



HAL
open science

MelanoDB: A dataset of clinical and molecular features of patients with advanced melanoma treated with MAPK inhibitors

Sarah Dandou, Julie A Vendrell, Jérôme Solassol, Baptiste Louveau, Céleste Lebbé, Samia Mourah, Florian Rambow, Eric Richard, Stanislas Du Manoir, Alain Mangé, et al.

► To cite this version:

Sarah Dandou, Julie A Vendrell, Jérôme Solassol, Baptiste Louveau, Céleste Lebbé, et al.. MelanoDB: A dataset of clinical and molecular features of patients with advanced melanoma treated with MAPK inhibitors. *Scientific Data*, 2025, 12 (1), pp.1144. <10.1038/s41597-025-05291-3>. <hal-05148978>

HAL Id: hal-05148978

<https://hal.science/hal-05148978v1>

Submitted on 7 Jul 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License



OPEN

DATA DESCRIPTOR

MelanoDB: A dataset of clinical and molecular features of patients with advanced melanoma treated with MAPK inhibitors

Sarah Dandou^{1,2}, Julie A. Vendrell³, Jérôme Solassol^{1,3}, Baptiste Louveau^{4,5}, Céleste Lebbé⁶, Samia Mourah^{4,5}, Florian Rambow^{7,8}, Eric Richard¹, Stanislas Du Manoir¹, Alain Mangé¹, Peter J. Coopman^{1,9}, Ovidiu Radulescu^{1,10} & Romain M. Larive^{1,10}✉

MAPK inhibitors (MAPKi) have revolutionized the treatment of patients with advanced melanoma. However, primary and acquired resistance mechanisms limit their efficacy. Predicting MAPKi response from the tumor baseline features remains challenging due to the limited size of patient cohorts. Therefore, we collected data from nine different patient cohorts (total $n = 417$ patients with advanced melanoma treated with MAPKi) to identify clinical and molecular features. Our curated dataset, named MelanoDB, includes whole or partial exome sequencing data for 191 patients, copy number alteration information for 66 patients, and gene expression data for 132 patients. We provide a web application to explore the integrated dataset and data distribution across the collected studies, and we share this dataset with the scientific community according to the Findable, Accessible, Interoperable, Reusable (FAIR) principles.

Background & Summary

Cutaneous melanoma represents only 10% of all skin cancer types, but advanced forms are responsible for 90% of deaths¹. Although only 2% to 5% of all detected cancers are cutaneous melanomas, it is a public health issue, particularly among white populations, due to the sharp rise of its incidence. Melanoma is resistant to conventional chemotherapy drugs, but the management of advanced melanoma has considerably progressed in the last decade following the development of effective systemic therapies, such as immune checkpoint inhibitors (ICI) and MAPK inhibitors (MAPKi). The discovery in the 2000s of valine 600 mutations in the protein kinase BRAF (40–50% of cutaneous melanomas) led to the development of BRAF inhibitors (BRAFi) that are currently used in combination with inhibitors of MEK (MEKi), another protein kinase of the MAPK pathway². However, in one third of patients, the tumor continues to grow after treatment start (primary resistance) and the 5-year recurrence-free survival rate is only 19% (acquired resistance)³. In parallel, immunotherapy with ICIs was first introduced to treat advanced non-BRAF-mutated melanoma. As ICIs show a more durable response than MAPKi, they are now supplanting MAPKi as first-line treatment of advanced BRAFV600-mutated melanoma⁴. However, the initial response rate to ICIs is inferior than to MAPKi. Therefore, to support therapeutic decision making, robust prediction of the tumor response to MAPKi is crucial. This requires the identification of

¹IRCM, Université de Montpellier, ICM, Inserm, Montpellier, France. ²LPHI, Université de Montpellier, CNRS, Montpellier, France. ³Laboratoire de Biologie des Tumeurs Solides, Département de Pathologie et Oncobiologie, Centre Hospitalier Universitaire (CHU) Montpellier, Université de Montpellier, Montpellier, France. ⁴Department of Pharmacology and Tumor Genomics, Hôpital Saint Louis, Assistance Publique-Hôpitaux de Paris, F-75010, Paris, France. ⁵Université Paris Cité, INSERM UMR-S 976, Team 1, Human Immunology Pathophysiology & Immunotherapy (HIPI), F-75010, Paris, France. ⁶Université Paris Cité, AP-HP Dermato-oncology and CIC, Cancer institute AP-HP Nord Paris Cité, INSERM U976, Saint Louis Hospital, Paris, France. ⁷Department of Applied Computational Cancer Research, Institute for AI in Medicine (IKIM), University Hospital Essen, Essen, Germany. ⁸University Duisburg-Essen, Essen, Germany. ⁹CNRS-Centre National de la Recherche Scientifique, 1919 Route de Mende, F-34293, Montpellier, France. ¹⁰These authors contributed equally: Ovidiu Radulescu, Romain M. Larive. ✉e-mail: romain.larive@umontpellier.fr

Reference	Number of patients	Number of patients with available molecular features	Sample number	Data source (Identifier)
Blateau <i>et al.</i> ⁹	53	SNV: 53	Pre-treatment: 53	Personal communication ¹⁶
Catalanotti <i>et al.</i> ⁵	66	SNV: 66 CNA: 66	Pre-treatment: 66	cBioPortal web browser ³⁸ ; cBioPortal data ²⁹⁻³¹
Hugo <i>et al.</i> ¹²	44	GEx: 36	Pre-treatment: 14 On treatment: 12 Progression: 10	Clinical data: extracted from the article; GEO GSE6186 ³²
Kwong <i>et al.</i> ¹³	22	GEx: 44	Pre-treatment: 12 Progression: 32	Clinical data: extracted from the article; EGA EGA0000100092 ³³
Long <i>et al.</i> ¹⁴	9	GEx: 23	Pre-treatment: 9 Progression: 9 Post-treatment: 5	Clinical data: extracted from the article; GEO GSE61992 ³⁴
Louveau <i>et al.</i> ⁸	24	SNV: 24 GEx: 24	Pre-treatment: 24	Personal communication ¹⁷
Rizos <i>et al.</i> ¹⁵	24	GEx: 51	Pre-treatment: 21 Progression: 24 Post-treatment: 6	Clinical data: extracted from the article; GEO GSE50509 ³⁵
Van Allen <i>et al.</i> ⁵	45	SNV: 78	Pre-treatment: 39 Progression: 39	cBioPortal web browser ³⁶ cBioPortal data ²⁹⁻³¹
Yan <i>et al.</i> ⁷	130	GEx: 72	Pre-treatment: 72	Clinical data: extracted from the article GEX: extracted from the article
Total	417	SNV: 191 CNA: 66 GEx: 132	Pre-treatment: 310 On treatment: 12 Progression: 66 Post-treatment: 11	MelanoDB ¹⁸

Table 1. Summary of the patient cohorts and data sources. CNA, Copy Number Alteration; GEx, Gene Expression; SNV, Single Nucleotide Variant.

biomarkers that can be measured before treatment initiation. Moreover, the comparison of the tumor molecular profile before treatment start and after recurrence contributes to deciphering the mechanisms of acquired resistance mechanisms. In clinical practice, the systematic molecular profiling of tumors has allowed the acquisition of a large amount of data that can be exploited for these purposes.

However, despite all these efforts, the integration, standardization, harmonization and structuring of clinical metadata with multi-omic data is far from being the norm in this field. The skin cutaneous melanoma cohort of The Cancer Genome Atlas (TCGA_SKCM) includes a substantial number of patients ($n = 470$) with genomic, epigenomic, transcriptomic and proteomic data from treatment-naive samples. Unfortunately, the information on treatments is not enough detailed (i.e. number of patients who received pharmaceutical therapy and number of patients who received radiation therapy) and does not allow selecting patients who received MAPKi. Moreover, most previous studies failed to identify molecular biomarkers for MAPKi response prediction due to the small size of their patient cohorts⁵⁻⁹. To date, there is no dataset that includes and combines clinical and molecular data from patients with advanced cutaneous melanoma treated with MAPKi for the simultaneous analysis of several harmonized cohorts. This lack also prevents the use of machine learning methods because they require a larger number of patients and rigorously annotated data.

