

Studying clonal dynamics in response to cancer therapy using high-complexity barcoding

Hyo-eun C Bhang¹, David A Ruddy², Viveksagar Krishnamurthy Radhakrishna¹, Justina X Caushi¹, Rui Zhao^{3,4}, Matthew M Hims¹, Angad P Singh¹, Iris Kao², Daniel Rakiec², Pamela Shaw², Marissa Balak², Alina Raza¹, Elizabeth Ackley², Nicholas Keen¹, Michael R Schlabach¹, Michael Palmer², Rebecca J Leary¹, Derek Y Chiang¹, William R Sellers¹, Franziska Michor^{3,4}, Vesselina G Cooke¹, Joshua M Korn^{1,5} & Frank Stegmeier^{1,5}

Resistance to cancer therapies presents a significant clinical challenge. Recent studies have revealed intratumoral heterogeneity as a source of therapeutic resistance. However, it is unclear whether resistance is driven predominantly by pre-existing or *de novo* alterations, in part because of the resolution limits of next-generation sequencing. To address this, we developed a high-complexity barcode library, ClonTracer, which enables the high-resolution tracking of more than 1 million cancer cells under drug treatment. In two clinically relevant models, ClonTracer studies showed that the majority of resistant clones were part of small, pre-existing subpopulations that selectively escaped under therapeutic challenge. Moreover, the ClonTracer approach enabled quantitative assessment of the ability of combination treatments to suppress resistant clones. These findings suggest that resistant clones are present before treatment, which would make up-front therapeutic combinations that target non-overlapping resistance a preferred approach. Thus, ClonTracer barcoding may be a valuable tool for optimizing therapeutic regimens with the goal of curative combination therapies for cancer.

Although targeted cancer therapies often yield impressive initial responses, tumors frequently develop resistance^{1–3}. Resistance is generally thought to occur through the acquisition of *de novo* mutations during cancer therapy. An alternative hypothesis posits that rare resistant clones exist in the tumor mass before treatment, and that these clones drive tumor relapse after therapeutic challenge. This hypothesis has significant clinical implications, as in such cases resistance to many cancer drugs could be considered a *fait accompli*, and diagnostic and therapeutic strategies would need to be modified to better detect and target pre-existing resistant clones. Specifically, sequential resistance treatments based on the results of a post-treatment biopsy would not address the diversity of intratumoral resistance mechanisms and hence would lead only to short-term treatment benefits. This hypothesis would require a deeper sampling of resistance mechanisms and the elucidation of non-cross-resistant combination therapies that could be applied with curative intent in the frontline therapeutic setting.

Intratumoral clonal heterogeneity is thought to have a role in therapeutic resistance^{4–6}, and some studies suggest that rare genetic subclones may exist before therapy^{6–10}. However, one drawback of the approaches used in previous studies is their limited detection sensitivity. The resolution of standard next-generation sequencing (NGS) applications, for instance, is limited to a sensitivity of 0.1% allelic fraction because of sequencing error rates¹¹. Given that the detectable

tumor burden is estimated to be approximately 10⁹ tumor cells at the time of diagnosis⁷, this level of resolution is clearly insufficient to comprehensively assess pre-existing cancer subpopulations. In addition, epigenetic heterogeneity would not be detected by cancer genome-sequencing approaches. Current NGS approaches also lack the ability to track the fate of individual cells in heterogeneous cancer-cell populations over time.

We reasoned that cellular barcoding, which has been used to trace lineage during hematopoietic stem cell differentiation^{12–14}, could be used to address this question and overcome the limited sensitivity of current NGS approaches. We developed a high-complexity DNA barcode library, ClonTracer, that allows for the labeling of more than 1 million individual cells with a unique sequence tag. The molecular barcodes then enable the tracking of individual cancer cells en masse and provide a means to monitor clonal population dynamics in response to therapeutic challenge.

RESULTS

Development of ClonTracer barcode library

The complexity of previously reported DNA barcode libraries was restricted to approximately 10⁵ unique barcodes^{12–14} and was thus insufficient to uniquely tag individual cells in complex tumor-cell populations. We therefore set out to engineer a DNA barcode library with more than 10⁷ unique barcodes using a 30-nucleotide-long semi-random

¹Oncology Disease Area, Novartis Institutes for Biomedical Research, Cambridge, Massachusetts, USA. ²Translational Clinical Oncology, Novartis Institutes for Biomedical Research, Cambridge, Massachusetts, USA. ³Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. ⁴Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. ⁵These authors contributed equally to this work. Correspondence should be addressed to F.S. (frank.stegmeier@novartis.com).

Received 16 September 2014; accepted 17 March 2015; published online 13 April 2015; doi:10.1038/nm.3841

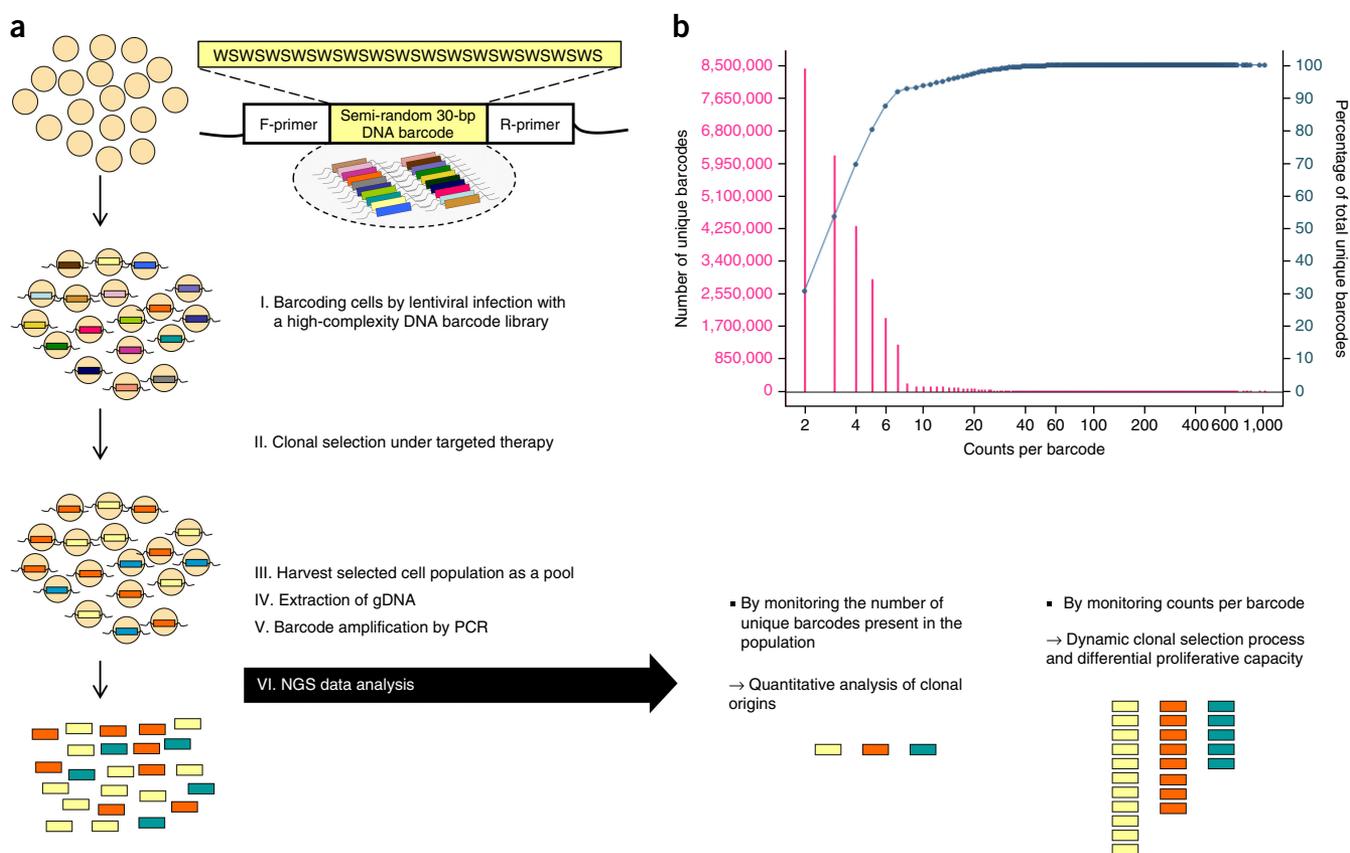


Figure 1 Design and characterization of the high-complexity ClonTracer DNA-barcoding technology. **(a)** Schematic depiction of the barcoding strategy used to track clonal dynamics in response to targeted therapies. The ClonTracer library consists of semi-random 30-bp-long DNA barcodes with 15 repeats of A or T (“W” for weak)–G or C (“S” for strong). Cell lines of interest were transduced with the ClonTracer library at low multiplicity of infection (m.o.i.) to label each cell with one unique barcode per cell (I). Each color represents a unique barcode. The barcoded cell population was treated with targeted therapies until a resistant clonal population emerged (II), and the selected population was harvested as a clonal pool (III). Barcode sequences were PCR-amplified from genomic DNA (gDNA) (IV,V) and subjected to NGS. First, the number of unique barcode sequences provided a quantitative readout of the number of clones the population originated from, and second, the counts per unique barcode offered a measure of the relative abundance of each clone (VI). **(b)** The library complexity was estimated by NGS of the library plasmid pool at a depth of approximately 161 million reads. The barcode distribution in the library is plotted in pink. More than 27 million unique barcodes are present in the ClonTracer library, and computational modeling projected approximately 73 million unique barcodes (**Supplementary Fig. 1f**). The barcode library did not show any significant bias, as 95% of the 27 million observed barcodes were detected with 2–12 reads. The cumulative wealth distribution of the unique barcodes is shown in blue.

DNA barcode design (**Fig. 1a** and **Supplementary Tables 1–3**; also see Online Methods). Each barcode was designed to have a balanced GC content (50%) to ensure uniform PCR-amplification efficiency (**Supplementary Fig. 1a–e**). Deep sequencing confirmed that the library was evenly distributed, with 95% of the barcodes detected with 2–12 reads (**Fig. 1b**). Computational modeling to fit the sequencing depth to the number of barcodes observed with NGS projected that the library contained approximately 73 million unique barcodes (**Supplementary Fig. 1f**). The high barcode complexity of this library enables the labeling and tracking of millions of individual cells in complex cancer-cell populations. Comprehensive validation demonstrated the highly quantitative nature of the ClonTracer system (**Supplementary Discussion** and **Supplementary Fig. 1g–w**).

