Chemotherapy induces canalization of cell state in childhood B-cell precursor acute lymphoblastic leukemia

Virginia A. Turati1, José Afonso Guerra-Assunção1,4, Nicola E. Potter2,3,4, Rajeev Gupta1, Simone Ecker1, Agne Daneviciute1, Maxime Tarabichi2, Amy P. Webster1, Chuling Ding1, Gillian May1, Chela James1, John Brown1, Lucia Conde1, Lisa J. Russell4, Phil Ancliff5, Sarah Inglott5, Giovanni Cazzaniga6, Andrea Biondi7, Georgina W. Hall8, Mark Lynch9, Mike Hubank2,10, Iain Macaulay1, Stephan Beck1, Peter Van Loo3, Sten E. Jacobsen12,13, Mel Greaves2, Javier Herrero1 and Tariq Enver1

Comparison of intratumor genetic heterogeneity in cancer at diagnosis and relapse suggests that chemotherapy induces bottleneck selection of subclonal genotypes. However, evolutionary events subsequent to chemotherapy could also explain changes in clonal dominance seen at relapse. We therefore investigated the mechanisms of selection in childhood B-cell precursor acute lymphoblastic leukemia (BCP-ALL) during induction chemotherapy where maximal cytoreduction occurs. To distinguish stochastic versus deterministic events, individual leukemias were transplanted into multiple xenografts and chemotherapy administered. Analyses of the immediate post-treatment leukemic residuum at single-cell resolution revealed that chemotherapy has little impact on genetic heterogeneity. Rather, it acts on extensive, previously unappreciated, transcriptional and epigenetic heterogeneity in BCP-ALL, dramatically reducing the spectrum of cell states represented, leaving a genetically polyclonal but phenotypically uniform population, with hallmark signatures relating to developmental stage, cell cycle and metabolism. Hence, canalization of the cell state accounts for a significant component of bottleneck selection during induction chemotherapy.

There is increasing evidence that the evolution of cancers and their responses to treatment are shaped by the complex interplay of their inherent genetic and epigenetic heterogeneities. However, the relative contributions of these factors to the phenotypes of treatment-resistant tumor cells remain poorly understood.

Intratumor genetic heterogeneity has been observed in all cancers. It evolves through branched trajectories and can be both spatially segregated and highly dynamic. The existence of genetically variegated subclones might explain resistance to chemotherapy and subsequent relapse. Indeed, comparisons of genetic variegation in paired samples from the same patient at diagnosis and relapse have found differences in clonal architecture, suggesting that chemotherapy may deterministically select for specific genetic variants. However, relapse typically occurs several months or even years after presentation, and often long after treatment has ceased. Thus, any insights into the biology of chemoresistance provided by such analyses are retrospective. Although the cellular substrate for relapse of any cancer clearly comprises those tumor cells that remain post-treatment, the complex—and often subclinical—process of post-therapy disease re-establishment provides ample opportunity for these cells to proliferate and undergo further genetic evolution that is not the direct result of treatment-induced selection processes. Hence, relapse samples may no longer be directly representative of the genetic landscape that prevailed immediately after therapy.

In addition to intratumor genetic heterogeneity, phenotypic heterogeneity—broadly encapsulated by epigenetic (nongenetic) influences on gene expression—might also provide a substrate for selection during both radio- and chemotherapy. However, whether this occurs independently of variability in genetic makeup remains unknown.

Very few studies have explored the genetic and epigenetic landscape of the disease residuum that persists during treatment. This is largely due to difficulties in obtaining and analyzing the appropriate material when cell numbers are limiting, because sampling tumors is difficult or—particularly in the context of hematological malignancies—because universal tumor cell-surface markers are lacking. Furthermore, as each patient has a genetically unique tumor that is treated only once, directly resolving deterministic from stochastic selection is not possible, although in principle it could be inferred from large cohorts of patients.

As a result of its relatively simple genetics and its restriction to a well-characterized blood lineage, BCP-ALL has proven...
paradigmatic in illuminating many fundamental principles of cancer biology. Although treatment of BCP-ALL is based on a multi-drug extended protocol lasting up to 36 months, maximum cytotoxicity (typically much greater than 99%) is achieved during the first cycle of treatment (induction chemotherapy, lasting 28 d). After this, most patients will be deemed in complete remission, without detectable disease. In some patients, however, a small number of leukemic cells, almost always subclinical and known as minimum or measurable residual disease (MRD), remain post-induction, and these cells presumably provide the substrate for relapse. Molecular quantification of this residual disease provides the single most potent indicator of long-term clinical outcome, clearly demonstrating the clinical importance of this cell compartment.

To overcome the technical and logistic hurdles outlined above, we used a patient-derived xenograft (PDX) model of BCP-ALL, whereby leukemic cells from children were transplanted into multiple xenograft recipients, each of which developed almost identical leukemias. Mice were then treated with vincristine (a microtubule poison) and dexamethasone (a glucocorticoid), mimicking the first 28 d of BCP-ALL treatment. These two drugs form the long-standing core of the standard multdrug regimens in clinical use, which also typically include the enzyme l-asparaginase. Additional compounds, acting primarily through targeting nucleic acid and its synthesis (for example, doxorubicin, cytarabine, methotrexate, 6-mercaptopurine, and so on), are variably used during induction chemotherapy, and always used during the subsequent intensification and maintenance phases. We then deployed a wide range of bulk and single-cell resolution assays to assess at high resolution the genotype and phenotype of the surviving cells.

Results

Defining the clonal architecture of BCP-ALL. To date, genetic variation in BCP-ALL has been inferred by bulk sequencing and/or multicolor FISH (mFISH) using a limited range of probes. To understand the true extent of intratumoral genetic heterogeneity, and to dissect the probable order in which mutations are acquired, we performed high-resolution, single-cell whole-genome sequencing (scWGS).

We developed a new script for PicoPlex whole-genome amplification (WGA) using the Fluidigm C1 platform, which is suitable for analyzing small cell numbers (Extended Data Fig. 1). PicoPlex chemistry efficiently detects copy-number aberrations (CNAs), the most common genomic lesions in BCP-ALL. We studied clonal complexity in bone marrow (BM) samples taken at diagnosis (that is, before treatment) from six cases of childhood BCP-ALL, analyzing 756 cells in total.

Most cases displayed branching clonal architecture with 4–19 descendant subclones (Fig. 1). Alongside commonly recurring lesions, we identified a number of rarer or previously unreported subclonal CNAs (that is, APOBEC3B amplification, MYC amplification, and DNMT3A deletion). In each case, we observed multiple independent lesions targeting the same locus. These were characterized by distinct genetic breakpoints and associated with different sets of CNAs, and so were segregated on to separate branches of the leukemias’ phylogenetic trees.

This points to parallel evolution as a highly pervasive mechanism of genomic diversification in BCP-ALL, consistent with previous reports that serial and sustained RAG- and activation-induced cytidine deaminase-mediated deletions contribute to the genetic heterogeneity observed in this leukemia. Furthermore, it suggests that genetically distinct subclones might exhibit similar fitness during the development of BCP-ALL, implying that some mutations might be evolutionarily neutral.

Genotype and resistant states do not co-segregate in BCP-ALL.

We next analyzed the relationship between leukemic genotypes and disease-relevant cell states, asking to what extent a given genotype corresponds to a specific phenotype. We focused on CD19+/CD34+/CD38− cells, which comprise a rare subpopulation present in leukemia, but so far undetected in healthy blood or BM. This compartment is enriched in developmentally primitive, quiescent cells that have previously been suggested to hold leukemia-initiating cell (LIC) potential and to be linked to increased chemoresistance in patients.

To simultaneously assess genotype and phenotype of individual cells, we used previously published bulk WGS data of five unsorted diagnostic leukemias with different karyotypic abnormalities, to identify patient-specific genetic alterations present at varying allele frequencies, allowing us to group cells into genetically similar subclones and infer a phylogenetic, clonal architecture for each patient. We then isolated CD19+/CD34+/CD38− and unsorted control cells and used single-cell quantitative (q)PCR to assess in a targeted fashion the presence/absence of each patient-specific genomic alteration (nonsynonymous single-nucleotide variants (SNVs) or CNA mutations), as well as expression of the proliferation marker Ki67 and other differentiation markers. This analysis validated that CD19+/CD34+/CD38− cells are, as anticipated, predominantly noncyling (see representative data in Extended Data Fig. 2a,b).

In 4 of the 5 cases, all but one of 18 subclones were identified in the CD19+/CD34+/CD38− cells, in proportions similar to those in the corresponding bulk leukemic cells. In one case (patient (Pt.)11), CD19+/CD34+/CD38− cells were restricted to a single minor subclone, representing ~5% of the total bulk population (Fig. 2, Extended Data Fig. 2c–d and Supplementary Table 1). We therefore conclude that all genotypes in BCP-ALL equally populate cell states associated with nonproliferative status (here defined as Ki67−) and early B-cell differentiation.

Intratumor genetic heterogeneity is unaffected by treatment. Next, we explored the implications of this lack of segregation between genotypes and phenotypes canonically associated with resistance in BCP-ALL shaped by the selective pressure of induction chemotherapy. We addressed this using PDXs (Fig. 3a), whereby BM cells from BCP-ALL patients were transplanted into multiple mice and treatment outcomes assessed by tracking clonal dynamics longitudinally in all recipients, thus directly comparing clonal variation before (BM, aspiration) and after (total BM) identical chemotherapy treatment (Extended Data Fig. 3a).

MFISH was used to score the co-occurrence of CNAs and/or translocation events in >250 cells, and thereby group cells into genetic subclones (Fig. 3b,c and Extended Data Fig. 3b–d). We used probes for AML1 (RUNX1), TEL (ETV6), TEL–AML1 (ETV6–RUNX1), CDKN2A (p16) and PAX5 (BSAP), which are among the most frequently altered genes in childhood BCP-ALL. Using the Jensen–Shannon index of divergence, which measures the similarity between two probability distributions, we verified that, at day 0, the clonal compositions of the leukemias in all mice engrafted with the same primary sample were effectively the same (Fig. 3d,e).

Next, we treated half the mice with vincristine and dexamethasone over 28 d, with the remaining mice as controls. These drugs are key components of systemic BCP-ALL induction chemotherapy, which also includes l-asparaginase, and sometimes additional agents targeting DNA replication through distinct mechanisms. Treatment with vincristine and dexamethasone produced cytoreduction in the PDX model comparable, if not superior (>10 log), to that observed in patients (Extended Data Fig. 3e–h). Mice engrafted with cells from high-risk patients showed both slower kinetics of response and a higher percentage of residual cells after treatment than mice transplanted with cells from low-risk patients (Extended Data Fig. 3i).