Therefore, according to the Findable, Accessible, Interoperable and Reusable (FAIR) principles, we propose MelanoDB, a curated dataset that assembles clinical and molecular features of patients with advanced melanoma treated with MAPKi. We collected these data from the literature and from clinical data repositories, and formatted them to gather as many common characteristics as possible for all patients. Statistical methods and data-driven methods require a sufficient amount of data to detect specific features in complex systems. With MelanoDB, we provide a key resource to identify relationships between molecular alterations and clinical-pathological parameters in patients with melanoma. The first objective was to add value to the clinical data described in different published studies by combining them, thereby increasing the size of the combined dataset. The second objective was to provide a clinical and molecular dataset in order to identify a molecular signature associated with resistance to MAPKi and to understand the mechanisms underlying MAPKi effects. As an illustrative example, we leveraged part of our MelanoDB dataset (4 studies, 184 patients) to demonstrate the predictive potential of baseline genomic data, and the value of integrating them in the development of predictive models for the response duration to MAPKi in the context of advanced melanoma¹⁰. To date, we collected clinical data on a total of 417 patients with advanced melanoma treated with MAPKi (5 additional studies, making a total of 9 studies), including whole or partial exome sequencing data for 191 patients, copy number alterations for 66 patients and gene expression information for 132 patients (Table 1). We harmonized the datasets, tested the homogeneity of the clinical and molecular data, and when needed, applied solutions to correct batch effects (Fig. 1). All scripts needed to generate the MelanoDB dataset and the used data sources are open source. The generated dataset is user-friendly, via a visual web application that allows users to navigate the integrated data and their distribution among the collected studies¹¹.

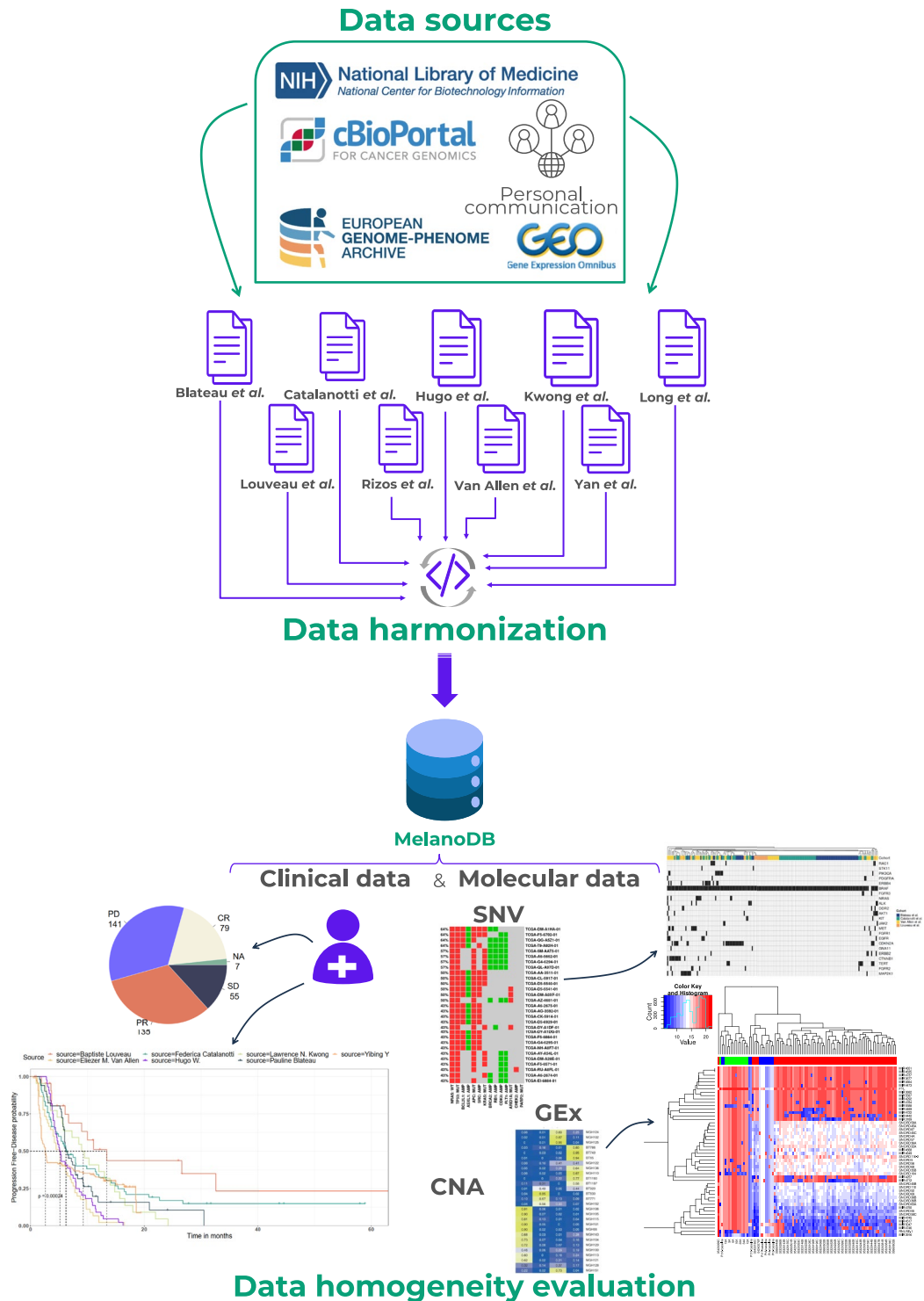


Fig. 1 Schematic overview of the MelanoDB dataset development. SNV, Single Nucleotide Variant; CNA, Copy Number Alteration; GEX, Gene Expression.

Methods

Patients, samples and clinical characteristics. We generated our multi-study dataset by aggregating clinical and molecular data from patients with BRAF^{V600}-mutated stage III or IV melanoma treated with MAPKi as mono-therapy (BRAFi) or bi-therapy (BRAFi and MEKi). Our data collection was reviewed by the Research Ethics Committee of the Montpellier University (UM 2022-009bis). Each center was responsible for handling the ethical, legal, and privacy issues of the relevant data.

Genes	
AKT1	MAP2K1
ALK	MET
BRAF	NRAS
CDKN2A	PDGFRA
CTNNB1	PIK3CA
DDR2	POLE
EGFR	PTEN
ERBB2	RAC1
ERBB4	SMAD4
FGFR1	STK11
FGFR2	TERT
FGFR3	
GNA11	
GNAQ	
GNAS	
H3F3A	
H3F3B	
HIST1H3B	
HRAS	
IDH1	
IDH2	
JAK2	
KIT	
KRAS	

Table 2. List of the selected genes for mutational analysis (next generation sequencing) in Blateau *et al.*⁹.

