Supplementary Methods

PATIENT SAMPLES

DNA was extracted from samples obtained from first-line CML patients with the Agilent DNA extraction kit #200600. For patients from the PACE trial, genomic DNA was extracted with the QIAamp RNA blood kit (Qiagen) via a custom protocol at MolecularMD (Portland, OR). Mutation gains in the PACE trial were defined by Sanger sequencing of baseline samples and end of treatment samples or surrogate end of treatment samples (last visit prior to treatment discontinuation). Our analysis of mutation gains in PACE patients was restricted to appropriately consented patients with both baseline and end of treatment sequencing results. Specifically, among all PACE patients, we required (i) availability of both baseline and end of treatment Sanger sequencing, (ii) on-treatment mutation gain, and (iii) appropriate consent for high-sensitivity molecular genetics testing. Specific consent language varied among trial sites, and overall 30 out of 49 potentially eligible patients had completed consent which was considered to allow Duplex Sequencing. "Treatment failure" in first-line CML patients (1,2) was defined according to European LeukemiaNet guidelines (3) (e.g. BCR-ABL1/ABL1>10% at 6 months or >1% at 12 months). BCR-ABL1/ABL1 transcript ratios were determined by RT-PCR. In the TOPS trial, beta-2-microglobulin was used as the control gene rather than ABL1. These values were converted to BCR-ABL1/ABL1 as previously described (4). Patient information is shown in **Supplementary Table 4**.

DUPLEX SEQUENCING

Duplex Sequencing of the ABL1 gene from genomic DNA was performed essentially as previously described (5,6). Duplex Sequencing data analysis has been described (5), and

software for data analysis can be found at the following website:

https://github.com/loeblab/Duplex-Sequencing All mutations were individually inspected in the Integrative Genomics Viewer (7). Germline polymorphisms (i.e. variants having allele fraction of 50%-100% which are listed in dbSNP) were excluded from analysis. All mutations identified are shown in **Supplementary Table 6.** Mutation frequency was calculated by dividing the total number of unique mutations identified by the total number of Duplex nucleotides obtained. Only unique mutations were counted in the numerator, as observing the same mutation twice in a sample with a low mutation burden is unlikely to reflect two independent mutational events, but rather is expected to indicate a single mutational event with subsequent clonal expansion. This calculation likely under-estimates the mutation frequency in the RT-PCR amplified samples, as the high error rate of RT-PCR resulted in mutations in a large fraction of the positions in the sequenced target.

RT-PCR

To determine the extent of artefactual mutations introduced by RT-PCR amplification, nested reverse transcription PCR of ABL1 exons 4-8 (codons 199-456) was performed as previously described (8), followed by Duplex Sequencing without hybrid capture (9). Read depth was poor at amino acid positions 265-274. These positions (which encompass only a single previously reported resistance mutation, L273M), were excluded from analysis, resulting in 733 nucleotide positions that were evaluated for each of the RT-PCR samples. RT-PCR error frequency was determined by dividing the total number of unique mutations identified by the total number of nucleotides sequenced (e.g. 323 unique mutations seen upon sequencing of 6.1×10^{6} nucleotides corresponds to an error frequency of 5.3×10^{-5} errors per nucleotide sequenced).

PHASING OF COMPOUND MUTATIONS

Ion Torrent PGM sequencing of a 400bp region of the ABL1 kinase domain was performed on all baseline samples for which Sanger sequencing presented an ambiguous interpretation of the sequencing result, and in all patients with putative triple mutants. We used the method of Deininger et al. (10) to determine the phasing of point mutations in individual patients. Compound mutations were not phased in Duplex Sequencing analyses, as Duplex Sequencing utilizes relatively short fragments of randomly sheared DNA (average fragment size ~200 nucleotides).

