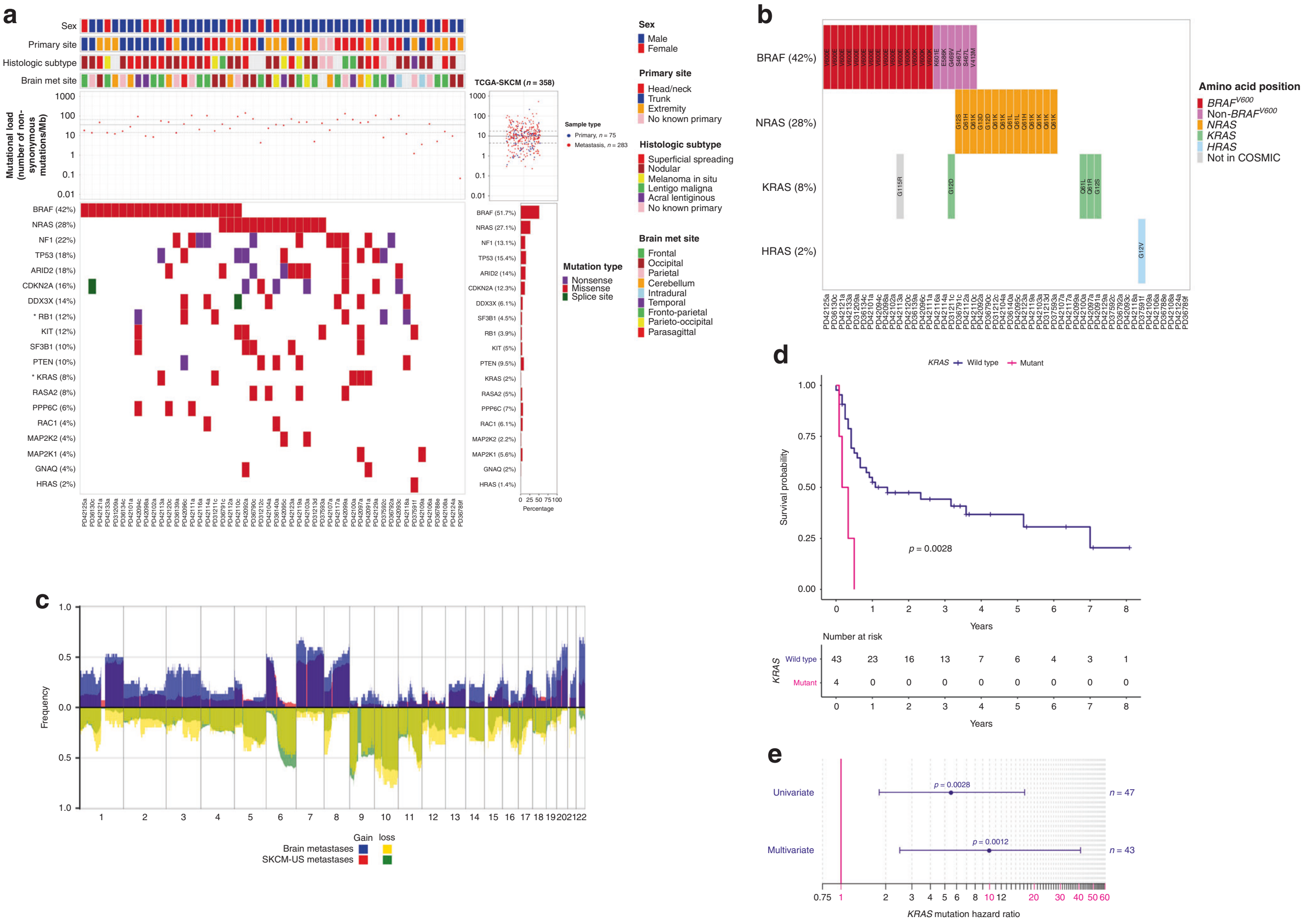


Fig. 1 Tile plot of melanoma-driver mutations in the early brain metastasis discovery cohort ($n = 50$). **a** The mutational profiles of early brain metastases are indicated. Mutational load was calculated as the number of non-synonymous mutations per Mb, median is indicated by the solid horizontal grey line and the 95% confidence interval by the dashed lines (median 38 mutations/Mb, 95% CI 14.2–67.1). The genes shown carry non-synonymous mutations within the selected melanoma drivers outlined in Hayward et al. ($n = 19$) and are ordered according to their mutation frequency within this cohort. The corresponding mutational load and gene mutational frequencies in the SKCM-TCGA dataset ($n = 358$) are indicated. *KRAS*; 8% (4/50) in our dataset vs 2% (7/358) in the SCKM-TCGA collection, $p = 0.0227$, logistic regression Wald *t* test, *RB1*; 6/50 12% in our dataset vs 3.9% 14/358 in SKCM-TCGA, $p = 0.019$, logistic regression Wald *t* test. **b** Focussed tile plot from **a**, highlighting the mutated amino acid positions within the *RAS* signalling genes. As expected, mutations in *NRAS* were mutually exclusive to $BRAF^{V600}$ hotspots. The four hotspot *KRAS* mutations were also mutually exclusive to $BRAF^{V600}$ and to mutations in *NRAS* and *HRAS*. Note that, although we identified five KRAS mutations within the early brain metastasis discovery cohort, the $KRAS^{G115R}$ mutation shaded in grey (occurring in association with a $BRAF^{V600E}$-driver mutation in sample PD42113a) is exceptionally rare[7] and was not considered pathogenic. **c** Copy number profile of the early melanoma metastasis discovery cohort ($n = 30$) overlaid onto the copy number profile of SKCM-TCGA ($n = 337$). The non-overlaid plots are shown in Supplementary Fig. 3. **d** Kaplan–Meier survival curves showing overall survival probabilities from resection of brain metastasis (defined as the time from the resection of the brain metastasis to last follow-up (right-censored) or death from any cause) as a function of time for the hotspot *KRAS*-mutant ($n = 4$) vs *KRAS* wild-type ($n = 43$) patients (3 *KRAS* wild-type patients did not have survival data available). Patients with *KRAS*-mutant tumours had significantly worse overall survival from resection of brain metastasis than *KRAS* wild-type patients, median 3 vs 12 months ($p = 0.003$, univariate Cox regression). **e** Forest plot comparing *KRAS*-mutant vs wild-type hazard ratio for survival from resection of brain metastasis in univariate (HR 5.58, 95% CI 1.80–17.24, $p = 0.003$, $n = 47$) and multivariate (HR 10.01, 95% CI 2.49–40.98, $p = 0.0012$, $n = 43$) Cox proportional hazards regression models. Multivariate correction was undertaken for gender, centre and age at resection of brain metastasis as well as *BRAF* and *NRAS* mutation status.

as part of routine clinical care. All cases were ethically approved by the local Institutional Review Boards, as well as by the Sanger Institute's human materials and data management committee. All samples and clinical details are listed in Supplementary Table 1.

The clinical and mutation data from The Cancer Genome Atlas (SKCM-TCGA)[5] was downloaded from the cBioPortal. The melanoma cases from the Memorial Sloan Kettering MSK-IMPACT data set (SKCM-MSK-IMPACT) were extracted from the publication by Zehir et al.[6] (Supplementary Methods and Supplementary Table 2).

DNA sequencing

Exome capture of the discovery cohort was performed using Agilent SureSelect All Exon V5 baits. Paired-end sequencing was performed using the Illumina HiSeq (Illumina, San Diego, CA, USA) platform at the Wellcome Sanger Institute. MuTect (v1.1.7) and Sequenza (v2.1.2) were used to call somatic single nucleotide variants (SNVs) and copy number aberrations, respectively. Melanoma-driver SNVs called in the whole-exome-sequenced discovery cohort were orthogonally validated (with an aliquot of the same DNA) using a custom gene panel designed to capture ($n = 287$) cancer-driver genes identified from analysis of the TCGA and ICGC cohorts (Supplementary Methods and Supplementary Table 3, ELID ID: 0822402). Panel sequencing of the 18 samples in the external validation cohort was performed using custom pull-down and sequencing of 549 key melanoma and related cancer-driver genes (Supplementary Methods and Supplementary Table 4, ELID ID: 3065404).

Tests of equality of proportions. Wald *t* tests for logistic regression parameters were used to test the equality of mutational frequencies in the discovery cohort and the reference data sets. Similar conclusions were obtained by means of Chi-square and Fisher's exact tests, see Supplementary Methods.

Survival analyses. Kaplan–Meier plots were used to compare survival of *KRAS* mutations within the discovery cohort and The Cancer Genome Atlas (SKCM-TCGA). Between-group differences in instantaneous risk were assessed by fitting univariate and multivariate Cox proportional hazards regression models and defining 95% confidence intervals (CIs) for relevant hazard ratios (HRs). Multivariate models were controlled for further predictors including sex, age, centre, *BRAF* and *NRAS* mutation status (as well as primary tumour characteristics where relevant), see Supplementary Methods for details.

RESULTS

Fifty patients who developed brain metastases as their first site of visceral disease spread were enrolled as part of the discovery cohort and were represented by a relatively high proportion of thin (T1–T2) ($n = 25$, 50%) and non-ulcerated ($n = 26$, 52%) primary melanomas (Supplementary Table 1).

Mutations in *BRAF* were detected in 21 (42%) tumours, of which 15/21 (71%) were in the V600 hotspot (Fig. 1a, b). *NRAS* mutations were identified in 14 (28%) tumours and were all in hotspot positions on exons 2 (codons 12 and 13) and 3 (codon 61) and mutually exclusive from $BRAF^{V600}$ mutations. Comparing the mutational landscape of brain metastases to that of cutaneous melanomas in the SKCM-TCGA dataset (see Supplementary Methods), *KRAS* was the most significantly enriched driver gene in our dataset, mutated in 8% (4/50) vs 2% (7/358) in the entire SCKM-TCGA collection ($p = 0.0227$, logistic regression Wald *t* test). Note that, although we identified 5 *KRAS* mutations within the discovery cohort, the $KRAS^{G115R}$ mutation (occurring in association with a $BRAF^{V600E}$-driver mutation in sample PD42113a) is exceptionally rare[7] and was not considered pathogenic (Fig. 1b). The mutation frequency of *KRAS* was also significantly enriched relative to the frequency of extracranial melanoma metastases; 8% (4/50) in our dataset vs 2.1% (6/274) in extracranial melanoma metastases in SKCM-TCGA ($p = 0.0413$, logistic regression Wald *t* test, see Supplementary Methods). Further, only 1.6% (3/186) of melanoma cases in the Memorial Sloan Kettering MSK-IMPACT dataset were *KRAS* mutant, significantly lower than in our early brain metastasis discovery cohort ($p = 0.0327$, logistic regression Wald *t* test). The odds of observing a *KRAS* mutation in a given sample within the early brain metastases discovery cohort was approximately fourfold higher than in these three reference datasets (Supplementary Fig. 1). Mutations in *KRAS* had a high variant allele frequency (median 0.77 (0.50–0.86), indicating that they likely represent clonal (early) driver mutations (Supplementary Tables 5 and 6). Of note, three extracranial metastases available for sequencing from two patients with *KRAS*-mutant brain metastases also harboured the same brain-metastatic *KRAS* mutations, suggesting that *KRAS* mutations were concordant with extracranial metastases (see Supplementary Methods). Notably, mutations in *KRAS* were in hotspot codons 12 and 61 and mutually exclusive from other mutations in the mitogen-activated protein kinase (MAPK) signalling genes including $BRAF^{V600}$, *NRAS* and *HRAS*, and this pattern of mutually exclusivity was also observed in *KRAS*-mutant melanomas within the SKCM-TCGA and SKCM-MSK datasets (Fig. 1b and Supplementary Table 6).

We conducted a further custom pull-down validation experiment on selected melanoma-driver mutations within the discovery cohort and confirmed 56/60 (93%) to be somatic mutations (see Supplementary Methods). We also conducted another external validation experiment, analysing a further 18 early metastases independently acquired from two different neurosurgical centres (see Supplementary Methods). This revealed that 1 brain metastasis (5.6%) harboured a $KRAS^{G13C}$ mutation, which was also mutually exclusive from mutations in the RAS signalling genes (*BRAF*/*NRAS*/*HRAS*) (Supplementary Fig. 2 and Supplementary Table 5). The copy number landscape of the early brain metastases discovery cohort proved remarkably similar to that of SKCM-TCGA cohort (Fig. 1c and Supplementary Fig. 3).

All patients with *KRAS*-mutant brain metastases succumbed to disease, with a median overall survival from resection of brain metastasis of only 3 months, compared to 12 months in patients with resected *KRAS* wild-type brain metastases (HR 10.01, 95% CI 2.49–40.98, $p = 0.0012$, $n = 43$, covariate corrected Cox proportional hazards model, Fig. 1d, e and Supplementary Fig. 4). Melanoma patients with tumours harbouring *KRAS* mutations or amplifications represented in the SCKM-TCGA dataset were also associated with worse overall survival compared to *KRAS* wild-type melanomas (HR 2.59, 95% CI 1.21–5.55, $p = 0.015$, $n = 352$, univariate Cox regression), although this did not meet the threshold for statistical significance after correction of clinical covariates likely due to the limited sample size (HR 2.04, 95% CI 0.88–4.75, $p = 0.098$, $n = 322$, multivariate corrected Cox proportional hazards regression, Supplementary Fig. 5 and Supplementary Table 2).

DISCUSSION

This analysis represents the largest survey of mutation profiles of melanoma brain metastases. Consistent with the landmark melanoma sequencing studies (primarily based on extracranial metastases),[5,6] early melanoma brain metastases were dominated by a high mutational burden (with a predominance of $C > T$ nucleotide transitions at dipyrimidines) and a similar frequency of the key driver mutations, including *BRAF* (42%), *NRAS* (28%), *NF1* (22%) and *TP53* (18%). This is the first study to show significant enrichment of *KRAS* mutations in melanoma brain metastases as well as an association of *KRAS* mutations with adverse outcomes. The predominance of *KRAS* mutations in codons 12, 13 and 61 as well as the mutual exclusivity with other key drivers of MAPK signalling suggests that these likely represent important drivers in this context.

The RAS family of GTPases consists of genes including *NRAS*, *KRAS* and *HRAS* mutated in 25, 2 and 1% of melanomas, respectively.[5] *NRAS*-mutant melanomas are recognised to be more aggressive and associated with poorer outcomes; however, very little is known about *KRAS*-mutant melanoma.[8] *KRAS*-mutant early brain metastases in our study generally emanated from thin and non-ulcerated primary melanomas (Supplementary Table 7). Hence, *KRAS* detection might in future be used to 'upstage' a subgroup of lower-risk patients not currently offered routine surveillance and/or adjuvant therapy potentially avoiding the devastating impact of brain metastases. Mutations in *KRAS* were clonal and concordant with extracranial disease, which suggests that these mutations are present within the primary tumour; however, further studies will be required to confirm this.

The MAPK and phosphoinositide-3 kinase (PI3K) pathways are the two key downstream signalling pathways through which constitutively activated RAS exerts its pro-tumorigenic effects. MAPK pathway activation and brain metastases are inextricably connected and *BRAF* and *NRAS* mutations are associated with an increased risk of brain metastasis.[9] In the same way, the PI3K/AKT pathway has been mechanistically linked with the development of brain metastases and analyses of patient-matched pairs of brain and extracranial metastases have revealed that brain metastases have higher levels of activated AKT and lower expression of *PTEN*, a finding also observed using immunohistochemistry.[10] Hotpot *KRAS* mutations are known to activate *EGFR* signalling pathways, which in-turn is associated with an increased risk of brain metastases in non-small cell lung cancer.[11]

Transgenic mouse models have established that oncogenic *Kras* can induce naevi and be a founder event in melanomagenesis.[12] In one study, a *Kras*^{G12D} allele was combined with alleles of *p53* or *Lkb1* and with a melanocyte-specific Cre driver to generate a model that developed melanoma with a penetrance of 100%.[13] In this study, metastases were identified in lymph node, lung, liver and spleen but not in kidney or brain. It is therefore important to consider that, while we observe an increased frequency of *KRAS* mutations associated with early brain metastases, it is also plausible that *KRAS* mutations may play a more general role in metastases. Well-conducted in vivo studies will be needed to further uncover the potential for site-specific metastatic tropism of specific *KRAS* variants.

The retrospective nature of this analysis could feasibly introduce a degree of selection bias, in particular by only identifying those patients with operable early brain metastasis we might have excluded a larger patient demographic with more widespread disease. Emerging evidence indicates that metastatic outgrowth may also depend on the interplay between cancer cells and the host stroma; however, such tumour-cell extrinsic factors would not be fully captured by this analysis. The identification of *KRAS* mutations as a predictive biomarker for the development of early brain metastases will ultimately require prospective validation in larger cohorts employing multivariate models, particularly assessing the predictive value of these mutations in relation to other clinical covariates.[3]

In summary, our analyses indicate that the patterns of melanoma recurrence may be at least partially determined by the tumour mutational profile and that up to 8% of patients developing early brain metastases may have tumours driven by oncogenic *KRAS* mutations. This observation has implications for deciphering the biology of site-specific metastatic pathogenesis and, if validated in larger prospectively curated cohorts, might influence prognosis, surveillance and interventions in patients carrying these somatic alterations.

ACKNOWLEDGEMENTS

First and foremost, we sincerely thank the patients involved in this study. We also thank the Cancer, Ageing and Somatic Mutation Program team at the Sanger Institute for facilitating QC and DNA sequencing, particularly Claire Hardy, Stephen Gamble and Elizabeth Anderson; Ingrid Ferreira for performing the histopathological analyses for the University of Queensland cases; and Sofia Chen, James Gilbert and Matthew Garnett for their help with the custom capture baits. Assistance from colleagues at Melanoma Institute Australia and Royal Prince Alfred Hospital is gratefully appreciated. We thank Carla Daniela Robles-Espinoza and Alastair Droop for their critical review of the data and manuscript. A preprint of this manuscript titled 'Hotspot *KRAS* mutations in brain metastases at the first metastatic recurrence of cutaneous melanoma' has been deposited in *bioRxiv*.

AUTHOR CONTRIBUTIONS

R.R. coordinated the clinical and molecular data extraction, sequencing, analysed the data and wrote the paper; P.F. coordinated the samples and clinical data extraction from both the MIA and The Wellington School of Medicine and co-wrote the paper. C.T. and P.E. identified the appropriate cases from across New Zealand and extracted all the samples and associated clinical data. U.M. identified the appropriate cases from New York University Medical Centre and extracted all the associated clinical data. J.M.S. and M.R.D. identified the appropriate cases from the University of Queensland and extracted all the associated clinical data. B.S. performed the neurosurgical procedures within the Melanoma Institute Australia. K.W. performed the bioinformatic analyses including somatic variant and copy number calling and plotted the copy number profiles. K.H. provided copy number calls on SKCM-TCGA samples. D.-L.C. independently reproduced the statistical analyses, including the tests of equality of proportions and Cox regression analyses. S.R.L. (The University of Queensland), I.O. (New York University School of Medicine), G.V.L. and R.A.S. (The Melanoma Institute Australia) funded and facilitated the provision of patient samples, patient's data and materials, reviewed manuscript and provided senior input from the respective comprehensive cancer centres. I.O. was the first to describe 'brain metastases developing as the first site of visceral relapse'[4] and provided expert input on this phenomenon. C.P. and P.C. provided clinical supervision including critical review of the manuscript. D.J.A. provided overall supervision on all aspects of the study as well as critical review of the manuscript. All authors approved the final version.

ADDITIONAL INFORMATION

Ethics approval and consent to participate All cases were ethically approved by the local Institutional Review Boards (REC approval reference numbers; HREC/RPAH/444, 16/CEN/149, 10362, 11/NE/0312, HREC 2005/022 and 16/CEN/149 for the six participating academic centres), as well as by the Sanger Institute's human materials and data management committee. All patients provided written informed consent in accordance with the Declaration of Helsinki. All samples and clinical details are listed in Supplementary Table 1.

Data availability All the whole-exome and targeted sequencing data (including raw sequencing files, variant calls and copy number calls) have been deposited at the European Genome-Phenome Archive (https://www.ebi.ac.uk/ega/ at the EBI) under study accession ID EGAS00001002107.

Competing interests G.V.L. is consultant advisor to Amgen, Aduro, Array, BMS, MERCK, Novartis, Roche and Pierre-Fabre. R.A.S. has received fees for professional services from MSD, BMS, Novartis, Myriad and NeraCare. D.J.A. is a consultant for Microbiotica. All other authors report no conflicts of interest.

Funding information This work was supported by Cancer Research UK and the Wellcome Trust, NHMRC Program Funding to S.R.L. (APP1113867) and NHMRC Program Funding to R.A.S. and G.V.L. (APP1093017). R.A.S. and G.V.L. are supported by NHMRC Practitioner Fellowships. G.V.L. is supported by the University of Sydney Medical Foundation. The funders/sponsors had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript and decision to submit the manuscript for publication.

Supplementary information is available for this paper at https://doi.org/10.1038/s41416-020-01090-2.

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

REFERENCES

1. Ajithkumar, T., Parkinson, C., Fife, K., Corrie, P. & Jefferies, S. Evolving treatment options for melanoma brain metastases. *Lancet Oncol.* **16**, e486–e497 (2015).
2. Long, G. V., Atkinson, V., Lo, S., Sandhu, S., Guminski, A. D., Brown, M. P. et al. Combination nivolumab and ipilimumab or nivolumab alone in melanoma brain metastases: a multicentre randomised phase 2 study. *Lancet Oncol.* **19**, 672–681 (2018).
3. Haydu, L. E., Lo, S. N., McQuade, J. L., Amaria, R. N., Wargo, J., Ross, M. I. et al. Cumulative incidence and predictors of CNS metastasis for patients with American Joint Committee on Cancer 8th Edition stage III melanoma. *J. Clin. Oncol.* **38**, 1429–1441 (2020).
4. Ma, M. W., Qian, M., Lackaye, D. J., Berman, R. S., Shapiro, R. L., Pavlick, A. C. et al. Challenging the current paradigm of melanoma progression: brain metastasis as isolated first visceral site. *Neuro Oncol.* **14**, 849–858 (2012).
5. Cancer Genome Atlas Network. Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
6. Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
7. Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N. et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
8. Jakob, J. A., Bassett, R. L. Jr., Ng, C. S., Curry, J. L., Joseph, R. W., Alvarado, G. C. et al. NRAS mutation status is an independent prognostic factor in metastatic melanoma. *Cancer* **118**, 4014–4023 (2012).
9. Adler, N. R., Wolfe, R., Kelly, J. W., Haydon, A., McArthur, G. A., McLean, C. A. et al. Tumour mutation status and sites of metastasis in patients with cutaneous melanoma. *Br. J. Cancer* **117**, 1026–1035 (2017).

10. Chen, G., Chakravarti, N., Aardalen, K., Lazar, A. J., Tetzlaff, M. T., Wubbenhorst, B. et al. Molecular profiling of patient-matched brain and extracranial melanoma metastases implicates the PI3K pathway as a therapeutic target. *Clin. Cancer Res.* **20**, 5537–5546 (2014).
11. Tan, L., Wu, Y., Ma, X., Yan, Y., Shao, S., Liu, J. et al. A comprehensive meta-analysis of association between EGFR mutation status and brain metastases in NSCLC. *Pathol. Oncol. Res.* **25**, 791–799 (2019).
12. Milagre, C., Dhomen, N., Geyer, F. C., Hayward, R., Lambros, M., Reis-Filho, J. S. et al. A mouse model of melanoma driven by oncogenic KRAS. *Cancer Res.* **70**, 5549–5557 (2010).
13. Liu, W., Monahan, K. B., Pfefferle, A. D., Shimamura, T., Sorrentino, J., Chan, K. T. et al. LKB1/STK11 inactivation leads to expansion of a prometastatic tumor subpopulation in melanoma. *Cancer Cell* **21**, 751–764 (2012).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020

The mutational landscape of melanoma brain metastases presenting as the first visceral site of recurrence

Rabbie *et al.* 2020

Supplemental Methods and Figures

Supplemental Methods

Patient enrolment

Patients with surgically resected brain metastases and no other visceral sites of metastatic disease (assessed prior to neurosurgery by CT or MRI imaging), were eligible for inclusion. Patients with available archival paraffin embedded melanoma brain metastases suitable for DNA extraction were selected from prospectively maintained databases from; The Melanoma Institute of Australia (n=34), The Wellington School of Medicine (n=8), New York University School of Medicine (n=4) and Cambridge University Hospitals (n=4) (total 50, discovery cohort). Samples from patients selected from the Auckland region and The University of Queensland Australia (9 and 9 cases respectively) made up the external validation cohort. Two additional patients (comprising 3 additional samples) from the discovery cohort with hotspot *KRAS* mutations who had available extracranial tumour tissue for molecular analysis were also sequenced for the presence of concordant extracranial *KRAS*-mutations (patient MBM_Disc_23 sample PD42097c regional lymph node, and patient MBM_Dis_40 samples PD31211d/PD31211e both extracranial skin metastases, clinical details in **Supplementary Table 1** and sequencing data in **Supplementary Data**). All neuro-resections were undertaken between 2008 and 2018 at the respective academic neurosurgical centres as part of routine clinical care. All cases were ethically approved by the local Institutional Review Boards (REC approval reference numbers; HREC/RPAH/444, 16/CEN/149, 10362, 11/NE/0312, HREC 2005/022 and 16/CEN/149 for the six centres outlined above respectively), as well as by the Sanger Institute’s human materials and data management committee. All samples and clinical details are listed in **Supplementary Table 1**.

Extraction of clinical details

Patient demographics, primary tumour characteristics (date of primary diagnosis, Breslow thickness, ulceration, mitotic rate, N stage), date of diagnosis of brain metastasis, date of neurosurgical resection, sites of surgically resected brain metastases as well as date of last follow-up or death, were extracted primarily from prospectively maintained clinical research databases and, if needed, through further review of the clinical record. None of the patients received systemic therapies (included targeted and immunotherapies) prior to the resection of brain metastases.

Extraction and quality assessment of DNA and RNA

For each tumour, a hematoxylin and eosin-stained (H&E) section was reviewed by a consultant histopathologist (PF, RAS, CT and PE) for confirmation of the histopathological diagnosis and to identify appropriate regions for DNA extraction. All samples were obtained as either 1.0 mm diameter cores or 4um tissue sections micro-dissected from the original FFPE block. Germline DNA was extracted from micro-dissected adjacent normal tissue where available (n=36). Genomic DNA was extracted using the QIAamp FFPE Tissue kit from Qiagen according to manufacturer's instructions.

Whole exome-sequencing of the discovery cohort

Exome capture was performed using the Agilent SureSelect Human All Exon V5 platform. Paired-end sequencing was performed using the Illumina HiSeq (Illumina, San Diego, CA, USA) platform at the Wellcome Sanger Institute to generate 75 bp paired-end reads. Sequencing reads were aligned using BWA-MEM (v0.7.17-r1188)[1] to the human reference genome hs37d5. PCR duplicates, secondary read alignments, and reads that failed Illumina chastity (purity) filtering were flagged and removed prior to running variant and copy number calling. For tumour samples (n=16) where no matching germline DNA was available, a panel of 39 FFPE-extracted normals was used to filter germline variants as well as artefacts. To ensure that tumour and matched normal samples had the best reciprocal genotype match, SAMtools mpileup, followed by BCFtools gtcheck were run to detect sample concordance, potential sample swaps and contamination. The resulting median sequencing coverage in the brain metastatic samples from the discovery cohort (excluding PCR duplicates) was 48x (range 11-134x) in the tumour samples and 47x (range 6-149x) in the germline samples.

MuTect (v1.1.7)[2] and Strelka (v2.9.2)[3] were used to call somatic SNVs and indels, respectively. Prior to running Strelka, Manta (1.5.0)[4] was run and candidate large indels were used as input to Strelka. The minimum base quality score for somatic and germline variant calling was set to Phred 30. The Ensembl Variant Effect Predictor[5] was used to predict the effect of variants on genes and proteins, relative to the gene build in Ensembl version 97. To remove artefacts, MuTect variant calls were filtered using a tiered approach, such that only the SNVs that met the following criteria were reported (1) Lower coverage samples (<40x); Variant Allele Frequency (VAF) >0.1. (2) All other depth samples; requirement of at least 2 alternative bases on each strand, total depth >=5, and VAF>=0.1 Variant calls found in the gnomAD database[6] were removed if the global population variant allele frequency was greater than or equal to 0.01. Mutational load was calculated as the number of non-synonymous mutations per Mb. The alignments for all reported variants in melanoma driver genes (**Fig. 1a**) were visually inspected using JBrowse and IGV.

Of note extracranial tumour tissue was available and whole-exome sequenced for two hotpot *KRAS* mutant patients. Patient MBM_Disc_23 sample PD42097c - regional lymph node also carried a *KRAS*^{Q61R} mutation, and patient MBM_Dis_40 samples PD31211d/PD31211e - extracranial skin metastases also carried *KRAS*^{G12D} mutations, both concordant with the *KRAS* mutation in the matched brain metastases (raw data deposited in **Supplementary Data**).

Copy number profiling of the discovery cohort

Sequenza (v2.1.2)[7] was used to estimate tumour cellularity and ploidy from 30 tumour-normal pairs in the discovery WES cohort (36/50 discovery cohort samples were paired with germline; minus 5 samples with coverage < 20x and one sample with a noisy CNV profile PD36788e), as well as to calculate allele-specific copy number profiles. For each sample, the default best-fit solution was used for the cellularity and ploidy estimates. To call copy number gain or loss in each sample, the neutral copy number was set as the weighted mean copy number of the segments, rounded to the nearest whole number.

Targeted panel sequencing of the external validation cohort

Panel sequencing of the 18 samples in the external validation was performed using custom pull-down and sequencing of 549 key melanoma and related cancer driver genes (**Supplementary Table 3**). A custom capture probe was designed using Agilent Technologies' online software 'Sure Select Design Wizard' (ELID Design ID: 3065404). DNA capture libraries were created using native DNA (paired-end, average insert size 150bp). Libraries were multiplex sequenced using the Illumina HiSeq platform, excluding reads from PCR duplicates. Only samples with an average depth of >10x (excluding PCR duplicates) across the entire bait were included. SAMtools mpileup was used to identify non-reference bases in: *BRAF* (10 loci covering hotspot SNVs in codons: 469, 586, 594, 599, 600, 601), *NRAS* (12 loci covering hotspot SNVs in codons: 12, 13, 59, 60, 61, 146), *HRAS* (7 loci covering hotspot SNVs in codons: 12, 13, 61) and *KRAS* (15 loci covering hotspot SNVs in codons: 11, 12, 13, 61, 117, 146). Thus, providing a total of 44 loci for interrogation in each of the 18 (tumour-only) samples. Only non-reference variants (with minimum based quality 30 and mapping quality 10) supported by at least 2 alternate bases (from reads not marked as PCR duplicates) are reported. The median sequencing coverage (excluding PCR duplicate reads) across all these hotspot loci in the 18 external validation samples was 44x, with 700/792 (88%) of loci covered with ≥ 10 bases.

Orthogonal validation of SNVs in the discovery cohort

Melanoma driver single nucleotide variants (SNVs) called in the whole-exome sequenced discovery cohort were orthogonally validated (with an aliquot of the same DNA) using a custom gene panel designed to capture (n=287) cancer driver genes in solid tumours within TCGA and ICGC (**Supplementary Table 4**, ELID ID: 0822402). A variant called in the discovery cohort (through whole-exome sequencing) that was also present in the orthogonal validation custom panel (with minimum base quality 30 and mapping quality 10) and supported by at least 1 alternate bases in the validation, is reported as validated somatic. With these criteria, 56/60 (93%) of the substitutions were validated somatic (true positives) (**Supplementary Data**). The (n=4) SNVs called in the discovery cohort that were not validated in the orthogonal validation may represent subclonal mutations that require greater depth rather than false positive calls.

Comparison of the mutational profiles to cutaneous melanomas represented in TCGA and the MSK-IMPACT datasets

In order to compare the mutational rates and frequencies to our dataset (represented by one sample per patient), the clinical and mutation data from The Cancer Genome Atlas (SKCM-TCGA)[8], were downloaded from the cBioPortal[9,10]. Samples were filtered to a single sample per patient giving a total of 358 samples from 358 patients, of which all samples had appropriate SNV data. The demographic characteristics of these 358 SKCM-TCGA patients were largely comparable to our cohort, whereby (n=220, 62%) were male and the median age was 57 years (95% CI 47-71 years) (**Supplementary Table 2**). The majority of these samples (n=283, 79%) were classified as 'metastasis', as opposed to 'primary' (n=75, 21%). The tumour sites were mainly from skin, subcutaneous or nodal metastatic sites, including 144 (40%) classified as from 'extremities', 134 (37%) from 'truncal' locations, 26 (7%) from the 'head and neck', 23 (6%) were 'regional lymph nodes' and the remainder were from other (less frequent) metastatic sites. Only 5 tumour samples in SKCM-TCGA were classified as from the 'brain' and these were excluded when we subsetted to the 'extracranial' metastatic melanoma comparator (n=274).

The MSK-IMPACT dataset was extracted from the publication by Zehir *et al*[11]. Samples were also filtered to a single sample per patient giving a total of 186 samples from 186 patients, all of which had appropriate SNV data. These samples were all labelled as 'cutaneous melanomas', and the majority (n=161, 87%) were classified as 'metastasis' as opposed to 'primary' (n=25, 13%). All primary tumours in this dataset (n=25) were classified as from 'skin' whereas the majority of metastatic samples were from 'regional lymph node' (n=25, 16%), 'lung' (n=24, 15%) and 'in-transit' (n=17, 11%) and the remainder were from other (less frequent) metastatic sites.

Both SKCM-TCGA and SKCM-MSK-IMPACT datasets included only cutaneous melanomas, in particular any melanomas within these datasets from acral, mucosal and other rarer sites were excluded. All SNVs reported from these datasets were non-synonymous mutations. We utilized the mutation calls provided by these resources and did not perform any additional filtering. The somatic mutational rate was calculated as the number of non-synonymous mutations divided by 30Mb, assuming that an average exome has 30Mb in protein-coding genes with sufficient coverage.

Copy number profiling of the SKCM-TCGA cohort

Copy number calls were generated from single nucleotide polymorphism (SNP data) within the SKCM-TCGA dataset using allele-specific copy number analysis (ASCAT version), as previously described.[12]

Statistical methods

All statistical analysis and graphics were generated using R version 3.5.3 (R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/).

Tests of equality of proportions: Logistic regression Wald t-tests (function glm) were used to test the equality of mutational frequencies in our cohort and in the reference datasets. Similar conclusions were obtained when considering Chi-square tests (function chisq.test in R) and Fisher's exact tests (function fisher.test in R). Furthermore, Monte Carlo simulations and considering scenarios corresponding to the observed numbers of successes and failures and based on 25,000 samples showed that the type I error of the chosen statistical methods lies between 0.02 to 0.03 instead of 0.05, suggesting that the conclusions drawn in the results are conservative.

Survival analyses: Overall survival from resection of brain metastasis was defined as the time from the date of resection of the brain metastasis to last follow-up (right-censored) or death from any cause. Overall survival from primary tumour was defined as the time from the date of resection of the primary tumour to last follow-up (right-censored) or death from any cause. Kaplan-Meier survival curves were estimated using the R function surv.fit (‘Survival’ package version 3.1-11). Univariate and multivariate hazard ratios (HR), 95% confidence intervals (95% CI) and corresponding p-values were obtained by fitting Cox proportional hazards regression models by means of the function coxph (‘Survival’ package version 2.4.2.). The multivariate analyses for overall survival from resection of brain metastases was controlled for sex, age at resection of brain metastasis, *BRAF* and *NRAS* mutation status as well as centre. Multivariate analyses for overall survival from primary tumour controlled for all the above, as well as primary tumour depth and ulceration. Multivariate analyses in TCGA dataset controlled for sex, age, *BRAF* and *NRAS* mutation status, Neoplasm disease stage as well as the number of non-synonymous mutations. Model check consist of proportional hazard chi-square tests by means of the function cox.zph ('Survival' package), assessment of the c-statistics defined by means of the function concordance ('Survival' package), and deviance and Schoenfeld residual analyses obtained by means function ggcoxdiagnostics ('survminer' package, version 0.4.6).

Supplementary data

All the whole exome and targeted sequencing data (including raw sequencing files, variant calls and copy number calls) have been deposited at the European Genome-Phenome Archive (https://www.ebi.ac.uk/ega/ at the EBI) under study accession ID EGAS00001002107. Each individual dataset is available under the following dataset accession ID's:

Discovery cohort raw whole-exome sequencing files: EGAD00001005981
Discovery cohort variant calls: EGAD00001005982
Discovery cohort copy number calls: EGAD00001005983
Orthogonal validation for discovery cohort, raw sequencing files: EGAD00001005984
External validation cohort, raw sequencing files: EGAD00001005985

Supplemental Table Legends

Supplementary Table 1. Demographic and clinical characteristics of all samples in the discovery cohort. The table also includes the matched germline samples, as well as the details from three extracranial samples relating to this cohort (the latter with concordant *KRAS* mutations). These are indicated in the column "Tissue type". Cambridge: Cambridge University Hospitals, MIA: Melanoma Institute of Australia, Queensland: The University of Queensland Australia. SSM: Superficial spreading melanoma, NM: Nodular melanoma, LMM: Lentigo-maligna melanoma.

Supplementary Table 2. Demographic and clinical characteristics of all samples in the SKCM-TCGA cohort.

Supplementary Table 3. List of cancer driver genes (n=549) included in the custom capture bait used on the (n=18) external validation samples.

Supplementary Table 4. List of cancer driver genes (n=278) included in the custom capture bait used on the orthogonal validation of driver SNVs (n=60) from the discovery cohort.

Supplementary Table 5. Clinical and mutational characteristics of the patients with hotspot *KRAS* mutations (including both patients from the discovery and validation cohorts). Of note all hotspot *KRAS* mutations were identified in patients with either thin (T1/T2) or no prior history of primary melanoma, and from primary tumours in chronically sun-exposed locations. Mutations in *KRAS* also had a high variant allele frequency, indicating that these likely represent clonal driver mutations, which is further supported by their concordance in extracranial metastases.

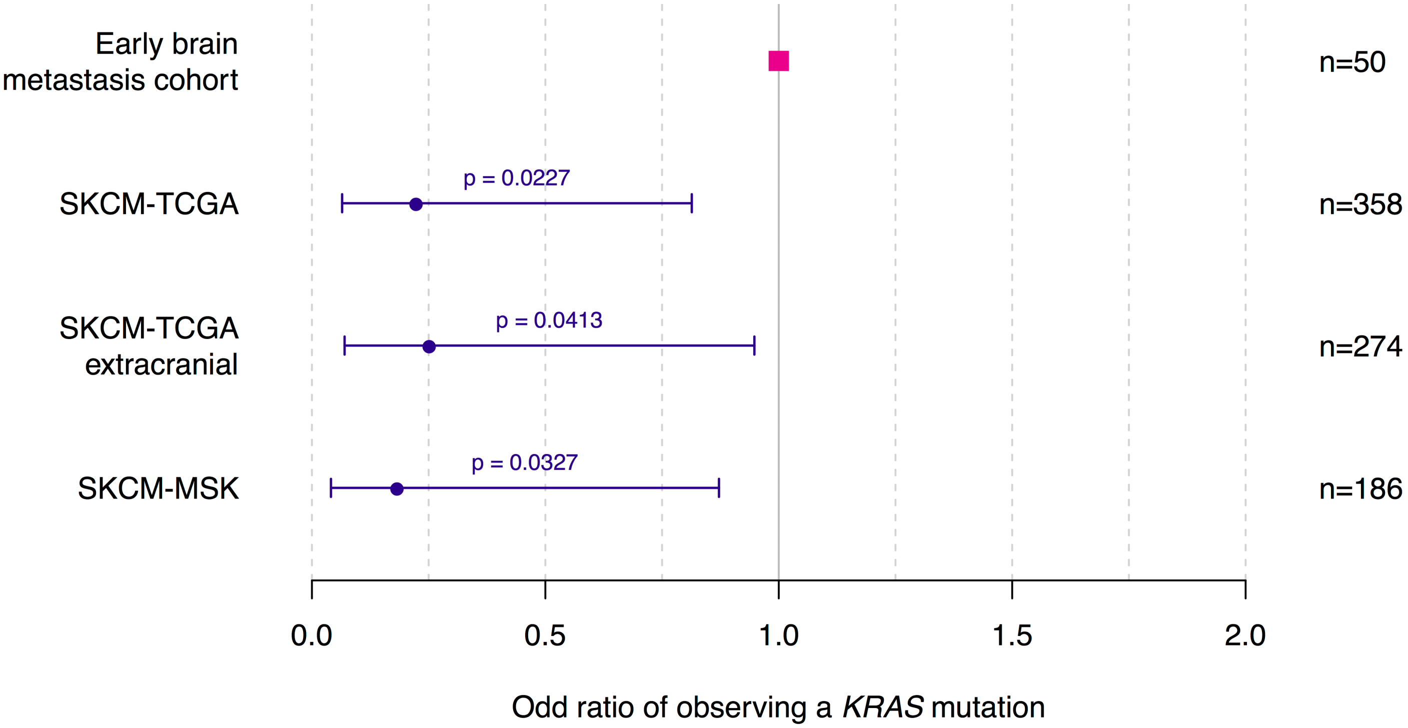
Supplementary Table 6. Clinical and mutational characteristics of the patients with hotspot *KRAS* mutations in SKCM-TCGA (n=6) and SKCM-MSK-IMPACT (n=2). *KRAS* mutations were mutually exclusive to *BRAF/NRAS/HRAS* hotspot mutations in both these datasets.

Supplementary Table 7. Clinical characteristics of hotspot *KRAS*-mutant (n=4) vs *KRAS*-wild type (n=46) patient within the discovery cohort.

References

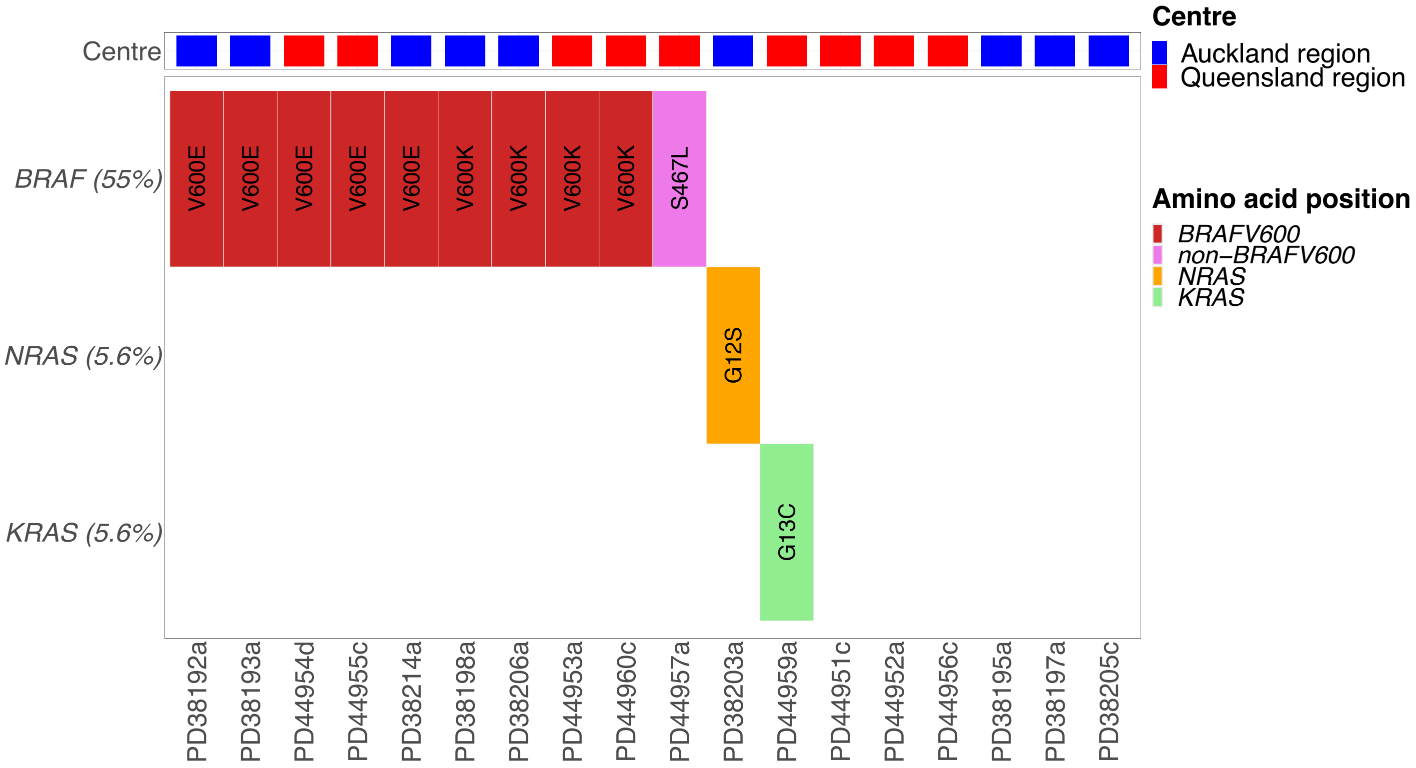
1. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv* **1303** (2013).
2. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219 (2013).
3. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811-1817 (2012).
4. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222 (2016).
5. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
6. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291 (2016).
7. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* **26**, 64-70 (2015).
8. Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681-1696 (2015).
9. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* **2**, 401-404 (2012).
10. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* **6**, pl1 (2013).
11. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* **23**, 703-713 (2017).
12. Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W *et al.* Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A* **107**, 16910-16915 (2010).