- Blateau *et al.*⁹: From 2014 to 2019, 53 patients underwent DNA sequencing at the University Hospital of Montpellier (Montpellier, France). Patients eligible for this retrospective study were diagnosed either for primary or recurrent metastatic melanoma. The noninterventional study was conducted in accordance with local ethical guidelines and was reviewed by the Ethical Committee from the Montpellier University Hospital (March 2019). Samples were obtained following research project approval by the Institutional Review Board from CRB-CHUM (BB-0033-00031). An approved informed consent statement for participation and data sharing was acquired for all patients.
- Catalanotti *et al.*⁶: Authors analysed pre-treatment tumors from 66 patients of Memorial Sloan Kettering Cancer Center (New York, NY, USA) with stage IV or unresectable stage III $BRAF^{V600E/K}$ melanoma.
- Hugo *et al.*¹²: Authors analysed patient-matched normal tissues, pre-treatment tumors and under-treatment disease- progressive tumors from patients with stage IV or stage IIIc $BRAF^{V600}$ melanoma. Patients were enrolled at University of California, Los Angeles; University of California, San Francisco; and Vanderbilt-Ingram Cancer Center.
- Kwong *et al.*¹³: 27 patients with metastatic stage IV melanoma and positively genotyped for a BRAFV600 mutation were enrolled in clinical trials.
- Long *et al.*¹⁴: Patients included in this study had $BRAF^{V600E/K}$ -mutant metastatic melanoma, had not received prior immunotherapy and were treated with dabrafenib plus trametinib as part of phase I/II or III clinical trials.
- Louveau *et al.*⁸: 27 patients initiating combined treatment of BRAFi + MEKi as first- or following-line therapy and followed at Saint Louis Hospital (Paris, France) were included in this retrospective study. All patients were enrolled in MelBase, a multicenter French clinical database with a biobank dedicated to the prospective follow-up of advanced melanoma patients since March 2013. MelBase was approved by the French ethics committee (CPP Ile-de-france XI, n°12027, 2012) and registered in the NIH clinical trials database (NCT02828202). Written informed consent was obtained from all patients for their participation, including collection, analysis and sharing of their data.
- Rizos *et al.*¹⁵: Patients included in this study had $BRAF^{V600E/K/R}$ -mutant metastatic melanoma, had not received prior MAPK inhibitors or immunotherapy, and were treated with either dabrafenib (150 mg twice daily or total daily dose of at least 300 mg daily) or vemurafenib (960 mg twice daily) as part of a clinical trial or an access program.
- Van Allen *et al.*⁵: Biopsy samples were secured from 45 patients who received vemurafenib or dabrafenib monotherapies to obtain research biopsies before and after resistance developed to systemic therapy at the University Hospital of Essen (Essen, Germany).
- Yan *et al.*⁷: Data from patients treated with cobimetinib combined with vemurafenib or vemurafenib in various clinical trials were included in the analysis.

Genes		
ABL1	GNAQ	RAC1
AKT1	GRIN2A	RAF1
AKT2	GRM3	RASA2
AKT3	HOXD8	RET
ALK	HRAS	RPS27
ARAF	IDH1	SERPINB3
BRAF	IGF1R	SNX31
BTK	JAK2	STAT3
CCND1	KDR	STK11
CDK4	KIT	STK19
CDKN2A	KRAS	TACC1
CD274	MAP2K1	TERT
CTNNB1	MAP2K2	TRRAP
CXCR4	MDM2	WT1
DDR1	MET	
DDR2	MITF	
DDX3X	MRPS31	
DUSP4	NOTCH1	
EGFR	NOTCH2	
ERBB2	NRAS	
ERBB3	PDGFRA	
ERBB4	PDGFRB	
EZH2	PHLPP1	
FBXW7	PIK3CA	
FERMT3	PIK3CG	
FGFR1	PIK3R1	
FGFR2	PIK3R2	
FGFR3	PLCG2	
FLT3	PPP6C	
GNA11	PTPN11	

Table 3. List of the selected genes for mutational analysis (next generation sequencing) in Louveau *et al.*⁸.

Data harmonization. We extracted source data from the literature or from repositories, when available. As data repositories are not generalized among publishers, some studies presented their results without data availability. Therefore, we contacted the authors of Blateau *et al.*⁹ and Louveau *et al.*⁸ to retrieve the data of interest^{16,17}. We did not modify the source files, and we harmonized the data during the dataset generation process. To facilitate multi-omics analysis and the merging of the different sets, we assigned a unique identifier (patientID) to the data of each patient from the different cohorts. As the original studies were carried out for different purposes, the types of data collected (SNV, CNA, GEx) and the time of sampling were different among studies (Table 1).

Data Records

All datasets files were formatted in the MelanoDB dataset¹⁸, a data collection on patients with advanced melanoma treated with MAPKi.

patients.csv table: Currently, clinical data include information from 417 patients. When available in the original dataset, the following features (specified by the column headings) are provided for each patient.

- Sex: sex of the patient (male, female).
- Age: age of the patient.
- AJCC_stage (American Joint Committee on Cancer): indicator associated with cancer progression; it describes the tumor location, size, how far it has grown in the surrounding tissues and whether it has spread to nearby lymph nodes or to other distant body parts¹⁹. It is determined by a clinician through microscope analysis of biopsies including tumor lesions and some surrounding healthy tissue. Staging is usually done using the Clark levels (e.g. level 1: melanoma confined to the epidermis; Level 2: melanoma has invaded the papillary dermis)²⁰. Moreover, the AJCC staging system uses three key pieces of information for the Tumor-Node-Metastasis (TNM) classification. Using these data, a patient is assigned to a stage group (from 0 to IV); stage IV describes a melanoma that has spread to distant sites of the body (e.g. soft tissues, distant lymph nodes, lung, liver, brain, bone).
- M_stage: cancer stage indicator. It describes the melanoma spreading to distant sites in the body (e.g. M0: no distant metastasis; M1: distant metastasis). It is further evaluated based on the location of the distant

Genes bait v3								
ABL1	CDC73	EPHA3	FOXL2	JAK3	MTOR	PIK3CD	SHQ1	YAP1
ABL2	CDH1	EPHA4	FUBP1	JUN	MYB	PIK3CG	SMAD4	YES1
AKT1	CDH11	EPHA5	GATA1	KCNJ5	MYC	PIK3R1	SMARCA4	ZRSR2
AKT2	CDK12	EPHA6	GATA2	KDM5C	MYCL1	PIK3R2	SMARCB1	
AKT3	CDK4	EPHA7	GATA3	KDM6A	MYCN	PIK3R3	SMO	
ALK	CDK6	EPHA8	GLI1	KDR	MYD88	PKM2	SOCS1	
ALOX12B	CDK8	EPHB1	GLI3	KEAP1	NCOA2	PLK2	SOX2	
APC	CDKN1A	EPHB2	GNA11	KIT	NF1	PNRC1	SPOP	
AR	CDKN2A	EPHB3	GNAQ	KLF6	NF2	PPP2R1A	SRC	
ARAF	CDKN2B	EPHB4	GNAS	KRAS	NFE2L2	PRDM1	SRSF2	
ARHGAP26	CDKN2C	EPHB6	GOLPH3	LDHA	NFKB1	PREX2	STAG2	
ARID1A	CEBPA	ERBB2	GRIN2A	LGR6	NFKB2	PRKAA2	STK11	
ARID2	CHEK1	ERBB3	GRM3	LMO1	NKX2-1	PRKAR1A	SUFU	
ASXL1	CHEK2	ERBB4	GSK3B	MAGI2	NOTCH1	PRKCI	TBK1	
ATM	CIC	ERG	HDAC2	MAP2K1	NOTCH2	PTCH1	TEK	
ATRX	CREBBP	ESR1	HIF1A	MAP2K2	NOTCH3	PTEN	TERT	
AURKA	CRKL	ETV1	HMGA2	MAP2K4	NOTCH4	PTPN11	TET1	
BAP1	CRLF2	ETV6	HNF1A	MAP3K8	NPM1	PTPRD	TET2	
BCL2L1	CSF1R	EZH2	HRAS	MCL1	NRAS	PTPRS	TGFBR2	
BCL2L11	CTNNB1	FAM123B	HSP90AA1	MDM2	NTRK1	PTPRT	TMPRSS2	
BCL6	CYLD	FAM46C	IDH1	MDM4	NTRK2	RAF1	TNFAIP3	
BCOR	DAXX	FAS	IDH2	MED12	NTRK3	RARA	TNFRSF14	
BIRC2	DDR2	FAT1	IGF1R	MEF2B	PAK7	RB1	TOP1	
BRAF	DICER1	FBXO11	IGFBP7	MEN1	PALB2	REL	TP53	
BRCA1	DIS3	FBXW7	IKBKE	MET	PARK2	RET	TP63	
BRCA2	DNMT1	FGFR1	IKZF1	MITF	PARP1	RICTOR	TSC1	
BUB1B	DNMT3A	FGFR2	IL7R	MLH1	PAX5	RNF43	TSC2	
CARD11	DNMT3B	FGFR3	INPP4A	MLL	PBRM1	ROR2	TSHR	
CBL	E2F3	FGFR4	INPP4B	MLL2	PDGFRA	ROS1	U2AF1	
CBLB	EGFR	FH	INSR	MLL3	PDGFRB	RPTOR	VHL	
CCND1	EIF4EBP1	FLCN	IRS1	MLST8	PHOX2B	RUNX1	WAS	
CCNE1	EP300	FLT1	IRS2	MPL	PIK3C2G	SDHB	WNK1	
CD79B	EPHA10	FLT3	JAK1	MSH2	PIK3CA	SETD2	WT1	
CDC42EP2	EPHA2	FLT4	JAK2	MSH6	PIK3CB	SF3B1	XPO1	