PREDICTION OF RESISTANCE POPULATION SIZES

To be eligible for a population size prediction, patients had to have a visible reduction in BCR-ABL1 levels and a visible rebound. A visible difference was defined as a reduction of one log of molecular response and one log of molecular relapse. Of 30 patients who had a change in mutation status at the end of treatment, 12 gained and lost one log of molecular response, and had quantitatively measurable molecular response (i.e. no CMR where BCR-ABL1/ABL1 levels are undetectable) at all time points. Starting from the last local minimum of molecular response, we then fit the log-linear model:

$$\log_{10}(Y) = mt + b$$

where *Y* is the BCR-ABL1/ABL1 measurement, *t* is time, and *m* and *b* are parameters to describe the response. Note that BCR-ABL1/ABL1 levels are directly proportional to cell number (11). Assuming exponential growth, which has been shown in 2 prior studies (12,13), i.e.

$$\frac{dY}{dt} = kY$$

it can be shown that

$$k = \frac{m}{\log_{10}(e)} \ .$$

The same log-linear model was used to derive the decay parameter for the sensitive regime of the molecular response.

In order to determine if resistant cells were present at the start of treatment, simulations were run using a joint stochastic-deterministic branching process model. The tau-leaping method (14,15) was used to stochastically simulate subpopulations of fewer than 10^4 cells. The birth-death dynamics of larger subpopulations were modeled using deterministic ordinary differential equations for the sake of computational efficiency, as was done in Bozic *et al*. (16). The differential equations used to describe the stochastic birth-death process are as follows:

$$\frac{dS}{dt} = \alpha S - \beta S$$

$$\frac{dR}{dt} = \alpha R - \beta R + \mu S$$

The net decay of the sensitive population and net growth of the resistant populations as determined by the log-linear fit of the molecular response were used to determine the division and death frequencies of both subpopulations. The mutational frequency was the same as prior work and is set to be 4e-7 (13). The stem cell size (i.e. the proportion of the sensitive population capable of seeding a resistant population) was assumed to be 1/1000 of the measured tumor burden (a highly conservative measure considering the discussion later in this supplement). The initial sensitive population was the measured white blood cell count at the

start of treatment assuming 5 L of peripheral blood in the body. To compare the patient response with the predicted response given the counterhypothesis that there were no resistant cells at the start of treatment, the initial resistant population was set to zero. With these parameters and initial conditions, the branching process model was simulated 100 times for each patient using MATLAB.

The simulated responses were then translated from subpopulation size in cells to equivalent BCR-ABL1/ABL1 levels using the equation:

$$Y = \frac{RT}{P}$$

where R is the predicted resistant population size, T is the final measured transcript level, and P is the final measured tumor burden.

For nearly all patients, the predicted molecular response lagged several hundred days behind the observed response. In fact, in many instances, spontaneously arising resistant clones failed to seed a clinically-detectable subpopulation at all. These results suggest that the counter hypothesis that there were no pre-existing resistant cells at the start of treatment is unlikely to be true.

PREDICTION OF MUTATIONAL EVENTS

Patient specific simulations

Each patient entering the PACE trial with a baseline mutation that is detectable by Sanger sequencing has a leukemia population size P in the peripheral blood that can be estimated because each patient had a known white blood cell (WBC/liter) count w, a starting disease allele frequency given by nested PCR from Sanger sequencing S, and an initial allele frequency