Clonal dynamics in a non-small cell lung cancer model treated with erlotinib

We sought to use the ClonTracer system to monitor clonal dynamics in response to targeted therapies. The non-small cell lung cancer cell line HCC827 harbors an activating epidermal growth factor receptor (EGFR) mutation (exon 19 deletion) that confers sensitivity to the

EGFR inhibitor erlotinib^{15–17}, and it has been widely used to study mechanisms of resistance to this drug^{8,18–21}. We transduced HCC827 cells with the ClonTracer library with the aim of labeling approximately 1 million cells to deeply sample the potential heterogeneity of the cancer cell population. Using NGS, we confirmed that the barcode complexity was 0.92 million. We reasoned that having multiple replicates with comparable starting barcode representations would provide a means to distinguish pre-existing from *de novo* acquired-resistance clones. If resistance were mostly driven by *de novo* alterations, distinct barcoded populations would emerge in independent replicates. By contrast, if pre-existing clones were the major source of resistance, one could expect the selective enrichment of the same sets of barcodes (labeling the same pre-existing subpopulations) in replicate experiments. To explore this, we expanded and plated the barcoded HCC827 cell pool in multiple replicates with 20-fold library representation (i.e., 18 million cells each) to minimize stochastic loss of barcodes during plating and to ensure comparable starting barcode representations in each replicate. Similar to the parental HCC827 cells, the barcoded HCC827 population was highly sensitive to erlotinib (**Supplementary Fig. 1u**), and upon prolonged treatment, actively proliferating resistant

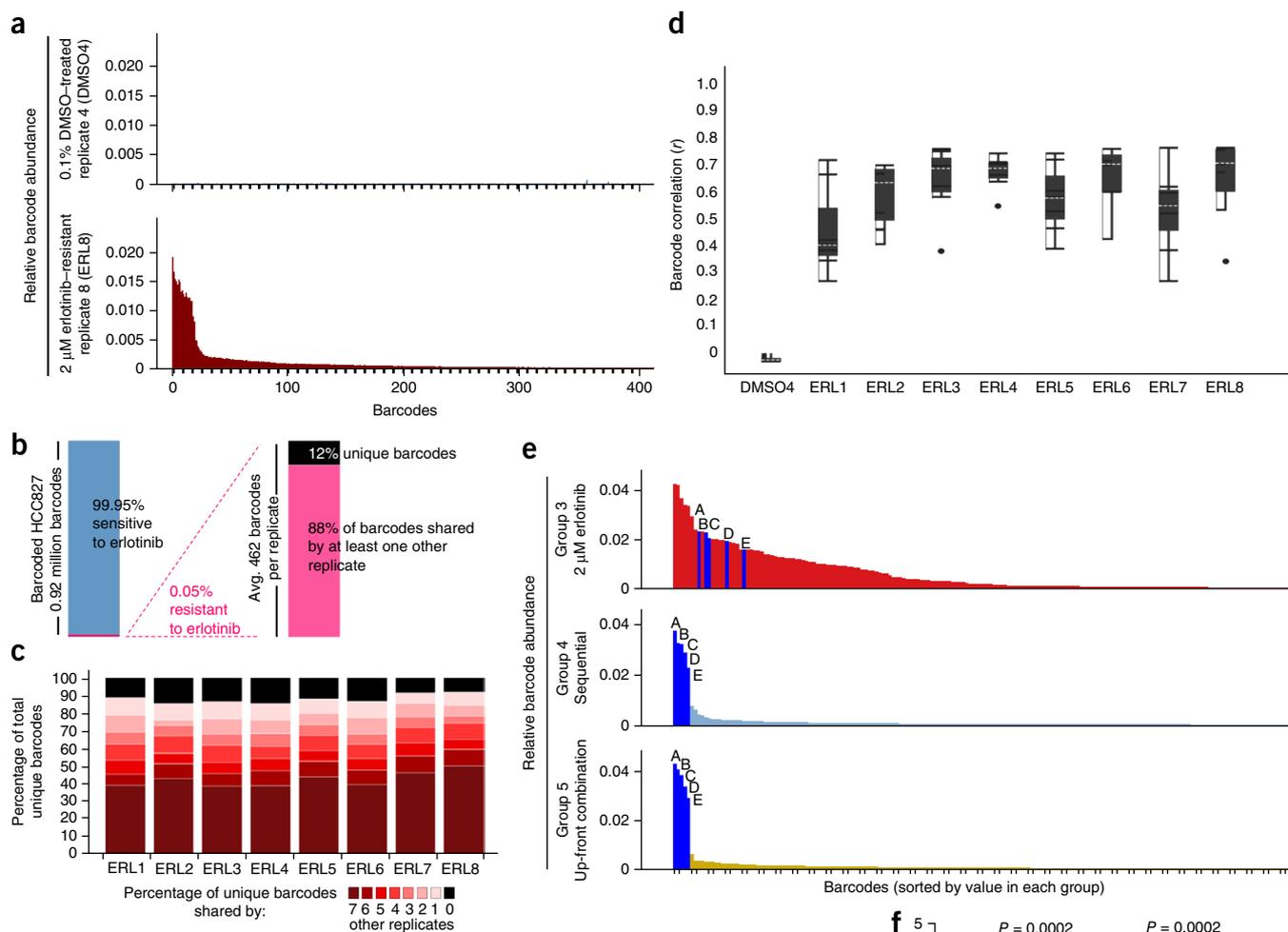


Figure 2 ClonTracer barcoding technology demonstrates the pre-existence of erlotinib-resistant subpopulations in the HCC827 cell line. **(a)** Barcode distribution (fraction of total barcode reads) of one representative replicate each from groups treated with vehicle (0.1% DMSO) and 2 μM erlotinib. The x -axes of the histograms are identical; each bar represents one unique barcode. **(b)** An average of 462 unique barcodes (0.05%) were enriched in each replicate after erlotinib treatment. Of these, on average 88% of the enriched barcodes from each replicate were shared by at least one other replicate. **(c)** Percentage of enriched barcodes in each replicate that were also found in other replicates (ERL1–8). Different colors denote the number of other replicates in which the barcodes were identified. **(d)** Pearson correlation confirmed that the overlap of enriched barcodes among erlotinib-treatment replicates was statistically significant (median Pearson's correlation coefficient $r = 0.56$). P values are presented in **Supplementary Table 5**. White dashed lines denote median values. Boxes show first and third quartiles, and whiskers show data within 1.5 \times the interquartile range (IQR) from the first and third quartiles. Dots denote data more than 1.5 \times the IQR from the first or third quartile. **(e)** The combination of erlotinib and crizotinib significantly reduced the barcode complexity of the resistant population. Barcoded HCC827 cells were treated with 2 μM erlotinib for 36 d (group 3), 2 μM erlotinib for 36 d followed by 0.2 μM crizotinib for 7 d (group 4) or a combination of 2 μM erlotinib and 0.2 μM crizotinib for 36 d (group 5). Histograms show the average relative abundance sorted by average fraction values within each group. The five most enriched clones in groups 4 and 5 are labeled as A–E (blue). Individual replicates of groups 4 and 5 are shown in **Supplementary Figure 5**. **(f)** *MET* copy number was assessed by quantitative PCR for groups 3–5 and control groups treated with vehicle (group 1) or 0.2 μM crizotinib (group 2) and normalized to group 1. Bars show mean \pm s.d. P value, unpaired two-tailed t -test.

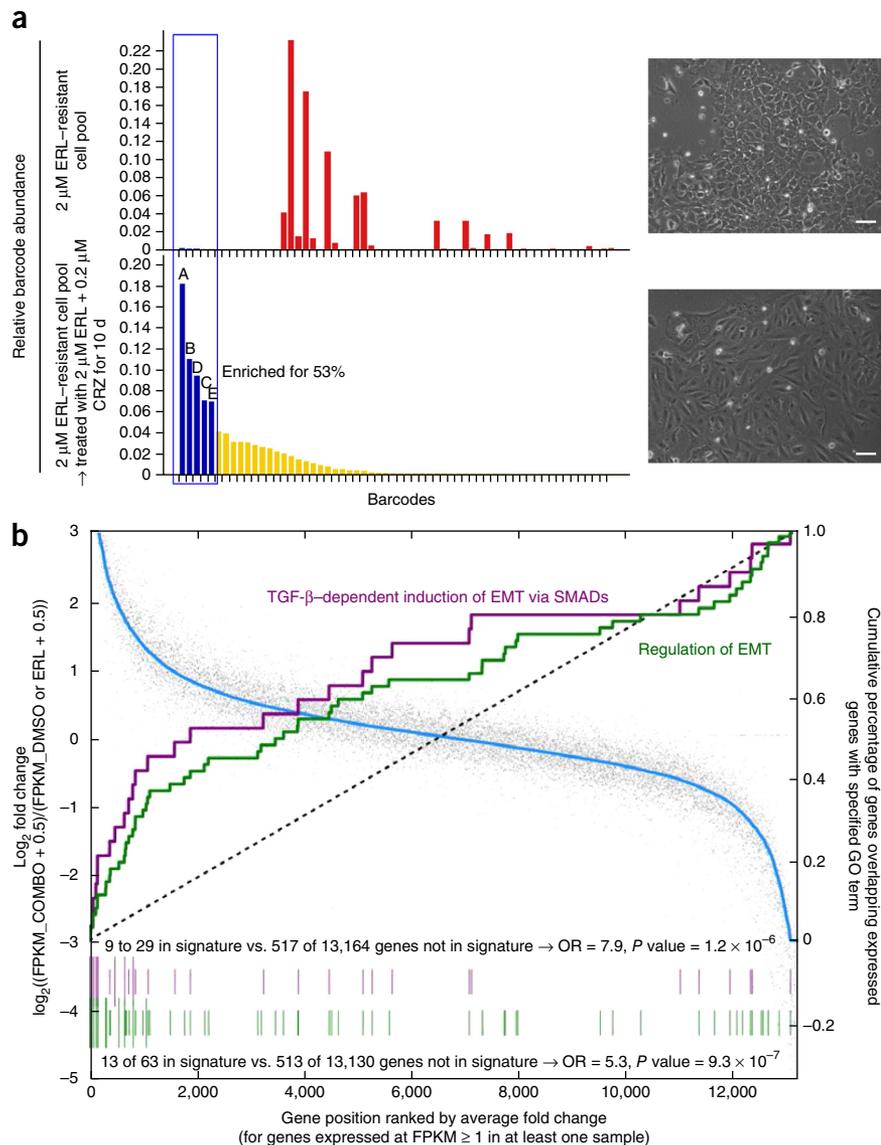
clones emerged (**Supplementary Fig. 2**). Comparison of the barcode distributions in cell populations treated with either vehicle (0.1% DMSO) or 2 μM erlotinib (a concentration that represents a clinically relevant drug exposure²²) revealed marked enrichment of a subset of barcodes in the erlotinib-treatment group (**Fig. 2a**). Using a threshold based on the maximum single-barcode frequency in DMSO-treated control cells (**Fig. 2a** and **Supplementary Fig. 3**), we found that an average of 462 barcodes were enriched in erlotinib-treated replicates, with a remarkable consistency of 388–503 barcodes enriched across

the eight different replicates (**Supplementary Table 4**). Given the projected barcode complexity of 0.92 million in the vehicle-treated population, the enrichment of 462 barcodes indicates that approximately 0.05% of the starting clonal population contributed to erlotinib resistance (**Fig. 2b**).

When we compared the barcodes enriched across the eight erlotinib-treated replicates, we found that approximately 90% of the barcodes in each replicate were shared by at least one other replicate (**Fig. 2b,c**). Moreover, approximately 40% of the enriched barcodes

Figure 3 The erlotinib–crizotinib dual-resistant subpopulations in HCC827 cells are pre-existing and predetermined and display features of epithelial–mesenchymal transition (EMT).

