nature research

Corresponding author(s): Stein Aerts

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐ Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

In this paper we used 9 single-cell data sets, a bulk Omni-ATAC data set with 8 samples (melanoma cell lines after SOX10KD on different time points), public epigenomic and transcriptomic, 2 ChIP-seq data sets:

1. Simulated single-cell epigenomes from the hematorpoietic system: Single-cell epigenomes were simulated using 13 bulk ATAC-seq profiles from different cell types in the hematopoietic lineage from Corces et al. (2016). Data was downloaded from GEO GSE74912. Bulk reads were mapped to the genome (hg19-Gencode v18) using STAR (v2.5.1), with the options applying the parameters --alignIntronMax 1, --aslignIntronMin 2 and --alignMatesGapMax 2000. MACS2 (version 2.0.10) was used with the option --qvalue 0.001 and nomodel. Called peaks from all cell types were merged using a custom script that combines overlapping and concatenated regions into a single feature. These peaks were used as candidate regulatory regions for cisTopic, after blacklisting (http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg19-human/). We removed regions overlapping the blacklist with at least 40% overlap. Bulk BAM files were used for simulating single cell BAM files using a custom script. For each simulation, a specific value for the desired read depth is provided. For each cell, the given number of reads is randomly sampled from the BAM file, resulting in a new single-cell epigenome profile. These BAM files were also used as input for cisTopic.

2. Simulated single-cell epigenomes from melanoma cell lines: Single-cell epigenomes were simulated using bulk H3K27Ac data from 14 cell lines (11 from Verfaille, Imrichová & Kalender-Atak et al., 2015, GSE60666; and 4 generated in this study). Bulk reads were mapped to the genome (hg19-Gencode v18) using Bowtie2 2.1.0. The sensitive-local setting for Bowtie2 was used to correct for a high percentage of mismatches at the start of a read, prompting the removal of the first five base pairs of each read. MACS2 (version 2.0.10) was used with the option --qvalue 0.001, and using as control the combination of inputs from 5 melanoma lines (A375, MM011, MM032, MM047,

MM057). Called peaks from all cell lines were merged using a custom script that combines overlapping and concatenated regions into a single feature. These peaks were used as candidate regulatory regions for cisTopic, after blacklisting (http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg19-human/). We removed regions overlapping the blacklist with at least 40% overlap. Bulk BAM files were used for simulating single cell BAM files using a custom script. For each simulation, a random vector with values between the top and bottom of the coverage interval must be provided. For each cell, the given number of reads is randomly sampled from the BAM file, resulting in a new single-cell epigenome profile. These BAM files were also used as input for cisTopic.

3. scATAC-seq from FAC-sorted populations from the hematopoietic system: Data was downloaded from GEO GSE96772. The single-cell reads were first cleaned for adapters using fastq-mcf using fastq-mcf (as part of ea utils; v1.1.2-686). Read quality was then checked using FastQC (v0.11.5). Paired-end reads were mapped to the human genome (hg19-Gencode v18) using STAR (v2.5.1) applying the parameters --alignIntronMax 1, --aslignIntronMin 2 and --alignMatesGapMax 2000. Mapped reads were filtered for quality using SAMtools (v1.2) view with parameter -q4, sorted with SAMtools sort and indexed using SAMtools index. Duplicates were removed using Picard (v1.134) MarkDuplicates using OPTICAL_DUPLICATE_PIXEL_DISTANCE=2500. We used as input for cisTopic the BAM files and the regions defined by Buenrostro et al., 2018, resulting in a count matrix with 2,755 cells and 488,825 regions.

4. scnmC-seq in human neuronal populations from the hematopoietic system: Data (raw methylation values per genomic bin))was downloaded from http://brainome.org Luo et al. 2017).

5. sciATAC-seq Mouse Cell Atlas: Data (binarized peak by cell matrix and annotations) was downloaded from http://atlas.gs.washington.edu/mouse-atac/ .

6. scTHS-seq in the human brain: Data was downloaded from GSE97887 (Lake et al., 2017). Single-cell epigenomes were merged into a matrix with 34,520 cells and 287,381 regulatory regions (after blacklisting http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg38-human/). We removed regions overlapping the blacklist with at least 40% overlap.

7. snDrop-seq in the human brain: Data was downloaded from GSE97930 (Lake et al., 2018), as a DGE matrix (and annotations). Cells containing less than 800 genes expressed were removed, resulting in a matrix with 15,884 cells.

8. scATAC-seq in the mouse brain: Data was retrieved from GEO GSE100033. This data set contains 3,034 cells and 139,504 regions.

