TECHNICAL REPORT
https://doi.org/10.1038/s41588-019-0424-9

nature genetics

High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer

Kevin Grosselin[1,2,9], Adeline Durand[3,4], Justine Marsolier[3,4], Adeline Poitou[1,10], Elisabetta Marangoni[4], Fariba Nemati[4], Ahmed Dahmani[4], Sonia Lameiras[5], Fabien Reyal[4,6,7], Olivia Frenoy[1,11], Yannick Pousse[1], Marcel Reichen[1,12], Adam Woolfe[1], Colin Brenan[1,8], Andrew D. Griffiths[2,13*], Céline Vallot[3,4,13*] and Annabelle Gérard[1,13*]

Modulation of chromatin structure via histone modification is a major epigenetic mechanism and regulator of gene expression. However, the contribution of chromatin features to tumor heterogeneity and evolution remains unknown. Here we describe a high-throughput droplet microfluidics platform to profile chromatin landscapes of thousands of cells at single-cell resolution. Using patient-derived xenograft models of acquired resistance to chemotherapy and targeted therapy in breast cancer, we found that a subset of cells within untreated drug-sensitive tumors share a common chromatin signature with resistant cells, undetectable using bulk approaches. These cells, and cells from the resistant tumors, have lost chromatin marks—H3K27me3, which is associated with stable transcriptional repression—for genes known to promote resistance to treatment. This single-cell chromatin immunoprecipitation followed by sequencing approach paves the way to study the role of chromatin heterogeneity, not just in cancer but in other diseases and healthy systems, notably during cellular differentiation and development.

The emergence of resistance to chemotherapy and targeted therapies is a major challenge for the treatment of cancer. Deep sequencing and single-cell approaches have highlighted the importance of genetic intra-tumor heterogeneity in tumor evolution[1–3], and shown that genetic heterogeneity within untreated tumors is a key factor in tumor resistance[4]. However, in many cases genetic mechanisms driving resistance have not been found, pointing to a role for non-genetic mechanisms[5–7]. Transcriptional and epigenetic mechanisms are anticipated to play a role in the adaptation of cancer cells confronted with environmental, metabolic or therapy-related stresses[8,9]. Recent studies, using single-cell RNA sequencing (scRNA-seq), indicate that emergence of transcriptional subclones on treatment may account for the adaptation of cancer cells to therapeutic pressure[9,10]. In contrast, only a few studies have tracked the clonal evolution of epigenetic alterations, analyzing DNA methylation at the population level[11,12], suggesting that DNA methylation alterations and genetic mutations can share a common evolutionary track[11]. Recently developed approaches for single-cell methylome profiling[13,14] might also help explore the complexity of DNA methylation in tumors. Single-cell assay for transposase-accessible chromatin using sequencing (ATAC-seq) has recently enabled the detailed characterization of open chromatin states in vivo, in *Drosophila* and human embryos[15,16], but unlike chromatin immunoprecipitation followed by sequencing (ChIP-seq), such approaches do not fully capture the diversity of chromatin states found in the human genome[17,18]. For example, unlike ATAC-seq, ChIP-seq can capture repressive as well as active chromatin states. However, until now, insufficient coverage has limited the applications of single-cell chromatin profiling with ChIP-seq[19], preventing the study of the heterogeneity of chromatin states in complex biological systems such as tumors.

To explore intra-tumor heterogeneity of chromatin states, we developed a droplet microfluidics single-cell chromatin immunoprecipitation followed by sequencing (scChIP-seq) approach to profile chromatin landscapes of thousands of cells at single-cell resolution with a coverage of up to 10,000 loci per cell. Using cell lines, we show that our approach can reveal cell identities from single-cell chromatin landscapes with high accuracy, and identify discriminating chromatin features, either permissive (H3K4me3) or repressive (H3K27me3) for transcription, between groups of single cells. Applying our methodology to breast cancer patient-derived xenograft (PDX) samples, we characterize the diversity of chromatin landscapes within stromal and tumor cell populations. Notably, when comparing pairs of sensitive and resistant PDXs, we show that a fraction of cells in untreated, drug-sensitive tumors display the same H3K27me3 chromatin landscape as resistant cells, revealing a facet of tumor heterogeneity at the level of chromatin features.

Results

Droplet microfluidics workflow for scChIP-seq. To study the intra-tumor heterogeneity of chromatin states, we developed a high-throughput scChIP-seq approach that combines droplet microfluidics with single-cell DNA barcoding technologies (Fig. 1a and Supplementary Figs. 1–4). We profiled histone post-translational modifications, from chromatin states that are permissive (H3K4me3) or repressive (H3K27me3) for transcription, at single-cell resolution with an average coverage of up to 10,000 unique loci per cell. The microfluidics workflow includes live monitoring of

[1]HiFiBiO SAS, Paris, France. [2]Laboratoire de Biochimie, CNRS UMR8231, ESPCI Paris, PSL University, Paris, France. [3]CNRS UMR3244, Institut Curie, PSL University, Paris, France. [4]Translational Research Department Institut Curie, PSL University, Paris, France. [5]Institut Curie Genomics of Excellence Platform, PSL University, Paris, France. [6]INSERM U932, Institut Curie, PSL University, Paris, France. [7]Surgical Department, Institut Curie, Paris, France. [8]HiFiBiO Inc., Cambridge, MA, USA. [9]Present address: Broad Institute of MIT and Harvard, Cambridge, MA, USA. [10]Present address: Unicancer, Paris, France. [11]Present address: Institut Pasteur, Paris, France. [12]Present address: Bayer AG, Leverkusen, Germany. [13]These authors jointly supervised this work: Andrew D. Griffiths, Céline Vallot, Annabelle Gérard. *e-mail: andrew.griffiths@espci.fr; celine.vallot@curie.fr; a.gerard@hifibio.com

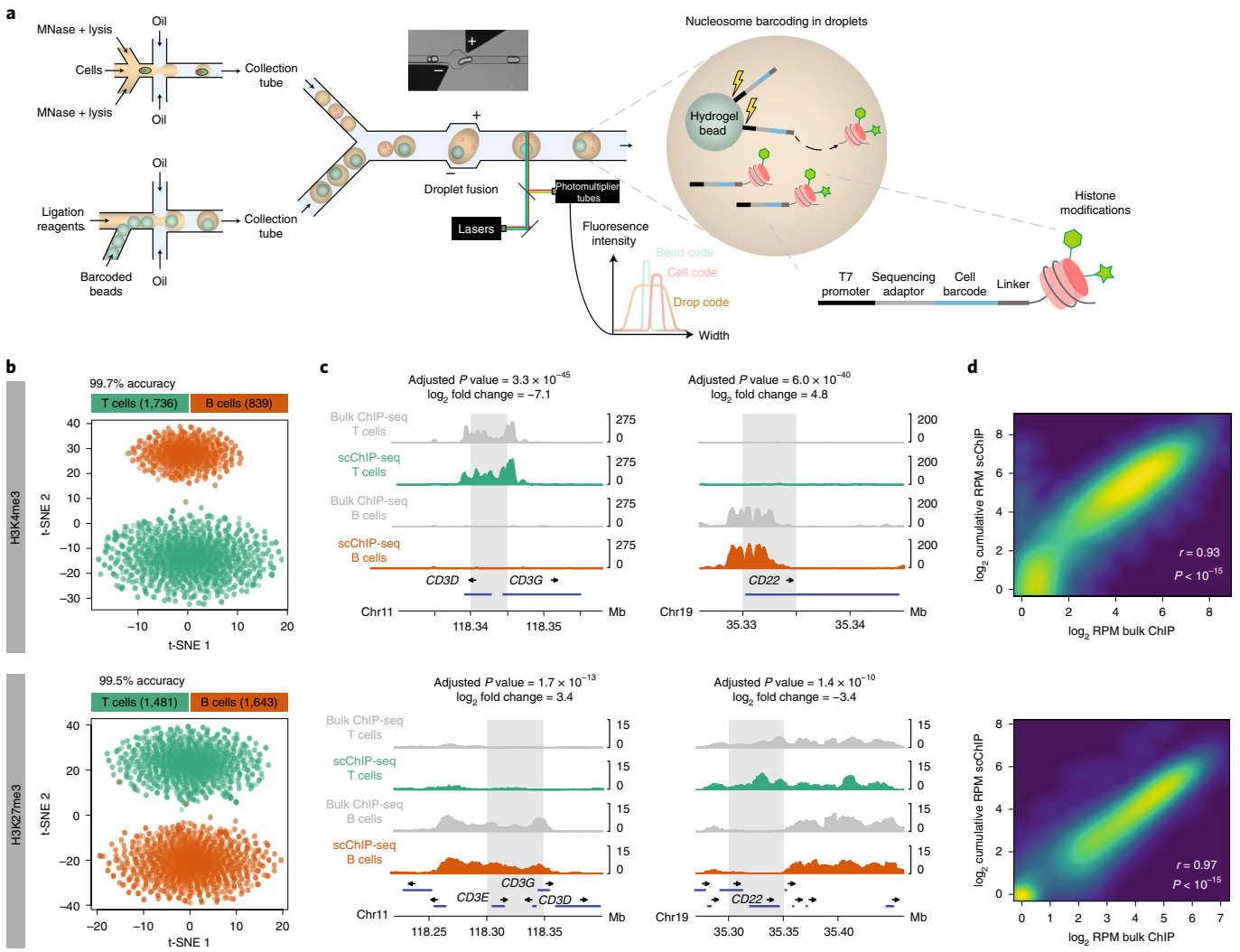

Fig. 1 | Reconstructing cell-type-specific chromatin states from single-cell ChIP-seq profiles. a, Overview of the microfluidics scChIP-seq workflow (see Methods). **b**, t-SNE plots representing the H3K4me3 and H3K27me3 scChIP-seq datasets from human B and T lymphocytes separately indexed in droplets using hydrogel beads carrying both single-cell and cell-type-specific barcodes. The points are colored according to the cell-type-specific barcode sequence. Accuracy represents the agreement between the classification by consensus clustering of scChIP-seq data (Supplementary Fig. 8a) and known cell identity, assessed by the cell-type-specific barcodes. **c**, Snapshots of differentially enriched loci (Supplementary Fig. 8b) of bulk profiles along with cumulative single-cell profiles for each cell type. Differentially bound regions identified using a two-sided Wilcoxon signed-rank test are indicated in gray with the corresponding adjusted *P* values and $\log_2$ fold change (H3K4me3: $n = 1{,}736$ T cells and 839 B cells; H3K27me3: $n = 1{,}481$ T cells and 1,643 B cells). **d**, Scatterplots displaying $\log_2$ RPM (reads per million mapped reads) enrichments in cumulative single-cell versus bulk ChIP-seq data, calculated within 5 kb genomic bins for H3K4me3 ($n = 642{,}098$ bins) and 50 kb for H3K27me3 ($n = 64{,}455$ genomic bins) (see Methods). Pearson's correlation scores and *P* values were computed genome-wide.

droplets (Supplementary Fig. 5a,b), which enables, in three steps, the controlled production of 145 ± 10 pl droplets containing both nucleosomes from individual cells and a hydrogel bead carrying barcoded adapters (Fig. 1a and Supplementary Videos 1 and 2). The system allows coencapsulation of approximately 33% of the input number of cells with a single hydrogel bead, that is, typically 5,000 cells out of 15,000 starting cells (Supplementary Fig. 5c). To confirm that barcodes were unique to a single cell, we performed an experiment with a mixture of mouse and human cell lines, which showed that 97% of the barcodes were unambiguously assigned to a single species (Supplementary Fig. 6), which is consistent with the percentage of occupied droplets containing single cells (95%; see Methods).