(a) The same five barcodes (A–E) from **Figure 2e** became enriched in an independent repeat experiment using combination treatment with erlotinib (ERL) and crizotinib (CRZ). Erlotinib-resistant HCC827 cell pools were treated with either erlotinib (top) or the erlotinib–crizotinib combination (bottom) to enrich for c-Met–independent subpopulations. Barcodes for each histogram were plotted in order of abundance in the combination-treated population. NGS revealed that barcodes A–E constituted 53% of the entire cell population after erlotinib–crizotinib treatment. Images (right) display the cellular morphology of erlotinib-resistant and erlotinib–crizotinib combination-treated cell populations. Scale bars, 1 μm . (b) Gene expression profiling by RNA-seq revealed the upregulation of EMT pathways in the subpopulation of HCC827 cells resistant to erlotinib–crizotinib. Two EMT-related signatures were significantly overexpressed in the erlotinib–crizotinib dual-resistant HCC827 subpopulation compared to the cells resistant to erlotinib only. Hash marks indicate the presence of a gene in the signature; taller hash marks denote genes that were more significantly (fourfold) upregulated in combination-treated versus single-agent-treated cells. Gray dots indicate the fold change versus single-agent or DMSO treatment; the blue solid line shows the average fold change. All multiexon genes expressed at ≥ 1 fragment per kilobase of exon per million fragments mapped (FPKM) in any sample are shown. The dashed line indicates the uniform distribution of genes in a pathway, and solid lines indicate the actual distribution (purple, TGF β -dependent induction of EMT (Gene Ontology (GO) term 2_2996); green, regulation of EMT (GO term 2_3018)). *P* values shown are based on Fisher's exact test.



were shared across all eight erlotinib treatment replicates (**Fig. 2c**). Given that only 0.05% of HCC827 cells contributed to the resistant population, the observed barcode overlap across the independent replicates was highly statistically significant on the basis of a Pearson correlation test (mean *r* value = 0.56, *P* < 1×10^{-10}). By contrast, low correlation was observed between erlotinib- and DMSO-treated samples (mean *r* value = -0.029, *P* = 0.36) (**Fig. 2d** and **Supplementary Table 5**). A small fraction of the resistant clones (about 10% of the total erlotinib-resistant population) in each replicate did not share any barcodes with other replicates (**Fig. 2b,c**). Although we cannot completely exclude the possibility of stochastic loss of some barcodes during plating, it is most likely that these clones acquired *de novo* alterations after barcoding and thus were not shared among other replicates. Collectively, the ClonTracer results strongly indicate that the vast majority of erlotinib-resistant clones in the HCC827 population were pre-existing and selected during erlotinib treatment.

Pre-existing clones exhibit *MET* amplification and epithelial–mesenchymal transition

Prior studies suggest that *MET* amplification confers resistance to erlotinib in cell lines and patients^{8,23,24}, raising the possibility that the

barcoded cells enriched in response to erlotinib might comprise a pre-existing *MET*-amplified subpopulation. To test this hypothesis, we examined the effects of the c-Met inhibitor crizotinib on the barcoded HCC827 population. Notably, the addition of crizotinib to erlotinib strongly reduced the barcode complexity in both sequential treatments (group 4) and treatment with a combination of erlotinib and crizotinib (group 5), decreasing the size of the resistant population to only five major clones (**Fig. 2e**). These findings indicate that the majority of pre-existing resistant clones in HCC827 cells are c-Met dependent and can be eradicated by crizotinib treatment. Consistent with this notion, the erlotinib-resistant population that was not exposed to crizotinib (group 3; **Fig. 2e**) displayed marked *MET* amplification as measured by quantitative PCR (**Fig. 2f**), but *MET* amplification was not detected in the crizotinib cotreatment groups (groups 4 and 5; **Fig. 2f**). Mapping of the lentiviral integration sites in several single-cell clones did not reveal any evidence of insertional mutagenesis (**Supplementary Table 6**), and copy-number analyses confirmed that these erlotinib-resistant pre-existing clones displayed *MET* amplification (**Supplementary Fig. 4**). Although we cannot completely exclude the possibility that the pre-existing clones were primed to acquire *MET* amplification, the identification by fluorescence *in situ*

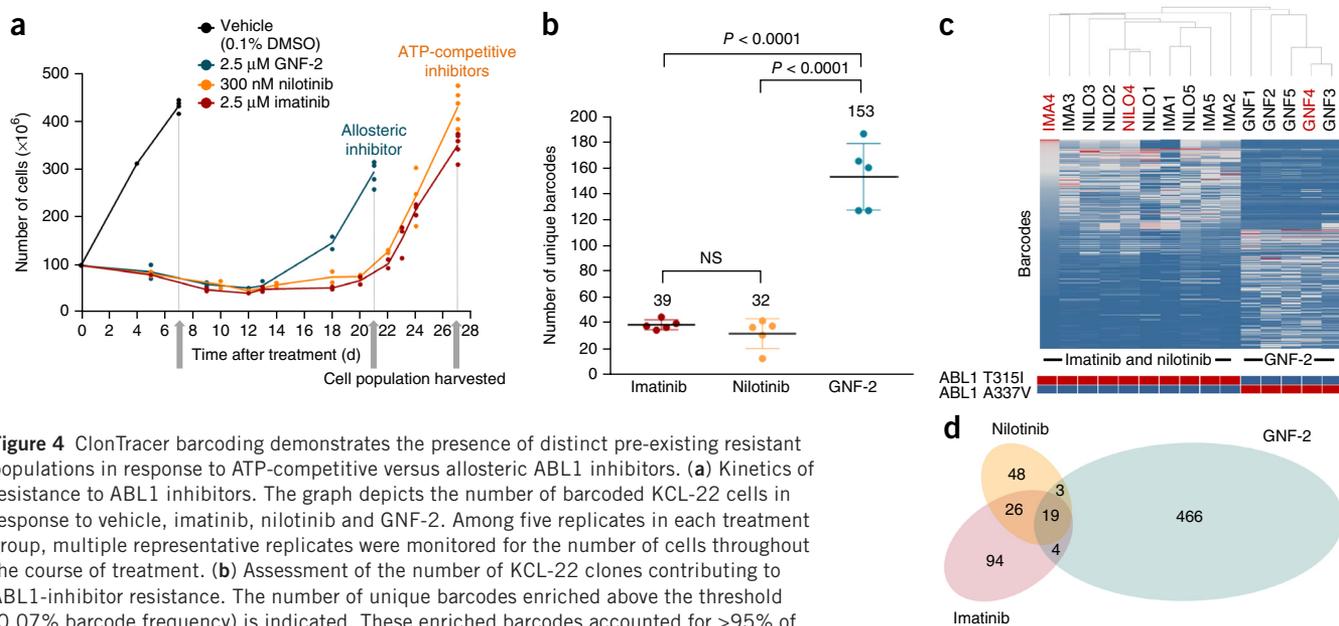


Figure 4 ClonTracer barcoding demonstrates the presence of distinct pre-existing resistant populations in response to ATP-competitive versus allosteric ABL1 inhibitors. **(a)** Kinetics of resistance to ABL1 inhibitors. The graph depicts the number of barcoded KCL-22 cells in response to vehicle, imatinib, nilotinib and GNF-2. Among five replicates in each treatment group, multiple representative replicates were monitored for the number of cells throughout the course of treatment. **(b)** Assessment of the number of KCL-22 clones contributing to ABL1-inhibitor resistance. The number of unique barcodes enriched above the threshold (0.07% barcode frequency) is indicated. These enriched barcodes accounted for >95% of the total barcode population in each of the drug-treated replicates (**Supplementary Fig. 9a**). Results are presented as mean \pm s.d. P value, unpaired two-tailed t -test; NS, not significant. **(c)** The barcode enrichment pattern was correlated within each treatment group but distinct between catalytic and allosteric ABL1-inhibitor groups. Shown is unsupervised hierarchical clustering of the 100 most enriched barcodes from replicates 1–5 of the imatinib (IMA), nilotinib (NILO) and GNF-2 (GNF) groups. We used the weighted pair group method with averaging, with Euclidean distance measures. Sequenom assays detected ABL1 T315I mutation in every replicate of the imatinib and nilotinib groups, whereas A337V mutation was observed in every replicate of the GNF-2 group (bottom). Red, mutation detected; blue, mutation not detected. Estimated allelic frequencies are presented in **Supplementary Table 10b**. One replicate from each group (marked in red) was subjected to RNA-seq (**Supplementary Table 10a**). **(d)** Venn diagram displaying the number of overlapping barcodes found in each ABL1 inhibitor-resistant population. Barcodes enriched above the threshold in at least one replicate were included.

hybridization of rare HCC827 cells harboring *MET* amplification⁸ is consistent with the notion of pre-existing *MET*-amplified clones driving resistance in this cell line. Despite the considerable reduction in barcode complexity, the erlotinib–crizotinib combination treatment failed to eradicate all of the resistant clones. The fact that the same five clones (**Fig. 2e**) were found to be enriched in all nine independent replicates of both sequential-treatment and combination-treatment groups (**Supplementary Fig. 5**) strongly suggests that they existed before drug treatment. Moreover, the ClonTracer approach provided a quantitative measurement of the frequency of these rare pre-existing dual-resistant clones, which constituted approximately 0.0005% of the parental population.

We next wanted to investigate the molecular mechanisms that confer dual resistance to erlotinib and crizotinib in HCC827 cells. We noted that the combination treatment enriched for cells with mesenchymal rather than epithelial morphology (**Fig. 3a** and **Supplementary Fig. 6**). Consistent with the morphology of these cells, gene expression profiling by RNA-seq revealed upregulation of epithelial–mesenchymal transition (EMT) pathways (**Fig. 3b** and **Supplementary Tables 7** and **8**). EMT has been implicated in resistance to several targeted agents, including erlotinib^{25–27}, which suggests that EMT might have mediated, at least in part, the dual erlotinib–crizotinib resistance in the five *c-Met*–independent barcoded clones. RNA-seq of the resistant population did not reveal obvious genetic variants that could be linked to EMT. Although we cannot exclude the presence of genetic alterations outside of the coding regions, such as promoter mutations, that could drive an EMT phenotype, it is also possible that this cellular change is driven by nongenetic mechanisms, such as epigenetic alterations. In either case, the barcoding data suggest that this dual-resistant population was pre-existing or predetermined, rather than *de novo* acquired.

Rare pre-existing subclones drive resistance to ABL1 inhibitors

We next used the ClonTracer system to monitor the therapeutic response of a chronic myeloid leukemia (CML) model to ABL1 inhibitors. KCL-22 is a CML cell line derived from a patient in blast crisis and harbors the *BCR-ABL1* translocation²⁸. The catalytic (ATP-competitive) ABL1 inhibitors imatinib and nilotinib have dramatically improved outcomes for CML patients, but the development of resistance remains a clinical challenge^{29,30}. Additionally, allosteric ABL1 inhibitors, such as GNF-2, that target the myristoyl binding pocket rather than the catalytic site of ABL1, have recently been described^{31,32}. To compare the clonal responses to catalytic and allosteric inhibitors, we transduced KCL-22 cells with the ClonTracer library, expanded them and passaged them into multiple replicates for each treatment group: vehicle (0.1% DMSO), imatinib, nilotinib and GNF-2 (**Supplementary Figs. 1v,w** and **7**). Resistance to GNF-2 consistently emerged with slightly faster kinetics than resistance to imatinib and nilotinib (**Fig. 4a**). An average of 153 clones were found in the GNF-2-resistant cell population in each replicate, whereas averages of only 39 and 32 barcodes were detected in the imatinib- and nilotinib-resistant populations, respectively (**Fig. 4b** and **Supplementary Figs. 8** and **9a**). Given that the starting KCL-22 population had a projected barcode complexity of 2.67 million (Online Methods), these findings indicate that small subpopulations of KCL-22 (i.e., 0.001% in the case of nilotinib or imatinib and 0.006% for GNF-2) drive resistance to these targeted agents. Moreover, the relative barcode frequency indicates that the faster kinetics of resistance to GNF-2 compared to resistance to nilotinib or imatinib can probably be explained by a greater number of clones contributing to GNF-2 resistance in this cell line.

