Phenotypic deconvolution in heterogeneous cancer cell populations using drug screening data

A. Köhn-Luque[*,1], E. M. Myklebust[*,1], D. S. Tadele[†,2,3,4], M. Giliberto[5,6], J. Noory[7], E. Harivel[1,8], P. Arsenteva[1,9], S.M. Mumenthaler[10], F. Schjesvold[6,11], K. Taskén[5,6], J. M. Enserink[2,3,12], K. Leder[‡,13], A. Frigessi[‡,1, 14], and J. Foo[‡,7]

[1] *Oslo Centre for Biostatistics and Epidemiology, Faculty of Medicine, University of Oslo, Oslo, Norway*
[2] *Department of Molecular Cell Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway*
[3] *Centre for Cancer Cell Reprogramming, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway*
[4] *Translational Hematology and Oncology Research, Cleveland Clinic, Cleveland, USA*
[5] *Dept. of Cancer Immunology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway*
[6] *KG Jebsen Center for B-Cell Malignancies, Institute for Clinical Medicine, University of Oslo, Oslo, Norway*
[7] *Institute for Mathematics and its Applications, School of Mathematics, University of Minnesota, Minnesota, USA*
[8] *ENSTA, Institut Polytechnique de Paris, Paris, France*
[9] *Institut de Matématiques de Bourgogne, Universite de Bourgogne, Dijon, France*
[10] *University of Southern California (Lawrence J. Ellison Institute for Transformative Medicine of USC), Los Angeles, USA*
[11] *Oslo Myeloma Center, Department of Hematology, Oslo University Hospital, Oslo, Norway*
[12] *Section for Biochemistry and Molecular Biology, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway*
[13] *College of Science and Engineering, University of Minnesota, Minnesota, USA*
[14] *Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway*

January 17, 2022

[*] These authors contributed equally to this work.
[†] present address: Department of Medical Genetics, Oslo University Hospital, Oslo, Norway
[‡] These authors contributed equally to this work.

Abstract

Tumor heterogeneity is an important driver of tumor recurrence, as treatments that initially elicit clinical responses can select for drug-tolerant tumor subpopulations, leading to the outgrowth of resistant clones and cancer treatment failure. Profiling the drug-response heterogeneity of tumor samples using traditional genomic deconvolution methods has yielded limited results, due in part to the imperfect mapping between genomic variation and functional characteristics. Here, by introducing an underlying population dynamic model of tumor subclonal response to therapy, we enable the phenotypic deconvolution of bulk drug-response data into component subpopulations, and the estimation of their differential drug sensitivities and population frequencies. We used this method, called DECIPHER, to perform deconvolution on tumor drug screening data generated both experimentally and *in silico*. This study demonstrates how mechanistic population modeling can be leveraged to develop statistical frameworks for profiling phenotypic heterogeneity from bulk tumor samples and to perform individualized patient treatment predictions.

Introduction

Most human tumors display a striking amount of phenotypic heterogeneity in features such as gene expression, morphology, metabolism, and drug response. This diversity fuels tumor evolution and adaptation, and it has been correlated with higher risks of treatment failure and tumor progression [1, 2, 3, 4, 5, 6, 7, 8]. Indeed, treatments that initially elicit clinical response can select for drug-tolerant tumor subpopulations, leading to outgrowth of resistant clones and tumor recurrence. Additionally, the heterogeneity and composition of tumors is known to vary widely between patients, underscoring the need for more personalized approaches to cancer therapy that profile and address intra-tumor heterogeneity and its evolutionary consequences. Towards this goal, recent advances in single-cell genomic profiling of tumor samples have enabled the assessment of the genetic variability within tumor cell populations. However, single-cell technologies are often limited by large measurement errors, incomplete coverage, and small sample availability, which leads to challenges in capturing the temporal dynamics crucial for understanding response to therapies. Furthermore, the mapping between genotypic and phenotypic variation is far from perfect: not all variation in cellular drug response can be explained by genetic mechanisms, and divergent genetic profiles can lead to similar treatment responses [9, 10].

Another important approach to designing individualized treatment strategies is personalized drug sensitivity screening, a procedure in which patient tumor samples are tested for functional responsiveness to a library of drugs using high throughput *in vitro* drug sensitivity assays. In these assays, cells are treated with various concentrations of a drug and the number of viable cells is measured at one or more fixed time points. The resulting data are normalized and fitted to produce viability curves, whose summary characteristics (e.g. IC_{50}, EC_{50}, AUC) are used to compare drug sensitivity across multiple drugs and/or

cell populations [11, 12, 13, 14, 15]. Increasingly, such drug screens are used as a tool in personalized medicine to evaluate and rank the potential efficacy of therapeutic agents on a patient's disease cell population. However, the interpretation of these cell viability curves and associated metrics are confounded by the presence of cellular heterogeneity within the population. In particular, the presence of multiple subclones with divergent drug response characteristics may result in an intermediate drug sensitivity profile that does not accurately represent any individual cell type within the population [16]. Developing techniques to detect the presence of subpopulations with distinct drug sensitivity profiles is crucial for achieving effective treatment strategies.

In this work, we develop a novel methodology for detecting the presence of cellular subpopulations with differential drug responses, using standard bulk cell viability assessment data from drug screens. Our method, DECIPHER: DEConvolution Into PHenotypes from cancEr drug scReens, detects the presence and composition fractions of distinct phenotypic components in the tumor sample and quantifies the sensitivity of each subpopulation to a specific mono-therapy. It utilizes statistical tools in combination with a novel underlying population dynamic model describing the evolution of a heterogeneous mixture of tumor cells with differential drug sensitivity over time. We validate DECIPHER using simulated tumor drug screening data as well as measurements of drug response in known mixture experiments of cancer cell lines. We then use this method to profile the population drug response heterogeneity in multiple myeloma patient samples, and we demonstrate how these results can be used to produce personalized predictions of tumor response to therapy. This methodology can be applied across cancer types and therapies to characterize the drug-response heterogeneity within tumors.

Results

DECIPHER method

Given a set of experimental drug-screen data on a bulk tumor sample, DECIPHER solves a series of optimization problems to identify individual subpopulations within the sample and estimate their frequencies and drug sensitivities. This problem is challenging because it requires simultaneous characterization of the number of individual subpopulations present, their frequencies in the population and their drug response characteristics, all based on noisy observations of the total cell population. Our solution to this problem is enabled by the introduction of a mixture population dynamic model of the tumor in which the growth rate dependence on drug concentration follows a Hill-type functional form (see equation (3)).