Detection of single-cell chromatin landscapes in vitro. Next, we validated the efficiency and accuracy of the scChIP-seq procedure to recapitulate cell identity from the single-cell distribution of H3K4me3 and H3K27me3 modifications. Human Ramos (B cell) and Jurkat (T cell) cells were processed separately (as shown in Fig. 1a) by using two independent sets of barcoded adapters, and, after ligation of the adapters in droplets, the barcoded nucleosomes were pooled and immunoprecipitated. For H3K4me3 and H3K27me3, we achieved an average coverage of 1,630 and 1,633 unique reads per cell, respectively, and a high correlation across technical and biological replicates (see Methods; Supplementary Fig. 7a–c, $r = 0.96$ and 0.98 with $P < 10^{-15}$, respectively), with no batch effect (Supplementary Fig. 7d,e). For both single-cell chromatin profiling experiments, we identified by consensus clustering two stable clusters corresponding to each cell line (Fig. 1b and Supplementary Fig. 8a), matching cell identity with a specificity of over 99.7% and 99.5%, respectively, for the H3K4me3 and H3K27me3 profiles. Aggregated single-cell

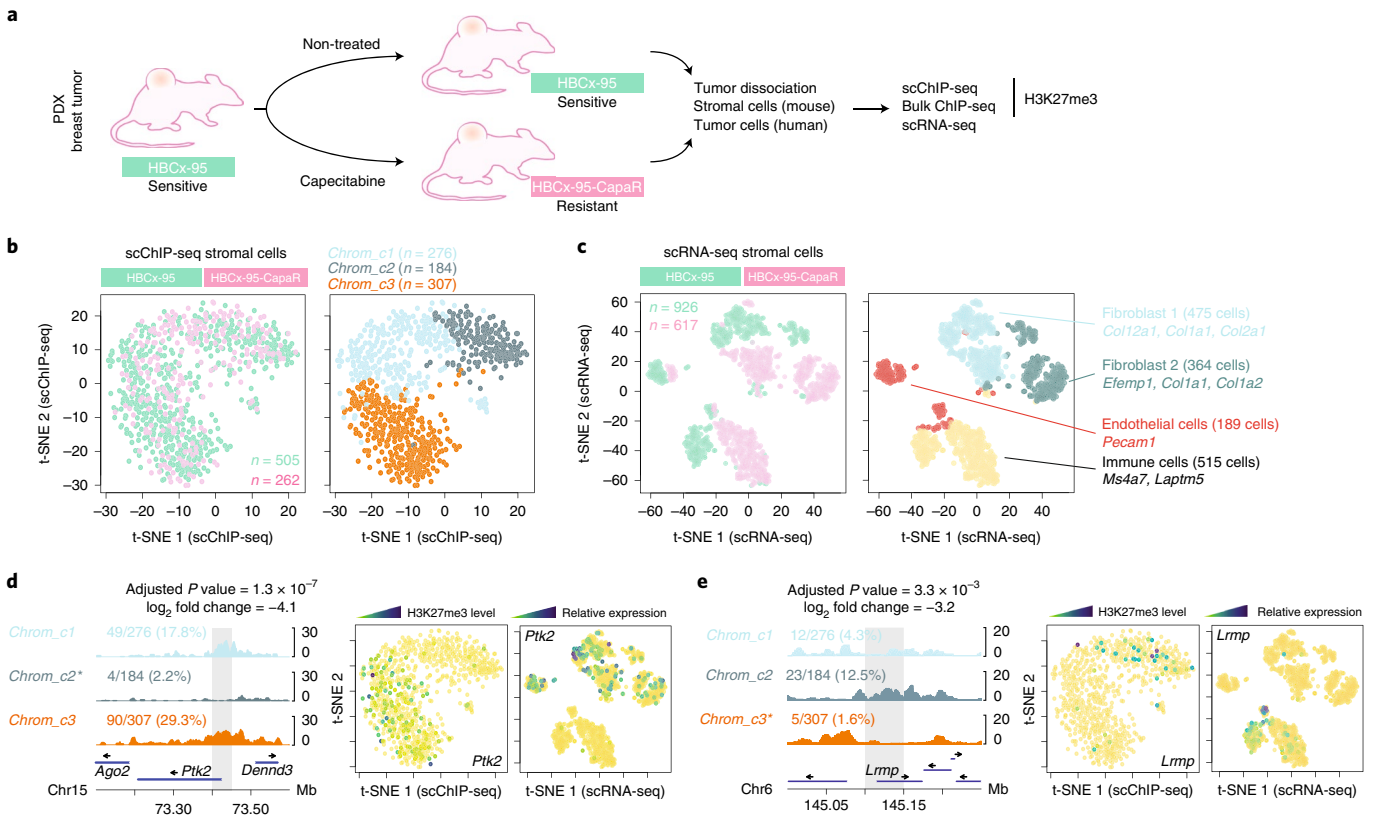

Fig. 2 | Single-cell ChIP-seq profiling of mouse stromal cells reveals cell identities from H3K27me3 chromatin landscapes. a, Scheme representing the generation of a PDX model of acquired resistance to chemotherapy (Supplementary Fig. 9a) and experiments. The sample color code is green for HBCx-95 and pink for HBCx-95-CapaR. **b**, t-SNE representation of scChIP-seq datasets, colored according to tumor of origin (sensitive or resistant to capecitabine, left panel) or consensus clustering results (Supplementary Fig. 10g). **c**, t-SNE representation of scRNA-seq datasets, colored according to tumor of origin or consensus clustering results (Supplementary Fig. 12b). Marker genes identified by differential expression analysis are indicated for each subpopulation (Supplementary Fig. 12c). **d**,**e**, Left panels: snapshots of genomic regions, significantly depleted in H3K27me3 in cells from cluster *Chrom_c2* (*Ptk2*) or *Chrom_c3* (*Lrmp*) versus other cells, with cumulative single-cell profiles for each cell type and cluster ($n = 276$, 184 and 307 single cells for *Chrom_c1*, *Chrom_c2* and *Chrom_c3*, respectively). The number and percentage of cells with H3K27me3 enrichment within each cluster are indicated above the tracks. Differentially bound regions identified using a two-sided Wilcoxon signed-rank test are highlighted in gray with the corresponding adjusted *P* value and $\log_2$ fold change; the reference cluster is indicated with an asterisk. Middle panels: t-SNE representation of scChIP-seq datasets; the points are colored according to the H3K27me3 enrichment signal in each cell for the *Ptk2* and *Lrmp* loci. Right panels: t-SNE representation of scRNA-seq datasets; the points are colored according to the expression signal in each cell for *Ptk2* and *Lrmp*.

profiles recapitulated the bulk ChIP-seq profiles with high accuracy (Fig. 1c,d, $r = 0.93$ and 0.97 with $P < 10^{-15}$ for H3K4me3 and H3K27me3, respectively; Supplementary Fig. 7f). We could identify through differential analysis permissive and repressive chromatin features specific to Ramos and Jurkat cells (Supplementary Fig. 8b). Focusing on H3K4me3, which accumulates around transcription start sites (TSSs), we identified concordant lineage-specific sets of genes as being enriched among chromatin features specific to each cell line (Supplementary Fig. 8c). These results demonstrate that the scChIP-seq procedure is a robust method to detect chromatin features at the single-cell level, to classify single cells with a high accuracy according to their chromatin landscape and to identify discriminating chromatin features between cell populations.

Diversity of H3K27me3 landscapes in stromal cells in vivo. We then used scChIP-seq to interrogate the heterogeneity of chromatin marks within a triple-negative breast tumor model of acquired resistance to chemotherapy. After two cycles of treatment of xenograft derivatives from an individual patient, initially responsive to capecitabine[20], we obtained a tumor with acquired resistance to capecitabine, HBCx-95-CapaR (Supplementary Fig. 9a). We profiled the H3K27me3 landscape at single-cell resolution for both sensitive and resistant xenografts, HBCx-95 and HBCx-95-CapaR (Supplementary Fig. 9b), and also performed scRNA-seq to evaluate transcriptional heterogeneity within the same cell suspension (Fig. 2a). Cumulative single-cell chromatin profiles matched the bulk ChIP-seq profiles (Supplementary Fig. 9c,d, $r = 0.89$ with $P < 10^{-15}$).

We first studied the diversity of chromatin profiles within stromal cells ($n = 1{,}766$ mouse cells with an average coverage of 3,535 unique reads per cell). Consensus clustering approaches (Supplementary Fig. 10) showed that stromal cells stably grouped in three chromatin-based populations according to H3K27me3 profiling, *Chrom_c1*, *Chrom_c2* and *Chrom_c3*, irrespectively of whether the PDX was sensitive or resistant to treatment (Fig. 2b), thus arguing against a potential batch effect. By comparing chromatin features between groups of cells (Supplementary Fig. 11), we identified loci with specific H3K27me3 enrichment for *Chrom_c2* and *Chrom_c3* populations ($n = 1{,}581$ and $n = 1{,}282$, respectively, with adjusted $P < 0.01$ and absolute $\log_2$ fold change > 1; Supplementary Table 1), and to a lesser extent for cluster *Chrom_c1* ($n = 122$).

In parallel, scRNA-seq analysis identified four populations of stromal cells (Fig. 2c and Supplementary Fig. 12): two groups of cells of fibroblast origin (with the specific markers *Col12a1* and *Efemp1*); endothelial cells (*Pecam1*); and macrophage cells (*Ms4a7*).