9. scATAC-seq during an EMT-like transition in melanoma: We generated scATAC-seq data on different time points (0, 24, 48 and 72h) for two melanoma cell lines (MM057 and MM087) upon SOX10 KD, which triggers an EMT-like cell state transition. The reads from scATAC-seq samples were first cleaned for adapters using fastq-mcf using fastq-mcf (as part of ea utils; v1.1.2-686). Read quality was then checked using FastQC (v0.11.5). Paired-end reads were mapped to the human genome (hg19-Gencode v18) using STAR (v2.5.1) applying the parameters --alignIntronMax 1, --aslignIntronMin 2 and --alignMatesGapMax 2000. Mapped reads were filtered for quality using SAMtools (v1.2) view with parameter -q4, sorted with SAMtools sort and indexed using SAMtools index. Duplicates were removed using Picard (v1.134) MarkDuplicates using OPTICAL_DUPLICATE_PIXEL_DISTANCE=2500. To filter out cell of bad quality, transcription start site aggregation plots were made using a custom script and cell having a low signal-to-noise profile were removed from further analyses. This lead to a final of 598 good quality cells over 8 Fluidigm C1 runs. Bam files of good quality single cells were aggregated per condition and peaks were called on these aggregated samples using MACS2 (v2.1.1) callpeak using the parameters --nomodel and --call-summits. The peak files per condition were merged (78,661 peaks in total before blacklisting) and filtered using the blacklisted regions on http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg19-human/, resulting a total of 78,262 peaks. We removed regions overlapping the blacklist with at least 40% overlap. This peak file was used, together with the BAM files of the good single cells as, input for cisTopic. To visualise the aggregated cells per sample, normalised bedGraphs were produced by genomeCoverageBed (as part of bedtools; v2.23.0) using as scaling parameter (-scale) size factors obtained from DEseq2 (v1.18.1). BedGraphs were converted to bigWigs by the bedtools suit functions bedSort to sort the bedGraphs, followed by bedGraphToBigWig to create the bigWigs.

10. Omni-ATAC during an EMT-like transition in melanoma: Adapter sequences were trimmed from the fastq files using fastq-mcf (as part of ea utils; v1.04.807). Read quality was then checked using FastQC (v0.11.5). Reads were mapped to the human genome (hg19-Gencode v18) using STAR (v2.5.1) applying the parameters --alignIntronMax 1 and --aslignIntronMin 2. Mapped reads were filtered for quality using SAMtools (v1.2) view with parameter -q4, sorted with SAMtools sort and indexed using SAMtools index. Peaks were called using MACS2 (v2.1.1) callpeak using the parameters --nomodel and --call-summits on the 8 conditions separately. A count matrix was generated by using featureCounts (as part of Subread; v1.4.6) of all separate bam files on the merged peak file (after conversion of the merged peak bed file to a gff format using a custom script). Normalised bedGraphs were produced by genomeCoverageBed (as part of bedtools; v2.23.0) using as scaling parameter (-scale) size factors obtained from DEseq2 (v1.18.1). BedGraphs were converted to bigWigs by the bedtools suit functions bedSort to sort the bedGraphs, followed by bedGraphToBigWig to create the bigWigs, which were used in IGV for visualisation.

11. Public signatures: Layer-specific regions from the mouse brain were taken from Gray et al. (2017), interneuron signatures (Mo et al., 2015); from GEO GSE63137 and the dentate gyrus signature (Su et al., 2017) from GEO GSE82010. scRNA-seq oligodendrocyte and astrocyte signatures for the mouse were obtained from Habib et al. (2017) and methylation signatures for the mouse and the human brain were retrieved from Mo et al. (2017) and Kozlenkov et al. (2018), respectively.

12. ChIP-seq data sets: GATA2 ChIP-seq peaks (GEO GSE32465) were downloaded as bed files from ChIP-Atlas (with q < 1E-20). SOX10 ChIP-seq data was downloaded as raw fastq files from GEO (GSE61965; Laurette et al., 2015) and was mapped to the human genome using Bowtie2 (v2.1.0) and peaks were called by MACS2 (v2.1.1).