Figure 1 provides an overview of the DECIPHER workflow. First, a tumor sample is extracted, divided, and exposed to a panel of therapeutic compounds at a range of concentrations. For each drug, the population size counts are measured at a series of time points for each concentration and replicate. This data is then used as the input to

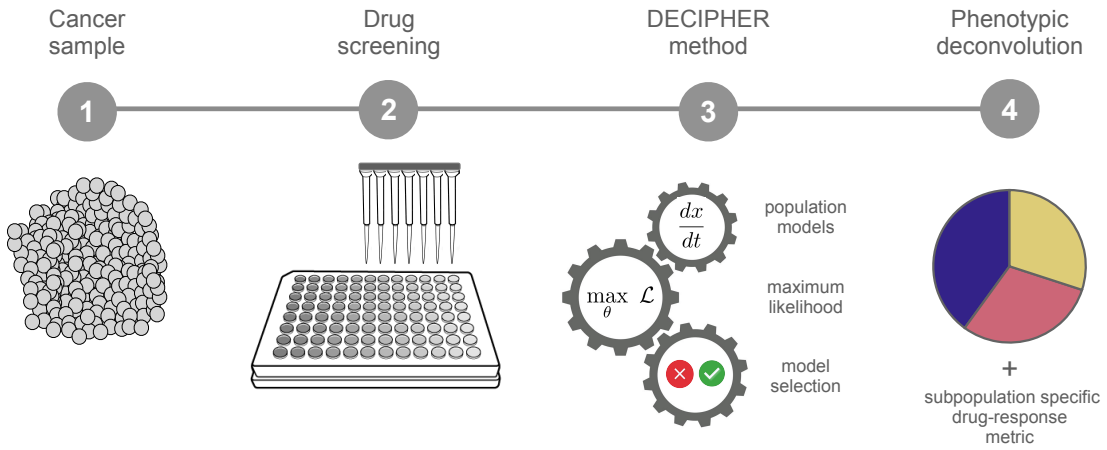

Figure 1: The DECIPHER workflow: 1) A cancer sample is taken from a patient. 2) Drug screening is performed. 3) DECIPHER method is applied to the data. 4) Phenotypic deconvolution is performed, with mixture fractions and drug-response metrics for each subpopulation as the result.

DECIPHER, which uses a maximum likelihood estimation procedure for estimating the parameters of the underlying population dynamic model for each candidate number of subpopulations. Then, a model selection process is performed to identify the number of subpopulations present and to estimate the mixture fractions and drug sensitivities of each subpopulation. Details are provided in the Online Methods.

Validation in synthetic populations

To quantify the performance of DECIPHER in mixtures of 1, 2, and 3 populations, three synthetic populations were designed to have drug-response properties similar to cell lines observed in in-vitro experiments. Using the model for data generation described in the online methods (section *Generation of synthetic population data*), synthetic data was generated for 9 different mixture compositions of the three populations. The synthetic mixtures were exposed to 17 concentrations of the simulated drug, and the bulk cell populations were measured at 9 equidistant points in time. The simulated drug concentrations were chosen to cover the entire range where the growth rates of the populations were affected by changes in the drug concentration. To simulate measurement error, random noise was added to each bulk cell count. Data from 4 replicates of the experiment were used to perform the inference.

To measure drug sensitivity, DECIPHER utilizes the growth rate-associated metric GR_{50}, introduced in [17] and defined as the concentration at which the population growth rate is reduced by half of the maximum observed effect, as it provides a robust metric for comparing drug-response across cellular subpopulations (online methods section *Calculation of GR_{50} values*). To assess the accuracy of the DECIPHER deconvolution analysis, (i) the true mixture fractions were compared with the deconvolution mixture fraction estimates, and (ii) the GR50 obtained from the deconvolution analyses was compared with the true GR50 region. The use of a region, or range of values, for the true GR50 reflects the inherent limitation from sampling discrete concentrations in experimental data; it is only possible to ascertain that the GR50 is somewhere between the closest two sampled concentration levels, and the finer the sampling resolution, the smaller the range of uncertainty.

Figure 2a shows true mixture compositions and GR50 values compared to DECIPHER's estimates for the 9 cases, in an experiment where the noise terms were sampled independently from a Gaussian distribution with mean 0 and base noise level of 5 % (standard deviation of the mean initial cell count). Additional sensitivity tests evaluating DECIPHER performance on synthetic data with varying noise levels (up to 50 %) are discussed in the section *DECIPHER-recommended experimental design and limitations* and data are provided in the Supplementary Information. To place these noise levels in the context of expected noise levels from experimental drug screen data, the standard deviation to mean ratio reported from several common automated or semi-automated cell counting techniques ranges from 1-15 % [18, 19]. For example, counts obtained via a trypan blue

exclusion-based Vi-CELL® XR Cell Viability Analyzer (Beckman Coulter) had noise levels consistently less than 5.3 % across several cell lines [18], while those obtained via a Countess®Automated Cell Counter (Invitrogen) fell in the range 11-14.3 %. Cells counts obtained using the Cellomics ArrayScan high content screening platform in another set of experiments (used in this work) had standard deviation to mean ratios of 1-5.6 % [19].

Figure 2a demonstrates that DECIPHER inferred the mixture fractions within 2 percentage points for mixtures of 1, 2, and 3 populations at the 5 % noise level. The GR50 values were inferred precisely within the true GR50 region for all mixtures of 1 and 2 populations, and also for an equal mixture of 3 populations. In the case with 3 populations in a 40:30:30 mixture, one of the estimated GR50 values is off by 1 GR50 region, and in the 3-population mixture with a 60:20:20 mixture, all three estimated GR50 values are off by 1 GR50 region.

Validation with cell line experiments

Next, to investigate the performance of our method in the experimental setting, mixtures of cell populations with differential drug sensitivity were constructed and subjected to drug screen experiments. The resulting bulk cell population readings at varying drug concentrations, time points, and replicates were used as inputs to DECIPHER.

Imatinib-sensitive and -resistant Ba/F3 cells. We tested monoclonal and mixture populations of isogenic Ba/F3 murine cell lines that were stably transformed with either the wild-type BCR-ABL fusion oncogene or with BCR-ABL-T315I, which contains a point mutation that confers increased resistance to the Abl tyrosine kinase inhibitor imatinib. Note that expression of these oncogenes renders cells addicted to BCR-ABL activity [20]. Monopopulations and mixtures of these two cell lines were treated with 11 different concentrations of imatinib, and the bulk cell population sizes were quantified at 14 time points. Using this bulk population data, DECIPHER was able to correctly assess the number of component subpopulations (see Supplementary Figure S17). As shown in Figure 2b, DECIPHER also estimated the fraction of the population belonging to each subcomponent at the start of the drug screen as well as the drug sensitivity (GR50) of each subpopulation; the estimates demonstrated good agreement with the known mixture proportions and independently assessed GR50 ranges of the monoclonal T315I+/- populations.