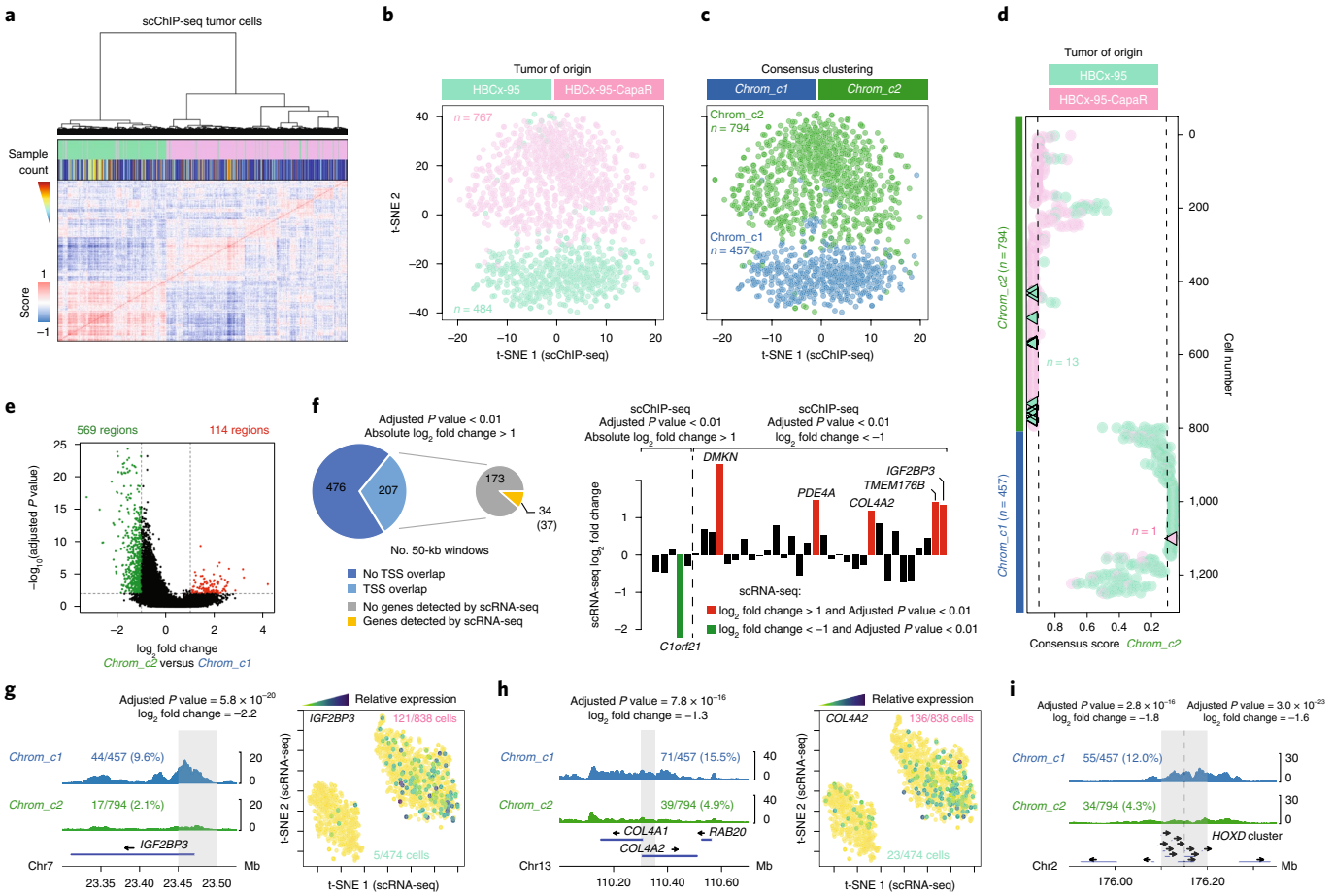

Fig. 3 | Sensitive and drug-resistant specific H3K27me3 chromatin landscapes in PDX model of triple-negative breast cancer treated with capecitabine. **a**, Hierarchical clustering and corresponding heatmap of cell-to-cell Pearson's correlation scores for the scChIP-seq datasets. The sample color code is green for HBCx-95 ($n=484$ single cells) and pink for HBCx-95-CapaR ($n=767$ single cells) and the unique read count is indicated above the heatmap. **b**,**c**, t-SNE representation of the scChIP-seq datasets; cells are colored according to the tumor of origin (**b**) and consensus clustering segmentation (**c**) (Supplementary Fig. 13c). **d**, Item consensus score with regard to *Chrom_c2;* a score of 1 corresponds to a cell highly representative of the *Chrom_c2* cluster. The dotted lines represent an item consensus score of 0.9 relative to *Chrom_c2* (left line) or *Chrom_c1* (right line). The green cells originate from HBCx-95, the pink cells from HBCx-95-CapaR. The triangles highlight cells with a consensus score > 0.9 and in opposition to their tumor of origin. **e**, Volcano plot representing adjusted *P* values (two-sided Wilcoxon signed-rank test) versus $\log_2$ fold changes for differential analysis comparing chromatin enrichment between *Chrom_c2* and *Chrom_c1* ($n=569$ enriched and 114 depleted regions with thresholds of 0.01 for adjusted *P* value and 1 for absolute $\log_2$ fold change). **f**, Left panel: pie chart representing the number of significantly differentially enriched windows (two-sided Wilcoxon signed-rank test) overlapping a TSS and with detectable transcription. Right panel: $\log_2$ expression fold change between cells from HBCx-95-CapaR and HBCx-95 for all detected genes ($n=37$) within differentially enriched loci. The barplot is colored according to the $\log_2$ fold change and the associated adjusted *P* value (black for adjusted *P* values > 0.01, green for significantly underexpressed and red for significantly overexpressed). **g**,**h**, Left panels: the aggregated H3K27me3 chromatin profiles for each cluster are shown for *IGF2BP3* and *COL4A2* ($n=457$ and 794 single cells for *Chrom_c1* and *Chrom_c2*, respectively). Differentially bound regions identified using a two-sided Wilcoxon signed-rank test are highlighted in gray with the corresponding adjusted *P* value and $\log_2$ fold change. The number and the percentage of cells with H3K27me3 enrichment within each cluster are indicated above the tracks. Right panels: t-SNE plots representing scRNA-seq datasets; the points are colored according to the cell expression signal for *IGF2BP3* or *COL4A2*. **i**, Aggregated H3K27me3 chromatin profiles for the *HOXD* locus depleted in H3K27me3 in *Chrom_c2*, but with no detectable transcription with scRNA-seq ($n=457$ and 794 single cells for *Chrom_c1* and *Chrom_c2*, respectively). Differentially bound regions identified using a two-sided Wilcoxon signed-rank test are highlighted in gray with the corresponding adjusted *P* value and $\log_2$ fold change. The number and the percentage of cells with H3K27me3 enrichment within each cluster are indicated above the tracks.

To further compare the identity of populations inferred from both approaches, we focused on genes with a TSS located within 1 kb of chromatin features specific to *Chrom_c1*, *Chrom_c2* and *Chrom_c3*. Loci specifically devoid of H3K27me3 in cells from *Chrom_c2*, that is, permissive for transcription only in *Chrom_c2*, were significantly enriched in genes involved in epithelial-to-mesenchymal transition (adjusted $P=2.8\times10^{-3}$), such as *Col4a1* (Supplementary Fig. 11b), or in apical junction (adjusted $P=9.0\times10^{-2}$), such as *Ptk2* (Fig. 2d), both signatures being characteristic of fibroblast expression programs. Similarly, we found loci devoid of H3K27me3 specific to *Chrom_c3* enriched in genes from immune expression programs (adjusted $P=5.2\times10^{-2}$; Supplementary Fig. 11a), such as *Lrmp* (Fig. 2e). The scRNA-seq data further confirmed the expression pattern of these markers within the PDX samples (right panels, Fig. 2d,e). We could not identify relevant genes associated to the few chromatin marks characteristic of *Chrom_c1*, either suggesting that these cells were less efficiently captured with our scChIP-seq procedure or that this cluster of cells