Although treatment markedly reduced overall leukemic burden, most genetic subclones were still detectable (Fig. 3f,g), suggesting that they display broadly similar fitness to chemotherapy. Where we
Fig. 1 | Defining the clonal architecture of BCP-ALL at single-cell resolution. Phylogenetic trees mapping the evolutionary histories of six diagnostic leukemias as inferred by scWGS. Each clone is represented by a gray circle with a diameter reflecting the size of the clone within the leukemia. White circles represent inferred ancestors (not detected by the analysis). Branches linking the subclones to one another are color coded by their copy-number events. Events that appear in more than one branch within each tree are marked with the same letter. Below each tree, a dot matrix shows the combination of copy-number events for each leaf clone. Linked dots represent copy-number events shared through a common ancestor. Disconnected dots correspond to cases where the same event happens in more than one lineage. Due to a technical failure of the flow sorter, only nine cells from Pt.2 were available for WGS. Normal early B-cell progenitors from two umbilical cord blood units were used as controls to normalize read counts and exclude potential technical artefacts.
observed differences in the size of specific subclones, these changes were (1) also detected in control mice and (2) inconsistent between recipients and between disease sites (BM and spleen) within an individual mouse, highlighting their stochastic rather than their deterministic nature (Fig. 3g and Extended Data Fig. 3j). We quantified the overall extent of intratumor genetic heterogeneity before and after chemotherapy by estimating the Shannon entropies for the subclonal composition of untreated and treated PDX recipients. This entropy index provides a measurement of a sample's diversity and richness in species (in this case the different genotypes), which

---

| Patient 7 | Patient 8 | Patient 9 |
|-----------|-----------|-----------|
| IKZF1: d | GPR156 het | CCCDC110 het |
| VPREB1: d | CCNA1E het | CYCL127.6 homo |
| IGHH-CRLF2: f | DNAH2 het | SLC3A1 het |
| ZNF260: d | ARHGAP6 homo | OBSCN het |
| LEMD3 i | C1orf101 het | KIAA1012 het |
| MUC17: m | MATN3 het | LRFN5 het |
| DST: m | CD163L7 het | ANO3 het |
| ANKFN1: m | TRIM36 het | SIM het |
| TSNAIXIP1: m | GPR156 het | NSD2 het |
| GMYAS: m | DNAH2 het | VPREB1: d |
| BMPS: m | CCNA1E het | SERP2: d |
| DEL deletions | DEL12.12 | IGH-CRLF2: f |
| DEL1.1 | DEL12.11 | ZNF260: d |
|          |           | LEMD3: i |
|          |           | MUC17: m |
|          |           | DST: m |
|          |           | ANKFN1: m |
|          |           | TSNAIXIP1: m |
|          |           | GMYAS: m |
|          |           | BMPS: m |

**Fig. 2 | Genotype and cell states associated with resistance do not co-segregate in BCP-ALL.** a–c. Subclonal genetic architecture of bulk (unsorted) and noncycling (CD34+/CD19+/CD38−) diagnostic leukemic cells from three patients with childhood BCP-ALL (Pt. 7 (a), Pt. 8 (b) and Pt. 9 (c)), as inferred by single-cell, multiplex, targeted qPCR. Informative patient-specific qPCR probes detecting either nonsynonymous SNVs or CNAs were selected based on WGS data of the same diagnostic samples. Some 80–300 cells were analyzed per compartment, and a threshold of ~2% was implemented to control for false-positive events. d, deletion; f, fusion; h, heterozygous; homo, homozygous; i, insertion; m, mutation.
also takes into account how evenly distributed these are. We found that chemotherapy did not substantially change the Shannon entropies in recipients, suggesting that it has little impact on the genetic complexity of residual disease (Fig. 3h,i).

Although scWGS captures the most common recurrent genetic alterations in childhood BCP-ALL and affords high throughput, its resolution is limited by the number of markers that can be tested in a single experiment and might therefore underestimate genetic complexity. Higher-resolution approaches should better elucidate the dynamics of minor subclones, and so we analyzed leukemic cells from a third patient by scWGS. We used sequencing data to consolidate individual cells into subclones defined solely by the copy-number status of the markers assessed by mFISH. To further subclassify these clones, we then extended the CNA to full genome resolution. We found a high level of agreement between the two approaches. Most cells could be classified into mFISH-defined subclones. Only cells belonging to a single subclone, diploid with respect to all loci assessed by mFISH, remained unassigned, suggesting that mFISH provides a good estimate of the overall subclonal composition of childhood BCP-ALL. This notwithstanding, whole-genome information allowed us to further divide each clone identified by mFISH into up to 13 smaller subclones (Fig. 4a).

Next, we investigated at whole-genome resolution the clonal dynamics of leukemic cells in PDX recipients undergoing treatment. The Jensen–Shannon index of divergence confirmed that leukemic cells in the most recipients (four out of five) sampled by BM aspiration before treatment shared the same distribution of subclones as the primary material from which they were derived. Enrichment for a subset of minor diagnostic subclones in one of the recipients suggested a possible sampling error at the time of injection (Fig. 4b,c).

We treated two of the mice, using two as controls. The fifth mouse was cycled early due to unrelated complications. High-resolution scWGS analysis of the immediate post-treatment residuum confirmed that treatment-resistant cells in childhood BCP-ALL are as genetically heterogeneous as untreated cells. Of all the clones observed before chemotherapy, only one minor clone (frequency <3% at day 0) became undetectable in both treated mice (Fig. 4d,e and Extended Data Fig. 4). This notwithstanding, a slight increase in Shannon entropy’s index values suggested a more uniform distribution of subclones post-treatment (Fig. 4f). This resulted from partial expansion of a few minor subclones present at very low frequency before chemotherapy. Although our data do not categorically exclude the possibility that particular genotypes might confer some degree of enhanced resistance to chemotherapy, they strongly suggest very limited selection of genetic subclones and reinforce our prior conclusion that the overall genetic complexity of childhood BCP-ALL remains intact in the post-chemotherapy residuum.

Transcriptionally driven phenotypes determine resistance. The absence of bottleneck genetic selection after chemotherapy for BCP-ALL suggests that resistance operates through alternative mechanisms, and we envisaged two distinct scenarios: either survival is purely stochastic and some cells escape treatment by chance, or convergent evolution underpins selection of specific cell state(s) with reduced chemotherapy sensitivity. These state(s) may pre-exist before treatment or be induced by it. To distinguish between these hypotheses, we tested whether leukemic cells that survive treatment are transcriptionally distinct from their treatment-naïve counterparts, through small cell number bulk RNA-sequencing (RNA-seq) on treatment-naïve and -resistant cells from multiple mice engrafted with cells from three patients with different cytogenetics (Fig. 5a).

Principal component analysis (PCA) demonstrated significant transcriptional differences in leukemic cells retrieved from treated and untreated mice (Extended Data Fig. 5a).

We reasoned that this transcriptional signature was probably a composite of the unique intrinsic transcriptional program(s) of cells with reduced sensitivity to chemotherapy, which had the potential to be retained throughout treatment and subsequent relapse, and a ‘generic’ stress-related, and more transient, transcriptional response to cytotoxic therapy. To discriminate these distinct elements, we compared the transcriptomes of xenografted leukemic cells under five different experimental conditions involving two rounds of transplantation and treatment, namely: (1) chronically treated cells: both primary and secondary recipients were treated for 7 d; (2) ‘acutely treated cells’: primary recipients untreated, secondary recipients treated; (3) ‘treatment withdrawn cells’: primary recipients treated, secondary recipients untreated (designed to address the reversibility of transcriptional phenotypes associated with a short exposure to chemotherapy); (4) ‘untreated controls’: cells were not exposed to chemotherapy in primary or secondary recipients; and (5) ‘relapse cells’: recipients were treated for 4 weeks and then allowed to relapse ‘in situ’ in the absence of further transplantation (Fig. 5b).

PCA clustered samples according to their exposure to chemotherapy, with PC1 discriminating ‘untreated’ and ‘treatment withdrawn cells’ from ‘acutely’ and ‘chronically treated’ cells (Supplementary Fig. 5b). Compared with untreated cells, acutely treated cells showed the highest number of differentially expressed genes. As hypothesized, pathways involved in response to cellular stress (for example, inflammatory response and complement activation) were expressed at significantly higher level in acute than in chronically treated cells (Supplementary Table 2 acute versus chronic comparison). We hence refined our definition of the core gene expression program of residual disease, by focusing on pathways that were altered across all four treatments and appeared to be similarly deregulated in both acutely and chronically treated cells. In line with previous reports by our group and others, this analysis identified a striking downregulation of pathways involved in

---

**Fig. 3** | Treatment in a mouse model of childhood BCP-ALL does not affect the overall extent of intratumor genetic heterogeneity. **a.** Treatment model: NSG mice received an intramedullary injection (right leg) of primary leukemic cells from pediatric patients with ALL; 12 weeks after injection (day 0 (d0)), all mice underwent a BM aspiration (right leg). On engraftment, mice were randomly assigned to either control or treatment groups. Treated mice received dexamethasone (6 mg l−1 continuous administration in the drinking water) and vincristine (0.50 mg kg−1 intraperitoneally (i.p.) weekly) for 4 weeks. After treatment, the BM (from the femur, tibia and pelvis from both sides) of each mouse was harvested. **b,c.** Bar plots showing the subclonal composition of each engrafted diagnostic Pt. 3 (b) or Pt. 2 (c) leukemia. Clonal makeup was determined by mFISH (n=1,186 cells across n=10 Pt.3-engrafted mice and n=2,268 cells across n=12 Pt.2-engrafted mice) using probes for TEL (ETV6), AMLL (RUNXI), PAX5 and CDKN2A (p16). Cells were grouped into genetically similar subclones based on their combinatorial copy-number status, and each subclone was then assigned a unique identifier number and color code (Extended Data Fig. 3c,d). **d,e.** Heatmap showing the results of all possible pairwise comparisons of the clonal composition of individual Pt. 3 (d) and Pt. 2 (e) engrafted leukemias using the Jensen–Shannon divergence index (bounded between 0 and 1). **f.** Bar plots showing the BM clonal composition (Clon. comp.) of two control and four treated mice (identified by the tag number shown above the chart) engrafted with Pt.3 cells at day 0 (BM aspiration) and day 28 (total BM harvest). Below the bar plots are the corresponding, pairwise, Jensen–Shannon divergence analysis results. **g.** Same as e but for mice engrafted with Pt.2 cells. In this case, the spleen was also analyzed at day 28. **h,i.** Shannon entropy was used to quantify the overall diversity of the matching pre- and post-treatment specimens for each Pt.3 (h) and Pt.2 (i) recipient.
cell-cycle regulation, cell activation and cell metabolism. Our data also suggest that, after treatment, signatures associated with earlier stages of the differentiation hierarchy are upregulated in residual leukemic cells (Fig. 5c). To test whether these phenotypic changes had any functional impact on the biology of resistant cells, we used a limiting dilution approach to quantify LICs in treated and untreated PDX recipients and found that cells from treated recipients had increased tumorigenicity (Extended Data Fig. 5c,d).