Erlotinib-sensitive and -resistant NSCLC cells. Additionally, two EGFR-mutant non-small cell lung cancer (NSCLC) lines, HCC827 and H1975, were considered for their differential sensitivity to the drug compound erlotinib. The mutation T790M, which is present in H1975 cells but not in HCC827 cells, confers increased resistance to erlotinib. Monopopulations and mixtures of the erlotinib-sensitive and -resistant NSCLC cell lines were treated with four drug concentrations and total cell population count was assessed at 0, 24, 48, and 72 hours with four replicates [19]. Figure 2c demonstrates DECIPHER's results on this bulk data. DECIPHER was able to correctly assess when populations were monoclonal, as well as to detect the presence of two populations in the bulk drug response data from mixed

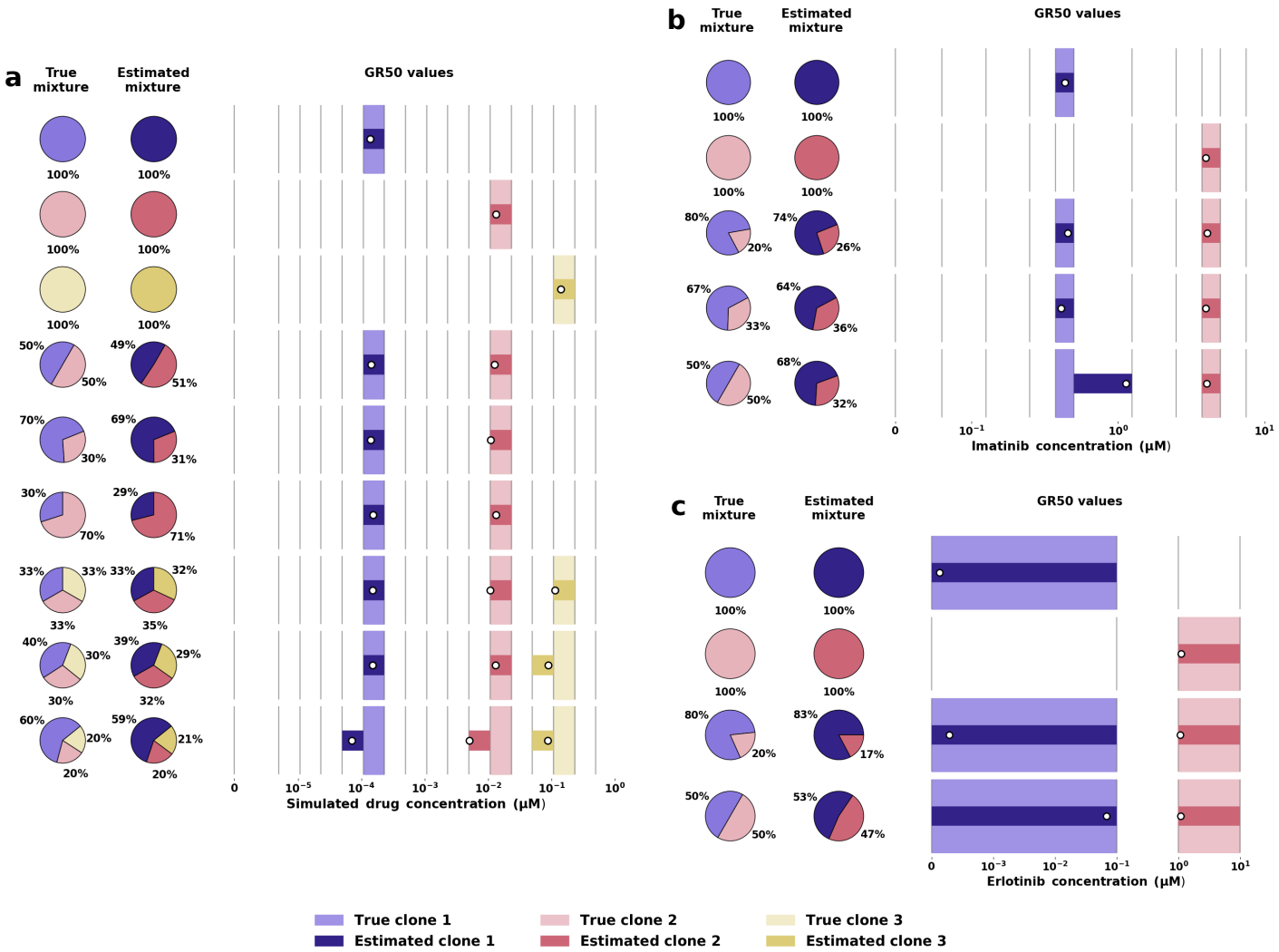

Figure 2: **DECIPHER validation on simulated data and cell-line derived experimental data.** **a**, True and estimated mixture fractions and GR50 values for synthetic data. For each row, the "True mixture" pie chart shows the mixture fractions used to generate the data; the "Estimated mixture" pie chart shows the mixture fractions estimated by DECIPHER; in the "GR50 values" panel, the vertical grey lines show the sampled drug concentrations; the true GR50 region of each subpopulation has been marked by coloring the region between the two adjacent observed concentrations in the color belonging to that population (see main text for more details); for each estimated population, the estimated GR50 is marked by a white dot, with the region between the adjacent observed concentrations colored in the estimated clone's color. **b**, True and estimated mixture fractions and GR50 values for Ba/F3 murine cell line data. **c**, True and estimated mixture fractions and GR50 values for NSCLC cell data.

populations. Furthermore, using the bulk mixture response data, DECIPHER accurately estimated the mixture fractions and GR50 values of each component subpopulation. The reference GR50 ranges were independently assessed on monoclonal HCC827 and H1975 cell populations.

Deconvolution analysis of Multiple Myeloma patient samples

Next, DECIPHER was used in a clinical scenario to deconvolve twenty drug sensitivity screens performed on five Multiple Myeloma (MM) patient samples. MM is a clonal B-cell malignancy characterized by abnormal proliferation of plasma cells in the bone marrow. The median survival time of MM patients is about 6 years, with a disease course typically marked by multiple recurrent episodes of remission and relapse [21]. Drug responses and relapses are currently unpredictable, largely due to unknown complex clonal compositions and dynamics under treatment [22, 23].

Bone marrow samples were taken from each patient, processed, and screened with a set of MM clinically-relevant drugs, as illustrated in Figure 3a and described in online methods [24]. To perform the drug screens, samples from each patient were subjected to treatment at varying concentrations with a subset of the following drugs: Dexamethasone, Ixazomib, Melflufen, Selinexor, Thalidomide, and Venetoclax. We note that screening data for all drugs for each patient was not available; Figure 3b-g shows the set of patient samples treated by each drug and summarizes the results of DECIPHER deconvolution analysis on each set of drug screen data.

Inter-patient similarities in subpopulation GR50s. In all cases DECIPHER identified either one or two subpopulations; details of model selection results are shown in supplementary Figure S16. For example, Figure 3b shows that for patient MM1420, DECIPHER estimates that 87 % of the cells are resistant to Dexamethasone. This matches the clinically observed response, as the patient was refractory to Dexamethasone treatment in vivo. Interestingly, for all drugs used except Dexamethasone, the inferred subpopulations across patient samples share comparable GR50 values, although the proportions of these subpopulations may vary between patients (see Figure 3c-g). For example, for three patient samples treated with Venetoclax, DECIPHER inferred one more-sensitive and one more-resistant population (Figure 3e). However, the estimated proportions of the more-resistant populations (shown in the plot by the right-pointing arrows) varied from 23% up to 58%. We hypothesized that subpopulations with similar GR50s across patients may in some cases be driven by similar genetic alterations. To investigate this, we also characterized the samples with inferred heterogeneous compositions for the presence of high-risk genomic abnormalities, including Gain(1q21) (2/3) and several mutations co-existing in the same screened sample (MM36). Interestingly, we noticed that the proportion of MM cells from two samples (MM1420 and MM195) harboring the aberration gain (1q21) (approximately 50 %) was similar to the DECIPHER-inferred mixture fractions for the more-resistant clone in the same two samples (50% and 58%, respectively). This supports our hypothesis that

these subclones, which have similar levels of drug tolerance in different patients, may be driven the same alterations, and it is consistent with previous findings showing Gain(1q21) as negative predictor for Venetoclax efficacy in MM. This analysis provides genetic evidence that supports DECIPHER's ability to profile phenotypic drug response heterogeneity.