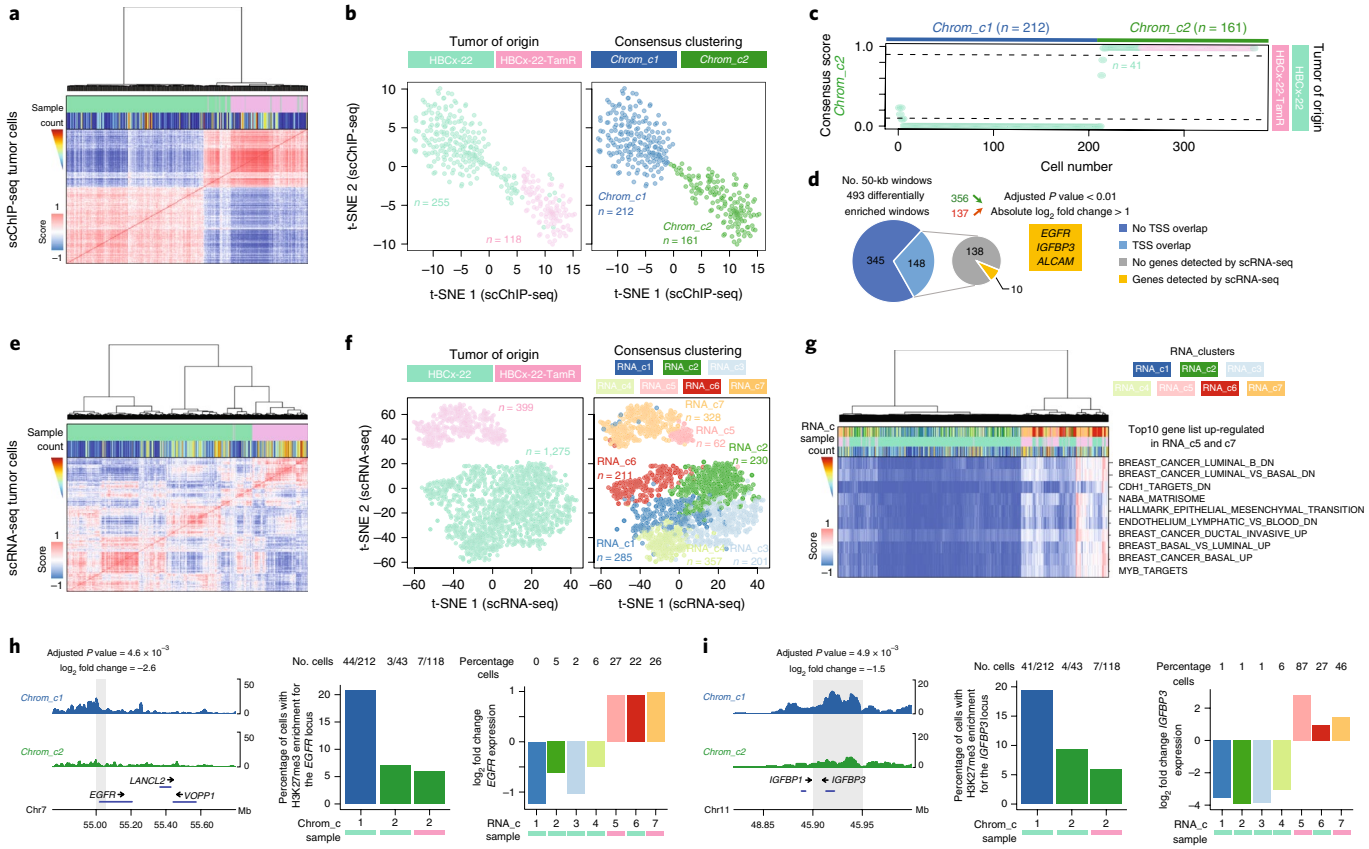

Fig. 4 | A fraction of cells from sensitive tumor share H3K27me3 chromatin features with resistant cells in a model of luminal estrogen receptor-positive PDX treated with tamoxifen. a, Hierarchical clustering and corresponding heatmap of cell-to-cell Pearson's correlation scores for scChIP-seq datasets. The tumor of origin is indicated in green for HBCx-22 ($n = 255$ single cells) and pink for HBCx-22-TamR ($n = 118$ single cells); the unique read count is indicated above the heatmap. **b**, t-SNE representation of the scChIP-seq datasets; cells are colored according to the tumor of origin (left) and consensus clustering segmentation (right) (Supplementary Fig. 14c,d). **c**, Item consensus score with regard to *Chrom_c2*. A score of 1 corresponds to a cell highly representative of the *Chrom_c2* cluster. The dotted lines represent an item consensus score of 0.9 relative to *Chrom_c2* (upper line) or *Chrom_c1* (lower line). **d**, Pie chart representing the number of significantly differentially enriched (adjusted *P* value < 0.01, two-sided Wilcoxon signed-rank test, and absolute $\log_2$ fold change >1) windows overlapping a TSS and with detectable transcription. **e**, Hierarchical clustering and corresponding heatmap of cell-to-cell Pearson's correlation scores for the scRNA-seq datasets. The tumor of origin is indicated in green for HBCx-22 ($n = 1,275$ single cells) and pink for HBCx-22-TamR ($n = 399$ single cells); the UMI count is indicated above the heatmap. **f**, t-SNE representation of the scRNA-seq datasets; the cells are colored according to the tumor of origin (left) and consensus clustering segmentation (right) (Supplementary Fig. 14f). **g**, Hierarchical clustering of average expression scores per cell for each of the top ten upregulated pathways (with the lowest adjusted *P* values) in HBCx-22-TamR versus HBCx-22. The tumor of origin, RNA cluster and unique read count are indicated above the heatmap. **h**,**i**, Left panels: snapshots for *EGFR* and *IGFBP3* loci of aggregated H3K27me3 chromatin profiles for each cluster ($n = 212$ and 161 single cells for *Chrom_c1* and *Chrom_c2*, respectively). Differentially bound regions identified using a two-sided Wilcoxon signed-rank test are highlighted in gray with the corresponding adjusted *P* value and $\log_2$ fold change. Middle panels: barplots displaying the percentage of cells with H3K27me3 enrichment in each cluster. The corresponding number of cells is indicated above the barplots. For each cluster, the origin of the cells (green for HBCx-22 and pink for HBCx-22-TamR) is indicated below. Right panels: barplots displaying the average $\log_2$ fold change in *EGFR* and *IGFBP3* expression levels for cells in each cluster versus all remaining cells. The percentage of cells, within each cluster, with detectable *EGFR* or *IGFBP3* expression is indicated above the barplot. For each cluster, the origin of the cells (green for HBCx-22 and pink for HBCx-22-TamR) is indicated below.

shared chromatin features with both *Chrom_c2* and *Chrom_c3*. Indeed, half of the cells from this cluster share with immune-like cells H3K27me3 enrichment for *Ptk2* (Fig. 2d). Altogether our scChIP-seq approach revealed the existence of three H3K27me3 chromatin landscapes within mouse stromal cells, two of which matched the transcriptomic signatures identified with scRNA-seq (fibroblast and immune-like signatures).

Heterogeneity of chromatin landscapes in breast tumors. Next, we studied the heterogeneity of chromatin profiles among tumor cells from the same pair of triple-negative breast tumor samples ($n = 4,331$ cells from HBCx-95 and HBCx-95-CapaR, with average coverage of 5,161 unique reads per cell). We removed from the analysis loci affected by copy number variations, as identified from bulk DNA profiles, to focus on chromatin alterations (Supplementary Fig. 13a). Based on both chromatin and transcriptomic profiles, cells clustered primarily according to their sensitive or resistant tumor origin (Fig. 3a–c and Supplementary Fig. 13b,c). While the chromatin profiles of sensitive cells were largely homogeneous, distinct chromatin states within the resistant population were apparent (Fig. 3a), suggesting that heterogeneous populations of resistant cells, with distinct chromatin features, emerged. However, consensus clustering also showed that 3% of cells from the untreated tumor ($n = 13$ out of 484) robustly classify with resistant cells (Fig. 3d, consensus score > 0.9), suggesting that they share common chromatin features. Resistant-like and resistant cells, corresponding to

Chrom_c2, displayed a high number of loci depleted in H3K27me3 compared to sensitive cells from *Chrom_c1* (Fig. 3e,f, $n=569$ loci with depleted versus 114 with enriched H3K27me3, adjusted $P<0.01$ and absolute $\log_2$ fold change >1, 30% overlapping a TSS; Supplementary Table 2). Loci specifically devoid of H3K27me3 in cells from *Chrom_c2* were enriched in genes that are targets of the Polycomb complex (Supplementary Fig. 13d), indicating that we were observing a demethylation of expected enhancer of zeste homolog 2 targets. We could only detect transcription within 5% of these loci, either due to the absence of transcription or to insufficient sensitivity of the scRNA-seq procedure. Within these loci, six genes were significantly deregulated according to scRNA-seq, and all in accordance to their H3K27me3 chromatin states (Fig. 3f and Supplementary Fig. 13e). We also identified *IGF2BP3*, a gene known to promote resistance to chemotherapy[21] (Fig. 3g) and regions with markers of epithelial-to-mesenchymal transition (*COL4A2*, *HOXD* cluster; Fig. 3h,i), which induces resistance to chemotherapy[22,23].

Detection of a 'resistant-like' chromatin subclone in untreated tumor. In addition, we profiled a pair of luminal estrogen receptor-positive breast PDXs: HBCx-22, responsive to tamoxifen; and HBCx-22-TamR, resistant to tamoxifen[24]. To obtain a high average coverage of 10,228 unique reads per cell, we limited the number of encapsulated cells ($n=822$ tumor cells; Supplementary Fig. 14a,b). Tumor cells displayed two major chromatin profiles related to their tumor of origin. However, 16% ($n=41$ out of 255) of cells within the sensitive tumor shared chromatin features with all resistant cells (Fig. 4a–c and Supplementary Fig. 14c). Thus, chromatin features characteristic of resistant cells were already found in rare cells from the sensitive tumor.

Differential analysis of chromatin features revealed that resistant-like and resistant cells (*Chrom_c2*) have predominantly lost H3K27me3 marks compared to sensitive-like cells (Supplementary Fig. 14d; $n=356$ loci with depleted versus 137 with enriched H3K27me3, Supplementary Table 3). Loci specifically devoid of H3K27me3 in cells from *Chrom_c2* were enriched in gene targets of the Polycomb complex and genes of basal-like signatures of the mammary epithelium (Supplementary Fig. 14e). With scRNA-seq, we could only detect transcription in 2% of differentially enriched windows, and significant differential expression for three genes, all showing transcription activation in a fraction of resistant cells, mirroring their loss of H3K27me3 enrichment: *EGFR*, a gene implicated in resistance to tamoxifen[25,26]; *IGFBP3;* and *ALCAM* (Fig. 4d,h,i and Supplementary Fig. 14g).

Parallel scRNA-seq analysis of the same cell suspension identified several clusters within the resistant and sensitive tumor (Fig. 4e,f and Supplementary Fig. 14f). While no cells from the sensitive tumor clustered with the resistant cells, we show that cells from the RNA_c6 cluster, originating from the sensitive tumor (corresponding to 211 out of 1,275 cells (17%)), display activation of pathways characteristic of resistant tumor cells, among which basal-like gene signatures and signature of epithelial-to-mesenchymal transition (Fig. 4g). These observations independently suggest that non-genetic features common to resistant cells, either at the transcriptomic or chromatin level, are already found in cells from the sensitive tumor. Both single-cell measurements point toward the activation of basal-like gene signatures, but through different sets of genes.

Discussion

Profiling histone modifications at the single-cell level with high coverage, up to 10,000 loci on average per cell, was instrumental to reveal the presence of relatively rare chromatin states within tumor samples. Our scChIP-seq approach enables the segmentation of cell populations solely based on their chromatin landscape, and the identification of key chromatin features of each subpopulation. Using this approach, we identified loci depleted for the transcriptional repressive mark H3K27me3 in a population of resistant and resistant-like cells, including genes known to promote resistance to chemotherapy or targeted therapy, highlighting the potential to discover new drug targets and biomarkers for patient stratification. Notably, both models share the H3K27 demethylation, and transcription activation of a gene of the *IGFBP* family, in the insulin-like growth factor signaling pathway, which plays a key role in breast cancer and drug resistance[27]. Previous studies have reported epigenomic reprogramming of cancer cells following endocrine therapies[28,29], but this study suggests that rare cells with chromatin features characteristic of resistant cancer cells exist before treatment. In the future, it will be interesting to further probe the potential role of spontaneous heterogeneity of chromatin states in untreated cells in the acquisition of drug resistance.

In our model of acquired resistance to tamoxifen, scRNA-seq and scChIP-seq revealed the existence of a fraction of cells within the sensitive tumor with transcriptional and epigenetic features common to resistant cells. Interestingly, the two approaches did not identify the same subset of genes, but revealed common gene signatures characteristic of a switch from luminal to basal cell identity. Such differences could highlight the poor sensitivity of current scRNA-seq approaches or indicate that a fraction of genes is poised for transcription, but not transcribed. Loss of repressive chromatin marks such as H3K27me3 could change the chromatin to a permissive state and might correspond to a priming event preceding changes in transcription. Altogether, our scChIP-seq system can probe the role of heterogeneity and dynamics of chromatin states within any complex biological system; in addition to cancer, it can be applied to other diseases and healthy systems, notably to study cellular differentiation and development.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41588-019-0424-9.

Received: 19 March 2019; Accepted: 19 April 2019;
Published online: 31 May 2019

References

1. Roth, A. et al. Clonal genotype and population structure inference from single-cell tumor sequencing. *Nat. Methods* **13**, 573–576 (2016).
2. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
3. McGranahan, N. & Swanton, C. Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell* **168**, 613–628 (2017).
4. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **15**, 81–94 (2018).
5. Salgia, R. & Kulkarni, P. The genetic/non-genetic duality of drug 'resistance' in cancer. *Trends Cancer* **4**, 110–118 (2018).
6. Gardner, E. E. et al. Chemosensitive relapse in small cell lung cancer proceeds through an EZH2-SLFN11 axis. *Cancer Cell* **31**, 286–299 (2017).
7. Shaffer, S. M. et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).
8. Rathert, P. et al. Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature* **525**, 543–547 (2015).
9. Kim, C. et al. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell* **173**, 879–893.e13 (2018).
10. Horning, A. M. et al. Single-cell RNA-seq reveals a subpopulation of prostate cancer cells with enhanced cell-cycle-related transcription and attenuated androgen response. *Cancer Res.* **78**, 853–864 (2018).
11. Mazor, T. et al. DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. *Cancer Cell* **28**, 307–317 (2015).
12. Aryee, M. J. et al. DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. *Sci. Transl. Med.* **5**, 169ra10 (2013).
13. Clark, S. J. et al. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat. Protoc.* **12**, 534–547 (2017).