Supplementary Figure 1



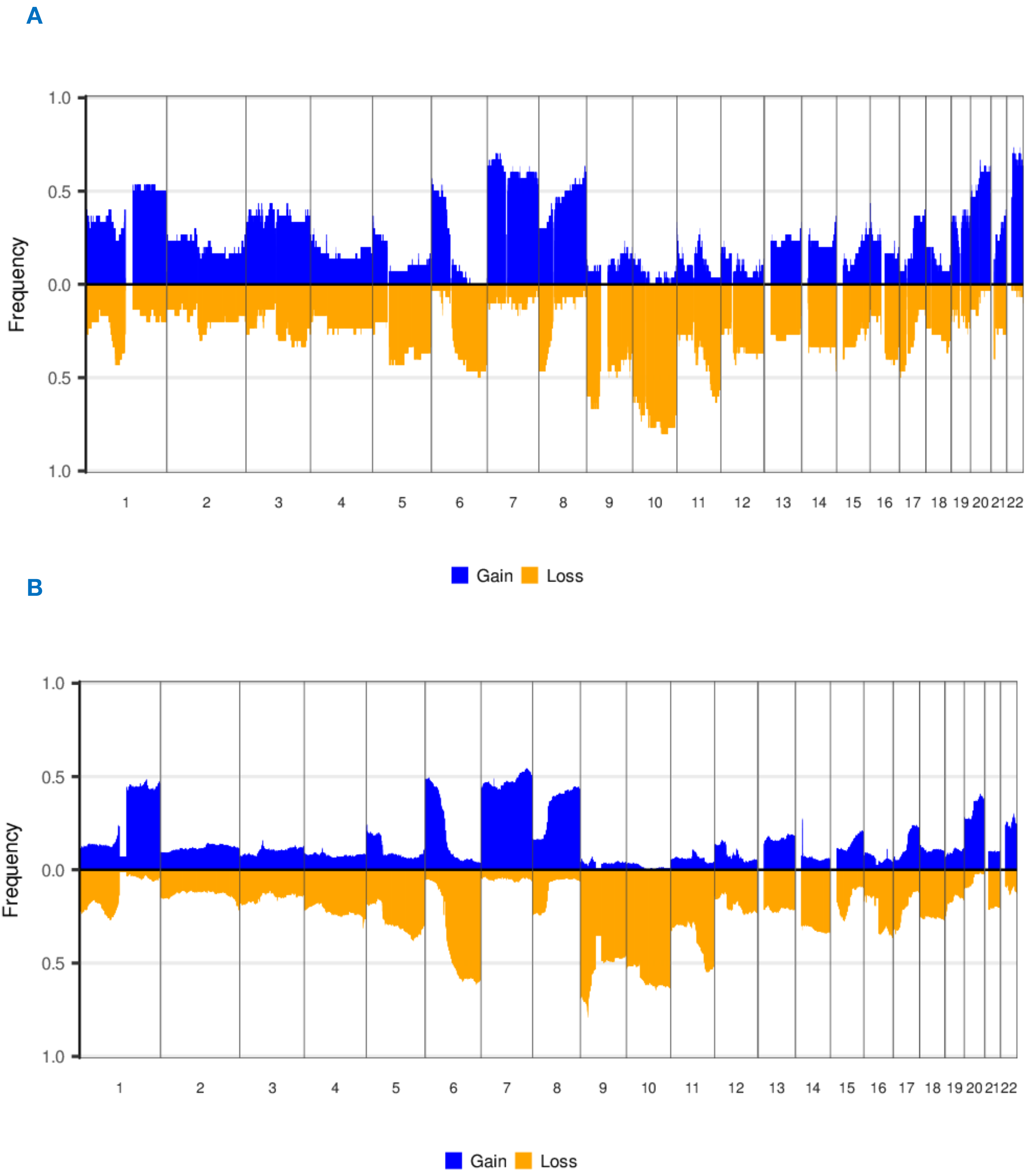
Supplementary Fig. 1. Odds ratio plot showing the odds of observing a *KRAS* mutation in the three reference datasets relative to the early brain metastases discovery cohort (baseline). Dots correspond to logistic regression odd ratio estimates and range to the corresponding 95% confidence interval. The odds of observing a *KRAS* mutation in a given sample within the discovery cohort was ~4-fold higher than in these three reference datasets, this difference is significant.

Supplementary Figure 2



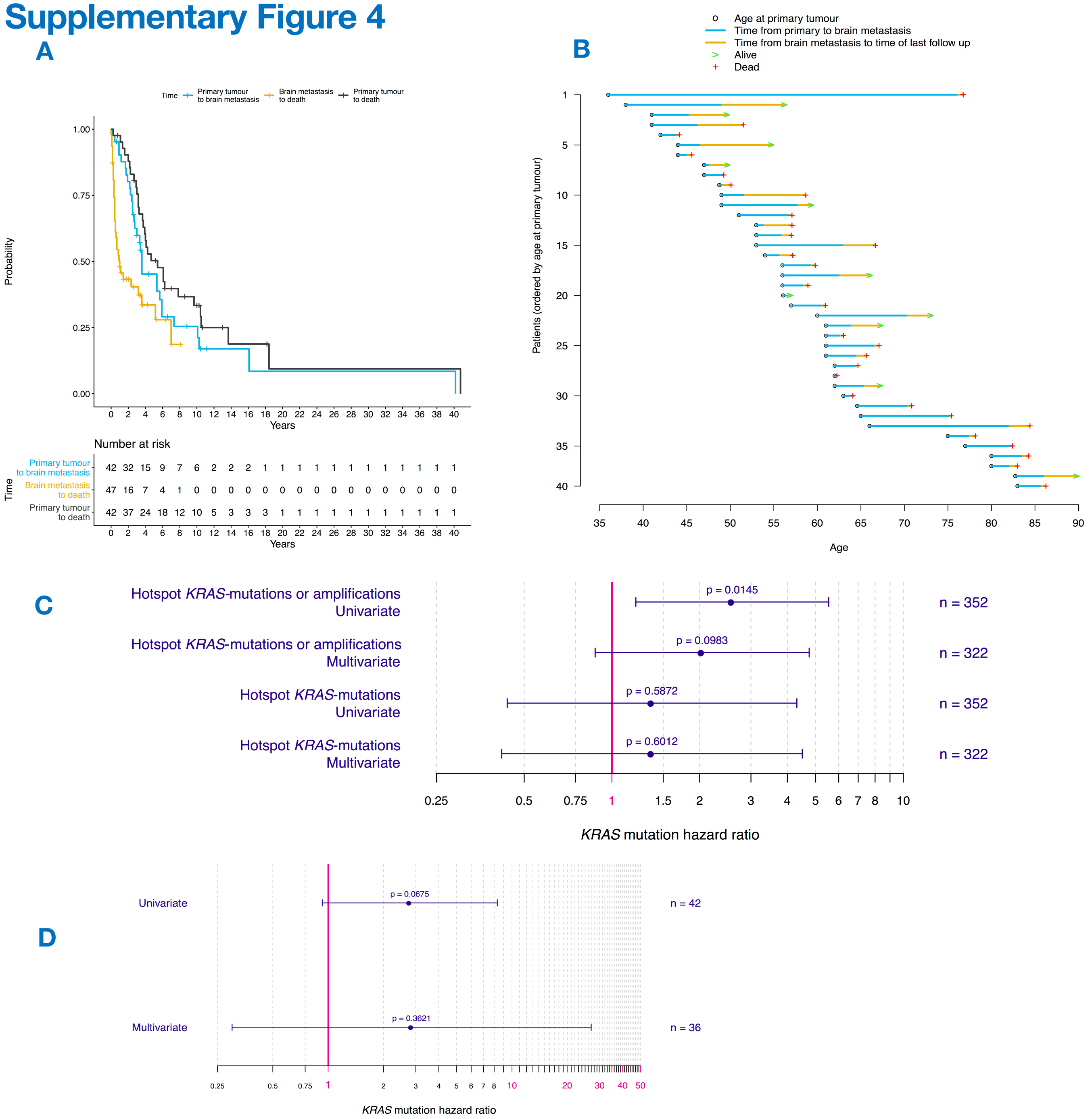
Supplementary Fig. 2. Tile plot of mutated positions within the RAS signalling genes in the external validation cohort (n=18). The *KRAS*^{G13C} mutation was mutually exclusive from both *BRAF* and *NRAS* hotspot mutations, in keeping with the findings from the discovery cohort (**Fig. 1B**).

Supplementary Figure 3



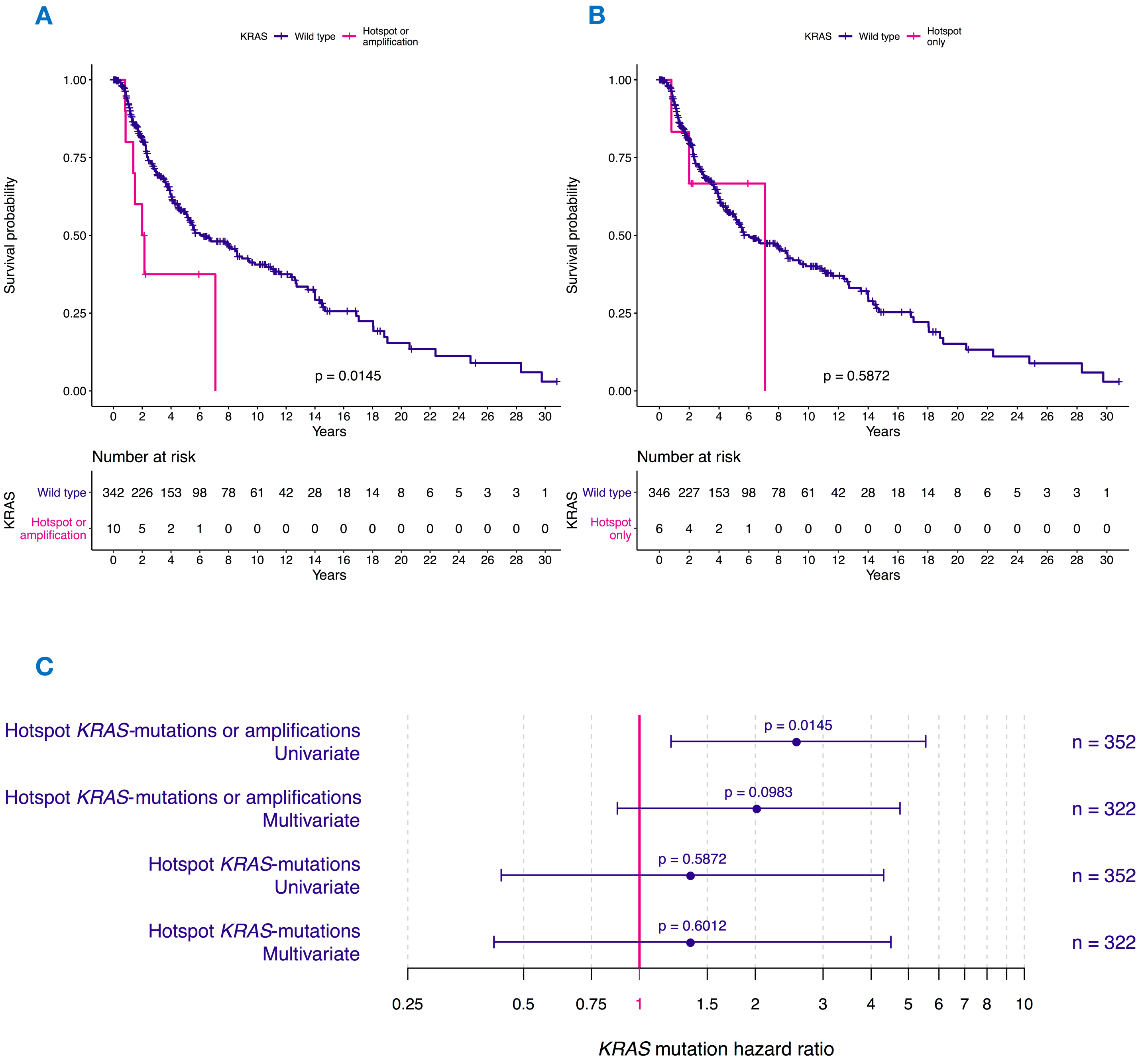
Supplementary Fig. 3. Copy number profile of (A) Early melanoma brain metastases from the discovery cohort (n=30) and (B) Cutaneous melanomas from The Cancer Genome Atlas (SKCM-TCGA, n=337).

Supplementary Figure 4



Supplementary Fig. 4. Survival plots of the early brain metastasis discovery cohort. A) Kaplan-Meier survival plots showing, Black line: survival from primary disease (defined as the time from the resection of the primary tumour to last follow-up (censored) or death from any cause, indicated in black) as a function of time (median survival from primary disease 65.0 months, 95% CI 45.0-125.0 months). Survival data from primary disease was only available on 42 patients (3 patients had no history of primary melanoma, 5 patients had no survival information). Orange line: survival from resection of brain metastasis, defined as the time from the resection of brain metastasis to last follow-up (censored) or death from any cause as a function of time. Median 12.0 months, 95% CI 6.0-62.0 months. Data from resection of brain metastasis was available on 47 patients. Blue line: Time to progression from primary tumour to brain metastases. **B)** Timeline summary for (n=40) patients. Patients are ordered according to the age at primary tumour resection (y-axis). **C)** Impact of *KRAS* mutational status on overall survival from primary tumour as a function of time. **D)** Forest plot comparing *KRAS*-mutant versus *KRAS*-wild-type overall survival from primary tumour in univariate (HR 2.78, 95% CI 0.93-8.32, p=0.0675, n=42) and multivariate (HR 2.85, 95% CI 0.30-26.97, p=0.3621, n=36) Cox proportional hazards regression models. Multivariate correction was undertaken for gender, centre, age at primary tumour resection, T-stage and ulceration of primary tumour, as well as *BRAF* and *NRAS* mutation status.

Supplementary Figure 5



Supplementary Fig. 5. Impact of *KRAS* mutational status on overall survival in SKCM-TCGA. A) Kaplan-Meier curves for overall survival in *KRAS* hotspot mutations or amplifications (n=10) versus *KRAS*-WT (n=342) melanoma in SKCM-TCGA. **B)** Overall survival in *KRAS* hotspot mutation only (n=6) versus *KRAS*-WT (n=346) melanoma in SKCM-TCGA. **C)** Forest plot showing *KRAS*-mutant survival in SKCM-TCGA (hotspot mutations or amplifications) in univariate (HR 2.59, 95% CI 1.21-5.55, p=0.015, n=352) and multivariate analyses (HR 2.04, 95% CI 0.88-4.75, p=0.098, n=322). As well *KRAS*-mutant survival in SKCM-TCGA (hotspot mutations alone) in univariate (HR 1.37, 95% CI 0.44-4.31, p=0.587, n=352) and multivariate analyses (HR 1.37, 95% CI 0.41-4.50, p=0.601, n=322). Generated using Cox proportional hazards regression models. Multivariate correction was undertaken for stage, sex, age at diagnosis of primary, non-synonymous mutation count as well for *BRAF* and *NRAS* mutation status.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-020-18060-0 OPEN

Check for updates

Multi-site clonality analysis uncovers pervasive heterogeneity across melanoma metastases

Roy Rabbie[1,2,13], Naser Ansari-Pour[3,13], Oliver Cast[4], Doreen Lau[5], Francis Scott[5], Sarah J. Welsh[2], Christine Parkinson[2], Leila Khoja[6], Luiza Moore[7,8], Mark Tullett[9], Kim Wong[1], Ingrid Ferreira[1], Julia M. Martínez Gómez[10], Mitchell Levesque[10], Ferdia A. Gallagher[5], Alejandro Jiménez-Sánchez[4], Laura Riva[1], Martin L. Miller[4], Kieren Allinson[8], Peter J. Campbell[7], Pippa Corrie[2], David C. Wedge[3,11,12✉] & David J. Adams[1✉]

Metastatic melanoma carries a poor prognosis despite modern systemic therapies. Understanding the evolution of the disease could help inform patient management. Through whole-genome sequencing of 13 melanoma metastases sampled at autopsy from a treatment naïve patient and by leveraging the analytical power of multi-sample analyses, we reveal evidence of diversification among metastatic lineages. UV-induced mutations dominate the trunk, whereas APOBEC-associated mutations are found in the branches of the evolutionary tree. Multi-sample analyses from a further seven patients confirmed that lineage diversification was pervasive, representing an important mode of melanoma dissemination. Our analyses demonstrate that joint analysis of cancer cell fraction estimates across multiple metastases can uncover previously unrecognised levels of tumour heterogeneity and highlight the limitations of inferring heterogeneity from a single biopsy.

---

[1] Experimental Cancer Genetics, The Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK. [2] Cambridge Cancer Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [3] Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [4] Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, UK. [5] Department of Radiology, School of Clinical Medicine, University of Cambridge, Box 218, Cambridge Biomedical Campus, Cambridge, UK. [6] Institute of Immunology and Immunotherapy, College of Medical and Dental Sciences, Vincent Drive, University of Birmingham, Birmingham, UK. [7] The Cancer, Ageing and Somatic Mutation Programme, Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK. [8] Department of Pathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [9] St Richard's Hospital, Spitalfield Lane, Chichester, UK. [10] Department of Dermatology, University of Zurich, University of Zurich Hospital, Gloriastrasse 31, CH-8091 Zurich, Switzerland. [11] Oxford NIHR Biomedical Research Centre, Oxford, UK. [12] Manchester Cancer Research Centre, University of Manchester, Manchester, UK. [13]These authors contributed equally: Roy Rabbie, Naser Ansari-Pour. ✉email: david.wedge@manchester.ac.uk; da1@sanger.ac.uk

Large-scale sequencing studies in cutaneous melanoma have revealed the complex mutational landscape of the disease[1–4]. However, few studies have explored the temporal and spatial evolution of molecular alterations acquired during disease progression. Such findings may inform risk and our understanding of the mode of metastatic spread, with implications for future patient management. Current methods for reconstructing evolution from bulk sequencing data rely on computational approaches to identify sets of mutations that are present in a similar proportion of cells within the tumour[4]. For each set of mutations, the fraction of cancer cells (cancer cell fraction, CCF) carrying them may be estimated from their allele frequencies by adjusting for purity and copy number. Recent studies have used these algorithms to infer the evolutionary relationship between cell populations across multiple samples, correlating these insights with changes in disease progression or therapy response. As neoplastic cells proliferate, some of their daughter cells can acquire mutations that convey a selective advantage, allowing them to become precursors for new tumour cell lineages[5]. In the metastatic context, dissemination of cells from multiple lineages may cause admixtures of cell populations to spread between different metastases, likely with different CCFs at each site. By clustering mutations according to their CCFs across multiple samples simultaneously, it is possible to identify cell populations from the same lineage spread across multiple sites. Further, by comparing these cell populations based on their CCFs across multiple sites simultaneously, it is possible to derive their ancestral relationship. For example, if one cell population is ancestral to another, its CCF must be greater in at least one sample and greater than or equal to the CCF of the descendant cell population in all other samples, when assuming the infinite sites assumption[6]. It should be noted that by constructing trees from clusters of mutations we avoid potentially inaccurate inferences arising from the construction of sample trees when samples are an admixture of cells from multiple lineages[7]. Moreover, joint analysis of CCFs across multiple samples enables the identification of complex intermixtures of cell populations spread across multiple samples from a primary tumour, as well as complex patterns of tumour cell metastasis[8–14]. Other approaches harnessing sophisticated biogeographic models to reconstruct clonal relationships across multiple samples have also provided detailed spatio-temporal insights of tumoural evolution[15].

Throughout this study, we refer to mutations (and mutation clusters) observed in all tumour cells within a sample as ‘clonal’, those found in a subset of tumour cells as ‘subclonal’ and those found clonally in all samples from the same patient as ‘truncal’. We note that the term ‘trunk’ is used here in the same sense as the term ‘root branch’ in the phylogenetic literature. The term ‘intra-tumour heterogeneity’ (ITH) has been previously used to refer to heterogeneity identified from single or multi-sampling of tissue from a primary tumour. In this paper we extend the definition of ITH to ‘intra-patient tumour heterogeneity’, using it to refer to the observation of variants within a tumour that are non-truncal, including variants that may be clonal within some individual samples.

The mutational load of melanoma is one of the highest among all malignancies[16] and, as somatic mutations provide an insight into a cancer’s initiation and evolution, genome sequencing studies can provide valuable insights into the progression of the disease. Using targeted panel sequencing of 263 cancer driver genes across 12 primary melanomas matched with regional metastases, Shain and colleagues[17] demonstrated that whilst some primary melanomas and matching regional metastases have pathogenic mutations in just one branch of the phylogenetic tree, there were no driver mutations exclusive to metastases (i.e., not shared with the primaries)[17]. By further showing that most somatic alterations (point mutations and copy number changes) were shared, the authors concluded that primary melanomas and melanoma metastases tend to select for the same set of pathogenic mutations. One feature of such studies is that the clonal composition of each sample is determined using the presence or absence of mutations in each sample. However, this type of modelling also relies on the estimation of clone frequencies, which is vital for the identification of two or more clones per sample and for accurate phylogenetic reconstructions[18]. A recent whole-exome sequencing (WES) study of 86 distant metastases obtained from 53 patients used variant allele frequency (VAF, proportion of reads supporting a mutant allele in parallel sequencing data) of shared vs. private mutations in each lesion to infer the likely clonal status of private mutations within each sample[19]. Although many private mutations were subclonal, this study found polyclonal seeding (defined as a sample harbouring subclonal mutations from 2 or more clonal lineages each of which is also found in another tumour site, thus representing multiple seeding events by two or more genotypically distinct cells[8]) to be a rare event[19]. A picture has therefore emerged whereby the majority of mutations in melanoma metastases are truncal and shared by all progeny. Leading up to the formation of a primary melanoma, a stepwise model of progression has been proposed, which includes selection for particular advantageous molecular alterations (including copy number aberrations), facilitating the sequential transition through successive stages[20,21]. Although this model is well-established for the progression of pre-malignant precursor lesions to invasive primary melanomas[22] the evidence for its ubiquity in metastatic progression is less conclusive.

Multi-site sequencing studies in melanoma have thus far been based on a small number of single nucleotide variants (SNVs) falling in coding exons, with gene panels focussed on SNVs in known cancer genes[17,19,23–26]. While the high depth of sequencing used in these studies enables the detection of rare variants, the number of variants detected will be orders of magnitude lower than that from whole genome sequencing (WGS) and some clonal lineages may therefore go undetected. The VAF can also be affected by contributions from alleles in stroma and infiltrating immune cells, as well as the presence of both the mutated and wildtype alleles in the tumour. Importantly, changes to the copy number of a locus may also alter the VAF dramatically and, if not accounted for, will result in inaccurate clonal frequency estimates, giving a misleading picture of the clonal structure of a tumour[18]. For example, a mutation that has occurred on a chromosome that is subsequently duplicated is carried by two out of three chromosomal copies, whereas a mutation that occurred after the gain is carried by one out of three copies. Indeed, whole-genome duplication and other copy number aberrations have been shown to vary across melanoma metastases from the same patient, evolutionary changes that may not be evident from the analysis of SNVs alone[19]. Inferring clonality from allelic frequencies therefore requires an integrative approach harnessing the most sensitive sequencing technologies, while considering measures of tumour ploidy and purity.

In this study, we present a genome-wide analysis of multiple melanoma metastases sampled at autopsy from a treatment-naïve patient. Using multi-sample clonality analyses across 13 whole-genome sequenced metastases from this patient, as well as multi-site analyses of whole-exome sequenced metastases from a further 7 patients, we identify clusters of co-occurring truncal, clonal and subclonal mutations across multiple samples, and uncover the chronological order of genomic alterations. We show that metastases in different organs may have distinct clonal lineages and reveal that melanoma metastases harbour previously unrecognised levels of ITH.

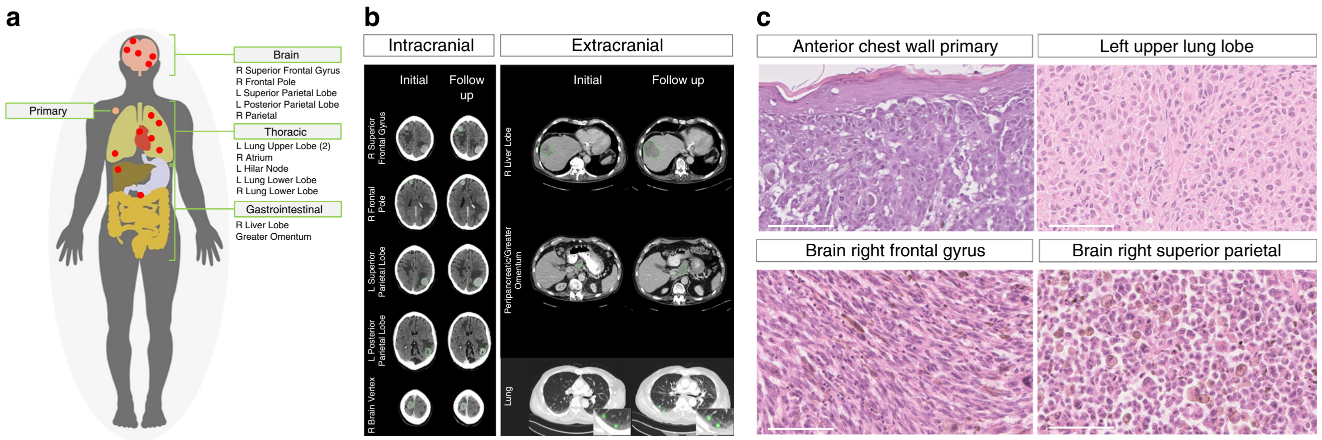


Fig. 1 Clinical presentation and sequalae of index autopsy case. a Sites of 13 metastases sampled during the autopsy and the anterior chest wall cutaneous primary melanoma. **b** Axial CT imaging from the brain, chest and abdomen before (left) and after (right) whole brain radiotherapy (imaging 5 weeks apart). Brain CT images represent the following metastatic sites (from top to bottom); right superior frontal gyrus, right frontal pole, left superior parietal, left posterior parietal and right brain vertex (the latter corresponding to the 'right parietal' sample labelled in the remainder of the text). Chest/abdomen CT images represent the following metastatic sites (from top to bottom); right lobe of liver, peripancreatic/greater omentum (corresponding to the 'greater omentum' sample labelled in the remainder of the text) and right lower lung. **c** Histological analyses (Hematoxylin-eosin images), from top left to bottom right; primary melanoma from the anterior chest wall, distant metastasis from; the left upper lobe of the lung, right frontal gyrus of the brain and right superior parietal lobe of the brain. Morphological appearances differ across the tumours, including varying cellular morphology and pigmentation. One representative region from each tumour sample is shown. The thick white line represents 100 μm in 40-fold magnification.

Results

A clinical course characterised by multi-organ metastases. Our index case was a 71-year-old male of European descent with no relevant family history, who initially presented with a 1.2 mm Breslow thickness non-ulcerated, Clark level 3 superficial spreading melanoma which was resected from the anterior chest wall with a wide local excision (Fig. 1a). The patient declined a sentinel lymph node biopsy and staging scans were clear for distant metastases. Five years later, the patient presented to the emergency department with sudden onset receptive dysphasia and dyspraxia. A contrast computerised tomography (CT) head scan showed multiple enhancing lesions in both cerebral hemispheres with adjacent vasogenic oedema consistent with metastases (Fig. 1b). A staging contrast CT also showed multiple lung, liver and retroperitoneal lymph node metastases (Fig. 1b). A biopsy of one of the liver lesions confirmed metastatic melanoma. Mutation-specific immunohistochemistry for *BRAF*^{V600E} did not detect the mutated protein, but subsequent targeted panel sequencing identified an activating *BRAF*^{V600R} mutation. However, in view of his poor overall performance status and on discussion with the patient and his family, he chose to be managed with best supportive care. The patient consented to undergo a research autopsy as part of the ethically-approved MelResist study (see 'Methods' section). He received corticosteroids with marked improvement in neurological symptoms and underwent whole-brain radiotherapy 30 Gy in 10 fractions. A repeat staging CT scan 2 weeks after completing radiotherapy revealed stable brain metastases but widespread progression of the extracranial disease (Fig. 1b and Supplementary Fig. 1). He died four weeks later.

During a research autopsy, metastases were identified macroscopically in the brain, lung, liver and retroperitoneum, as well as the right atrium, the latter was identified as an 18 mm polypoid lesion arising from the endocardial surface. The cause of death was identified as a saddle pulmonary embolus. In total, 13 metastases were sampled at autopsy and a further 2 samples were obtained from the archived anterior chest wall primary melanoma for further molecular analyses (Fig. 1 and Supplementary Table 1). Histopathological analyses of the metastases confirmed metastatic melanoma. Morphological heterogeneity was observed between metastases based on cellular features (ranging from epithelioid to spindle cell), as well as in the degrees of pigmentation and necrosis (Fig. 1c).

Melanomas are dominated by UV-induced clonal mutations. Whole genome sequencing of the 13 metastatic tumours sampled at autopsy, which were sequenced to a median depth of 38x, revealed a union list of ~118,000 SNVs. We detected 1993 putative somatic indels, of which 10 were frameshifts and common to all metastases (see 'Data availability' section). All 13 metastases carried an activating missense *BRAF*^{V600R} mutation (c.1798_1799delGTinAG), as well as mutations in the melanoma driver genes *PTEN*^{A43T} and *MAP2K1*^{G128S} (the latter has not been previously reported in the COSMIC database[27]), and a splice-site variant in *ARID2*, all of which were truncal across all metastases. We further explored the clonal architecture using the Cancer Cell Fraction (CCF), determined by adjusting the variant allele frequencies of SNVs for copy number aberration (CNA) status and the extent of normal cell contamination (i.e., purity)[4]. Briefly, multidimensional Bayesian Dirichlet Process-based mutation clustering (ndDPClust[4,28]) was used to identify truncal, clonal and subclonal mutation clusters based on the CCF of the union list of somatic SNVs across all 13 metastases. Given that clustering analyses were initially undertaken on the metastatic tumours, subsequent references to metastatic-truncal and metastatic non-truncal mutation clusters apply only to those identified in the metastases of the index autopsy case. Using this approach, we found that >90% of all somatic variants in the metastases were metastatic-truncal, with only one additional cluster which represents at least 1% of the SNVs ($N = 1651$, 1.35%). The large metastatic-truncal cluster was dominated by C > T transitions at dipyrimidines (characteristic of UV-induced mutational damage[29]) and was shared across all metastases, implying that ITH was absent (Fig. 2a). We next filtered for artefactual clusters and SNVs within regions of ambiguous copy number status across all samples (see 'Methods' section) and subtracted variants assigned to the major metastatic-truncal cluster which uncovered a union list of 2247 unique non-truncal variants (Fig. 2b). Overall, 22/2247 of these variants fell within the

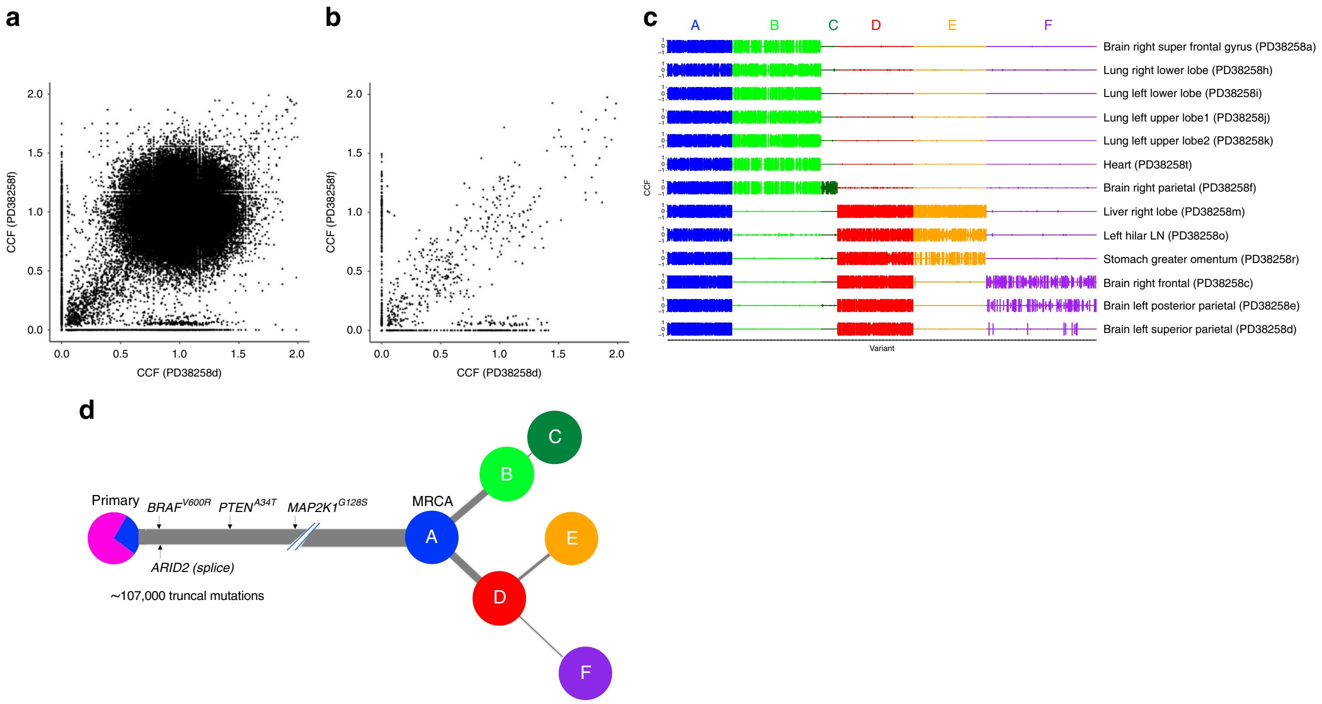


Fig. 2 Subtracting the clonal cluster revealed subclonal diversification in the index autopsy case. a Point estimates of the cancer cell fraction (CCF) are represented here as black dots across two representative samples in a density plot (left super parietal (PD38258d) and right parietal brain metastases (PD38258f), represented across the *X* and *Y* axes, respectively). We observed a large cluster of mutations at (1,1), corresponding to single nucleotide variants (SNVs) present in all the cells in both sites (CCF = 1), indicating truncal variants. **b** After removing the metastatic truncal variants, the metastatic non-truncal mutation clusters uncovered subclonal mutations. Although there are still a small number of SNVs at CCF 1, these did-not belong to the metastatic truncal cluster of mutations, representing those mutations found clonally in all metastases. **c** The clusters of metastatic non-truncal mutations are represented in a CCF distribution plot, wherein rows reflect samples and columns reflect alphabetically and colour-assigned mutation clusters ($n = 6$). The x-axis represents the SNVs within each cluster and the y-axes represent CCF; the latter has been plotted in positive and negative directions for improved visualisation of clusters. **d** Alphabetically and colour-assigned mutation clusters from **c** are represented in circles, where branch length is proportionate to the number of SNVs in the cluster (filtering SNVs occurring in regions with different CNA status across samples reduced 118,000 SNVs to 107,000 truncal SNVs indicated here) and branch thickness is proportionate to the average CCF of the cluster (across all the samples). Mutations that have occurred before the most recent common ancestor (MRCA) are carried by all tumour cells and define the metastatic truncal cluster of mutations. Truncal mutations in the melanoma driver genes *BRAF*^{V600R}, *PTEN*^{A34T}, *MAPK2K1*^{G128S} (the latter represents a mutation not previously reported in the COSMIC database[27]) as well as an ARID2 splice-site variant are indicated on the trunk of the tree (the order of the driver mutations displayed on the trunk is arbitrary). The primary tumour colour shading represents the mutation clusters detected in the primary (using targeted sequencing), whereby the blue represents the subclonal mutation cluster within the primary (at average CCF of 0.27), which became fixed and truncal across all metastases.

protein-coding region of the genome of which 14 were protein-altering (all missense). None of these variants were in established cancer driver genes listed in the Catalogue of Somatic Mutations In Cancer (COSMIC)[27]. We undertook a validation experiment by custom capture pull-down sequencing of protein-coding metastatic-truncal SNVs present across all 13 metastases (selected as either cancer driver or loss-of-function SNVs, $N = 652$), as well as all 2247 metastatic non-truncal SNVs. In this way, 99% of metastatic-truncal SNVs and 92% of all the metastatic non-truncal SNVs were observed to be true variants, instilling confidence in the downstream phylogenetic reconstructions based on these SNVs (see 'Methods' section).

Mutational cluster analyses reveal distinct clonal lineages. Applying ndDPClust to the metastatic non-truncal variants from the index autopsy case revealed 6 distinct mutation clusters with variable distribution across all metastases (Fig. 2c). Assessing the distribution of these clusters, as well as the CCF distribution within each cluster across the metastases, we were able to reconstruct a phylogenetic tree (see 'Methods' section for further details). The mutation clusters showed clear lineage separation, such that samples harbouring clusters from one lineage were mutually exclusive from samples harbouring clusters in the opposing lineage, supporting a clear bifurcation after the most recent common ancestor (MRCA) (Fig. 2c). In particular, cluster B (light green) was present (and clonal) in the first 7 samples belonging to the first clonal lineage and absent in the latter 6 samples belonging to the second clonal lineage, which in-turn were represented by cluster D (red). These mutually exclusive clonal clusters at the first bifurcation suggest that alternative tree solutions are very unlikely (Fig. 2d).

We next assessed the distribution of mutation clusters per sample in order to reconstruct sample-level phylogenetic trees for the metastases (Supplementary Fig. 2). Sample-level trees represent subtrees of the overall phylogenetic tree including just those clones seen within each metastasis, however, in doing this we were able to segregate the samples based on their respective clonal lineage. Analysis in this way revealed two clear lineages, representing distinct waves of metastatic seeding. Interestingly, the brain metastases were represented on both lineages suggesting there were at least two waves of spread to the central nervous system, whereas the lung metastases were derived from a single lineage (Supplementary Fig. 2). In summary, we found that mutation clusters from multiple distinct clones were present across multiple metastatic tumours. This approach has revealed convincing evidence of polyclonal seeding in metastatic melanoma. This finding however could not have been otherwise resolved had it not been for the removal of the dominant cluster of truncal variants, which masked this complex phylogenetic architecture.

Metastatic truncal mutations are subclonal in the primary. We next analysed the representation of metastatic SNVs within two tissue blocks from the original cutaneous primary of the index case resected five years earlier, with the aim of tracing back the ancestral clones. We used targeted panel sequencing with a median coverage of 40×, to ascertain whether selected SNVs identified in the genome-sequenced metastatic tumours were also present in the primary, requiring at least 2 supporting reads reporting the alternative allele to call an SNV in the primary (see 'Methods' section). We found that 573/652 (88%) of the selected metastatic truncal variants could also be detected in the primary, whereas only 1 of the metastatic non-truncal cluster variants was identified in the primary at this depth (see 'Methods' section). By selecting the 144/652 metastatic truncal SNVs present in diploid regions, we were not only able to estimate purity of the two primary samples based on VAF density of SNVs (see 'Methods' section), we were also able to run ndDPClust on both primary tumour samples by assuming diploidy in SNVs within the primary tumours. In addition to the main clonal cluster, we further identified that the two primary tumour samples harboured the same subclone represented by 37 SNVs (at CCF 0.25 95% CI 0.22–0.37 and 0.29 95% CI 0.18–0.34 for samples PD38258u and PD38258v, respectively). No known drivers were uniquely present in this primary tumour subclone. However, we did identify a nonsense variant in *IL1R1*, a gene which is thought to act as a tumour suppressor[30]. This subclonal cluster within the primary tumour might correspond to the lineage that originated the metastases (Fig. 2d). However, the small number of evidential variants warrants further studies in both primary and metastatic melanomas.

Lineage diversification from analyses of 7 further patients. In order to assess whether lineage diversification was detectable in further cases, we undertook whole-exome sequencing (WES) of 19 melanoma metastases matched with germline blood samples from an additional 7 patients with metastatic melanoma who had consented to take part in the MelResist study. All samples were obtained from clinically or radiologically progressing disease sites, either at the time of first distant relapse, or following systemic therapy with MAP-kinase directed therapies or immune checkpoint inhibitors (Supplementary Table 2). We identified an average of 598 (range of 108–2088) non-synonymous coding variants (including missense, nonsense and splice-region mutations), and 7 (range 2–15) frameshift variants per patient, both totalled across all samples within each patient (see 'Data availability' section). Six out of 7 patients had metastases arising from a cutaneous primary and 1 patient (MultiSite_WES_Patient1) had an acral primary melanoma. Their metastases (MultiSite_WES_Patient1), as expected, carried a particularly low number of SNVs (only 298 non-synonymous coding variants totalled across metastases). All 7 patients had metastases harbouring an activating *BRAF*^{V600E} driver mutation and, in accordance with previous reports[17,19], all melanoma drivers were represented on the trunks (rather than the branches) of the phylogenetic trees (except for a previously unreported *TP53*^{R141C} mutation in patient MultiSite_WES_Patient1) (Fig. 3). We again used ndDPClust to cluster SNVs according to their respective CCFs (see 'Methods' section). We identified 2–10 distinct clusters per patient with clear evidence of lineage diversification across 6 out of 7 patients, evidenced by the presence of mutation clusters in mutually exclusive subsets of samples (Supplementary Fig. 3). By reconstructing sample-level phylogenetic trees, we identified distinct clonal lineages within each patient and found evidence of polyclonal seeding in two patients (Supplementary Fig. 4). Given that lineage diversification was detected in 6 out of 7 cases (including the acral melanoma patient) based on WES, which has a much lower genomic resolution than WGS, and with as little as two samples per patient in most cases (Supplementary Table 2), we provide strong evidence that ITH is likely to be pervasive in melanoma metastases.

Subclonal APOBEC signature mutations. We extracted mutational signatures from the 13 whole-genome sequenced metastases collected from the index case[31]. As expected, all samples were dominated by signature 7 reflecting UV-induced mutagenesis (Supplementary Fig. 5). Within the pool of 2247 non-truncal mutations, however, we found evidence of non-UV induced mutational signatures, including signatures 2 and 13, which represent the action of the APOBEC family of cytidine deaminases (which enzymatically modify single-stranded DNA)[32] (Supplementary Fig. 6A). We found that whilst signature 7 dominated the truncal cluster, it was absent from the branches of the evolutionary tree, which were characterised by the APOBEC mutational signatures suggesting this process might be implicated in later stages of clonal evolution (Supplementary Fig. 6B). Interestingly skin cancer has been shown to have the fifth highest *APOBEC3B* expression rank[33]. However, the dipyrimidine-focused C-to-T mutation pattern of UV eclipses an *APOBEC3B* deamination signature, which we have only uncovered here by separating out the truncal mutations.