Treatment response prediction using DECIPHER estimates. The utility of these phenotypic deconvolutions as initial states for predicting and optimizing patient-specific treatment schedules remains to be systematically explored. Here, as a proof of concept, we present a mathematical model to illustrate how to use the DECIPHER estimates of population frequencies and differential drug sensitivities to predict the treatment outcome for the three patients exposed to Venetoclax. For easier comparison, we assumed that all three patients start with a total of 10^{12} abnormal plasma cells. Figure 4 demonstrates how the same treatment dose, 2 μM of Venetoclax, assumed constant over the simulation for simplicity, leads to highly disparate treatment outcomes in patients with distinct phenotypic heterogeneity profiles uncovered by DECIPHER. In particular, we note that to observe the predicted relapse in patient MM36, simulations have to be run for a much longer time (3000 days) than for the other two patients. See the Online Methods section for a description of the used model and its parameterisation.

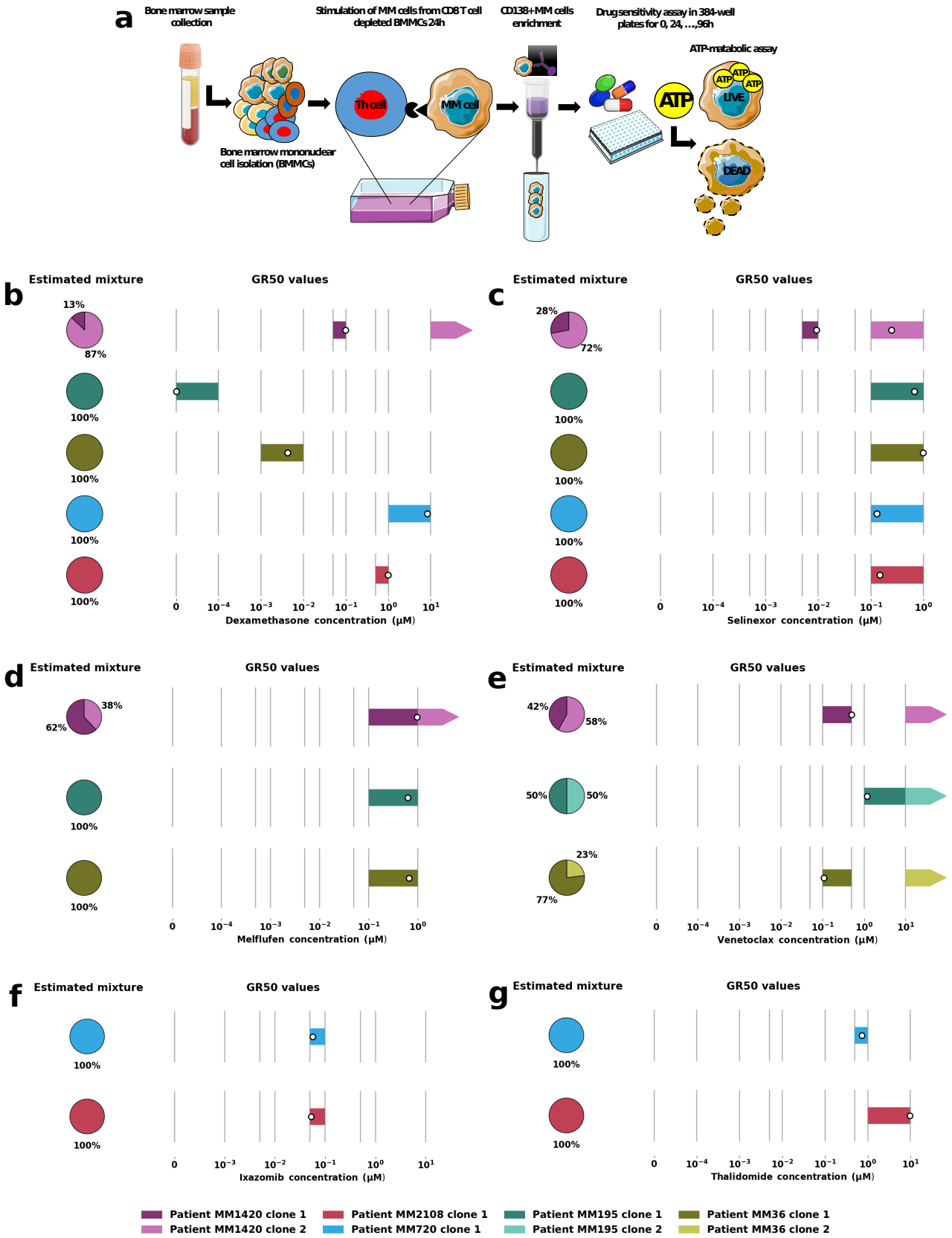

Figure 3: Phenotypic deconvolution of drug screens from MM patient samples. **a**, Illustration of the experimental protocol described in online methods [24]. Illustration created by the authors using smart.servier.com and biorender.com. **b-g**, Inferred mixtures and GR50 values for 5 MM samples with respect to 5 drugs: b) Dexamethasone, c) Selinexor, d) Melflufen, e) Venetoclax, f) Ixazomib, g) Thalidomide. For each row, the "Estimated mixture" pie chart shows the mixture fractions estimated by DECIPHER; In the "GR50 values" panel, the vertical grey lines show the observed drug concentrations; For each estimated population, the estimated GR50 is marked by a white dot, with the region between the adjacent observed concentrations colored in the estimated clone's color. If the inferred GR50 value of a population was higher than the highest observed concentration value, the estimated GR50 is instead marked by an arrow pointing towards the right from the highest observed concentration.

DECIPHER-recommended experimental design and limitations

We next performed a computational study using synthetic drug screen data to identify experimental design strategies that enhance DECIPHER accuracy, and to explore the limitations of the method.

Prioritization of experimental efforts: increasing the number of drug concentrations, time points, or replicates? We first considered the relative importance of experimental resolution in drug concentration, time points, and replicates in DECIPHER performance. Figure S2 shows the average gain in accuracy for a mixture of 2 populations (one sensitive, one resistant) when either the replicates R, the number of concentrations N_c or the number of time points N_t are increased while the others are held constant at the value three. To compare the accuracies, 27 two-sided t-tests were made, since 3 effects (increasing R, N_c, and N_t) were compared pairwise at 3 sample sizes (5, 9 and 17), in 3 different comparison measures. To account for multiple testing, the family-wise error rate was controlled to be below 0.05 using Bonferroni correction.