14. Angermueller, C. et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13**, 229–232 (2016).
15. Li, L. et al. Single-cell multi-omics sequencing of human early embryos. *Nat. Cell Biol.* **20**, 847–858 (2018).
16. Cusanovich, D. A. et al. The *cis*-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018).
17. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
18. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
19. Rotem, A. et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172 (2015).
20. Marangoni, E. et al. Capecitabine efficacy is correlated with TYMP and RB1 expression in PDX established from triple-negative breast cancers. *Clin. Cancer Res.* **24**, 2605–2615 (2018).
21. Lederer, M., Bley, N., Schleifer, C. & Hüttelmaierr, S. The role of the oncofetal IGF2 mRNA-binding protein 3 (IGF2BP3) in cancer. *Semin. Cancer Biol.* **29**, 3–12 (2014).
22. Zheng, X. et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530 (2015).
23. Fischer, K. R. et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).
24. Cottu, P. et al. Acquired resistance to endocrine treatments is associated with tumor-specific molecular changes in patient-derived luminal breast cancer xenografts. *Clin. Cancer Res.* **20**, 4314–4325 (2014).
25. Massarweh, S. et al. Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function. *Cancer Res.* **68**, 826–833 (2008).
26. Ciupek, A. et al. Androgen receptor promotes tamoxifen agonist activity by activation of EGFR in ERα-positive breast cancer. *Breast Cancer Res. Treat.* **154**, 225–237 (2015).
27. Denduluri, S. K. et al. Insulin-like growth factor (IGF) signaling in tumorigenesis and the development of cancer drug resistance. *Genes Dis.* **2**, 13–25 (2015).
28. Nguyen, V. T. et al. Differential epigenetic reprogramming in response to specific endocrine therapies promotes cholesterol biosynthesis and cellular invasion. *Nat. Commun.* **6**, 10044 (2015).
29. Magnani, L. et al. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *Proc. Natl Acad. Sci. USA* **110**, E1490–E1499 (2013).

Acknowledgements

We thank members of the Griffiths and Vallot laboratories and HiFiBiO for helpful discussions. We thank B. E. Bernstein (Massachusetts General Hospital, Broad Institute) and R. Nicol (Whitehead Institute, Broad Institute) for advice on technology development, J. Waterfall (Institut Curie) and B.E.B. for critical discussion on the manuscript and the Institut Pierre-Gilles de Gennes for the use of the clean room facilities. We also thank patients for their approval to use their tumor samples. This work was supported by the Investissements d'Avenir program via Agence Nationale de la Recherche grant nos. ANR-10-NANO-02, ANR-10-IDEX-0001-02 PSL, ANR-10-LABX-31, ANR-10-EQPX-34, ANR-10-EQPX-03, ANR-14-CE10-0013_CellectChIP and ANR-10-INBS-09-08 (France Génomique Consortium). K.G. was supported by a Cifre PhD funded by ANRT and HiFiBiO, fellowship no. 2015/0159. C.V. was supported by the ATIP-Avenir program, by the Plan Cancer program and by the Institut Curie SiRIC program, SiRIC grant nos. INCa-DGOS- 4654 and INCa-DGOS-Inserm_12554. The scRNA-seq experiments were supported by the PSL University grant 'Single-Cell Sequencing'. Next-generation sequencing was performed by the ICGex platform of the Institut Curie and the iGenSeq platform of the Institut du Cerveau et de la Moelle épinière.

Author contributions

The scChIP-seq experimental method was conceived by K.G., C.B. and A.G., developed by K.G. and A.G. for the cellular and molecular biology and developed by K.G., A.D.G and A.G. for the microfluidics. The scChIP-seq analytical methods were conceived and developed by K.G. and C.V.; K.G., C.B., A.D.G., C.V. and A.G. designed the experiments. K.G., A.D., F.N., A.D., A.P. and O.F. conducted the experiments. The PDX experiments were designed and interpreted by J.M., F.R., A.D. and E.M.; S.L. performed the sequencing experiments. Y.P. performed the barcoded bead preparation and quality controls. M.R. contributed to the microfluidics chip design. K.G., A.W. and C.V. performed the data analysis. K.G., A.D.G., C.V. and A.G. wrote the manuscript with input from all authors.

Competing interests

Patents have been filed on some aspects of this work. The inventors may receive payments related to the exploitation of these patents under their employer's rewards to inventor scheme. A.D.G. and C.B. are cofounders of HiFiBiO.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0424-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.D.G., C.V. or A.G.

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

Methods

Cell lines. Jurkat cells (T18-125; ATCC), an immortalized line of human T lymphocytes, and Ramos cells (CRL-1596; ATCC), an immortalized line of human B lymphocytes, were grown in Roswell Park Memorial Institute (RPMI) 1640 medium (catalog no. 61870-010; Thermo Fisher Scientific) supplemented with 10% heat-inactivated fetal bovine serum (catalog no. 16140071; Thermo Fisher Scientific) and 1% penicillin-streptomycin (catalog no. 15140122; Thermo Fisher Scientific). Mouse M300.19 cells (gift from B. Moser), an immortalized line of mouse pre-B lymphocytes, were grown in RPMI 1640 medium supplemented with 10% fetal bovine serum (HyClone SH30070.03; Thermo Fisher Scientific), 1% penicillin-streptomycin, 1% Gibco L-glutamine (catalog no. 25030081; Thermo Fisher Scientific) and 5×10^{-5} M Gibco 2-mercaptoethanol (catalog no. 21985023; Thermo Fisher Scientific).

PDXs. Female Swiss nude mice were purchased from Charles River Laboratories and maintained under specific-pathogen-free conditions. Their care and housing were in accordance with institutional guidelines and the rules of the French Ethics Committee (project authorization no. 02163.02). A PDX model of luminal breast cancer (HBCx-22) was previously established at Institut Curie from untreated, early-stage luminal breast cancer with informed consent from the patient[30]. Acquisition of a resistant phenotype for a derivative of HBCx-22, HBCx-22-TamR, was previously established and maintained as described previously[24]. A PDX from a residual triple-negative breast cancer post-neoadjuvant chemotherapy (HBCx-95) was previously established at Institut Curie with informed consent from the patient[20,24]. Mice with HBCx-95 xenografts ($n=6$) were treated with capecitabine (Xeloda; Roche Laboratories) orally at a dose of 540 mg kg d^{-1}, 5 d a week for 6 weeks. Relative tumor volumes (mm^3) were calculated as described previously[31] (Supplementary Fig. 9a). Mice with recurrent tumors were treated for a second round of capecitabine when the PDX reached a volume > 200 mm^3 (mice nos. 35, 40 and 33). Mouse no. 40 did not respond to capecitabine; the PDX specimen was extracted at 1,100 mm^3 and tagged as HBCx-95-CapaR.

Before scChIP-seq, scRNA-seq and bulk ChIP-seq, PDXs were digested at 37 °C for 2 h with a cocktail of collagenase I (catalog no. 11088793001; Roche) and hyaluronidase (catalog no. H3506; Sigma-Aldrich) as described previously[32]. Cells were further individualized at 37 °C using a cocktail of 0.25% trypsin-Versen (catalog no. 15040-033; Thermo Fisher Scientific), dispase II (catalog no. D4693; Sigma-Aldrich) and DNase I (catalog no. 11284932001; Roche). eBioscience red blood cell lysis buffer (catalog no. 00-4333-57; Thermo Fisher Scientific) was then added to degrade red blood cells. To increase the viability of the cell suspension, dead cells were removed using the Dead Cell Removal Kit (Miltenyi Biotec). Cells were resuspended in PBS/0.04% UltraPure BSA (catalog no. AM2616; Thermo Fisher Scientific).

scChIP-seq. The overall workflow is shown in Supplementary Fig. 1.

Microfluidics chips. Four microfluidics chips were used: (1) to compartmentalize single cells with lysis reagents and micrococcal nuclease in droplets; (2) to produce hydrogel beads; (3) to compartmentalize single hydrogel beads in droplets; and (4) for one-to-one fusion of droplets containing digested nucleosomes (from single lysed cells) with droplets containing single hydrogel beads (Supplementary Fig. 2). All chips were manufactured using soft photolithography in polydimethylsiloxane[33] (Sylgard) as described[34]. Masters were made using one layer of SU-8 photoresist (MicroChem). The list depth of the photoresist layer for device I was 40.8 ± 1 µm, for device II was 30.0 ± 1 µm and for device III was 34.0 ± 1 µm. For device IV, the list depth was 45.0 ± 1 µm and electrodes were prepared by melting a 51 In/32.5 Bi/16.5 Sn alloy (Indium Corporation) into the electrode channels[35]. Microfluidics devices were treated the day of the experiment with 1% v/v 1H,1H,2H,2H-perfluorodecyltrichlorosilane (catalog no. AB111155; abcr) in Novec HFE7100 fluorinated oil (3M) to prevent droplets wetting the channel walls.

Microfluidics operations. Droplet formation, fusion and fluorescence analysis were performed on a dedicated droplet microfluidics station, similar to Mazutis et al.[34]. The continuous oil phase for all droplet microfluidics experiments was Novec HFE-7500 fluorinated oil (3M) containing 2% w/w 008-FluoroSurfactant (RAN Biotechnologies).

Cell compartmentalization and chromatin digestion. Cells were centrifuged (300*g*, 5 min at 4 °C), labeled by 20 min incubation with 1 µM Calcein AM (catalog no. C3099; Thermo Fisher Scientific) and resuspended in cell suspension buffer, comprising DMEM/F12 (Thermo Fisher Scientific) supplemented with 30% Percoll (catalog no. P1644; Sigma-Aldrich), 0.1% Pluronic F-68 non-ionic surfactant (catalog no. 24040032; Thermo Fisher Scientific), 25 mM HEPES pH 7.4 (catalog no. 15630080; Thermo Fisher Scientific) and 50 mM NaCl. Cells were resuspended to give an average number of cells per droplets λ of 0.1, resulting in 90.48% of empty droplets, 9.05% of droplets containing one cell and only 0.46% containing two or more cells due to the Poisson distribution of the cells in droplets[36]. Overall, we determined experimentally that among non-empty droplets, 95.16% contained one cell and 4.84% contained two or more cells, close to the expected values (94.92 and 5.08%, respectively). The cells were co-flowed in a microfluidics chip (Supplementary Fig. 2) with digestion buffer containing lysis buffer (107.5 mM Tris-HCl pH 7.4, 322.5 mM NaCl, 2.15% Triton X-100, 0.215% deoxycholate and 10.75 mM $CaCl_2$), 2 µM Sulforhodamine B sodium salt (catalog no. S1402-5G; Sigma-Aldrich), 4 µM DY-405 (catalog no. 405-00; Dyomics), protease inhibitor cocktail and 0.2 U $µl^{-1}$ micrococcal nuclease enzyme (catalog no. EN0181; Thermo Fisher Scientific). Droplets were produced by hydrodynamic flow focusing[37] with a nozzle 25 µm wide, 40 µm deep and 40 µm long. The flow rates (150 µl h^{-1} for both aqueous phases, 850 µl h^{-1} for the continuous oil phase) were calibrated to produce 45 pl droplets. The droplets were collected in a collection tube (1.5 ml Eppendorf tube filled with HFE-7500 fluorinated oil) and then incubated at 37 °C for 20 min.