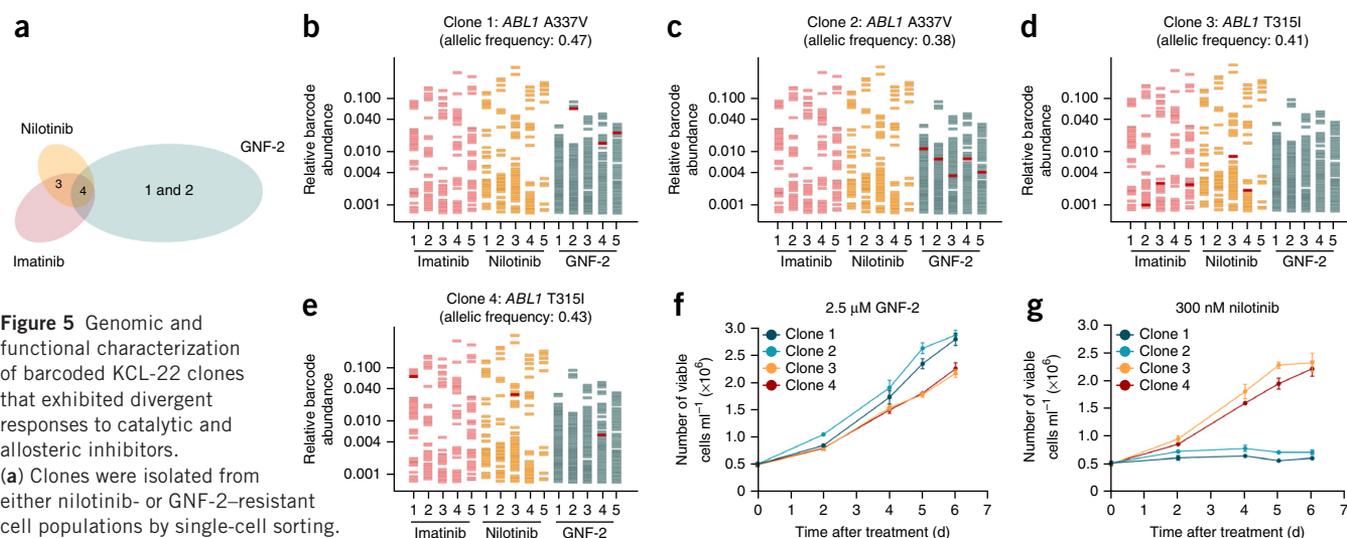


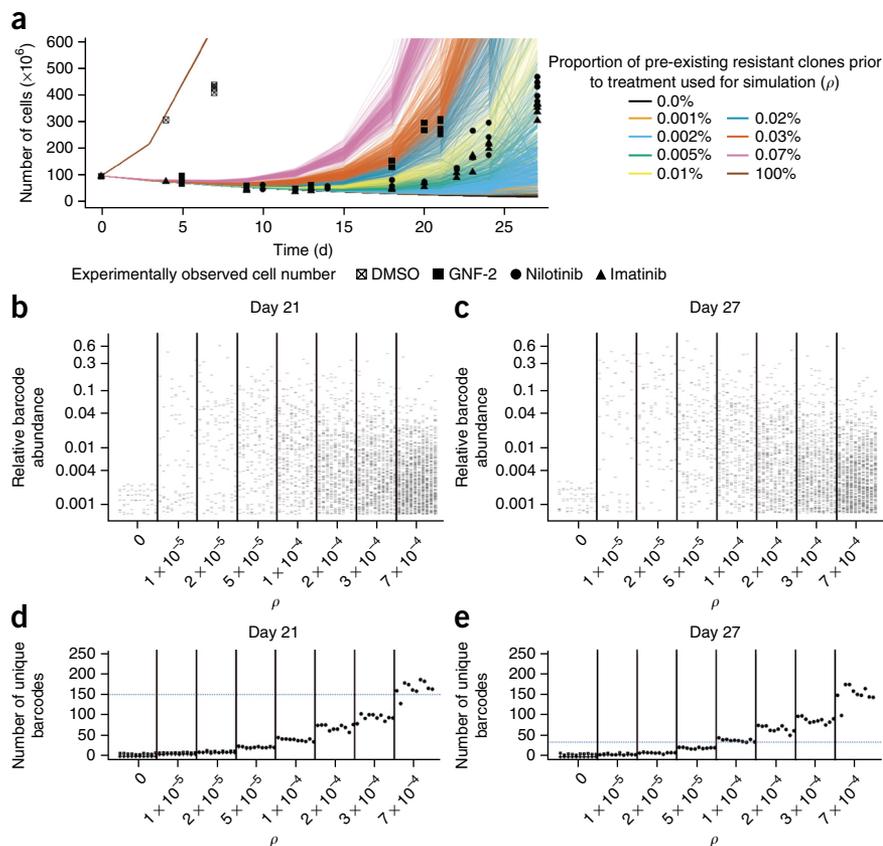
Figure 5 Genomic and functional characterization of barcoded KCL-22 clones that exhibited divergent responses to catalytic and allosteric inhibitors. (a) Clones were isolated from either nilotinib- or GNF-2-resistant cell populations by single-cell sorting. Barcode sequencing of the individual clones enabled the selection of four clones that showed differential resistance under ABL1-inhibitor treatments according to the experiment displayed in **Figure 4**. The Venn diagram illustrates in which resistance groups each clone or barcode was observed to be enriched: resistant to GNF-2 only (clones 1 and 2), resistant to nilotinib and imatinib but not to GNF-2 (clone 3) or resistant to all three drugs (clone 4). (b–e) The barcode sequence and *ABL1* mutation status of clones 1–4 are indicated at the top of each panel. The barcode plots illustrate the relative abundance (fraction of total barcode reads) of the indicated clone (marked in red) in the independent replicates of the three treatment groups (five replicates per group, presented along the x-axis), which correspond to the replicates shown in **Figure 4c**. Each horizontal bar represents a unique barcode, and the position on the y-axis indicates its relative abundance as measured by the fraction of reads (the most enriched clones are at the top). Red bars indicate the positions of specific clones (clones 1–4 in b–e, respectively) relative to the positions of other clones that were present above the threshold (i.e., fraction of 0.07%) in a particular treatment replicate. (f,g) The growth kinetics of clones 1–4 in the presence of 2.5 μM GNF-2 (f) or 300 nM nilotinib (g). Error bars show mean \pm s.d.

To examine the pattern of resistance across different replicates and various ABL1 inhibitors, we performed unsupervised hierarchical clustering on the 100 most enriched barcodes from each replicate across all three drug-treatment groups (**Fig. 4c**). Strikingly, the independent replicates of nilotinib and imatinib treatment formed a cluster and showed highly similar patterns of barcode enrichment. By contrast, although the GNF-2 replicates clustered with one another, they displayed a very divergent barcode pattern compared to that of the catalytic-inhibitor replicates. The fact that barcodes enriched in the replicates of each treatment group clustered with one another strongly suggests that a significant fraction of the emergent resistant clones were pre-existing. Moreover, the clustering of nilotinib- and imatinib-treated replicates indicates that similar pre-existing clones drove resistance to these two catalytic inhibitors. By contrast, the GNF-2-resistant clones appeared to be largely non-overlapping with those enriched by the catalytic inhibitors (**Fig. 4c,d**, **Supplementary Fig. 9b–e** and **Supplementary Table 9**), suggesting that resistance may be driven by different mechanisms. Consistent with this notion, genomic analyses revealed that all of the imatinib- and nilotinib-treatment replicates harbored T315I mutations in *ABL1* on the population level, whereas all five GNF-2 replicates displayed an A337V variant in *ABL1* (**Fig. 4c** and **Supplementary Table 10**). The T315I mutation in the ABL1 catalytic site has been reported as a major resistance mechanism in clinical studies with imatinib and nilotinib^{29,33,34}, indicating that the KCL-22 line recapitulates resistance mechanisms of clinical relevance. No clinical resistance data are available for allosteric ABL1 inhibitors; however, preclinical studies in Ba/F3 engineered systems have shown that alterations in the myristoyl pocket, such as at the Ala-337 and Ala-344 residues, can render ABL1 insensitive to GNF-2 (ref. 31). The ClonTracer results indicate that a pre-existing subpopulation of KCL-22 cells harboring A337V mutations accounted for a considerable fraction of the GNF-2-resistant clones. However, although the

majority of clones resistant to catalytic and allosteric ABL1 inhibitors were non-overlapping, at the concentrations used in this experiment (six- to eightfold above the half-maximal inhibitory concentration (IC₅₀) for each drug), a small subpopulation of cells exhibited dual resistance to both catalytic and allosteric inhibitors (**Fig. 4d**).

In order to further explore the molecular mechanisms of ABL1-inhibitor resistance, we isolated individual clones that were resistant to GNF-2 (clones 1 and 2), to imatinib and nilotinib (clone 3), or to all three compounds (clone 4) (**Fig. 5a**). The barcode-sequence information of each clone enabled us to track cellular behavior within the pooled cell populations in the experiment described in **Figure 4** under treatment with the three ABL1 inhibitors (**Fig. 5b–e**). RNA-seq revealed that clone 3 harbored the T315I gatekeeper mutation in *ABL1*, consistent with its resistance to imatinib and nilotinib (**Fig. 5a,d**). Clones 1 and 2 both carried an A337V mutation and were strongly resistant to GNF-2 treatment (**Fig. 5f** and **Supplementary Fig. 10a**) but were sensitive to the catalytic ABL1 inhibitors (**Fig. 5g** and **Supplementary Fig. 10b**), which explained their selective enrichment of these barcodes under treatment with GNF-2 but not with imatinib or nilotinib (**Fig. 5b,c**). Interestingly, we found that clone 4, which was resistant to both catalytic and allosteric ABL1 inhibitors, also harbored a T315I mutation but had no other apparent alterations in *ABL1* (**Fig. 5a,e**). KCL-22 clones with T315I mutations (clones 3 and 4) showed an approximately 50-fold increase in IC₅₀ for GNF-2 compared to the parental KCL-22 cell population (**Supplementary Figs. 7 and 10a**). Although both A337V and T315I mutations were able to mediate resistance to GNF-2, we noted that there was a slight but consistent increase in the relative growth rate of A337V mutant (clones 1 and 2) compared to T315I mutant clones (clones 3 and 4) when challenged with 2.5 μM GNF-2, the concentration used for the barcoding experiments (**Fig. 5f** and **Supplementary Fig. 10a**). Mathematical modeling (**Supplementary Discussion**) revealed that

Figure 6 Mathematical modeling of KCL-22 clonal dynamics under treatment with ABL1 inhibitors. **(a)** Modeled effects of varying proportions of pre-existing resistant barcodes or clones on KCL-22 population dynamics under treatment with DMSO or ABL1 inhibitors. ρ , fraction of resistant clones before treatment (i.e., pre-existing). Symbols represent experimentally measured numbers of viable cells (displayed in **Fig. 4a**). Colored lines denote the predicted population dynamics for different ρ values. For each ρ value, various combinations of mutation rates ($u = 10^{-7}$, 10^{-8} and 10^{-9} per cell division), birth rates ($b = 0.00-0.30$) and death rates ($d = 0.06-0.36$) of drug-sensitive cells were tested (Online Methods). Independent model predictions for days 0, 3, 7, 9, 12, 15, 18, 21, 24 and 27 are connected by lines. **(b,c)** Simulation of the effects of different ρ values on barcode-abundance distribution at day 21 with GNF-2 **(b)** and day 27 with nilotinib or imatinib **(c)**. Barcodes with fractions greater than 0.07% are shown. Each column represents a simulation run; ten simulations were performed for each ρ value. **(d,e)** Model-predicted number of barcodes exceeding 0.07% for the indicated ρ values at day 21 **(d)** and day 27 **(e)**. The number of total barcodes (■) and the number of barcodes corresponding to resistant clones (+) are depicted. Blue dashed lines indicate the average number of unique barcodes observed experimentally in groups treated with GNF-2 **(d)** and imatinib or nilotinib **(e)** described in **Figure 4b**. $b = 0$, $d = 0.06$ and $u = 1 \times 10^{-9}$ used in **b-e**; additional simulations are presented in **Supplementary Figure 13** and **Supplementary Data Sets 1** and **2**.



this growth difference accounted for the observed shift in clonal distribution in favor of A337V versus T315I mutant clones during GNF-2 treatment and probably could explain why some of the T315I mutant clones, such as clone 3, were not found in any of the GNF-2-treated replicates (**Fig. 5d**). Collectively, these data suggest that interclonal competition due to subtle differences in relative fitness can significantly skew the clonal composition of cancer cell populations.