Table 4. List of the selected genes for mutational analysis (next generation sequencing) in Catalanotti *et al.*⁶ bait v3.

metastasis (e.g. M1a: the cancer has spread to other parts of the skin, soft-tissue sites or distant lymph nodes; M1b: the cancer has spread to the lungs).

- LDH: Lactate dehydrogenase concentration. A biological marker normally present in the blood and other body tissues. High LDH level may indicate tissue damage or the presence of cancer cells.
- PFS (Progression-free survival): PFS has been defined as the time from MAPKi start to the date of progression on MAPKi or the date of death. If no progression was observed, patients were censored at the date of the last follow-up.
- PFS_status: it refers to the event occurrence (i.e. cancer progression on treatment). This feature is set to 1 when the event occurs. For all patients from the studies by Hugo *et al.*¹², and Kwong *et al.*¹³, PFS_status was set to 1 (observation of progression) because all patients with melanoma included in these studies experienced disease progression.

For all patients in the study by Van Allen *et al.*⁵, PFS_status was set to 1 (observation of progression) because the authors focused only on patients with melanoma resistant to MAPKi (early and late resistance).

For all patients in the study by Yan *et al.*⁷, the original variable (PFS censor) was inverted (i.e. values [0,1] changed to [1,0]) in line with the other studies because the 0 value initially referred to the appearance of resistance (all patients with progressive disease patients initially had a 0 status).

- OS (Overall survival): OS was measured as the time from the MAPKi start to the date of death; patients alive at the last follow-up were censored at the date of the last documented contact.
- OS_status: it refers to the occurrence of the OS event (i.e. the patient's death). The feature is set to 1 when the event occurs.

For the patients from the study by Yan *et al.*⁷, the original reference variable (OS censor) was inverted (i.e. values [0,1] changed to [1,0]) in line with the other studies, because the 0 value initially referred to death.

bait v4	bait v5
CBLC	ARID1B
MAP3K1	ATR
SMAD2	CBLC
SMAD3	CCND2
	CCND3
	CTCF
	ERCC2
	ERCC5
	FAT4
	FKBP1A
	HLA-A
	IRF4
	MAP3K1
	MAPK1
	MRE11A
	NBN
	PMS2
	POLE
	PPP6C
	RAC1
	RAD50
	RPS6KB1
	SMAD2
	SMAD3
	SYK

Table 5. List of the selected genes for mutational analysis (next generation sequencing) in Catalanotti *et al.*⁶ bait v4 and v5.

- drug: the kinase inhibitor(s) used to slow and/or stop tumor progression (dabrafenib, trametinib, vemurafenib, cobime- tinib).
- BOR (Best overall response): BOR was recorded from the study treatment start to disease progression/recurrence. Patients who died or were lost to follow-up before the first assessment or who had incomplete imaging were graded as not evaluable (NE).
For the patients from the studies by Hugo *et al.*¹², and Kwong *et al.*¹³, the BORs were interpreted according to the criteria used to determine the objective tumor response of target lesions²¹ because only information on lesion changes (e.g. disappearance, sum of diameter increase/decrease, insufficient shrinkage or increase) was recorded.
- BRAF_mut (BRAF mutation): it refers to the detected BRAF mutation (V600E, V600K). As activating BRAF mutations are present in nearly 50% of malignant melanomas and contribute to the constitutive activation of the MAPK pathway, their presence is systematically investigated to choose a therapeutic strategy that may or may not include targeted treatments (MAPKi).
- brain_metastasis: it indicates the presence of metastasis in the patient's brain (no/yes).
- immunotherapy_treatment: it indicates whether the patient received immunotherapy concomitantly or after MAPKi treatment (no/yes).
- pre_MAPKi_treatment: it indicates whether the patient received cancer therapy before inclusion in the study (no/yes). The pre-treatment type also is specified (e.g. chemotherapy, interleukin-2, interferon).
- CNA_data: it indicates the presence of information on copy number alterations (CNA) for a given patient.
- SNV_data: it indicates the presence of information on single nucleotide variants (SNV) for a given patient in the dataset. As Catalanotti *et al.*⁶ used three different gene lists for their cohort (baits), the list used for each patient is specified.
- GEx_data: it indicates the presence of gene expression data for this patient in the dataset.

cnas.csv table: The CNA records include 66 pre-treatment samples from the Catalanotti *et al.*⁶ set. Each sample ID is associated with the initial patientID. Each record uses the Hugo Gene Nomenclature Committee (HGNC) symbol and the value indicates a DNA gain/amplification or loss/deletion compared with the reference genome. The following fields detail the column headings.HGNC: Hugo Gene Nomenclature Committee symbol associated with a sequenced gene.

- Value: CNA value.
- Temporality: time of sample collection relative to the study treatment (pre-treatment; on treatment; progression; post- treatment).

Genes
BRAF
RAF1
ARAF
PDGFRB
IGF1R
MET
HGF
KIT
EGFR
ERBB2
MAP3K8
MKI67
E2F2
RB1
CDK2
CDK4
CDK6
CCND1
RRM2
BCL2
BCL2L1
BCL2L11
BMF
MCL1
BAD
PTEN
CDKN1A
CDKN1B
CDKN2A

Table 6. List of the selected genes for mRNA expression analysis by quantitative PCR in Louveau *et al.*⁸.

snvs.csv table: The dataset also include 50,331 SNV records from Bateau *et al.*⁹, Catalanotti *et al.*⁶, Louveau *et al.*⁸ and Van Allen *et al.*⁵. For patients in the studies by Bateau *et al.*⁹, Catalanotti *et al.*⁶, and Louveau *et al.*⁸, only partial sequencing was carried out. The lists of targeted genes for each study are indicated in Tables 2–5. For patients in the study by Van Allen *et al.*⁵, whole exome sequencing was performed. When available in the original source, the following features (specified by the column headings) are provided for each sample.