Production of hydrogel beads carrying barcoded DNA adapters. Hydrogel beads carrying barcoded DNA adapters were produced by split-mix synthesis using a method similar to that described previously[38,39]. Briefly, polyethylene diacrylate (PEG-DA) hydrogel beads containing streptavidin acrylamide were produced; barcoded primers were added to the beads by split-and-pool synthesis using ligation. PEG-DA hydrogel beads were produced using the microfluidics device indicated in Supplementary Fig. 2, essentially as described by Zilionis et al.[38]. The 9 pl droplets were produced at a 4.5 kHz frequency and were exposed at 200 mW cm^{-2} with a 365 nm ultraviolet light source (ac475-365; OmniCure) to trigger gel bead polymerization. Recovered gel beads were washed ten times with washing buffer (100 mM Tris pH 7.4, 0.1% v/v TWEEN 20). Twelve million PEG-DA beads were incubated in a 500 µl final volume for 1 h at room temperature with 50 µM of a photo-cleavable biotinylated double-stranded DNA oligonucleotide (see SEQ1 in Supplementary Table 4) and then distributed into a 96-well microplate, each well containing 5 µl at 5 µM of a double-stranded DNA with a specific first index (index 1), for split-and-pool synthesis by ligation using T7 DNA ligase (catalog no. M0318; New England Biolabs) according to the manufacturer's instructions. At each round of split-and-pool, the hydrogel beads were pooled and washed as described[38]. Repeating this splitting and pooling process three times in total (adding three indexes) results in 96^3 combinations, which generates approximately 8.8×10^5 different barcodes. After adding the last index, the beads were pooled and a common double-stranded DNA oligonucleotide (SEQ2 in Supplementary Table 4) was ligated to the beads. Each bead carries on average approximately 5×10^7 copies of a unique barcode (see Supplementary Fig. 3 for quality controls of the single-cell barcodes).

Compartmentalization of hydrogel beads. The barcoded hydrogel beads were labeled by 30 min incubation with 10 µM Cy5-PEG3 biotin (catalog no. FP-1M1220; Interchim) and washed with washing buffer (100 mM Tris pH 7.4, 0.1% v/v TWEEN 20), then suspended in bead mix (62.5 mM EGTA, 2 mM dNTPs, 1 mM ATP, 0.5 µM Sulforhodamine B). Barcoded hydrogel beads were co-flowed using the microfluidics device indicated in Supplementary Fig. 2, with ligation mix (2× ligation buffer, 2 mM ATP, 1 µM Sulforhodamine B, 100 mM EGTA, 0.38 U $µl^{-1}$ Fast-Link DNA ligation Kit (catalog no. LK0750H; Lucigen) and EndRepair mix (4× ligation buffer, 4 mM deoxyribonucleotide triphosphate, 1 µM Sulforhodamine B, 0.08 U $µl^{-1}$ Fast-Link DNA ligation Kit, 0.15× End-It End-Repair Kit (catalog no. ER0720; Lucigen)). The reinjection of close-packed barcoded hydrogel beads[40] resulted in 65 ± 5% of the droplets containing a single bead. The flow rates (150 µl h^{-1} for the beads, 75 µl h^{-1} for both ligation and End-Repair buffers, 150 µl h^{-1} for the continuous oil phase) were calibrated to produce 100 pl droplets.

Fusion of beads and cell droplets. Droplets containing fragmented chromatin and droplets containing barcoded hydrogel beads were reinjected into a microfluidics device with two aqueous inlets and one oil inlet for droplet fusion (Supplementary Fig. 2). The paired droplets were electrocoalesced[41] using an electrical field generated by applying 100 V AC (square wave) at 5 kHz across electrodes embedded in the microfluidics device; 75 ± 5% of the droplets were correctly paired and fused.

Nucleosome barcoding in droplets. Fused droplets were collected and exposed for 90 s at 200 mW cm^{-2} with a 365 nm ultraviolet light source. The ligation was performed at 16 °C overnight. The emulsion was then broken by adding one volume of 80/20 v/v HFE-7500/1H,1H,2H,2H-perfluoro-1-octanol (catalog no. 370533; Sigma-Aldrich). The aqueous phase containing barcoded nucleosomes was diluted by adding ten volumes of lysis dilution buffer (50 mM Tris-HCl pH 7.4, 1% Triton X-100, 0.1% deoxycholate, 37.5 mM EDTA, 37.5 mM EGTA, 262.5 mM NaCl and 1.25 mM $CaCl_2$) and centrifuged for 10 min at 10,000*g* at 4 °C. The soluble aqueous phase was used for chromatin immunoprecipitation.

Immunoprecipitation of barcoded nucleosomes. Dynabeads Protein A magnetic particles (catalog no. 10001D; Thermo Fisher Scientific) were washed in blocking buffer comprising PBS supplemented with 0.5% TWEEN 20, 0.5% BSA fraction V and incubated for 4 h at 4 °C in 1 ml blocking buffer with 2 µl of anti-H3K4me3 antibody (catalog no. 07-473; EMD Millipore) and 2.5 µl of anti-H3K27me3 antibody (C36B11)(catalog no. 9733; Cell Signaling Technology). After incubation, the particles were suspended with the barcoded nucleosomes and incubated at 4 °C overnight. Magnetic particles were washed as described previosuly[19] and immediately processed to prepare the sequencing library.

Sequencing library preparation and sequencing. Concatemers of barcodes were digested by PacI restriction enzyme (catalog no. R0547; New England Biolabs), following the manufacturer's instructions. Immunoprecipitated chromatin was then treated with RNase A, DNase and protease-free (catalog no. EN0531; Thermo Fisher Scientific) and with Proteinase K (catalog no. EO0491; Thermo Fisher Scientific). DNA was eluted from the magnetic particles with one volume of elution buffer (1% SDS, 10 mM Tris-HCl pH 8, 600 mM NaCl and 10 mM EDTA). Eluted DNA was purified with 1× Agencourt AMPure XP beads (catalog no. A63881; Beckman Coulter) and eluted with RNase/DNase-free water. Barcoded nucleosomes were amplified using in vitro transcription with the MegaScript T7 Transcription Kit (catalog no. AM1334; Thermo Fisher Scientific). The resulting amplified RNA was purified using 1× Agencourt RNAClean XP Kit (catalog no. A66514; Beckman Coulter) and reverse transcribed using SEQ4 (Supplementary Table 4). After RNA digestion, DNA was amplified by PCR using SEQ5 (Supplementary Table 4 and Supplementary Fig. 4). The final product was size-selected by gel electrophoresis; scChIP-seq libraries were sequenced on an NextSeq 500 (Illumina) using Mid Output runs (150 cycles). Cycles were distributed as follows: 50 bp (read no. 1) were assigned for the genomic sequence and 100 bp (read no. 2) were assigned to the barcode. The first four cycles of read no. 2 were dark cycles to prevent low complexity failure during cluster identification.

Bulk ChIP-seq. The bulk ChIP-seq experiments were performed as described previously[42] on 10^6 cells from cell suspensions obtained from HBCx-22, HBCx-22-TamR, HBCx-95 and HBCx-95-CapaR using anti-H3K27me3 antibody; 2 ng of immunoprecipitated and input DNA were used to prepare the sequencing libraries using the Ovation Ultralow V2 DNA-Seq Library Preparation Kit (NuGEN) according to the manufacturer's instructions. Bulk ChIP-seq libraries were sequenced on a HiSeq 2500 (Illumina) in rapid run mode SE50.

scRNA-seq. Approximately 3,000 cells from each cell suspension, HBCx-22, HBCx-22-TamR, HBCx-95 and HBCx-95-CapaR, were loaded on a Chromium Single Cell Controller Instrument (10x Genomics) according to the manufacturer's instructions. Samples and libraries were prepared according to the manufacturer's instructions. Libraries were sequenced on a HiSeq 2500 (Illumina) in rapid run mode, using paired-end 26–98 bp sequencing.

scChIP-seq data analysis. Sequencing data were analyzed with Python v.2.7.12 and R v.3.3.3.

Demultiplexing cellular barcodes. Barcodes were extracted from reads no. 2 by first searching for the constant 4-bp linkers found between the 20-mer indexes of the barcode allowing up to one mismatch in each linker (see Supplementary Fig. 3 and Supplementary Table 5 for the barcode structure). If the correct linkers were identified, the three interspersed 20-mer indexes were extracted and concatenated together to form a 60-bp non-redundant barcode sequence. A library of all 884,736 combinations of the three sets of 96 indices (96^3) was used to map the barcode sequences using the sensitive read mapper CUSHAW v.3.0.3 (ref. [43]). Each set of indexes was error-correcting because it takes more than an edit distance of 3 to convert one index into another. Therefore, we set a total mismatch threshold of 3 across the entire barcode, with two or fewer per index to avoid misassigning sequences to the wrong barcode ID. In a second, slower step, sequences that could not be mapped to the CUSHAW3 index library were split into their individual indexes and each index compared against the set of 96 possible indexes, allowing up to two mismatches in each individual index. Any sequences not assigned to a barcode ID by these two steps were discarded.

Alignment, filtering and normalization. Reads no. 1 were aligned to the mouse mm10 and human hg38 reference genomes using Bowtie v.1.2.2 (ref. [44]) by keeping only reads having no more than one reportable alignment and two mismatches. Raw reads are distributed according to a bimodal distribution (Supplementary Figs. 7, 9 and 14), the lower peak most probably corresponding to droplets with barcoded beads but without cells[18], and the right peak corresponding to droplets with cell + bead, thereby setting a read count cutoff to define barcodes associated to a cell. For subsequent analysis, we kept barcodes with a unique (post-PCR duplicate removal) read count above this cutoff. To remove PCR duplicates, for each barcode (that is, cell), all the reads falling in the same 150-bp window were stacked into one as reads possibly originating from PCR duplicates or from the same nucleosome. We generated a coverage matrix and metrics from these de-duplicated reads, referred to as 'unique reads' in the text.