Data analysis

The cisTopic code (v0.1.0) is available at: https://github.com/aertslab/cisTopic. The R version used during this work, including for the development and testing of cisTopic, is v3.4.3 (Bioconductor v3.6). The versions of the dependencies with which cisTopic analyses have been run are: Rsubread (v1.28.1), GenomicRanges (v1.30.3), stats (v3.4.3), lda (v1.4.2), doSNOW (v1.0.16), plyr (1.8.4), Rtsne (v0.13), Umap (v0.2.0.0), destiny (v2.6.2), fastcluster (v1.1.24), NMF (v0.23.6), DT (v0.4), fitdistrplus (v1.0-9), RcisTarget (v1.1.2), rGREAT (v1.11.1), ChIPseeker (v1.14.2), ggplot2 (v2.2.1), data.table (v1.11.4), feather (v0.3.1), AUCell (v1.3.3), TxDb.Hsapiens.UCSC.hg19.knownGene (v3.2.2), TxDb.Hsapiens.UCSC.hg38.knownGene (v3.4.0) and org.Hs.eg.db (v3.5.0). For the method comparison on simulated data, chromVAR (v1.0.1), scABC (v0.99.0), Cicero (v1.0.1), SCRAT (v0.99.0) and the BROCKMAN pipeline (bash (v.1.0) and BrockmanR (0.0.0.9000)) were used. For the analysis of single-cell RNA-seq data in the human brain, SCENIC (v0.1.5)

was used (with RcisTarget (v1.2.0), AUCell (v1.5.2), and GRNBoost2 (v0.1.3) with default parameters). Homer (v4.9; findMotifs.pl) and RSAT peak-motifs (http://rsat.sb-roscoff.fr/peak-motifs_form.cgi) were used for de novo motif discovery; while for known motifs we have used the cisTarget collection (v8, available as a feather file for genes or regions at: https://resources.aertslab.org/cistarget/). Additionally, best CRM scores per region were calculated with our custom script (v0.1.0; https://github.com/aertslab/primescore/tree/master/src/make_feature_table.py); compatible with the Cluster-Buster version 2b810d (https://github.com/weng-lab/cluster-buster/). For differential motif enrichment comparisons, R packages MAST (v1.4.1) and ggtern (v3.0.0) were used. STAMP (v1.3) was used for motif clustering. Coverage heatmaps were done using a custom script that takes a BAM file as an input and outputs a matrix of either read depth or fragment start sites based on a bed input, an aggregation plot and a joined aggregation plot/heatmap. We used MUSCLE (v3.8.31) for aligning mouse and human brain conserved regions. TOUCAN (http://dev.bits.vib.be:8888/toucan.js/) was used for visualization of motifs and CRMs in the regions.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

- Public data sets
+ GSE74912 (Corces et al., 2016): Bulk ATAC-seq profiles from 13 cell-types from the hematopoeitic system
+ GSE60666 (Verfaille, Imrichová & Kalender-Atak et al., 2015): H3K27Ac in a collection of 10 melanoma cell lines
+ http://brainome.org (Luo et al., 2017): scnmC-seq data from human neurons
+ http://atlas.gs.washington.edu/mouse-atac/ (Cusanovich et al., 2018): sciATAC-seq in several mouse tissues (Mouse Cell Atlas)
+ GSE96772 (Buenrostro et al., 2018): scATAC-seq in FAC-sorted populations from the hematopoietic system
+ https://chip-atlas.org : GATA2 Chip-seq data
+ GSE97887 (Lake et al., 2017): scTHS-seq in the human brain
+ GSE97930 (Lake et al., 2017): snDrop-seq in the human brain
+ GSE100033 (Preissl et al., 2018): scATAC-seq in the mouse prefrontal cortex
+ Gray et al. (2017): Mouse brain layer-specific regions
+ GSE63137 (Mo et al., 2015): Interneuron ATAC-seq regions
+ GSE82010 (Su et al., 2017): Dentate gyrus signature
+ Habib et al. (2017): Oligodendrocyte and astrocyte scRNA-seq signatures
+ Mo et al. (2017): Mouse brain methylation profiles
+ Kozlenkov et al. (2018): Human methylation signatures
+ http://mousebrain.org: scRNA-seq from the excitatory neuron layers
+ GSE61965 (Laurette et al., 2015): SOX10 ChIP-seq in melanoma

- This work (GSE114557)
+ H3K27Ac profiles of 3 melanoma cell lines: A375, MM029 and MM032
+ Omni-ATAC during an EMT-like transition in melanoma (8 samples): Two cell lines (MM057 and MM087) under SOX10KD and sampling on different time points (0h, 24h, 48h, 72h)
+ scATAC-seq during an EMT-like transition in melanoma (598 samples): Two cell lines (MM057 and MM087) under SOX10KD and sampling on different time points (0h, 24h, 48h, 72h)