We find that for accuracy in the mixture parameter, increasing the number of concentrations or time points gives significantly higher precision than increasing the number of replicates to the same amount. Similarly, to enhance accuracy in the GR50 value of the sensitive population, increasing either the number of concentrations or number of time points gives significantly higher precision compared to increasing the number of replicates by the same number. In addition, increasing the number of concentrations to 9 or 17 is significantly better than increasing the number of time points similarly. No significant differences were found for estimating the GR50 of the resistant population.

Noise level. We also studied how increasing levels of measurement noise in the data (e.g. in cell counting) impact the precision of the deconvolution results. Results of these tests are shown in Supplementary figures S3, S4, S5 and S6, where the same synthetic data with increasing levels of measurement noise were used as inputs to DECIPHER. We found that for noise levels up to a standard deviation equal to 20% of the initial cell count, DECIPHER is able to correctly deconvolve the bulk response signal into the correct components. Beyond this noise level mixture fractions are off by more than 10% in 2-population mixtures, and populations may go undetected in 3-population mixtures. Supplementary figures S9, S10, S11, S12, and S13 (corresponding to Figure 2a and Supplementary figures S3, S4, S5 and S6) show how model selection was performed in these cases.

Small mixture fractions. To determine how small population fractions DECIPHER is able to detect, inference was performed on simulated data with a range of small mixture fractions, with a noise level of 5% of the initial cell count. We found that in 2-population mixtures, DECIPHER was able to detect populations at frequencies as low as 1 percent. In 3-population mixtures, DECIPHER was able to detect populations with mixture fractions of 3 percent and higher. At noise level of 5%, the estimated mixture parameters were within 1% of the true value and the estimated GR50 values were always within two GR50 regions of the true value. Supplementary Figure S7 shows these results. The figure also

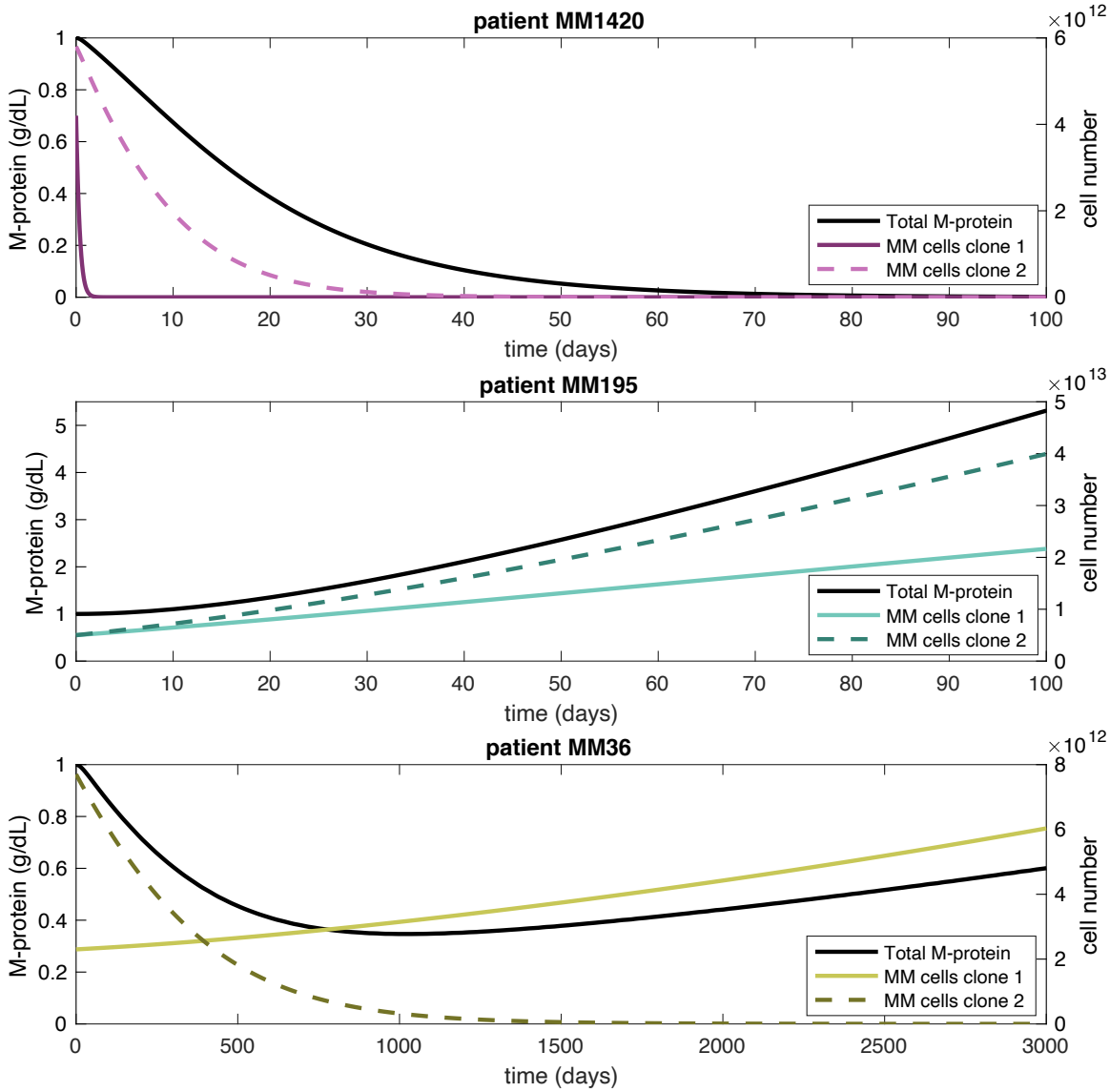

Figure 4: Proof-of-concept modeling of Multiple Myeloma disease dynamics under Venetoclax treatment for three patients using DECIPHER deconvolution results. The estimated mixture and drug-response parameters obtained by DECIPHER (see Figure 3e) define the initial percentage of cells and drug-response for each clone and patient. Cells from both clones are assumed to produce monoclonal protein (M-protein), which can be used as a proxy for tumor burden. For easier comparison, we assume that all three patients start with a total of 10^{12} abnormal plasma cells (cell number shown in the right y-axes) and $1g/dL$ M-protein (shown in the left y-axes). All three patients are exposed to 2 μM of Venetoclax. See Online Methods for description of mathematical model.

shows that it is harder to detect two small populations mixed with a large population (bottom row), than it is to infer one small population mixed with two larger ones (fifth row). Supplementary Figure S14 shows how model selection performed for these cases.

Subpopulation similarity. We performed computational experiments to determine the degree of similarity between component subpopulations beyond which DECIPHER was unable to detect distinct populations. We tested a set of 2 similar mixed populations, at a noise level of 5% of the initial cell count. We found that DECIPHER was able to detect populations whose GR50 values were as close as 2 GR50 regions apart. For such close populations, the estimate of the mixture parameters were within 2% of the true value and the estimated GR50 values were within 1 GR50 region of the true value, even for mixtures as unbalanced as 90:10. The results are shown in Figure S8. The figure's third, sixth and ninth rows show that the inferred GR50 values may overlap or swap position if the true GR50 values are less than 2 GR50 regions apart. The figure's eighth row shows that for mixtures of 5% or smaller, the inferred GR50 values can overlap even when the true GR50 values are 2 GR50 regions apart. Figure S15 shows how model selection performed for these cases.