Although the global transcriptome of cells treated for 7 d and then re-transplanted (that is, ‘withdrawn’) reverted toward that of treatment-naive cells, the reversion was incomplete. In addition, ‘relapsed’ leukemic cells from a separate branch of the experiment...
Fig. 4 | ScWGS confirms limited selection of genetic subclones in response to chemotherapy. a, Phylogenetic tree of Pt.1 diagnostic leukemia as inferred by scWGS. Background shading is used to highlight clones previously defined by mFISH. N/A, not available. b, Subclonal genetic composition of five mice engrafted with Pt.1 leukemia at day 0 (before treatment) and Pt.1 diagnostic disease. c, Heatmap showing all Jensen–Shannon divergence index-based comparisons of clonal compositions between pairs of engrafted mice at day 0. d, e, TimeScape plots visualizing tumor evolution over time in two (grouped) control (d) and treated (e) mice. The plots show the clonal composition at three timepoints: (1) the primary diagnostic leukemia, (2) at day 0 in the xenografts (BM aspirate) and (3) at day 28 in the xenografts (total BM harvest). Clonal identities as detailed in Supplementary Fig. 4c. f, Table summarizing the Shannon entropy values of each control (Ctrl) and treated engrafted leukemia at days 0 and 28.
that more closely mimicked the clinical situation retained many transcriptional characteristics of resistant cells, even if harvested 6–8 weeks after the end of treatment, at a time when the leukemia was fully re-established. These cells showed persistent downregulation of G2M checkpoint, E2F and MYC1 gene target signatures, and upregulation of the pro-B-cell differentiation signature compared with treatment-naive cells (Fig. 5c).

Together, our data further define the nature of the polyclonal leukemic cells that escape chemotherapy in childhood BCP-ALL; they have features associated with both primitive differentiation status and quiescence and exhibit increased LIC potential. Furthermore, some of the key transcriptional features of the chemotherapy-resistant post-treatment residuum persist through in vivo relapse.

We next asked to what extent the gene expression programs associated with resistance are encoded through epigenetic modifications. We performed bulk DNA methylation analysis on the same matching pre- and post-treatment samples that were used for RNA-seq. We defined a core set of resistance-associated methylation differences, identified as regions of the genome that were consistently differentially methylated in pre- (day 0) and post-treatment (day 28) samples while remaining unchanged in control mice. Of 669 significantly differentially methylated regions (DMRs), the vast majority mapped in promoter regions (430 in total). Reassuringly, several DMRs spanned genes previously described as drivers of cancer, BCP-ALL progression and drug resistance (for example, HOXA9, MAPR3, ADAM12, ELAVL4 and PD4DE; representative Fig. 5d and Supplementary Table 3).

We then performed an integrative analysis of our RNA-seq and methylene data using the functional epigenetic modules (FEM) algorithm1, which identifies gene modules of coordinated differential methylation and expression in the context of a human interactome. Several coordinated modules in resistant cells encoded known regulators of early B-cell differentiation, cell cycle and methylation itself (that is, MME/C10D, E2F1 and EZH2), suggesting that resistance phenotypes, at least in part, extend to the epigenome level (Fig. 5e–g).

Epigenetic variability provides a substrate for selection. We next used single-cell RNA-seq (scRNA-seq) to explore the transcriptional heterogeneity of BCP-ALL and its contribution to treatment resistance. We examined two diagnostic leukemias (Pt.1 and Pt.2) and cells from Pt.1 PDxs, whose genotypes had previously been characterized at scWGS resolution. We found considerable transcriptional heterogeneity in untreated primary leukemias, with expression of genes involved in proliferation, metabolism (oxidative phosphorylation), apoptosis (p53 pathway) and differentiation status being the most variable between cells (Extended Data Fig. 6a,b). Most pre-treatment diagnostic and xenografted leukemic cells showed promiscuous expression of signatures characteristic of distinct hematopoietic lineages and differentiation stages (Extended Data Figs. 6c and 7) and, when assigned to their most closely related normal counterpart2, we found that BCP-ALL cells can map transcriptionally to any point of the hierarchy (Extended Data Fig. 6d and Fig. 7d).

PCA showed that, in contrast to untreated cells, treated cells occupied a well-defined, limited projection space, suggesting greater transcriptional homogeneity (Fig. 6a), as further confirmed by statistical analysis. Furthermore, treated cells expressed significantly fewer genes per cell, and at a lower level, than their untreated counterparts, suggesting global transcriptional repression, in line with the bulk RNA-seq results (Fig. 6b). We validated the greater heterogeneity of untreated cells using Uniform Manifold Approximation and Projection (UMAP), which classified untreated cells into three distinct clusters while assigning treated cells to a single subpopulation (Extended Data Fig. 8a), and used the R toolkit (Seurat) to identify markers defining each cluster (Extended Data Fig. 8b,c).

Of note, even though most signaling cascades were downregulated, resistant cells retained, and even upregulated, signatures corresponding to Hedgehog signaling, Tirosh stemness and tumor necrosis factor (TNF)-α via the nuclear factor κ-light-chain-enhancer of activated B cells (NF-κB) pathway (Extended Data Fig. 9), previously implicated in the maintenance of cancer stem cells14–16.

In line with these findings, our methylene analysis also showed that RNA polymerase II transcription factor activity was the most significantly deregulated gene set after treatment (identified by pathway enrichment analysis; \(P = 1.85 \times 10^{-37}\)), and methylation variance (MV) was also significantly reduced compared with untreated cells. This was true even when taking into account the reduced sensitivity inherent to any bulk assay (Fig. 6c,d).

By calling SNVs directly in the scRNA-seq data and comparing their frequency in the paired pre- and post-treatment samples, we validated our earlier CNA-based finding that chemotherapy did not primarily select for genotypes, and demonstrated that those very same transcriptionally homogeneous cells that survive the transcriptional bottleneck are as mutationally heterogeneous as the untreated disease (Fig. 6e). This lack of selection was true even when the identified variants (2,155 in total, previously reported in the COSMIC dataset) were clustered by affected gene, demonstrating that no specific mutated gene is preferentially enriched either before (day 0) or after (day 28) treatment (Fig. 6f).
Only a subset of G0 cells is chemoresistant. G0 nonproliferative cell state and resistance to cytotoxic drugs have previously been linked in cancer in general, and specifically in the context of BCP-ALL28,30,36. We therefore explored at single-cell resolution the contribution of cell-cycle and differentiation state to the intercellular heterogeneity of leukemic cells and chemoresistance. To address whether all quiescent leukemic cells displayed equally reduced sensitivity to treatment, we analyzed the relative expression of gene sets
associated with G1/S and G2/M cell cycle phases in untreated and treated cells, to assign each to its corresponding cell-cycle stage. As expected, although untreated cells could be at any stage of the cell cycle, treated cells were restricted to G0 (Fig. 7a). Surprisingly, however, G0 cells also accounted for approximately 70% of the leukemic cells harvested from xenograft recipients before treatment (Fig. 7b). As our treatment model achieved very large levels of cytoreduction (>10x log), this suggests that quiescent cells resistant to cytotoxic drugs probably represent a specific and rare subset of a wider quiescent compartment—operationally defined by molecular profiling, that is, for the most part, sensitive to treatment.

Consistent with this, in cell-cycle scatter plots, untreated and treated G0 cells only partially overlap (Fig. 7a), indicating that more than one quiescent state might exist in childhood BCP-ALL. We explored this further by looking at MYC and E2F1, which are among the genes most differentially expressed between treated and untreated cells (Extended Data Fig. 10a), and have previously been implicated in the regulation of quiescence depth and the propensity to proliferate in normal fibroblasts. Analysis of co-expression showed that progressively shallower quiescence in leukemic cells is defined by the concomitant upregulation of Myc and (Rb)-E2F signaling. Once a Rb-E2F-dependent restriction point (or expression plateau) has been reached, cells begin to proliferate, suggesting that, as in normal cells, a MYC-dependent E2F switching threshold regulates the propensity to enter the cell cycle in childhood BCP-ALL (Fig. 7c).

Treatment enriched for cells expressing the multi-lymphoid progenitor (MLP) signature (Fig. 7d), which identifies the earliest precursors that have both lymphoid (B, T and natural killer (NK)) and myelomonocytic potential, whereas high expression of other signatures (CMP, GMP and ETP) was never observed in any deeply quiescent untreated or treatment-resistant cells (Fig. 7d and Extended Data Fig. 10b). Accordingly, expression of markers associated with the ETP lineage, as well as to a lesser extent, with the CMP, MEP and GMP lineages, positively correlates with the signatures of cycling cells (that is, MYC, E2F and G2M) (Fig. 7e).

Thus, G0 cells in childhood BCP-ALL appear to vary in quiescence depth and, although the untreated disease comprises predominantly shallow quiescent cells interspersed with rare deeply quiescent cells, the chemoresistant population is highly populated by the latter, which are also developmentally distinct. Crucially, the existence of rare phenotypic substates within the G0 compartment, with different sensitivities to treatment, resonates well with the observation that, in the clinical setting, only a miniscule percentage of leukemic cells survives chemotherapy, typically there being <10^{-4}–10^{-5} residual leukemic cells in the BM of MRD-positive patients. Treatment selects a rare pre-existing cell state. We next asked whether leukemic cells with the transcriptional phenotype associated with chemoresistance were present before treatment. We used the TSCAN algorithm to cluster cells with similar gene expression profiles, irrespective of treatment status, and ordered these clusters along an inferred pseudotime trajectory, based on the average expression values of all genes. Most untreated and treated cells were positioned at opposite ends of the resulting minimum spanning tree, with the more heterogeneous untreated cells being distributed over a more significant number of clusters than treated cells (Extended Data Fig. 10c). Expression of genes involved in cell-cycle status and differentiation varied as a function of cells’ position along the inferred pseudotime axis (Extended Data Fig. 10d,e). Strikingly, a subpopulation of cells in the untreated leukemia had transcriptional features more closely related to those of the chemoresistant cells that persisted after treatment than to the rest of the untreated sample (Fig. 8a), suggesting that some cells were in a pre-existing resistant state (PRS) that closely resembles the chemoresistant cells.

When compared with the rest of the untreated cells, these PRS cells represented a more homogeneous population enriched for markers characteristic of chemoresistant, treated cells, including deep transcriptional quiescence and the expression of lineage markers associated with a primitive developmental state (Fig. 8b,c). Unsupervised heatmap clustering of the cells based on the most variable genes and t-distributed stochastic neighbor embedding (t-SNE) clustering also independently confirmed the presence of PRS cells within the untreated disease (Fig. 8d and Extended Data Fig. 10f). Of note, the t-SNE analysis also provided insight into the spatial relationship across untreated, PRS and treated cells, and identified a subset of cells that, although previously defined as PRS by pseudotime analysis, now lay outside the limits of the PRS occupied area. Such cells might present an additional intermediate cell state that might acquire full phenotypic resistance only on treatment exposure. It is interesting that PRS cells display enhanced expression of mitochondrial genes, which may contribute to their capacity to withstand chemotherapy (Extended Data Fig. 10g).

To assess the relevance of these findings to clinical practice, we performed scRNA-seq analysis of matching diagnosis, MRD and relapse specimens from two patients undergoing standard BCP-ALL treatment (including with l-asparaginase). Of note, residual leukemic cells in primary MRD specimens typically account for <10^{-3}–10^{-4} cells, and there are no universal markers that can be used to separate malignant cells from healthy hematopoietic cells, making the direct interrogation of this disease compartment extremely challenging.

We first used the xenograft data to generate a signature of resistant and PRS cells by differential gene expression (DGE) analysis of treated versus untreated noncycling cells, and of PRS versus untreated noncycling cells. Our signature consisted of the top 50 genes, after pre-emptively filtering out both untreated cycling cells and cell-cycle-related genes to ensure that the DGE analysis results would not solely reflect cell-cycle differences between the populations. To account for the very small size of the PRS compartment, which makes direct comparison with larger populations statistically
challenging, and to ensure weighting of the signature in favor of genes important for PRS cells, we also filtered out genes differentially expressed between treated and PRS cells.