Gene expression analyses reveal clustering within organs. Gene expression analyses of 11 metastases from the index autopsy case further revealed differences between metastases found across different organs, with principal component analysis (PCA) separating metastases sampled from the brain and lung (Fig. 4a, Supplementary Fig. 7A, B). Interestingly, metastases seeding within the brain clustered together by PCA, despite phylogenetic inferences indicating these likely emanated from differing lineages (Supplementary Fig. 2). In order to identify the tumour-specific genes and biological processes uniquely associated with brain metastases in this patient and mitigate for any potential influence of cellular contamination from the surrounding stromal/immune cells (Fig. 4a), of which purity analyses from copy number calls suggested were <15% (see Supplementary Table 1), we intersected the genes (Fig. 4b) and pathways (Fig. 4c) differentially expressed between both brain metastases ($n = 5$) vs. normal tissue (from the patients' normal brain and lung tissue) (Supplementary Fig. 7C), with those between brain ($n = 5$) vs. lung metastases ($n = 4$) (Supplementary Fig. 7D). The gene *PLEKHA5* was significantly upregulated (log-fold change 4.5, FDR-adjusted p-value < 0.003) in brain vs. lung metastases, as well as in brain metastases vs. normal tissues (log-fold change 5.3, FDR-adjusted p-value < 0.004). This guanine nucleotide exchange factor has previously been shown to be upregulated in a cell line model of melanoma brain metastasis (cerebrotropic A375Br cells vs. parental A375P cells) and silencing of *PLEKHA5* expression decreased in-vitro potential of these cells to cross the blood-brain barrier[34]. Gene set enrichment analyses also showed significant enrichment of the oxidative phosphorylation KEGG pathway in both brain vs. lung metastases (normalised enrichment score 4.65, FDR-adjusted p-value < 0.0001) and in brain metastases vs. normal tissues (normalised enrichment score 3.17, FDR-adjusted p-value < 0.0001) (Fig. 4c). This is consistent with recent analyses implicating the upregulation of oxidative phosphorylation

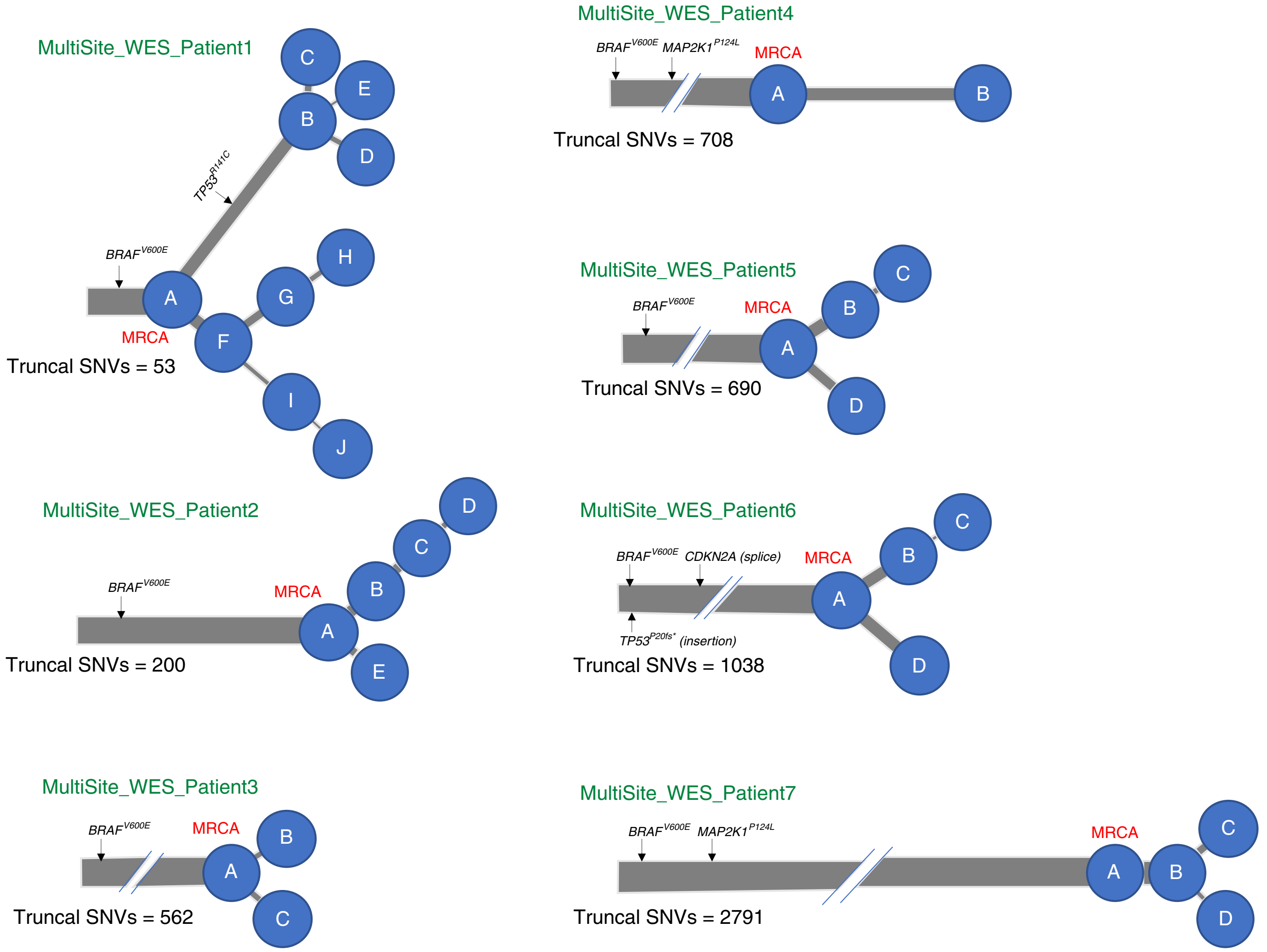


Fig. 3 Multi-dimensional clustering across metastases from 7 patients uncovers divergent lineages. Phylogenetic trees of 7 metastatic melanoma patients are indicated, with branch lengths proportional to the number of SNVs within the mutation cluster and branch thickness proportional to the average CCF of the cluster (across all the samples within each patient). The number of truncal SNVs is indicated underneath the respective phylogenetic trees, most recent common ancestor (MRCA) is labelled in red. All patients harboured an activating *BRAF*^{V600E} driver mutation. In keeping with previous reports[17,19], all other melanoma drivers were represented on the trunk (rather than the branches) of the phylogenetic trees, except for one patient (MultiSite_WES_Patient1) harbouring a previously unreported *TP53*^{R141C} mutation which was unique to one branch and completely absent in the three samples on the opposing branch (despite adequate coverage, >40x).

in patient-matched brain vs. extracranial metastases, as well as further functional studies demonstrating that inhibition of this pathway resulted in increased survival in both implantation xenografts and spontaneous murine models of melanoma brain metastases[35].

Immune cell estimation of bulk tumoural mRNA from the above mentioned 11 metastases (as previously described[36]), further revealed evidence of distinct tumour-immune microenvironments (Fig. 5). The brain metastases in particular had relatively few immune cells compared to lung and other extracranial metastases, which might corroborate studies suggesting the brain represents a relatively immuno-privileged organ[35,37]. Inflammatory macrometastases in the brain expressed high levels of transcripts for activated M2 macrophages (including significant upregulation of the macrophage marker *CD163* in brain vs. lung metastases, log-fold change 5.5, FDR-adjusted p-value < 0.004) which have been described as having anti-inflammatory or tumour supporting activities, including in malignant brain tumours (although it is important to recognise that it may be difficult to distinguish between microglia and macrophages, both of monocyte lineage, using these methods)[38]. In summary, therefore, despite having overall similar mutational landscapes, we identified distinct tumour-immune microenvironments by gene expression analyses, which likely reflects differences in the tumour microenvironment ecosystem[39].

Discussion

Although melanoma is associated with a large number of somatic SNVs, the genomic diversity of melanoma metastases has previously been reported to be low, with most SNVs expected to be shared across tumours[17]. Recently, the International Cancer Genome Consortium Pan-Cancer Analysis of Whole Genomes (PCAWG) initiative[40] leveraged WGS data to infer evolutionary relationships across multiple cancer types, further showing that metastatic melanomas may be monophyletic, with a single clone appearing to seed metastases and, when compared with other cancers, may uniquely lack ITH[41]. In our study, analyses of clonal structure from multi-site whole-genome sequenced melanoma metastases provided a powerful method to detect mutation clusters and a unique insight into clonal evolution. In agreement

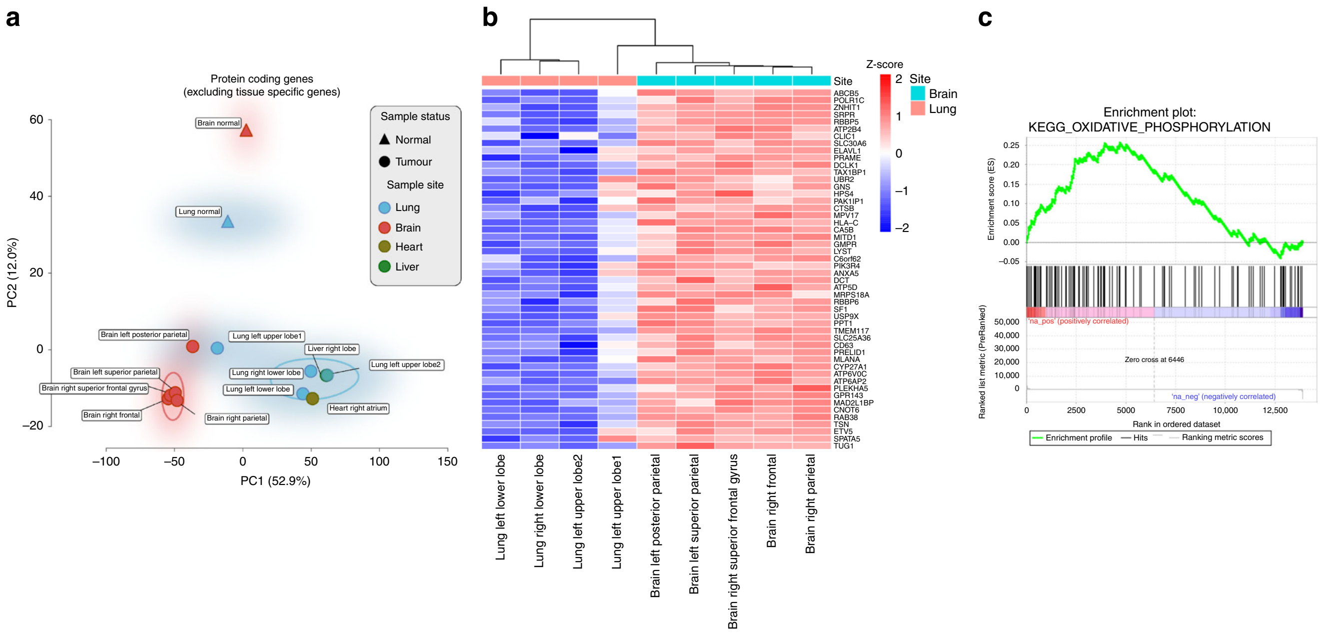


Fig. 4 Gene expression analyses reveals regional separation of site-specific metastases. a Principal component analysis of protein-coding gene expression across all samples. A regional separation can be seen between the brain ($n = 5$, coloured red circles) and lung metastases ($n = 4$, coloured blue circles) on PC1 along the x-axis (these samples are circled using a kernel density estimation), which accounted for greater variance than the separation between the tumour ($n = 11$) and normal samples ($n = 2$) by PC2 along the y-axis (53% vs. 12%, respectively), indicating that expression patterns are at least partially influenced by cellular contamination from the surrounding stromal cells. **b** Heatmap showing the top 50 intersecting genes that are differentially expressed between both brain vs. lung metastases and brain metastases vs. normal tissue comparisons (see "Methods" section). Z-score scale indicates normalised gene expression and represents the number of standard deviations away from the mean (red-blue denoting high-low normalised gene expression, respectively). **c** Pre-ranked gene set enrichment analysis of genes associated with metastases to the brain revealed biological pathways that might have been implicated in brain metastases. The enrichment plot provides a graphical view of the enrichment score for a gene set. The top portion of the plot shows the running enrichment score for the gene set as the analysis walks down the ranked list. The middle portion of the plot shows where the members of the gene set appear in the ranked list of genes and the bottom portion of the plot shows the value of the ranking metric as you move down the list of ranked genes. Oxidative phosphorylation was the most statistically significant over expressed MSigDB KEGG pathway, enriched in both the brain vs. normal tissue and brain vs. lung comparisons (FDR corrected p-value < 0.0001, calculated by constructing a histogram corresponding to null enrichment scores see 'Methods' section), and has recently been linked to melanoma brain metastases in both human and murine analyses[35].

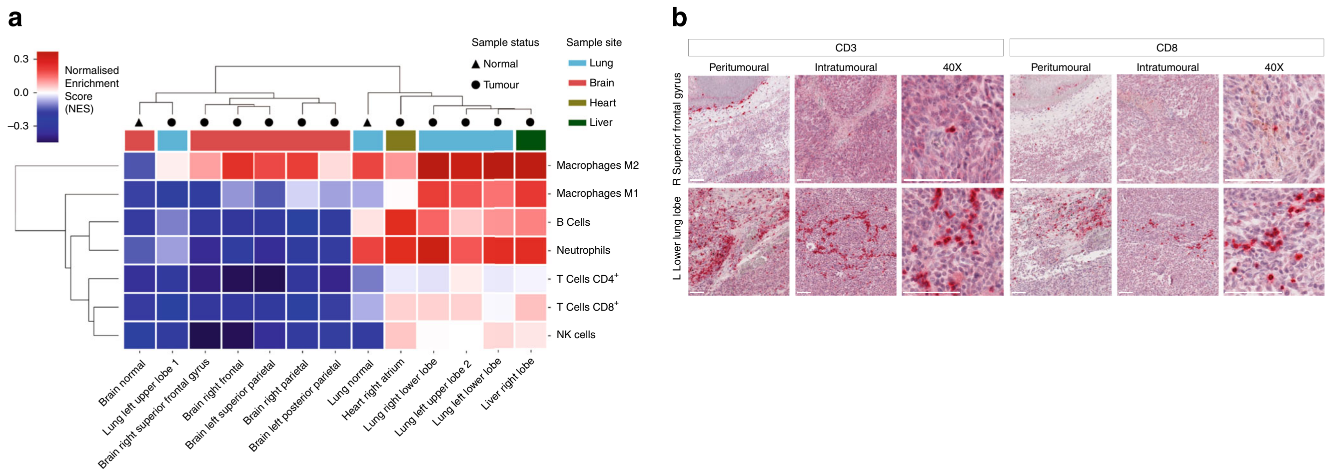


Fig. 5 Regional differences in immune cell representation; brain vs. lung metastases. a Immune cell deconvolution of mRNA using hierarchical clustering of Consensus$^{\mathrm{TME}}$ scores[66]. Scale indicates high-low normalised enrichment scores (red-blue). Both the brain normal and metastases were relatively immuno-privileged when compared to lung normal and lung metastases (note that one of the four lung metastases, left upper lobe 1, was an outlier clustering with the immune-sheltered brain samples). **b** Immunohistochemical (IHC) validation against CD3 and CD8 in a representative brain and lung metastasis. One representative region from each tumour sample is shown. Chromogen red indicates positive immune cell staining. The thick white line represents 100 µm in 10-fold (peritumoural and intratumoural) and 40-fold magnification, respectively.

with previous literature, we identified a single cluster of truncal variants ubiquitously represented across all metastases and representing >90% of all somatic SNVs. Metastatic truncal variants dominated downstream phylogenetic reconstruction analyses. Initial analyses therefore showed that ITH appeared to be absent and that metastases were derived from a single parental clone harbouring the majority of genetic alterations. However, by subtracting this dominant cluster of variants, we were able to identify non-truncal clones and subclones. Assessment of the representation of these clones across the metastatic tumours revealed the ancestral relationships between metastases and uncovered a phylogenetic structure that reframes currently accepted models of metastatic dissemination. In particular, these patient-matched primary and metastatic tumours provided the power to detect SNVs that were present only in a subset of sequenced samples, thereby increasing the power of these reconstruction approaches.

We found evidence of lineage diversification across metastatic melanoma exomes from a further 6 out of 7 patients (including in one case of metastases from an acral primary), indicating that, even with a much lower sequencing breadth, and nearly two orders of magnitude fewer SNVs (relative to whole-genome sequencing), detailed clonal lineages could still be inferred, and pervasive ITH was observed. The detection of ITH using lower-resolution WES from archival formalin-fixed paraffin embedded (FFPE)-derived samples is particularly relevant to clinical practice, where the majority of samples are still stored in paraffin, and where custom pull-down is much more readily available than whole-genome sequencing approaches[42]. It is therefore our impression that previous studies suggesting that melanoma metastases lack heterogeneity may have been confounded either by the use of VAF as a surrogate for CCF[43], or by the lack of power to separate subclones through single sample analyses[41] (rather than by the limits of resolution of targeted sequencing approaches). Our analyses should therefore serve as a cautionary tale in future phylogenetic analyses that still define trunk and branch mutations by the presence or absence of shared variants and that do not consider CCF calculations (adjusting somatic VAFs with tumour purity and CNA status). In a previous single-patient WGS study analysing a primary acral melanoma and its concurrent ipsilateral inguinal lymph node, a wide spectrum of SNVs and copy number alterations were found to be shared between the primary and metastatic tumour, however, the phylogenetic architecture could not be fully reconstructed[44]. By harnessing the power of CCF calculations across 6 metastases from our acral melanoma patient, we identified divergent lineages. A recent detailed multi-regional clonality analysis in uveal melanomas has also found multiple driver mutations in the branches of the phylogenetic trees, suggesting that these melanomas also continue to evolve as they progress from primary to metastatic disease[45]. Therefore, we postulate that multi-sample analyses may reveal that ITH is characteristic of other melanoma subtypes, although further studies in these rarer subtypes are warranted.

Analysing skin/subcutaneous metastases in 8 patients with cutaneous melanoma, Sanborn and colleagues previously showed that locoregional relapses arose from different cellular subpopulations of the primary tumour[46]. Our analyses support these findings, and show that lineage diversification is associated with both locoregional, as well as more distant metastatic spread. The phylogenetic trees were dominated by long trunks, with smaller branches representing subclonal diversification (palm tree resemblance) (Fig. 2d). Driver mutations generally arose before subclonal diversification and were found primarily on the long trunks of the trees (Figs. 2d and 3). This contrasts with recent reports in prostate cancer[11], where branching generally occurred throughout the tumours' evolutionary trajectories, and with studies of various other tumour types reporting the frequent occurrence of subclonal driver mutations, but is concordant with previous studies of melanoma[17,19]. Interestingly, a single subclonal mutation cluster (cluster F, shown in purple in Supplementary Fig. 2) was found subclonally in brain metastases from the index autopsy case. However, in keeping with previous analyses[19,46,47], polyclonal seeding was generally a rare event in this cohort (Supplementary Figs. 2 and 4). Although the index patient underwent whole-brain radiotherapy six weeks prior to the autopsy, we did not detect differences in mutational signatures between the brain and extracranial metastases[31,48] (see 'Data availability' section), while the lack of prior systemic therapy further supports the mutational processes being reflective of evolutionary changes during dissemination, rather than the result of treatment.

The catalogue of somatic mutations in cancer is the aggregate outcome of exposure to one or more mutational processes. Each process generates mutations characterised by a specific combination of nucleotide changes and nucleotide contexts, therefore providing a signature that can be used for its identification[31,49]. Studies by Shain and colleagues suggested that UV is the dominant factor associated with the initiation of precursor lesions and dominates every stage of tumour evolution, from the progression of pre-malignant lesions to primary melanoma and to metastases[20,21]. Interestingly however, other studies have reported a reduction of the proportion of mutations associated with the UV-induced mutational signature in branch (non-truncal) mutations, suggesting that mutations arising later in melanoma progression may occur as a result of increased activity of other mutational processes[17,19]. This is consistent with lung cancer analyses, where subclonal lineages acquired mutations that lacked the tobacco-smoking signature and were replaced with mutations associated with APOBEC cytidine deaminase activity[9]. Consistent with these analyses, we found that whilst the truncal cluster of mutations in the autopsy index case was dominated by signature 7, this was absent in the subclonal mutation clusters which also appeared to be replaced by APOBEC-associated mutations. Further studies will be required to ascertain whether these observations can unravel new biological insights.

In summary, through leveraging the power of clonality analyses across multiple whole-genomes we were able to identify rich clonal architectures and uncover ITH of melanoma metastases obtained at autopsy of a single patient, a structure which would not have been evident through single-site reconstructions. Using the same approach, we found further evidence of divergent lineages in whole-exome sequenced metastases obtained from 6 out of 7 additional melanoma patients, one of which was an acral melanoma, suggesting that this is independent of sequencing breadth or depth. Our ability to detect distinct clonal lineages was greatly enhanced by leveraging the power from multiple samples and uncovers conclusive evidence of ITH in melanoma metastases. Future large-scale studies incorporating clonal analyses across multiple metastases will be required to further delineate how these tumours evolve, and provide insights into whether interrupting this process could contribute to patient management.

Methods

Patient enrolment. All patients were recruited to the MelResisist prospective non-interventional study sponsored by Cambridge University Hospitals NHS Foundation Trust. The study was approved by the National Research Ethics Committee (NREC) North East on the 17th October 2011 IRAS project ID 66161 and REC reference 11/NE/0312. All patients provided written informed consent to take part. All cases were also ethically approved by the Sanger Institute's human materials and data management committee. The research autopsy was conducted 48 h after the patient's death, during which time the body had been stored at 4 °C. Sixteen 1cm-diameter core biopsies were sampled from the centre of each metastatic tumour and snap frozen in liquid nitrogen at −80 °C. These were used for the extraction of bulk DNA and RNA, as well as for the creation of stained H&E slides for direct histopathological assessment of the sequenced regions. The remaining multi-site exome-sequenced cases were also identified from the prospective Mel-Resist trial, and were selected based on their availability of banked multi-site metastases for molecular interrogation. A total of 7 patients with 21 metastases (median 2, range 2–6 metastases per patient) were identified for this analysis. All samples and clinical details are listed in the Supplementary Tables 1 and 2.

Extraction and quality assessment of DNA and RNA. Histopathological assessments were performed by two consultant histopathologists (KA and MT), who confirmed that all tumours were composed of >90% neoplastic cells. Macro-dissection of fresh tumour cores from the autopsy case was performed with a sterile scalpel. DNA and RNA were extracted from the fresh tumour cores using the AllPrep combined DNA/RNA Mini Kit (Qiagen Ltd.) according to manufacturer's recommendations. All the multi-site exome-sequenced cases were obtained as 1.0 mm diameter cores micro-dissected from the original FFPE block. Genomic DNA was extracted from the FFPE cores using the QIAamp FFPE Tissue kit from Qiagen according to manufacturer's instructions. Germline DNA was extracted from peripheral blood mononuclear cells collected before death from all cases, using the DNeasy Blood and Tissue Kit (Qiagen). To confirm that the tumours and germline DNA were derived from the same patient Fluidigm genotyping was performed. All DNA samples were quantified using the PicoGreen dsDNA Quantification Reagent according to manufacturer's recommendations (Invitrogen). The structural integrity of DNA was checked by gel electrophoresis. RNA quantity and quality were assessed using Agilent's 2100 bioanalyzer.

Laser capture microdissection of the cutaneous primary from the autopsy case. Two FFPE tissue blocks from the index autopsy patient's archival primary tumour (cutaneous melanoma from the anterior chest wall, samples PD38258u and PD38258v) were processed into 5 µm histology sections, deparaffinised with ethanol thrice and stained with Gill's haematoxylin for 20 s. Malignant melanocytes from each section were isolated by a histopathologist (LM) using laser capture microdissection and collected in separate Eppendorf tubes. These were then lysed with lysis buffer ATL and digested with proteinase k (Qiagen Ltd.). Extraction of nucleic acids was performed using the QIamp DNA FFPE extraction kit (Qiagen Ltd.) according to manufacturer's recommendation.

Whole genome sequencing and somatic variant detection. Paired-end sequencing of the metastatic tumours and matched normal was performed on the Illumina X10 platform at the Wellcome Trust Sanger Institute to generate 150 base-pair reads. Sequencing reads were aligned using BWA-MEM (v0.7.12)[50] to the human reference genome (NCBI build GRCh37). The resulting sequencing coverage ranged from 33-fold to 43-fold (median 38-fold). Caveman (v1.11.2)[51] and Pindel (v2.2.4)[52] were used to call somatic SNVs and indels, respectively. The minimum base quality score for somatic variant calling was set to Phred 30. ANNOVAR[53] was used to predict the effect of variants on genes and to assign rsIDs for known variants based on dbSNP Human Build 150. The alignments for all variants are reported in the 'Data availability' section. To call rearrangements we applied the BRASS (breakpoint via assembly) algorithm, which identifies rearrangements by grouping discordant read pairs that point to the same breakpoint event (github.com/cancerit/BRASS). BRASS rearrangements were used to search for balanced inversions, which have been previously associated with radiation-induced mutagenesis[48], and were particularly relevant to explore in the radiotherapy-treated brain metastases from the index case.

Whole exome-sequencing of multi-site metastases cases. Exome capture was performed using Agilent's SureSelect bait. Paired-end sequencing was performed using the Illumina HiSeq platform at the Wellcome Trust Sanger Institute to generate 75 bp reads. Sequencing reads were aligned using BWA-MEM (v0.7.12)[50] to the human reference genome GRCh37. PCR duplicates, secondary read alignments, and reads that failed Illumina chastity (purity) filtering were flagged and removed prior to running variant and copy number calling. The resulting sequencing coverage after filtering ranged from 33-fold to 95-fold (median 52-fold) in the tumoural samples and median 54-fold across the germline blood samples. Caveman (v1.11.2)[51] and PINDEL (v2.2.4)[52] were used to call somatic SNVs and indels, respectively. The minimum base quality score for somatic and germline variant calling was set to Phred 30. ANNOVAR[53] was used to annotate SNVs (based on Caveman) and indels (based on PINDEL) for functional classification and to assign rsIDs for known variants based on dbSNP Human Build 150 (see 'Data availability' section).

Copy number aberration (CNA) profiling. Segmental copy number information was derived for each of the 13 metastatic tumours using the Battenberg algorithm (v3.2.2)[10]. This was also used to estimate tumour cellularity and ploidy, and calculate allele-specific copy number profiles[12]. Sequenza (v2.1.2)[54] was used to estimate tumour cellularity and ploidy from the tumour-normal pairs in the multi-site FFPE-extracted exome sequenced cohort, as well as to calculate allele-specific copy number profiles (see 'Data availability' section). For each sample, the best Sequenza solution was chosen after visual inspection of both the best-fit solution (with the maximum log posterior probability) and alternative solutions.

Validation of metastatic truncal and non-truncal SNVs from the index whole-genome sequencing case. Validation was performed using custom pull-down and sequencing of the key mutations identified across the 13 metastases from WGS analysis. The validation experiment was enriched to cover all 2247 metastatic non-truncal variant positions, 652 manually selected metastatic truncal variant positions (identified as either cancer driver mutations or with loss-of-function mutations from the truncal cluster). A 340kbp custom capture probe was designed using Agilent Technologies' online software Sure Select Design Wizard. The highest-stringency repeat masking was used (where possible), as well as a tiling density of 2× and maximum performance boosting (replicating any orphan or GC-rich baits by a higher factor). Agilent ELID ID: 3184291. DNA capture (paired-end, average DNA fragment size 158 bp) libraries were created using native DNA, testing DNA from all 13 whole-genome sequenced metastatic tumours. Libraries were multiplex sequenced to a median depth of 40× on the Illumina MiSeq platform. A variant called in the WGS experiment that was also present in the validation study and supported by at least 2 alternate bases in the validation, is reported as validated somatic.

With these criteria, 7429/7502 (99%) of the metastatic truncal substitutions and 6223/6750 (92%) of the metastatic non-truncal substitutions were validated somatic (the denominator represents the sum of all the SNVs called across all of the 13 samples in the WGS data, excluding those SNVs where coverage in the validation experiment was <30×, see 'Data availability' section). Only 7/7502 of the metastatic truncal mutation calls made in the WGS were not detected in the validation experiment, and 104/6750 (1.5%) of the metastatic non-truncal mutation calls were not detected in the validation experiment. Subsetting only to the metastatic non-truncal substitutions included in the six mutation clusters which passed QC and on which the phylogenetic tree was based (Fig. 2c, d), we found 2127/2231 (95%) SNVs across all 13 metastases were validated (as above, the denominator represents the sum of all the SNVs called across all of the 13 metastases in the WGS data, excluding those SNVs where coverage in the validation experiment was <30×). The breakdown of the validation rate per cluster; Cluster A: 158/192 (82%), Cluster B: 820/837 (98%), Cluster C: 21/24 (88%), Cluster D: 694/720 (96%), Cluster E: 261/276 (95%), Cluster F: 173/182 (95%).

Targeted sequencing of the archival primary. The same baits and custom pull-down experiment described above were used to sequence the validation set SNVs within the index autopsy case's primary tumour. DNA was extracted from the two tumour blocks from the same chest wall primary and variants supported by at least 2 alternate bases were called in the primary (samples PD38258u and PD38258v).

Gene expression analyses. RNA expression of metastatic tumours from the index autopsy case was determined using the human Affymetrix Clariom D Pico assay (see 'Data and software availability' section). Arrays were analysed using the SST-RMA algorithm on the Affymetrix Expression Console Software. Expression was determined using the Affymetrix Transcriptome Analysis Console. Median absolute deviation normalisation (of probe-level data) was implemented. The Tissue-specific Gene Expression and Regulation (TiGER) database was used to filter out 600 tissue-specific genes[55]. Differential expression was performed using the R package limma (v3.36.1)[56]. Preranked GSEA (GSEA-P) was implemented using the GenePattern module *GSEAPreranked* (v6.0.10)[57]. Hallmark gene sets were downloaded from the MSigDB database[58]. Rank metric was calculated as the sign of log2-FCs calculated using the limma pipeline. The pipeline calculates an enrichment score (ES) that reflects the degree to which a gene set is overrepresented at the extremes (top or bottom) of the entire ranked list. The score is calculated by walking down the list, increasing a running-sum statistic when a gene in the set is encountered and decreasing it when a gene not in the set is encountered. The enrichment score is the maximum deviation from zero encountered in the walk and corresponds to a weighted Kolmogorov–Smirnov-like statistic[57]. The significance of an observed enrichment score (ES) is assessed by comparing the enrichment score with a set of scores computed with randomly assigned phenotypes, which generates a histogram of the corresponding null enrichment scores. The nominal P value is then calculated by using the positive (or negative) portion of the distribution corresponding to the sign of the observed enrichment score. To calculate the false discovery rate, the ES is normalised to account for the size of the gene set yielding a normalised enrichment score. The proportion of false positives is then calculated using the false discovery rate (FDR)[59] corresponding to each normalised enrichment score (NES). The FDR is the estimated probability that a set with a given NES represents a false positive finding; it is computed by comparing the tails of the observed and null distributions for the NES[57]. In order to mitigate the influence of cellular contamination from the surrounding stromal cells in identifying particular genes and biological processes uniquely associated with brain metastases in this patient, we intersected the genes (FDR-adjusted P-value < 0.005 and $-1 < \log$ fold-change>1) and pathways (FDR-adjusted p-value < 0.01) significantly differentially expressed between both brain metastases ($n = 5$) and normal tissue (normal samples extracted from the brain and lung, $n = 2$) (Supplementary Fig. 7C), and also between the brain ($n = 5$) and lung metastases

($n = 4$) (Supplementary Fig. 7D). Single sample gene set enrichment analysis (ssGSEA) was employed using the GSVA R package (v1.32.0) to determine the relative enrichment of each of the HALLMARK pathways across samples[60].

Immune cell deconvolution. A consensus approach, Consensus$^{\mathrm{TME}}$ was used to generate cell-type specific estimates of immune cell infiltration from bulk tumour RNA gene expression profiles. This leverages information from multiple gene sets and immune cell expression matrices to build a compendium of robust gene sets for each immune cell type[36]. These genes were further filtered to ensure each has a negative correlation with tumour purity within The Cancer Genome Atlas. Gene sets specific for human skin cutaneous melanoma (SKCM) were used. Finally, single sample gene set enrichment analysis was applied to our Consensus$^{\mathrm{TME}}$ gene sets to generate normalised enrichment scores for each cell type in each sample. This method has previously been thoroughly benchmarked in a pan-cancer setting[36].

Extraction of mutational signatures. SigProfilerMatrixGenerator python packages[61] was used to extract mutational signatures, generating 96 possible mutation types and used to plot mutational profiles. The sigproSS python package (v0.0.0.26)[62] was used to determine the proportion of mutations in each sample attributable to specific COSMIC signatures identified by Alexandrov et al.[31]. sigproSS was also run on the non-truncal mutation clusters defined in the phylogenetic tree (Supplementary Fig. 5).

Computerised tomography analysis of tumour volume. Regions of interest were outlined over the entire area of visible tumours on post-contrast CT scans (2 mm thin sections) by a radiologist (FS) using the OsiriX medical imaging software (Pixmeo SARL, Switzerland). Tumour volume was calculated by multiplying the area of tumour outlined on each CT image by the slice thickness.

Immunohistochemistry. IHC staining was performed on 5 µm FFPE sections, extracted from the same frozen sections from which DNA and RNA were extracted. Slides were deparaffinised in series of xylene and hydrated in a series of descending ethanol. Heat-induced antigen retrieval was performed using TRIS-EDTA (pH = 9), followed by immunostaining performed on the Leica Bond III autostainer (Leica Biosystems). Antibodies used included mouse monoclonal anti-human CD3 (DAKO, clone F7.2.38, dilution 1:50) and mouse monoclonal anti-human CD8 (DAKO, clone C8/144B, dilution 1:25) at 1 h RT. DAKO REAL$^{\mathrm{TME}}$ alkaline phosphatase and chromogen red detection system was used for secondary detection of positive staining. Stained slides were counter-stained with haematoxylin and cover-slipped for review. Image acquisition was performed on the Hamamatsu whole slide scanner at 40-fold magnification.

Statistical analysis and informatics approaches. All statistical analysis and graphics were generated using R version 3.0.1 (R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/). Alignment viewing was performed using Jbrowse and IGV.

Analysis of Intra-patient tumour heterogeneity (ITH) and phylogenetic tree reconstruction. To model the clonal structure across all multi-site tumour samples per patient (at WGS, WES and targeted sequencing levels), we used a previously described computational framework[63]. This approach is an SNV-centric ITH analysis which is described below. In the first step, CCF is estimated for each SNV. By taking into account VAF, CNA status of the SNV locus and purity of the tumour sample under analysis, mutation copy number[63], which is the product of CCF and number of SNV-bearing chromosomal segments, was calculated. CCF is then estimated from mutation copy number by adjusting for the number of SNV-bearing chromosomes, as assessed by a binomial distribution maximum likelihood test[63]. The CCF represents the fraction of tumour cells carrying a mutation, and accounts for differences in tumour purity and copy number[4]. SNVs were removed from further analysis if loss of heterozygosity or any other altered CNA status could explain the complete loss of SNV or its differential VAF in other samples. This filtering is essential to eliminate pseudo-heterogeneity being called among the multiple related samples. The second step is to cluster SNVs based on their CCF by using the Bayesian Dirichlet process-based clustering in a multidimensional mode (ndDPClust (https://github.com/Wedge-Oxford/dpclust)[4]) implemented based on DPClust v2.2.8 (https://github.com/Wedge-Oxford/dpclust) to identify clonal and subclonal clusters across multiple samples of the same patient. Other algorithms including that developed by El-Kebir et al.[64] could also be used to infer the evolutionary history of multiple metastatic tumours (see Alves et al.[15]). However, this requires equivalent data from the matched primary tumour, which was not feasible in this case. The DP clusters (identified as local peaks in the posterior mutation density) are then defined as clonal and subclonal according to their CCF peaks (with an expectation of one cluster at CCF of 1 representing clonal variants). Within individual samples, SNVs are annotated as clonal if they are assigned to the cluster with CCF of 1 and subclonal if assigned to a cluster with lower CCF. SNVs are annotated as truncal when they are clonal across all samples from a patient.

The third step is to reconstruct patient-level phylogenetic trees based on all samples. To determine the most likely phylogenetic tree solution, we applied a previously described mathematical framework[4,10]. Specifically, we applied the previously reported sum and crossing rules[65]. Briefly, the sum rule operates upon the premise that if the CCFs of 2 mutation clusters in any sample add up to more than the CCF of their shared ancestral cluster, they must be collinear. The crossing rule states that if 2 mutation clusters B and C are descendants of mutation cluster A, and if cluster B has higher CCF than cluster C in one sample and cluster C has higher CCF than cluster B in another sample, clusters B and C must be branching. Any mutation cluster that violates these two principles is likely to be an artefact and thus removed from tree reconstruction. It should be noted that the sum rule and crossing rule only strictly apply when the infinite sites assumption is assumed. The model states that each mutation only occurs once during the lifetime of a tumour and never reverts to normal[6].

Given that all metastatic samples were clonally related, only one phylogenetic tree was constructed for each patient. Individual sample trees are subtrees of the overall phylogenetic tree, which include just those clones observed within a single sample. One of the strengths of multi-region sampling is that it exerts a greater inferential restriction on possible phylogenies, since the above stated principles must be simultaneously obeyed across all samples from a patient. We reconstructed the phylogenetic trees for all WGS-based and WES-based patients using clusters representing at least 1% of the clustered SNVs. In tree visualisation, the relative branch lengths were made proportional to the fraction of all SNVs assigned to a cluster and the width of each branch was made proportional to the mean CCF of that cluster across all samples of the patient.

Simulations of phylogenetic trees with variable subclonal heterogeneity. In order to further validate the lineage divergence observed in the index autopsy case, six simulations were undertaken. Briefly, we simulated trees with a trunk of 100,000 SNVs along with a variable set of non-truncal SNVs across four metastases. We used the same genome-wide coverage distribution (Poisson distribution; lambda = 34) and a similar range of purity as those observed in our WGS dataset for the four tumours (0.7–0.95). Sequencing coverage at each locus was sampled from the Poisson distribution and VAF was simulated by sampling the number of mutant reads from a binomial distribution based on the simulated coverage and success rate adjusted by purity and ploidy to give the desired CCF distributions (clonal and subclonal). Three bifurcations were simulated with equal SNV burden on each branch with the first bifurcation represented by two mutually exclusive clonal clusters (see Supplementary Fig. 8). In addition, two second-step branches were assigned as subclonal with different mean CCFs (0.7 vs. 0.3). The first simulation involved only the truncal cluster with no branching clusters (total SNV = 100,000) while the other five simulations included all the other six clusters but with varying SNV burden on each branch, ranging from low to high SNV burden (i.e., 50, 100, 150, 200, and 500 SNVs). This resulted in the total SNV count to range from 100,300 to 103,000 from simulation two to six.

ndDPClust was not only able to detect all clonal and subclonal clusters at all SNV burden levels (including, for instance, distinguishing 50 variants with mean CCF of 0.7 unique to one sample from 50 clonal variants shared between two samples), but it also did not assign any of the non-truncal variants to the truncal cluster. Moreover, in the truncal-only simulation, ndDPClust did not call any non-truncal clusters from the noise incorporated in generating the truncal cluster (especially at this relatively low WGS coverage). This analysis demonstrates that the non-truncal signals detected in our WGS dataset are detectable by this method and unlikely to be a result of noise in the data.

Analysis of intra-patient tumour heterogeneity in the primary tumour of the index autopsy case. As we only had targeted sequencing data on the two FFPE-based thin primary samples, and not WGS, we could not confidently call CNAs in these samples. To estimate CCF, we used the protein altering truncal SNVs ($N =$ 652) restricted to those that were in regions of the genome that were diploid in all metastatic samples ($N = 144$, 22.2%; closely matching the global proportion of SNVs in diploid regions in all metastatic samples i.e., 24.6%) and diploidy was assumed in the primary samples for those loci. The density distribution of VAF was obtained for both primary samples and the peak with the highest VAF was inferred as the clonal set of variants. Purity was then estimated as twice the VAF of the clonal peak. With inferred CNA status and estimated purity, CCF for each SNV was estimated and ndDPClust was run on both primary samples to detect ITH.

Driver mutation analyses. Melanoma drivers were identified as the 20 defined genes with relevant biological evidence outlined in a recently published seminal study[2]. Based on ndDPClust results, driver mutations present in the truncal cluster were assigned as truncal and those present in a non-MRCA cluster which formed a branch were assigned to that branch.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The targeted, whole genome and Affymetrix raw sequencing data have been deposited at the European Genome-Phenome Archive (EGA) (https://www.ega-archive.org at the

European Bioinformatics Institute). Data on all somatic SNVs, indels, inversions and copy number calls for both the index WGS autopsy case and the multi-site WES cases have also been deposited at the EGA under the following accession ID's: EGA Study ID: EGAS00001001348 [ega-archive.org]. Index_autopsy_case: Tumour/normal BAM files for all WGS data. EGAD00001005072. Index_autopsy_case: Caveman, pindel, battenberg and brass calls. EGAD00001005483. Index_autopsy_case: Affymetrix gene expression data. EGAD00010001717. MultiSite_WES_Patients1-7: Tumour/normal BAM files for all WES data. EGAD00001005421. MultiSite_WES_Patients1-7: Caveman, pindel and sequenza calls for all WES data. EGAD00001005487. EGA Study ID: EGAS00001003531 [ega-archive.org]. Index_autopsy_case: Targeted pulldown validation in support of the WGS analysis and targeted pulldown of the primary tumour. EGAD00001005073. The remaining data are available in the Article, Supplementary Information or available from the authors upon request.