given by genomic DNA pull down and Duplex Sequencing *D.* Thus, the leukemia population size in the peripheral blood *P* is approximately $P = w\left(\frac{S}{100}\right)\left(\frac{D}{50}\right) * volume(Liters)$. Patient specific mutation burdens *B* are measured as the absolute count of unique mutations *m* (regardless of allele frequency) per genome (*g*) per bp sequenced (*b*) and is described as: $B = m/(gb)$. B is an estimate of the success rate of the discovery of new mutations as more and more unique genomes are sequenced. However, since m is rare (much less than 0.5), we modeled the distribution of the success probabilities using the beta distribution: B $\sim Beta(m, gb)$. We did this because the Clopper-Pearson Exact Confidence Interval can be derived from a beta distribution. Thus simulating B using the beta distribution gives a conservative estimate of the distribution of potential success rates of mutation discovery in an individual patient. The success rate of mutation discovery *B* and the population size *P* can be used to parameterize a binomial distribution: $Total\ Mutations \sim Bin(B; \widehat{m}, P)$. We then modeled the occurrence of mutational events $\widehat{m}$ as a Bernoulli process. We simulated the estimated total number of mutational events $\widehat{m}$ in a population size *P* using patient specific estimates of *m, g* and *b.* to initiate a simulation in R we used the rbeta() and rbinom() functions in the {stats} package Each unique mutational event $\dot{m}$ within the total number of estimated mutational events $\widehat{m}$ is then assigned randomly to a particular mutation at a particular codon position c_{1j}……c_{bj} where b is the total length the kinase domain that is measured by Duplex Sequencing. At each position c_i 3 different nucleotide substitutions denoted as *j* may take place. Thus, the total number of unique mutational events that a mutation can be assigned to is 3*b. The likelihood of each of the 12 nucleotide substitutions on the sense strand (A>C, T>G, etc.) is approximated from all of the Duplex Sequencing data across all of the patients sequenced. This gives us an estimate of

the biases of mutational processes within the ABL1 gene within Ph+ malignancies *in situ.* The number of each type of mutational event X_{ij} is then simulated according to a multinomial distribution:$\Pr([p_{ij} \ldots\ldots p_{bj}]; [X_{ij \ldots\ldots} X_{bj}], \hat{m})$. We recorded the absolute number of total mutational events, the number of predicted double mutants, the proportion of codon substitutions observed and the specific number of T315I, E255K, E255V and other mutations. These are supplied per patient per simulation and were used to generate the boxplots in figure 4. Notably, we simulate mutational events, not population sizes. Clearly, a lack of fitness at a particular residue, and the exact generation at which a given mutation occurs is necessary to simulate the population structure of resistance. We do however know that resistance mutations like T315I do not have large fitness disadvantages, so the number of mutational events at the T315I position will tend to underestimate the population size because it will not account for rare situations when the mutational event occurs early in clonal expansion. Probabilities are provided in Supplementary Table 6.

LEUKEMIA INITIATING CELL FRACTION

Leukemia initiating cell fractions can be experimentally defined by determining the number of patient cells needed for engraftment in immunodeficient mice. Injection of cells from human CP-CML patients into sub-lethally irradiated SCID mice requires a cell dose of >8E7 for efficient engraftment, while BP-CML cells engrafted at all tested levels (1-4x10^7 cells) (17). In a second study, >=5x10^7 CP-CML cells are needed for consistent engraftment into NOD/SCID mice (18), with almost no engraftment at a dose of 2x10^7 cells.

In contrast to the large number of CP-CML cells required for engraftment, John Dick and colleagues have reported that cells from Ph+ ALL patients engraft in immunodeficient mice with

high efficiency (19). Samples from 18 out of 20 patients studied engrafted with a cell dose of <=1x10^{6}. 10 of the patient samples led to rapidly progressive disease after transplantation; these samples engrafted all recipient mice tested, even when transplanted at near-limiting doses. Limiting dilution analysis performed across multiple patient samples was consistent with leukemia initiating cell fractions ranging from 1/10 to 1/1x10^{6}. Consistent with the high burden of leukemia initiating cells indicated by this study, in a murine model of Ph+ ALL, as few as 20 cells can produce a rapidly fatal disease within weeks (20).

The fraction of leukemia initiating cells can also be estimated by mathematical modeling. CP-CML is observed to have a biphasic response after initiation of therapy with an initial, relatively rapid drop in tumor burden by 2-5 orders of magnitude followed by a much slower decrease in tumor burden. Under the assumption that the initial phase turnover of differentiated cells while the second phase reflects turnover of leukemia initiating cells, Werner et al. (21) determined the number of leukemia-initiating cells in individual patients, which resulted in an estimate of 1 leukemia-initiating cell for every 10^{4}-10^{12} total cells. Ph+ ALL, in contrast, does not have a biphasic response curve, which implies a much higher burden of leukemia initiating cells.