Testing the expression of the generated signature in individual cells from the clinical specimens demonstrated that, before any treatment, the diagnostic disease contains a rare population of cells
**Fig. 7** Resistant BCP-ALL cells represent a small subset of G0 cells with MLP-like transcriptional features. 

**a**, Scatter plot showing the classification of n = 611 cells to noncycling/G0 and different stages of the cell cycle, based on the relative expression of gene sets associated with the G1/S (x axis) and G2/M (y axis) phases. Gene sets are obtained from ref. 34 and available in Supplementary Table 4. 

**b**, Bar plot showing the percentages of cycling and G0 cells within untreated leukemia at day 0 (n = 351 cells). Treated cells were found exclusively in the G0 compartment. 

**c**, Scatter plot showing the classification of noncycling/G0 n = 611 cells as deep quiescent (Q) and shallow quiescent, based on the relative expression of gene sets associated with E2F signaling (x axis) and MYC signaling (y axis). The 95th percentile of the expression of each signature in treated cells was used as a reference to establish the approximate cutoff values for deep quiescence. 

**d**, Nonlinear dimensionality reduction and visualization t-SNE analysis. Each cell is assigned to its closest normal hematopoietic cell lineage based on the expression of known marker genes (signature with the highest z-score average). 

**e**, Correlation matrix looking at the relationship between different gene signatures associated with cell cycle and lineage. Correlated expression patterns are displayed in blue and anticorrelated in red.
with high expression of the xenograft-based PRS/treated signature. Crucially, in agreement with our idea that treatment selects for cells in this state, this population is considerably enlarged in the matching post-induction chemotherapy day-28 MRD samples (Fig. 8c). Altogether, therefore, our data paint a direct and clinically relevant picture of the prime mechanisms of selection operating during the critical phase of induction therapy in BCP-ALL, which entail selection for a rare pre-existing cell state rather than a specific genotype.

Discussion
We have sought to understand the principles shaping the genetic and epigenetic landscapes of the leukemic residuum that persists in children with BCP-ALL immediately after the completion of induction chemotherapy, as modeled through in vivo transplantation and treatment. These cells are typically subclinical but are ultimately responsible for disease recurrence.

Using analyses linking the genotype and the cell state of individual leukemic cells to their phenotype, we have tested the prevailing notion that intratumor genetic heterogeneity is the primary determinant of resistance to chemotherapy in BCP-ALL. Surprisingly, but in line with the results of a similar recent study of colorectal cancer, this is not the case; rather, phenotypic heterogeneity is the primary source of escape.

Our treatment studies deployed xenografts, a model that is particularly well suited to the study of B-cell malignancies. The primary leukemic cells that engraft recipients retain the range of genotypic and phenotypic heterogeneities of the leukemia from which they are derived, and are routinely used to test new drugs. Inevitably, the chemotherapy regimen we used in our experiments is not a perfect replica of that utilized in children, which always includes l-asparaginase, and sometimes additional cytotoxic drugs with distinct mechanisms of action. In addition, induction chemotherapy in BCP-ALL patients is followed by 24–36 months of consolidation/maintenance treatment, aimed at preventing disease recurrence. As this lengthy protocol would be problematic to fully recapitulate in xenografted mice, whether genetic selection occurs during these later phases of treatment remains to be assessed. Nevertheless, induction chemotherapy is where maximum cytoreduction occurs, and dexamethasone and vincristine, which we administered to PDXs in line with standard clinical schedules, together comprise the backbone of this crucial phase of treatment in BCP-ALL. Furthermore, our in vivo treatment regimen achieved a degree of cytoreduction and disease clearance that is comparable, if not superior, to that observed in clinical practice, and we were able to confirm some of our key findings directly in patients receiving complete clinical regimens.

The model also allowed us to independently treat cells from our key findings directly in patients receiving complete clinical regimens. The results in clinical practice, and we were able to confirm some of our key findings directly in patients receiving complete clinical regimens. The model also allowed us to independently treat cells from the same leukemia multiple times, something that is not possible in patients, and therefore afforded insights into the deterministic versus stochastic nature of selection at the level of the genotype of individual tumor cells.

Our finding that induction chemotherapy in BCP-ALL does not primarily select tumor genotypes deterministically, suggests that most subclones have similar fitness in response to the selective pressure of treatment. This resonates with our observation that there is no clear co-segregation between BCP-ALL phenotypes classically associated with treatment resistance, such as the absence of Ki67 expression, and genotype. Although in some instances we observed minor differences in clonal size post-treatment, which could reflect either stochastic drifts or potentially modest differences in treatment sensitivity, none was of a magnitude compatible with major bottleneck selection acting on genotypic differences, as has been inferred by earlier studies of diagnosis and relapse.

However, as alluded to above, the persistence of genetic heterogeneity provides a plausible substrate for the later diversification and expansion of specific genetic subclones during the evolution of relapse, either on- or off-treatment (intensification/maintenance). Thus, although leukemic subclones of all genotypes survive induction chemotherapy to a broadly similar extent, those that either recover/expand/proliferate fastest (either deterministically or stochastically), or stochastically acquire de novo mutations that confer increased resistance to subsequent treatment (maintenance and consolidation), will probably dominate at relapse.

In the absence of significant genetic selection during induction treatment, our transcriptomic analysis nevertheless uncovered a severe bottleneck at the cell-state level. Leukemic cells in newly diagnosed patients span a spectrum of cell states varying in phenotypic traits such as cell-cycle, metabolic status and differentiation stage. Furthermore, analogous to normal hematopoietic stem and progenitor cells, they exhibit promiscuous expression of genes associated with distinct blood lineages and maturation stages. In contrast, MRD-like cells generated through in vivo treatment capture a unique cellular phenotype characterized by both an early multi-lymphoid developmental stage and diminished expression of MYC, E2F and their target genes.

Furthermore, we have demonstrated a previously unappreciated correlation between cell-cycle status and differentiation stage, and identified gene modules of coordinated differential methylation and expression involved in regulating these essential phenotypic traits that might explain their persistence throughout in vivo relapse. Crucially, we find that cells with the same transcriptional phenotype as MRD-like cells are already present at very low numbers in untreated leukemias, in both xenografted and direct clinical specimens.

Given the limited sample size and possible limitations associated with the use of a model system, it will be interesting to see how the relative contribution to resistance of genetic versus epigenetic heterogeneity plays out in a larger cohort of patients. In light of the strong association observed in clinical practice between cytogenetics and prognosis, such a cohort should span all cytogenetic subtypes, and encompass patients with varied time from diagnosis to relapse, because the clinical observation that some patients with BCP-ALL relapse during treatment whereas others relapse off-treatment might suggest different underlying selection processes. Whether or not our findings apply more broadly to different tumor types remains
to be assessed. Genetic heterogeneity might play a more predominant role in other cancers, particularly in fast-evolving tumors and/or spatially segregated solid tumors. Deterministic selection for genotypes is also likely to underlie treatment resistance in the context of targeted therapy, where earlier studies have indeed led to the identification of subsets of genetic variants directly involved in sensitivity to relevant inhibitors. Nevertheless, these findings argue for the importance of exploring the immediate post-treatment disease landscape and evaluating cell state as a determinant of resistance in a broader range of hematological and solid malignancies, and further advocate epigenetic cell state as a candidate target of any subsequent therapeutic interventions.

|          | Cycling (%) | ShallowQ (%) | DeepQ (%) |
|----------|-------------|-------------|-----------|
| PRS      | 0           | 23.3        | 76.7      |
| Untreated| 28.2        | 68.7        | 3.1       |
| Treated  | 0           | 4.9         | 95.1      |

|          | ETP (%) | CMP (%) | GMP (%) | Pro-B (%) | Early B (%) | MLP (%) |
|----------|---------|---------|---------|-----------|-------------|---------|
| PRS      | 0       | 0       | 0       | 43.3      | 33.3        | 23.3    |
| Untreated| 9.1     | 8.9     | 0.5     | 18.2      | 59.8        | 3.6     |
| Treated  | 0       | 0       | 0       | 47.9      | 22.1        | 30.1    |

Pseudotime-only PRS
Methods

Patients and samples. Diagnostic childhood BCP-ALL BM samples (Supplementary Table 5) were obtained from human participants aged 3–18 years on informed consent and approval by the relevant research ethics committees at John Radcliffe Hospital, Oxford, UK, and Centro Ricerche Tettamanzi, Clinica Pediatrica Universitaria Milano Bicocca, Italy.

BM reconstitution assay. Primary childhood BCP-ALL mononuclear cells isolated by ficoll gradient centrifugation were transplanted into 8–12-week-old NOD/SCID/IL2Rg−/− (NSG) sublethally irradiated mice (either females or males) via intramedullary injection. To minimize possible adverse effects of sublethal irradiation, mice were administered acidic water for a week before the procedure, and Baytril (resuspended at 25.5 mg kg−1 in the drinking water) for the 2 weeks after it. Sublethal irradiation was achieved with a single dose of 37.5 Gy. Each mouse received 2×10^7 primary leukemia cells resuspended in 40 μl of phosphate-buffered saline (PBS) + 5% fetal bovine serum (FBS). In the case of secondary limiting dilution assays, a specified equal dose of treated and control leukemia cells harvested from the BM of primary recipients was injected. Mice, 12 weeks post-injection, were sampled by BM aspiration, and the percentage of human engraftment was evaluated by flow cytometry (hCD45+/hCD45−/mCD45+).

At the same time, human cells were also FACS sorted for downstream applications. Mice displaying at least 70% human engraftment were randomly assigned to either control or treatment groups (see below for details on the treatment protocol). After 28d, tibias, femurs, pelvies, spleens and brains were harvested. Total BM and spleen cellularity were estimated through the Sysmex XN-300 Automated Hematology Analyzer, and all remaining cells were then stained for FACS sorting.

In vivo treatment protocol. All in vivo experiments were performed in strict accordance with the UK Home Office regulations. Mice were treated with pharmacological-grade vincristine and dexamethasone (vincristine sulfate 1 mg ml−1 injection (Hospira) and dexamethasone 2 mg tablets (Auden Mckenzies)). The treatment regimen was first optimized on mice engrafted with commercially available, healthy donor cord blood cells (Lomza) and consisted of 0.50 mg kg−1 of vincristine administered weekly intraperitoneally, and 6mg kg−1 of dexamethasone supplemented to the drinking water (continuous administration). Unless otherwise stated, treatment was administered for a total of 4 weeks.

Flow cytometry. In the case of material freshly harvested from mice, primary cells were treated with ACK buffer (0.15 M NH4Cl, 1.0 mM KHCO3, 0.1 mM ethylenediaminetetraacetic acid) to lyse red blood cells before staining. Primary patient samples were brought from liquid nitrogen to room temperature and washed in Dulbecco’s PBS. Single-cell suspensions were washed with FACS buffer (PBS + 10% FBS) and stained (15 min, 4 °C) with the appropriate fluorochrome-conjugated antibodies (Supplementary Table 6) diluted to optimal working concentration in PBS + 10% FBS. A wash in FACS buffer was performed before analysis. Appropriate unstained, single-color and fluorescence minus one controls were used for compensation set-up and to define the gating strategy. Leukemic cells were defined as hCD45+/p16−/CD34−cells as CD45+/CD34+/CD38−. Data were analyzed with FlowJo v.8.6 (Tree Star) software.