Received: 18 November 2019; Accepted: 27 July 2020;
Published online: 27 August 2020

References

1. The Cancer Genome Atlas Network. Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
2. Hayward, N. K. et al. Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175–180 (2017).
3. Krauthammer, M. et al. Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat. Genet.* **44**, 1006–1014 (2012).
4. Bolli, N. et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* **5**, 2997 (2014).
5. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
6. Beerenwinkel, N., Schwarz, R. F., Gerstung, M. & Markowetz, F. Cancer evolution: mathematical models and computational inference. *Syst. Biol.* **64**, e1–e25 (2015).
7. Alves, J. M., Prieto, T. & Posada, D. Multiregional tumor trees are not phylogenies. *Trends Cancer* **3**, 546–550 (2017).
8. Gundem, G. et al. The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).
9. de Bruin, E. C. et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* **346**, 251–256 (2014).
10. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
11. Wedge, D. C. et al. Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nat. Genet.* **50**, 682–692 (2018).
12. Yates, L. R. et al. Genomic evolution of breast cancer metastasis and relapse. *Cancer Cell* **32**, 169–184.e7 (2017).
13. Jamal-Hanjani, M. et al. Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
14. Noorani, A. et al. Genomic evidence supports a clonal diaspora model for metastases of esophageal adenocarcinoma. *Nat. Genet.* **52**, 74–83 (2020).
15. Alves, J. M., Prado-López, S., Cameselle-Teijeiro, J. M. & Posada, D. Rapid evolution and biogeographic spread in a colorectal cancer. *Nat. Commun.* **10**, 5139–5139 (2019).
16. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
17. Shain, A. H. et al. Genomic and transcriptomic analysis reveals incremental disruption of key signaling pathways during melanoma evolution. *Cancer Cell* **34**, 45–55.e4 (2018).
18. Dentro, S. C., Wedge, D. C. & Van Loo, P. Principles of reconstructing the subclonal architecture of cancers. *Cold Spring Harb. Perspect. Med.* **7**, a026625 (2017).
19. Birkeland, E. et al. Patterns of genomic evolution in advanced melanoma. *Nat. Commun.* **9**, 2665 (2018).
20. Shain, A. H. et al. The genetic evolution of melanoma from precursor lesions. *N. Engl. J. Med.* **373**, 1926–1936 (2015).
21. Shain, A. H. & Bastian, B. C. From melanocytes to melanomas. *Nat. Rev. Cancer* **16**, 345–358 (2016).
22. Colebatch, A. J. & Scolyer, R. A. Trajectories of premalignancy during the journey from melanocyte to melanoma. *Pathology* **50**, 16–23 (2018).
23. Harbst, K. et al. Molecular profiling reveals low- and high-grade forms of primary melanoma. *Clin. Cancer Res.* **18**, 4026–4036 (2012).
24. Harbst, K. et al. Molecular and genetic diversity in the metastatic process of melanoma. *J. Pathol.* **233**, 39–50 (2014).
25. Harbst, K. et al. Multiregion whole-exome sequencing uncovers the genetic evolution and mutational heterogeneity of early-stage metastatic melanoma. *Cancer Res.* **76**, 4765–4774 (2016).
26. Anaka, M. et al. Intratumoral genetic heterogeneity in metastatic melanoma is accompanied by variation in malignant behaviors. *BMC Med. Genom.* **6**, 40 (2013).
27. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–d947 (2019).
28. Wedge lab. DPClust R package. https://github.com/Wedge-Oxford/dpclust. Accessed March 2019.
29. Brash, D. E. UV signature mutations. *Photochem. Photobiol.* **91**, 15–26 (2015).
30. Dagenais, M. et al. The Interleukin (IL)-1R1 pathway is a critical negative regulator of PyMT-mediated mammary tumorigenesis and pulmonary metastasis. *Oncoimmunology* **6**, e1287247 (2017).
31. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
32. Smith, H. C., Bennett, R. P., Kizilyer, A., McDougall, W. M. & Prohaska, K. M. Functions and regulation of the APOBEC family of proteins. *Semin Cell Dev. Biol.* **23**, 258–268 (2012).
33. Burns, M. B., Temiz, N. A. & Harris, R. S. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* **45**, 977–983 (2013).
34. Jilaveanu, L. B. et al. PLEKHA5 as a biomarker and potential mediator of melanoma brain metastasis. *Clin. Cancer Res.* **21**, 2138–2147 (2015).
35. Fischer, G. M. et al. Molecular profiling reveals unique immune and metabolic features of melanoma brain metastases. *Cancer Discov.* **9**, 628–645 (2019).
36. Jimenez-Sanchez, A., Cast, O. & Miller, M. L. Comprehensive benchmarking and integration of tumor microenvironment cell estimation methods. *Cancer Res.* **79**, 6238–6246 (2019).
37. Quail, D. F. & Joyce, J. A. The microenvironmental landscape of brain tumors. *Cancer Cell* **31**, 326–341 (2017).
38. Wei, J., Gabrusiewicz, K. & Heimberger, A. The controversial role of microglia in malignant gliomas. *Clin. Dev. Immunol.* **2013**, 285246 (2013).
39. Wang, M. et al. Role of tumor microenvironment in tumorigenesis. *J. Cancer* **8**, 761–773 (2017).
40. Campbell, P. J. et al. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
41. Dentro, S. C. et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. Preprint at biorxiv https://www.biorxiv.org/content/10.1101/312041v5 (2020).
42. De Paoli-Iseppi, R. et al. Comparison of whole-exome sequencing of matched fresh and formalin fixed paraffin embedded melanoma tumours: implications for clinical decision making. *Pathology* **48**, 261–266 (2016).
43. Gartner, J. J. et al. Comparative exome sequencing of metastatic lesions provides insights into the mutational progression of melanoma. *BMC Genom.* **13**, 505 (2012).
44. Turajlic, S. et al. Whole genome sequencing of matched primary and metastatic acral melanomas. *Genome Res.* **22**, 196–207 (2012).
45. Shain, A. H. et al. The genetic evolution of metastatic uveal melanoma. *Nat. Genet.* **51**, 1123–1130 (2019).
46. Sanborn, J. Z. et al. Phylogenetic analyses of melanoma reveal complex patterns of metastatic dissemination. *Proc. Natl Acad. Sci. USA* **112**, 10995–11000 (2015).
47. Brastianos, P. K. et al. Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov.* **5**, 1164–1177 (2015).
48. Behjati, S. et al. Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **7**, 12605 (2016).
49. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
50. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv:1303.3997 (2013).
51. Jones, D. et al. cgpCaVEManWrapper: simple execution of CaVEMan in order to detect somatic single nucleotide variants in NGS data. *Curr. Protoc. Bioinform.* **56**, 15.10.1–15.10.18 (2016).
52. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
53. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
54. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).
55. Liu, X., Yu, X., Zack, D. J., Zhu, H. & Qian, J. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinform.* **9**, 271 (2008).
56. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
57. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

58. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
59. Reiner, A., Yekutieli, D. & Benjamini, Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* **19**, 368–375 (2003).
60. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7 (2013).
61. Bergstrom, E. N. et al. SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. *BMC Genom.* **20**, 685 (2019).
62. Alexandrov lab. SigProfilerSingleSample. https://pypi.org/project/sigproSS/ Accessed June 2019.
63. Yates, L. R. et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.* **21**, 751–759 (2015).
64. El-Kebir, M., Satas, G. & Raphael, B. J. Inferring parsimonious migration histories for metastatic cancers. *Nat. Genet.* **50**, 718–726 (2018).
65. Jiao, W., Vembu, S., Deshwar, A. G., Stein, L. & Morris, Q. Inferring clonal evolution of tumors from single nucleotide somatic mutations. *BMC Bioinform.* **15**, 35 (2014).
66. Jiménez-Sánchez, A., Cast, O. & Miller, M. L. Comprehensive benchmarking and integration of tumour microenvironment cell estimation methods. *Cancer Res.* **3560**, 2018 (2019).

Acknowledgements

This work was supported by Cancer Research UK and the Wellcome Trust. The MelResist study is supported by Addenbrooke's Charitable Trust. D.C.W. is supported by the Li Ka Shing foundation and the NIHR Oxford Biomedical Research Centre. The Affymetrix transcriptome sequencing was funded by the Addenbrooke's Charitable Trust pump priming grant ID SWAI/138. We would like to thank the MelResist team for facilitating this research and providing the ethical framework for post-mortem donations, including Emily Barker, Doreen Milne, Catherine Wilson and Myfanwy Nicholas. The Cambridge Brain Bank for facilitating all aspects of the autopsy, particularly Jenny Wilson. The Tissue Bank team at Addenbrooke's Hospital for helping to extract, store and deliver the samples including Martin Bromwich, Emily Daniels, Elizabeth Cromwell and Beverley Haynes. We thank Yvette Hooks for sectioning the primary tumour and digitally scanning the H&E images. The Cancer, Ageing and Somatic Mutation Programme team at the Sanger Institute for facilitating QC and DNA sequencing, particularly Claire Hardy, Stephen Gamble and Elizabeth Anderson. Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. Sam Behjati for his help in searching for balanced inversions (signature of ionising radiation). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. Finally, we would like to sincerely thank the patients involved in this study and particularly the index patient and their family. Their complete dedication to advancing melanoma research despite an unpredictable and aggressive disease course dominated every clinical encounter. Their altruism and dedication to help has inspired every aspect of this work and lays the foundation for further studies of this kind.

Author contributions

R.R. collected the clinical and molecular data, analysed the data and wrote the paper; N.A.P. analysed the DNA sequencing data, undertook the phylogenetic analyses and co-wrote the paper; O.C. helped analyse the Affymetrix expression data and performed the Consensus^{TME} clustering. D.L., F.S. and F.A.G. analysed the imaging data. D.L. also helped format the immune cell IHC images in Fig. 5. M.T. performed the histopathological analyses for the WES multi-site metastases patients. L.M. performed the laser-capture microdissection for the index patients' primary tumour. I.F. helped select H&E images for Fig. 1. K.W. analysed the copy number profiles for WES multi-site cases with sequenza. J.M.M.G. and M.L. performed the immune cell IHC on the autopsy index patient's samples. L.R. ran the mutational signature analyses. M.M. and A.J.S. provided supervision on the expression analyses and Consensus^{TME} clustering. K.A. performed the autopsy and histopathological assessments of the index patients' samples. L.K. helped review clinical aspects of the manuscript. P.C., C.P. and S.J.W. provided overall clinical supervision including in study set up, patient recruitment, as well as critical review of the manuscript. D.C.W. provided overall study supervision, particularly relating to the phylogenetic analyses and contributed to all sections of the manuscript. D.J.A. provided overall study supervision and critical review of the manuscript. All authors approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-18060-0.

Correspondence and requests for materials should be addressed to D.C.W. or D.J.A.

Peer review information *Nature Communications* thanks David Posada, Alan Hunter Shain and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

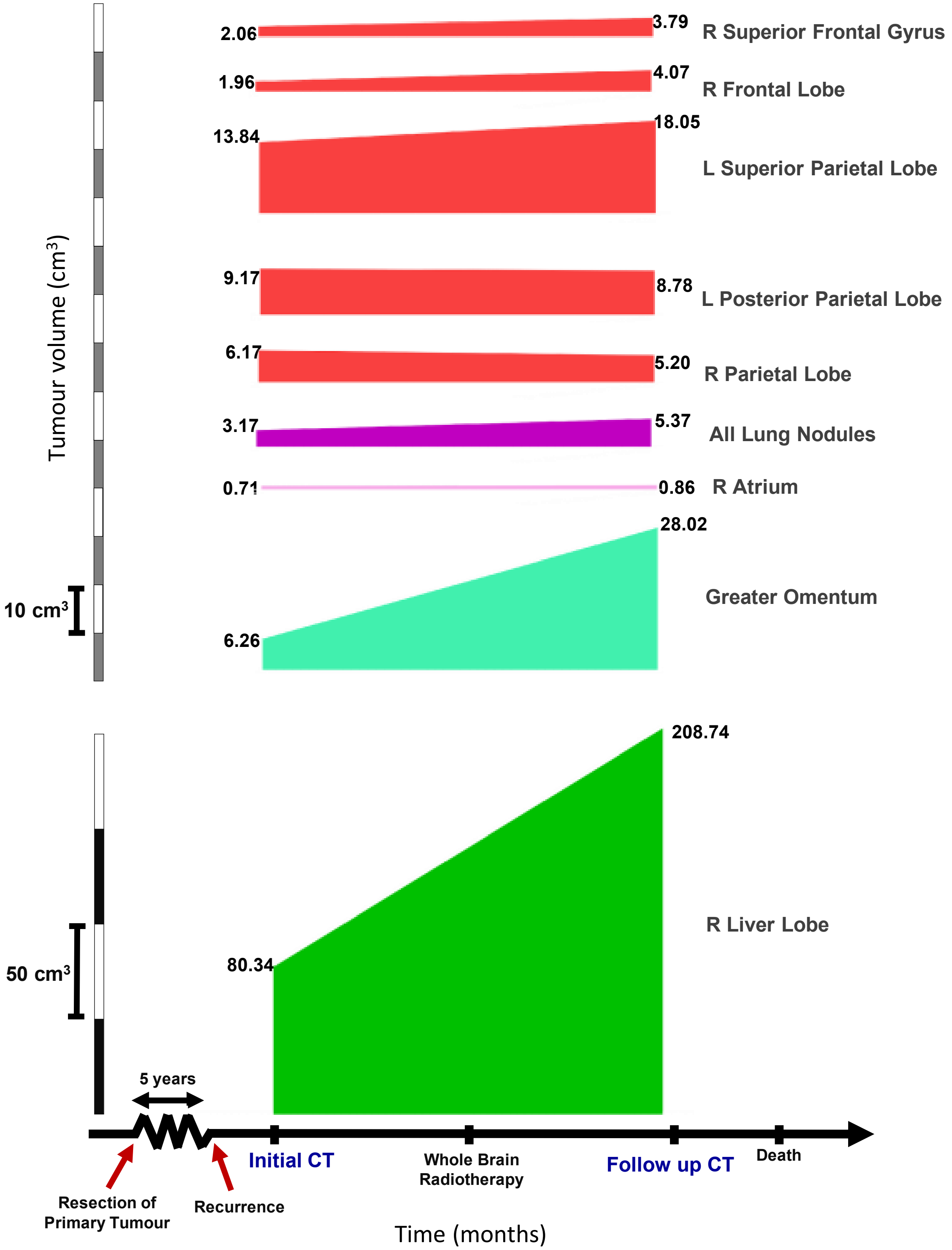
© The Author(s) 2020

Multi-site clonality analysis uncovers pervasive heterogeneity across melanoma metastases.

Rabbie *et al.* 2020

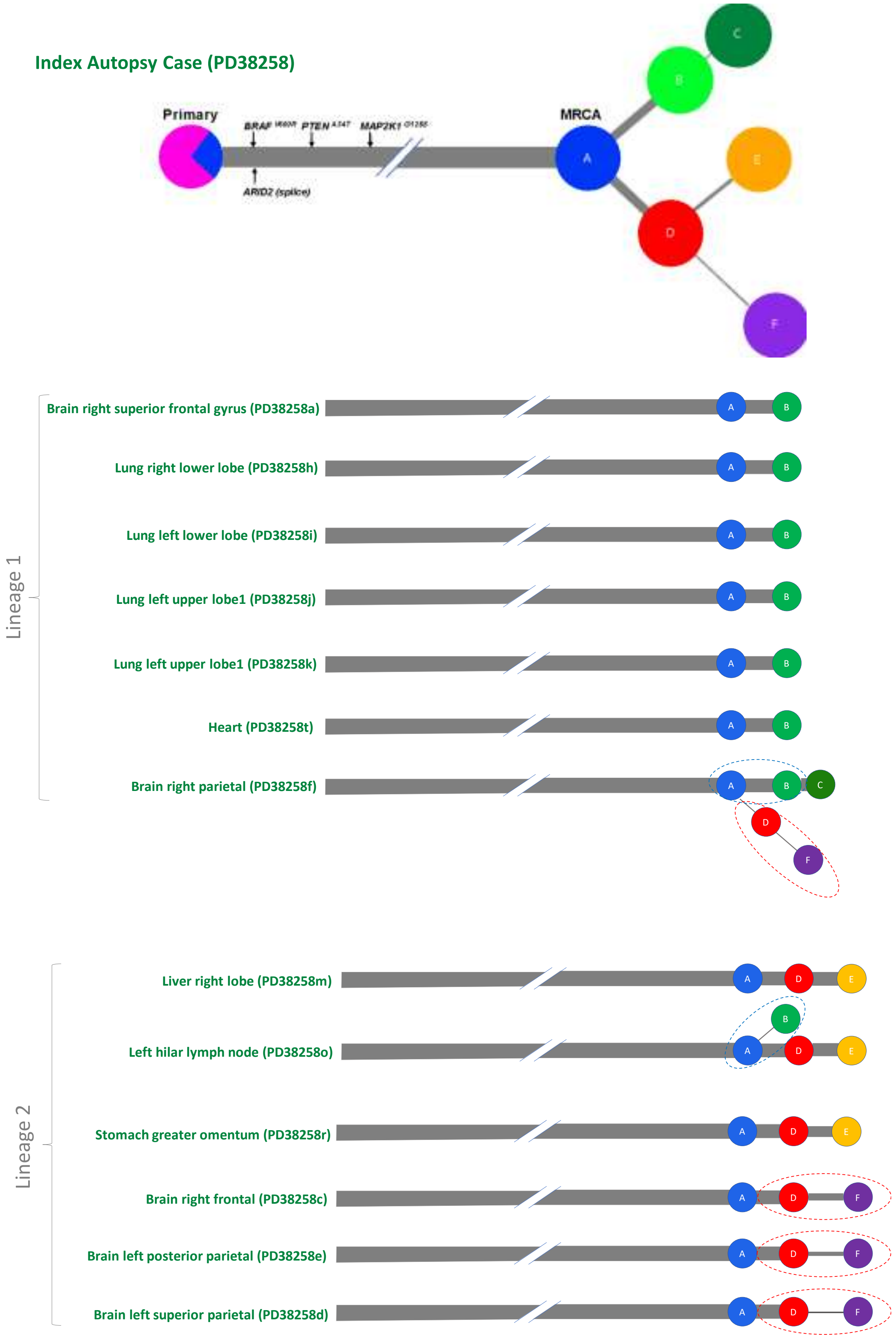
Supplementary Information

Supplementary Figure 1



Supplementary Fig 1. Representative clinical timeline of the index autopsy case, demonstrating rapid progression from the first appearance of metastatic disease. The volume changes of target lesions between interval CTs performed 5 weeks apart are shown, with the follow-up imaging taken two weeks after the completion of whole-brain radiotherapy. This showed only minimal intracranial, but extensive extracranial disease progression.

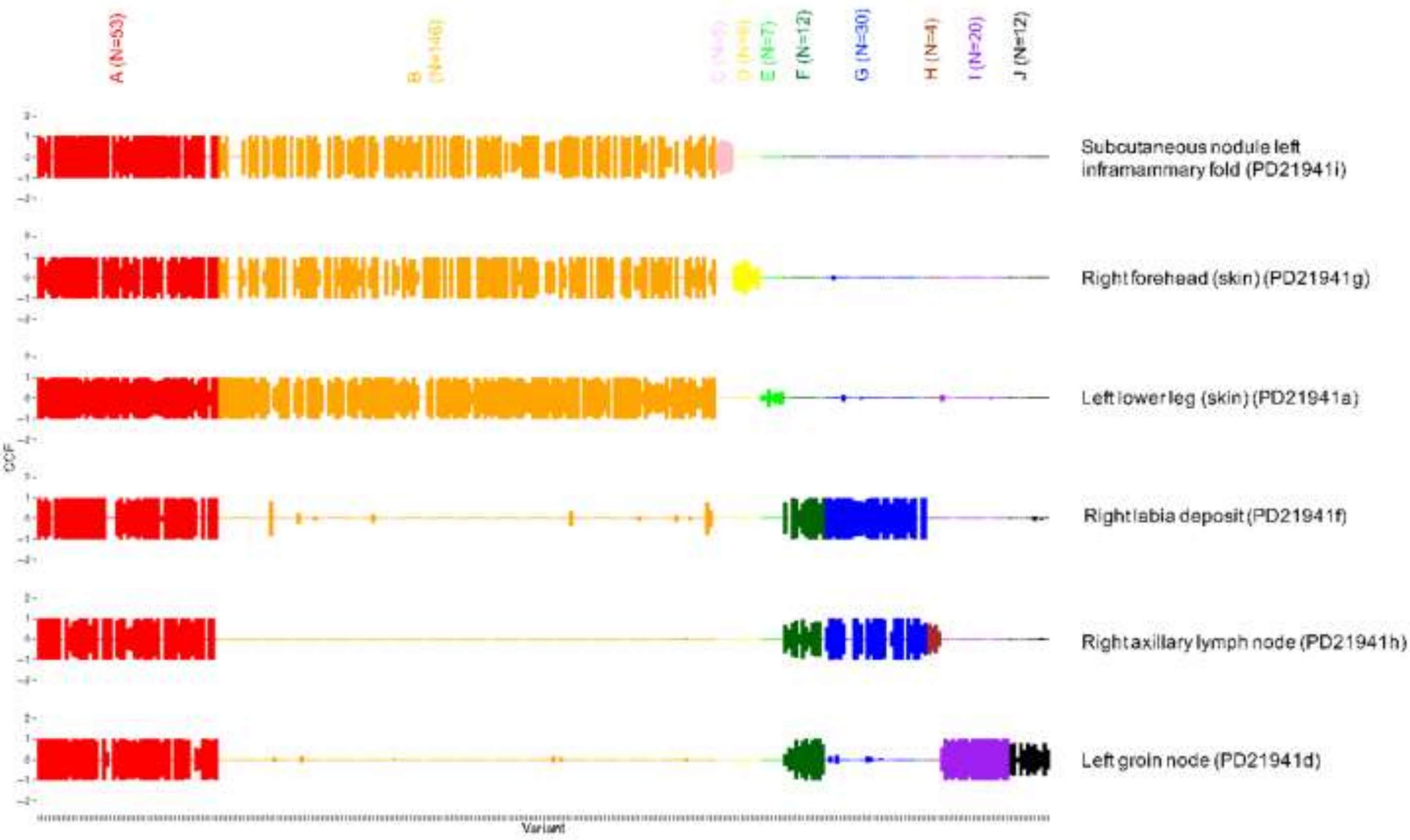
Supplementary Figure 2



Supplementary Fig 2. Sample-level phylogenetic tree for the index autopsy case. Each tree represents a subtree of the overall phylogenetic tree (**Fig. 2D**) including just those subclones seen within that particular sample. However in doing this we were able to segregate the samples based on their respective clonal lineages. We observed two clear lineages, representing distinct waves of metastatic seeding depicted here as the lineage 1 and 2 emanating from clusters B (light green) and clusters D (red) respectively. Dotted ovals represent evidence for polyclonal seeding. Subclones within each oval are found with differing CCFs in 2 or more samples.

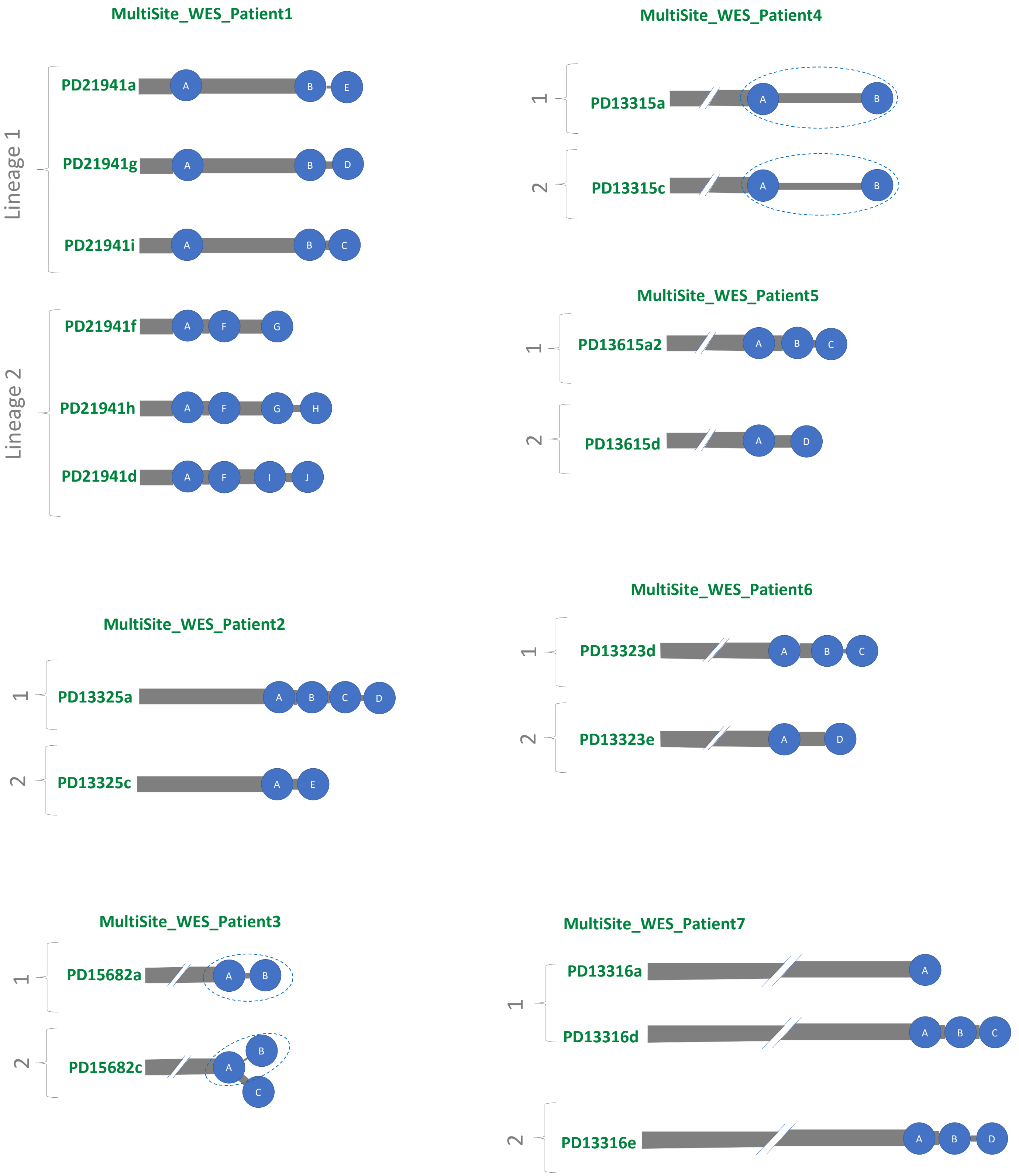
Supplementary Figure 3

MultiSite_WES_Patient1



Supplementary Fig 3. CCF distribution plot for whole-exome sequenced patient MultiSite_WES_Patient1. Rows reflect samples and columns reflect alphabetically and colour-coded mutation clusters (number of SNVs within each cluster is indicated at the top). This shows that clusters B (yellow) and F (green) were clonal in mutually exclusive samples and represent mutually exclusive clonal phylogenies at the first bifurcation of the phylogenetic tree.

Supplementary Figure 4



MultiSite_WES_Patient1

MultiSite_WES_Patient4

MultiSite_WES_Patient5

MultiSite_WES_Patient6

MultiSite_WES_Patient2

MultiSite_WES_Patient3

MultiSite_WES_Patient7

Supplementary Fig 4. Sample-level phylogenetic tree for multi-site whole-exome sequenced cases. The respective branched lineages are depicted for each patient. Dotted ovals represent evidence for polyclonal seeding. Subclones within each oval are found with differing CCFs in 2 or more samples. Only two patients (MultiSite_WES_Patient3 and MultiSite_WES_Patient4) displayed polyclonal seeding.

Supplementary Figure 5


Index Autopsy Case (PD38258)

Brain right superior frontal gyrus (PD38258a)

Lung right lower lobe (PD38258h)

Lung left lower lobe (PD38258i)

Lung left upper lobe1 (PD38258j)

Lung left upper lobe1 (PD38258k)

Heart (PD38258t)

Brain right parietal (PD38258f)

Brain right frontal (PD38258c)

Brain left posterior parietal (PD38258e)

Brain left superior parietal (PD38258d)

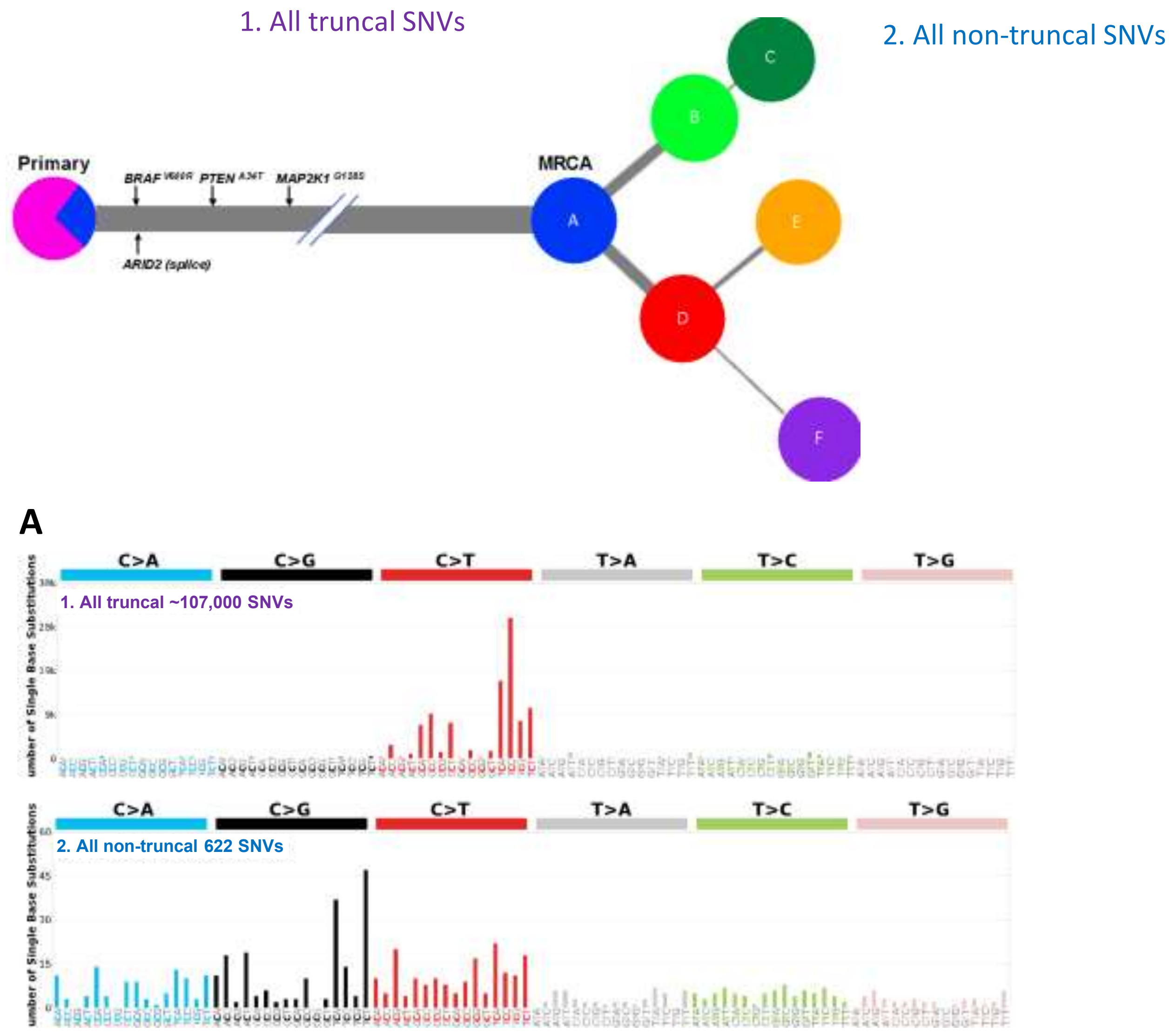
Liver right lobe (PD38258m)

Left hilar lymph node (PD38258o)

Stomach greater omentum (PD38258r)

Supplementary Fig 5. Mutational signatures for all SNVs from the index autopsy case. Shows the mutational profile using the conventional 96 mutation type classification as described by Alexandrov and colleagues[31,49]. This classification is based on the six substitution subtypes: C>A, C>G, C>T, T>A, T>C, and T>G. Further, each of the substitutions is examined by incorporating information on the bases immediately 5' and 3' to each mutated base generating 96 possible mutation types. Here we show the signature profiles including all SNVs from all 13 WGS metastases which as expected, were dominated by signature 7.

Supplementary Figure 6

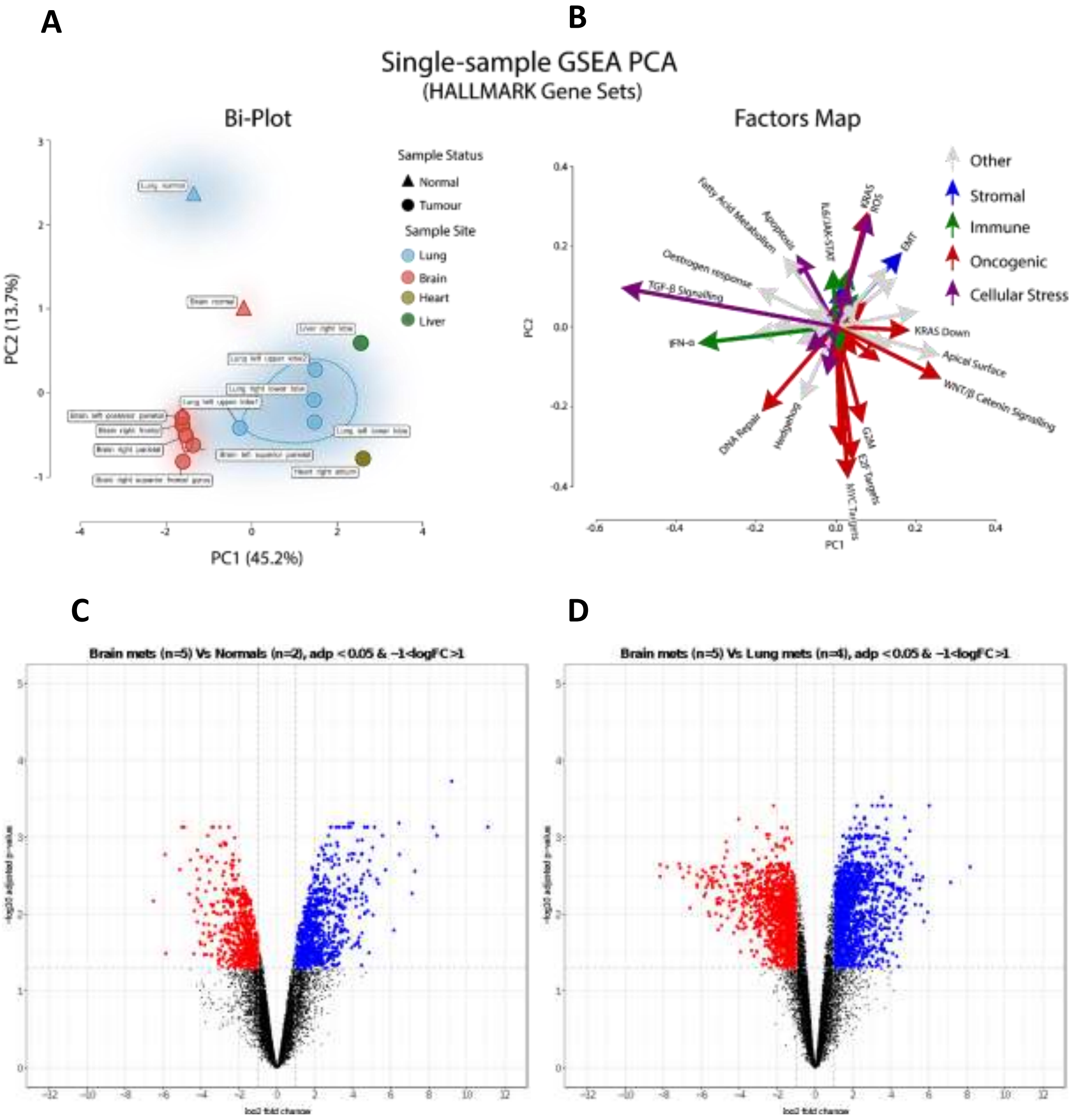


B

Cluster Groups	Global NMF Signatures	Similarity
1. All truncal SNVs	Signature Subs-01 (0.90%) Signature Subs-05 (9.01%) Signature Subs-07 (90.09%)	1.00
2. All non-truncal SNVs	Signature Subs-01 (9.65%) Signature Subs-02 (3.22%) Signature Subs-05 (5.95%) Signature Subs-13 (16.08%) Signature Subs-40 (65.11%)	0.95

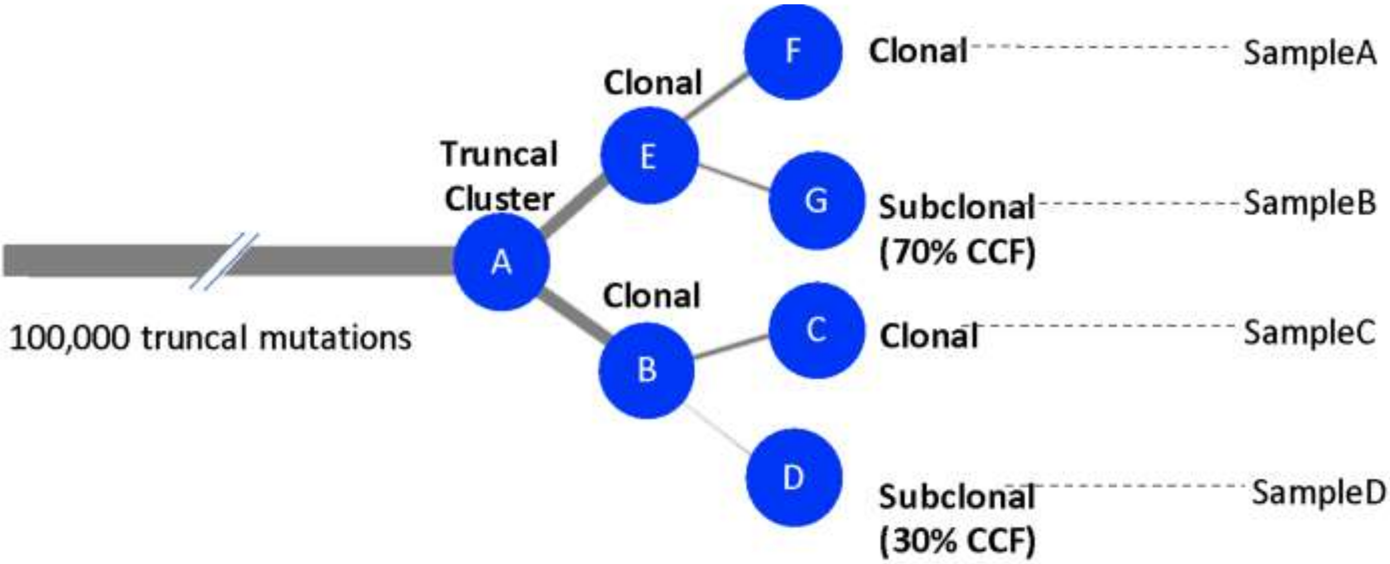
Supplementary Fig 6. Mutational signatures from the index autopsy case. A) Here we show the mutational profiles including; all truncal SNVs in the tree (n~107,000 SNVs) and all non-truncal SNVs (from the six mutation clusters B-F) in the tree (n=622 SNVs). B) The Global NMF signatures shown represent the 'best fit' signatures across all SNVs and the individual percentages for each signature is the proportion which that signature represents. The cosine similarity reports how closely these signatures together mirror the context of all SNVs within that cluster group. As expected, the mutational signature for all truncal SNVs is dominated by signature 7 (90% of SNVs are represented by this signature) whilst this is entirely absent from the non-truncal clones, which are represented by the APOBEC mutational signatures (2 & 13). The non-highlighted signatures (1, 5 and 40) shown in black represent relatively featureless ("flat") oncogenic signatures found in most cancer types and do-not as yet define any distinguishing biological processes[49].

Supplementary Figure 7



Supplementary Fig 7. Differential expression and principal component analyses from the index autopsy case. A) Principal component analysis (PCA) of single sample gene-set enrichment (ssGSEA) using the hallmark gene sets (see **Methods**)[51]. A similar pattern of regional separation (represented in **Fig. 4A**) was observed between the brain and lung metastases, which again are separated from the corresponding patient-matched normal organ control samples. Samples are circled using a kernel density estimation. B) Principal component feature loadings (magnitude and direction) of (A) are shown in the variables factor map. Each co-ordinate in (B) reflects the correlation coefficient of the biological process to principal components 1 (x-axis) and 2 (y-axis) from (A). Vectors are coloured according to the major biological classification of Hallmark gene sets. This revealed that PC2 (on the y-axis), explaining the variation between the tumour and normal samples (represented by circles and triangles in (A) respectively), is primarily represented by the up-regulation of oncogenic processes highlighted with a red arrow pointing (downwards) towards the tumour samples. C) Volcano plot of the genes differentially expressed between the brain metastases (n=5) versus the patient-matched normal tissue (one sample from brain and lung respectively, n=2). Each dot represents one gene, dots above the dotted line are considered statistically significant (FDR-adjusted p-value <0.005 calculated in limma, see methods) and are shaded according to fold-change cut-offs (log fold-change < -1 coloured in red, and log fold-change >1 in blue). D) Volcano plot of the genes differentially expressed between the brain (n=5) versus lung metastases (n=4). Dots represent genes, coloured in the same format as (C). FDR-corrected p-values are calculated in limma, see methods.

Supplementary Figure 8



Supplementary Fig 8. Simulations of phylogenetic tree reconstructions from the index autopsy case with variable subclonal heterogeneity.

Supplementary Table 1. Summary of key clinical and sample details relating to the index autopsy case. The raw sequencing files and variant/CNV calls are all deposited in data availability.

Patient	MelResist_PatientID	Sex	Melanoma subtype	Date of initial melanoma diagnosis	TSTAGE (at initial melanoma diagnosis)	NSTAGE (at initial melanoma diagnosis)	MSTAGE (at initial melanoma diagnosis)	Date of trial registration	Number of lines of therapy	Treatment name	Treatment start date	Treatment end date	WGS_SampleID	AffySampleID	Tumour/Normal	Sample_resection_date	Sample_type	Site	Depth	Purity_Estimate
Index_autopsy_case	01_123	Male	Cutaneous	07/2012	T2a	N0	M0	10/2017	1	Whole-brain RT	07/2017	'07/2017	PD38258a	PR38258a	Tumour	09/2017	Brain	Brain_right_superior_frontal_gyrus	33.5134	0.94
Index_autopsy_case	01_123	Male	Cutaneous										PD38258b	NA	Normal	NA	Buffy coat	Germline_buffy_coat	37.9218	NA
Index_autopsy_case	01_123	Male	Cutaneous										PD38258c	PR38258c	Tumour	09/2017	Brain	Brain_right_frontal	36.21725	0.728
Index_autopsy_case	01_123	Male	Cutaneous										PD38258d	PR38258d	Tumour	09/2017	Brain	Brain_left_superior_parietal	37.81692	0.856
Index_autopsy_case	01_123	Male	Cutaneous										PD38258e	PR38258e	Tumour	09/2017	Brain	Brain_left_posterior_parietal	36.1521	0.821
Index_autopsy_case	01_123	Male	Cutaneous										PD38258f	PR38258f	Tumour	09/2017	Brain	Brain_right_parietal	41.0643	0.857
Index_autopsy_case	01_123	Male	Cutaneous										PD38258h	PR38258h	Tumour	09/2017	Lung	Lung_right_lower_lobe	43.87209	0.511
Index_autopsy_case	01_123	Male	Cutaneous										PD38258i	PR38258i	Tumour	09/2017	Lung	Lung_left_lower_lobe	41.63439	0.89
Index_autopsy_case	01_123	Male	Cutaneous										PD38258j	PR38258j	Tumour	09/2017	Lung	Lung_left_upper_lobe1	33.03544	0.883
Index_autopsy_case	01_123	Male	Cutaneous										PD38258k	PR38258k	Tumour	09/2017	Lung	Lung_left_upper_lobe2	39.93535	0.709
Index_autopsy_case	01_123	Male	Cutaneous										PD38258m	PR38258m	Tumour	09/2017	Liver	Liver_right_lobe	36.84432	0.884
Index_autopsy_case	01_123	Male	Cutaneous										PD38258o	NA	Tumour	09/2017	L hilar lymph node	LN_left_hilar	39.42053	0.422
Index_autopsy_case	01_123	Male	Cutaneous										PD38258r	NA	Tumour	09/2017	Stomach	Stomach_greater_omentum	38.36801	0.586
Index_autopsy_case	01_123	Male	Cutaneous										PD38258t	PR38258t	Tumour	09/2017	Heart	Heart_right_atrium	40.30025	0.823
Index_autopsy_case	01_123	Male	Cutaneous										PD38258u	NA	Tumour	07/2012	Primary	Primary_anterior_chest_wall1	NA	NA
Index_autopsy_case	01_123	Male	Cutaneous										PD38258v	NA	Tumour	07/2012	Primary	Primary_anterior_chest_wall2	NA	NA

Supplementary Table 2. Summary of key clinical and sample details relating to the multisite whole-exome sequenced cases. The raw sequencing files and variant/CNV calls are all deposited in data availability.