- Figures
+ Fig 1: GSE96772
+ Fig 2: GSE97887, GSE100033,
+ Fig 3: scATAC-seq during an EMT-like transition in melanoma (598 samples) (GSE114557), GSE61965
+ Fig S1: GSE74912
+ Fig S2: GSE74912
+ Fig S3: GSE74912
+ Fig S4: GSE60666, H3K27Ac profiles of 4 melanoma cell lines: A375, HS895T, MM029 and MM032 (GSE114557)
+ Fig S5: GSE60666, H3K27Ac profiles of 4 melanoma cell lines: A375, HS895T, MM029 and MM032 (GSE114557)
+ Fig S6: http://brainome.org (Luo et al., 2017)
+ Fig S7: http://atlas.gs.washington.edu/mouse-atac/ (Cusanovich et al., 2018)
+ Fig S8: GSE96772
+ Fig S9: GSE96772, https://chip-atlas.org
+ Fig S10: GSE97930, GSE97887
+ Fig S11: GSE100033, GSE63137, GSE82010 , Gray et al. (2017),
+ Fig S12: GSE97887, GSE100033, GSE97930, Gray et al. (2017), Habib et al. (2017), Mo et al. (2017), Kozlenkov et al. (2018)
+ Fig S13: Omni-ATAC during an EMT-like transition in melanoma (8 samples) and scATAC-seq during an EMT-like transition in melanoma (598 samples) (GSE114557)
+ Fig S14: GSE97887, GSE100033, scATAC-seq during an EMT-like transition in melanoma (598 samples) (GSE114557)
+ Fig S15: GSE74912

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

- cisTopic analysis
No statistical method was used to predetermine sample size. We show the application of cisTopic to 9 epigenomic datasets, mainly from single-cell ATAC-seq. These datasets were selected to cover different case studies: a simulated data set from ATAC-seq profiles (using different coverages) (Corces et al., 2016), a simulated data set from H3K27Ac profiles (using different coverages) (Verfaille et al., 2015), scATAC-seq on FAC-sorted populations from the development of the hematopoietic system (Buenrostro et al., 2018), a methylation data set (Luo et al., 2017), the largest scATAC-seq data set to date (Cusanovich et al., 2018), two data sets from the brain (human (Lake et al., 2017) and mouse (Preissl et al., 2018)) and a novel time-series scATAC-seq data set during SOX10KD in melanoma cell lines.

- scATAC-seq
No statistical method was used to predetermine sample size. We performed scATAC-seq on one replicate of each condition as the single cells themselves can already serve as technical replicate and as this is also the convention for scATAC-seq studies. We also performed bulk OmniATAC-seq on the same conditions and saw strong correlation between the bulk and single-cell results for the same conditions.

- OmniATAC-seq
No statistical method was used to predetermine sample size. OmniATAC-seq was done on one replicate of each condition. In addition, we have also performed one replicate of FastATAC-seq on the same conditions, which reproduced the profiles found by our OmniATAC-seq experiment (data not shown in the manuscript).

- H3K27Ac ChIP-seq
No statistical method was used to predetermine sample size. We performed H3K27Ac ChIP-seq on three different melanoma lines (one replicate for each one). We considered this as sufficient as their profiles were comparable to previously generated FAIRE-seq data (Verfaillie et al., 2015 and in-house).

Data exclusions

- cisTopic analysis
No data were excluded in these analyses.

- scATAC-seq
Samples with low signal on TSS aggregation plots were disregarded for the analysis, as described in the methods section. These most probably represent reactions chambers in the microfluidic devices that did not contain a cell. This exclusion was pre-established; it was purely based on the raw signal and was performed prior to any further analyses.

- OmniATAC-seq
No data were excluded from the analyses.

- H3K27Ac ChIP-seq
No data were excluded from the analyses.

Replication

- cisTopic analysis
cisTopic reliably identified the expected cell types (plus some novel cell types) in all analysed datasets, as well as regulatory topics that match with previous literature. Models were run several times with different parameters (e.g. random seeds, not shown), and resulted in consistent conclusions.

- scATAC-seq
We performed OmniATAC-seq on the same conditions to verify the single-cell data. OmniATAC-seq samples reproduced the findings by scATAC-seq.

- OmniATAC-seq
We have performed FastATAC-seq on the same conditions (a methods similar to OmniATAC-seq, although on a lower cell number and with lower signal-to-noise ratio). FastATAC-seq samples on the same timepoints reproduced the original findings reliably (data not shown in the manuscript).

- H3K27Ac ChIP-seq
We compared the profiles obtained by the H3K27Ac ChIP-seq with FAIRE-seq data (Verfaillie et al., 2015 and in-house). FAIRE-seq profiles correlated with the H3K27Ac profiles (data not shown in the manuscript).

Randomization

- cisTopic analysis

Randomization

Not relevant. Each analysis was independent.

- scATAC-seq
Not relevant. Groups were determined by the set time points and cells at each time point were processed in parallel and thus were highly randomized.