Modeling clonal dynamics in response to ABL1 inhibitors

The ClonTracer data provide a quantitative measure of subclonal fitness in response to different cancer therapies. In an attempt to model the population dynamics observed in KCL-22 cells in response to different drug treatments, we designed a stochastic mathematical model informed by the experimental data (Online Methods). The model is based on a multitype birth-death process in which each uniquely barcoded clone is considered to consist of two cell types: drug-sensitive cells and drug-resistant cells (**Supplementary Fig. 11**). The model was implemented as an exact stochastic computer simulation in which clonal growth and death kinetics under various growth conditions (treatment with DMSO, GNF-2, imatinib or nilotinib) were derived from the experimental data (**Figs. 4a** and **5f,g**, **Supplementary Fig. 12**, **Supplementary Tables 11** and **12** and Online Methods). To model the effect of varying levels of pre-existing resistant clones, we implemented simulations using different proportions of pre-existing resistant clones, where ρ indicated the fraction of resistant clones in the starting population before treatment. This mathematical model accurately recapitulated the pattern of population rebound in response to GNF-2 treatment when ρ was set at 0.02%–0.03%. Similarly, ρ values of

0.005%–0.01% yielded the patterns observed for imatinib- and nilotinib-treated cells (**Fig. 6a** and Online Methods). Importantly, the introduction of newly emerging resistant clones to this mathematical model, even at a mutation rate as high as 10^{-7} per cell division³⁵, did not lead to significantly different results, indicating that *de novo* mutations in the absence of pre-existing resistant cells ($\rho = 0$) could not explain the observed growth patterns (**Fig. 6a**). In addition, testing of different values of ρ revealed that the simulated pattern and relative barcode enrichment closely resembled the observed patterns at $\rho > 0.005\%$ for imatinib- or nilotinib-resistant populations and at $\rho > 0.03\%$ for GNF-2-resistant populations, but not at $\rho = 0$ (**Fig. 6b–e** and **Supplementary Fig. 13**). Thus, the mathematical modeling strongly supported the notion of pre-existing resistant clones.

DISCUSSION

The ClonTracer barcode system described here enables the tracking of more than 1 million individual cells in cancer models, and thus provides a means to monitor clonal dynamics in response to targeted therapies at high resolution. This resolution affords the ability to detect very rare pre-existing cancer-cell subclones at a resolution of 1 in 1 million. Moreover, in contrast to targeted detection methods, such as digital PCR or RainDance PCR^{36–38}, the ClonTracer technology can be used to detect pre-existing resistant clones without prior knowledge of the underlying molecular resistance mechanism. Using this approach in two clinically relevant resistance models, we obtained direct evidence for the presence of pre-existing subpopulations that become selected in response to therapeutic challenge. The demonstration of rare pre-existing resistant clones in this study might explain the

clinically observed, and often rapid, emergence of resistance to targeted therapies^{1–3}. Notably, the pre-existing resistant cells in the cancer models studied here constituted very small subpopulations—for example, approximately 0.05% in the case of erlotinib-resistance in HCC827 cells and 0.001% for nilotinib or imatinib resistance in KCL-22 cells. Given that the sequencing error rate¹¹ limits the resolution of standard NGS-based methods to a sensitivity of approximately 0.1% allele frequency, it is likely that such rare pre-existing subclones would be missed by standard NGS approaches. Indeed, sequencing of parental KCL-22 cells failed to identify pre-existing T315I mutant clones^{39–42}. These findings have important clinical implications, as they constitute a strong argument for the development and application of more sensitive sequencing technologies, such as barcode-based NGS strategies^{43–46}, or allele-specific detection methods, such as BEAMing⁴⁷, for the detection of rare pre-existing clones that might plant the seeds for clinical relapse.

We show here that the ClonTracer system can be used to quantify the frequency of pre-existing clones resistant to various cancer therapies and thus is a valuable tool for the preclinical assessment of novel cancer drugs. In addition, the ClonTracer approach affords the ability to prioritize combination therapies that target non-overlapping resistant subpopulations, characterized by non-overlapping barcode patterns, in an attempt to prevent relapse due to rare pre-existing resistant clones.

Given their largely distinct resistance patterns, the use of a combination of catalytic and allosteric ABL1 inhibitors appears to be a promising way to suppress relapse by targeting pre-existing T315I and A337V mutant clones, respectively. The relatively low potency of GNF-2, however, will make it challenging to clinically achieve GNF-2 concentrations that are sufficient to suppress pre-existing T315I clones (Supplementary Fig. 14). Thus the development and clinical testing of more potent allosteric ABL1 inhibitors, such as ABL001 (ClinicalTrials.gov, NCT02081378), holds great promise for preventing the emergence of resistance in CML. The findings of our study also have important implications for the optimal scheduling of combination therapies. Mathematical modeling indicated that, especially in genomically unstable cancers with mutation rates greater than 10^{-8} per cell division, new ‘pre-existing’ dual-resistant clones could readily develop during the expansion of clones resistant to first-line therapies (Supplementary Fig. 15) (refs. 48–50). Thus, treatments that have non-overlapping resistance mechanisms should be given concomitantly or in close sequence to maximally suppress the emergence of resistance.

In future studies, it will be interesting to apply the ClonTracer system to *in vivo* models, which might provide additional insights regarding the contribution of cancer cell autonomous mechanisms versus tumor microenvironmental factors to relapse. Even though it is not possible to apply the ClonTracer system in patients, the broad tropism of lentiviral transduction should allow for the barcoding of patient-derived xenograft models. More generally, the ability to track single cells in longitudinal studies with the ClonTracer system should facilitate the investigation of cancer stem cells and clonal evolution processes such as metastasis.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).

ACKNOWLEDGMENTS

We thank A. Wylie, J. Donovan, S. Zhu, E.R. McDonald, Z. Jagani and B. Firestone for helpful discussion. We also thank J. Brugge, K. Polyak, J. Williams and J. Kuiken for critical reading of the manuscript and K. Yu for technical assistance. R.Z. and F.M. gratefully acknowledge support from the Dana-Farber Cancer Institute Physical Sciences-Oncology Center (U54CA143798).

AUTHOR CONTRIBUTIONS

H.C.B. and F.S. conceived the project and designed the experiments. H.C.B., D.A.R., J.M.K., M.R.S. and F.S. designed and developed the ClonTracer cellular barcoding system, and H.C.B. constructed the ClonTracer barcode library. H.C.B. and J.X.C. performed the biological experiments and analyzed the data. H.C.B., V.K.R., J.M.K., A.P.S. and D.Y.C. analyzed NGS and RNA-seq data. D.A.R., I.K., D.R., M.M.H., A.R., E.A. and R.J.L. directed or performed NGS and RNA-seq. V.K.R. and J.M.K. performed bioinformatic analysis. R.Z. and F.M. performed mathematical modeling of clonal dynamics. P.S., M.B. and M.P. performed or directed *MET* copy-number analysis and ABL1 Sequenom mutation analysis. H.C.B., V.K.R., R.Z. and J.M.K. prepared figures and tables. H.C.B., R.Z., W.R.S., F.M., V.G.C., J.M.K. and F.S. wrote and edited the manuscript. W.R.S. and N.K. contributed to oversight of and advice on the overall project. F.S. and J.M.K. provided overall project leadership.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests; details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Ohashi, K., Maruvka, Y.E., Michor, F. & Pao, W. Epidermal growth factor receptor tyrosine kinase inhibitor-resistant disease. *J. Clin. Oncol.* **31**, 1070–1080 (2013).
- Holohan, C., Van Schaeybroeck, S., Longley, D.B. & Johnston, P.G. Cancer drug resistance: an evolving paradigm. *Nat. Rev. Cancer* **13**, 714–726 (2013).
- Weisberg, E., Manley, P.W., Cowan-Jacob, S.W., Hochhaus, A. & Griffin, J.D. Second generation inhibitors of BCR-ABL for the treatment of imatinib-resistant chronic myeloid leukaemia. *Nat. Rev. Cancer* **7**, 345–356 (2007).
- Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
- Morrison, C.D. *et al.* Whole-genome sequencing identifies genomic heterogeneity at a nucleotide and chromosomal level in bladder cancer. *Proc. Natl. Acad. Sci. USA* **111**, E672–E681 (2014).
- Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
- Diaz, L.A. Jr. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
- Turke, A.B. *et al.* Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell* **17**, 77–88 (2010).
- Landau, D.A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
- Misale, S. *et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532–536 (2012).
- Robasky, K., Lewis, N.E. & Church, G.M. The role of replicates for error mitigation in next-generation sequencing. *Nat. Rev. Genet.* **15**, 56–62 (2014).
- Lu, R., Neff, N.F., Quake, S.R. & Weissman, I.L. Tracking single hematopoietic stem cells *in vivo* using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* **29**, 928–933 (2011).
- Naik, S.H. *et al.* Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496**, 229–232 (2013).
- Wu, C. *et al.* Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. *Cell Stem Cell* **14**, 486–499 (2014).
- Lynch, T.J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
- Paez, J.G. *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
- Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. USA* **101**, 13306–13311 (2004).
- Amann, J. *et al.* Aberrant epidermal growth factor receptor signaling and enhanced sensitivity to EGFR inhibitors in lung cancer. *Cancer Res.* **65**, 226–235 (2005).
- Sharma, S.V. *et al.* A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80 (2010).
- Ohashi, K. *et al.* Lung cancers with acquired resistance to EGFR inhibitors occasionally harbor BRAF gene mutations but lack mutations in KRAS, NRAS, or MEK1. *Proc. Natl. Acad. Sci. USA* **109**, E2127–E2133 (2012).
- Shien, K. *et al.* Acquired resistance to EGFR inhibitors is associated with a manifestation of stem cell-like properties in cancer cells. *Cancer Res.* **73**, 3051–3061 (2013).