For each cell, reads were binned in non-overlapping 50 kb bins for H3K27me3, known to accumulate over broad genomic regions, and 5 kb genomic bins for H3K4me3, known to accumulate in narrow peaks around TSSs, spanning the genome, to generate an $n \times m$ coverage matrix with n barcodes and m genomic bins. We combined coverage matrices for each of our four analyses from the following samples: (1) Ramos and Jurkat (Fig. 1); (2) mouse cells from HBCx-95 and HBCx-95-CapaR (Fig. 2); (3) human cells from HBCx-95 and HBCx-95-CapaR (Fig. 3); and (4) human cells from HBCx-22 and HBCx-22-TamR (Fig. 4). We first removed cells with a total number of unique reads within the upper percentile, considered as outliers, and filtered out genomic regions not represented in at least 1% of all cells. By principal component analysis, we could group cells independently of coverage only if cells had at least 1,600 unique reads per cell (Supplementary Fig. 10a). For all subsequent analyses, we excluded cells with lower coverage. Coverage matrices were then normalized by dividing counts by the total number of reads per cell and multiplying by the average number of reads across all cells.

Unsupervised clustering of scChIP-seq profiles. Normalized matrices were reduced by principal component analysis ($n = 50$ first components selected for further analysis). To improve the stability of our clustering approaches, we further limited our analysis to cells displaying a Pearson's pairwise correlation score above a threshold t with at least 1% of cells. Threshold t was defined as the upper percentile of the Pearson's pairwise correlation scores for a randomized dataset (see as an example Supplementary Fig. 10). We used consensus clustering (Bioconductor ConsensusClusterPlus v.1.46.0 package[45]) to examine the stability of the clusters and compute an item consensus score for each cell. We established consensus partitions of the dataset in k clusters (for $k = 2, 3, \ldots$), on the basis of 1,000 resampling iterations (80% of cells) of hierarchical clustering, with Pearson's dissimilarity as the distance metric and Ward's method for linkage analysis. The optimal number of clusters (k) was chosen to maximize the intra-cluster correlation score. Clustering results were visualized with t-distributed stochastic neighbor embedding (t-SNE) plots[46]. To visualize chromatin profiles of subpopulations, we aggregated reads of single cells within each cluster and created enrichment profiles using the R package Sushi v.1.12.0 (ref. [47]).

Differential analysis of scChIP-seq profiles. To identify differentially enriched regions across single cells for a given cluster, we performed a non-parametric two-sided Wilcoxon rank-sum test comparing normalized counts from individual cells from one cluster versus all other cells. We tested for the null hypothesis that the distribution of normalized counts from the two compared groups have the same median, with a 95% confidence interval. We limited our analysis to the windows selected for unsupervised analysis described earlier. P values were corrected for multiple testing using the Benjamini–Hochberg procedure[48]. Genomic regions were considered as 'enriched' or 'depleted' for H3K27me3 or H3K4me3 if adjusted P values were lower than 0.01 and the absolute $\log_2$ fold change > 1. For H3K4me3, for each 5-kb bin, we identified genes with the closest TSS and computed the corresponding distance (0 in the case of bins directly overlapping TSS) using bedtools v.2.17 (ref. [49]) and the reference annotation of the human transcriptome Gencode_hg38_v26, limited to protein_coding, antisense and long non-coding RNA genes. We applied hypergeometric tests to identify gene sets from the Molecular Signatures Database (MSigDB) v.5 (ref. [50]) overrepresented within differentially enriched bins, correcting for multiple testing with the Benjamini–Hochberg procedure. For H3K27me3 scChIP-seq analysis, we used peak annotation from bulk ChIP-seq datasets to further annotate our 50-kb windows and corresponding genes. For each window, we kept for subsequent analyses (scRNA-seq comparison and gene enrichment analyses) genes with a TSS overlapped by a peak in any condition. For mouse cells, we interrogated the c5_GO and c7_hallmark gene lists (converting human gene names to mouse gene names); for human tumor cells, we interrogated the c2_curated and c7_hallmark gene lists.

Bulk ChIP-seq data analysis. Reads were aligned to the mouse mm10 and human hg38 reference genomes using Bowtie v.1.2.2; the tool bamcmp[51] was used to separate human from mouse sequences. Subsequent analysis was performed as explained previously[42]. Only uniquely mapping reads were kept for the analysis; in addition, PCR duplicates were removed using Picard Tools (https://broadinstitute.github.io/picard/). Data were binned in 5-kb (H3K4me3) or 50-kb (H3K27me3) consecutive genomic windows. For each window, $\log_2$ reads per million mapped reads (RPM) were computed.

scRNA-seq data analysis. Single-cell sequencing files were processed using the Cell Ranger Single Cell Software Suite v.1.3.1 to perform quality control, sample demultiplexing, barcode processing and single-cell 3′ gene counting (http://software.10xgenomics.com/single-cell/overview/welcome) using the University of California Santa Cruz mouse (mm10) and human (hg19) transcriptome and genome with default parameters. A total of 2,728 cells with an average coverage of 30,166 reads per cell (1,564 human and 1,191 mouse cells) for HBCx-22, 1,746 cells with an average coverage of 41,166 reads per cell (753 human and 1,013 mouse cells) for HBCx-22-TamR, 1,184 cells with an average coverage of 160,583 reads per cell (545 human and 647 mouse cells) for HBCx-95 and 2,087 cells with an average coverage of 38,345 reads per cell (861 human and 1,242 mouse cells) for HBCx-95-CapaR were analyzed. Further analysis was performed in R v.3.3.3 using custom R scripts. Any cell with more than 10% of mitochondrial unique molecular identifier (UMI) counts was filtered out. We only kept cells with a total UMI count below 100,000 and total detected genes below 6,000 and over 1,000. We then only kept genes with at least one transcript in at least two cells. Using the R package scater v.1.2.0, scRNA-seq count matrices were normalized for coverage and transformed by the relative log expression method[52]. Using annotations from the R package ccRemover v.1.0.4 (ref. [53]), we removed genes related to the cell cycle from subsequent clustering analyses to group cells according to cell identity and

not cell cycle-related phenomena. Barnes–Hut approximation to t-SNE was then performed on the $n = 50$ first principal components to visualize cells in a two-dimensional space. Clusters were identified using consensus clustering as for the scChIP-seq analyses described earlier. We identified genes that were differentially expressed between clusters using edgeR GLM statistical models[54]. For stromal mouse cells, clusters were identified according to the differential expression of hallmark genes. For Fig. 4f, we studied the ten most significantly enriched pathways among overexpressed genes in cells from HBCx-22-TamR versus cells from HBCx-22. For each cell of the HBCx-22 and HBCx-22-TamR samples, we computed an average expression score for each of these ten pathways and performed hierarchical clustering.

Statistics. *Correlation tests.* To compare genome-wide the distribution of H3K27me3 and H3K4me3 enrichment obtained with the bulk versus single-cell approach, we estimated the association between enrichment values for the 50-kb or 5-kb windows, respectively, and tested for the null hypothesis of the association being null. We used the Pearson's product moment correlation coefficient, in the range (−1, 1) with 0 indicating no association, and computed the associated *P* value using cor.test function in R.

Differential enrichment tests. To identify differentially enriched regions across single cells for a given cluster, we performed a non-parametric two-sided Wilcoxon rank-sum test comparing normalized counts from individual cells from one cluster versus all other cells. We tested for the null hypothesis that the distribution of normalized counts from the two compared groups have the same median, with a 95% confidence interval. *P* values were corrected for multiple testing using the Benjamini–Hochberg procedure[48].

Pathway analysis. We used hypergeometric tests to identify gene sets from the MSigDB v.6.2, which are overrepresented among the lists of genes with a significant depletion in H3K27me3 (or enrichment in H3K4me3) over their TSS. For each gene list of the MSigDB, we used the phyper function in R to evaluate the overlap between the reference list and the list of significantly depleted/enriched genes (sample size dependent on the length of each gene list), correcting for multiple testing with the Benjamini–Hochberg procedure[48].

Copy number profiles of bulk tumor cells. The R package HMMcopy v.1.16.0 (ref. [55]) was used to correct for copy number variation in non-treated versus resistant xenograft models. Reads from bulk input ChIP-seq samples were binned in 0.5 Mb non-overlapping regions spanning the genome. Regions with a deviation to the mean greater than $n = 2$ s.d. were removed for analysis (Supplementary Fig. 13a and 14b).

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

Data availability

All sequencing files and processed count matrices were deposited with the Gene Expression Omnibus under accession number GSE117309.

Code availability

Codes are available from the GitHub repository (https://github.com/vallotlab/scChIPseq).

References

30. Cottu, P. et al. Modeling of response to endocrine therapy in a panel of human luminal breast cancer xenografts. *Breast Cancer Res. Treat.* **133**, 595–606 (2012).
31. Marangoni, E. et al. A new model of patient tumor-derived breast cancer xenografts for preclinical assays. *Clin. Cancer Res.* **13**, 3989–3998 (2007).
32. Petit, V. et al. Optimization of tumor xenograft dissociation for the profiling of cell surface markers and nutrient transporters. *Lab. Invest.* **93**, 611–621 (2013).
33. Duffy, D. C., McDonald, J. C., Schueller, O. J. & Whitesides, G. M. Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). *Anal. Chem.* **70**, 4974–4984 (1998).
34. Mazutis, L. et al. Single-cell analysis and sorting using droplet-based microfluidics. *Nat. Protoc.* **8**, 870–891 (2013).
35. Siegel, A. C., Bruzewicz, D. A., Weibel, D. B. & Whitesides, G. M. Microsolidics: fabrication of three-dimensional metallic microstructures in poly(dimethylsiloxane). *Adv. Mater.* **19**, 727–733 (2007).
36. Clausell-Tormos, J. et al. Droplet-based microfluidic platforms for the encapsulation and screening of mammalian cells and multicellular organisms. *Chem. Biol.* **15**, 427–437 (2008).
37. Anna, S. L., Bontoux, N. & Stone, H. A. Formation of dispersions using "flow focusing" in microchannels. *Appl. Phys. Lett.* **82**, 364–366 (2003).
38. Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017).
39. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
40. Abate, A. R., Chen, C. H., Agresti, J. J. & Weitz, D. A. Beating Poisson encapsulation statistics using close-packed ordering. *Lab Chip* **9**, 2628–2631 (2009).
41. Chabert, M., Dorfman, K. D. & Viovy, J. L. Droplet fusion by alternating current (AC) field electrocoalescence in microchannels. *Electrophoresis* **26**, 3706–3715 (2005).
42. Vallot, C. et al. Erosion of X chromosome inactivation in human pluripotent cells initiates with *XACT* coating and depends on a specific heterochromatin landscape. *Cell Stem Cell* **16**, 533–546 (2015).
43. Liu, Y., Popp, B. & Schmidt, B. CUSHAW3: sensitive and accurate base-space and color-space short-read alignment with hybrid seeding. *PLoS ONE* **9**, e86869 (2014).
44. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
45. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
46. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
47. Phanstiel, D. H., Boyle, A. P., Araya, C. L. & Snyder, M. P. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics* **30**, 2808–2810 (2014).
48. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B* **57**, 289–300 (1995).
49. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
50. 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).
51. Khandelwal, G. et al. Next-generation sequencing analysis and algorithms for PDX and CDX models. *Mol. Cancer Res.* **15**, 1012–1016 (2017).
52. McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).
53. Barron, M. & Li, J. Identifying and removing the cell-cycle effect from single-cell RNA-sequencing data. *Sci. Rep.* **6**, 33892 (2016).
54. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
55. Lai, D., Ha, G. & Shah, S. HMMcopy: copy number prediction with correction for GC and mappability bias for HTS data. R package v.1.8.0 (2016); http://bioconductor.org/packages/release/bioc/html/HMMcopy.html

nature research

Corresponding author(s): Andrew D. Griffiths, Céline Vallot, Annabelle Gérard

Last updated by author(s): Apr 3, 2019

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.