- OmniATAC-seq
Not relevant. Groups were determined by the set time points.

- H3K27Ac ChIP-seq
Not relevant. Groups were determined by the cell line.

Blinding

- cisTopic analysis
Not relevant. Analyses were performed on the accessibility matrix alone, without taking into account the cell types, cell lines, time points or any other phenotypic information. These information was only used a posteriori for validation of the results obtained.

- scATAC-seq
No blinding was used. However, no prior knowledge was used to perform cisTopic and to define the different topics.

- OmniATAC-seq
No blinding was used.

- H3K27Ac ChIP-seq
No blinding was used.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

ChIP-seq was performed using the Anti-histone H3 acetyl K27 antibody (ab4729, Abcam).

Validation

The antibody has been validated by the manufacturers using peptide arrays and on HeLa cells (http://www.abcam.com/histone-h3-acetyl-k27-antibody-chip-grade-ab4729.html).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

The cell lines (A375, MM001, MM011, MM029, MM031, MM032, MM034, MM047, MM057, MM074, MM087, MM099, MM118, SKMEL5) were kindly provided by Ghanem-Elias Ghanem (ULB) (sample delivered under MTA).

Authentication

The A375 line was authenticated by verifying the presence of three mutations in A375 cells according the ATCC, namely BRAF homozygous c.1799T>A, CDKN2A homozygous c.181G>T and CDKN2A homozygous c.205G>T in RNA-seq data from the same line. Similarly, we verified MM087 by the lack of the BRAF c1799T>A and NRAS Q61L mutations and the presence of mutations in TP53 (chr17:7578268 A>C) and CTNNB1 (chr3:41281413 T>A); MM057 by the lack of the BRAF c1799T>A and TP53 (chr17:7578268 A>C) mutations, and presence of NRAS Q61L mutation. For MM032 and MM029, there is no mutational profile available, but all the MM lines were received under a MTA from the lab of Ghanem-Elias Ghanem, which is the primary source of these short-term patient-derived cultures.

Mycoplasma contamination

The cell lines were tested regularly for mycoplasma contamination. Results were negative at the time of the experiment. Also, ATAC-seq is very sensitive to mycoplasma infections and an infection would have been clear from the data due to low mapping to the human genome.

Commonly misidentified lines
(See ICLAC register)

No misidentified cell lines have been used in this study.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	GEO accession GSE114557 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114557).
Files in database submission	GSE114557 contains H3K27Ac ChIP-seq of 3 melanoma lines, 8 OmniATAC-seq data sets and 1 single-cell ATAC-seq data set of 762 single cells. Files in the series: - H3K27Ac ChIP-seq: Raw fastq files, bigWigs and Bed files - scATAC-seq: Raw fastq files (paired end), processed counts matrix and bigWigs of the aggregate per condition - OmniATAC-seq: Raw fastq files, processed counts matrix and bigWigs
Genome browser session (e.g. UCSC)	The H3K27Ac ChIP-seq data is accessible to view in UCSC via this link: http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&hubUrl=http://ucsctracks.aertslab.org/cistopic/hub.txt

Methodology

Replicates	- We performed H3K27Ac ChIP-seq on three different melanoma lines (one replicate for each one). We considered this as sufficient as their profiles were comparable to previously generated FAIRE-seq data (Verfaillie et al., 2015 and in-house).
Sequencing depth	For A375, MM029 and MM032, respectively: - Number of reads: 193,350,020; 113,253,892; 80,897,992 - Uniquely mapped reads: 43,704,340; 27,550,922; 17,174,430 - Length of reads: 51 base pairs - All reads are single-end
Antibodies	ChIP-seq was performed using the Anti-histone H3 acetyl K27 antibody (ab4729, Abcam)
Peak calling parameters	We used as control the merged control profiles of five cell lines, namely A375, MM011, MM032, MM047 and MM057. We ran MACS2 (v2.0.10) with the command 'callpeak -t ChIPseq_file -c Control_file -g hs --qvalue 0.001 -n name'
Data quality	We used the -qvalue 0.001 parameter to specify the minimum FDR to call only significant regions. We obtained 36,075 peaks for A375; 46,549 peaks for MM029; and 32,102 peaks for MM032.
Software	Bulk reads were mapped to the genome (hg19-Gencode v18) using Bowtie2 2.1.0. The sensitive-local setting for Bowtie2 was used to correct for a high percentage of mismatches at the start of a read, prompting the removal of the first five base pairs of each read. MACS2 (version 2.0.10) was used with the option --qvalue 0.001, and using as control the combination of inputs from 5 melanoma lines (A375, MM011, MM032, MM047, MM057).