- HGNC: Hugo Gene Nomenclature Committee symbol associated with the sequenced gene.
- HGVS_p: this code refers to a change in DNA sequence (variant) according to the HGVS nomenclature standards (e.g. c.4375 C > T).
- HGVS_p_short: HGSV code shortened, the three-letter codes for amino acids are replaced by their single-letter codes.
- Uniprot_id: Uniprot identifier corresponding to the sequence.
- Consequence: predicted consequence of the DNA sequence change on the resulting mRNA transcript or protein.
- Variant_classification: translational effect of the variant allele.
- Variant_type: type of mutation. Double nucleotide polymorphism (DNP), triple nucleotide polymorphism (TNP) and oligo-nucleotide polymorphism (ONP) describe the alteration of two, three and four (or more) consecutive nucleotides, respectively.
- Chromosome: number of the chromosome, without the “chr” prefix, that contains the gene.
- Start_position: lowest numeric position of the reported variant in the reference genome sequence. Mutation start coordinate (1-based coordinate system).
- End_position: highest numeric genomic position of the reported variant in the reference genome sequence. Mutation end coordinate (inclusive, 1-based coordinate system).
- Strand: genomic strand of the reported allele. Variants are reported on the positive genomic strand.
- Ref_allele: the plus strand reference allele at this position. It includes “-” for an insertion.
- Tumor_allele_1: primary genotype data. Tumor sequencing (discovery) allele 1. If similar to the Ref_allele, then no mutation. “-” for a deletion represents the variant. A novel inserted sequence does not include the flanking reference bases.

Feature	Values	Total (%)	Bleateau <i>et al.</i> ⁹	Catalanotti <i>et al.</i> ⁶	Hugo <i>et al.</i> ¹²	Kwong <i>et al.</i> ¹³	Louveau <i>et al.</i> ⁸	Long <i>et al.</i> ¹⁴	Rizos <i>et al.</i> ¹⁵	Van Allen <i>et al.</i> ⁵	Yan <i>et al.</i> ⁷
Total number of patients	—	417	53	66	44	22	24	9	24	45	130
Sex	Male	243 (58%)	28 (53%)	41 (62%)	31 (70%)	12 (54%)	11 (46%)	6 (67%)	16 (67%)	22 (49%)	76 (58.5%)
	Female	174 (42%)	25 (47%)	25 (38%)	13 (30%)	10 (45%)	13 (54%)	3 (33%)	8 (33%)	23 (51%)	54 (41.5%)
Age	Median (range), years	54 (17–94)	57 (21–90)	54.5 (21–83)	58.5 (29–84)	54.5 (25–74)	51.5 (18–94)	46 (27–76)	59 (17–84)	51 (25–76)	53 (17–86)
Lactate dehydrogenase	Elevated	115 (27.6%)	23 (43%)	24 (36.5%)	—	—	9 (37.5%)	3 (33%)	9 (37.5%)	—	47 (36%)
	Normal	159 (38%)	25 (47%)	38 (57.5%)	—	—	—	6 (67%)	13 (54%)	—	77 (59%)
	NA	143 (34.3%)	5 (9%)	4 (6%)	44 (100%)	22 (100%)	15 (32.5%)	—	2 (8%)	45 (100%)	6 (5%)
Drug type	vemurafenib	204 (49%)	4 (7.5%)	48 (72%)	37 (84%)	7 (32%)	—	—	5 (21%)	31 (69%)	72 (55%)
	dabrafenib + trametinib	94 (22.5%)	42 (79%)	8 (12%)	—	14 (64%)	21 (87.5%)	9 (100%)	—	—	—
	dabrafenib	48 (11.5%)	2 (4%)	6 (9%)	7 (16%)	—	—	—	19 (79%)	14 (31%)	—
	vemurafenib + cobimetinib	69 (16.5%)	4 (7.5%)	4 (6%)	—	—	3 (12.5%)	—	—	—	58 (45%)
	NA	1 (0.24%)	1 (2%)	—	—	—	—	—	—	—	—
Best Overall Response	progressive disease	141 (34%)	40 (75%)	10 (15%)	—	—	2 (10%)	—	4 (16.6%)	7 (15.5%)	78 (60%)
	stable disease	55 (13%)	3 (6%)	13 (20%)	9 (20.5%)	4 (18%)	4 (16.5%)	5 (55.5%)	5 (21%)	12 (27%)	—
	partial response	135 (32%)	5 (9%)	31 (47%)	31 (70.5%)	16 (73%)	10 (41.5%)	4 (45.5%)	13 (54%)	25 (55.5%)	—
	complete response	79 (19%)	3 (6%)	8 (12%)	4 (9%)	1 (4.5%)	8 (33%)	—	2 (8.3%)	1 (2%)	52 (40%)
	NA	7 (2%)	2 (4%)	4 (6%)	—	1 (4.5%)	—	—	—	—	—
Progression-free survival	Median (range), months	5.4 (0.5–68.3)	6.1 (0.7–30.5)	5.75 (1.6–58.9)	5.2 (2.8–16.3)	9.25 (1–24.2)	7.85 (0.8–68.3)	3.7 (2.6–9.3)	5.45 (1.6–22.4)	5.2 (1.2–16.5)	2.6 (0.5–19.4)
Overall survival	Median (range), months N.A.	12.2 (0.6–85.8) 160 (38%)	10.5 (0.7–55.8) —	14.5 (1.8–66.6) —	— 44 (100%)	18.9 (4.9–85.8) 3 (14%)	25.6 (3.4–69.4) —	10.8 (3.1–40.6) —	13.95 (5.1–47.5) —	— 45 (100%)	6.75 (0.6–20.7) 68 (52%)
Vital status	1=dead	175 (42%)	37 (70%)	45 (68%)	—	16 (73%)	11 (46%)	7 (78%)	17 (70%)	—	42 (32%)
	0=alive	81 (19.4%)	15 (28%)	21 (32%)	—	3 (14%)	13 (54%)	2 (22%)	7 (29%)	—	20 (15%)
	Lost to follow-up	3 (0.72%)	—	—	—	3 (14)	—	—	—	—	—
	NA	158 (38%)	1 (2%)	—	44 (100%)	—	—	—	—	45 (100%)	68 (52%)

Table 7. Patient features and treatment outcomes. For categorical features, numbers and percentages in the cohort are given. For continuous features, the median of the distribution in each cohort is given. NA, Not Available.

- Tumor_allele_2: primary genotype data. Tumor sequencing (discovery) allele 2. If similar to the Ref_allele, then no mutation. “-” for a deletion represents the variant. A novel inserted sequence does not include the flanking reference bases.
- Temporality: time of sample collection relative to the study treatment (pre-treatment; on treatment; progression; post-treatment).

gene_expressions.csv table: Gene expression (GEx) records include 8 641 391 samples records from Hugo *et al.*¹², Kwong *et al.*¹³, Long *et al.*¹⁴, Louveau *et al.*⁸, Rizos *et al.*¹⁵ and Yan *et al.*⁷. For Louveau *et al.*⁸, gene expression analysis was performed for a reduced list of genes. The list of targeted genes is available in Table 6. When available in the original source, the following features (specified by the column headings) are provided for each sample.

- HGNC: Hugo Gene Nomenclature Committee symbol associated with the sequenced gene.
- GeneID: gene identifier from the gene database of NCBI RefSeq genomes.
- Value: gene expression value; values are adjusted according to the study protocol (different for each study; see the Methods section).
- Temporality: time of sample collection relative to the study treatment (pre-treatment; on treatment; progression; post-treatment).

gene_expressions_matrix.csv view: Gene expressions matrix extracted from the Gene expression (GEx) table, with sample identifiers as columns and gene names as rows. The view has been generated to facilitate visualization of gene expression within the samples.