Discussion

Understanding the phenotypic heterogeneity of human tumors, especially in terms of drug response, is essential in treatment planning and prognosis prediction. The optimization of treatment regimens is a long-standing area of research in the mathematical oncology community [25, 26, 27, 28]; however, the initial state of the tumor, which strongly influences optimal treatment strategies, is typically unknown. The DECIPHER method enables the detection of tumor subclonal populations, as well as estimation of their frequencies and drug sensitivities. The resulting deconvolved tumor profile can be fed, as an initial state, into mathematical models of tumor dynamics to predict treatment response (see Figure 4) and identify optimal treatment regimens.

Although the mathematical structure of the phenotypic deconvolution problem bears a resemblance to classical clustering based on observing individual responses, a vital difference is that in our setting the observed data is a combined signal from the entire population with unknown mixture frequencies and components. This statistical problem is also similar to the problem considered in blind source separation in digital signal processing, in which one attempts to recover individual source components from a mixture of signals (see e.g., [29]). However, a key assumption in this classic problem is the independence of the constituent components, an restriction that is not needed for DECIPHER. Interaction between individual populations, e.g. due to resource limitation or phenotypic switching, can be incorporated within the DECIPHER framework (see Supplementary Information section entitled Model Extension to Interacting Populations). The mathematical structure used in DECIPHER can also be applied to perform deconvolution analyses for cellular

response to many other external stimuli, such as intercellular signaling, the environmental pH level, mechanical forces and many others. To achieve this, the underlying population dynamic model of drug response used in DECIPHER can be replaced with another mechanistic or machine-learning derived model describing response to other stimuli. DECIPHER produces a heterogeneity profile tailored to each patient sample for each drug in a drug-screen panel. While this information is useful for identifying successful single-agent therapies and for optimizing or designing their therapeutic schedules, combination therapy design requires joint deconvolution analyses that elucidate the mapping between heterogeneity profiles for multiple drugs. This task will necessitate additional data from combination drug screens, and further methodological development in experimental design to identify tractable subsets of combination screening experiments that are necessary for identifying these joint deconvolution profiles. We plan to address this problem in future work.

The precision of DECIPHER depends on the amount of observation noise in the data. For the exponential growth model, normally distributed noise with a standard deviation of up to 20% of the initial cell count can be tolerated, while higher noise levels lead to errors in model selection and decreased accuracy in mixture fractions and GR50 estimates. This is especially seen in the 3-population mixtures, and it is expected that the problem would be aggravated in mixtures of more than 3 populations. We note that the standard deviation to mean ratio reported from several of the most common automated or semi-automated cell counting techniques ranges from 1-15 % [18, 19].

At moderate noise levels (standard deviation to mean ratio of 5%), DECIPHER was able to detect subpopulations as small as 1% of the total population in 2-population mixtures, while in 3-population mixtures the smallest detectable population fraction was 3 %. The precision is reduced when subpopulations have very similar GR50 values and the resolution of experimental drug concentrations does not distinguish well between them, but for predicting treatment response, distinguishing subpopulations that are almost identical is of limited clinical importance. Additionally, our study suggests that in terms of data resolution and prioritization of experimental effort, increasing the number of observed concentrations improves accuracy the most, followed by the number of time points, and then the number of replicates.

In this work we have demonstrated that DECIPHER can provide vital insights into the diversity of drug response amongst tumor cells. This framework, enabled by mixture population dynamic modeling of response to therapy, utilizes bulk drug screen data and alleviates the need for costly single-cell methods in profiling tumor heterogeneity. Although we focus here on drug-response heterogeneity, the DECIPHER framework can also be applied to detect and profile heterogeneous response to general environmental stimuli in cell populations, such as stromal content, nutrient/oxygen deprivation, and epigenetic modifiers. Accurate, efficient techniques for profiling of heterogeneity across multiple axes are important foundations for personalized treatment decision-making.

Online Methods

Model of dose-dependent population dynamics

DECIPHER relies upon an underlying model of heterogeneous tumor population dynamics *in vitro*. The growth of a single population of cells with homogeneous drug response is modelled by

$$X(t,d) = X(0) * \exp\left[t\left(\alpha + \log H(d)\right)\right], \tag{1}$$

where $X(t,d)$ is the number of cells at time t under drug concentration d, $X(0)$ is the initial population size, α is the intrinsic growth rate of the population in the absence of drug, and $H(d)$ is a classic sigmoidal function describing the dependence of the population growth rate on drug concentration d:

$$H(d) = b + \frac{1-b}{1+\left(\frac{d}{E}\right)^n}. \tag{2}$$

The parameters of this function control the shape of the sigmoid: $b \geq 0$ reflects the maximum effect of the drug, E is the log concentration at which 50 percent of the maximum effect is achieved, and $n > 0$ controls the steepness of the response. This novel form of the growth rate, $r(d) \equiv \alpha + \log H(d)$, is chosen so that the predicted cell viability curve, which is the treated viable cell population size normalized by the untreated viable cell population size at a fixed time, exhibits the standard Hill-shaped dependence on drug concentration that is empirically observed in viability assays [16]. Supplementary Figure S1 demonstrates that this model accurately recapitulates experimental cell viability dependence on drug concentration in two BCR-ABL positive Ba/F3 cell lines (with and without the T315I mutation) treated with the tyrosine kinase inhibitor imatinib. Note that since we are studying *in vitro* populations prior to confluence, an exponential growth model is appropriate.

To extend the monoclonal growth model (1) to a population composed of several subpopulations, each with a specific own drug response dynamics, we denote the growth parameters of the i-th subpopulation by α_i, b_i, E_i, n_i. Then the model of a cell population with S subpopulations under drug concentration d at time t is:

$$Z(t,d) = Z(0)\sum_{i=1}^{S} \pi_i \exp\left[t(\alpha_i + \log H(d; b_i, E_i, n_i))\right] \tag{3}$$

where $Z(0)$ is the total initial population and π_i is the initial mixture fraction of the i-th subpopulation ($\sum_{i=1}^{S} \pi_i = 1$). Here the parameters of the Hill function $H(d; b_i, E_i, n_i)$ are explicitly noted to emphasize the individual drug response profile for each subpopulation. Under this formulation, we need to estimate, on the basis of the drug screen data, the unknown parameters: S, and $\{(\pi_i, \alpha_i, b_i, E_i, n_i); i \in \{1, \ldots, S\}\}$. Note that in this case,

the heterogeneous population is modelled as a mixture of populations in which individual subpopulations are assumed to grow independently. In the Supplementary Information, we consider a case in which interaction between subpopulations is incorporated.