PARAMETER SWEEPS FOR PH+ ALL VERSUS CP-CML

To examine the complete landscape of pre-existing mutational events, we used the pbinom() function in R to calculate low and high estimates of the confidence intervals for mutation burden across all patients in both disease type. Simulations were performed as in individual patients, except that we explored the full range of possible parameters that are relevant in the different diseases. To simulate the phenotypic heterogeneity, we accounted for the leukemia initiating fraction. So instead of *P* we adjust by the stem cell proportion to create a smaller

effective population size P_s . For Ph+ ALL we simulated mutation burdens between $2X10^{-7}$ and $3X10^{-6}$. Leukemia initiating fractions between 1/10 and $1/10^6$ and population sizes between 10^8 and 10^{11}. For CP-CML we simulated mutation burdens between $4X10^{-8}$ and $6x10^{-7}$. Leukemia initiating fractions between 1/1000 and $1/10^7$ and population sizes between 10^8 and 10^{11}. We assumed that mutation burdens, which were measured in the bulk population, are the same as those in the sub-population of cells which comprise the leukemic cell population. Simulation values are given as Supplementary Data 1.

To provide for a more quantitative comparison of the role that stem cells may play in resistance in Philadelphia positive malignancies, we sought to go beyond parameter exploration and do statistical tests for the role of the tumor initiating fraction across all mutation burdens measured in this study. To do this we decided on a single metric of initiating fraction, cells that are capable of re-seeding an immune-compromised mouse. This experiment has been performed by the Dick lab for both CP-CML and Ph+ ALL. We reasoned that a single experimental definition of initiating fraction performed by the same group would give a reasonable approximation of of initiating fraction in these two diseases. Thus, we performed simulations as described above, but instead of exploring all possible parameters, we sampled the tumor initiating cell fraction from an eCDF generated from the engraftment assays of Notta et al 2011 (for Ph+ ALL) and Lewis et al. 1998 (for CP-CML). Population size in the simulations was 10^{11} and mutation burden was that observed across CML and ALL. At all levels of mutation burden, we simulated 1000 individuals. We plotted the proportion of these individuals containing pre-existing T315I mutations in Figure 5C. We did this for both Ph+ ALL and CP-CML

eCDFs of initiating cell numbers. We then compared the difference in counts (simulations with a pre-existing T315I mutation at baseline) for the two disease types using a Fisher's Exact Test.

PREDICTION OF COMPOUND RESISTANCE

Putative ponatinib compound resistance mutations were taken from prior reports (22,23). The total number of compound mutations for any simulation parameter set is recorded during a simulation. This incidence of any compound mutation is then corrected by the probability that a compound mutation is resistant. The estimated probability of compound resistance utilizes the single site probabilities derived for equation 6 $[p_{ij} \ldots\ldots p_{bj}]$. The set of compound resistance mutations that were considered potential resistance mutations is I315M, T315I/E255V, T315I/E255K, T315I/M351T, T315I/G250E, Q252H/T315I, Y253H/T315I, F317L/T315I, Y253H/E255V, E255V/F317I. I315M is considered a compound mutation as two distinct nucleotide substitutions need to occur for this variant to arise. To calculate the total probability of compound resistance, individual mutations were assumed to be independent. Under this assumption, the probability of compound resistance is simply the product of the probability of the mutation at each position. The total probability of any compound resistance is then the sum of these individual compound resistance probabilities. $\sum_{k=1}^{L} p_i p_b$Where L is the total number of putative compound resistance mutations, and $p_i p_b$ represents the product of the probability of each of the 2 independent resistance alleles.

ESTIMATION OF THE PROBABILITY OF DISCOVERING T315I

We approximated the process as a pure birth discrete generation branching process. The number of cells N can be thought of as an effective number of generations. This effective

generation number is defined as the number of measurable productive generations. This is approximately:

$$\sim Log2(population\ size)$$

and it is bounded by the integer values above or below the result.