Technical Validation

Homogeneity of the clinical data. The patient features and treatment outcomes of the nine patient cohorts did not show any significant imbalance despite the variable cohort sizes (from 130 patients for the largest cohort to 9 patients for the smallest cohort) (Table 7, Fig. 2a). All patients were treated with MAPKi: 252 (61%) with BRAFi alone (vemurafenib or dabrafenib) and 163 (39%) with a combination of BRAFi and MEKi (dabrafenib + trametinib or vemurafenib + cobimetinib) (Fig. 2b). The patients’ age distribution was not significantly different among studies (Fig. 2c). 269 patients (66%) experienced an objective response (complete response, partial response or stable disease), and 141 patients (34%) experienced progressive disease despite the

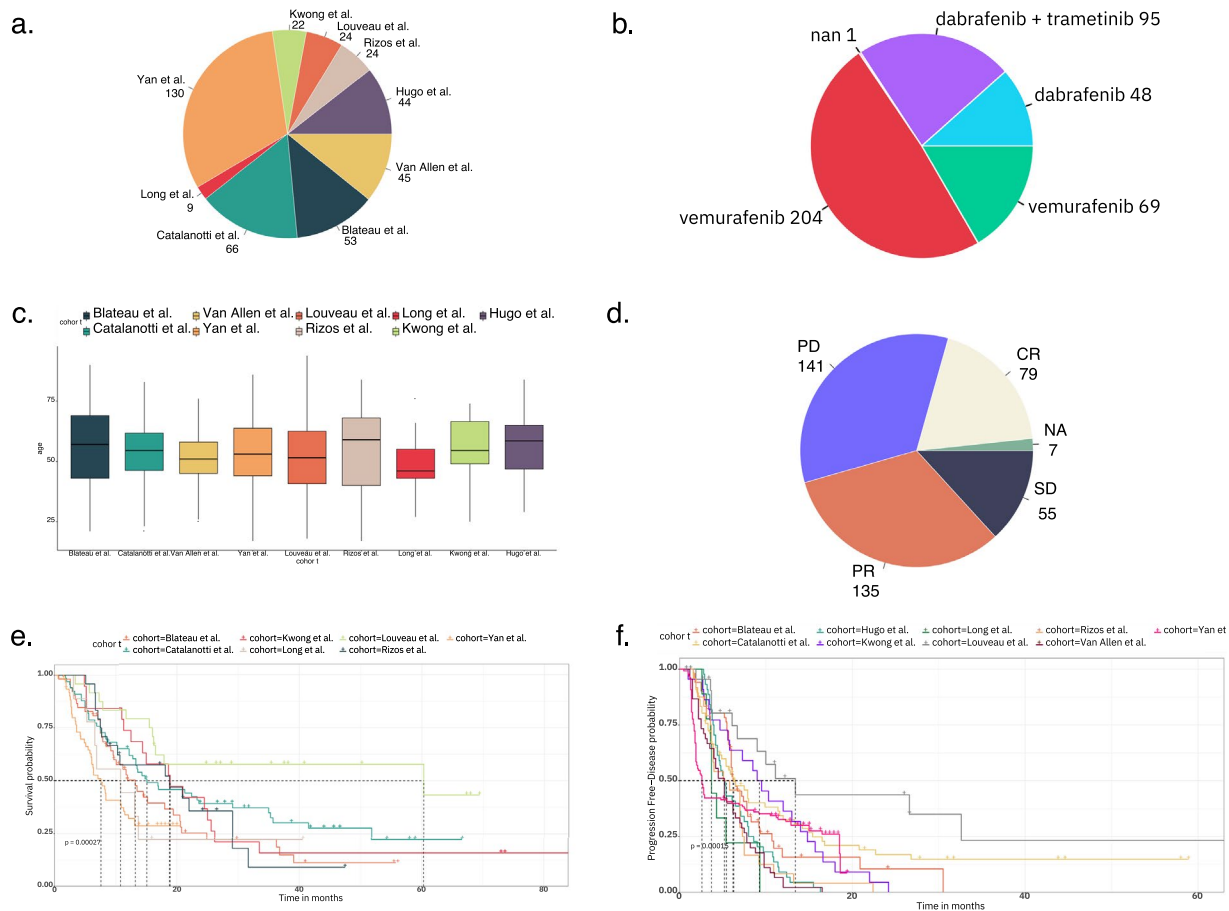


Fig. 2 Patient features and treatment outcomes. **(a)** Number of patients in the different cohorts (from the studies integrated in MelanoDB). **(b)** Number of patients who received the indicated MAPKi as mono-therapy (dabrafenib or vemurafenib) or in combination (dabrafenib + trametinib or vemurafenib + cobimetinib). **(c)** Boxplot showing the age distribution of patients in the different cohorts. Patient age was not significantly different among cohorts (p -value > 0.05 ; ANOVA). **(d)** Best overall response classification of patients according to the Response Evaluation Criteria in Solid Tumors (RECIST). CR, Complete Response; PR, Partial Response; SD, Stable Disease; PD, Progressive Disease. **(e)** Kaplan-Meier curves of overall survival in the indicated patient cohorts. **(f)** Kaplan-Meier curves of progression-free survival in the indicated patient cohorts. The Log-rank test was used to compare the Kaplan-Meier curves followed by the Holm-Bonferroni test to correct for multiple comparisons (**e,f**).

Features	Van Allen <i>et al.</i> ⁵	Blateau <i>et al.</i> ⁹	Catalanotti <i>et al.</i> ⁶	Louveau <i>et al.</i> ⁸
DNA sequencing method	WES	Targeted (35 genes)	Targeted (300 genes)	Targeted (74 genes)
Mutated genes	5.32%	5%	18.8%	2%

Table 8. Genomic features. Summary of the DNA sequencing methods and percentage of mutated genes in the four cohorts with SNV data. WES, whole-exome sequencing.

treatment (primary resistance) (Fig. 2d). This distribution was comparable to what reported in clinical studies on MAPKi efficacy^{22,23}.

We also investigated the survival outcomes to validate the data homogeneity across cohorts. Figure 2e,f show the Kaplan-Meier curves using the OS and PFS data, respectively. Comparison of the OS curves with the log-rank test highlighted a significant difference ($***p < 0.0005$), indicating a difference at least between two of the curves. Pairwise comparisons revealed a lack of homogeneity between the Louveau *et al.* and Yan *et al.* curves compared with the others. When we performed the same analysis without these two cohorts, we did not detect any significant difference ($p = 0.134$). The Yan *et al.* cohort included only patients with complete response or rapid progression. The Louveau *et al.* cohort included only 24 patients and was a real-life cohort. These two specific sampling designs may explain the observed differences in OS. Similarly, the log-rank test of the PFS curves highlighted a significant difference ($***p < 0.0005$). Pairwise comparisons revealed a lack of homogeneity between the Louveau *et al.* and Van Allen *et al.* curves. Van Allen *et al.* focused on the analysis