Estimation procedure

As input, DECIPHER takes bulk tumor sample drug screening observations, in the form of total cell counts at a series of time points and drug concentrations. A variety of experimental techniques is commonly used to generate such observations of cell population counts in drug screening. For example, tetrazolium reduction assays (e.g. MTT, MTS), protease viability markers (e.g. GF-AFC), ATP assays (e.g. Cell Titer-Glo), and more recently developed real-time assays (e.g. Real-Time Glo, live-cell imaging) [30, 31]. The DECIPHER methodology is capable of using experimental input from any of these assays, as long the measurements provide viable cell count or a proxy quantity (e.g. fluorescence intensity) that is proportional to the cell number. Generally, real-time techniques may yield superior deconvolution results due to a reduction in the total noise of the data set.

Given a set of experimental drug-response data on a bulk tumor sample, DECIPHER solves a series of optimization problems to deconvolve and characterize individual subcomponents of the bulk sample in terms of varying drug sensitivity profiles. In particular, each experimental observation, denoted by $\mathcal{O}_{j,k,r}$, corresponds to a cell population number measured under drug concentration $d(j)$ where $j \in \{1, \ldots, C\}$, time point t_k, where $k \in \{1, \ldots, T\}$, and replicate $r \in \{1, \ldots, R\}$. We denote the total set of observations by $\mathcal{O}$.

For simplicity, we will first assume that there are S subpopulations. We then denote the parameter set by $\mathcal{P}_S$, and we use the function

$$f(d, t; \mathcal{P}_S) = X(0) \sum_{i=1}^{S} \pi_i \exp\left[t\left(\alpha_i + \log H\left(d; b_i, E_i, n_i\right)\right)\right].$$

to model the cell population at drug dose d and time t. Our statistical model of experimental observations will be based on this deterministic model. In particular, we model each experimental observation as an independent standard Gaussian random variable with mean $f(d(j), t(k); \mathcal{P}_S)$ and standard deviation $\sigma\left(d(j), t(k)\right)$. Note that we allow the standard deviation σ to vary with dose and time. This is because at low doses, and high times we expect a larger variances due to the larger cell counts. Therefore we define

$$\sigma(d, t) = \begin{cases} \sigma_H, & d \leq D_L \text{ and } t \geq T_L \\ \sigma_L, & \text{otherwise.} \end{cases}$$

Our standard deviation is thus characterized by four parameters, $\sigma = (\sigma_L, \sigma_H, D_L, T_L)$. We will denote the set of time-dose observations where we use standard deviation σ_H by I_H, and the set where we use σ_L by I_L, and we denote their cardinalities by $|I_H|$ and $|I_L|$.

Assuming S subpopulations we can use this model to write the log-likelihood as

$$\begin{aligned} L\left(\mathcal{P}_S, \sigma; \mathcal{O}\right) = & -\frac{1}{2} R|I_H| \log(2\pi\sigma_H) - \frac{1}{2} R|I_L| \log(2\pi\sigma_L) \\ & - \frac{1}{2\sigma_H^2} \sum_{r=1}^{R} \sum_{(j,k)\in I_H} \left(\mathcal{O}_{j,k,r} - f\left(d(j), t(k); \mathcal{P}_S\right)\right)^2 \\ & - \frac{1}{2\sigma_L^2} \sum_{r=1}^{R} \sum_{(j,k)\in I_L} \left(\mathcal{O}_{j,k,r} - f\left(d(j), t(k); \mathcal{P}_S\right)\right)^2 . \end{aligned} \tag{4}$$

For a fixed S, we thus compute the maximum likelihood estimates of the model parameters by solving the optimization problem

$$(\hat{\mathcal{P}_S}, \hat{\sigma}) = \text{argmax}_{\mathcal{P}_i, \sigma} L\left(\mathcal{P}_S, \sigma; \mathcal{O}\right). \tag{5}$$

Model selection using the elbow method. To infer the number of subpopulations in the mixture, DECIPHER is fitted to the data repeatedly, for each number of subpopulations S in $S = \{1, 2, \ldots, S_{\max}\}$ in turn, and the S negative log-likelihood values are recorded. We then plot the negative log-likelihood values as a decreasing function of S, and observe the number of subpopulations corresponding to which the negative log-likelihood does not decrease significantly further. This means that no useful increase in model accuracy is gained by including another additional population. This point of inflection of the negative log-likelihood is called the elbow of the curve. The optimal number of populations is then chosen by the experimenter through visual inspection. The resulting estimate $\hat{\mathcal{P}}_{\hat{S}}$ contains the inferred population's drug response substructure: the estimated number of populations along with the estimated mixture frequency and estimated drug sensitivity GR_{50} of each subpopulation. This method is known as the elbow method, and it is a well-known heuristic for model selection in cases where the model fit generally increases with complexity. Model selection is shown in Supplementary figures S17.

Optimization methodology

The maximum likelihood estimate of the parameters $\hat{\theta}_{\text{MLE}}$ was obtained by maximizing the log-likelihood (4), subject to constraints that were placed on the range of each parameter. This constrained optimization problem was performed using the function `fmincon` from the MATLAB Optimization Toolbox [32], with the default interior-point optimization method. To combat converging to suboptimal local minima, the log-likelihood was maximized repeatedly and independently, by starting from N_{optim} different random initial positions for the parameter θ, sampled uniformly within their allowed range (except for the parameter E, which was sampled log-uniformly within the bounds). Among the N_{optim} minima, the one with the highest log-likelihood value was chosen as estimate $\hat{\theta}_{\text{MLE}}$.

Calculation of GR_{50} values

The viability curve and associated metrics of drug response (e.g. IC_{50}, EC_{50}) typically exhibit dependence on the timing of data collection (see [17]). We form a growth rate curve by inferring the growth rate $r(d)$ at each tested dose level d. In contrast to the viability curve the growth rate curve does not have a hidden dependence on the duration of the experiment, assuming exponential growth. Once the parameters of model (3) are estimated for each subpopulation using the inferential procedures above, the GR_{50} for each subpopulation can be explicitly determined using the set of parameters (α_i, b_i, E_i, and n_i). Following Sorger and et al [17] we characterize dose-response of clones with a GR_{50} value. This number represents the dose at which the cellular growth rate experiences half of its total reduction. In particular, suppose that we are interested in a homogeneous population with the growth rate at dose d given by

$$GR(d; \alpha, b, E, n) = \alpha + \log H\left(d; b, E, n\right).$$

Note that we will generally suppress the dependence on parameters and simply write $GR(d)$. If the maximum dose administered is d_m, and the minimum dose administered is 0, then the median growth rate is $r_m = (GR(0) + GR(d_m))/2$. We then define the GR_{50} as the dosage that results in this growth rate, i.e., the value d such that $GR(d) = r_m$. We can then solve to obtain

$$GR_{50} = E\left(\frac{e^{r_m-\alpha} - 1}{b - e^{r_m-\alpha}}\right)^{1/n}.$$

Generation of synthetic population data

By defining a number of populations S and a parameter set $\mathcal{P}_S \equiv \{\pi_i, \alpha_i, b_i, E_i, n_i : i \in \{1, \ldots, S\}\}$, synthetic data can be generated in a deterministic manner with equation (3). Table 1 shows the parameters $\{\alpha_i, b_i, E_i, n_i : i \in \{1, 2, 3\}\}$ of the blue, red and yellow populations in figures 2, S3, S4, S5 and S6.