22. Costa, D.B., Kobayashi, S., Yeo, W.L. & Hamada, A. Serum concentrations of Erlotinib at a dose of 25 mg daily. *J. Thorac. Oncol.* **5**, 1311–1312 (2010).
23. Bean, J. *et al.* MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. *Proc. Natl. Acad. Sci. USA* **104**, 20932–20937 (2007).
24. Engelman, J.A. *et al.* MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science* **316**, 1039–1043 (2007).
25. Sequist, L.V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci. Transl. Med.* **3**, 75ra26 (2011).
26. Ware, K.E. *et al.* A mechanism of resistance to gefitinib mediated by cellular reprogramming and the acquisition of an FGF2-FGFR1 autocrine growth loop. *Oncogenesis* **2**, e39 (2013).
27. Byers, L.A. *et al.* An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clin. Cancer Res.* **19**, 279–290 (2013).
28. Kubonishi, I. & Miyoshi, I. Establishment of a Ph1 chromosome-positive cell line from chronic myelogenous leukemia in blast crisis. *Int. J. Cell Cloning* **1**, 105–117 (1983).
29. Soverini, S. *et al.* BCR-ABL kinase domain mutation analysis in chronic myeloid leukemia patients treated with tyrosine kinase inhibitors: recommendations from an expert panel on behalf of European LeukemiaNet. *Blood* **118**, 1208–1215 (2011).
30. Soverini, S. *et al.* Unraveling the complexity of tyrosine kinase inhibitor-resistant populations by ultra-deep sequencing of the BCR-ABL kinase domain. *Blood* **122**, 1634–1648 (2013).
31. Adrián, F.J. *et al.* Allosteric inhibitors of Bcr-abl-dependent cell proliferation. *Nat. Chem. Biol.* **2**, 95–102 (2006).
32. Zhang, J. *et al.* Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. *Nature* **463**, 501–506 (2010).
33. Gorre, M.E. *et al.* Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* **293**, 876–880 (2001).
34. Lange, T. *et al.* The quantitative level of T315I mutated BCR-ABL predicts for major molecular response to second-line nilotinib or dasatinib treatment in patients with chronic myeloid leukemia. *Haematologica* **98**, 714–717 (2013).
35. Kunkel, T.A. & Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* **69**, 497–529 (2000).
36. Robbins, C.M. *et al.* Copy number and targeted mutational analysis reveals novel somatic events in metastatic prostate tumors. *Genome Res.* **21**, 47–55 (2011).
37. Tewhey, R. *et al.* Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat. Biotechnol.* **27**, 1025–1031 (2009).
38. Zhong, Q. *et al.* Multiplex digital PCR: breaking the one target per color barrier of quantitative PCR. *Lab Chip* **11**, 2167–2174 (2011).
39. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
40. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
41. Garnett, M.J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
42. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, p11 (2013).
43. Schmitt, M.W. *et al.* Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. USA* **109**, 14508–14513 (2012).
44. Shiroguchi, K., Jia, T.Z., Sims, P.A. & Xie, X.S. Digital RNA sequencing minimizes sequence-dependent bias and amplification noise with optimized single-molecule barcodes. *Proc. Natl. Acad. Sci. USA* **109**, 1347–1352 (2012).
45. Kinde, I. *et al.* Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. *Sci. Transl. Med.* **5**, 167ra164 (2013).
46. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K.W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. USA* **108**, 9530–9535 (2011).
47. Diehl, F. *et al.* BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat. Methods* **3**, 551–559 (2006).
48. Bozic, I. *et al.* Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife* **2**, e00747 (2013).
49. Komarova, N.L. & Wodarz, D. Drug resistance in cancer: principles of emergence and prevention. *Proc. Natl. Acad. Sci. USA* **102**, 9714–9719 (2005).
50. Iwasa, Y., Michor, F. & Nowak, M.A. Evolutionary dynamics of escape from biomedical intervention. *Proc. Biol. Sci.* **270**, 2573–2578 (2003).

ONLINE METHODS

High-complexity DNA library construction. Approximately 31.5 nmol of 115-bp-long oligonucleotide containing a semi-random 30-bp-long barcode sequence (i.e., 15 repeats of A/T (W)–G/C (S)) and a flanking primer pair for barcode amplification were synthesized by Integrated DNA Technology. The double-stranded oligonucleotides were generated by extension reaction and cloned into lentiviral vector pRSI9-U6-(sh)-UbiC-TagRFP-2A-Puro (Cellecra) at the ClaI-XhoI site at an estimated vector:insert ratio of 1.35×10^{12} : 8.10×10^{12} molecules. Approximately 24% of the ligated products were transformed into MegaX DH10B T1^R electrocompetent cells (Life Technologies), and the entire transformed bacteria were inoculated into 500 ml LB medium containing 100 µg/ml carbenicillin. Deep sequencing of the plasmid library pool at a depth of approximately 161 million reads revealed more than 27 million unique barcodes observed at least twice. On the basis of these data, computational modeling projected a maximum complexity of more than 73 million unique barcodes (**Supplementary Fig. 1**). The barcode library was relatively evenly distributed, with 95% of the 27 million observed barcodes detected with 2–12 reads. Furthermore, the most represented barcode constituted only 0.000037% of the total library, and the barcodes represented by more than 12 reads composed only 5% of the library (**Fig. 1b**). To prevent transcription of DNA barcode sequences that might potentially cause biological effects, the U6 promoter was removed and an RNA polymerase III terminator site was inserted upstream of the WS barcode. Detailed sequence information can be found in **Supplementary Table 1**. TagRFP and a puromycin-resistant gene of the pRSI9-U6-(sh)-UbiC-TagRFP-2A-Puro vector were used as selection markers to isolate barcoded cell populations.

Barcoding of cancer cell lines. HCC827 and KCL-22 cell lines were cultured in RPMI-1640 medium supplemented with 10% FBS and 1% penicillin-streptomycin. Cells were barcoded by lentiviral infection using 0.8 µg/ml polybrene. After a 24-h incubation with virus, infected cells were selected with puromycin. To ensure that the majority of cells were labeled with a single barcode per cell, for lentiviral infection we used a target m.o.i. of approximately 0.1, corresponding to 10% infectivity after puromycin selection. Infected cell populations were expanded in culture for the minimal time period to obtain a sufficient number of cells to set up replicate experiments.

Compound-resistance studies in HCC827 and KCL-22 cells. HCC827 cells were treated with 2 µM erlotinib (a concentration that represents a clinically relevant drug exposure²²) for 36 d, starting 1 d after plating. Drug-resistant cell colonies were observed by 36 d in all cases. An immediate cytotoxic effect was observed within 72 h in response to erlotinib treatment, obviating the need for splitting or passaging, which further reduced the chance of random barcode loss during passaging. Therefore, the control groups were treated with 0.1% DMSO and harvested after 6 d until cells were confluent to avoid passing and to keep the baseline clonal population as close as possible to that of the erlotinib-treatment group. For combination- or sequential-treatment experiments with the HCC827 line, cells were treated with 0.2 µM crizotinib. KCL-22 cells were treated with three different ABL1 inhibitors with two distinct mechanisms of action: (i) ATP-binding site inhibitors imatinib (2.5 µM) and nilotinib (300 nM) and (ii) allosteric inhibitor GNF-2 (2.5 µM). We chose 5.5–7.8-fold IC₅₀ of each inhibitor as a final concentration for treatment in order to apply comparable selective pressure and avoid off-target effects that can occur at higher drug concentrations (**Supplementary Fig. 7**). For control cell populations, the replicate flasks were treated with 0.1% DMSO until the cells reached confluence. All the cell lines used in the study were obtained from the Cancer Cell Line Encyclopedia collection, and cell line authentication was confirmed by single-nucleotide polymorphism testing at multiple time points: (i) after the cell lines were barcoded by lentiviral infection and (ii) after the drug-resistant subpopulations of cells were selected. We estimated the barcode complexity of starting (pretreatment) cell populations by applying a degree-six polynomial function to analysis of the cumulative wealth distribution on the barcode NGS results for the vehicle-treated populations.

Barcode amplification. At the end of the treatment period, cells from each replicate treated with either vehicle or drug(s) were harvested as a clonal pool

and snap-frozen. Genomic DNA was extracted from the frozen cell populations with a DNeasy Blood & Tissue Kit or QIAamp DNA Blood Maxi Kit (Qiagen). We used PCR to amplify the barcode sequence for NGS by introducing Illumina adaptors and 5-bp-long index sequences. The sampling of sufficient template coverage was ensured by parallel PCR reactions. For each PCR reaction, up to 2 µg of genomic DNA was used as a template. The sequence information for the primers used for barcode amplification can be found in **Supplementary Table 2**. Labeling each sample with 1 of the 20 unique indices enabled us to multiplex and sequence up to 20 samples at once.

Isolation of drug-resistant KCL-22 single clones. Single clones were isolated from an extra replicate of the drug-resistant cell population by FACS. To identify a KCL-22 clone that was resistant to both catalytic and allosteric ABL1 inhibitors, we isolated single clones from a GNF-2-resistant pooled cell population by FACS. Stamped replicate plates of GNF-2-resistant clones were treated with either 300 nM nilotinib or a combination of 300 nM nilotinib and 2.5 µM GNF-2. Clone 4 showed cross-resistance in both nilotinib cotreatment conditions, whereas clones 1 and 2 showed sensitivity to nilotinib treatments. Clones 1, 2 and 4 from the original plate that were maintained under 2.5 µM GNF-2 without any exposure to catalytic inhibitors were used for further analysis. Clone 3 was isolated from a 2.5 µM imatinib-resistant cell population. Genomic DNA was extracted using a QIAamp DNA Blood 96 kit or DNeasy Blood & Tissue Kit (Qiagen), and the barcode sequence was amplified using the primers listed in **Supplementary Table 3** for Sanger sequencing. The growth rate of each clone along with that of the barcoded KCL-22 parental pool was measured with a Vi-Cell XR Cell Viability Analyzer (Beckman Coulter) over the 6-d period under 0.1% DMSO (vehicle), 2.5 µM imatinib, 300 nM nilotinib or 2.5 µM GNF-2. Both the viable cell number and the total (viable + dead) cell number were recorded. The IC₅₀ of each clone for the three ABL1 kinase inhibitors was determined by Cell TiterGlo assay (Promega) after 5 d of compound treatment. Luminescent measurements were normalized to the vehicle-treated controls.

Next-generation barcode sequencing. PCR-amplified products were quantified using the Standard Sensitivity NGS Fragment Analysis Kit (Advanced Analytical Technologies) on the Advanced Analytical Fragment Analyzer Automated CE System. Fragment sizes and concentrations of PCR amplicons were analyzed using Advanced Analytical PROSize 2.0 software. PCR amplicons were then prepared at 10 nM and loaded at 3.5 pM for sequencing on the Illumina HiSeq2500 sequencer in Rapid Mode using the 50 Cycle TruSeq Rapid SBS Kit, TruSeq Rapid SR Cluster Kit, and HiSeq Rapid SR Flow Cell (Illumina). Samples were sequenced at 66 cycles for read 1 and 7 cycles for the i7 index read in Rapid Mode. PCR amplicons were also prepared at 10 nM and loaded at 3.25 pM for sequencing on the Illumina HiSeq2500 sequencer in High Output Mode using the 50 Cycle TruSeq v3 SBS Kit, TruSeq SR Cluster Kit v3 cBot, and HiSeq SR Flow Cell (Illumina). Samples were sequenced at 51 cycles for read 1 and 7 cycles for the i7 index read in High Output Mode.