Karyotype preparation. Flow-sorted cells (up to 500×10^6) were collected in a 1.5-ml screwcap tube filled with 10–200 μl of FACS buffer (PBS + 10% FBS). Cells were centrifuged at 400g for 10 min and, after removing the supernatant, resuspended in 700 μl of pre-heated KCl (5.6 g in 1 l of purified water) and incubated at 37 °C for 15 min. Cells were then prefixed by adding 300 μl of ice-cold methanol–acetic acid fixing solution (three parts methanol and one part glacial acetic acid). Tubes were mixed by inversion and centrifuged at 400g for 10 min. After removing the supernatant, cells were fixed by adding 1 ml of ice-cold fixing solution dropwise while holding the tube on a vortex (speed 7 or 8). Cells were left to rest for 5 min before spinning down at 400g for 10 min. If needed, cells were stored at 4 °C (short term) or −20 °C (long term).

Cytokine aquarius probes for mFISH. Locus-specific FISH probes targeting chromosomes 9, 12, and 21 were customized and manufactured by CytoCell. Probes were directly labeled with different fluorochromes, allowing for the simultaneous detection of four different fluorescent signals on hybridization to target sequences (Supplementary Table 7).

The mFISH protocol. After karyotype preparation, frozen cell suspensions were pelleted by centrifugation at 400g for 15 min, washed in 800 μl of freshly prepared ice-cold fixing solution (prepared as above) and centrifuged again. Supernatants were removed and cell pellets were resuspended in 5–50 μl of ice-cold methanol–acetic acid fixing solution. Cells were spotted down onto a moist slide and allowed to air dry for a few minutes. Spotted slides were immersed in 2x saline–sodium citrate (SSC) for 5 min and washed 3x (1 min) in distilled water at room temperature (RT). Slides were then incubated in Pepsin Working Solution (Sigma–Aldrich; stock solution: 100 mg ml−1 diluted in distilled water (35 μl aliquots stored at −20 °C)). The working solution was 0.05 mg ml−1 diluted in 10 mM HCl applied at 37 °C for 2 min and washed 3x (1 min) in distilled water at RT. Slides were dehydrated through immersion in ethanol series (70%, 85%, 100%, 2 min each at RT) and air dried. Pre-denaturation of the probe and the sample was achieved by spotting warm probe mix (5 μl per slide, pre-heated for 5 min at 37 °C) over heated slides (5 min on a heat block set to 37 °C). A 22×22 mm2 coverslip (VWR) was applied over the spotted area, bubbles were removed by gentle pressure, and the coverslip was sealed with rubber glue (Fixogum) and left to dry. Simultaneous denaturation of probe and sample was achieved by placing slides on a hybridizer (ThermoBrite) at 75 °C for 2min, and was followed by overnight incubation in the hybridizer at 37 °C. The next day, cover slips were removed, and a series of post-hybridization washes was performed. Slides were incubated in 0.4x SSC at 72 °C for 2 min, followed by 30 s in 2x SSC/0.1% Igepal at RT. On air drying, slides were stained with 5 μl of Cytoseq DAPI/anti-fade and covered with a 22×22 mm2 coverslip.

Microscopes and probe visualization. A Zeiss Axio Observer z1 Apotome fluorescence microscope was used for the imaging and scoring of mFISH samples. The microscope was set up with commercially available DAPI (filter set 49, 488049-9901-000) and fluorescein isothiocyanate (FITC, filter set 38 HE, 498038-9901-000) filters purchased from Zeiss, and customized DEAC (49302), Texas Red (49008), GOLD (49034) and FITC/TXR (59022) filters purchased from Chroma. Slides were scanned overnight using Metafer (MetaSystem), an automated high-throughput software specialized in FISH image acquisition and processing. For each sample, 100–600 interphase nuclei were scanned and 100–250 cells manually scored for the presence of the ETV6–RUNX1 fusion gene, in combination with deletion (hemizygous or homozygous) or amplification of ETV6 (TEL), RUNX1 (AML1), PAX5 and CDKN2A (p16).

Establishing cutoff levels. Each probe hybridization efficiency was quantified, scoring unexpectedly abnormal signal patterns in two types of positive control samples: (1) unenriched peripheral blood mononuclear cells from normal peripheral blood and (2) CD19+ enriched cells from the BM of two normal karyotypes. In each case, >200 nuclei were scored for the presence/absence of the TEL–AML1 fusion gene as well as copy-number variants involving ETV6 (TEL), RUNX1 (AML1), PAX5 and CDKN2A (p16) genes. The mean percentage of cells with loss or amplification of any single gene signal was 0–1.87%. Combinatorial cutoff levels were obtained by two independent approaches: (1) visually scoring slides for the coexistence of any two lesions (cutoff of 0.45%) and (2) mathematical calculation of the likelihood of any two events happening at the same time based on scoring data collected for each probe (cutoff of 0.35%). Based on these results (Supplementary Table 8), a conservative threshold of 2% was set for a clone with any single additional CNA (compared with its closest predecessor) to be called; an additional requirement was also set requiring at least three cells with a given genetic makeup to be scored within a sample.

Small cell number RNA-seq. Small cell number RNA-seq from equivalent cell numbers (400 cells) was performed as described in Boiers et al.2.

Single-cell gene expression (qPCR) analysis. Single cells were sorted into 96-well PCR plates as described in Potter et al.25. Complementary DNA synthesis from sorted single cells and subsequent target probe pre-amplification with genes of interest were performed as described by Maignan et al.7. Gene probes used in these experiments included: CD4, CD58, CD64, CD14, CD123, p16, p21, CD68, CD11b, CD69, CD71, TRFR, CD3e, RUNX1 and ETV6–RUNX1 (TrapMan gene expression assays, Thermo Fisher Scientific).

ScRNA-seq. ScRNA-seq experiments were performed as previously described in Ghori et al.29. For each experiment, 1,500 single hCD45+ viable ( Hoechst, 332358 negative) leukemic cells were flow sorted directly into a 10- to 17-μm diameter C1-Integrated Fluidic Circuit (IFC; Fluidigm).

ScWGS. Fifteen hundred single hCD45+ viable (Hoechst, 332358 negative) leukemic cells were flow sorted directly into a 10- to 17-μm diameter C1-Integrated Fluidic Circuit (IFC; Fluidigm) preloaded with 3.5 μl of PBS + 0.5% bovine serum albumin (BSA). Post-sorting, the total volume was measured and measured set to PBS + 0.5% BSA, after which 1 μl of C1 Cell Suspension Resagent was added to the final solution. Each C1 IFC capture site was carefully examined using the EVOS FL Auto Imaging System (Thermo Fisher Scientific). An automated scan of all capture sites was also obtained for future reference. Cell lysing and whole-genome amplification from single cells were performed on the C1 Single-Cell Auto Prep IFC using the PicoPLEX WGA Kit (Rubicon Genomics). To this end, a customized script was generated through the C1 Script Builder.

Methylation analysis. Genomic DNA was isolated from phenol–chloroform extraction of primary human CD45+ BM cells harvested from untreated and treated mice. DNA quantitation and purity were determined by Nanodrop and Qubit. The EZ DNA methylation kit (Zymo Research, bisulfite conversion) was used for bisulfite conversion following the manufacturer’s instructions (final elution in 6 μl). Samples were analyzed at the UCL Genomics facility using the Infinium Human Methylation EPIC array (Illumina).
Processing of bulk RNA-seq. Bulk RNA-seq samples were processed using a nextflow pipeline (https://github.com/UCL-BLIC/rnaseq), which runs FastQC, TrimGalore!, STAR and featureCounts, to clean, map the reads to the human reference GRCh38.p12 and quantify gene expression using the Ensembl v84 annotations. Analyses were performed within the R statistical computing framework v3.5.1, using packages from Bioconductor v3.10 (https://Bioconductor.org). The DEseq2 Bioconductor package was used for outlier detection, normalization and differential gene expression analyses. PCA was derived using the top 1,000 most variable genes after DEseq2 vt transformation.

Processing of scRNA-seq. We used STAR to map the reads to the GRCh38 reference human genome, as included in Ensembl v84, RSEM to quantify transcript and gene expression abundance, scater for quality assessment and scan for normalization. All count data, metadata and intermediate results were kept within the SingleCellExperiment R object. Unless otherwise specified, all analyses were performed using log(transformed normalized counts). We used the t-SNE method implemented in the Rtsne package. The Seurat package was used to build the UMAP, cluster the cells and identify cluster gene markers (only among genes detected in at least 25% of the cells). Pseudotime reconstruction was obtained using the TSCAN package with ten clusters, as suggested by the SC3 package. The PRSamTREATED signature was derived using the SingleR package with the top 50 genes in a 3-way comparison, where we removed any genes identified as differentially expressed between PRS and treated cells.

Gene sets and gene set enrichment analysis. Gene signatures of potential biological interest were retrieved from the Hallmark dataset (MSigDB v6.1). These were complemented with gene signatures defining the different stages of hematopoietic lineage differentiation. Gene signatures identifying the cell-cycle marker were also included (Supplementary Table 4). This combined set of gene signatures was used for enrichment analysis, lineage determination, cell-cycle state determination, single-cell variant model analysis and correlation analysis between these different factors. The IGSEA package was used to determine gene set enrichment for each of the signatures.

Lineage classification for single cells. For hematopoietic lineage determination, we followed a previously published approach for assigning a lineage to each cell. In more detail, to achieve normal distribution, data were re-normalized as z-scores of log(TPM + 1); the lineage score of each signature within a given cell was then computed as the average value of all signature genes. In the single-lineage analysis, each cell was assigned to the lineage with the highest score. For multilineage analyses, a threshold of z-score = 0.75 was used as a cutoff for a positive lineage call.

Processing of scWGS. ScWGS data were processed using a nextflow pipeline (https://github.com/UCL-BLIC/nf-ginkgo; commit 2fb20ac) which ran at the time: FastQC v0.11.8, MultiQC v1.6, TrimGalore! v0.5.0, bwa mem v0.7.12-r1039, samtools v1.9 for sorting and indexing, Picard MarkDuplicates v2.18.9-SNAPSHOT, BEDTools v2.27.1 to generate BED files with read information and Ginkgo (https://github.com/robertoubkalhali/ginkgo; commit 892b2e9) to bin the reads, GC-correct and normalize the counts, produce per-cell quality control plots and initial copy-number (CN) profiles. Data were re-normalized using single细胞 obtained from two different normal samples (one male, one female) initially processed as above. Bins characterized as deletions, amplifications or displaying excessive variability among control cells were excluded. Initial CN events were detected for each sample using the multi-sample PCF (R package copynumber v1.14.0) with γ = 5. These CN events were manually inspected and compared with the raw data, and CN calls using single-sample PCF (γ = 10) to confirm the edges, adjust the CN thresholds (usually 0.5 for homozygous deletion, 1.5 for heterozygous deletion, 2.5 for 1-copy gain, and so on) and resolve possible overlaps between CN events. To mitigate the effect of choosing one common CN threshold for all cells, we sampled around each CN threshold (using a truncated normal distribution, s.d. = 0.05) and around each cell CN value using a normal distribution, s.d. = 0.025. Then CN calls and the number of the best 100 out of 10,000 trials to reduce the total number of clones in each sample. The results were then manually inspected to confirm the soundness of the selected thresholds and clone assignments. The final tree is inferred using getMinimumArborescence from the optrees v1.0r package. Pt.1’s PDX tree was trimmed to remove leaf nodes with a single cell. Fishplots were drawn using the R package Timescape v1.10.