Patient	MelResist_PatientID	Sex	Melanoma subtype	Date of initial melanoma diagnosis	TSTAGE (at initial melanoma diagnosis)	NSTAGE (at initial melanoma diagnosis)	MSTAGE (at initial melanoma diagnosis)	Date of trial registration	Number of lines of therapy	Treatment name	Treatment start date	Treatment end date	WES_SampleID	Tumour/Normal	Sample_resection_date	Sample_type	Site	Target_Region_Coverage_AfterPCRDupRemoval
MultiSite_WES_Patient1	MR01_034	Female	Acral	12/2003	**T4b**	N0	M0	03/2014	2	Roche/RAF/MEK Inhibitor	07/2010	09/2011	PD21941i	Tumour	06/2015	Distant skin/subcutanoues	Subcutaneous nodule left inframammary fold	34.5
MultiSite_WES_Patient1	MR01_034	Female	Acral							Ipilimumab	10/2011	12/2011	PD21941f	Tumour	04/2015	Distant skin/subcutanoues	Right labia deposit	37.4
MultiSite_WES_Patient1	MR01_034	Female	Acral										PD21941d	Tumour	09/2014	Distant lymph node	Left groin node	39.3
MultiSite_WES_Patient1	MR01_034	Female	Acral										PD21941e	Tumour	04/2015	Distant skin/subcutanoues	Right shoulder deposit	53.7
MultiSite_WES_Patient1	MR01_034	Female	Acral										PD21941h	Tumour	04/2015	Distant lymph node	Right axilla	61.8
MultiSite_WES_Patient1	MR01_034	Female	Acral										PD21941a	Tumour	03/2014	Distant skin/subcutanoues	Left Lower limb	67.8
MultiSite_WES_Patient1	MR01_034	Female	Acral										PD31210b	Normal	NA	Normal	Normal	107.5
MultiSite_WES_Patient2	MR01_014	Male	Cutaenous	07/2012	T1b			11/2014	0	None			PD13325b	Normal	11/2012	Normal	Normal	63.2
MultiSite_WES_Patient2	MR01_014	Male	Cutaenous										PD13325a	Tumour	11/2012	Primary	Right forearm	54
MultiSite_WES_Patient2	MR01_014	Male	Cutaenous										PD13325c	Tumour	08/2014	Distant skin/subcutanoues	Right head	94.7
MultiSite_WES_Patient3	MR01_020	Male	Cutaenous	07/2012	T4b	N0	M0	03/2013	3	Vemurafenib	03/2013	05/2013	PD15682b	Normal	NA	Normal	Normal	74.2
MultiSite_WES_Patient3	MR01_020	Male	Cutaenous							Ipilimumab	05/2013	'08/2013	PD15682c	Tumour	05/2013	Regional lymph node	Axillary lymph node	64.8
MultiSite_WES_Patient3	MR01_020	Male	Cutaenous							Dacarbazine, Cisplatin, Vinblastine	11/2008	12/2009	PD15682a	Tumour	03/2013	Regional lymph node	Left axilla lymph node	68.7
MultiSite_WES_Patient4	MR01_003	Male	Cutaenous	03/2004	T4a	N1b	M1a	02/2012	4	Dacarbazine	04/2004	07/2004	PD13315b	Normal	NA	Normal	Normal	71.9
MultiSite_WES_Patient4	MR01_003	Male	Cutaenous							Ipilimumab	09/2011	11/2011	PD13315c	Tumour	02/2012	Distant skin/subcutanoues	Left back	42.2
MultiSite_WES_Patient4	MR01_003	Male	Cutaenous							Vemurafenib	02/2012	11/2013	PD13315a	Tumour	02/2012	Distant skin/subcutanoues	Right chest	44.5
MultiSite_WES_Patient4	MR01_003	Male	Cutaenous							Nivolumab	12/2013	05/2016						
MultiSite_WES_Patient5	MR01_010	Male	Cutaenous	10/2004	T2b	N0	M0	07/2012	3	Post-op RT	04/2010	04/2010	PD13615a2	Tumour	07/2012	Regional skin/subcutaneous	Skin unspecified	33.3
MultiSite_WES_Patient5	MR01_010	Male	Cutaenous							Vemurafenib	07/2012	01/2013	PD13615b	Normal	NA	Normal	Normal	35.1
MultiSite_WES_Patient5	MR01_010	Male	Cutaenous							Ipilimumab	01/2013	02/2013	PD13615d	Tumour	01/2013	Regional skin/subcutaneous	Left lower back	48.9
MultiSite_WES_Patient6	MR01_012	Male	Cutaenous	01/2010	TX	N0		11/2012	1	Dabrafenib + trametinib	10/2012	06/2013	PD13323d	Tumour	10/2012	Distant lymph node	Left groin node	46
MultiSite_WES_Patient6	MR01_012	Male	Cutaenous										PD13323b	Normal	NA	Normal	Normal	79.8
MultiSite_WES_Patient6	MR01_012	Male	Cutaenous										PD13323e	Tumour	04/2013	Distant lymph node	Left inguinal node	51.3
MultiSite_WES_Patient7	MR01_004	Male	Cutaenous	10/2010	T3b	N3	M0	04/2012	5	RT (40Gy/20#)	05/2011	06/2011	PD13316b	Normal	NA	Normal	Normal	74.9
MultiSite_WES_Patient7	MR01_004	Male	Cutaenous							Vemurafenib	04/2012	01/2013	PD13316a	Tumour	04/2012	Distant skin/subcutanoues	Left posterior ear	42.4
MultiSite_WES_Patient7	MR01_004	Male	Cutaenous							Ipilimumab	01/2013	03/2013	PD13316d	Tumour	05/2013	Distant skin/subcutanoues	Left supraclavicular fossa	52.2
MultiSite_WES_Patient7	MR01_004	Male	Cutaenous							Pazopanib	07/2013	09/2013	PD13316e	Tumour	11/2013	Regional skin/subcutaneous	Left scalp	63.6
MultiSite_WES_Patient7	MR01_004	Male	Cutaenous							Dabrafenib + trametinib	11/2013	02/2014						

Supplementary tables key

Header	**Code**	**Explanation**
Patient	Patient	Patient identifier as labelled in this study
MelResist_PatientID	MelResist_PatientID	Patient code within the clinical trial (Melresist)
Sex	Male	Male
Sex	Female	Female
Melanoma subtype	Cutaneous	Cutaneous/acral primary melanoma subtype
Melanoma subtype	Acral	
Date of initial melanoma diagnosis	Date of initial melanoma diagnosis	Date of initial melanoma diagnosis
TSTAGE (at initial melanoma diagnosis)	TX	Primary tumour cannot be assessed
TSTAGE (at initial melanoma diagnosis)	T0	No evidence of primary tumour
TSTAGE (at initial melanoma diagnosis)	Tis	Melanoma in situ
TSTAGE (at initial melanoma diagnosis)	T1a	Melanomas 1.0mm or less without ulceration & mitosis
TSTAGE (at initial melanoma diagnosis)	T1b	Melanomas 1.0mm or less with ulceration & mitosis
TSTAGE (at initial melanoma diagnosis)	T2a	Melanomas 1.01 - 2.0mm without ulceration
TSTAGE (at initial melanoma diagnosis)	T2b	Melanomas 1.01 - 2.0mm with ulceration
TSTAGE (at initial melanoma diagnosis)	T3a	Melanomas 2.01 - 4.0mm without ulceration
TSTAGE (at initial melanoma diagnosis)	T3b	Melanomas 2.01 - 4.0mm with ulceration
TSTAGE (at initial melanoma diagnosis)	T4a	Melanomas more than 4.0mm without ulceration
TSTAGE (at initial melanoma diagnosis)	T4b	Melanomas more than 4.0mm with ulceration
NSTAGE (at initial melanoma diagnosis)	NX	Patients in whom the regional lymph nodes cannot be assessed
NSTAGE (at initial melanoma diagnosis)	N0	No regional lymph node metastasis detected
NSTAGE (at initial melanoma diagnosis)	N1a	Melanoma cells in one lymph node with micrometastasis
NSTAGE (at initial melanoma diagnosis)	N1b	Melanoma cells in one lymph node with macrometastasis
NSTAGE (at initial melanoma diagnosis)	N2a	Melanoma cells in 2 or 3 lymph nodes with micrometastasis
NSTAGE (at initial melanoma diagnosis)	N2b	Melanoma cells in 2 or 3 lymph nodes with macrometastasis
NSTAGE (at initial melanoma diagnosis)	N2c	Melanoma cells in 2 or 3 lymph nodes with intransit met(s)/ satellite(s) without metastatic nodes
NSTAGE (at initial melanoma diagnosis)	N3	Four or more metastatic lymph nodes, or intransit met(s)/ satellite(s) with metastatic nodes
MSTAGE (at initial melanoma diagnosis)	M0	No detectable evidence of distant metastasis
MSTAGE (at initial melanoma diagnosis)	M1a	Metastasis to skin, subcutaneous, or distant lymph nodes with normal LDH
MSTAGE (at initial melanoma diagnosis)	M1b	Metastasis to lung with normal LDH
MSTAGE (at initial melanoma diagnosis)	M1c	Metastasis to all other visceral sites with normal LDH or Any distant metastasis to any site with elevated LDH
Date of trial registration	Date of trial registration	Date of trial registration
Number of lines of therapy	Number of lines of therapy	Number of lines of therapy (up to data extract 15.11.19)
Treatment name	Treatment name	Treatment name
Treatment start date	Treatment start date	Treatment start date
Treatment end date	Treatment end date	Treatment end date
WGS_SampleID	WGS_SampleID	Sample IDs for whole-genome sequencing data (index autopsy case only, though the two samples from primary tumour were sequenced with custom capture pull-down)
WES_SampleID	WES_SampleID	Sample IDs for whole-exome sequencing data
AffySampleID	AffySampleID	Affymetrix sample IDs (index autopsy case only)
Tumour/Normal	Tumour	Sample is a tumour
Tumour/Normal	Normal	Sample is normal (germline)
Sample_resection_date	Sample_resection_date	Date tumour sample was resected from patient
Sample_type	Sample_type	Type of sample
Site	Site	Anatomical site of sample
Depth	Depth	Average depth across the entire whole genome (index autopsy case only)
Purity_Estimate	Purity_Estimate	Estimate of tumour purity from copy number call (Battenberg, index autopsy case only)
Target_Region_Coverage_AfterPCRDupRemoval	Target_Region_Coverage_AfterPCRDupRemoval	Depth after excluding PCR dups, QC failed reads, supplementary and secondary read alignments

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-21207-2 OPEN

Check for updates

Tumour gene expression signature in primary melanoma predicts long-term outcomes

Manik Garg[1], Dominique-Laurent Couturier[2], Jérémie Nsengimana[3,4], Nuno A. Fonseca[5], Matthew Wongchenko[6], Yibing Yan[6], Martin Lauss[7], Göran B. Jönsson[7], Julia Newton-Bishop[3], Christine Parkinson[8], Mark R. Middleton[9], D. Timothy Bishop[3], Sarah McDonald[10], Nikki Stefanos[10], John Tadross[10], Ismael A. Vergara[11,12], Serigne Lo[11,12,13], Felicity Newell[14], James S. Wilmott[11,12], John F. Thompson[11,12,15], Georgina V. Long[11,12,16], Richard A. Scolyer[11,12,17], Pippa Corrie[8], David J. Adams[18,19], Alvis Brazma[1,19] & Roy Rabbie[8,18,19]✉

Adjuvant systemic therapies are now routinely used following resection of stage III melanoma, however accurate prognostic information is needed to better stratify patients. We use differential expression analyses of primary tumours from 204 RNA-sequenced melanomas within a large adjuvant trial, identifying a 121 metastasis-associated gene signature. This signature strongly associated with progression-free (HR = 1.63, $p = 5.24 \times 10^{-5}$) and overall survival (HR = 1.61, $p = 1.67 \times 10^{-4}$), was validated in 175 regional lymph nodes metastasis as well as two externally ascertained datasets. The machine learning classification models trained using the signature genes performed significantly better in predicting metastases than models trained with clinical covariates ($p_{\mathrm{AUROC}} = 7.03 \times 10^{-4}$), or published prognostic signatures ($p_{\mathrm{AUROC}} < 0.05$). The signature score negatively correlated with measures of immune cell infiltration ($\rho = -0.75$, $p < 2.2 \times 10^{-16}$), with a higher score representing reduced lymphocyte infiltration and a higher 5-year risk of death in stage II melanoma. Our expression signature identifies melanoma patients at higher risk of metastases and warrants further evaluation in adjuvant clinical trials.

---

[1] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridgeshire, UK. [2] Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, UK. [3] University of Leeds School of Medicine, Leeds, United Kingdom. [4] Biostatistics Research Group, Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. [5] CIBIO/InBIO-Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Rua Padre Armando Quintas, 4485-601 Vairão, Portugal. [6] Oncology Biomarker Development, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080, USA. [7] Lund University Cancer Center, Lund University, Lund, Sweden. [8] Cambridge Cancer Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [9] Oxford NIHR Biomedical Research Centre and Department of Oncology, University of Oxford, Oxford, UK. [10] Department of Pathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [11] Melanoma Institute Australia, The University of Sydney, North Sydney, NSW, Australia. [12] Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia. [13] Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia. [14] QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia. [15] Discipline of Surgery, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia. [16] Royal North Shore and Mater Hospitals, Sydney, Australia. [17] Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital and New South Wales Health Pathology, Sydney, NSW, Australia. [18] Experimental Cancer Genetics, The Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK. [19] These authors contributed equally: David J. Adams, Alvis Brazma, Roy Rabbie. ✉email: rr13@sanger.ac.uk

Cutaneous melanoma (CM) accounts for 75% of skin cancer-related deaths, and the incidence has been increasing worldwide[1]. Most patients present with primary tumours and the majority will be cured by local surgery. Outcomes for patients with metastatic melanoma have improved radically over the past 10 years with the introduction of new systemic therapies[2], although median survival of patients has remained at ~3 years. Importantly, of those patients who ultimately die of melanoma a significant proportion originally presented with early-stage disease[3], suggesting that there is a subgroup of these patients who have aggressive tumours. Thus, optimal management of early melanoma is key to improving outcomes.

Patients with resected AJCC stage III melanoma are now eligible for adjuvant immune checkpoint inhibitors, as well as *BRAF*-targeted therapies, based on randomised trials confirming a reduction in the risk of relapse and improved overall survival (OS)[4–7]. Clinical trials are underway to evaluate similar therapies in resected stage IIB/C patients[8], whose outcomes reflect that of untreated stage IIIA/B melanoma[9]. As such, the number of patients eligible for the treatment with adjuvant therapies over the coming years is expected to increase substantially. These modern anti-cancer drugs are high cost and carry a risk of both life-changing and life-threatening toxicities, so there is a growing desire to more accurately predict those patients at high risk of recurrence in whom intervention is expected to be beneficial so-as-to avoid over-treating patients likely to have been cured of their disease by surgery alone.

Gene expression signatures have the potential to improve the prediction of the biological behaviour of melanoma by objectively defining "high risk" on a molecular level[10]. Previous transcriptomic analyses of CM identified patterns of gene expression associated with survival independent of AJCC stage[11]. Building on these data, Gerami et al. first reported a proprietary prognostic gene expression profile (GEP) test utilising a 31-gene panel (28 discriminating and 3 control genes) for use in patients with CM (Decision-Dx Melanoma$^{\mathrm{TM}}$)[12]. The test uses quantitative reverse transcriptase polymerase chain reaction technology to measure the expression of individual genes from formalin-fixed paraffin-embedded (FFPE) primary melanomas to provide a binary classification of low (class 1) or high (class 2) risk for developing metastases within 5 years of diagnosis (with A and B subclasses to further stratify risk)[13]. The signature's performance has since been evaluated in a number of retrospective clinical studies evaluating recurrence-free survival[14,15]. A further recent unsupervised clustering analysis based on 677 primary melanoma transcriptomes (generated using the Illumina DASL array platform) embedded within a population-controlled cohort study from the Leeds Melanoma Cohort (LMC) identified a six-class 150 gene prognostic signature (herein referred to as LMC_150)[16]. The signature uniquely demonstrated prognostic relevance (melanoma-specific survival; MSS) in patients with stage I primary melanoma and further predicted poor outcomes in patients undergoing immunotherapy[16]. Heterogeneity in key aspects of the aforementioned studies (including varying trial design, sample type, sequencing platforms and primary outcome measures) may partially explain the small number of overlapping genes ($n = 4$) between both sets of signatures. Furthermore, owing to a lack of prospective data proving the clinical utility of such prognostic molecular tools[17], there are currently no established prognostic biomarkers able to accurately identify truly high-risk patients.

Using patient samples and long-term clinical outcome data from one of the largest adjuvant melanoma trials[18,19], we undertook RNA sequencing of the primary tumour matched with robust prospective clinical data to uncover a molecular signature that could be used to predict patient outcomes. This was then validated in two externally ascertained datasets.

Results

Prognostic signature generated using covariate-corrected differential expression. The structure of the datasets and analyses are depicted in Supplementary Fig. S1. Principal component analysis (PCA) showed that primary CMs ($n = 204$) and melanoma spread to local lymph nodes (LNs; $n = 175$) clustered separately, suggesting an impact of the microenvironment on tumour gene expression (Supplementary Fig. S2; see "Methods" section "Visualization of inherent distribution of samples"). We therefore decided to treat these as separate datasets, focussing our analyses on the primary melanoma samples followed by a validation of our results in the regional LN metastases from this dataset. We conducted a differential expression analysis, identifying differences in gene expression levels in primary tumours between those patients with and without distant metastasis over a minimum of 6 years follow up, while controlling for a key set of variables that were independently associated with distant metastases, including Stage (AJCC 7th edition[20], herein referred to as "stage"), Breslow thickness, ECOG; Eastern Cooperative Oncology Group Performance Status and the experimental adjuvant therapy (Supplementary Fig. S3a and Supplementary Table S1; see "Methods" sections "Clinical covariate selection" and "Differential expression analysis"). Our analyses revealed 197 significantly differentially expressed genes (DEGs, FDR-adjusted p value <0.1) associated with metastases (Supplementary Figs. S1 and S3b). These DEGs were further filtered to remove pseudogenes ($n = 39$) and those genes not identified within the LMC DASL array ($n = 37$)[16] to enable external validation of our signature (Supplementary Fig. S1). We were therefore left with 121 DEGs, which made up our core prognostic signature herein referred to as "Cam_121" (Supplementary Data 1).

Signature added incremental prognostic value when combined with conventional clinical staging. In order to explore the relationship between Cam_121 gene expression and prognosis, we first performed univariate Cox regression using the weighted Cam_121 expression score (see "Methods" section "Survival analyses") as a predictor. We found that the Cam_121 signature significantly associated with both OS (hazard ratio (HR) = 1.64 (95% CI 1.30–2.07), $p = 3.56 \times 10^{-5}$) and progression-free survival (PFS; HR = 1.63 (95% CI 1.31–2.02), $p = 8.92 \times 10^{-6}$; Fig. 1). In order to evaluate whether the signature score contributed independent prognostic information while controlling for conventional clinical covariates, multivariate Cox regression analyses were performed (Fig. 1c). The signature score was significantly associated with both OS (HR = 1.61 (95% CI 1.26–2.07), $p = 0.000167$) and PFS (HR = 1.63 (95% CI 1.29–2.07), $p = 5.24 \times 10^{-5}$) in multivariate Cox regression models.

In order to avoid the risk of overfitting, we further tested the performance of Cam_121 in an (entirely separate) set of samples from regional LN metastasis embedded within this dataset ($n = 143$). We found that the weighted signature score was also significantly associated with both OS (HR = 1.72 (95% CI 1.37–2.14), $p = 1.53 \times 10^{-6}$) and PFS (HR = 1.75 (95% CI 1.43–2.16), $p = 1.10 \times 10^{-7}$) in multivariate Cox regression models (Fig. 2a–c). Thereby indicating that Cam_121 could also be relevant as a prognostic tool after the resection of stage III (regional LN positive) melanoma.

We further tested the signatures' prognostic power in four externally acquired independent validation datasets from; (i) the LMC ($n = 677$)[16]; (ii) The Cancer Genome Atlas (TCGA-

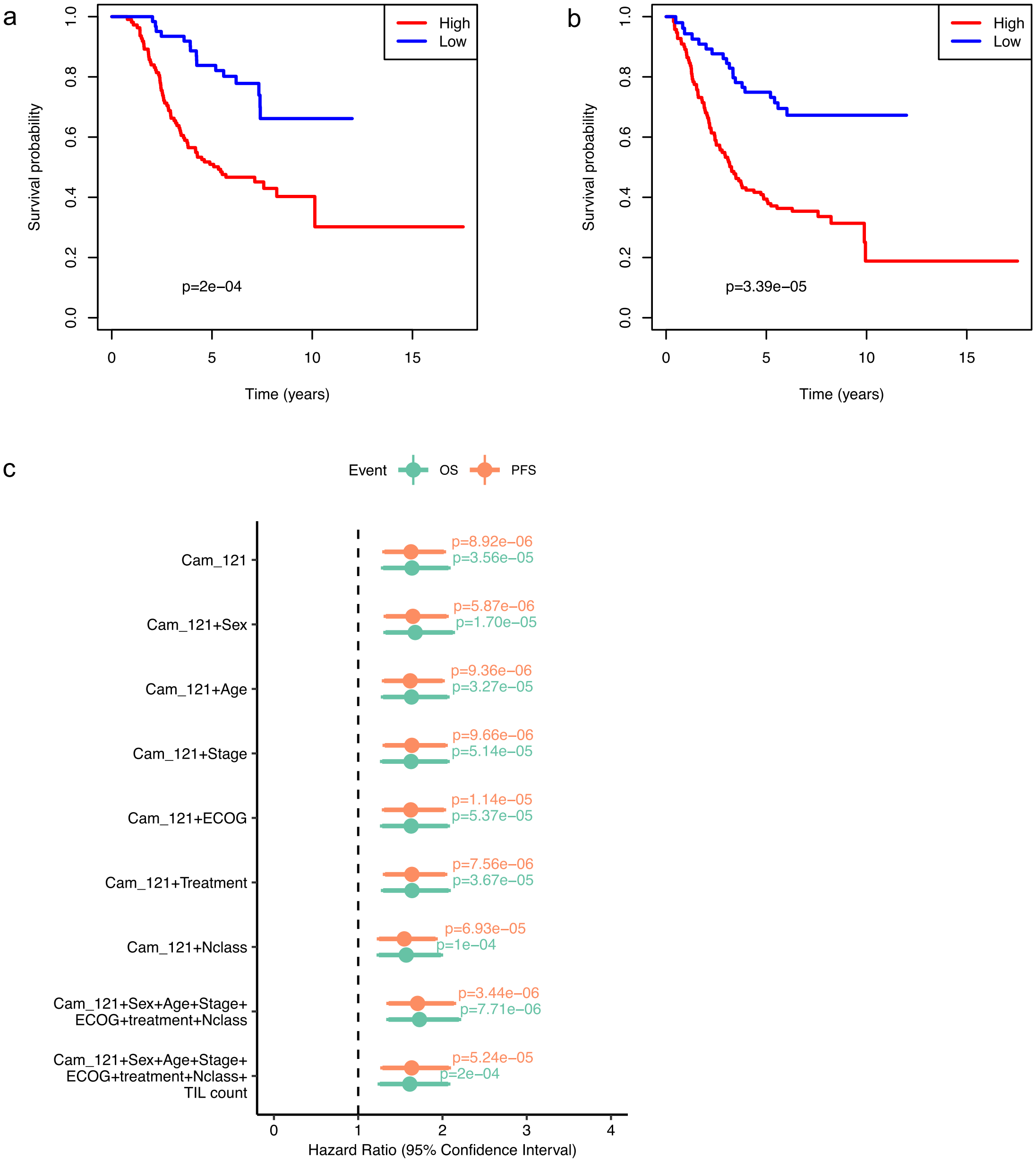


Fig. 1 The Cam_121 gene expression signature is strongly associated with survival in uni- and multivariate Cox regression analyses (AVAST-M primary melanoma cohort; $n = 194$). Kaplan-Meier survival plots comparing the survival probabilities (y-axes) as a function of time in years (x-axes) of groups with high and low "Cam_121" (quantile 0.33 split) for outcomes **a** overall survival (OS) and **b** progression-free survival (PFS). The p value of a two-sided logrank test comparing the survival distributions of both groups are indicated. **c** Forest plot indicating the hazard ratio (HR) estimates relating to the Cam_121 signature when predicting OS (green) and PFS (orange) by means of Cox proportional hazard models controlling for different (sets of) clinical variables (y-axis). The HR estimates are indicated by the dots at the centre of the error bars; the horizontal error bars correspond to the 95% confidence intervals of the HR. The two-sided Wald t test p values corresponding to the signature "Cam_121" parameter are indicated for each model and outcome. ECOG Eastern Cooperative Oncology Group Performance Status. TIL count Tumour-infiltrating lymphocyte count.

SKCM[21]; skin = 159 and LN = 216); (iii) the Lund Primary Melanoma Cohort[22] ($n = 223$) and (iv) the Australian Melanoma Genome Project[23]. Validation within the LMC confirmed that Cam_121 was associated with melanoma-specific survival in both univariate (HR = 1.49 (95% CI 1.27–1.74), $p = 5 \times 10^{-7}$) and multivariate Cox regression models (HR = 1.7 (95% CI), $p =$ 0.001, Fig. 2d, e). Owing to a lack of power when considering true primary melanomas within TCGA-SKCM dataset ($n = 87$), samples from primary tumours were considered together with regional cutaneous relapsed tumours (defined herein as "Skin TCGA-SKCM" ($n = 159$)) and were tested separately from the regional LN samples ($n = 216$). Cam_121 was associated with OS in both univariate (skin: HR = 1.50, 95% CI = (1.15, 1.96), $p =$ 0.00273; LN: HR = 1.28, 95% CI = (1.06, 1.56), $p = 0.0109$) and multivariate survival analyses (skin: HR = 1.59, 95% CI = (1.17, 2.17), $p = 0.00348$; LN: HR = 1.28, 95% CI = (1.03, 1.59), $p =$ 0.0256) in these two external datasets (Fig. 2f–i). A third external validation was attempted using the Lund Primary Melanoma Cohort[22] ($n = 223$); however (owing to the use of less sensitive sequencing technologies within the study), only 24 of the Cam_121 genes were identified within this dataset (Supplementary Fig. S5b). Nonetheless, this 24-gene signature was

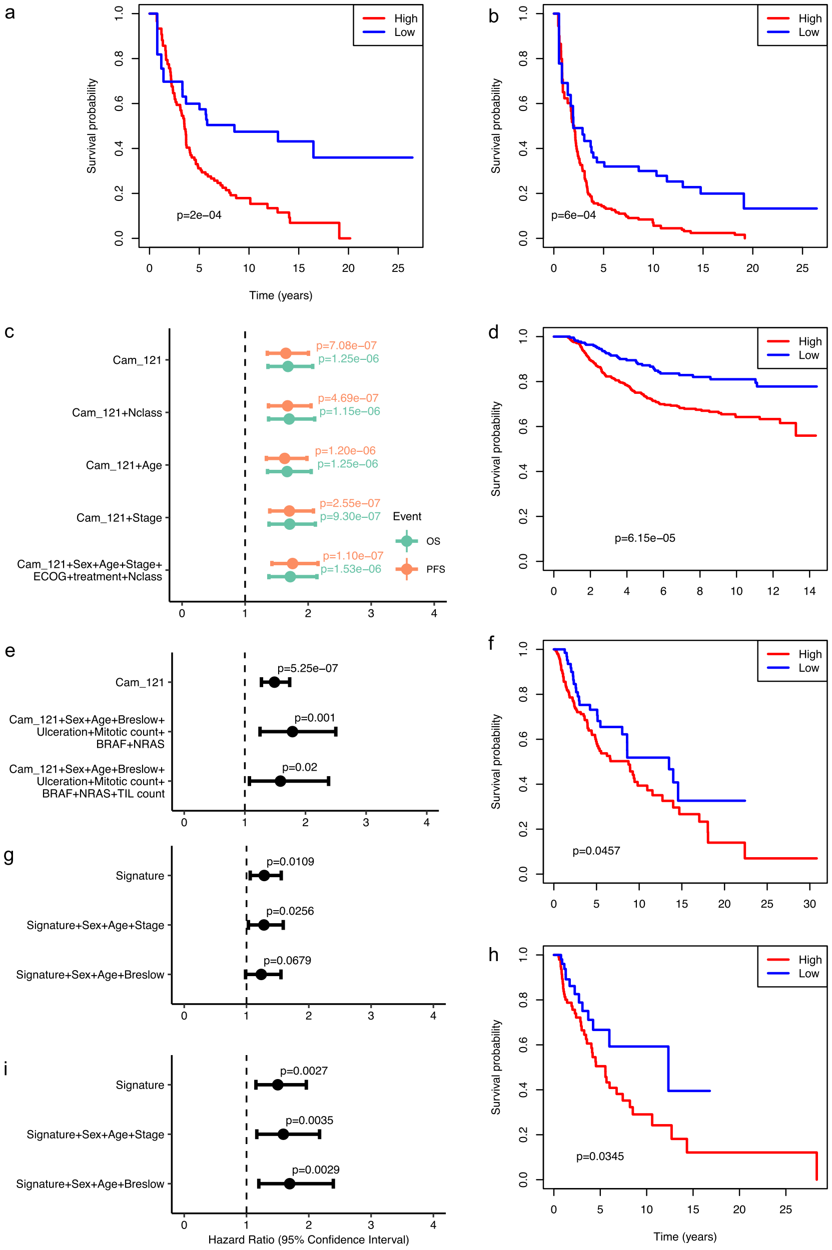


Fig. 2 Validation of the Cam_121 in further datasets. We explored the Cam_121 signature in the entirely separate lymph node (LN) samples from the AVAST-M Cohort (**a**–**c**; *n* = 143), as well as in three independently acquired external datasets including the Leeds Melanoma Cohort (**d**, **e**; *n* = 677), TCGA-SKCM LN (**f**, **g**; *n* = 212) and TCGA-SKCM Skin (**h**, **i**; *n* = 156). AVAST-M LN Cohort: Kaplan-Meier (K-M) survival plots comparing the survival probabilities (*y*-axes) as a function of time in years (*x*-axes) of groups with high and low "Cam_121" (quantile 0.33 split) for outcomes **a** overall survival (OS) and **b** progression-free survival (PFS). The *p* value of a logrank test comparing the survival distributions of both groups is indicated on each curve. **c** Forest plot indicating the hazard ratio (HR) estimates (dots at the centre of the error bars) and corresponding 95% confidence intervals (horizontal error bars) related to the Cam_121 signature when predicting OS (green) and PFS (orange) by means of Cox proportional hazard models when controlling for different (sets of) clinical variables (*y*-axis). The two-sided Wald *t* test *p* values corresponding to the signature "Cam_121" parameter are indicated for each model and outcome. ECOG Eastern Cooperative Oncology Group Performance Status. Leeds Melanoma Cohort: **d** K–M curve comparing the melanoma-specific survival probabilities (*y*-axis) of groups with high and low "Cam_121" (quantile 0.33 split) through time in years (*x*-axis) and *p* value of a two-sided logrank test comparing the survival distributions. **e** Forest plot showing the HR estimates (dots at the centre of the error bars) and 95% confidence intervals (horizontal error bars) of the HR estimates corresponding to the continuous signature "Cam_121" parameter when predicting melanoma-specific survival by means of different Cox proportional hazard models (*y*-axis). Multivariate correction was undertaken for sex, age, Breslow thickness, ulceration, mitotic count, as well as *BRAF* and *NRAS* mutation status and (in the final row), correction was also undertaken for the tumour-infiltrating lymphocyte (TIL) score. TCGA-SKCM Cohort: **f**, **h** K–M curves comparing the overall survival probabilities of groups with high and low "Cam_121" (quantile 0.33 split) and *p* value of a two-sided logrank test comparing the survival distributions in TCGA-SKCM LN and TCGA-SKCM Skin datasets, respectively. TCGA-SKCM Cohort: **g**, **i** Forest plot showing the HR estimates (dots at the centre of the error bars) and 95% confidence intervals (horizontal error bars) corresponding to the of related to the continuous signature "Cam_121" parameter when controlling for different (sets of) clinical variables (*y*-axes) in the TCGA-SKCM LN and TCGA-SKCM Skin datasets, respectively. Multivariate correction was undertaken for sex, age, stage and Breslow thickness. The *p* values of two-sided Wald *t* tests corresponding to the signature "Cam_121" parameter are indicated for each model and outcome (no multiplicity correction used).

significantly correlated with PFS (HR = 1.67 (95% CI 1.06–2.62), $p = 0.03$ univariate Cox regression analyses), but not with OS ($p = 0.32$) (Supplementary Fig. S5a). A final validation was attempted using the Australia Melanoma Genome Project dataset[23], in which only 55 samples from a mixture of tissue sites (including primary tumours, regional LNs, distant metastases, in-transit metastases and others) were available for analysis. Cox regression parameter estimates showed the same trend as observed above when comparing the "high"/"low" risk Cam_121 cohorts in this dataset (Supplementary Fig. S4a), however, significance was not achieved. Our power calculations revealed that the sample size was too small to statistically detect the effect of interest with a high probability (Supplementary Fig. S4b; see power analysis in "Methods" section "Power calculation for the external validation datasets").

The published signature from Gerami et al.[12] (Decision-Dx Melanoma^{TM}; $n = 27$ genes) was not associated with OS in multivariate models in the AVAST-M primary melanoma dataset. The signature from Thakur et al.[16] (LMC_150; $n = 150$ genes) was associated with both OS and PFS, though the wide confidence intervals may in part be reflective of a higher proportion of stage III patients in the AVAST-M dataset (Supplementary Fig. S6).

Cam_121 predicts metastasis better than both clinical covariates and published prognostic signatures. We further sought to determine whether the Cam_121 gene expression signature outperforms key clinical covariates in predicting whether a primary melanoma would ultimately metastasise to distant body sites or not. For this, we developed separate machine learning (ML) classification models using the Cam_121 gene expression values as features, as well as using clinical covariates as features, with the aim of maximising the area under the sensitivity vs (1-specificity) curve (herein referred to as "AUROC" (area under the receiver operating characteristic curve)) as both the metrics are important in this case (see "Methods" section "Machine learning analysis"; Supplementary Fig. S15). Note that these clinical covariates were selected independently based on their level of association with distant metastases (Supplementary Table S1a). We found that models trained with the Cam_121 gene expression signature as features significantly outperformed the models trained with the clinical covariates as features ($p_{\mathrm{AUROC}} = 2.27 \times 10^{-3}$, $p_{\mathrm{sensitivity}} = 1.79 \times 10^{-3}$, $p_{\mathrm{specificity}} = 0.46$; Supplementary Table S2a), and this remained consistent across all ML classifiers (Fig. 3a, Supplementary Figs. S7 and S8, and Supplementary Table S2a). The classifier giving the highest AUROC with the prognostic signature gave better results across all three performance metrics than the classifier giving the highest AUROC with the clinical covariates alone: AUROC (0.67 ± 0.12 with the clinical covariates alone vs 0.83 ± 0.09 with the Cam_121 alone), sensitivity (0.58 ± 0.16 vs 0.75 ± 0.13) and specificity (0.71 ± 0.16 vs 0.73 ± 0.14; Fig. 3a and Supplementary Fig. S8).

In order to reduce the risk of bias that might result from feature selection (of both Cam_121 and clinical covariates) and training/testing from the same dataset, we went on to validate our findings in an entirely independent dataset represented by the regional LN samples from within the AVAST-M dataset ($n = 143$). In this model, the classifiers giving the highest AUROC for each set of features on the training data (AVAST-M primary melanoma data) were selected for further validation within regional LN samples. We found that the classification model developed using the Cam_121 gene expression signature as features (AUROC = 0.67) significantly outperformed the classification model developed using the clinical covariates alone (AUROC = 0.54; DeLong's test p value = 0.02, $z = 2.05$; Fig. 3b and Supplementary Table S2b). In particular, adding the signature to the clinical covariates (Cam_121 + clinical covariates) correctly predicted an additional three overlapping cases that were missed out by the model trained on the signature alone (49 overlapping cases in Fig. 3c vs 46 overlapping cases in Fig. 3d).

In order to test the performance of the published prognostic signatures from Gerami et al.[12] (Decision-Dx Melanoma^{TM}; $n = 27$ genes) and Thakur et al.[16] (LMC_150; $n = 150$ genes) in predicting metastases in an unbiased way, we repeated the training and testing of classification models in the entirely independent AVAST-M LN dataset ($n = 143$) from which no feature selection has been undertaken. We found that Cam_121 significantly outperformed the baseline clinical covariates (Cam_121 vs clinical covariates: $p_{\mathrm{AUROC}} = 7.03 \times 10^{-4}$, $z_{\mathrm{AUROC}} = 4.44$), as well as these two published signatures at the 5% significance level (Cam_121 vs LMC_150: $p_{\mathrm{AUROC}} = 0.02$, $z_{\mathrm{AUROC}} = 2.30$; Cam_121 vs Decision-Dx Melanoma: $p_{\mathrm{AUROC}} = 0.012$, $z_{\mathrm{AUROC}} = 2.45$; Fig. 3e and Supplementary Table S2c, d).

Cam_121 gene signature score performed significantly better than genes selected at random in predicting overall and progression-free survival. In light of reports suggesting that randomly selected genes may perform equally well in predicting prognosis as published signatures[24], we further tested the performance of Cam_121 against a signature of 121 randomly selected genes (from a pool of 19,434 protein-coding genes in our dataset and repeated 1000 times; see "Methods" section "Testing signature performance against randomly selected genes"). Cam_121 significantly outperformed random signatures across all measures of clinical efficacy including: OS ($p = 0.001$) and PFS ($p = 0.001$) in multivariate Cox regression models (Supplementary Fig. S9).

Stage II patients with a "high-risk" signature demonstrated a 33% risk of death at 5 years, a threshold for which adjuvant therapy could be considered. We envisage that one of the central clinical applications of a prognostic GEP might be to identify those patients with stage II melanoma who may be at higher risk of relapse or death and for whom adjuvant systemic therapies may be considered. In order to compare our data with the registration adjuvant melanoma trials[4–7], we measured the absolute risk of death at 5 years (calculated as the proportion of patients who died due to melanoma within 5 years from diagnosis; see "Methods" section "Determination of the weighted expression score cut-off to define "high" and "low" absolute risk of death at 5 years").

Analyses within the LMC cohort (where there was a higher preponderance of early-stage patients) revealed that stage II patients had a 27% (76/279) baseline absolute risk of death at 5 years. This risk rose to 33% (64/192) in those stage II patients with a high-risk-weighted Cam_121 expression score profile and dropped to 14% (12/87) in those stage II patients with a low-risk profile (Table 1). The stratification of high/low-risk cohorts in this context was based on a 0.33 quantile cut off of the weighted Cam_121 expression score and subsequent references to high/low Cam_121 risk groups refer to these weighted expression groups.

Per-gene analyses. In order to determine the relative influence of each gene and each baseline clinical covariate within the ML model, we analysed their feature importance scores and found that no single feature dominated the performance of the model (Supplementary Fig. S10), suggesting that it is the combination of all the features that yielded improved performance over the baseline clinical covariates (Fig. 3a, and Supplementary Figs. S7 and S8). In keeping with this, none of the Cam_121 genes proved

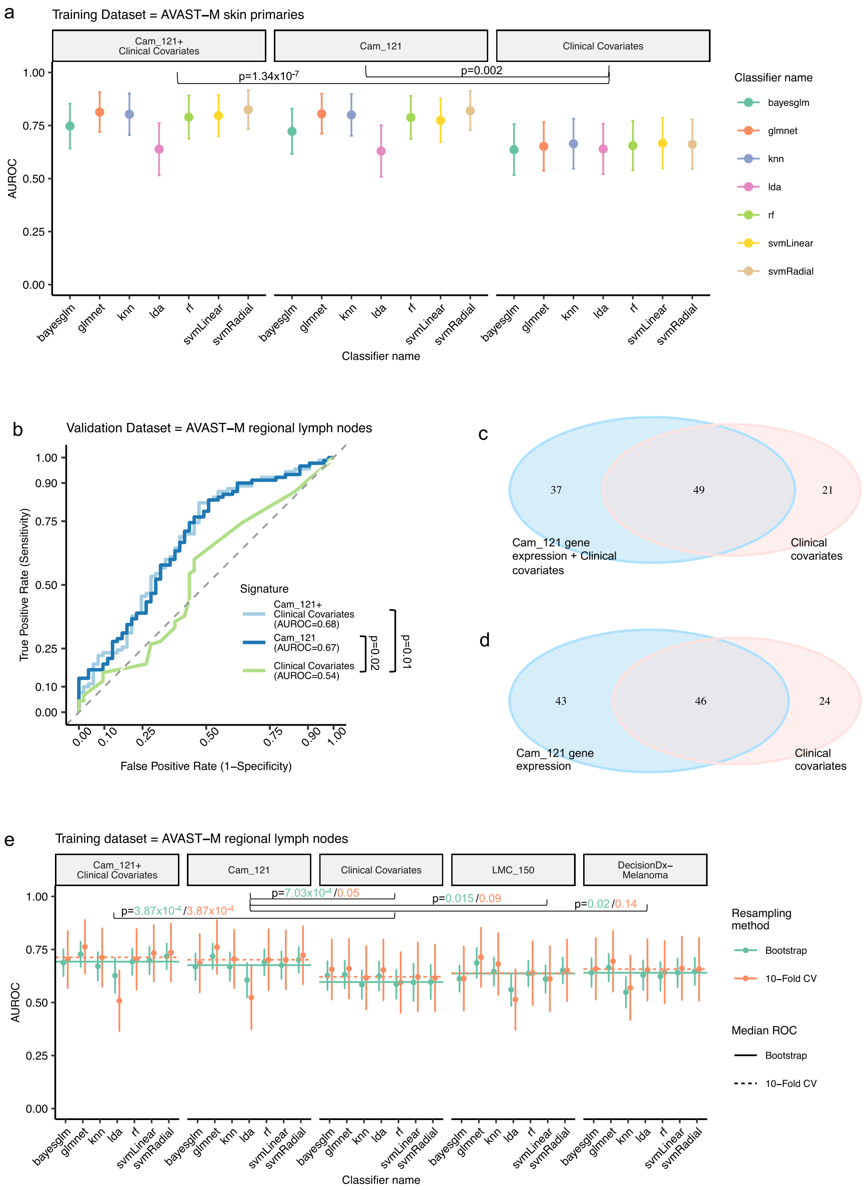


Fig. 3 The Cam_121 gene expression signature is predictive of metastases across multiple machine learning classification models and in an internal validation dataset (*n* = 143). a Plot showing the mean ± standard deviation of area under the ROC curve (AUROC) predicted by different classification models when trained using tenfold cross validation (CV; repeats = 1000) on the AVAST-M primary melanoma dataset (*n* = 194). The features used for training each classification model are indicated on the top grey panel. Within each panel, seven different machine learning classifiers were trained to predict metastases. Statistical comparison using one-sided two-sample Welch *t* tests are indicated (see also Supplementary Table S2a). **b** Area under the ROC curve plots, showing, for the best performing classification model selected in each panel of **a**, its performance on an entirely separate AVAST-M lymph node validation dataset (*n* = 143). The one-sided DeLong's test *p* value is reported for each comparison (see also Supplementary Table S2b). **c**, **d** Venn diagrams comparing the number of correctly predicted relapse outcomes (yes/no) of 143 patients from the models described in **b**. **c** Venn diagram showing the number of correctly predicted relapse outcomes specific to or common between "Cam_121 + clinical covariates" (blue) vs "Clinical covariates" (pink). **d** Venn diagram showing the number of correctly predicted relapse outcomes specific to or common between "Cam_121"(blue) and "clinical covariates" alone (pink). Out of a total of 143 patients, 30 were wrongly predicted by both the models in **c** and 36 were wrongly predicted by both the models in **d**. **e** Plot showing the performance of different classification models in predicting metastases in terms of AUROC (mean ± standard deviation) when trained on the AVAST-M lymph node dataset (*n* = 143). Within each panel, 14 different machine learning classifiers were trained to predict metastases: seven using tenfold CV (repeats = 1000) and seven using bootstrap resampling method (repeats = 1000). The two horizontal lines indicated within each panel denote the median AUROC of these seven classifiers, respectively. Statistical comparison using one-sided two-sample Welch *t* tests are indicated (bootstrap in green and tenfold CV in orange), see also Supplementary Tables S2c-d. Decision-Dx Melanoma: Decision-Dx MelanomaTM, LMC_150: Leeds Melanoma Cohort 150 gene signature. See "Methods" section "Machine learning analysis" for details about the classification algorithms. Source data are provided as a Source data file.

significant in the per-gene multivariate survival analyses after correcting for multiple testing (p value <0.05), further confirming that it is the combination of all the Cam_121 signature genes that provide the improved performance in predicting OS and PFS (Supplementary Figs. S11 and S12).

We undertook further multivariate Cox regression analyses for all protein-coding genes in this dataset ($n = 19{,}427$) and found no single gene was significantly associated with either PFS or OS after correcting for multiple testing (p value < 0.05) (Supplementary Data 2 and Supplementary Data 3).