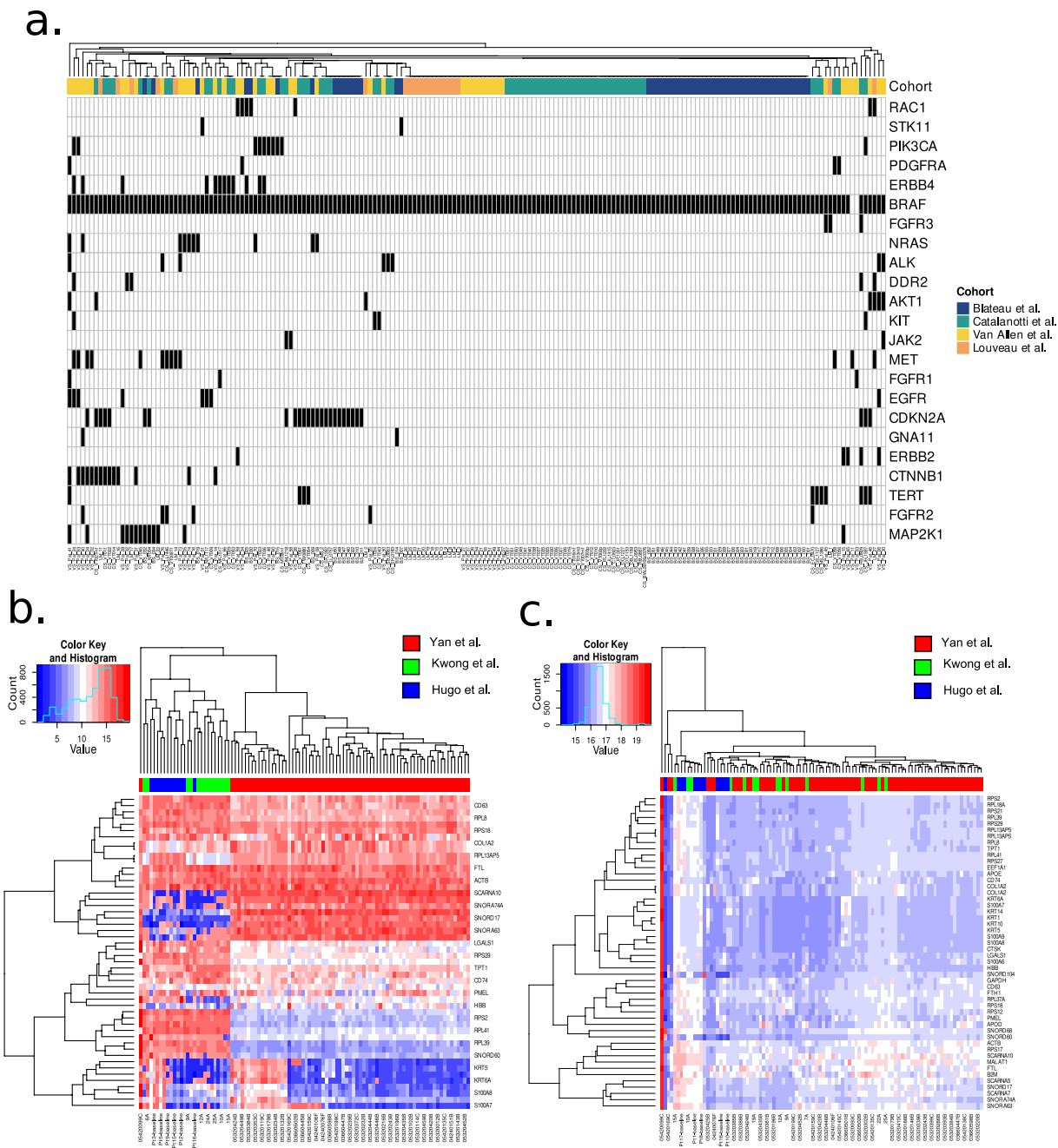


Fig. 3 Molecular features. **(a)** Hierarchical clustering of patients according to their tumor SNVs identified in a restricted number of genes that had been sequenced in all patients. Mutated genes are marked with a black box. Hierarchical clustering was performed with Rstudio. **(b,c)** Hierarchical clustering of patients according to their tumor gene expression profiles without **(b)** and with **(c)** batch correction.

of early resistance predictors using data from patients with very early and with late resistance. When we performed the same analysis without these two cohorts, we did not find any significant difference ($p = 0.309$). Overall, grouping these cohorts allowed us to overcome the cohort specific features linked to the inclusion criteria.

Homogeneity of the molecular data. To evaluate the homogeneity of SNV data among cohorts, we analyzed different characteristics of the dataset. Although the number of sequenced genes greatly varied among cohorts (from 35 genes to whole-exome sequencing), the percentage of mutated genes in tumors before treatment was comparable among three of the four cohorts with SNV data (Table 8). The mean percentage of mutated genes per patient was higher in the Catalanotti *et al.* cohort, although the number of sequenced genes was intermediate compared with the other cohorts. This higher percentage could be explained by the fact that only

cancer-associated genes were selected for sequencing. Then, we assessed the mutation distribution in a list of 23 genes that had been sequenced in all four studies. This small number was due to the fact that Louveau *et al.* sequenced only few genes. Using the mutation numbers in all patients across cohorts, we performed hierarchical clustering to identify possible patient groups based on their mutation profiles (Fig. 3a). One cluster included 60 patients with only *BRAF* mutations from all four cohorts. The remaining 98 patients displayed other SNVs and were not clustered according to their original cohort.

Then, we analyzed the gene expression data to assess their homogeneity among cohorts and to identify potential biases introduced by the cohorts and to propose a corrective solution if necessary. We compiled data from the Hugo *et al.*, Kwong *et al.* and Yan *et al.* studies because they used a similar method (mRNA expression profiling performed with Illumina HiSeq 2000). We used BatchQC, a tool that evaluates batch effects across conditions. This software helps to determine whether batch adjustment is required, and how correction should be applied before proceeding with the downstream analysis. We used the *inSilicoMerging* package²⁴ to check the validity of our approach. We visualized the ComBat-seq²⁵ adjustment at the gene-specific level. We first identified the genes the expression of which varied most among patients. We selected 500 genes with the highest variation across the dataset. We then investigated the cohort-driven batch effect with BatchQC. This showed that without batch correction, patients from the Yan *et al.* cohort clustered together and were separated from the two other cohorts (Fig. 3b). ComBat-seq adjustment corrected this batch effect (Fig. 3c).

Usage Notes

Data reuse and study limitations. Dataset tables are freely available in Zenodo¹⁸.

When reusing data extracted from MelanoDB, users must consider the relative differences in aims, design and patient composition among the different studies from which they were extracted. Some biases must be taken into account before data exploitation; however, our current dataset methodology and data homogenization have been carried out as meticulously as possible to avoid misinterpretations. Biases can be investigated using statistical methods and widely available software tools, such as Rstudio. Users must also take into account the fact that genomic sequencing data are not based on the same lists of genes depending on the cohort.

The MelanoDB dataset integrates multiple batches to provide a larger pool of data in order to increase the statistical power when using computational approaches. However, to avoid undesirable variations when re-using the data, users must pay attention to adjust the batch effects appropriately. Data quality control can be performed with the BatchQC software²⁶ that allows the rapid diagnosis of biases between batches by generating several visualizations, such as principal component analysis, dendrograms, heatmaps. To reuse the available gene expression data directly or to mix them with external sources, correction methods and/or software tools, such as ComBat-seq²⁵, are essential to make batches compatible.

Web browser. To visualize the different data types and their sources, we developed a user-friendly open source Streamlit web application that is freely available¹¹ to share this dataset with the scientific community according to the FAIR principles. This web application contains four different pages (patients, SNVs, CNAs and gene expression). Each page summarizes features that can be found in the associated tables through interactive graphs and visualizations.

Code availability

All codes used to generate the processed dataset, the codes used to produce the plots described in this article, and the Streamlit application are openly available at Github²⁷.