Illumina Infinium EPIC data preprocessing. DNA methylation levels were measured using Illumina Infinium Methylation EPIC BeadChips (EPIC array) according to the manufacturer’s protocol. Data preprocessing steps were carried out using methods incorporated in the R packages minfi v1.20.1 and R and ChAMP. First, we filtered out probes based on the following criteria: (1) detection P < 0.01; (2) head count of <3 in at least 5% of samples; (3) non-CG probes; (4) containing single-nucleotide polymorphisms with a minor allele frequency of at least 1% in the European population (1,000 Genomes Project) within 5-bp of the probed CG; (5) mapping to multiple genomic locations; and (6) mapping to sex chromosomes. The filtering procedure resulted in 763,949 out of 866,238 probes. Next, we applied peak-based correction (PBC) normalization, a peak-based method, to correct for probe bias in Illumina Infinium Methylation data and reduce technical variation. We adjusted for batch effects due to multiple samples processed on each slide using the Combat function of the R package SVA. To assess data quality and identify further potential batch effects or outlier samples, we performed singular value decomposition to determine components of variation and applied MDS, PCA and hierarchical clusterings at all steps of the data preprocessing procedure.

Quantification of DNA methylation. DNA methylation values are provided as either M values or β values. M values are log, of the ratio of the intensities of the methylated probe versus the unmethylated probe on the EPIC array; β values are calculated as the ratio of the methylated probe intensity and the overall intensity. All analyses of DNA methylation data were performed on M values; β values, which have more straightforward interpretability, were used for the visualization of DNA methylation data in figures (that is, 0–100% DNA methylation). To quantify DNA M values and correct for the dependency of variability measurements on the mean, we adapted the method of Alemi et al. as previously described.

Analysis of differential methylation. We applied a paired limma model to identify differentially methylated positions. Statistical significance was defined as Benjamini–Hochberg corrected P < 0.05 and an absolute fold-change ≥1. DMRs were identified by bumphunter, setting the minimum number of neighboring probes to 4, the maximum length of a DMR to 300 bp and the number of permutations to 250. Statistical significance was again defined as Benjamini–Hochberg corrected P < 0.05.

Analysis of differential MV. We applied a combined statistical approach based on DiffVar3, which is embedded in the framework of limma, and the MV as previously described. Pairwise tests were used to identify differentially variable positions. Statistical significance was defined as Benjamini–Hochberg-corrected P < 0.05 and MV difference ≥10% relative to the observed range of MV values. Differentially variable regions were identified by DMRcate based on limma, setting the individual Benjamini–Hochberg corrected P-value threshold to 0.25, and considering regions with a minimum of four neighboring Cpgs. Statistical significance was defined by the default region-based P-value cutoff determined by DMRcate.

Methylation array pathway analysis. We analyzed the biological functions and pathway enrichment of flanking genes with GREAT using the standard parameters: association rule = baas + extension (proximal 5-kb upstream, 1-kb downstream, up to 1-Mb extension); curated regulatory domains = included.

Statistics and reproducibility. Statistics calculations were carried out in either Prism (v8.0) or R (v3.4.3). Power calculations and Monte-Carlo simulations based on earlier data were used to inform the design of the in vivo experiments. Investigators were not blinded to allocation during the experiments and outcome assessment. Samples for transcriptional analysis were chosen based on the appropriate cellularity of the specimen. Mice assignment to either the control or the treatment group was randomized. Where appropriate, P-values were adjusted using the Benjamini–Hochberg method to control the type 1 error rate in the context of multiple testing. Any scWGS samples were excluded if they had a median absolute pairwise difference > 0.6, fewer than 500,000 reads and a median number of reads per gen smaller than 10, a Gini index > 0.35, or displayed unusual profiles in the quality control plots produced by Ginkgo. Three bulk RNA-seq samples were excluded from the analysis. These include one Pt.1 untreated, one Pt.2 acutely treated and one Pt.12 treated sample because they failed to cluster with the related samples in PCA. ScRNA-seq libraries with fewer than 1 million reads or fewer than 2,000 genes expressed were removed. In the downstream analysis, genes with more than 20 read counts in fewer than 6 cells (scRNA-seq) or samples (bulk RNA-seq) were also excluded.

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

Data availability
Bulk and single-cell DNA- and RNA-seq and methylation array data (Supplementary Table 9) that support the findings of the present study have been deposited in the European Genome–phenome Archive (EGA), which is hosted by the European Bioinformatics Institute and the Centre for Genomic Regulation, under accession number EGAD00001004047. Previously published WGS data from refs. 20–23 are available under accession identifiers EGAD00001000776 and EGAD00001001636. Source data are provided with this paper.

Code availability
Code used for this analysis is available at https://github.com/UCL-BLIC-analysis/Turai_NatCancer_2021.
63. Ritchie, M. E. et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
64. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
65. Shipson, B. & Oshlack, A. DiffVar: a new method for detecting differential variability with application to methylation in cancer and aging. *Genome Biol.* **15**, 465 (2014).
66. Peters, T. J. et al. De novo identification of differentially methylated regions in the human genome. *Epigenet. Chromatin* **8**, 2–16 (2015).
67. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
68. Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
69. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).
70. Lun, A. T. I., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* **5**, 2122 (2016).
71. Laurenti, E. et al. Cell stem cell hematopoietic stem cell function and survival depend on c-Myc and N-Myc activity. *Stem Cell* **3**, 611–624.
72. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

**Acknowledgements**

We thank all clinicians and the Blood Cancer UK Leukaemia Cell Bank for samples, Y. Guo for the assistance with the in vivo work, B. Gaal for sharing her MATLAB expertise, I. Titley and A. Ford for cell sorting and Ki67 analysis of single quiescent cells, and Fluidigm for technical support. We also thank the UCL-Cruck-funded flow cytometry, microscopy and genomics facility for access to instrumentation and for practical help with FACS analysis (W. Day, Y. Guo, B. Wilbourne and G. Morrow). We thank the UCL Legion and Myriad High Performance Computing Facility (Legion@UCL; Myriad@UCL), and associated support services, for the use of their services in completion of this work. The present study was supported by Blood Cancer UK project grants (to T.E., M.G. and S.E.W.J.) and partially supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (grant no. FC001202), the UK Medical Research Council (grant no. FC001202) and the Wellcome Trust (grant no. FC001202). V.A.T. was funded by grants from Children with Cancer UK and Blood Cancer UK, N.P. was founded by Gabriel’s Angels, and J.A.G.-A. and L.C. were funded by the Cancer Research UK-UCL Centre Award. M.T. is a post-doctoral fellow supported by the European Union’s Horizon 2020 research and innovation program (Marie Skłodowska-Curie grant agreement no. 747852-SIOMICS). P.V.L. is a Winton Group Leader in recognition of the Winton Charitable Foundation’s support toward the establishment of the Francis Crick Institute and M.G. was supported by the Wellcome Trust. L.J.R was funded by the Kay Kendall Leukaemia Fund and Leuka, G.C. was funded by the Italian Association for Cancer Research grants (IG2017 no. 20564 to A.B. and IG2015 no. 17593 to G.C.). A.B. and G.C. were funded by the Comitato Maria Letizia Verga and S.E.W.J. by the UK Medical Research Council (grant no. MC_UU_12009/5).

**Author contributions**

V.A.T., T.E., M.G. and S.E.J conceived the study. V.A.T., M.L., J.B., A.W. and M.H. conducted the methodology. V.A.T., N.P., I.M. and A.D. did the investigations. J.A.G.-A., M.T., P.V.L., S.B. and J.H. carried out the formal analysis. J.A.G.-A., A.W., S.E., C.D., C.J., C.L. and J.H. curated the data. S.E.J., G.W.H., A.B., L.R., S.I., P.A., G.C. and L.R. provided the resources. T.E., S.E.J and M.G. supervised the project. V.A.T., R.G., G.M., T.E. and M.G. wrote the manuscript. T.E. and M.G. acquired the funding.

**Competing interests**

M.L. was an employee of rpg Fluidigm Corporation at the time of the present study. The other authors declare no competing interests.

**Additional information**

Extended data is available for this paper at https://doi.org/10.1038/s43018-021-00219-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43018-021-00219-3.

Correspondence and requests for materials should be addressed to V.A.T. or T.E.

Peer review information *Nature Cancer* thanks the anonymous reviewers for their contribution to the peer review of this work.

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

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021
Extended Data Fig. 1 | Single cell whole genome sequencing analysis. 

**a.** Representative traces of two whole-genome sequenced cells. Mapped reads are plotted across all chromosomes to highlight variations in copy number status as changes in the ploidy of the cells ($n = 2$). 

**b.** Diagram explaining the data processing pipeline. 

**c.** Scatterplots comparing the average normalized, GC-corrected and ploidy-adjusted counts per bins for pairs of copy-number events, where each dot represents an individual cell ($n = 420$ cells) harvested from Pt1 PDXs (across all treatment conditions and timepoints). Red lines divide the plots in 4 quadrants based on the nominal thresholds for calling each CNA.
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Single cell gene expression analysis and mFISH analysis. 

a, Representative plots describing the flow cytometry sorting approach used to collect CD34+/CD38-/CD19+ cells. DAPI and forward scatter were used to identify viable cells (gate p1). To ensure that no dying cells were sorted, the p1 gate was further refined based on cells staining with the cell viability dye carboxyfluorescein succinimidyl ester (CFSE - FITC channel) (gate p2 - used to sort bulk cells). Positive vs negative combinations of specific antibodies were used to identify the populations of interest: CD34+/CD38-/CD19+ were collected from gate p4.