A high weighted Cam_121 score reflected a lymphocyte depleted tumour with worse clinical outcomes. In order to identify the biological processes reflected by the signature, we ran pre-ranked gene set enrichment analyses on genes ordered by their shrunken log-fold change from the covariate-corrected differential expression analysis (see "Methods" section "Gene set enrichment analyses"). In doing this, we found that the top five significantly (FDR-corrected p value <0.05) downregulated hallmark gene sets resulting from this analysis included interferon (IFN) gamma response, IFN alpha response, allograft rejection, inflammatory response and IL6-JAK-STAT3 signalling (Supplementary Fig. S13). Interestingly, when we ran gene set enrichment on DEGs derived from the (entirely separate) LN samples ($n = 143$; Supplementary Fig. S14), we observed significant (FDR-corrected p value <0.01) downregulation of the exact same immune-related processes (Supplementary Fig. S14c). Therefore, indicating that the differential expression analyses (with a predominance of downregulated genes in both the primary melanoma and regional LN datasets) also reflect a significant downregulation of key immune-mediated processes in the samples from patients that developed metastases (Supplementary Figs. S3b, S13a and S14c).

We next used the Angelova dataset[25] to deconvolute the expression of immune cell subtypes within each sample (see "Methods" section "Immune cell correlation analysis"). We found a negative correlation between the weighted signature score and multiple immune cell types (with the highest correlation found with activated B cells ($\rho_{\text{Distant-Metastases=No}} = -0.8$, exact two-sided $p_{\text{Distant-Metastases=No}} < 2.2 \times 10^{-16}$; Fig. 4a), T cells ($\rho_{\text{Distant-Metastases=No}} = -0.73$, $p_{\text{Distant-Metastases=No}} < 2.2 \times 10^{-16}$; Fig. 4c), as well the overall immune cell expression score ($\rho_{\text{Distant-Metastases=No}} = -0.75$, $p_{\text{Distant-Metastases=No}} < 2.2 \times 10^{-16}$; Fig. 4e). We also found that samples with a high weighted Cam_121 expression score were more likely to develop metastases than samples with a low weighted Cam_121 expression score (Fig. 4b, d, f and Supplementary Table S3). Although nine of the Cam_121 signature genes were common with the Angelova immune marker genes (*TUBB*, *AIM2*, *CASQ1*, *NTRK1*, *FASLG*, *CCR3*, *P2RY14*, *PRF1* and *CCR5*), we were able to demonstrate that samples segregated based on overall immune cell score, with low immune cell expression clustering with high weighted Cam_121 gene expression scores (Fig. 4g), using PCA.

We further explored the relationship between the weighted Cam_121 gene expression score and histopathologically assessed tumour-infiltrating lymphocyte (TIL) scores (applying independent scoring criteria with both the Clark[26] and Melanoma Institute Australia (MIA) scores[27], see "Methods" section "Tumour infiltrating lymphocyte analysis"). This further confirmed a significant negative correlation between the Cam_121 signature and TIL scores, such that a higher signature score equated to a relatively immune-deprived tumour with consequent worse clinical outcomes (Fig. 5). It is nonetheless important to point out that the Cam_121 signature retained its prognostic influence even following correction for pathologically assessed TIL scores, and this remained consistent both within the AVAST-M and the external validation dataset from the LMC (Figs. 1c and 2e, respectively).

Table 1 Clinical utility of GEP test 5-year melanoma-specific survival (Leeds Melanoma Cohort Data).

AJCC stage	GEP_class	Death	Total	Proportion
1	All	17	194	0.09
1	High	11	118	0.09
1	Low	6	76	0.08
2	All	76	279	0.27
2	High	64	192	0.33
2	Low	12	87	0.14
3	All	44	76	0.58
3	High	35	58	0.60
3	Low	9	18	0.50

Number and proportion of deaths per combination of AJCC stage and Cam_121 risk level estimates (based on 0.33 quantile cut off on weighted Cam_121 gene expression score expression of all 121 genes). Source data are provided as a Source data file.

Discussion. The ability to identify primary melanoma patients at risk for disease recurrence is an important unmet need and effective prognostic biomarkers that could serve to guide adjuvant therapy are lacking. The 31-GEP assay (Decision-Dx Melanoma™, Castle Biosciences) has been developed in an attempt to address this clinical dilemma, however, the majority of published studies evaluating its performance have been retrospective or prospective cohort studies without a comparator group[28], and its use has not been advocated in established clinical guidelines[17]. We sought to identify whether the expression of genes in a primary melanoma tumour could predict for distant metastasis and survival, analysing data acquired from one of the largest phase III prospective adjuvant melanoma clinical trials associated with long-term patient outcome data[19]. We used covariate-corrected differential expression analyses to identify 121 genes significantly associated with distant metastases, which made up our signature, and found that this added prognostic value in both the prediction of metastasis and survival. The prognostic relevance was further confirmed in two independent external validation cohorts. Immune cell deconvolution analyses revealed that the weighted Cam_121 expression score negatively correlated with multiple measures of lymphocyte infiltration, with a high weighted signature score reflecting a relatively cold tumour immune microenvironment with worse long-term prognosis. These findings were cross-validated using pathologically assessed TIL scores, as well as gene set enrichment analyses, the latter showing that differential expression analyses in both primary melanoma ($n = 194$) and LN ($n = 143$) datasets reflected downregulation of the same key immune-mediated processes in association with metastases. That this conclusion was reached using unbiased differential expression, reaffirms the central importance of the immune system in this context.

The melanoma microenvironment consists of multiple immune and stromal cells, which play a critical role in regulating both the initiation and development of disease. Several studies have demonstrated the association of lymphocyte infiltration with longer survival[29–31], as well as an inverse relationship between TIL grade and the presence of LN metastases[27,32], implying that evaluating the tumour microenvironment landscape may hold promise for prognostic biomarkers. However, only a limited number of studies have investigated the immune landscape in primary melanomas. A transcriptomic analysis of primary melanomas identified six distinct subgroups based on their expression of immune-related, keratin and beta-catenin pathway genes[33]. In this study, patients with low immune but high beta-catenin score (CIC4) had the poorest OS[33]. A recent study utilising high-throughput sequencing of T-cell receptor beta-chain in T2–T4 primary melanomas ($n = 199$) indicated that the T-cell fraction accurately predicted PFS and was independent of other key clinico-pathologic covariates[34]. Although in our study it was difficult to discern specific immune cell subtypes using bulk RNA sequencing, given that the weighted Cam_121 score was strongly negatively correlated with B cells, T cells and all immune cells ($\rho_{\text{Distant-Metastases=No}} = -0.8$, -0.73 and -0.75, respectively, with exact two-sided $p_{\text{Distant-Metastases=No}} < 2.2 \times 10^{-16}$) and that IFN pathways dominated gene set enrichment, we regard this as

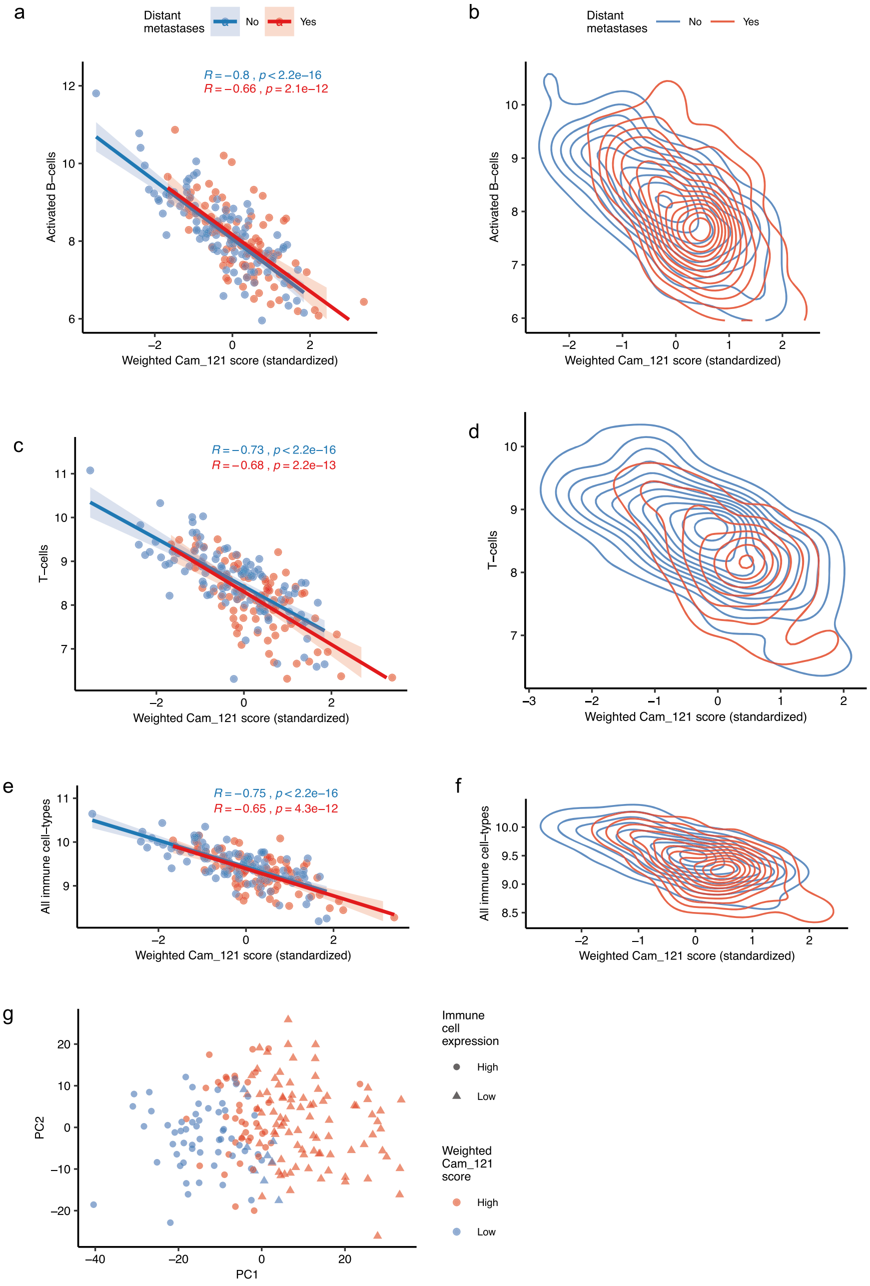


Fig. 4 Weighted Cam_121 score negatively correlates with immune cell expression scores, indicating that a lower weighted signature expression score is associated with a richer immune microenvironment and better prognosis. Scatterplots and density plots showing the relationship between the standardised weighted Cam_121 score (*x*-axes) and **a**, **b**, the median activated B-cell gene expression (*y*-axes), **c**, **d** the median T-cell gene expression (*y*-axes) and **e**, **f** the total immune score (median gene expression of all cell types derived from Angelova et al.[25], (*y* axes). Observations and lines of best fit are colour-coded according to their metastatic status, with red indicating relapse and blue indicating no relapse. The shaded region in the scatter plots of **a**, **c** and **e** corresponds to the 95% confidence interval of the line of best fit. The two-sided *p* values from the Pearson correlation coefficients are indicated for scatter plots. **g** Scatterplot of the scores of the observations on the two-first dimensions of a PCA analysis on the overall immune cell expression data. Observations are colour-coded according to their weighted Cam_121 expression score ("high"/"low" classification based on a quantile 0.33 split, indicated in red/yellow, respectively). Different symbols are used for observations with "high" (circles) and "low" (triangles) immune cell expression levels based on a median split. Source data are provided as a Source data file.

further evidence that a successful immune-mediated cytotoxic anti-tumour response exists in primary melanoma. Critically, we found that the signature retained its prognostic power even after correcting for TIL score, and it is our opinion that quantifying the expression of these key immune-mediated genes could potentially provide a more standardised and reproducible measure of immune activity. Furthermore, the prognostic relevance in both primary melanoma and LN datasets attests to the signatures' robustness. The challenge over the coming years will be to identify and validate a clinically relevant measure of lymphocytic abundance of relevance to primary CM, that can be easily implemented in real-life clinical practice. These studies will also need to consider aspects of cost-effectiveness, which have not been explored in this analysis.

Interrogating the LMC, we found that the GEP-designated high/low risk could be used to separate patients with ≥33% risk of

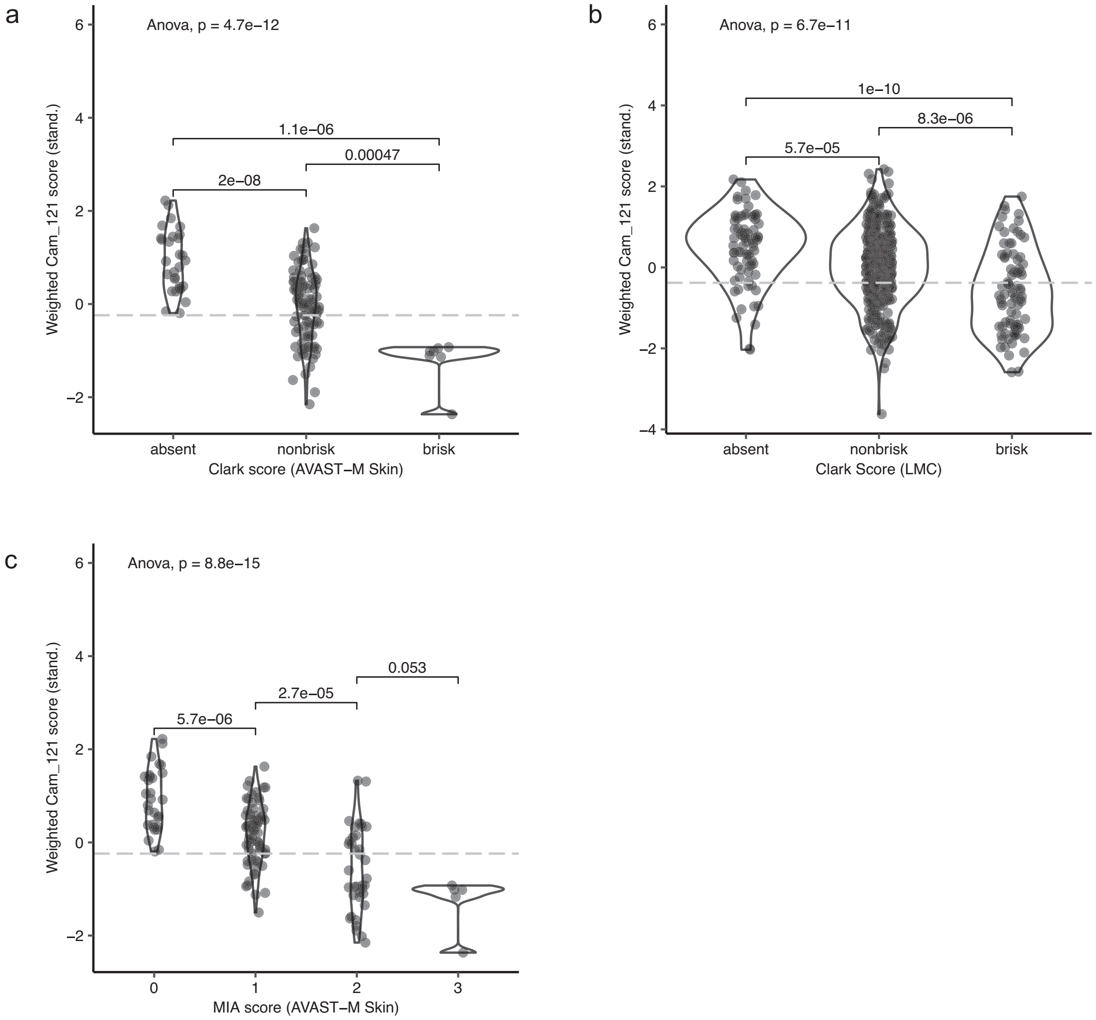


Fig. 5 Cam_121 negatively correlates with tumour-infiltrating lymphocyte counts. Violin plots showing the weighted (standardised) Cam_121 scores (*y*-axis) by levels of the TIL scores (x-axes) in the **a** AVAST-M primary melanoma dataset ($n = 137$) using Clark et al.[26] TIL scores, **b** Leeds Melanoma Cohort dataset ($n = 499$) using Clark et al. TIL scores and **c** AVAST-M primary melanoma dataset ($n = 139$) using MIA[27] TIL scores. The overall p values per plot are calculated using Fisher's ANOVA, and pairwise comparison p values were defined using two-sided Student's t test. The grey dotted line represents the 33% quantile cut off for the gene expression signature in each dataset. Source data are provided within the Source data file.

death at 5 years; a risk threshold for which we believe adjuvant systemic therapies could be considered. Conversely, it is envisaged that "low-risk" GEP profiles could also be used to "downstage" stage III patients for whom treatment might be unnecessary. There is substantial evidence supporting the importance of pre-treatment immune cell infiltration in eliciting anti-tumour responses with immunotherapy[35], however, it remains to be seen whether Cam_121 expression can predict therapeutic responses in this setting. Future well-designed prospective clinical trials will ultimately be required to examine whether Cam_121 can be used to better tailor adjuvant therapy for early-stage melanoma patients.

The present study has a number of advantages over previous analyses. First, the large sample size linked to a well-conducted prospective clinical trial enabled an objective assessment of the risk of distant metastases, in addition to the key survival measures of interest. Furthermore, the long duration of follow up (minimum of 6 years) in a cohort of patients predating modern approved adjuvant systemic therapies provided a unique insight into the "natural history" of primary CM. Finally, to our knowledge, this is the first large-scale biomarker analysis in primary melanoma to make use of data from comprehensive RNA sequencing. That such unbiased genome-wide assessment uncovered the dominance of immune-mediated genes reaffirms the central role of the host immune system's ability to respond to the tumour, resulting in immune editing or in some immune control.

It is important to point out that high-quality evidence guiding the best practice use of gene expression predictors, particularly in the context of early-stage CM are lacking. Future trials evaluating adjuvant therapies should examine both primary and locoregional melanoma samples using full RNA-sequencing technologies, to better characterise a molecular subtype/signature that could ultimately be used in conjunction with existing CM staging parameters and tailor future interventions more specifically to the individual. We believe that measures of lymphocytic infiltration should also be assessed. Ultimately such studies will need to show that randomising early-stage melanoma patients based on a high-risk Cam_121 GEP to an intervention (or a change in surveillance) leads to improved outcomes[36].

Our results indicate that the Cam_121 signature score complements conventional melanoma staging by contributing prognostically relevant information and could potentially be used to select early-stage melanoma patients at higher risk of relapse or death. Further carefully designed prospective clinical trials will help guide how molecular features can be incorporated with traditional clinico-pathologic features to best estimate individual risk and guide the optimal clinical use of molecular biomarkers.

Methods

AVAST-M melanoma cohort. This study made use of individual patient-level and transcriptomic data from the phase III adjuvant AVAST-M study, investigating the role of the angiogenesis inhibitor bevacizumab vs placebo in high-risk primary CM[18,19]. One thousand three hundred forty-three stage American Joint Committee on Cancer stage IIB (T3bN0M0 and T4aN0M0), IIC (T4bN0M0) and III (TxN1–3M0) cutaneous melanoma (seventh edition AJCC[20]) were recruited to the study over the period July 18, 2007–March 29, 2012, as previously described. The study (including the collection of DNA and RNA) was ethically approved in accordance with the Declaration of Helsinki (REC reference number 07/Q1606/15, 16th March 2007). Participants provided written informed consent to sampling of their tumour blocks during study recruitment (and prior to the investigational systemic therapy).

All study participants underwent a sentinel LN node biopsy, and if positive proceeded to a completion LN clearance as per the study protocol. Demographic (including gender, age, centre, as well as pathologic data (site of primary, Breslow depth, ulceration, LN involvement and *BRAF/NRAS* mutation by pyrosequencing)) was collected at the time of randomisation. Data were also collected on the timing, presence/absence and site/s of distant metastases (according to the findings from CT scanning). Data on overall and progression-free survival were collected with a minimum of 6 years follow up.

RNA-sequencing data was available on 204 primary melanoma samples of which 10 samples were removed from the downstream analyses owing to lack of data on all the clinical covariates, and 175 regional LN samples of which 32 samples removed from the downstream analyses owing to a lack of data on all clinical covariates (Supplementary Fig. S1).

Leeds Melanoma Cohort. A primary melanoma transcriptomic dataset from the LMC study (LMC DASL array) was used as independent replication. This represents a population-controlled cohort study, as previously described[16]. This study recorded data on MSS in 677 patients, calculated from the time of diagnosis to the time of last follow up or time of death from melanoma, whichever occurred first. The regression coefficient (beta) for each gene (reflecting differential expression in AVAST-M dataset) was used to generate a weighted signature score in the new dataset. Hence for further analysis, a per-sample weighted gene expression score for our Cam_121 gene signature was calculated by multiplying the expression value of each gene by its corresponding beta coefficient (Eq. 1) followed by z-score normalisation (zero mean–unit variance).

$$\text{Weighted signature score} = \sum_{i=1}^{n} \beta_i \times \text{gene}_i \tag{1}$$

where i ranges from 1 to number of genes in the signature and β corresponds to the beta coefficient of the respective gene obtained from DESeq2 analysis on AVAST-M melanoma cohort.

Lund Melanoma Cohort. Gene expression data on 223 primary tumours was generated using the Illumina DASL platform, as previously described[22]. Data on relapse-free survival as well as OS were collected. The DASL platform analysed 7752 genes and only 24 of the Cam_121 genes were present. Validation was undertaken using weighted signature scores as outlined above (Eq. 1).

The Cancer Genome Atlas-SKCM Cohort. The clinical and gene expression data from TCGA-SKCM[21], was downloaded from the cBioPortal[37]. The TCGA-SKCM dataset included only CMs, in particular any melanomas within these datasets from acral, mucosal and other rarer sites were excluded. Samples were filtered to a single sample per patient giving a total of 375 samples from 375 patients (including 87 primaries, 72 cutaneous relapses and 216 regional LNs).

Australia Melanoma Genome Project Cohort. All fresh frozen and FFPE samples were obtained in a method that was compliant with the relevant ethical regulations for work with human participants. The fresh-frozen tissue from the biospecimen bank of MIA[23]. All tissues and bloods form part of prospective collections of fresh-frozen samples accrued with written informed patient consent. The study was approved by the Sydney Local Health District RPAH zone ethics committee (Protocol No. X15-0454—previously X11-0289 and HREC/11/RPAH/444; Protocol No X17-0312—previously X11-0023 and HREC/11/RPAH/32; and Protocol No X15-0311—previously X10-0300 and HREC/10/RPAH/530). All samples were independently reviewed by expert melanoma pathologists to confirm the origin of each tumour from cutaneous skin.

Total RNA was extracted from fresh-frozen tissue using a mirVana miRNA Isolation Kit (Applied Biosystems, AM1560). RNA quality and presence of a small RNA fraction were measured using the Agilent 2100 RNA 6000 Nano and small RNA kits. RNA sequencing was performed using 1 µg of total RNA, which was converted into messenger RNA libraries using an Illumina mRNA TruSeq kit. RNA sequencing was performed using 2 × 75 bp paired-end reads on an Illumina Hiseq2000. Small RNA sequencing was performed using 1 µg of total RNA, which was converted into a small RNA libraries, size selection range 145–160 bp (RNA of 18–33 nucleotides) using Illumina's TruSeq Small RNA Library Preparation Kit and sequenced on an Illumina Hiseq2000 using 50 bp single-read sequencing with 1% control spiked in. RNA sequence reads were aligned to transcripts corresponding to ensemble 70 annotations using RSEM, raw sequences are available under study accession EGAS00001001552. Data from 55 samples from a mixture of tissue sites (including primary tumours, regional LNs, distant metastases, in-transit metastases and others) were available for this analysis.

mRNA extraction. Histopathological assessments of hematoxylin and eosin (H&E) stained slides were used to facilitate tumour sampling. Samples were consistently extracted from the least inflamed, least stromal regions of the invasive front of the tumour. RNA was extracted using the Roche High Pure FFPE RNA Micro Kit (cat# 04823125001; Genentech Biosciences) according to the manufacturer's recommendations. RNA quantity and quality were assessed using Agilent's 2100 bioanalyzer.

Expression data generation. Extracted RNA was sequenced on the Illumina exome-capture sequencing platform, using 50 base-pair paired-end sequencing. Quality control (QC) was performed using fastq_utils (https://github.com/nunofonseca/fastq_utils; v0.14.7) and FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; v0.11.7). The reads that passed QC were aligned to the reference genome (GRCh38) using TopHat2 (ref. [38]). Aligned reads were quantified using HTSeq[39]. Only those genes with more than five reads, as reported by HTSeq, in at least one sample were selected for further analysis. The sequencing data was of good quality with a median of ~50 million read-pairs/sample. A total of 446 tumour transcriptomes were profiled which included samples from: primary tumour ($n = 204$); LN ($n = 175$); local/distant relapse ($n = 58$) and uncategorised samples ($n = 9$). However, due to the clinical value of primary tumours in facilitating stratification at the earliest disease timepoint, we chose to focus our analyses on samples from cutaneous primaries ($n = 204$) and used the LN samples as an internal validation.

Clinical covariate selection. Firstly, the association between distant relapse (yes/no) and clinical covariates was studied both with or without controlling for length of follow up (defined as the time from diagnosis to last follow up) and for treatment (yes/no). When ignoring length of follow up and treatment, generalised Cochran-Mantel-Haenszel tests (R-package coin[40] v1.3-1) were used for nominal clinical predictors as they have the Pearson's Chi-square tests and Cochran-Armitage trend tests as special cases, when respectively considering the clinical covariate of interest as categorical or ordinal. For ordinal clinical covariates, we reported the "nominal/nominal" association results when the "nominal/ordinal" one was found less significant (as it is likely a sign that the assumption of linearity required by the ordinal test was not met). Mann–Whitney–Wilcoxon tests were used for continuous clinical covariates. When controlling for length of follow up and treatment, likelihood ratio tests comparing the fits of logistic regression models with and without the clinical predictor of interest were used. The p values were corrected for multiple testing using the Holm–Bonferroni method (Supplementary Table S1a). Note that, as the five-level stage variable was highly related to Nclass (Spearman correlation coefficient over 0.85), we picked the one with the lowest number of levels.

The variables two-level Breslow staging and two-level ECOG were significantly associated with relapse. The variable two-level treatment was found to be related to relapse, but is kept as control and the two-level EventMet was the variable of interest indicating whether the patient relapsed or not. Therefore, the covariates Stage, Breslow thickness, ECOG and treatment were accounted for in the design formula of DESeq2 (ref. [41]) without interactions and for further downstream analysis.

Secondly, the association between clinical covariates and OS (calculated from the time of diagnosis to the time of last follow up or death) and PFS (calculated from the time of diagnosis to the time of last follow up or death/progression to metastatic disease, whichever occurred first) was assessed by means of Cox proportional hazard models (R-package survival[42]). Both outcomes were considered as left-truncated due to delayed patient enrolment and right-censored due to loss of follow up or alive at the time of the end of the study. Six years was chosen as the minimum cut off for these analyses based on the original trial design.

Stage, sex, age and Nclass were significantly associated with both OS and PFS ($p < 0.05$; Supplementary Table S1b, c). The state of distant relapse ("EventMet") was the most important variable but was not of our interest, hence dropped. ECOG is a good predictor. Treatment was not significant ($p > 0.05$), but was kept in the analysis. Also for PFS, ulceration (Ulc) was borderline at 5% level and was dropped from further analysis. Therefore, stage, sex, age, Nclass, ECOG and treatment were corrected for in subsequent gene-level survival analyses. The AVAST-M primary melanoma dataset was also corrected for TIL counts (Clark Score, see also section "Tumour-infiltrating lymphocyte analysis").

Differential expression analysis. Differential expression analysis between primary tumours that became metastatic vs those that remained non-metastatic over the 6-year study period was performed using the package DESeq2 (ref. [41]; v1.18; R v3.6.1). The negative binomial models we considered controlled for the clinical covariates Stage, Breslow thickness, ECOG and treatment, as well as for the library size (offset). Raw read counts were provided as the input, with each column representing a sample and each row representing a gene, along with the categorical clinical information about each sample as colData. Samples with missing information for any of these four covariates were removed from the analysis, leaving 194 samples in total. The adjusted p value cut off (FDR) was set to 0.1 using the alpha parameter in DESeq2 results function and genes with FDR <0.1 were considered significantly differentially expressed.

Log-fold change shrinkage was applied using the lfcshrink function with apeglm method from the apeglm package[43] (v1.6.0; R v3.6). For visualisation and other downstream analysis, variance stabilising transformation (vst) was used by means of the DESeq2's varianceStabilizingTransformation function with option blind = FALSE.

Machine learning analysis. This section explains the steps followed to develop a ML classifier for each signature and to evaluate their performance in predicting relapse (yes/no). A summary of the pipeline is outlined in Supplementary Fig. S15. The following steps were conducted using the following packages; caret[44] v6.0-86, DESeq2 v1.28.1; R v4.0.2, snakemake[45] v5.17.0. The AVAST-M primary melanoma dataset ($n = 194$) was used for classification model development and the AVAST-M LN dataset ($n = 143$) was used to test the performance of the final model (internal validation). In an independent analysis, we used only the AVAST-M LN dataset for both training/testing ("Methods" section "Model development and selection").

Dataset preparation and pre-processing. The AVAST-M primary melanoma (training) dataset was prepared such that each column/feature contained information about all the patients/samples. These features could either be the expression values of the genes within the signature and/or the categorical clinical covariate metadata. In the latter case, the categorical clinical covariates were converted to numeric dummy variables using one-hot encoding. The clinical outcome data (relapse vs non-relapse) were used as labels for the analyses. The AVAST-M LN dataset (testing) dataset was pre-processed in the same way, such that the order of the features was preserved as in the training dataset and in case where the clinical covariates were used as features, they were converted to dummy variables from the train clinical covariates using the predict function in R.

In case of the gene expression data, vst transformation was applied to both training and testing dataset. To apply vst transformation on the testing dataset, mean-dispersion estimates learnt on the training dataset were used.

Next, the features corresponding to near-zero variance were removed using the default parameters of the trainControl function (R-package caret; i.e., freqCut = 95/5, uniqueCut = 10). The same feature(s) were removed from the testing dataset before evaluating the performance of the fully trained model.

Model development and selection. The aim of this analysis was to critically assess whether the Cam_121 gene expression signature (with or without clinical covariates as features) could outperform clinical covariates alone in predicting relapse (yes/no). We also compared this to the predictive power of two published prognostic signatures (LMC_150 (ref. [16]) and Decision-Dx Melanoma[12]) in an independent analysis. In this model, the training/testing was carried out on the AVAST-M LN dataset instead of AVAST-M primary melanoma dataset on which feature selection was performed for Cam_121 and clinical covariates. This was undertaken to reduce the risk of over-optimistic results that might arise from feature selection and testing from the same dataset. In developing a classification model for each of these five signatures of interest, seven different ML classifiers were considered, including; Bayesian generalised linear model[46] (bayesglm), Lasso and elastic-net regularised generalised linear model[47] (glmnet), k-nearest neighbour[48] (knn), linear discriminant analysis[49] (lda), random forest[50] (rf) and support vector machine[51] with linear (svmLinear) and radial kernel (svmRadial).

To avoid overfitting, repeated tenfold cross validation (repeats = 1000) was performed for model development and evaluation. Leave-one-out cross validation method (Supplementary Fig. S15) and bootstrap resampling method (Fig. 3e; repeats = 1000) were also tested to see if the choice of resampling method altered our results. This was implemented using the trainControl function (R-package caret).

At each training step, a random search was performed using 100 random (combinations of) hyperparameter(s) and the set of hyperparameter(s) leading to the largest maximum AUROC estimate on the training dataset was selected. Using this approach, we obtained 14 different classification models (7 classifiers × 2 resampling methods) for each of the five different signatures. To select the final best performing classification model for each signature, the model giving the highest AUROC value was selected for testing on the AVAST-M LN dataset.

When training the random forest classifier on the feature set Cam_121 with clinical covariates, we estimated the average feature importance score based on repeated tenfold CV and displayed them by means of boxplots using geom_boxplot function in R.

Gene expression vs covariate performance comparison per patient. In order to compare the performance of the signatures: "Cam_121 + clinical covariates", "Cam_121" and "clinical covariates" on the AVAST-M LN dataset ($n = 143$) on a per-patient basis, the labels predicted by their respective best performing classifiers were extracted (tenfold CV + svmRadial, tenfold CV + svmRadial and tenfold CV + svmLinear respectively). Venn diagrams obtained by means of the venn.diagram function from the VennDiagram package[52] (v1.6.20) in R were used to visualise overlaps.

Statistical analyses. To check whether the Cam_121 gene expression signature performed better in predicting relapse (yes/no) than models built on clinical covariates alone across multiple ($n = 7$) classifiers, Welch Two Sample t tests were used (R function t.test with option var.equal = FALSE, paired = FALSE and alternative = "greater") for each performance metric and each combination of signature and clinical covariate at the 5% level. The null hypothesis was that the true difference in mean performance across seven classifiers between both conditions ("clinical covariates alone" vs "signature with/without clinical covariates") equals 0, while the alternative hypothesis was that the true difference in means is >0.

To compare the AUROC obtained on the testing dataset, DeLong's tests[53] were used using the roc.test function with alternative "greater" from the pROC package[52] v1.16.2 in R, where the null hypothesis is two AUROC obtained from the model trained on gene expression as features and the model trained on clinical covariates as features are equal, while the alternate hypothesis is that the model trained on gene expression as features performs better than the model trained on clinical covariates as features. The p values and the z decision threshold values from the test were reported.

Determination of the weighted expression score cut-off to define "high" and "low" absolute risk of death at 5 years. Data from the LMC were used to calculate the absolute risk of death at 5 years (this dataset was chosen for this analysis due to the preponderance of early-stage patients; stage I = 194 samples; stage II = 279 samples; stage III = 76 samples). Five-year MSS was calculated such that those patients who died due to melanoma within 5 years of follow up were assigned event = "Yes" and those that did not were assigned event = "No". Those patients who did not yet die and were followed up for <5 years were removed from the analysis due to inadequate follow up.

The quantile cut offs 0.25, 0.33 and 0.5 were used to divide patients into high/low groups based on their corresponding weighted Cam_121 expression score. Absolute risk of death at 5 years was calculated as the ratio of patients where event = "Yes" to the total number of patients within each stage (I–III). The cut off giving the maximal separation (of absolute risk of death) between high/low groups was selected. This was achieved using a 0.33 quantile cut off of the weighed Cam_121 expression score and subsequent references to high/low Cam_121 risk groups refer to these high/low weighted stratification cohorts.

Survival analyses. For each sample, a vector of weighted signature expression scores was calculated by using Eq. 1 on the vst normalised gene expression data. The standardised scores were then used as a continuous predictor in Cox regression models fitted by means of the coxph function of the survival package[42] (v3.1-12) in R (v3.6.3). The HR (95% CI) and p values corresponding to the signature were reported in both univariate and multivariate analyses. Note that in case of the two published signatures, median gene expression scores were used instead of the weighted gene expression scores.

In order to display Kaplan–Meier (K–M) survival curves, samples were divided into "high"/"low" signature expression groups based on the 0.33 quantile cut off which we obtained from the absolute 5-year risk assessment in "Methods" section "Determination of the weighted expression score cut off to define "high" and "low" absolute risk of death at 5 years". Samples with weighted signature expression score greater than this cut off were assigned to the "high" group and those with weighted signature expression score lower than this cut off were assigned to the "low" group. Of note, we found that the 0.33 quantile cut off of the weighted gene expression scores were remarkably consistent across all the external datasets (data not shown). The survival distribution of both groups was finally compared by means of logrank tests using survfit function from the survival package (v3.1-12) in R (v3.6.3). The parallel processing was conducted using snakemake[45] v5.17.0. The K–M curves were plotted using plot function from the R-package graphics[54] v3.6.3 and ggsurvplot function from R-package survminer[55] v0.4.7.

Tumour immune microenvironment analysis. Sample-level gene expression data from the AVAST-M primary melanoma cohort was deconvoluted into infiltrating immune cell scores using the Angelova dataset[25]. This dataset reports 812 marker genes corresponding to 31 immune cell subtypes. Out of these 812 genes, 53 genes were missing from our 38,690 gene list. Therefore, two immune cell subtypes MDSC (myeloid-derived suppressor cells) and NK56_bright (natural-killer CD56$^{\mathrm{bright}}$ cells), with >1% of missing marker genes were removed from further analysis, leaving 719 marker genes corresponding to 29 cell types.

Immune cell correlation analysis. To perform correlation analysis for each cell type, the median of the corresponding marker genes' expression for each sample (y-axis) was plotted against the weighted Cam_121 expression score for that sample (*x*-axis). To calculate the overall immune score, the median of all 719 marker genes was used for the analysis.

A regression line was fitted through these points using the geom_smooth function (method = "lm") of the ggplot2 package[56] (v3.3.0) in R (v3.6.2). The Pearson's correlation coefficient (ρ) between the immune cell score and the signature, as well as the p value of the corresponding test of association were estimated by means of the function stat_cor of the ggpubr package[57] (v0.2.5) in R (v3.6.2). Samples were coloured according to their metastatic status. With red points indicating the samples obtained from patients who later metastasised (Yes) and blue points indicating the samples obtained from patients who didn't (No). The density plots were made using the geom_density_2d function of R-package ggplot2 (ref. [51]; v 3.3.0).

To confirm the relationship between the signature and immune cells, a PCA analysis was performed. For each cell type, the corresponding marker genes' expression was projected into the principal component space and the first two principal components explaining the maximum variance were plotted against each other. These samples were then coloured by the "high"/"low" weighted Cam_121 expression score groups obtained during the survival analysis and shaped by the "high"/"low" groups based on marker genes' median expression corresponding to that particular cell type. Here, to divide the samples into two independent categories based on their marker genes' expression, the median cut off was used, where the sample with marker genes' median expression value above its overall median value was assigned to the "high" group and that with marker genes' median expression value below its overall median value was assigned to the "low" group.

Gene set enrichment analyses. Preranked GSEA (GSEA-P) was implemented using the GSEAPreranked tool of the GSEA software from Broad Institute[58,59] (v4.0.2). Hallmark gene sets were downloaded from the MSigDB database[60] (v6.2.0). The genes were preranked according to their shrinked log-fold change values obtained in the differential expression analysis and the GSEAPreranked tool was run with default parameters with the enrichment statistic set to "classic".

Tumour-infiltrating lymphocyte analysis. H&E slides corresponding to each of the 194 samples were digitally scanned to 40× magnification using the Vectra Polaris scanner from AKOYA biosciences. TIL scores were double blindly evaluated by two experienced pathologists. Two different scoring methods were used including; (i) the Clark scoring[26] and (ii) the MIA system[27]. This resulted in an agreement of 56% and 40%, respectively. This lack of consistency in scoring (particularly within the "non-brisk" group) has been previously noted in the literature[61,62]. We used a third independent expert pathologist to assess those slides where the two pathologists failed to agree. After removing the slides with poor scan quality, we had Clark TIL scores for 133 primary tumours and MIA TIL scores for 135 primary tumours.

Once the scores were obtained, a violin plot was plotted between the scores and the standardised weighted Cam_121 score using the geom_violin function of ggplot2 package. The p values for pairwise comparisons were obtained using t test and the global p value was computed using ANOVA. This was implemented using the stat_compare_means function from the R-package ggpubr (v0.2.5)[57].

Visualisation of inherent distribution of samples. To visualise if the samples cluster by their metastatic status, PCA was performed on the primary melanoma samples ($n = 204$) and LN samples ($n = 175$) using 1000 most variable genes. This analysis was performed using the prcomp function from R-package stats[54] (v3.6.2), plotted using the qplot function from ggplot2 package[56] (v 3.2.1) and the scree plot was generated using the screeplot function, also using the R-package stats[54] (v3.6.2). The samples were further coloured by whether they metastasised or not and shaped by their tissue of origin.

Testing signature performance against randomly selected genes. In order to test the performance of our signature against randomly selected genes, random genes of the 121 gene length were selected from 19,434 protein-coding genes. The analysis was repeated using the exact same pipeline to compare its performance against our signature. This process was repeated 1000 times without replacement using 1000 different seeds and the p value testing the significance of our signature was defined as a left-tailed event for predicting OS/PFS survival. This analysis was inspired from the SigCheck package[63] (v2.14.0) in R (v3.5.1), whereby the plotting function sigCheckPlotSurvival was modified to accept scores generated from our analysis.

Power calculation for the external validation datasets. To assess whether significant validation of our signature was likely in the external validation datasets, we performed a simulation-based power analysis considering $R = 2500$ Monte Carlo samples. Simulation parameters, like the proportion of events for both outcomes, the hazard ratios and the predictor and right-censored time-to-event distributions, were based on the AVAST-M study. The log-normal and exponential distributions were respectively chosen to model time-to-relapse and time-to-death from time-to-relapse. The normal distribution was selected to model censoring for both outcomes.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw RNA-sequencing data (forward and reverse fastq files) has been made available at the European Genome-Phenome Archive at the EBI under the following dataset accession ID: EGAD00001006401. The source data underlying this paper are also available through the GitHub repository: Manikgarg/MelanomaTranscriptomics [https://github.com/Manikgarg/MelanomaTranscriptomics/tree/master/Source_Data)[64]. The clinical and gene expression data from The Cancer Genome Atlas (TCGA-SKCM), can be downloaded from the cBioPortal[37]. Data from the Leeds Melanoma Cohort[16], Lund Melanoma Cohort[22] and the Australia Melanoma Genome Project[23] are available from the source publications.The MSigDB database[60] gene set collections are available for download from http://www.gsea-msigdb.org/gsea/downloads.jsp#msigdb. Source data are provided with this paper.

Code availability

The code to reproduce the results is available at the GitHub repository: Manikgarg/MelanomaTranscriptomics (https://github.com/Manikgarg/MelanomaTranscriptomics/tree/master/scripts)[64].

Received: 10 March 2020; Accepted: 15 January 2021;
Published online: 18 February 2021

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **70**, 7–30 (2020).
2. Spain, L., Larkin, J. & Turajlic, S. New survival standards for advanced melanoma. *Br J Cancer* **122**, 1275–1276 (2020).
3. Whiteman, D. C., Green, A. C. & Olsen, C. M. The growing burden of invasive melanoma: projections of incidence rates and numbers of new cases in six susceptible populations through 2031. *J. Investig. Dermatol.* **136**, 1161–1171 (2016).
4. Tarhini, A. A. et al. Phase III study of adjuvant ipilimumab (3 or 10 mg/kg) versus high-dose interferon alfa-2b for resected high-risk melanoma: North American Intergroup E1609. *J. Clin. Oncol.* **38**, 567–575 (2020).
5. Robert, C. et al. Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study. *Lancet Oncol.* **20**, 1239–1251 (2019).
6. Maio, M. et al. Adjuvant vemurafenib in resected, BRAF(V600) mutation-positive melanoma (BRIM8): a randomised, double-blind, placebo-controlled, multicentre, phase 3 trial. *Lancet Oncol.* **19**, 510–520 (2018).
7. Long, G. V. et al. Adjuvant dabrafenib plus trametinib in stage III BRAF-mutated melanoma. *N. Engl. J. Med.* **377**, 1813–1823 (2017).
8. Luke, J. J. et al. KEYNOTE-716: Phase III study of adjuvant pembrolizumab versus placebo in resected high-risk stage II melanoma. *Future Oncol.* **16**, 4429–4438 (2020).
9. Poklepovic, A. S. & Luke, J. J. Considering adjuvant therapy for stage II melanoma. *Cancer* **126**, 1166–1174 (2019).
10. Bhutiani, N., Egger, M. E. & McMasters, K. M. Optimizing follow-up assessment of patients with cutaneous melanoma. *Ann. Surg. Oncol.* **24**, 861–863 (2017).
11. Jonsson, G. et al. Gene expression profiling-based identification of molecular subtypes in stage IV melanomas with different clinical outcome. *Clin. Cancer Res.* **16**, 3356–3367 (2010).
12. Gerami, P. et al. Gene expression profiling for molecular staging of cutaneous melanoma in patients undergoing sentinel lymph node biopsy. *J. Am. Acad. Dermatol* **72**, 780–5.e3 (2015).
13. Zager, J. S. et al. Performance of a prognostic 31-gene expression profile in an independent cohort of 523 cutaneous melanoma patients. *BMC Cancer* **18**, 130 (2018).
14. Gastman, B. R. et al. Identification of patients at risk for metastasis using a prognostic 31-gene expression profile in subpopulations of melanoma patients with favorable outcomes by standard criteria. *J. Am. Acad. Dermatol.* **80**, 149–157 (2018).
15. Vetto, J. T. et al. Guidance of sentinel lymph node biopsy decisions in patients with T1-T2 melanoma using gene expression profiling. *Future Oncol.* **15**, 1207–1217 (2019).
16. Thakur, R. et al. Transcriptomic analysis reveals prognostic molecular signatures of stage I melanoma. *Clin. Cancer Res.* **25**, 7424–7435 (2019).
17. Kovarik, C. L., Chu, E. Y. & Adamson, A. S. Gene expression profile testing for thin melanoma: evidence to support clinical use remains thin. *JAMA Dermatol.* **156**, 837–838 (2020).