Received: 27 January 2025; Accepted: 28 May 2025;

Published online: 04 July 2025

References

- Long, G. V., Swetter, S. M., Menzies, A. M., Gershenwald, J. E. & Scolyer, R. A. Cutaneous melanoma. *The Lancet* **402**, 485–502 (2023).
- Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
- Robert, C. *et al.* Five-Year Outcomes with Dabrafenib plus Trametinib in Metastatic Melanoma. *New England Journal of Medicine*. (2019).
- Larkin James *et al.* Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *New England Journal of Medicine* **381**, 1535–1546, <https://www.nejm.org/doi/pdf/10.1056/NEJMoa1910836> (2019).
- Van Allen, E. M. *et al.* The Genetic Landscape of Clinical Resistance to RAF Inhibition in Metastatic Melanoma. *Cancer Discovery* **4**, 94–109 (2014).
- Catalanotti, F. *et al.* PTEN Loss-of-Function Alterations Are Associated With Intrinsic Resistance to BRAF Inhibitors in Metastatic Melanoma. *JCO precision oncology* **1** (2017).
- Yan, Y. *et al.* Genomic Features of Exceptional Response in Vemurafenib ± Cobimetinib-treated Patients with BRAF V600-mutated Metastatic Melanoma. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research* **25**, 3239–3246 (2019).
- Louveau, B. *et al.* Baseline Genomic Features in BRAFV600-Mutated Metastatic Melanoma Patients Treated with BRAF Inhibitor + MEK Inhibitor in Routine Care. *Cancers* **11**, E1203 (2019).
- Blateau, P. *et al.* TERT Promoter Mutation as an Independent Prognostic Marker for Poor Prognosis MAPK Inhibitors- Treated Melanoma. *Cancers* (2020).
- Dandou, S. *et al.* Improved prediction of MAPKi response duration in melanoma patients using genomic data and machine learning. *NPJ precision oncology*. <https://doi.org/10.1038/s41698-025-00814-y> (2025).
- Dandou, S., Radulescu, O. & Larive, R. MelanoDB web viewer application: Visualisation of clinical and molecular features of advanced melanoma patients treated with MAPK inhibitors. https://www.ircm.fr/crcm_en/melanodb.html Accessed: 2025-03-23 (2024).

12. Hugo, W. *et al.* Non-genomic and Immune Evolution of Melanoma Acquiring MAPKi Resistance. *Cell* **162**, 1271–1285 (2015).
13. Kwong, L. N. *et al.* Co-clinical assessment identifies patterns of BRAF inhibitor resistance in melanoma. *The Journal of Clinical Investigation* **125**, 1459–1470 (2015).
14. Long, G. V. *et al.* Increased MAPK reactivation in early resistance to dabrafenib/trametinib combination therapy of BRAF-mutant metastatic melanoma. *Nature Communications* **5**, 5694 (2014).
15. Rizos, H. *et al.* BRAF Inhibitor Resistance Mechanisms in Metastatic Melanoma: Spectrum and Clinical Impact. *Clinical Cancer Research* **20**, 1965–1977 (2014).
16. Blateau, P. *et al.* Tert promoter mutation as an independent prognostic marker for poor prognosis mapk inhibitors-treated melanoma - dataset. <https://zenodo.org/records/14603634> Accessed: 2025-03-23 (2020).
17. Louveau, B. *et al.* Baseline genomic features in brafv600-mutated metastatic melanoma patients treated with braf inhibitor + mek inhibitor in routine care - dataset. <https://zenodo.org/records/14633948> Accessed: 2025-03-23 (2019).
18. Dandou, S., Radulescu, O. & Larive, R. MelanoDB: Dataset files of clinical and molecular features of advanced melanoma patients treated with MAPK inhibitors. <https://doi.org/10.5281/zenodo.15292916> Accessed: 2025-03-23 (2024).
19. Keung, E. Z. & Gershenwald, J. E. The eighth edition American Joint Committee on Cancer (AJCC) melanoma staging system: Implications for melanoma treatment and care. *Expert review of anticancer therapy* **18**, 775–784 (2018).
20. Clark's Level - an overview | ScienceDirect Topics. <https://www.sciencedirect.com/topics/medicine-and-dentistry/clarks-level>.
21. Eisenhauer, E. A. *et al.* New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). *European journal of cancer* **45**, 228–247 (2009).
22. Sosman, J. A. *et al.* Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *The New England Journal of Medicine* **366**, 707–714 (2012).
23. Long, G. V. *et al.* Adjuvant Dabrafenib plus Trametinib in Stage III BRAF-Mutated Melanoma. *New England Journal of Medicine* **377**, 1813–1823, <https://doi.org/10.1056/NEJMoa1708539> (2017).
24. Taminau, J. *et al.* Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. *BMC Bioinformatics* **13**, 335 (2012).
25. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: Batch effect adjustment for RNA-seq count data. *NAR Genomics and Bioinformatics* **2**, lqaa078 (2020).
26. Manimaran, S. *et al.* BatchQC: interactive software for evaluating sample and batch effects in genomic data. *Bioinformatics* **32**, 3836–3838, <https://doi.org/10.1093/bioinformatics/btw538> (2016).
27. Dandou, S. Github repository: MelanoDB, creation of a database collection of patients with advanced melanoma treated with MAPK inhibitors. <https://doi.org/10.5281/zenodo.15321518> Accessed: 2025-03-23 (2024).
28. Catalanotti, F. *et al.* Pten loss-of-function alterations are associated with intrinsic resistance to braf in- hibitors in metastatic melanoma. cbiportal. https://www.cbiportal.org/study/summary?id=skcm_vanderbilt_mskcc_2015 Accessed: 2025-03-23 (2017).
29. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* **2**, 401–404 (2012).
30. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling* **6**, p11 (2013).
31. de Bruijn, I. *et al.* Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Research* **83**, 3861–3867 (2023).
32. Hugo, W. *et al.* Non-genomic and immune evolution of melanoma acquiring mapki resistance. e.f. geo. <https://identifiers.org/geo/GSE65186> Accessed: 2025-03-23 (2015).
33. Kwong, L. N. *et al.* Co-clinical assessment identifies patterns of BRAF inhibitor resistance in melanoma. e.f. ega. <https://ega-archive.org/studies/EGAS00001000992>. Accessed: 2025-03-23 (2015).
34. Long, G. V. *et al.* Increased mapk reactivation in early resistance to dabrafenib/trametinib combination therapy of braf-mutant metastatic melanoma. e.f. geo. <https://identifiers.org/geo/GSE61992> Accessed: 2025-03-23 (2014).
35. Rizos, H. *et al.* Braf inhibitor resistance mechanisms in metastatic melanoma: spectrum and clinical impact. e.f. geo. <https://identifiers.org/geo/GSE50509> Accessed: 2025-03-23 (2014).
36. Van Allen, E. M. *et al.* The genetic landscape of clinical resistance to raf inhibition in metastatic melanoma. cBio- portal. https://www.cbiportal.org/study/summary?id=skcm_broad_brafresist_2012 Accessed: 2025-03-23 (2014).

Acknowledgements

The authors would like to thank Jean-Christophe Marine (Laboratory for Molecular Cancer Biology, Center for Cancer Biology, VIB, Leuven, Belgium; Department of Oncology, KU Leuven, Leuven, Belgium), Yibing Yan (Translational Medicine, Genentech, Inc., South San Francisco, California, USA) and Antoni Ribas (Jonsson Comprehensive Cancer Center at the University of California, Los Angeles UCLA, Los Angeles, CA, USA) for personal communications concerning their published data. This work has been supported by the INCa-Cancéropôle GSO (Programme Emergence Intelligence Artificielle N°2022- IA20 and Programme Emergence N°2018-E1), the Fondation ARC pour la recherche sur le cancer (Projets ARC 2022 PJA2 N°ARCPJA2022060005107), and the Ligue nationale contre le cancer (Comité du Gard JPB/GA/MV/39-2019). SD is supported by the Agence Nationale de la Recherche (Intelligence Artificielle en Santé et Environnement – AXIAUM N°ANR-20-THIA-0005-01) and the University of Montpellier/Ecole Doctorale CBS2.

Author contributions

R.M.L. ensured data collection and communication with the authors of the original studies from which the data were recovered, and verified the consistency of the data with the study objective. J.A.V., J.S., B.L., C.L. and S.M. shared data. S.D. ensured the creation of the dataset the homogenization of the integrated data and performed bioinformatics analyses. S.D. reviewed the bioinformatics analyses. R.M.L. and O.R. acquired funding and supervised the project. R.M.L. and S.D. interpreted and curated the data, and wrote the original draft. All authors reviewed and contributed to the manuscript.

Competing interests

The authors declare that they have no known competing financial interest or personal relationship that could have appeared to influence the work reported in this paper.

Additional information

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

Reprints and permissions information is available at 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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025