b, Representative heat map from a FLEXsix single-cell Q-PCR transcription panel showing the expression level of selected relevant transcripts in 11 single-cells sorted from the same patient as negative for CD38 expression. UBC is polyubiquitin-C used as a housekeeping gene control, CB8 is a single normal cord blood control cell. c and d, Phylogenetic trees showing the subclonal architectures of Pt10 and Pt11 bulk and CD34+/CD19+/CD38-low sorted cells.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | mFISH analysis. **a**, Flow-sorting strategy used to assess engraftment and isolate human leukemic cells (hCD45+) from mouse cells (mCD45+). **b**, Representative Multicolor FISH image. The subclonal variegation within each sample was assessed using fluorochrome-labelled custom-made probes detecting AML1 (RUNX1), TEL (ETV6), p16 and PAX5 genes. **c** and **d**, Tables showing the copy number status associated with each of the subclones identified within Pt2 and Pt3 engrafted leukemias. **e-f**, BM engraftment data before and after chemotherapy for Pt2 (n=11 control mice and n=8 treated mice) and Pt3 (n=10 control mice and n=13 treated mice) at d0 and d28 as assessed by flow cytometry. Note engraftment is expressed here in the conventional way as hCD45+/hCD45+/mCD45+; the >10 log cytoreduction achieved by chemotherapy is obscured in this representation due to the accompanying decrease in mouse CD45+ cells. Data are presented as mean values +/− SD. The statistical significance of each comparison (Paired t-test, two-tailed) is shown (p-value). **g**, Absolute numbers of human white blood cells (WBC) retrieved at d28 from both control (n=5 mice) and treated mice (n=8 mice) as quantified by the XP-300TM Automated Haematology Analyser (SYSMEX Co., Japan). Data are presented as mean values +/− SD. **h**, Dot plot showing treatment responses across different metastatic sites (brain, liver, spleen), with representative images of the reduction in spleen size following treatment (n=2 control mice and n=5 treated mice). Data are presented as mean values +/− SD. Error bars are shown for treated samples where n≥3. **i**, Treatment response kinetics of xenografts derived from high and low risk patients. Response was assessed by flow cytometry comparing human chimerism at d7 and d28 to d0 engraftment. In the plot each line represents a different patient (including some of the patients used for downstream analysis (pt1, pt2, pt3)). Data are presented as mean values +/− SD (n=3 mice). **j**, Graph tracking the size of individual subclones from Pt2 xenografted leukemias before (d0 BM aspiration) and after treatment (d28 BM and d28 spleen). Multiple lines of the same color highlight the kinetic of the same subclone across different treated xenografts.
Extended Data Fig. 4 | scWGS analysis. a-b, TimeScape plots visualizing the clonal composition of two untreated (Ctrl) and two treated mice at 3 timepoints: i) in the primary diagnostic leukemia ii) at d0 in the xenograft (BM aspirate) and ii) at d28 in the xenograft (total BM harvest). c, Hierarchical classification of the main clones identified across all analyzed Pt1 specimens (primary diagnosis and xenografts d0-d28); provides a legend for the color coding used for the TimeScape plots in panels a and b (left) as well as Fig. 4d-e.
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | Population RNAseq analysis. a-b, Graphical representations of the standard and multi-arm treatment models and resulting PCA analysis. In the standard model (a) matching d0 and d28 cells harvested from untreated (n = 37) and treated (n = 17) mice were analyzed. Untreated samples, shown in light blue, represent d0 cells from untreated mice and treated mice as well as d28 samples from untreated mice. In the multi-arm model (b), two consecutive rounds of transplantation and treatment (7 days) allowed the analysis of ‘acutely treated’ (n = 8 samples; green), ‘treatment withdrawn’ (n = 6 samples; dark yellow), ‘chronically treated’ (n = 5 samples; red), and ‘untreated control’ (n = 5 samples; dark blue) cells harvested from n = 3 mice each (see main text). ‘Relapse-like’ cells were also obtained through a separate arm of the experiment, whereby n = 3 mice were treated for 28 days and then let to relapse in situ in the absence of further treatment (6–8 weeks). c, Untreated and treated cells from primary recipients of Pt2 cells were transplanted into secondary recipients at three different limiting doses (100, 1*10³, 1*10⁴ cells). Engraftment (%) of BM, spleen and brain was assessed by flow cytometry and is shown in the dot plot (n = 4 mice for control; n = 5 mice for treated 100 and 1*10³ cells; n = 4 mice for treated 1*10⁴ cells). d, Estimated LIC frequencies for treated and untreated cell populations from each analyzed tissue (n = 4 mice for control; n = 5 mice for treated 100 and 1*10³ cells; n = 4 mice for treated 1*10⁴ cells) as calculated using the ELDA software⁶⁹. In the top plot the number of initially injected cells (x-axis) is plotted against the log fraction of non-responders, corresponding to mice with no engraftment (y-axis).
## Extended Data Fig. 6 | Single cell RNA sequencing analysis of fresh diagnostic samples.

### a-b

Analysis of the biological states, processes and lineages underpinning the intercellular transcriptional heterogeneity amongst diagnostic cells from either Pt1 or Pt2 using the bioinformatics tool scran to decompose gene expression heterogeneity into technical noise and a biologically relevant component. A list of genes ranked by their biologically meaningful variance was then used as input for enrichment analysis with the fGSEA R-package using the Hallmark pathway gene sets (MSigDB) and lineage markers defined in. The importance of each given signature in explaining variation amongst untreated cells is reflected in the normalized enrichment score (NES) reported next to the signature itself. P-values are obtained using 2-sided permutation test with 10,000 permutations. All shown processes have a Benjamini-Hochberg FDR-adjusted p-value $< 0.05$.

### c

Heatmap showing the expression of gene sets associated with different lineages and differentiation stages of the hematopoietic hierarchy. Each line of the map represents an individual diagnostic cell from either Pt1 or Pt2.

### d

Bar plots visualizing the percentage of individual diagnostic cells from either Pt1 or Pt2 assigned to their closest normal hematopoietic cell lineage genes (signature with the highest Z-score average). The analysis is based on the expression of known markers.

---

### Table: Gene ranks and NES for Patient 1 and Patient 2

| Pathway                          | Gene ranks | NES   |
|----------------------------------|------------|-------|
| MYC_TARGETS_V1                   | 2.65       |       |
| MTORC1_SIGNALING                 | 2.20       |       |
| DNA_REPAIR                       | 2.07       |       |
| FATTY_ACID_METABOLISM            | 1.83       |       |
| OXIDATIVE_PHOSPHORYLATION        | 1.90       |       |
| PROTEIN_SECRETION                | 1.71       |       |
| PI3K_AKT_MTOR_SIGNALING          | 1.70       |       |
| CHOLESTEROL_HOMEOSTASIS          | 1.55       |       |
| G2M_CHECKPOINT                   | 1.85       |       |
| PEROXISOME                       | 1.62       |       |
| INTERFERON_ALPHA_RESPONSE        | 1.65       |       |
| ANDROGEN_RESPONSE                | 1.59       |       |
| APOPTOSIS                        | 1.70       |       |
| HYPOXIA                          | 1.71       |       |
| KRAS_SIGNALING_UP                | 1.65       |       |
| UNFOLDED_PROTEIN_RESPONSE        | 1.65       |       |
| E2F_TARGETS                      | 1.78       |       |
| IL2_STAT5_SIGNALING              | 1.69       |       |
| TNFA_SIGNALING_VIA_NFKB          | 1.69       |       |
| UV_RESPONSE_UP                   | 1.56       |       |
| ALLOGRAFT_REJECTION              | 1.51       |       |
| COMPLEMENT                       | 1.48       |       |
| ProB                             | 1.56       |       |
| ADIPOGENESIS                     | 1.50       |       |
| GLYCOLYSIS                       | 1.43       |       |
| P53_PATHWAY                      | 1.44       |       |
| INTERFERON_GAMMA_RESPONSE        | 1.41       |       |

| Pathway                          | Gene ranks | NES   |
|----------------------------------|------------|-------|
| MYC_TARGETS_V1                   | 2.55       |       |
| OXIDATIVE_PHOSPHORYLATION        | 2.29       |       |
| G2M_CHECKPOINT                   | 2.13       |       |
| KRAS_SIGNALING_UP                | 2.01       |       |
| PROTEIN_SECRETION                | 1.95       |       |
| APOPTOSIS                        | 2.04       |       |
| TNFA_SIGNALING_VIA_NFKB          | 2.08       |       |
| TGF_BETA_SIGNALING               | 1.68       |       |
| UNFOLDED_PROTEIN_RESPONSE        | 1.91       |       |
| PEROXISOME                       | 1.95       |       |
| FATTY_ACID_METABOLISM            | 1.89       |       |
| HYPOXIA                          | 1.74       |       |
| UV_RESPONSE_UP                   | 1.87       |       |
| E2F_TARGETS                      | 1.87       |       |
| ANDROGEN_RESPONSE                | 1.65       |       |
| IL6_JAK_STAT3_SIGNALING          | 1.94       |       |
| PI3K_AKT_MTOR_SIGNALING          | 1.71       |       |
| IL2_STAT5_SIGNALING              | 1.71       |       |
| P53_PATHWAY                      | 1.71       |       |
| ADIPOGENESIS                     | 1.54       |       |
| ESTROGEN_RESPONSE_LATE           | 1.54       |       |
| G2M_CHECKPOINT                   | 1.54       |       |
| ProB                             | 1.54       |       |
| UV_RESPONSE_DN                   | 1.39       |       |
| GMP                              | 1.49       |       |
| ALLOGRAFT_REJECTION              | 1.42       |       |
| GLYCOLYSIS                       | 1.43       |       |
| INTERFERON_GAMMA_Response        | 1.44       |       |
| XENOBIOLOGIC_METABOLISM          | 1.37       |       |
| ETP                              | 1.36       |       |

---

**Legend:**
- **Pt1**: Pt1 samples.
- **Pt2**: Pt2 samples.
- **EarlyB**: Early B cell lineage.
- **ProB**: Pro-B cell lineage.
- **CMP**: Common myeloid progenitor.
- **GMP**: Granulocyte macrophage progenitor.
- **MLP**: Multi-lineage progenitor.
- **MEP**: Megakaryocyte-erythrocyte progenitor.

**Lineage:**
- **EarlyB**: Early B cell lineage.
- **ProB**: Pro-B cell lineage.
- **CMP**: Common myeloid progenitor.
- **GMP**: Granulocyte macrophage progenitor.
- **MLP**: Multi-lineage progenitor.
- **MEP**: Megakaryocyte-erythrocyte progenitor.