18. Corrie, P. G. et al. Adjuvant bevacizumab in patients with melanoma at high risk of recurrence (AVAST-M): preplanned interim results from a multicentre, open-label, randomised controlled phase 3 study. *Lancet Oncol.* **15**, 620–630 (2014).
19. Corrie, P. G. et al. Adjuvant bevacizumab for melanoma patients at high risk of recurrence: survival analysis of the AVAST-M trial. *Ann. Oncol.* **29**, 1843–1852 (2018).
20. Balch, C. M. et al. Final version of 2009 AJCC melanoma staging and classification. *J. Clin. Oncol.* **27**, 6199–6206 (2009).
21. Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681–1696 (2015).
22. Harbst, K. et al. Molecular profiling reveals low- and high-grade forms of primary melanoma. *Clin. Cancer Res.* **18**, 4026–4036 (2012).
23. Hayward, N. K. et al. Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175–180 (2017).
24. Venet, D., Dumont, J. E. & Detours, V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput. Biol.* **7**, e1002240 (2011).
25. Angelova, M. et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol.* **16**, 64 (2015).
26. Clark, W. H. Jr. et al. Model predicting survival in stage I melanoma based on tumor progression. *J. Natl Cancer Inst.* **81**, 1893–1904 (1989).
27. Azimi, F. et al. Tumor-infiltrating lymphocyte grade is an independent predictor of sentinel lymph node status and survival in patients with cutaneous melanoma. *J. Clin. Oncol.* **30**, 2678–2683 (2012).
28. Berman, B. et al. Appropriate use criteria for the integration of diagnostic and prognostic gene expression profile assays into the management of cutaneous malignant melanoma: an expert panel consensus-based modified delphi process assessment. *SKIN J. Cutan. Med.* **3**, 291–306 (2019).
29. van Houdt, I. S. et al. Favorable outcome in clinically stage II melanoma patients is associated with the presence of activated tumor infiltrating T-lymphocytes and preserved MHC class I antigen expression. *Int. J. Cancer* **123**, 609–615 (2008).
30. Krynitz, B., Rozell, B. L., Lyth, J., Smedby, K. E. & Lindelof, B. Cutaneous malignant melanoma in the Swedish organ transplantation cohort: a study of clinicopathological characteristics and mortality. *J. Am. Acad. Dermatol.* **73**, 106–113.e2 (2015).
31. Thomas, N. E. et al. Tumor-infiltrating lymphocyte grade in primary melanomas is independently associated with melanoma-specific survival in the population-based genes, environment and melanoma study. *J. Clin. Oncol.* **31**, 4252–4259 (2013).
32. Donizy, P. et al. Paucity of tumor-infiltrating lymphocytes is an unfavorable prognosticator and predicts lymph node metastases in cutaneous melanoma patients. *Anticancer Res.* **35**, 351–358 (2015).
33. Nsengimana, J. et al. beta-Catenin-mediated immune evasion pathway frequently operates in primary cutaneous melanomas. *J. Clin. Investig.* **128**, 2048–2063 (2018).
34. Pruessmann, W. et al. Molecular analysis of primary melanoma T cells identifies patients at risk for metastatic recurrence. *Nat. Cancer* **1**, 197–209 (2020).
35. Vilain, R. E. et al. Dynamic changes in PD-L1 expression and immune infiltrates early during treatment predict response to PD-1 blockade in melanoma. *Clin. Cancer Res.* **23**, 5024–5033 (2017).
36. Grossman, D. et al. Prognostic gene expression profiling in melanoma: necessary steps to incorporate into clinical practice. *Melanoma Manag.* **6**, Mmt32 (2019).
37. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
38. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
39. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
40. Hothorn, T., Hornik, K., van de Wiel, M. A. & Zeileis, A. Lego system for conditional inference. *Am. Stat.* **60**, 257–263 (2006).
41. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
42. Therneau, T. A package for survival analysis in R. R package version 3.1-12 https://CRAN.R-project.org/package=survival (2015).
43. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).
44. Max Kuhn. Contributions from Jed Wing, S.W., Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan and Tyler Hunt. caret: Classification and Regression Training. R package version 6.0-84. https://CRAN.R-project.org/package=caret (2019).
45. Köster, J. & Rahmann, S. Snakemake–a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).
46. Gelman, A., Jakulin, A., Pittau, M. G. & Su, Y. S. A weakly informative default prior distribution for logistic and other regression models. *Ann. Appl. Stat.* **2**, 1360–1383 (2009).
47. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
48. Thomas, C. & Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. theory* **13**, 21–27 (1967).
49. Fisher, R. A. The use of multiple measurements in taxonomic problems. *Ann. Eugen.* **7**, 179–188 (1936).
50. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
51. Boser, B., Guyon, I. & Vapnik, V. An algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152 (ACM, Pittsburgh, PA, 1992).
52. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
53. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
54. R Core Team. *R: a Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
55. Kassambara, A., Kosinski, M. & Biecek, P. survminer: Drawing Survival Curves using 'ggplot2'. https://cran.r-project.org/web/packages/survminer/index.html (2020).
56. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
57. Kassambara, A. ggpubr: 'ggplot2' Based Publication Ready Plots Vol. R package version 0.2.5 https://cran.r-project.org/web/packages/ggpubr/index.html (2020).
58. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
59. Mootha, V. K. et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
60. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
61. Saldanha, G., Flatman, K., Teo, K. W. & Bamford, M. A novel numerical scoring system for melanoma tumor-infiltrating lymphocytes has better prognostic value than standard scoring. *Am. J. Surg. Pathol.* **41**, 906–914 (2017).
62. Park, C. K. & Kim, S. K. Clinicopathological significance of intratumoral and peritumoral lymphocytes and lymphocyte score based on the histologic subtypes of cutaneous melanoma. *Oncotarget* **8**, 14759–14769 (2017).
63. Stark, R. & Norden, J. SigCheck: check a gene signature's prognostic performance against random signatures, known signatures, and permuted data/metadata. R package version 2.18.0. https://rdrr.io/bioc/SigCheck/ (2019).
64. Garg, M. Melanoma transcriptomics, v1.0.0 edn https://doi.org/10.5281/zenodo.4420033 (2020).

Acknowledgements

We would like to thank Neera Maroo, Leticia Campo and the Translational Histopathology Laboratory at the Oxford Department of Oncology for sample biobanking and processing, as well as digitally scanning the primary tumour H&E slides. Vivek Iyer for cross-checking the differential expression analyses. Wolfgang Huber and Janet M. Thornton for advice on machine learning analyses. Michael I. Love for advice on data normalisation for machine learning. Nicholas K. Hayward for sharing RNA-sequencing data from the Australia Melanoma Genome Project. Danish Memon for expert input on immune cell deconvolution and Andrea Manrique-Rincon for her help on the critical interpretation of these analyses. Alastair Droop for critical review of the manuscript. This work was supported by Cancer Research UK and the Wellcome Trust. Bevacizumab was supplied by Genentech pharmaceuticals. R.R. was partially supported by the 2019 Cycle for Survival's Equinox Innovation Award. N.A.F. was partially supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 668981. The Leeds Melanoma Cohort research was carried out with research funding CR-UK; C588/A19167, C8216/A6129 and C588/A10721 and NIH; CA83115. M.R.M. is supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. R.A.S. supported by a National Health and Medical Research Council of Australia (NHMRC) Program Grant and Practitioner Fellowship. G.V.L. is supported by the University of Sydney Medical Foundation. Support from the Ainsworth Foundation, the Cameron Family, as well as from colleagues at Melanoma Institute Australia and Royal Prince Alfred Hospital are also gratefully acknowledged.

Author contributions

R.R., A.B. and D.J.A. conceived the project. MRM co-ordinated the sample biobanking. M.W. and Y.Y. co-ordinated the mRNA extraction and RNA sequencing from the AVAST-M trial. R.R. co-ordinated the sequencing and clinical data extraction. N.A.F. derived the RNA-seq counts from raw sequencing reads. R.R., D.J.A., A.B. and M.G. developed the clinical questions and experiments and M.G. carried out the analysis. D.L.C. ran the statistical analyses. J.N., T.B. and J.N.B. validated the findings in the Leeds Melanoma Cohort. M.L. and G.B.J. validated the findings in the Lund Melanoma Cohort. P.C., M.R.M. and C.P. provided senior input on the translational scope of the project. S.M., N.S. and J.T. annotated histopathological slides with the tumour-infiltrating scores. I.V., S.L., F.N., J.S.W., J.F.T., G.V.L. and R.A.S. facilitated validation within the Australian Melanoma Genome Project. M.G. and R.R. wrote the manuscript. A.B. and D.J.A. provided overall study supervision. All authors approved the final manuscript.

Competing interests

P.C. is an advisory board member at Roche. M.W./Y.Y. are employees and stockholders at Roche/Genentech. D.J.A. is a consultant at Microbiotica. R.A.S. has received fees for professional services from Qbiotics, Novartis, MSD Sharp & Dohme, NeraCare, AMGEN Inc., Bristol-Myers Squibb, Myriad Genetics, GlaxoSmithKline. G.V.L. is consultant advisor for Aduro Biotech Inc, Amgen Inc, Array Biopharma inc, Boehringer Ingelheim International GmbH, Bristol-Myers Squibb, Highlight Therapeutics S.L., Merck Sharpe & Dohme, Novartis Pharma AG, QBiotics Group Limited, Regeneron Pharmaceuticals Inc, SkylineDX B.V. J.F.T. has received honoraria for advisory board participation from BMS Australia, MSD Australia, GSK and Provectus Inc, and travel support from GlaxoSmithKline and Provectus Inc. All other authors report no conflicts of interest.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21207-2.

Correspondence and requests for materials should be addressed to R.R.

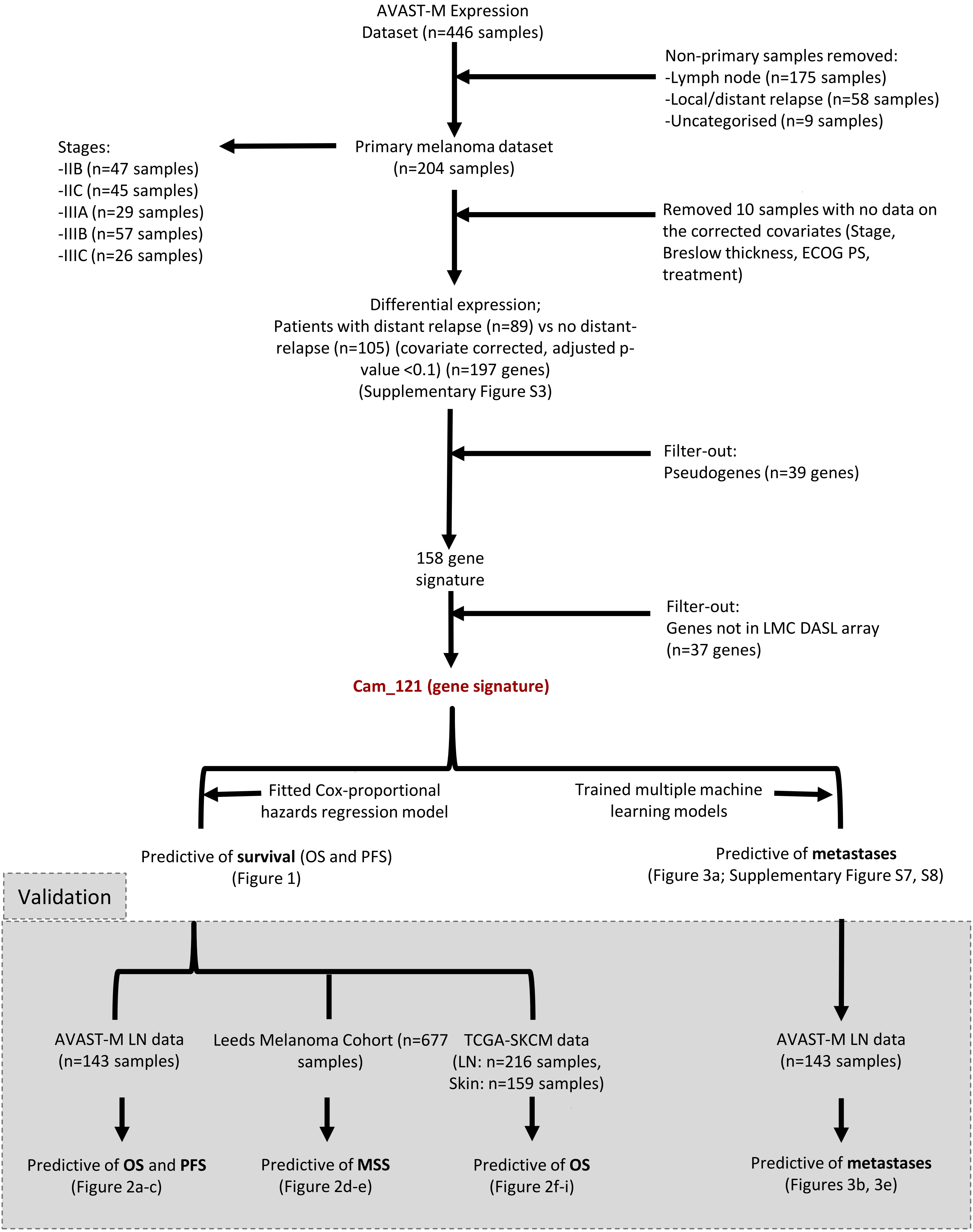
Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

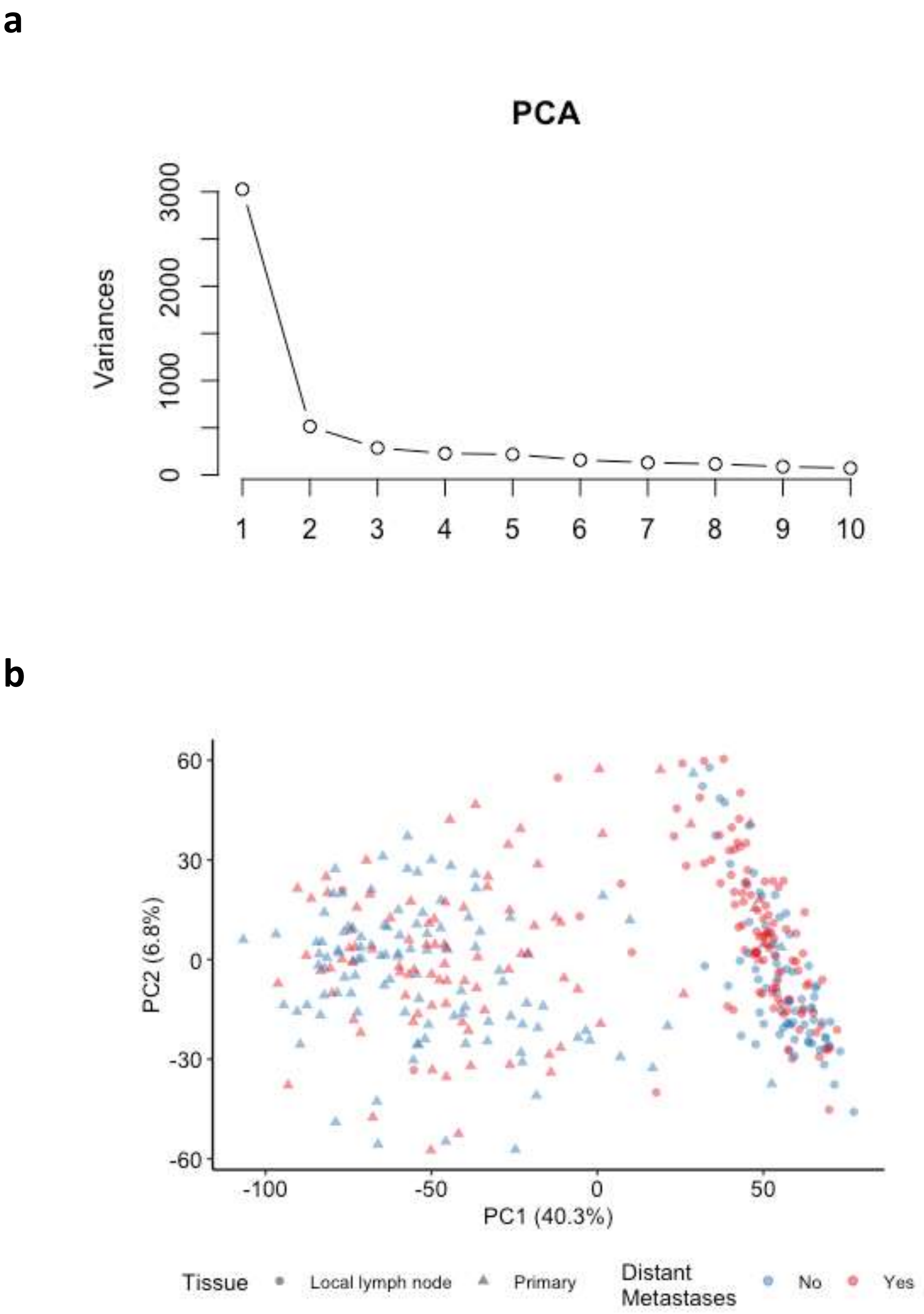
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

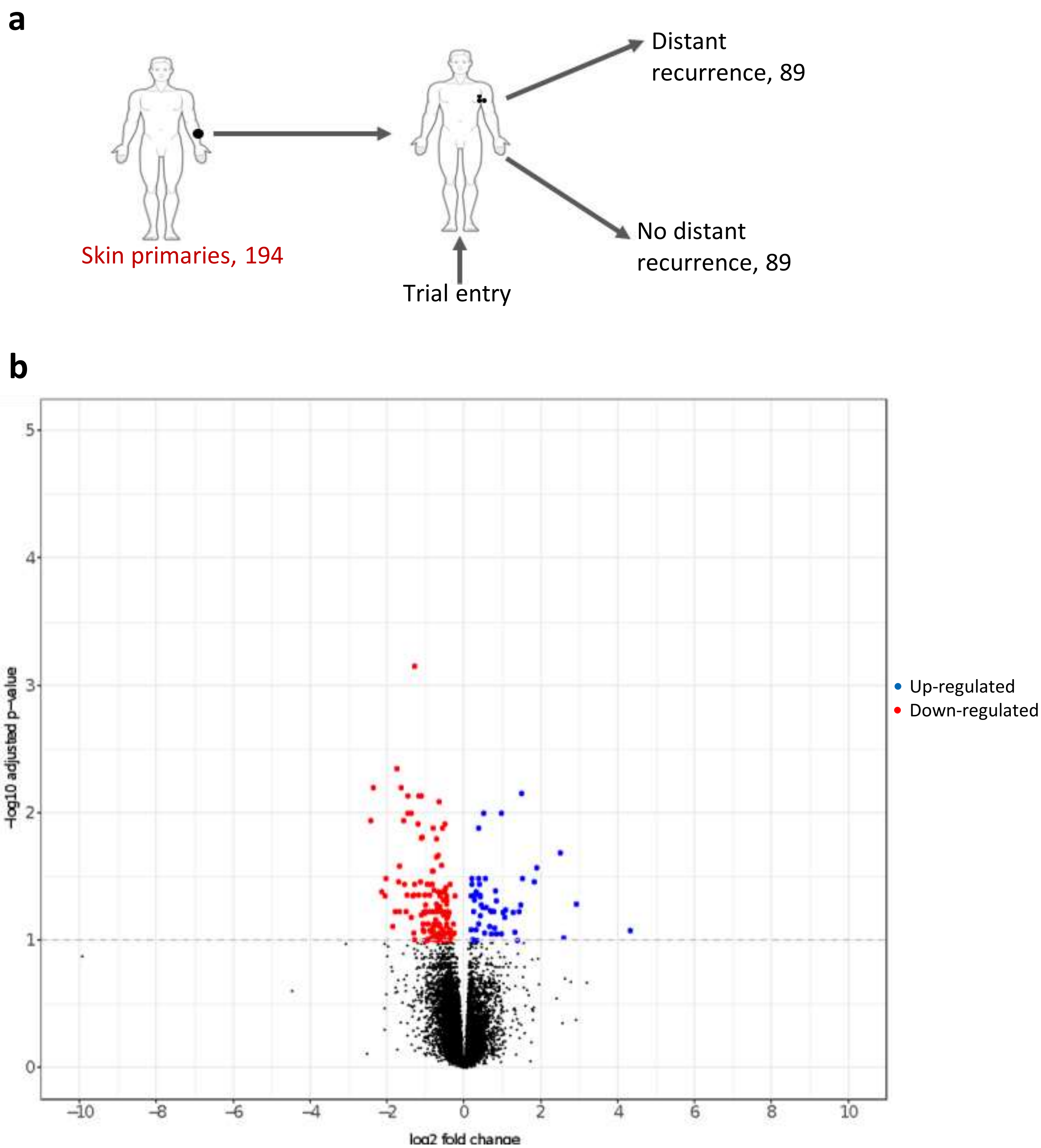
© The Author(s) 2021



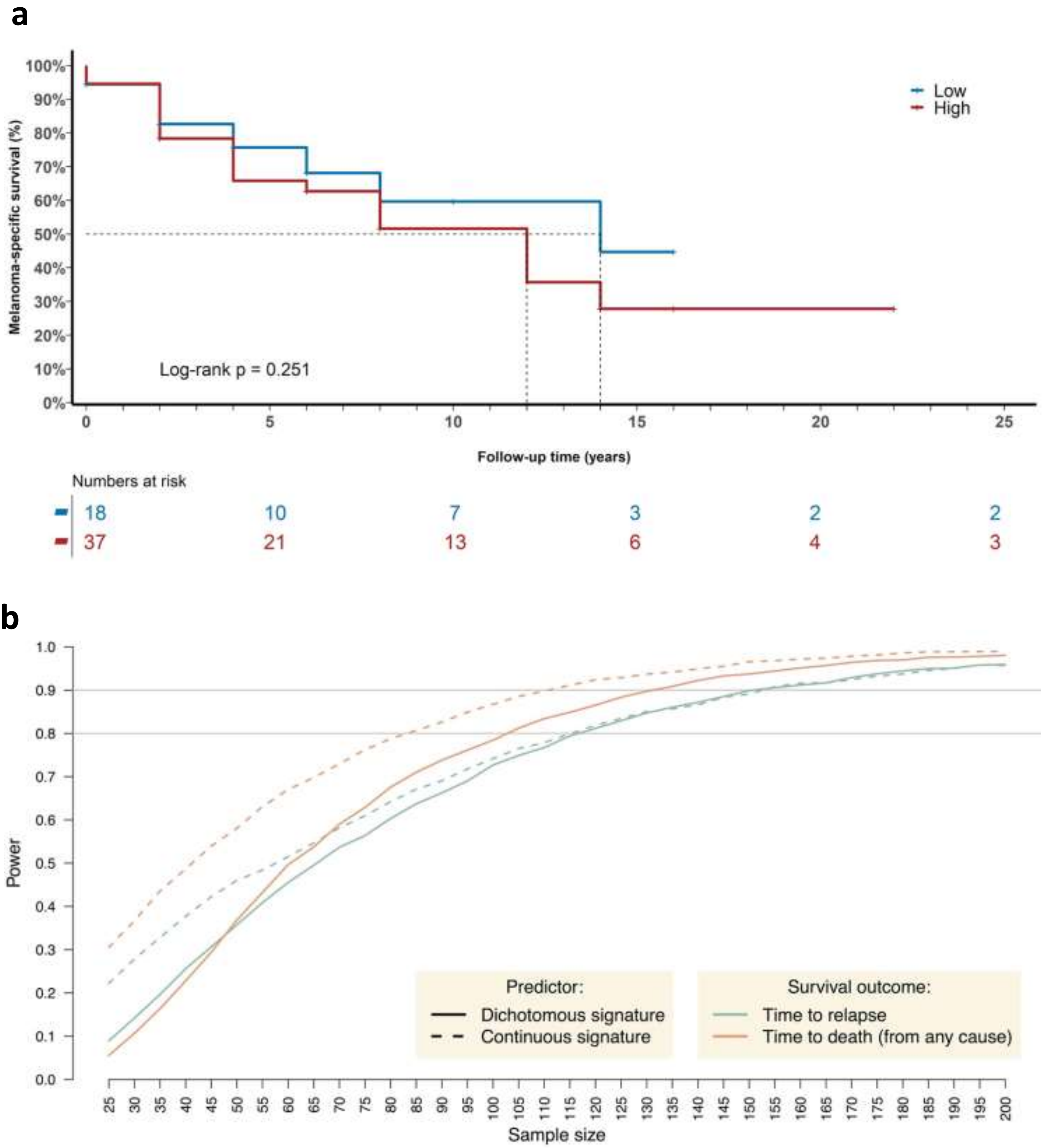
Supplementary Figure S1. Flow chart of the analysis. The figures corresponding to each specified analysis are indicated. OS: Overall survival; PFS: progression-free survival; MSS: melanoma-specific survival; LN: lymph node. ECOG: Eastern Cooperative Oncology Group Performance Status.



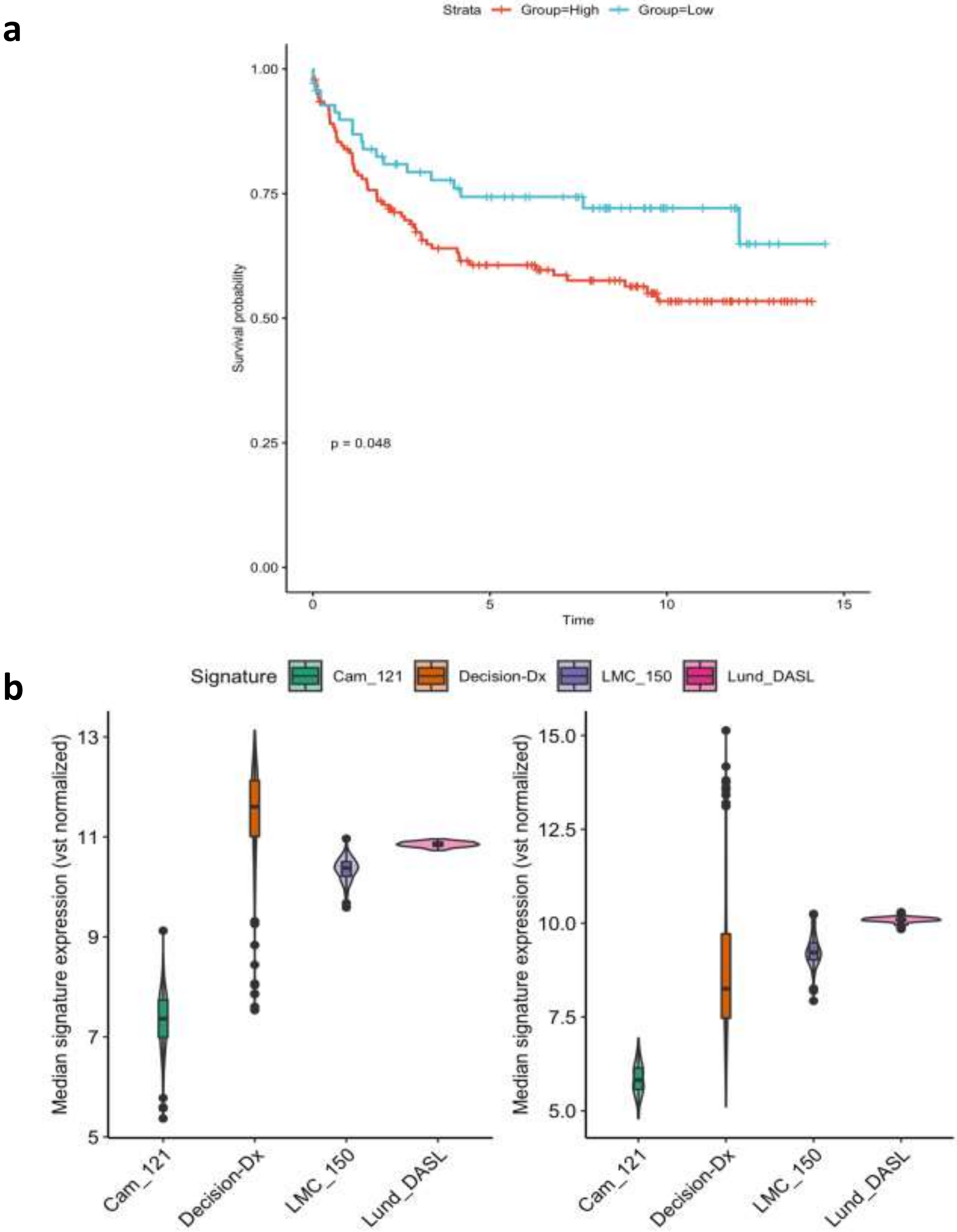
Supplementary Figure S2. PCA of the 1000 most variable genes, showing that samples separated by their tissue site of origin. a) Scree-plot showing the variance of the data (y-axis) explained by each of the 10 first principal components (x-axis). **b)** Biplot of the scores of the observations on the two first dimensions of a PCA analysis on the 1000 most variable genes. The samples are colour-coded according to their relapse status (yes in red, no in blue) and shaped according to their tissue of origin (primary melanoma samples in triangles, lymph nodes in circles). The x-axis of the biplot, accounting 40.3% of the total variance, shows that samples are clustered primarily based on the tissue site of origin. Source data are provided as a Source Data file.



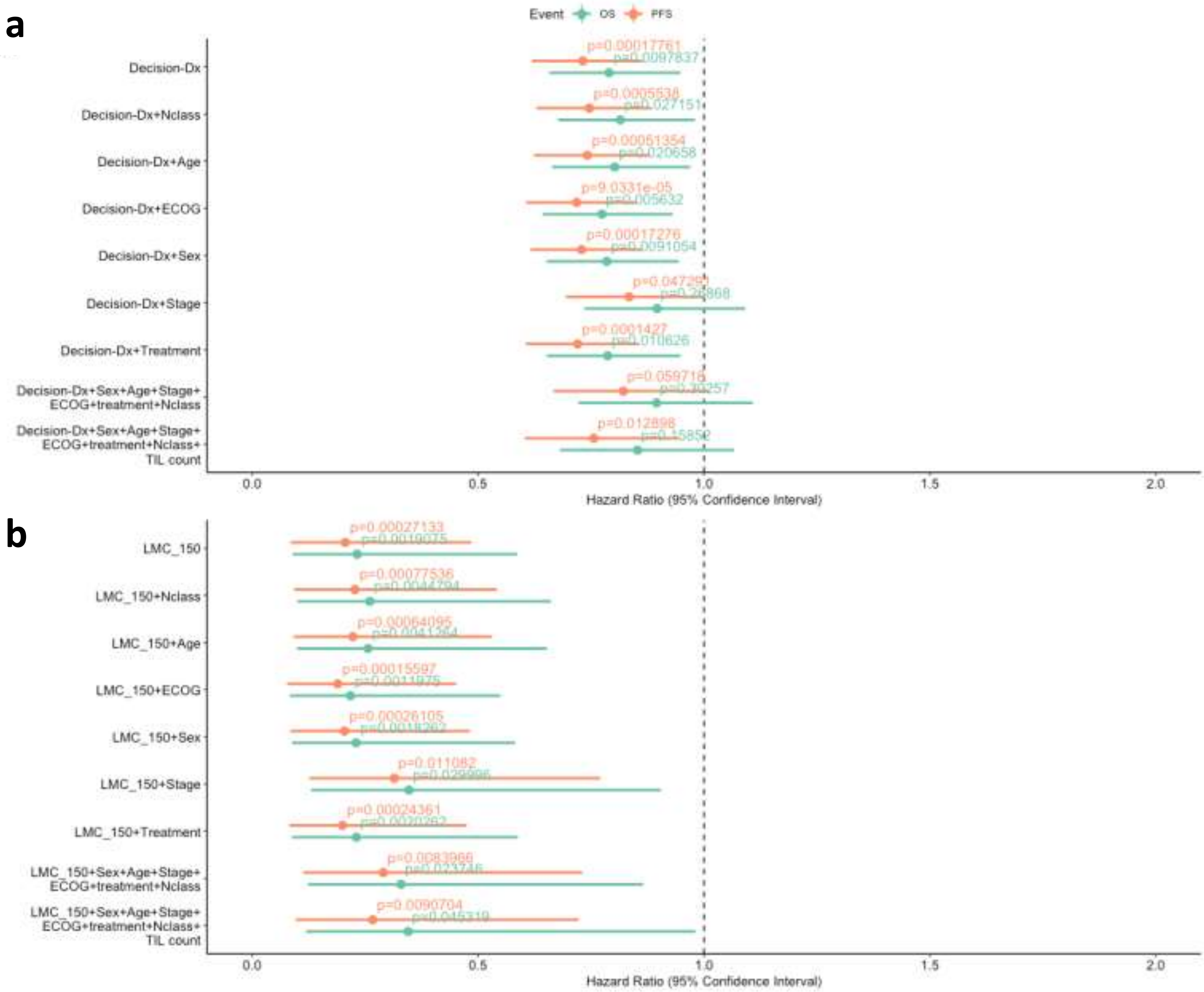
Supplementary Figure S3. Covariate-corrected differential expression analysis for primary melanoma samples (comparing metastases vs non-metastases). a) Schematic flowchart indicating the number of samples considered in the differential expression analysis. **b)** Volcano plot showing, for each gene, the -log 10 FDR corrected p-value (y-axis) and the corresponding log-fold change estimate within the differential expression analysis (x-axis). The (predominant) downregulated (144/197, 73.1%) and upregulated (53/197, 26.9%) DEGs are respectively coloured in red and blue for genes with a p-value < 0.1 after Benjamini and Hochberg multiplicity correction (this p-value cut-off is represented by the dashed horizontal grey line). Source data are provided as a Source Data file.



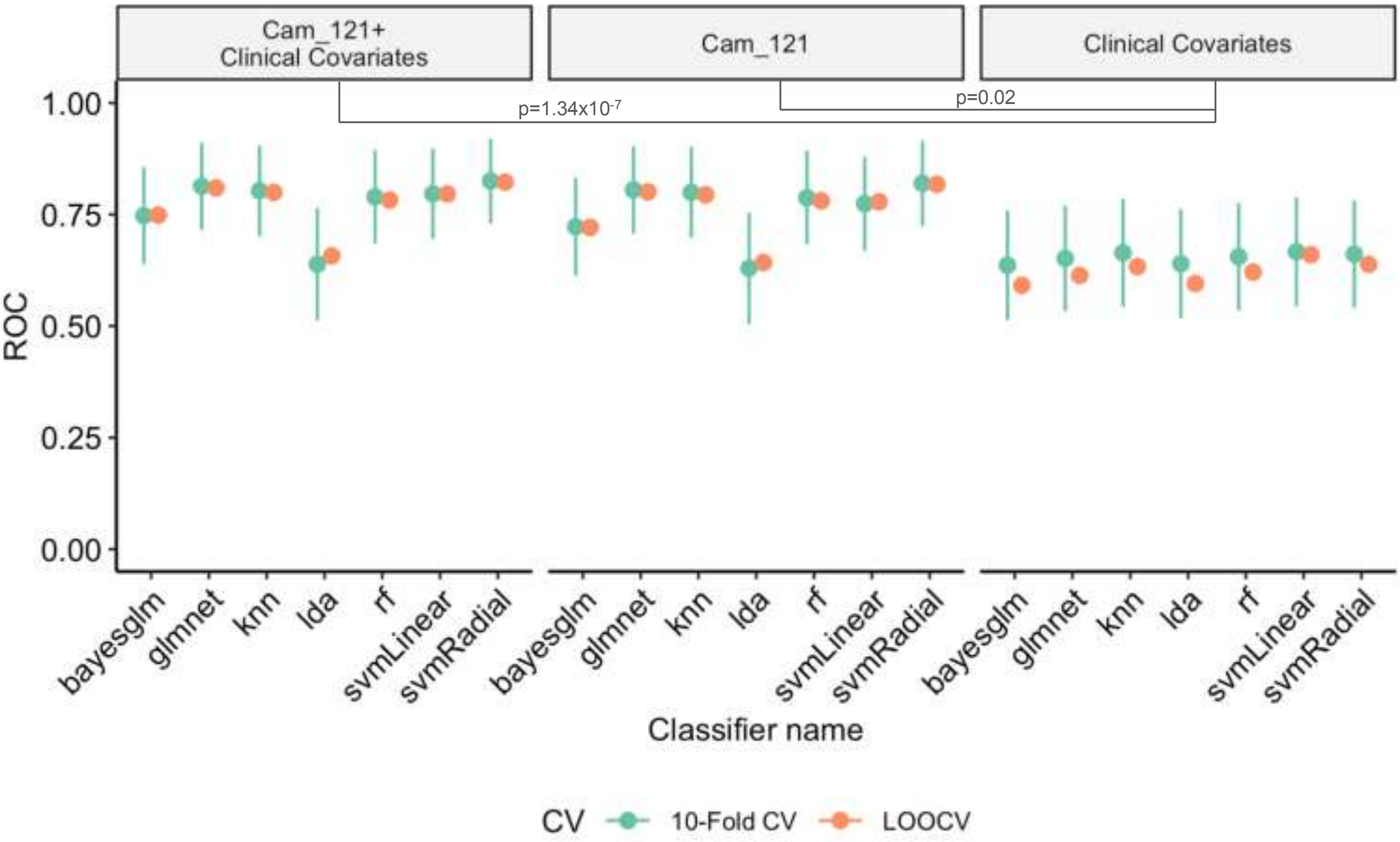
Supplementary Figure S4: Validation of Cam_121 within a third independently acquired external dataset (Australia Melanoma Genome Project, n=55). a) Kaplan-Meier curve comparing the melanoma specific survival probabilities (y-axis) of groups with high and low "Cam_121" (quantile 0.33 split) as a function of follow-up time in years (x-axis). Note that this dataset contains samples (n=55) from a variety of tissue sites (including primary tumours, regional lymph nodes, distant metastases, in-transit metastases and others) The p-value of a two-sided logrank test comparing the survival distributions of both groups is indicated. **b)** Simulation-based power analysis based on the AVAST-M study considering R=2500 Monte Carlo samples. The plot shows the power (y-axis) as a function of the sample size (x-axis) when modelling two survival outcomes, the time to relapse (green lines) and time to death from any cause (red lines), as a function of our signature, both considered as a dichotomous (solid line) and continuous (dashed line) predictor. A sample size greater than n=115 would be required to obtain a power greater than 80% in all cases.



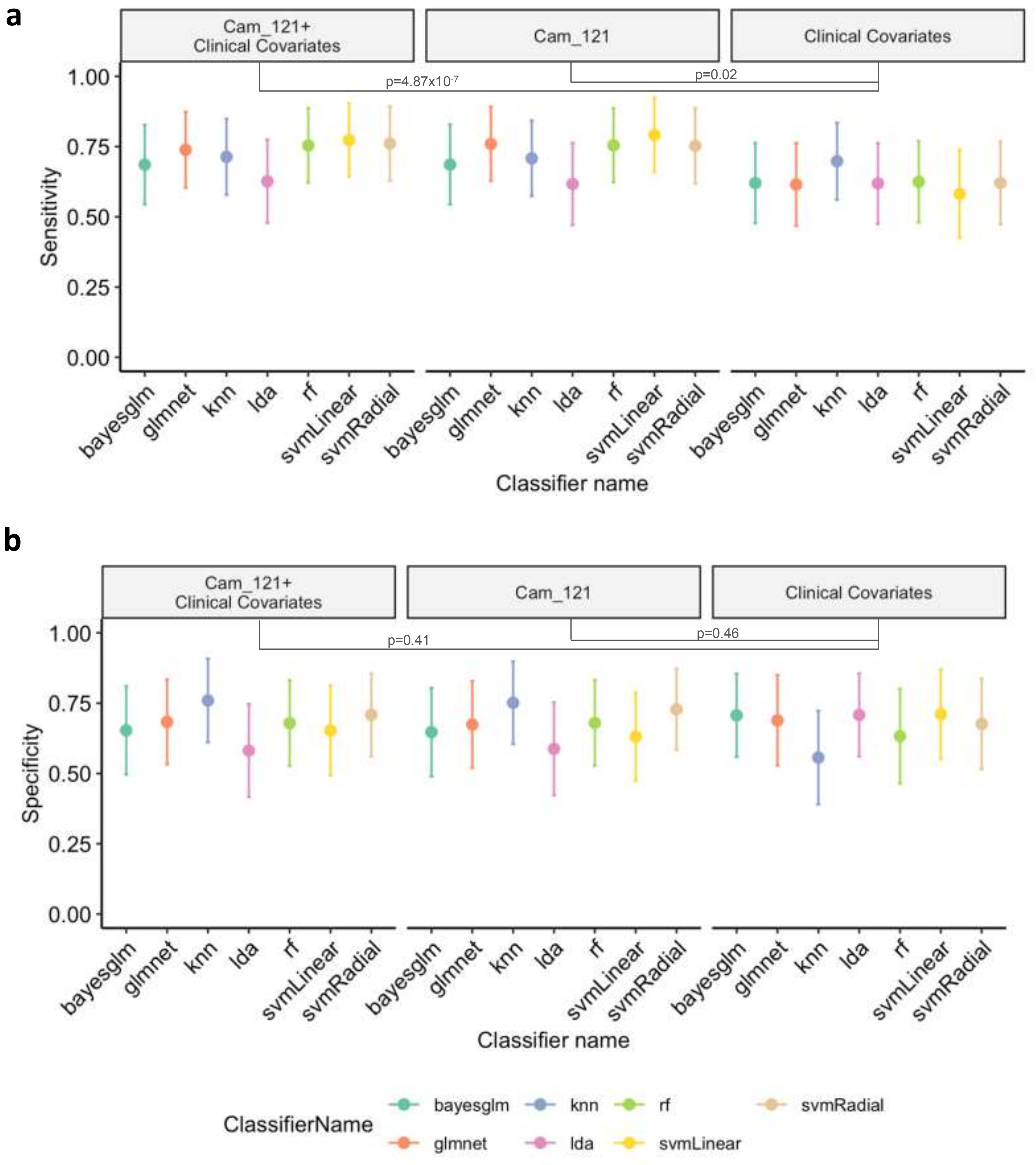
Supplementary Figure S5. Partial validation of Cam_121 within a fourth independently acquired external dataset (Lund Melanoma Cohort, n=223). a) Kaplan-Meier plot comparing the progression free survival probabilities (y-axis) of groups with high and low "Cam_121" (quantile 0.33 split) as a function of time in years (x-axis) within the Lund Melanoma Cohort (n=223) and p-value of a two-sided logrank test comparing the survival distributions. Note that only 24 genes out of 121 were identified within the Lund dataset. The Wald t-test p-values corresponding to the signature "Cam_121" is indicated. **b)** Violin plots showing the median gene expression values (normalized using variance stabilizing transformation (vst)) of the genes (y-axes) in each signature (x-axes). Note that for the signature "Lund_DASL" (pink), all the 7739 genes present in the Lund Melanoma Cohort were used for calculating the median expression value in each dataset. Left panel: median expression within the AVAST-M primary melanoma dataset (n=194), right panel: median expression within TCGA-SKCM Skin dataset (n=159). The median (vst normalized) expression values for each signature were plotted using the *geom_violin* and *geom_boxplot* function of the R-package ggplot2. The center of the box-plot denotes the median value; the bounds of the box denote the 25th percentile and the 75th percentile values; the dots denote the outliers lying ± 1.5 times beyond the interquartile range from the bounds of the box (shown by the minima and maxima of the box-plot).