Statistics

For all statistical analyses, confirm that the following items are present in the 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
☐	☒	A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Datasets from our manuscript have been generated for the purpose of the study, we did not collect additional datasets.
Data analysis	scChIP-seq and scRNA-seq sequencing data were analyzed using Python (v2.7.12) and R (v3.3.3). Specific R Bioconductor packages include ConsensusClusterPlus, Sushi, Scater, ccRemover and HMMcopy. Codes are available upon request. We also used Cushaw3, bowtie 1.2.2, bedtools 2.17, bamcmp, Picard tools and 10x Cell Ranger Single Cell Software Suite 1.3.1.

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. 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

The GEO accession number of the data produced in this manuscript is GSE117309. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117309 with the token wniliuuuntelzcj

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	No sample-size calculation was performed; the aim of our manuscript was to develop and validate a droplet-based scChIP-seq approach.
Data exclusions	We excluded cells from analyses based on pre-established Quality Control metrics, detailed in Methodology questions (based on minimal and maximum coverage, and correlation score to other cells).
Replication	We have validated the reproducibility of our single-cell chromatin profiling method by analyzing n=3 technical replicates for H3K4me3 and n=2 biological replicates, ie independent cell suspensions processed independently on our micro-fluidics system, for H3K27me3 (All attemps at replication were successful). We also confirmed that the clustering of scChIP-seq data is not driven by technical artifact (batch effect) but cell type-specific biological similarities.
Randomization	Randomization is not relevant to our study as it aims at comparing chromatin and transcriptional states between a sensitive and a resistant tumor specimen.
Blinding	No blinding was performed as the study aimed at comparing chromatin and transcriptional states between a sensitive and a resistant tumor specimen.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	anti-H3K4me3 [Millipore, # 07-473 , lot # 2664283, dilution 2/1000] anti-H3K27me3 (C36B11) [Cell Signaling Technology, # 9733, lot # 008, dilution 2.5/1000]
Validation	- Anti-trimethyl-Histone H3 (Lys4) Antibody is a rabbit polyclonal antibody for detection of Histone H3 trimethylated at lysine 4. This highly specific and well published antibody has been validated in ChIP, DB, WB, PIA, ChIP-seq (Merck Millipore's website) - Tri-Methyl-Histone H3 (Lys27) (C36B11) Rabbit mAb detects endogenous levels of histone H3 only when tri-methylated on Lys27. The antibody does not cross-react with non-methylated, mono-methylated or di-methylated Lys27. In addition, the antibody does not cross-react with mono-methylated, di-methylated or tri-methylated histone H3 at Lys4, Lys9, Lys36 or Histone H4 at Lys20. (Cell Signaling Technology's website)

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	In the study, cell lines were used to demonstrate and benchmark the performance of the single-cell ChIP-seq technology. For this purpose, we used: - Human Jurkat cells (ATCC, T18-125) - Human Ramos cells (ATCC, CRL-1596) - Mouse M300.19 cells was a gift from B. Moser

Authentication	Human cell lines were obtained from ATCC and not authenticated. Mouse M300.19 cells were not authenticated.
Mycoplasma contamination	The cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	8- to 12-week-old female Swiss nude mice purchased from Charles River and maintained under specific pathogen-free conditions.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	The care and housing of mice used in this study were in accordance with institutional guidelines and the rules of the French Ethics Committee (project authorization no. 02163.02).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Not applicable, study focuses on two patient samples which have been used to generate two patient-derived xenograft models.
Recruitment	The two PDX models come from two patients of the Institut Curie hospital.
Ethics oversight	We have obtained consent from all participants of the study

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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.*	The GEO accession number of the data produced in this manuscript is GSE117309. The following secure token has been created to allow review of record GSE117309 while it remains in private status: wniliuuuntelzcj
Files in database submission	- Raw sequencing files Jurkat-Ramos_scChIP_K27me3_R1.fastq.gz Jurkat-Ramos_scChIP_K27me3_R2.fastq.gz Jurkat-Ramos_scChIP_K4me3_R1.fastq.gz Jurkat-Ramos_scChIP_K4me3_R2.fastq.gz HBCx-95_scChIP_H3K27me3_R1.fastq.gz HBCx-95_scChIP_H3K27me3_R2.fastq.gz HBCx-95-CapaR_scChIP_H3K27me3_R1.fastq.gz HBCx-95-CapaR_scChIP_H3K27me3_R2.fastq.gz HBCx-22_scChIP_H3K27me3_R1.fastq.gz HBCx-22_scChIP_H3K27me3_R2.fastq.gz HBCx-22-TamR_scChIP_H3K27me3_R1.fastq.gz HBCx-22-TamR_scChIP_H3K27me3_R2.fastq.gz HBCx-95_scRNA.bam HBCx-95-CapaR_scRNA.bam HBCx-22_scRNA.bam HBCx-22-TamR_scRNA.bam - Processed data files CountTable_Jurkat_scChIP_K4me3.txt.gz CountTable_Ramos_scChIP_K4me3.txt.gz CountTable_Jurkat_scChIP_K27me3.txt.gz CountTable_Ramos_scChIP_K27me3.txt.gz CountTable_HBCx-95_scChIP_H3K27me3_hg38.txt.gz CountTable_HBCx-95-CapaR_scChIP_H3K27me3_hg38.txt.gz

CountTable_HBCx-95_scChIP_H3K27me3_mm10.txt.gz
CountTable_HBCx-95-CapaR_scChIP_H3K27me3_mm10.txt.gz
CountTable_HBCx-22_scChIP_H3K27me3_hg38.txt.gz
CountTable_HBCx-22-TamR_scChIP_H3K27me3_hg38.txt.gz
CountTable_HBCx-22_scChIP_H3K27me3_mm10.txt.gz
CountTable_HBCx-22-TamR_scChIP_H3K27me3_mm10.txt.gz
CountTable_HBCx-95_scRNA.tar
CountTable_HBCx-95-CapaR_scRNA.tar
CountTable_HBCx-22_scRNA.tar
CountTable_HBCx-22-TamR_scRNA.tar

- Supplementary files (barcode sequences used in scChIP-seq experiments)
GSE117309_BarcodeLibrary_scChIPseq.fasta.gz

Genome browser session
(e.g. UCSC)

- Genome browser session H3K4me3 and H3K27me3 Jurkat Ramos
https://genome.ucsc.edu/s/kevin%20grosselin/Grosselin_Jurkat_Ramos_hg38

- Genome browser session H3K27me3 HBCx-95 model stromal cells
https://genome.ucsc.edu/s/kevin%20grosselin/Grosselin_HBCx%2D95_mm10

- Genome browser session H3K27me3 HBCx-95 model tumor cells
https://genome.ucsc.edu/s/kevin%20grosselin/Grosselin_HBCx%2D95_hg38

- Genome browser session H3K27me3 HBCx-22 model tumor cells
https://genome.ucsc.edu/s/kevin%20grosselin/Grosselin_HBCx%2D22_hg38

Methodology

Replicates

We have validated the reproducibility of our single-cell chromatin profiling method by analyzing n=3 technical replicates for H3K4me3 and n=2 biological replicates for H3K27me3 (r = 0.96 and 0.98 with p < 10e-15 respectively). We also confirmed that the clustering of scChIP-seq data is not driven by technical artifact (batch effect) but cell type-specific biological similarities.

Sequencing depth

scChIP-seq libraries were sequenced on an Illumina NextSeq 500 MidOutput 150 cycles. Cycles were distributed as follows: 50 bp (Read #1) were assigned for the genomic sequence and 100 bp (Read #2) were assigned to read the single-cell barcode sequence.
Number of mapped reads refers to the number of reads aligned to the reference genome by keeping only reads having no more than one reportable alignments and 2 mismatches.
On the other hand, number of uniquely mapped reads refers to the number of reads obtained after de-multiplexing of the single-cell barcodes and removal of duplicated reads (see Methods).

- scChIP-seq H3K4me3 Jurkat Ramos
Number of raw reads: 64,396,048
Number of mapped reads (hg38): 42,659,092
Number of uniquely mapped reads: 10,877,052

- scChIP-seq H3K27me3 Jurkat Ramos
Number of raw reads: 72,069,685
Number of mapped reads (hg38): 47,123,836
Number of uniquely mapped reads: 8,121,495

- scChIP-seq H3K27me3 HBCx-95
Number of raw reads: 151,963,813
Number of mapped reads (mm10): 34,586,767
Number of uniquely mapped reads (mm10): 5,187,802
Number of mapped reads (hg38): 50,472,984
Number of uniquely mapped reads (hg38): 8,715,197

- scChIP-seq H3K27me3 HBCx-95-CapaR
Number of raw reads: 204,907,172
Number of mapped reads (mm10): 23,862,992
Number of uniquely mapped reads (mm10): 3,739,207
Number of mapped reads (hg38): 92,698,620
Number of uniquely mapped reads (hg38): 15,333,344

- scChIP-seq H3K27me3 HBCx-22
Number of raw reads: 114,823,151
Number of mapped reads (hg38): 45,798,204
Number of uniquely mapped reads (hg38): 5,990,857

- scChIP-seq H3K27me3 HBCx-22-TamR
Number of raw reads: 109,501,758
Number of mapped reads (hg38): 33,037,774
Number of uniquely mapped reads (hg38): 2,844,194

Antibodies	anti-H3K4me3 [Millipore, # 07-473 , lot # 2664283, dilution 2/1000] anti-H3K27me3 (C36B11) [Cell Signaling Technology, # 9733, lot # 008, dilution 2.5/1000]
Peak calling parameters	For single-cell ChIP-seq experiments, we do not perform peak calling, as there is no input sample. For control bulk specimen, we performed peak calling using MACS for H3K4me3 and Zerone for H3K27me3 with default parameters.
Data quality	Data quality for scChIP-seq was assessed with QC metrics detailed in the Methods section: average coverage per cell with standard deviation, correlation with bulk profiles and ability to detect cell populations with clustering algorithms.
Software	scChIP-seq datasets were analyzed with Python (v2.7.12) and R (v3.3.3). All thresholds and steps are detailed in the Method section of the manuscript. Codes available upon request.