Barcode-composition analysis. FASTQ files were prepared from the barcode-sequencing runs to count the number of reads and the fraction of barcodes in each sample. Reads were then filtered to keep those that (a) showed the WS × 15 pattern; (b) matched the expected sequence after the WS × 15 barcode for sequence libraries with lengths of ≥37 bp (**Figs. 2e,f, 3a, 4b–d and 5a–e**); and (c) had an estimated Phred quality score of at least 10 for all base pairs in the read, with an average Phred quality score greater than 30. All barcodes observed at least twice that passed these criteria were kept. Sets of barcodes were then merged to account for sequencing errors if either (a) the test barcode was a hamming distance of 1 from the more abundant barcode and observed at 1/8th of the count or (b) the test barcode was a hamming distance of 2 from the more abundant barcode and observed at 1/40th the count; this is similar to the approach used by Lu *et al.*¹². For those sequence libraries with lengths of exactly 30 bp (DMSO4 and ERL1–8 shown in **Fig. 2a–d**), the hamming distance also allowed insertions or deletions of 2 bp; because of the barcode design and our filtering steps, insertions or deletions of odd-numbered length or in the longer reads should already have been filtered out. After these steps, each barcode set was annotated with the sequence of the most abundant barcode, counts before merging,

counts after merging, and fraction with respect to the total count of all barcodes that passed the read filters.

Calculation of barcode overlap between samples. To establish the significance of sharing between samples treated with different drugs, we computed 'sharing ratios'. The sharing ratio was defined as the ratio of significant barcodes of one replicate seen in another to the total number of significant barcodes in that replicate. A barcode was called significant if it was seen in 0.021% of the total population in the HCC827 experiments (which was the highest fraction observed in DMSO-treated groups) and in 0.07% in the KCL-22 experiments (which accounted for <1% of the DMSO-treated population). A barcode was considered shared between two samples if it was significant in both samples and, to rule out possible errors due to low-level contamination during sequencing, the fraction of the barcode in the less abundant sample was at least 1% compared to the fraction in the more abundant sample. Further details can be found in **Supplementary Table 9** and **Supplementary Figure 9b–e**.

RNA-seq. Total RNA was isolated from snap-frozen cells using the RNeasy Plus Mini Kit (Qiagen) and was quantified using the Agilent RNA 6000 Nano Kit (Agilent Technologies) on the Agilent 2100 BioAnalyzer. One microgram of high-purity total RNA (defined as having an RNA integrity number greater than 7.0) was used as input for the Illumina TruSeq RNA Sample Prep Kit, Sets A/B (48Rxn) (Illumina). The gel-free protocol was employed for the TruSeq RNA Sample Prep Kit per the manufacturer's specifications and performed on the Beckman Coulter Biomek FXp robotics platform. The standard RNA-fragmentation profile was used as recommended by Illumina (94 °C for 8 min). The PCR-amplified RNA-seq library products were then quantified using the Fragment Analyzer Standard Sensitivity NGS Fragment Analysis Kit (Advanced Analytical Technologies). The samples were diluted to 10 nM in EB Buffer (Qiagen), denatured, and loaded at 2.75 pM on an Illumina HiSeq2500 in Rapid Run Mode using TruSeq Rapid PE Cluster Kit–HS and TruSeq Rapid SBS Kit–HS (200 cycle) reagents (Illumina). The RNA-seq libraries were sequenced at 100 bp paired-end with a 7-bp index using the standard Illumina primers. The sequence-intensity files were generated on an instrument using the Illumina Real Time Analysis software. The intensity files were demultiplexed and FASTQ files were created using the CASAVA 1.8.2 software suite.

RNA-seq analysis. RNA-seq analysis, mutation discovery, and expression quantification were performed according to previously published methods^{51,52}. Briefly, FASTQ files were preprocessed for base quality and to remove adaptor sequences. They were then aligned with bowtie2 (ref. 53) and a modified version of tophat1.3 (ref. 54) against a custom genome or transcriptome FASTA file. Mutations were then called using GATK best practices⁵⁵. FPKM values were generated using cufflinks version 2.0.2 (ref. 56) against known transcripts from UCSC Known Genes (<http://genome.ucsc.edu>).

RNA-seq pathway enrichment. Pathways derived from GO terms and transcription factor networks were analyzed for overrepresentation via a one-tailed interpolated Fisher's exact test, using genes that varied fourfold or more between treatment conditions. Benjamini–Hochberg correction was then applied to these *P* values⁵⁷.

qPCR copy-number assay. Copy-number analysis of select genes was performed in triplicate using TaqMan Copy Number Assays (Life Technologies) on a CFX96 Touch RealTime PCR Detection Platform (Bio-Rad). cMET TaqMan Copy Number Assay (#Hs02323823_cn, FAM-labeled, 20× concentrate) was duplexed with RNaseP (RPPH1) TaqMan Copy Number Reference Assay (#4401631, VIC-labeled, 20× concentrate) in TaqMan Genotyping Master Mix, and qPCR was conducted according to the manufacturer's instructions. Whole gDNA extract was included on each plate as a diploid normalization control (Promega). Real-time amplification plots and *C_t* calculations were derived from Bio-Rad CFX Manager v3.0 software according to the manufacturer's instructions.

Sequenom BCR-ABL1 mutation assay. DNA samples were genotyped with Sequenom iPLEX Pro (Sequenom). PCR primers were designed using Assay

Design Suite (Sequenom) to amplify the following target mutations: Y253H, T315I, A337V, F359V, P465F and V468F. Primer sequences are available in **Supplementary Table 12**. All PCR primers were screened against mouse DNA to confirm the absence of cross-reaction. As a result, Y253H, A337V, F359V and V468F reactions were run in multiplex, whereas T315I and P465F were run in single-plex. Following a shrimp alkaline phosphatase reaction to neutralize extra deoxyribose-containing nucleoside triphosphates, we generated allele-specific extension products using primers located 1 bp away from the target mutation. Extension products were transferred onto a SpectroCHIP via a MassArray Nanodispenser (Sequenom) and subjected to MALDI-TOF mass spectrometry. The resultant spectra were analyzed by Sequenom MassArray Typer4.0 software (Sequenom).

Mathematical modeling to predict clonal frequency of T315I and A337V mutant KCL-22 clones after GNF-2 treatment. We calculated the expected total abundance of A337V (denoted by A_{337V_t}) and T315I (denoted by T_{315I_t}) clones after 21 d of GNF-2 treatment using standard exponential growth equations:

$$A_{337V_t} = A_{337V_0} \exp(r_{A_{337V}}t)$$

$$T_{315I_t} = T_{315I_0} \exp(r_{T_{315I}}t)$$

$$\frac{A_{337V_t}}{T_{315I_t}} = \frac{T_{337V_0}}{T_{315I_0}} \times \frac{\exp(r_{A_{337V}}t)}{\exp(r_{T_{315I}}t)}$$

where *t* is the number of treatment days, A_{337V_0} and T_{315I_0} are the numbers of A337V and T315I clones at the start of treatment, and $r_{A_{337V}}$ and $r_{T_{315I}}$ are the growth rates of the two cell types per day. Under DMSO treatment, the calculated net growth rate of clone 1 was determined as 0.308 per day and was applied for A337V mutant clones ($r_{A_{337V;DMSO}}$) in the above equation, and the net growth rate of clone 4 was determined as 0.251 and was applied for T315I mutant clones ($r_{T_{315I;DMSO}}$) (**Supplementary Table 11**). In the vehicle-treated replicates (0.1% DMSO treatment for 7 d), the average barcode fraction of clone 1 in the entire barcoded KCL-22 cell population was 5.57×10^{-5} , whereas clone 4 was found to have a frequency of 8.34×10^{-6} . The abundance of each clone at day 0 was calculated as

$$A_{337V_0} = 5.57 \times 10^{-5} \times \exp(0.308(-7)) = 6.45 \times 10^{-6}$$

$$T_{315I_0} = 8.34 \times 10^{-6} \times \exp(0.251(-7)) = 1.43 \times 10^{-6}$$

This calculation provides the ratio of the abundance of clone 1 to that of clone 4 within the starting population before GNF-2 treatment as approximately 4.51:1 (A_{337V_0}/T_{315I_0}). Under GNF-2 treatment, the calculated net growth rate of clone 1 (0.288) was applied for A337V mutant clones ($r_{A_{337V;GNF-2}}$), and the net growth rate of clone 4 (0.250) was applied for T315I mutant clones ($r_{T_{315I;GNF-2}}$) (**Supplementary Table 11**). Therefore, this model predicted that the ratio of the abundance of A337V mutant clones to that of T315I clones should be approximately 10:1 after 21 d of GNF-2 treatment, as $(4.514/1) \times (\exp(0.288 \times 21)/\exp(0.250 \times 21)) = 10.01$.

Mathematical modeling. I. Stochastic mathematical model. We designed a nonhomogeneous continuous-time multitype birth–death process with mutations to model the population dynamics of cells (**Supplementary Fig. 11**). Each uniquely barcoded clone was considered to consist of two cell types: sensitive cells, whose number at time *t* is denoted by N_s , and resistant cells, whose number at time *t* is denoted by N_r . Their corresponding birth and death rates are denoted by b_s , d_s , b_r and d_r in the equations below. During each sensitive-cell division, a resistant mutation arose at rate *u*. This model was then implemented as an exact stochastic computer simulation. Let $\Theta = (b_s + d_s)N_s + (b_r + d_r)N_r$. Then the transition probabilities for the stochastic simulation are given as follows:

$$P(\text{total sensitive cells increase by 1 via cell division}) = \frac{b_s N_s (1 - \mu)}{\Theta}$$

$$P(\text{total sensitive cells decrease by 1 via cell death}) = \frac{d_s N_s}{\Theta}$$

P (total mutant cells increase by 1 via sensitive-cell division and mutation): $\frac{b_s N_s \mu}{\Theta}$

P (total mutant cells increase by 1 via mutant-cell division): $\frac{b_r N_r}{\Theta}$

P (total mutant cells increase by 1 via mutant-cell death): $\frac{d_r N_r}{\Theta}$

The inter-event time follows an exponential distribution with the rate parameter Θ . Given the low mutation rate, on the order of $u < 10^{-7}$ per cell division, as well as the small initial number of cells per barcode, our model ignores the scenario in which multiple individual mutations arise, either sequentially or in independent cells. To investigate the question of whether resistant mutations are pre-existing or acquired during treatment, we incorporated an additional parameter into our model; this parameter, ρ , represents the initial proportion of barcodes that are resistant to treatment. Note that this parameter represents not the proportion of all cells that are resistant to treatment, but the proportion of barcodes that label resistant populations. For each simulation run, a random set of barcodes was selected to be resistant for a given value of ρ . These barcodes were selected with weights proportional to the sizes of their populations. All cells with these barcodes were then deemed resistant to treatment.