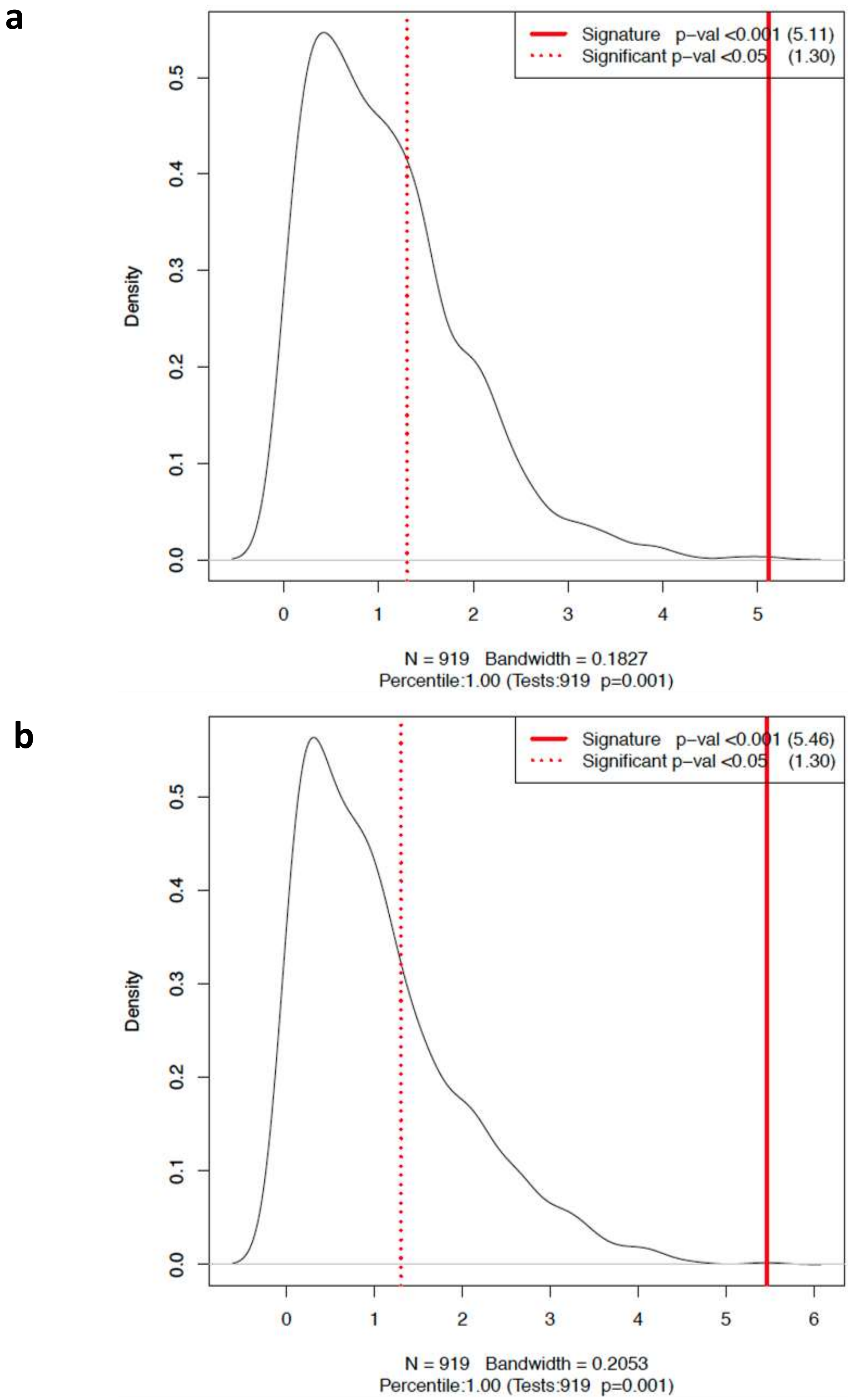
Supplementary Figure S6. Forest plot of univariate and multivariate survival analysis for the two previously published signatures (Gerami and LMC_150). Forest plot indicating the hazard ratio (HR) estimates (dots at the centre of error bars), corresponding to 95% confidence intervals of the HR estimates (horizontal error bars) and two-sided Wald t-test p-values related to the signature parameter when considering the signature definitions of **a)** Gerami and **b**) LMC_150 when predicting overall survival (green) and progression free survival (orange) by means of Cox proportional hazard models when controlling for different (sets of) clinical variables (y-axes). No multiplicity correction were used. ECOG: Eastern Cooperative Oncology Group Performance Status. TIL count; Tumour-infiltrating lymphocyte count.



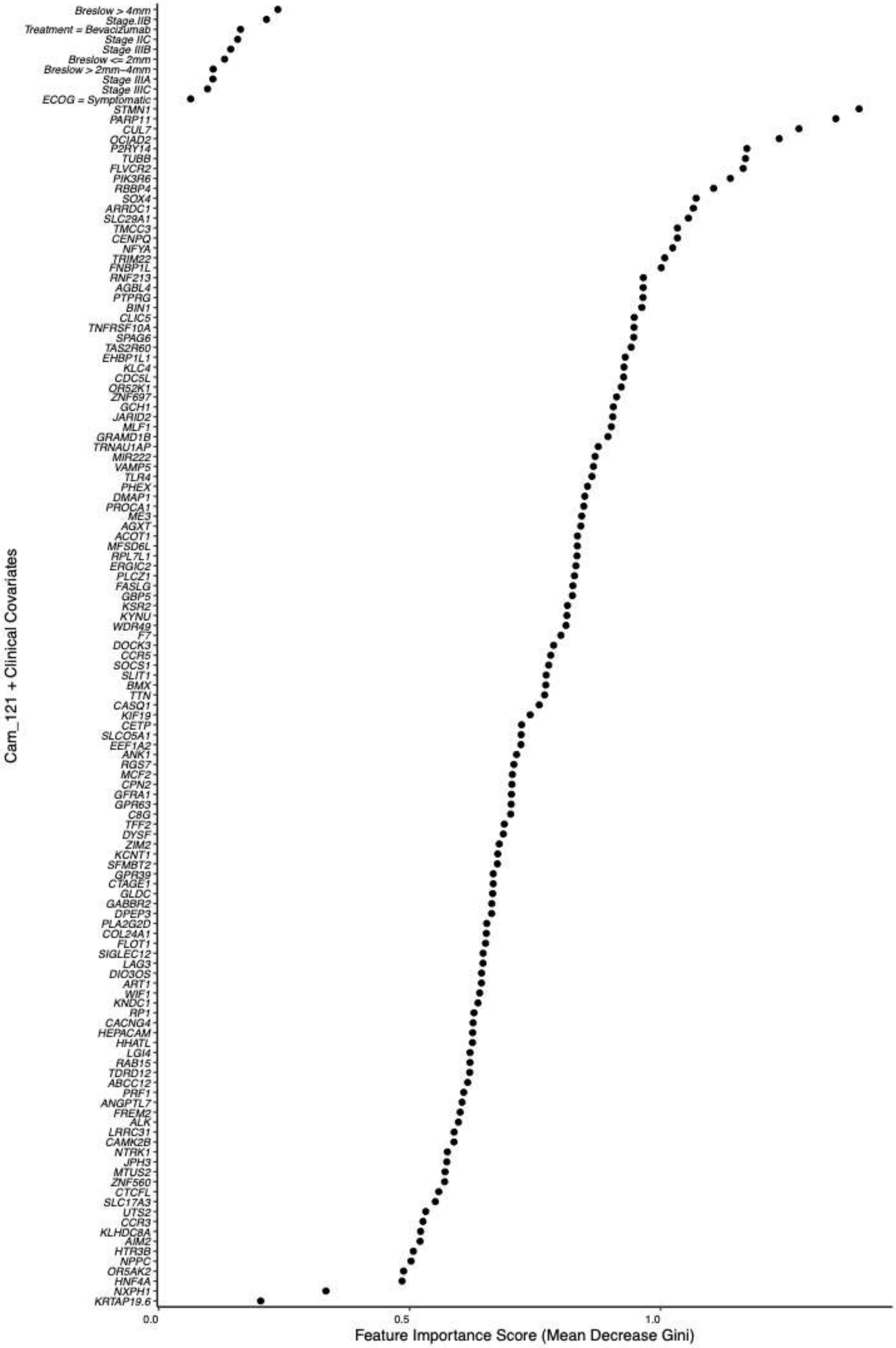
Supplementary Figure S7. Comparison of 10-Fold and Leave-one out cross validation within the machine learning model (training). Plot showing, for each set of predictors (panels) and 7 different classification models (x-axis), the mean ± sd AUROC obtained from 10-Fold cross validation in green (repeats = 1000) and the AUROC estimates from leave-one-out cross-validation in orange. The final classification remained consistent across both cross-validation approaches. The p-values of statistical comparisons using one-sided two-sample Welch t-tests are indicated (see also supplementary table S2a). See methods section 10 for details about the classification algorithms. Source data are provided as a Source Data file.



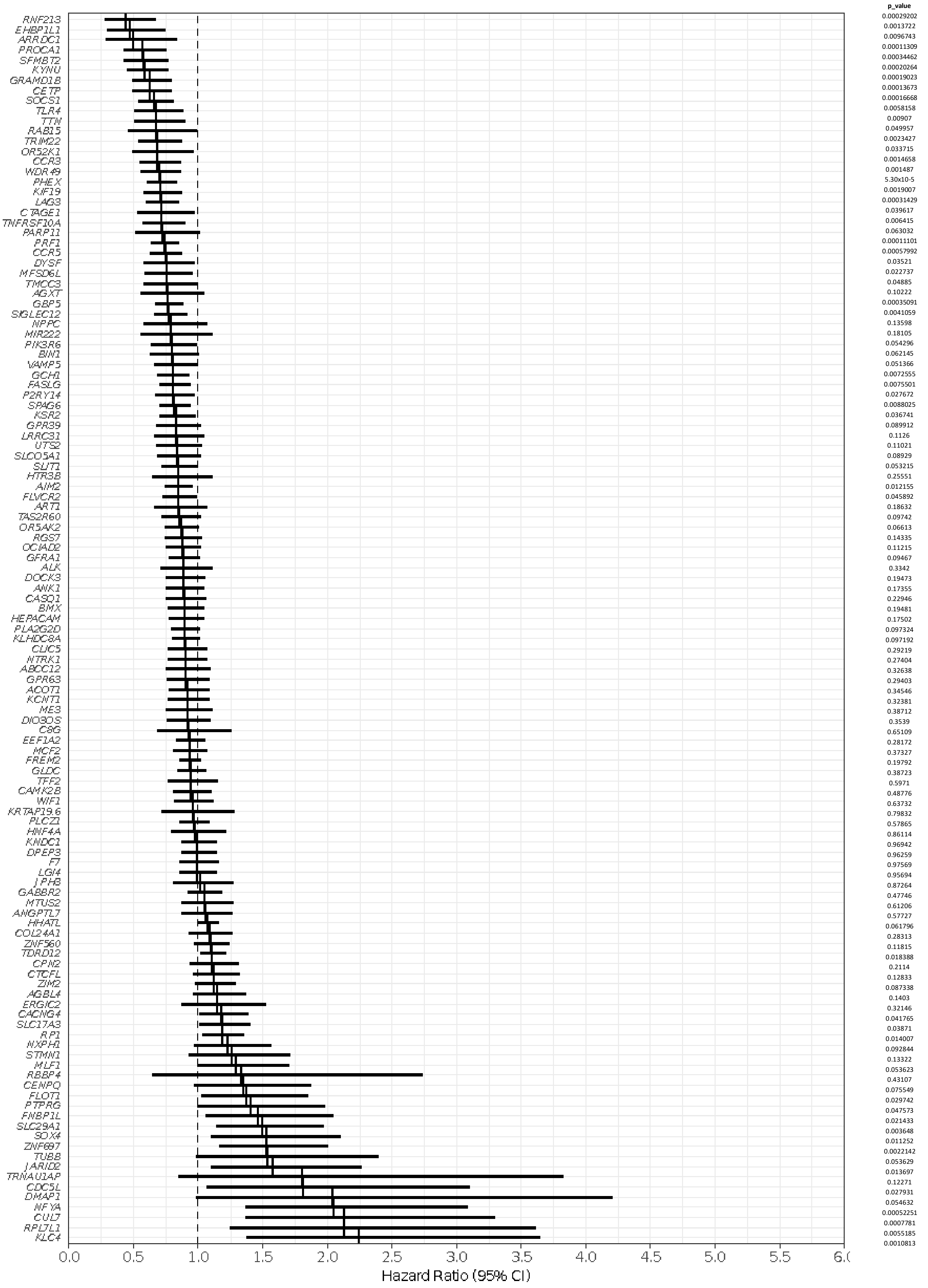
Supplementary Figure S8. Machine learning results across multiple classifiers, showing that Cam_121 outperforms clinical covariates using 10-Fold CV (repeats = 1000). Plots showing, for each set of predictors (columns), the mean ± sd scores of **a)** sensitivity **b)** and specificity (y-axes) of the 7 selected classifiers (x-axes). The p-values of statistical comparisons using one-sided two-sample Welch t-tests are indicated (see also supplementary table S2a). See methods section 10 for details about the classification algorithms. Source data are provided as a Source Data file.



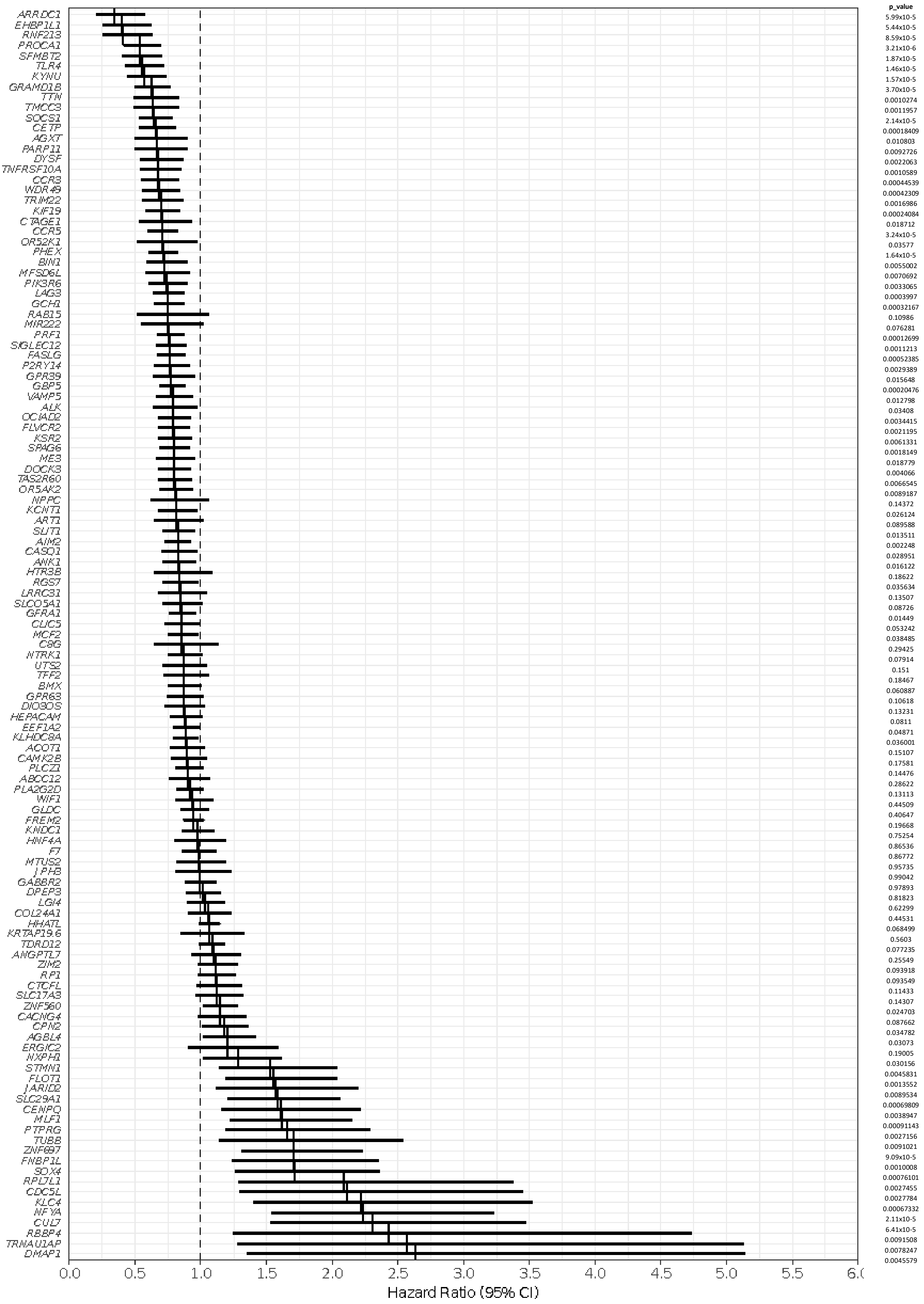
Supplementary Figure S9. The Cam_121 signature significantly outperforms 1000 sets of 121 randomly selected protein-coding genes. Plots showing the densities of the p-values corresponding to the Cam_121 gene signature when considering random signatures (smoothed densities) in terms of **a)** OS and **b)** PFS in multivariate Cox regression models. The one-sided p-values obtained with the Cam_121 signature appear as red vertical lines. The p-values are defined as the fraction of scores of random signatures which are greater than those observed using the (real) Cam_121 gene signature. The dotted red vertical lines correspond to -log10(0.05), i.e., to the 5% cutoff meaning that only 5% of the p-values (on the -log10 scale) corresponding to random signatures would be beyond this point if random signatures had no predictive effect. In **a)** and **b)** No multiplicity corrections were used to define the p-values indicated in the legend.



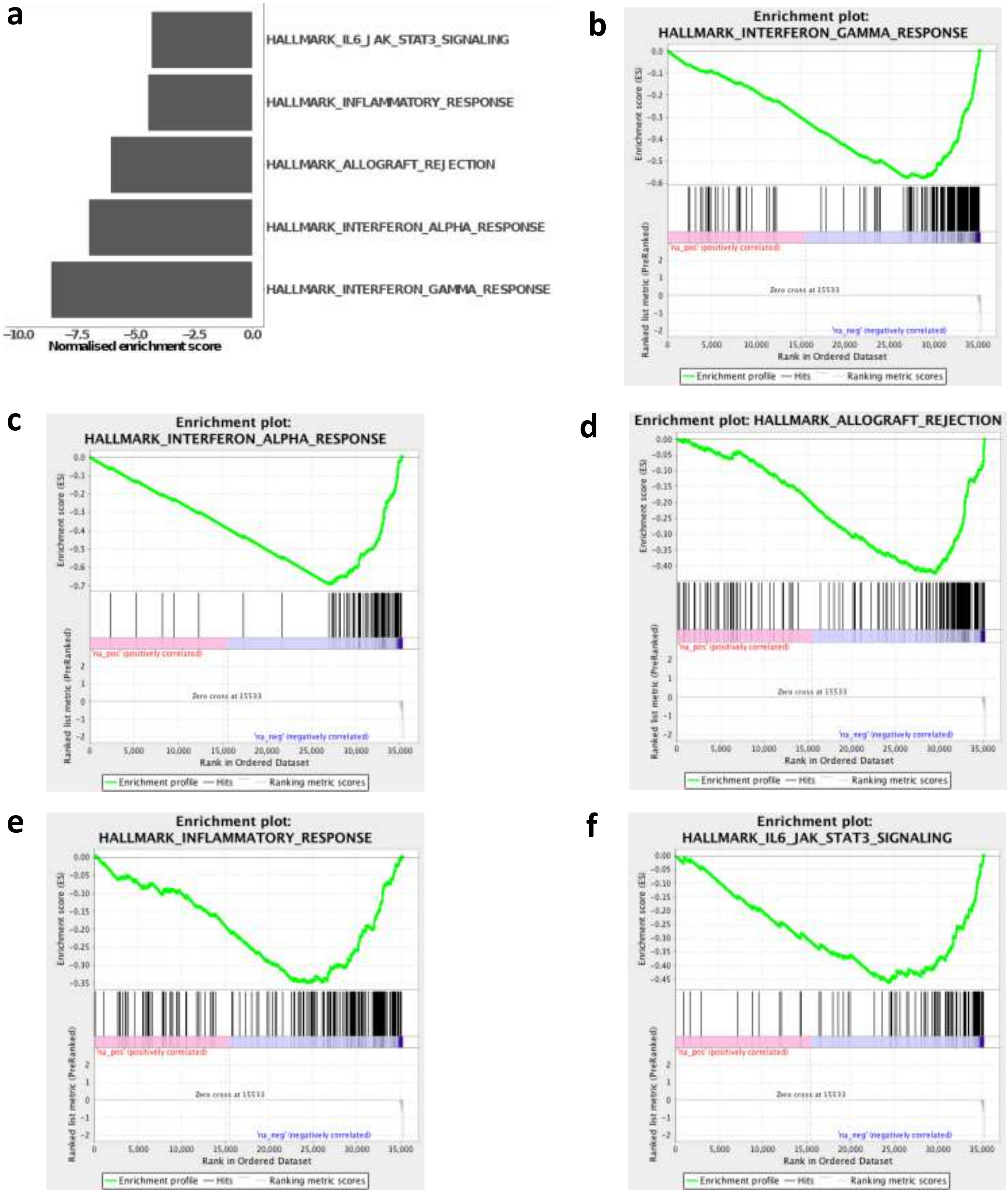
Supplementary Figure S10. Feature importance score (FIC) for each gene in Cam_121 showing similar scores across all genes within the random forest machine learning model suggesting that, rather than the dominance of any single gene, it is a composite analysis of all genes that contribute to the signatures' prognostic power. Genes are ranked in decreasing FIC order. The clinical covariates are indicated on the top for comparison. Source data are provided as a Source Data file.



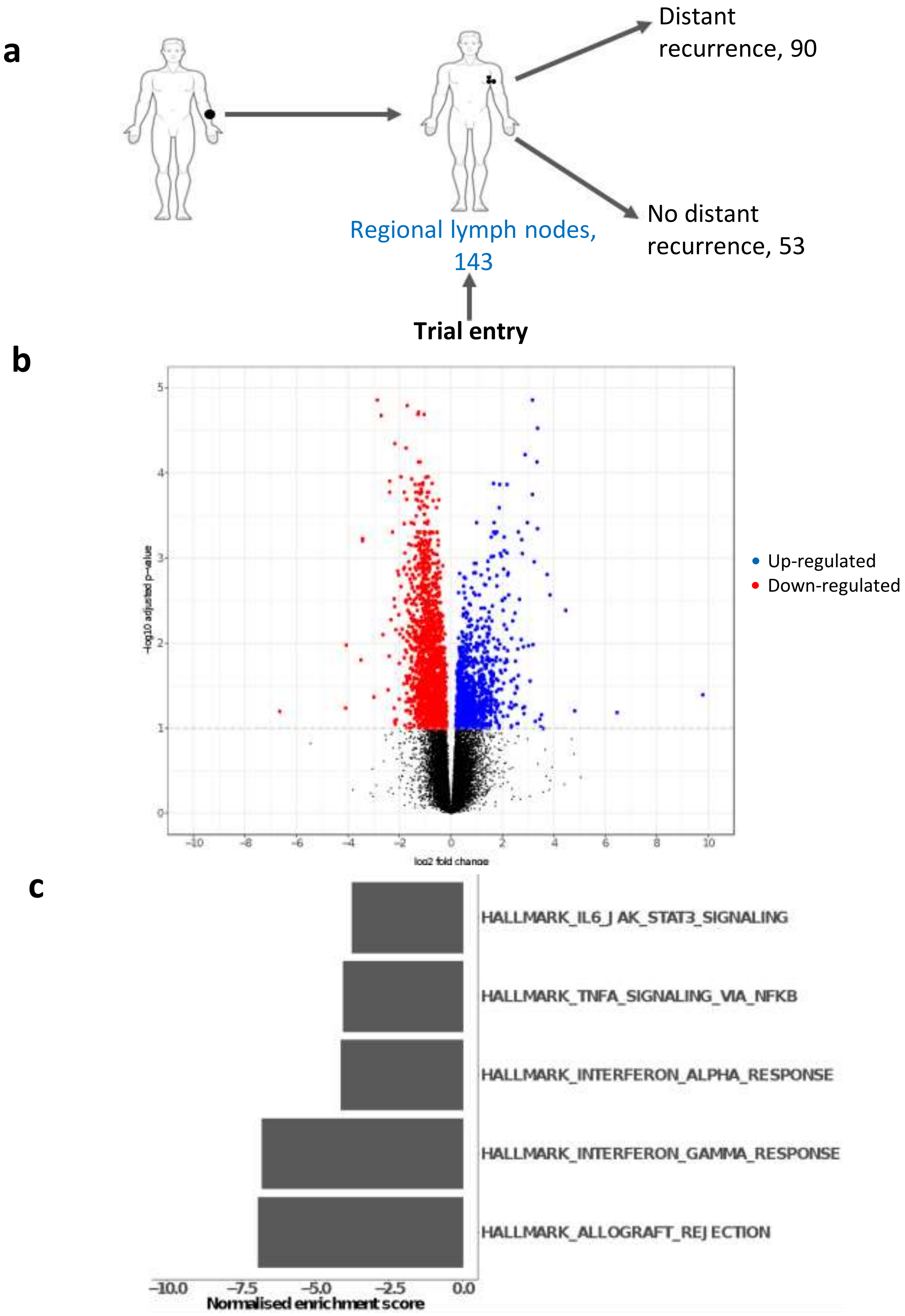
Supplementary Figure S11. Multivariate survival analysis (OS) for each Cam_121 signature gene. Forest plot indicating the hazard ratio (HR) estimates (vertical lines at the centre of error bars) and horizontal error bars corresponding to 95% confidence intervals of the HR estimates for each Cam_121 gene ranked in increasing order of HR. The corresponding two-sided Wald t-test p-values (before FDR correction) are indicated on the right. No single gene proved significant after correction for multiple testing (FDR corrected p-value<0.05). This therefore suggests that it is a composite analysis of all genes that contribute to the signatures' prognostic power rather than the dominance of any single gene. No multiplicity correction were used.



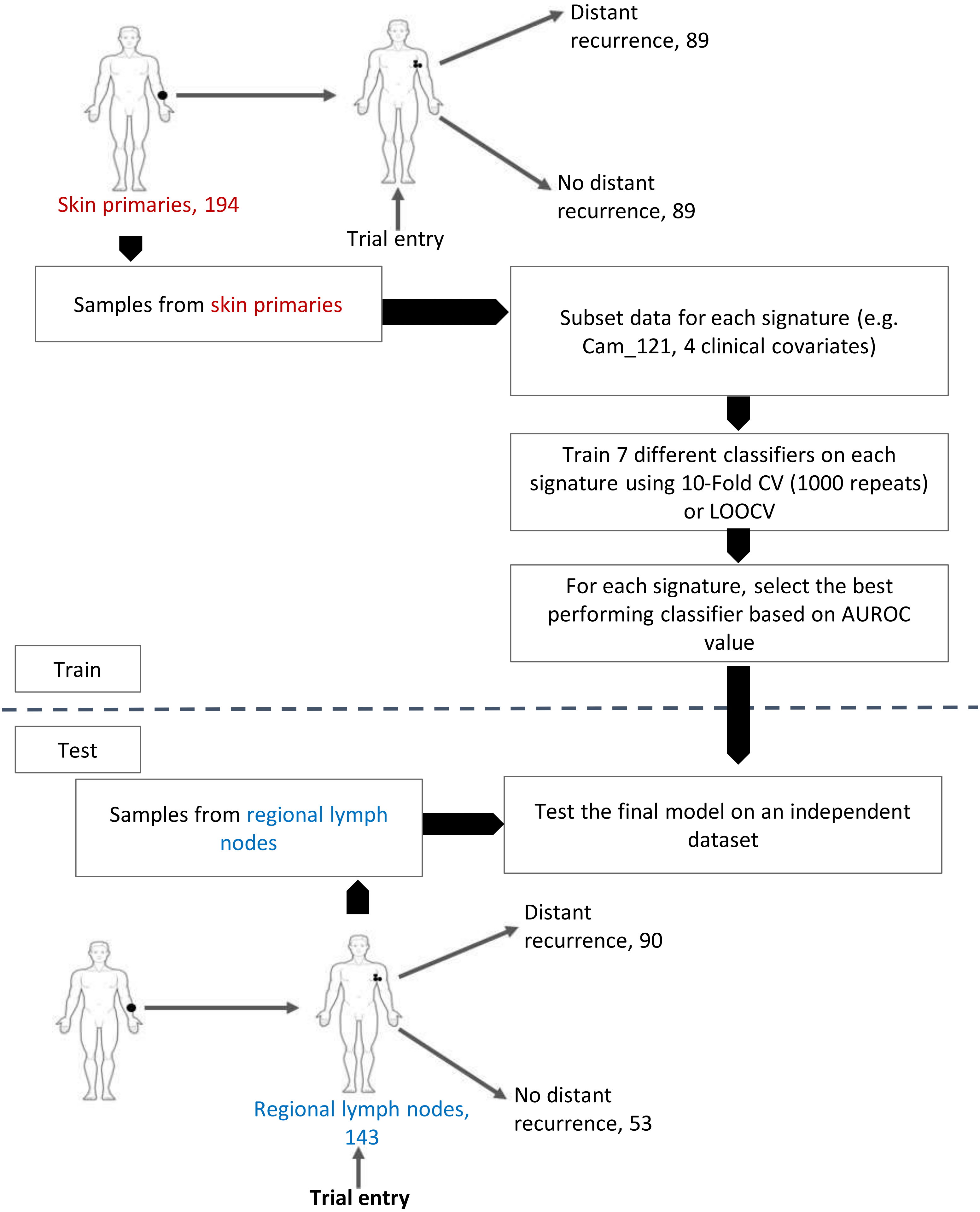
Supplementary Figure S12. Multivariate survival analysis (PFS) for each Cam_121 signature gene. Forest plot indicating the hazard ratio (HR) estimates (verrical lines at the centre of error bars) and horizontal error bars corresponding to 95% confidence intervals of the HR estimates for each Cam_121 gene ranked in increasing order of HR. The corresponding two-sided Wald t-test p-values (before FDR correction) are indicated on the right. No single gene proved significant after correction for multiple testing (FDR corrected p-value<0.05). This therefore suggests that it is a composite analysis of all genes that contribute to the signatures' prognostic power rather than the dominance of any single gene. No multiplicity correction were used.



Supplementary Figure S13. Results of gene set enrichment analysis showing downregulation of key immune-related pathways in primary melanoma samples destined for metastases. a) Barchart of the top five most significant down-regulated Hallmark gene sets ($p<0.01$ for all gene sets). **b-f)** The corresponding enrichment plots from (a). The top portion of the plot shows the running enrichment score for the gene set as the analysis walks down the ranked list. The middle portion of the plot shows where the members of the gene set appear in the ranked list of genes and the bottom portion of the plot shows the value of the ranking metric as one moves down the list of ranked genes. Source data are provided as a Source Data file.



Supplementary Figure S14. Results of differential expression and gene-set enrichment analysis comparing metastases vs non-metastases in lymph node samples (n=143), uncovering the same results as those obtained from primary melanoma samples. a) A schematic of the number of samples in the covariate corrected differential expression analysis. **b)** Volcano plot showing, for each gene, the -log 10 FDR corrected p-value (y-axis, FDR<0.1) and the corresponding log-fold change estimate within the differential expression analysis (x-axis). The (predominant) downregulated (1967/3022 (65.1%)) and up upregulated DEGs (1055/3022 (34.9%)) are respectively colour-coded in red and blue for genes showing a p-value < 0.1 after Benjamini and Hochberg multiplicity correction. **c)** Barplot showing downregulation of the same top five immune-related pathways as those identified in primary melanoma sample analyses (FDR corrected p-value<0.01 for all gene sets) (Supplementary Figure S13A). Source data are provided as a Source Data file.



Supplementary Figure S15. Schematic outline of the machine learning analysis. The AVAST-M primary melanoma dataset was used for training the machine learning and the AVAST-M lymph node dataset was used for testing the final model. Three different sets of features were considered for comparison: (i) considering the Cam_121 gene expression values only as features, (ii) one-hot encoded clinical covariates only as features, (iii) Cam_121 gene expression values as well as one-hot encoded clinical covariates as features. For each set of features, 7 different classifiers were trained using 10-Fold cross-validation (CV repeats = 1000) or leave-one-out cross validation (LOOCV) and the classifier giving the highest mean area under the ROC curve (AUROC) value was selected as the final model (see methods section 10).

Supplementary Table S1a: Analyses of the clinical covariates associated with metastases: relapsed (n=590) vs non-relapsed (n=753). Corrected for the length of follow-up and treatment. The likelihood ratio test p-values were calculated and corrected for multiple testing using Holm-Bonferroni method (see Methods section 8).

Clinical Covariate	p-value	Adjusted p-value
Stage	1.01x10-25	1.11x10-24
Nclass	1.07x10-21	1.07x10-20
ECOG	0.0036	0.03
Treatment	0.045	0.36
Age	0.076	0.53
BRAF	0.1	0.6
Sex	0.56	1
Site	0.79	1
Breslow thickness	0.22	1
Ulceration	0.68	1
NRAS	0.21	1

Supplementary Table S1b: Analyses of the clinical covariates associated with overall survival. The p-values from the partial-likelihood-ratio tests performed on the coxph model objects are reported.

Clinical Covariate	Pr(>|Chi|)
EventMet	4.90x10-157
Sex	4.61x10-3
Age	3.20x10-3
I(Age^2)	3.04x10-1
as.numeric(Nclass)	1.42x10-2
Stage	8.78x10-2
Site	7.70x10-1
Breslow thickness	6.43x10-1
Ulceration	8.40x10-1
Treatment	4.80x10-1
ECOG	2.14x10-2

Supplementary Table S1c: Analyses of the clinical covariates associated with progression-free survival. The p-values from the partial-likelihood-ratio tests performed on the coxph model objects are reported.

Clinical Covariate	Pr(>|Chi|)
EventMet	1.73x10-278
Sex	1.55x10-1
Age	1.03x10-1
I(Age^2)	6.61x10-1
as.numeric(Nclass)	5.81x10-3
Stage	1.76x10-3
Site	4.28x10-1
Breslow thickness	3.72x10-2
Ulceration	4.51x10-1
Treatment	8.68x10-1
ECOG	7.22x10-1

Supplementary Table S2a: Statistical analysis of mean classifier results shown in Figure 3a and Supplementary Figure S8. The mean performances of 7 classifiers for signatures were compared in a pairwise manner using one-sided two-sample Welch t-tests implemented in R by the function t.test() with parameters (var.equal = FALSE, paired = FALSE, alternative = "greater"). No multiplicity correction were used.

	ROC (p-value)		Sensitivity (p-value)		Specificity (p-value)	
Signature > Clinical covariates	p-value	z-statistic	p-value	z-statistic	p-value	z-statistic
Cam_121 + Clinical Covariates > Clinical Covariates	1.34E-07	8.05E+00	4.87E-07	6.5796	0.4079	0.2353
Cam_121 > Clinical Covariates	0.0023	4.2749	0.0018	3.8075	0.4634	0.0937

Supplementary Table S2b: Statistical analysis of ROC curves shown in Figure 3b. The ROC curves were compared using the roc.test() function in R with parameters (method="delong", paired=FALSE, alternative = "greater"). The one-sided DeLong's p-values and the test statistic for each comparison are reported. No multiplicity correction were used.

Signature > Clinical covariates	DeLong's p-value	DeLong's Z-statistic
Cam_121 + Clinical Covariates > Clinical Covariates	0.0136	2.2098
Cam_121 > Clinical Covariates	0.0202	2.0502

Supplementary Table S2c: Statistical analysis of mean classifier results shown in Figure 3e (bootstrap). The mean performances of 7 classifiers for signatures were compared in a pairwise manner using one-sided two-sample Welch t-tests implemented in R by the function t.test() with parameters (var.equal = FALSE, paired = FALSE, alternative = "greater"). No multiplicity correction were used.

	ROC		Sensitivity		Specificity	
Signature1 > Signature2	p-value	z-statistic	p-value	z-statistic	p-value	z-statistic
Cam_121 + Clinical Covariates > Clinical Covariates	0.000387	4.07E+00	3.47E-04	3.8534	0.6684	-0.4411
Cam_121 > Clinical Covariates	0.000703	4.4371	0.0283	2.1533	0.3698	0.3419
LMC_150 > Clinical Covariates	0.109007	1.3209	0.2835	0.5928	0.2565	0.6755
Decision-Dx Melanoma > Clinical covariates	0.105969	1.3439	0.1579	1.0642	0.4710	0.0749
Cam_121 > LMC_150	0.015221	2.4532	0.1208	1.2322	0.5786	-0.2030
Cam_121 > Decision-Dx Melanoma	0.020351	2.2962	0.2679	0.6385	0.4215	0.2023

Supplementary Table S2d: Statistical analysis of mean classifier results shown in Figure 3e (10-Fold cross validation). The mean performances of 7 classifiers for signatures were compared in a pairwise manner using one-sided Welch Two Sample t-tests implemented in R by the function t.test() with parameters (var.equal = FALSE, paired = FALSE, alternative = "greater")

	ROC		Sensitivity		Specificity	
Signature1 > Signature2	p-value	z-statistic	p-value	z-statistic	p-value	z-statistic
Cam_121 + Clinical Covariates > Clinical Covariates	3.87E-04	4.07E+00	3.47E-04	3.85E+00	6.68E-01	-4.41E-01
Cam_121 > Clinical Covariates	5.43E-02	1.83E+00	1.34E-02	2.53E+00	8.08E-01	-9.22E-01
LMC_150 > Clinical Covariates	4.86E-01	3.73E-02	9.52E-02	1.39E+00	8.06E-01	-9.16E-01
Decision-Dx Melanoma > Clinical covariates	1.71E-01	9.95E-01	1.07E-01	1.31E+00	7.28E-01	-6.26E-01
Cam_121 > LMC_150	8.71E-02	1.45E+00	2.55E-01	6.83E-01	4.88E-01	3.13E-02
Cam_121 > Decision-Dx Melanoma	1.35E-01	1.18E+00	1.38E-01	1.14E+00	6.93E-01	-5.22E-01

Supplementary Table S3: Distribution of clinical covariates across the high versus low risk signature groups. Please note that the corresponding two-sided statistical test used to obtain the p-values is indicated in the last column of the table. No multiplicity correction were used.

Variable	Levels	Statistics	Overall	Signature - high	Signature - low	P-value	Statistical test
Full dataset		n	194	130	64		
Sex	Male	n (%)	123 (63.4%)	83 (63.8%)	40 (62.5%)	0.9804	Chi-square test
	Female	n (%)	71 (36.6%)	47 (36.2%)	24 (37.5%)		
	Missing	n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
Age		Median (IQR)	57.0 (44.0-66.0)	57.5 (44.0-66.0)	56.0 (43.5-65.2)	0.9375	Welch's test
Site	Head and neck	n (%)	39 (20.1%)	25 (19.2%)	14 (21.9%)	0.4741	Chi-square test
	Lower limbs	n (%)	55 (28.4%)	41 (31.5%)	14 (21.9%)		
	Trunk	n (%)	71 (36.6%)	47 (36.2%)	24 (37.5%)		
	Upper limbs	n (%)	29 (14.9%)	17 (13.1%)	12 (18.8%)		
	Missing	n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
Bres	<= 2.0 mm	n (%)	31 (16.0%)	22 (16.9%)	9 (14.1%)	0.7347	Chi-square test
	>2-4mm	n (%)	63 (32.5%)	40 (30.8%)	23 (35.9%)		
	>4.0mm	n (%)	100 (51.5%)	68 (52.3%)	32 (50.0%)		
	Missing	n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
Ulc	Present	n (%)	104 (53.6%)	66 (50.8%)	38 (59.4%)	0.5543	Chi-square test
	Absent	n (%)	83 (42.8%)	57 (43.8%)	26 (40.6%)		
	Missing	n (%)	7 (3.6%)	7 (5.4%)	0 (0.0%)		
Nclass	N0	n (%)	80 (41.2%)	44 (33.8%)	36 (56.2%)	0.0115	Fisher's test
	N1a	n (%)	32 (16.5%)	25 (19.2%)	7 (10.9%)		
	N1b	n (%)	17 (8.8%)	11 (8.5%)	6 (9.4%)		
	N2a	n (%)	19 (9.8%)	17 (13.1%)	2 (3.1%)		
	N2b	n (%)	5 (2.6%)	3 (2.3%)	2 (3.1%)		
	N2c	n (%)	13 (6.7%)	7 (5.4%)	6 (9.4%)		

Supplementary Table S3: Distribution of clinical covariates across the high versus low risk signature groups (continued)

Variable	Levels	Statistics	Overall	Signature - high	Signature - low	P-value	Statistical test
	N3	n (%)	18 (9.3%)	16 (12.3%)	2 (3.1%)		
	Missing	n (%)	10 (5.2%)	7 (5.4%)	3 (4.7%)		
Nclass_binary	Positive	n (%)	104 (53.6%)	79 (60.8%)	25 (39.1%)	0.0046	Chi-square test
	Negative	n (%)	80 (41.2%)	44 (33.8%)	36 (56.2%)		
	Missing	n (%)	10 (5.2%)	7 (5.4%)	3 (4.7%)		
Stage	IIB	n (%)	45 (23.2%)	21 (16.2%)	24 (37.5%)	0.0145	Fisher's test
	IIC	n (%)	45 (23.2%)	30 (23.1%)	15 (23.4%)		
	IIIA	n (%)	29 (14.9%)	23 (17.7%)	6 (9.4%)		
	IIIB	n (%)	51 (26.3%)	37 (28.5%)	14 (21.9%)		
	IIIC	n (%)	24 (12.4%)	19 (14.6%)	5 (7.8%)		
	Missing	n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
Stage_binary	II	n (%)	90 (46.4%)	51 (39.2%)	39 (60.9%)	0.007	Chi-square test
	III	n (%)	104 (53.6%)	79 (60.8%)	25 (39.1%)		
	Missing	n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
BRAF	V600E	n (%)	75 (38.7%)	50 (38.5%)	25 (39.1%)	1	Chi-square test
	WT	n (%)	104 (53.6%)	69 (53.1%)	35 (54.7%)		
	Missing	n (%)	15 (7.7%)	11 (8.5%)	4 (6.2%)		
NRAS	Mutant	n (%)	35 (18.0%)	23 (17.7%)	12 (18.8%)	1	Chi-square test
	WT	n (%)	63 (32.5%)	40 (30.8%)	23 (35.9%)		
	Missing	n (%)	96 (49.5%)	67 (51.5%)	29 (45.3%)		
BRAF_NRAS_WT	Yes	n (%)	70 (36.1%)	47 (36.2%)	23 (35.9%)	1	Chi-square test
	No	n (%)	109 (56.2%)	72 (55.4%)	37 (57.8%)		
	Missing	n (%)	15 (7.7%)	11 (8.5%)	4 (6.2%)		

Supplementary Table S3: Distribution of clinical covariates across the high versus low risk signature groups (continued)

Variable	Levels	Statistics	Overall	Signature - high	Signature - low	P-value	Statistical test
treatment	Bevacizumab	n (%)	100 (51.5%)	68 (52.3%)	32 (50.0%)	0.8811	Chi-square test
	Observation	n (%)	94 (48.5%)	62 (47.7%)	32 (50.0%)		
	Missing	n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
	Missing	n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
ECOG	Symptomatic	n (%)	16 (8.2%)	13 (10.0%)	3 (4.7%)	0.2725	Fisher's test
	Asymptomatic	n (%)	178 (91.8%)	117 (90.0%)	61 (95.3%)		
	Missing	n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		
EventMet	Yes	n (%)	89 (45.9%)	71 (54.6%)	18 (28.1%)	0.0009	Chi-square test
	No	n (%)	105 (54.1%)	59 (45.4%)	46 (71.9%)		
	Missing	n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)		