---

**References:**
- Hallmark pathway gene sets (MSigDB).
- Pommeringer et al.
- Betteridge et al.
Extended Data Fig. 7 | Single cell RNA sequencing lineage analysis. Series of t-SNE plots of n = 448 untreated and n = 163 treated cells showing the positive (TRUE) or negative (FALSE) assignment of each lineage across different regions of the plot. Each cell is colored in red if it expresses genes associated with the named lineage (>0.75 average z-score over genes in the signature) or in grey if it does not express the signature or expresses it below the cut-off level. The plots do not show relative expression levels but aim to highlight how the same cell can have concomitantly high expression of gene markers characteristic of different lineages.
Extended Data Fig. 8 | Single cell RNA sequencing clustering analysis. a, UMAP clustering of the single cell RNA-seq xenograft data. The dimensionality reduction tool identifies 4 main clusters of cells. Clusters 0, 2 and 3 correspond to different states of untreated cells, while cluster 1 mostly comprises chemotherapy-treated cells. b, Heatmap showing the top positively differentially expressed features that characterize each of the UMAP clusters. c, UMAP plots visualizing the expression of selected biomarkers for each of the clusters. Images produced using the Seurat analysis package72.
Extended Data Fig. 9 | Boxplots of scRNAseq gene expression signatures. Plot showing the expression of relevant MSigDB 50 Hallmark gene sets across single cells grouped by treatment status (untreated (n = 448 cells) and treated (n = 163 cells)). The median expression per class is shown as a solid black line within each box. The box covers the 25th–75th percentiles range; whiskers cover all data points within a 1.5x the inter-quantile ranges from the bounds of the box; remaining data points are plotted as outliers.
Extended Data Fig. 10 | See next page for caption.
Extended Data Fig. 10 | Single cell RNA sequencing cell cycle, lineage and pseudotime analysis. **a**, Violin plots showing the expression of E2F1 and MYC genes in \( n = 448 \) untreated and \( n = 163 \) treated cells. Both genes are on average 10 log2 folds downregulated in treated cells. **b**, Stacked bar plot comparing the inferred lineage of each \( n = 448 \) untreated and \( n = 163 \) treated single cell with its cell cycle status as determined by the expression of G1/S, G2/M, MYC targets, and E2F targets signatures. **c**, TSCAN analysis displaying a PCA analysis of the data (\( n = 611 \) cells) and inferred pseudotime. The TSCAN package uses a cluster-based minimum spanning tree (MST) approach to order cells based on the gradual transition of their transcriptomes\(^{39}\). A line on the PCA highlights the constructed pseudo-temporal path. The table shows the number of treated and untreated cells within each cluster. The numbers of treated and untreated cells analysed are shown in the histograms at the top of the figure. **d**, Plot showing the distribution of deep, shallow quiescent and cycling cells (\( n = 611 \) cells), and lineage scores, across the TSCAN pseudotime clusters. **e**, Plots displaying the dynamic expression along the inferred pseudo-temporal path of representative genes with a known role in B-cell differentiation. Each cell (\( n = 611 \) cells) is shown as a circle (treated) or a triangle (untreated), the color of which reflects the cluster it belongs to (see panel **c**). **f**, Heatmap showing the top variable genes that contribute to clustering of \( n = 448 \) untreated and \( n = 163 \) chemotherapy-treated cells. PRS cells (\( n = 10 \) cells) as defined by this approach are evidenced by black arrows. **g**, Violin plots showing the expression of mitochondrial genes in treated, PRS and untreated cells as defined by either pseudotime (\( n = 163, 30 \) and 418 cells respectively; left plot) or t-SNE analysis (\( n = 163, 10 \) and 428 cells respectively; right plot).
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 our Editorial Policies 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  multicolor-FISH slides were scanned overnight using Metafer4 v3.12.7 (MetaSystems), an automated high-throughput software specialized in FISH images acquisition and processing.

Data analysis  For single cell WGS: FastQC v0.11.8; MultiQC v1.6; TrimGalore! v0.5.0; bwa mem v0.7.12-r1039; samtools v1.9 for sorting and indexing; Picard MarkDuplicates v2.18.9-SNAPSHOT; BEDTools v2.27.1 to generate BED files; Ginkgo (https://github.com/robertaboukhalil/ginkgo); commit 892b2e9) to generate initial profiles and QC plots; R package copyrumber v1.14.0; getMinimumArborescence from the R package optrees v1.0; R package Timescape v1.10.0 for the fishplots.

For bulk RNA-seq: FastQC v0.11.8; TrimGalore! v0.5.0; STAR v2.7.0d; featureCounts from R package subread v1.6.3; R package DESeq2 v1.26.0.

For single cell RNA-seq: STAR v2.5.2b; RSEM v1.2.31; R packages scater (v1.10.1) for PCA, scran v1.14.6, Rtsne v0.13, Seurat v3.1.4, SC3 v1.8.0, TSCAN v1.10.2, scater v1.14.6 and SingleR v1.0.6.

For Gene Set Enrichment Analysis, R package FGSEA v1.12 and R package msigdb v7.0.1 to download the MSigDB v7.0 Hallmark gene sets.

For EPIC methylation arrays: R packages minfi 1.28.4, ChAMP v2.12.4, sva v3.30.1 for comBat; limma v3.38.3, bumpHunter v1.24.5, missMethyl v1.16.0, DMRcate v1.18.0 and rGREAT v1.14.0

For Jensen-Shannon divergences: R package philantropy v 0.3.0.

For Shannon entropies: entropy shrink from R package entropy v1.2.1.
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-list.pdf

Life sciences study design

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

Sample size
Across the whole manuscript, 13 primary leukemias and more than 150 mice were analyzed. A series of power calculations and Monte-Carlo simulations based on earlier data (see Anderson et al. 2011) were used to inform the design of the in vivo experiments.

Data exclusions
scWGS samples were excluded if they had an MAPQ > 0.6, fewer than 500,000 reads, a median number of reads per bin smaller than 10; a Gini index > 0.35 or if they displayed unusual profiles in the QC plots produced by Ginkgo. Three bulk RNAseq samples were excluded from the analysis. These include one PT1 untreated, one PT2 acutely treated and one PT12 treated sample because they failed to cluster with the related samples in a PCA analysis. Single cell RNAseq libraries with fewer than 1 million reads or with expression detected for fewer than 2000 genes were removed. In the downstream analysis, genes with more than 20 read counts in fewer than 6 cells (scRNA-seq) or samples (bulk RNA-seq) were also excluded.

Replication
All experiments were designed to include appropriate numbers of biological replicates. Appropriate statistical analysis was conducted on the results and reported in the manuscript text/figures/figure legends or methods section. All findings were reproducible. One of mice engrafted with pt1 cells as part of our single-cell whole-genome-sequencing analysis, displayed enrichment for a subset of minor diagnostic subclones; suggesting that the pattern of engraftment in this recipient might be the result of sampling error at the time of injection. This difference is appropriately described in the text. The mouse was not excluded from analysis.

Randomization
Engrafted mice allocation into control or treatment cohorts was random.

Blinding
Investigators were not blinded to allocation during the experiments and outcome assessment.

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 and archaeology |
| ☑   | Animals and other organisms |
| ☑   | Human research participants |
| ☑   | Clinical data         |
| ☑   | Dual use research of concern |

**Methods**

| n/a | Involved in the study |
|-----|-----------------------|
| ☑   | ChiP-seq              |
| ☑   | Flow cytometry        |
| ☑   | MRI-based neuroimaging |
Antibodies

| Antibodies used                          |
|----------------------------------------|
| MscDA45, F11C, 1:100, Rat IgG2b, k, clone 30-F11, Biolegend, Cat. 303308 |
| hCD34-APC, 1:160, Ms IgG1, Clone H30, Biolegend, Cat. 204023           |
| hCD34-APC, 1:160, Ms IgG1, Clone BiRAMA-K3, DAKO-Agilent, Cat. C7238   |
| hCD38-APC/e-Fluor780, 1:80, Ms IgG1, clone k HIT-2, eBiosciences, Cat. 47-0389-42 |
| hCD19-PB, 1:50, Ms IgG1, clone LT19, Biorad, Cat. MCA1940PB           |
| hCD19-APC, 1:50, Ms IgG1k, clone HB19, Biolegend, Cat. 302212          |
| CFSE - Cell Labeling Kit, Abcam, Cat. ab113853                        |

Validation

To determine the best antibody concentration all antibodies were titrated on normal CB cells and leukemic cell lines. Panel optimization, set up and validation was performed using Thermo Fisher UltraComp ebeads (01-222-41) FMO controls. Single-cells hCD45+/mCD45- were sorted. An example of the gating strategy used in the cell sorting experiments is provided in Extended Data Fig.3a.

From the manufacturers’ website:
- Biolegend: ‘Each lot of this antibody is quality control tested by immunofluorescent staining with flow cytometric analysis.’ [Wu Y, et al. 2013; Peterson VM, et al. 2017; Zhu Y et al. 2017; Enggård J, et al. 2018].
- Biorad: ‘In house manufactured antibodies that have ‘flow cytometry’ indicated on the datasheet have been assessed to ensure application suitability. This includes testing the antibody for performance in flow cytometry using a relevant cell line, or primary cells, known to express the marker under appropriate conditions. They also have to meet internal criteria including batch-to-batch consistency for dilution, percentage of positive cells, and shift’ [de Rie et al. 1989; Bradbury et al. 1992; Horváth et al. 1998; Pezzutto et al. 1987].
- DAKO-Agilent: [Cho et al.2017; Sato et al.2017]
- eBioscience: ‘This antibody has been pre-titrated and tested by flow cytometric analysis of normal human peripheral blood cells.’ [Kong et al. 2017; Fernandez et al. 2014; Radomska et al. 2012].

Animals and other organisms

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

Laboratory animals

The study involved ex vivo transplantation of primary leukemic cells into 8-12 weeks old NOD/SCID IL2Rgamma (NSG) sub-lethally irradiated mice, both females and males. Mice were bred in isolators and maintained individually ventilated cages in a room set to a 14-hour light/10-hour dark cycle with temperatures of 18-23°C and 40-60% humidity.

Wild animals

The study did not involve any wild animals.

Field-collected samples

The study did not involve samples collected in the field.

Ethics oversight

All in vivo experiments were performed in strict accordance to the United Kingdom Home Office regulations (Project licence n. PFEC1FABA).

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

Diagnostic childhood BCP-ALL bone marrow (BM) samples were obtained from children diagnosed with BCP-ALL. Both female and male patients were enrolled in the study aged 3-18. Full patients demographics are available as a table in Supplementary Data Table 1.

Recruitment

Participants were not prospectively recruited for this study; rather specimens stored as part of a number of different clinical Biobanks across the UK and Italy were made available through clinician collaborators (see below). No selection bias has been identified thus far.

Ethics oversight

Samples were obtained upon informed consent and approval of relevant research ethics committees from patients at John Radcliffe Hospital, Oxford, UK and Centro Ricerca Tettamanzi, Clinica Pediatrica Universitaria Milano Bicocca, Italy.

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

Plots

Confirm that:
☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
☒ All plots are contour plots with outliers or pseudocolor plots.
☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation
In the case of material freshly harvested from mice, primary cells were treated with ACK buffer (0.15M NH4Cl, 1.0mM KHCO3, 0.1mM EDTA) to lyse red blood cells prior to staining. Primary samples were instead brought from liquid nitrogen to room temperature and washed in Dulbecco’s phosphate buffered saline (DPBS). Single cell suspensions were washed with FACS buffer (PBS supplemented with 1% FBS). Cells were stained (15min at 4°C) with the appropriate fluorochrome-conjugated antibodies diluted to optimal working concentration in PBS 10K FBS. A wash in FACS buffer was performed prior to analysis. Appropriate unstained, single color and FMO controls were used for compensation set-up (using the “automatic” compensation feature of BD FACSdiva software), and to define the gating strategy.

Instrument
Flow cytometry experiments were performed on LSR II, FortessaX20, FACS Aria III and FACS Aria Fusion (all BD Biosciences).

Software
Data was collected with BD Diva software and was analyzed with FlowJo v8.6 (Tree Star) software.

Cell population abundance
In all flow cytometry experiments cell were first gated based on FSC and SSC, doublet exclusion, and viability (based on Hoechst-33342 staining). Human-CD45+ cells in the BM of untreated mice at d0, untreated mice at d28 and treated mice at d0 represented 70-100% of all viable cells analyzed. human-CD45+ cells in the BM of treated mice at d28 represented 0.1-10% of total cells. The size of the CD34+, CD13+/CD38− compartment in fresh diagnostic samples varied greatly between individuals (0.1-5%).

Gating strategy
Leukemic cell harvested from the BM of xenografts were defined (and sorted) as hCD45+. Quiescent leukemic cell from fresh diagnostic samples were defined (and sorted) as CD34+/CD13+/CD38− represented.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.