Parameter estimation. We then designed an estimation method to determine the growth and death rates of the cell population from experimental data. We denote the total numbers of viable and dead cells by N_v and N_d , respectively, and the birth and death rates of cells by b and d . To describe the average behavior of the cell population, we used the following system:

$$\begin{aligned} \frac{dN_v}{dt} &= (b-d)N_v \\ \frac{dN_d}{dt} &= dN_v \end{aligned} \quad (1.1)$$

The analytical solution to this system of equations is given by

$$\begin{aligned} N_v(t) &= C_1 e^{(b-d)t} \\ N_d(t) &= \frac{C_1 d (e^{(b-d)t} - 1)}{b-d} + C_2 (b-d) \end{aligned} \quad (1.2)$$

where C_1 and C_2 are the initial numbers of viable and dead cells at time $t = 0$. To estimate the parameters of this model from the experimental data, we then implemented a grid-search algorithm to estimate the birth and death rates for the cell line cultures under treatment with DMSO, GNF-2, nilotinib and imatinib. We minimized the following objective function separately for each of the four treatment conditions (DMSO, GNF-2, nilotinib and imatinib):

$$\arg \min_{b,d} \sum_{i=0}^N \{ (N_v(t_i) - N_v^{\text{obs}}(t_i))^2 + (N_d(t_i) - N_d^{\text{obs}}(t_i))^2 \} \quad (1.3)$$

where $C_1 = N_v^{\text{obs}}(0)$ and $C_2 = N_d^{\text{obs}}(0)$.

Results. We applied the estimation algorithm outlined above to the barcoded KCL-22 parental cell population and clones 1–4 treated with DMSO, GNF-2, nilotinib or imatinib. **Supplementary Table 11** displays the estimated parameter values—the birth rates b and death rates d —for different cell types and growth conditions.

Note that treatment reduced birth rates without affecting death rates to a large extent. GNF-2 treatment lowered birth rates only slightly compared to DMSO treatment. In contrast, nilotinib and imatinib lowered birth rates significantly for clones 1 and 2 and, but only slightly for clones 3 and 4.

We then used these estimated parameter values in the stochastic model. To determine the birth rate of resistant cells for each treatment condition, we identified clones 1 and 2 as resistant to GNF-2 treatment and clones 3 and 4 as resistant to imatinib and nilotinib treatment. We then used the data in **Supplementary Table 11** to determine the average birth rates of resistant clones as follows:

we averaged the birth rates of clones 1 and 2 during GNF-2 treatment and the birth rates of clones 3 and 4 during imatinib and nilotinib treatment.

$$\hat{b}_r = \frac{0.294 + 0.316 + 0.300 + 0.270 + 0.286 + 0.256}{6} = 0.282 \approx 0.3$$

Similarly, we determined the death rate of resistant cells by calculating the average resistant-cell death rate as

$$\hat{d}_r = \frac{0.006 + 0.020 + 0.026 + 0.018 + 0.032 + 0.018}{6} = 0.020$$

Because of the lack of quantitative measurements of the numbers of both viable and dead parental cells during treatment, we were not able to estimate the birth and death rates of these cells under treatment in the same way. Instead, we determined the net growth rate, $b - d$, using the viable-cell growth data from day 0 to day 12, before the initiation of the rebound due to resistant clones (**Fig. 4a**). The net rate was estimated as $b - d = -0.06$. Although this net rate did not provide us with the birth and death rates individually, it did provide an additional constraint that helped us to refine the search space for the birth and death rates of parental cells in the presence of treatment. We then investigated a large range of birth rates spanning 0.30–0.00; the corresponding death rate was constrained to satisfy $b - d = -0.06$. In addition to birth and death rates, we also investigated a large range of mutation rates, from 10^{-7} to 10^{-9} per cell division. These values were chosen because the baseline mutation rate for genetically stable cells has been determined as 10^{-10} – 10^{-11} per base per cell division³⁵, and multiple genomic sites may be able to lead to resistance when mutated. Lastly, the proportion of resistant barcodes ρ was investigated for a range from 0.000 to 1.000.

Figure 6a shows the predicted population dynamics of cells when considering different fractions of pre-existing resistant barcodes. The model predictions show that the experimental data cannot be explained if we assume that there are no pre-existing resistant cells present at the initiation of treatment ($\rho = 0$). However, positive initial frequencies of resistant barcodes can recapitulate the pattern of population rebounds observed during treatment. We found that a pre-existing resistant cell frequency of $\rho \approx 0.02\%$ – 0.03% led to model predictions that provided a good match to the experimental data obtained from GNF-2-treated cells, whereas $\rho \approx 0.005\%$ – 0.01% led to model predictions that provided a good match to the experimental data obtained from imatinib- and nilotinib-treated cells. The addition of newly emerging resistant clones, even at rates as high as 10^{-7} per cell division, did not lead to significantly different results. Thus, the most important parameter for shaping the population dynamics of cells is the proportion parameter ρ : as ρ increases, the total cell number rebounds faster after initially responding to treatment. This initial response is caused by the decline of the sensitive cells, whereas the rebound is driven by outgrowth of resistant clones that existed before the initiation of treatment. For a given value of ρ , the model-predicted population dynamics overlap for different mutation, birth and death rates, suggesting that those parameters have less of an effect on the population dynamics than ρ .

We then investigated the distributions of barcodes in the model-predicted cell populations after 21 d of treatment with GNF-2 and 27 d of treatment with imatinib or nilotinib for different values of ρ (**Supplementary Fig. 13a–h**). As ρ increased, the number of barcodes exceeding the threshold of a fraction of 0.0007 increased, leading to more barcodes at greater frequencies. The distributions of barcodes were not sensitive to the mutation rate μ ; however, in the extreme case of no pre-existing resistant barcodes ($\rho = 0$), the barcode distributions were sensitive to the birth and death rates (**Supplementary Fig. 13a–h**). The complete barcode-distribution plots for all combinations of mutation, birth and death rates are shown in **Supplementary Data Set 1**. When the simulated distributions are compared to the observed distributions (**Fig. 5b–e**), the observed distributions better resemble the cases in which $\rho \geq 0.0001$, providing further evidence that resistance is likely to be pre-existing.

We also investigated whether the prevalent clones with barcode fractions greater than the threshold of 0.0007 were resistant to treatment (**Supplementary Fig. 13i–p**). Under the assumption of no pre-existing resistant barcodes ($\rho = 0$),

none of the most prevalent cells at day 21 or 27 were resistant to treatment. This result implies that newly emerging resistant clones are present only in small fractions—less than 0.0007—and thus would be insufficient to drive the population dynamics leading to a rebound. In contrast, for the simulation runs in which $\rho > 0$, a large percentage of barcodes were resistant (**Supplementary Fig. 13i–p**). The complete plots for the numbers of barcodes with fractions greater than 0.0007 for all combinations of mutation, birth and death rates are shown in **Supplementary Data Set 2**.

As an alternative assumption, we also investigated another method of selecting resistant barcodes (results not shown). In this method, resistant barcodes were selected randomly with equal weights. Although the parameter values for ρ needed in order to recapitulate the observed data differed between the two selection methods, the qualitative result remained the same: in order for the observed data to be recapitulated, the proportion parameter ρ had to be greater than 0.

In summary, our observations of the distributions of barcodes and the number of resistant barcodes consistently contradicted the assumption of no pre-existing resistant barcodes, providing rational evidence supporting pre-existing resistance in this CML cell line.

Discussion. Our mathematical model was able to recapitulate the observed population dynamics, and its predictions show that the observed population dynamics cannot be explained without pre-existing resistance in the CML cell line. Our stochastic model is based on a few assumptions. (1) Mutant cells arising from two or more independent mutation events in the same barcoded parental cells are considered to be phenotypically identical, with the same birth and death rates. This assumption simplifies our model and increases the simulation speed. However, our model can easily be extended to include scenarios in which parental cells give rise to multiple cell phenotypes, through either sequential accumulation of mutations or multiple independent mutations, if birth and death rates for each of the individual mutants are available. (2) Birth and death rates for the resistant cells were calculated using the average of all subclones resistant to GNF-2, nilotinib and imatinib. We did not perform simulations using the birth and death rates specific to each drug because of the similarity in the parameter estimates. (3) Our model ignores the possibility of interbarcode competition, assuming cells with different barcodes grow independently. Given the limited resources in the *in vitro* experimental setting, the actual underlying growth pattern might resemble logistic growth instead of exponential growth. However, the increased model complexity stemming from the inclusion of a carrying capacity would require detailed measurements of the carrying capacity and the parameters characterizing the extent of interbarcode competitions. Given that our model was able to capture the population dynamics and barcode distributions, we believe that it was not necessary to increase the model complexity by considering logistic growth.

II. Mathematical modeling of generation of ‘pre-existing’ mutations during clonal expansion of small cell population. We constructed a simple birth–death process with mutations to model the clonal evolution process starting from a small number of homogeneous cells ($N_w = 1,000$). The simulations were initiated with 1,000 wild-type cells and no mutant cells ($N_m = 0$). The corresponding birth and death rates for the two cell types are denoted by b_w, d_w, b_m and d_m , and

the mutation rate is denoted by u . During each elementary time step, one of the five scenarios may occur:

Total wild-type cells increase by 1 through cell division:

$$\frac{b_w N_w (1-u)}{(b_w + d_w) N_w + (b_m + d_m) N_m}$$

Total wild-type cells decrease by 1 through cell death:

$$\frac{d_w N_w (1-u)}{(b_w + d_w) N_w + (b_m + d_m) N_m}$$

Total mutant cells increase by 1 through division of wild-type cells:

$$\frac{b_w N_w u}{(b_w + d_w) N_w + (b_m + d_m) N_m}$$

Total mutant cells increase by 1 through mutant-cell division:

$$\frac{b_m N_m}{(b_w + d_w) N_w + (b_m + d_m) N_m}$$

Total mutant cells increase by 1 through mutant-cell death:

$$\frac{d_m N_m}{(b_w + d_w) N_w + (b_m + d_m) N_m}$$

The inter-event time follows an exponential distribution with the rate parameter $\lambda = (b_w + d_w) N_w + (b_m + d_m) N_m$. For the birth and death rates of wild-type cells, we used the rates obtained from *in vitro* experiments performed on the KCL-22 cell line, with $b_w = 0.356$ and $d_w = 0.030$. For the birth and death rates of mutant cells, we explored various combinations of birth and death rates ranging from 0.00 to 0.45 for birth rates and from 0.00 to 0.15 for death rates, at an increment size of 0.05. The scenarios in which mutant birth rates were smaller than the death rates were ignored, because in such cases mutant cells cannot expand their population sizes. We explored a large range of mutation rates from 10^{-6} to 10^{-10} per cell division. For each combination of birth, death and mutation rate, we performed 100 simulation runs, and each simulation was terminated after 45 time units.

The results from our simulations are summarized graphically in **Supplementary Figure 15**. We found that for certain combinations of birth, death and mutation rates, mutant cells reached significant sizes, with $N_m / (N_w + N_m) > 10^{-4}$, sufficient to drive the rebounding behaviors observed in **Figure 6a** to explain the observed trajectories.

51. Korpai, M. *et al.* An F876L mutation in androgen receptor confers genetic and phenotypic resistance to MDV3100 (enzalutamide). *Cancer Discov.* **3**, 1030–1043 (2013).
52. Rahal, R. *et al.* Pharmacological and genomic profiling identifies NF- κ B-targeted treatment strategies for mantle cell lymphoma. *Nat. Med.* **20**, 87–92 (2014).
53. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
54. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
55. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
56. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
57. Wiederschain, D. *et al.* Contribution of polycomb homologues Bmi-1 and Mel-18 to medulloblastoma pathogenesis. *Mol. Cell. Biol.* **27**, 4968–4979 (2007).