Thus, to estimate the cumulative probability of an individual detected mutation occurring at or below a certain detection threshold we use the equation:

$$C_{PT}=\sum_{i=1}^{log2(\frac{1}{Tolerance})} 2^{(i-1)} /(N-1)$$

1/Tolerance gives the effective population size when a mutation occurred. For instance, if the tolerance that is sequenced to is ~1 in 524,000 genomes, then the mutation must arise at or before the 19th effective generation. Thus, summing equation 9 from 1st generation (*i*) to the 19th generation (*i*) gives the cumulative probability of detecting a mutational event occurring at or before the 19th generation in a population of size *N*. Next, given the same population size *N* and the output of the simulations from equation 6 (parameterized for the average CP-CML patient 10^{11} cells and $3X10^{-7}$ mutation burden) we simulate the expected number of T315I mutational events that would exist in an average CP-CML leukemia (regardless of stem cell fraction) since the observed T315I mutation frequency is independent of cellular phenotype. This is approximately 20,000 T315I events. Next we model every T315I mutation as a binomial process where a success occurs if that mutation arises at or before the ith generation that defines the sequencing *Tolerance.* Thus, the probability of 0 T315I successes occurring before the threshold generation can be calculated from the binomial distribution by using C_{PT} as the probability of success in each of 20,000 independent mutation trials.

SENSITIVITY ANALYSIS

While we do not have mutation burden measurements in frontline Ph+ ALL patients in this study, it is difficult to hypothesize that they would have a substantially higher burden than treatment refractory Ph+ ALL patients, or less mutation burden than frontline CP-CML patients. Thus, we believe the present study can be interpolated to provide conservative biological upper and lower bounds for frontline Ph+ ALL mutation burden. To investigate the influence of mutation burden on our conclusions, we performed a sensitivity analysis by permuting CP-CML mutation burden measurements with Ph+ ALL stem cell fractions and vice versa. This result **(Supplementary Figure 7)**, suggests that the largest determinant of pre-existing resistance is the stem cell burden. In the presence of a CP-CML mutation burden, most Ph+ ALL simulation parameter sets still indicate that resistance mutations pre-exist. Thus, even without frontline Ph+ ALL data, if one uses CP-CML data as a conservative lower bound for mutation burden, we suggest that most plausible frontline Ph+ ALL parameter sets are likely to indicate pre-existing resistance.

Supplementary References

1. Cortes JE, Kantarjian HM, Goldberg SL, Powell BL, Giles FJ, Wetzler M, et al. High-dose imatinib in newly diagnosed chronic-phase chronic myeloid leukemia: high rates of rapid cytogenetic and molecular responses. Journal of Clinical Oncology. 2009;27:4754–9.

2. Baccarani M, Druker BJ, Branford S, Kim D-W, Pane F, Mongay L, et al. Long-term response to imatinib is not affected by the initial dose in patients with Philadelphia chromosome-positive chronic myeloid leukemia in chronic phase: final update from the Tyrosine Kinase Inhibitor Optimization and Selectivity (TOPS) study. International Journal of Hematology. 2014;99:616–24.

3. Baccarani M, Deininger MW, Rosti G, Hochhaus A, Soverini S, Apperley JF, et al. European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. Blood. 2013;122:872–84.

4. Iacobucci I, Galletti L, Amabile M, Soverini S, Baccarani M, Martinelli G. Monitoring BCR-ABL transcript levels by real-time quantitative polymerase chain reaction: a linear regression equation to convert from BCR-ABL/B2M ratio to estimated BCR-ABL/ABL ratio. Haematologica. 2007;92:429–30.

5. Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by Duplex Sequencing. Nature Protocols. 2014;9:2586–606.

6. Schmitt MW, Fox EJ, Prindle MJ, Reid-Bayliss KS, True LD, Radich JP, et al. Sequencing small genomic targets with high efficiency and extreme accuracy. Nature Methods. 2015;12:423–5.

7. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature Biotechnology. 2011;29:24–6.

8. Egan DN, Beppu L, Radich JP. Patients with Philadelphia-positive leukemia with BCR-ABL kinase mutations before allogeneic transplantation predominantly relapse with the same mutation. Biology of Blood and Marrow Transplantation. 2015;21:184–9.

9. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. Proceedings of the National Academy of Sciences U S A. 2012;109:14508–13.

10. Deininger MW, Hodgson JG, Shah NP, Cortes JE, Kim D-W, Nicolini FE, et al. Compound mutations in BCR-ABL1 are not major drivers of primary or secondary resistance to ponatinib in CP-CML patients. Blood. 2016;127:703–12.

11. Saldanha J, Silvy M, Beaufils N, Arlinghaus R, Barbany G, Branford S, et al. Characterization of a reference material for BCR-ABL (M-BCR) mRNA quantitation by real-time amplification assays: towards new standards for gene expression measurements. Leukemia. 2007;21:1481–7.

12. Stein AM, Bottino D, Modur V, Branford S, Kaeda J, Goldman JM, et al. BCR-ABL transcript dynamics support the hypothesis that leukemic stem cells are reduced during imatinib treatment. Clinical Cancer Research. 2011;17:6812–21.

13. Michor F, Hughes TP, Iwasa Y, Branford S, Shah NP, Sawyers CL, et al. Dynamics of chronic myeloid leukaemia. Nature. 2005;435:1267–70.

14. Cao Y, Gillespie DT, Petzold LR. Avoiding negative populations in explicit Poisson tau-leaping. The Journal of Chemical Physics. 2005;123:054104.

15. Cao Y, Gillespie DT, Petzold LR. Adaptive explicit-implicit tau-leaping method with automatic tau selection. The Journal of Chemical Physics. 2007;126:224101.

16. Bozic I, Reiter JG, Allen B, Antal T, Chatterjee K, Shah P, et al. Evolutionary dynamics of cancer in response to targeted combination therapy. Elife. 2013;2:e00747.

17. Sirard C, Lapidot T, Vormoor J, Cashman JD, Doedens M, Murdoch B, et al. Normal and leukemic SCID-repopulating cells (SRC) coexist in the bone marrow and peripheral blood from CML patients in chronic phase, whereas leukemic SRC are detected in blast crisis. Blood. 1996;87:1539–48.

18. Lewis ID, McDiarmid LA, Samels LM, To LB, Hughes TP. Establishment of a reproducible model of chronic-phase chronic myeloid leukemia in NOD/SCID mice using blood-derived mononuclear or CD34+ cells. Blood. 1998;91:630–40.

19. Notta F, Mullighan CG, Wang JCY, Poeppl A, Doulatov S, Phillips LA, et al. Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. Nature. 2011;469:362–7.

20. Williams RT, Besten den W, Sherr CJ. Cytokine-dependent imatinib resistance in mouse BCR-ABL+, Arf-null lymphoblastic leukemia. Genes & Development. 2007;21:2283–7.

21. Werner B, Scott JG, Sottoriva A, Anderson ARA, Traulsen A, Altrock PM. The Cancer Stem Cell Fraction in Hierarchically Organized Tumors Can Be Estimated Using Mathematical Modeling and Patient-Specific Treatment Trajectories. Cancer Research. 2016;76:1705–13.

22. Gozgit JM, Schrock A, Chen T-H, Clackson T, Rivera VM. Comprehensive Analysis of the in vitro potency of ponatinib, and all other approved BCR-ABL tyrosine kinase inhibitors (TKIs), against a panel of single and compound BCR-ABL mutants [Abstract]. Blood. 2013;122:3992.

23. Zabriskie MS, Eide CA, Tantravahi SK, Vellore NA, Estrada J, Nicolini FE, et al. BCR-ABL1 compound mutations combining key kinase domain positions confer clinical resistance to ponatinib in Ph chromosome-positive leukemia. Cancer Cell. 2014;26:428–42.