	α	b	E	n
Clone 1 (blue)	0.03	0.3	0.0001	3.0
Clone 2 (red)	0.03	0.4	0.01	3.0
Clone 3 (yellow)	0.03	0.5	0.1	3.0

Table 1: Parameters for simulated data.

For the synthetic validation, simulated data with initial population size of $Z_0 = 1000$ cells were generated for the following 9 mixtures of the three cell populations in table 1: [1,0,0], [0,1,0], [0,0,1], [0.5,0.5,0], [0.7, 0.3, 0], [0.3, 0.7, 0], [1/3, 1/3, 1/3], [0.4, 0.3, 0.3] and [0.6, 0.2, 0.2].

We chose 17 simulated drug concentrations. One equal to zero, the rest spaced log-linearly in a region designed to cover the GR50 values of the simulated populations. The simulated concentrations were: [0, 0.00000500, 0.0000108, 0.0000232, 0.0000500, 0.000108, 0.000232, 0.000500, 0.00108, 0.00232, 0.00500, 0.0108, 0.0232, 0.0500, 0.108, 0.232, 0.5] μM. Cell counts were measured at 12-hour intervals from 0 to 96 hours, and 4 replicates of the simulation were carried out, where the only difference between the replicates was the randomly sampled observation noise: A random noise term was added to each observed cell count, sampled from an independent and identically distributed (i.i.d.) Gaussian distribution with mean 0 and standard deviation ranging from 1 to 50% of the initial cell count. Any negative cell count caused by the additive noise was set to zero. This gives the following expression for the generated observation $\mathcal{O}_{j,k,r}$ with concentration number j at time k for replicate r:

$$\mathcal{O}_{j,k,r} = \max\Big(Z(t,d;\mathcal{P}) + \varepsilon_{j,k,r} \ , \ 0\Big), \qquad \varepsilon_{j,k,r} \sim \mathcal{N}(0,\sigma^2) \ \text{ i.i.d.} \tag{6}$$

Model of multiple myeloma under treatment

Inspired by [33] we present a mathematical model of M-protein levels of a multiple myeloma patient under treatment with an anti-cancer drug. This model assumes that the patient has two subpopulations of cancer cells with distinct responses to the drug. In particular the cancer cells and M-protein levels are governed by the following system of ordinary differential equations

$$\frac{dx}{dt} = \frac{r_1(d)x}{1+p(x+y)}, \tag{7a}$$

$$\frac{dy}{dt} = \frac{r_2(d)y}{1+p(x+y)}, \tag{7b}$$

$$\frac{dz}{dt} = r_3(x+y) - d_3 z, \tag{7c}$$

where x and y denote number of myeloma cells in subpopulations 1 and 2 respectively, and z denotes M-protein concentration in plasma. Parameters r_1 and r_2 are the net growth rates under treatment of subpopulations 1 and 2 respectively. We assume the net growth rates can be computed as

$$r_i(d) = \alpha_i + \log H(d; b_i, E_i, n_i), \quad i \in \{1,2\}, \tag{8a}$$

where $(\alpha_i, b_i, E_i, n_i)$ are the estimated parameters of subpopulation i using DECIPHER. The term $(1+p(x+y))^{-1}$ in equations 7a and 7b alters the growth rate of both subpopulations when the total number of cells increases. Parameters r_3 and d_3 are the production and decay rate of the M-protein, respectively. Inspired by [33], we use $p = 10^{-13}$, $r_3 = 0.07 * 10^{-13}$ and $d_3 = 0.07$.

Model parameter ranges

For model with $S = \{1, 2, 3, 4\}$ populations, the log-likelihood was maximized $N_{\text{optim}} = 1000$ times or more to combat local minima. For each maximization, the initial estimate was sampled from within the bounds on the parameter range, which were set to the values listed below for the different datasets.

The parameter ranges for the different settings were largely similar. Some differences occur due to different concentration scales in the different experiments or due to parameter estimates hitting the boundary of the domain, in which case the range was expanded. When working with the Ba/F3 cells we needed to adjust the lower bound on the parameter b. Due to the complexity of the optimization problem, the solver had a tendency to push b towards an unrealistically low value. To address this issue we used previous observations and derived a realistic lower bound on b. Denote the net growth rate of the cells by $\lambda = \beta - \mu$, where β is the birth rate and μ the death rate. From [34], we know that $\beta \leq .06$. We can thus write $\mu = \beta - \lambda \leq .06 - \lambda_{min} = d_0$, where λ_{min} is the minimum observed growth rate amongst all Ba/F3 cell line experiments. Thus the maximal possible death rate is d_0, and the minimal possible net growth rate is $-d_0$. Next note that according to our growth rate model, as the dose d goes to infinity the growth rate decreases to the lower limit $\alpha + \log(b)$. Therefore we know that $\alpha + \log(b) \geq -d_0$. We again use that $\alpha \leq .06$, and based on observed data we set $\lambda_{min} = .04$ and get $d_0 = 0.2$. However to account for any possible errors in the method we increase d_0 to be 0.07. This then gives us the lower bound $\log(b) \geq -0.08$ or equivalently $b \geq 0.878$.

NSCLC data.

$p_i \in [0, 1]$ with the inequality constraint $\sum_{s=1}^{S-1} p_i \leq 1$

$\alpha_i \in [0, 0.1]\ hours^{-1}$

$b \in [0, 1]\ hours^{-1}$

$E \in [0, 50]\ \mu M$

$n \in [0, 50]$

$\sigma_L, \sigma_H \in [0, 5500]$

Ba/F3 data.

$p_i \in [0, 1]$ with the inequality constraint $\sum_{s=1}^{S-1} p_i \leq 1$

$\alpha_i \in [0, 0.06]\ hours^{-1}$

$b \in [0.878, 1]\ hours^{-1}$, see comment below.

$E \in [0, 50]\ \mu M$

$n \in [0.001, 20]$

$\sigma_L, \sigma_H \in [0, 2500]$

Synthetic data.

$p_i \in [0, 1]$ with the inequality constraint $\sum_{s=1}^{S-1} p_i \leq 1$

$\alpha_i \in [0, 0.1]\ hours^{-1}$

$b_i \in [0.27, 1]\ hours^{-1}$

$E_i \in [10^{-6}, 0.5]\ \mu M$

$n_i \in [0.01, 10]$

$S \in [0, 4]$

$\sigma_L, \sigma_H \in [10^{-6}, 5000]$

Multiple Myeloma data.
For the multiple myeloma patient data, an inital parameter range was chosen for all patients. Then if one or more of the inferred parameters happened to lie on or near the upper or lower bound, the parameter range was increased for that patient until the estimate was no longer on the bound. Therefore, the parameter for the E and σ variables are different for some of the patients.

$p_i \in [0, 1]$ with the inequality constraint $\sum_{s=1}^{S-1} p_i \leq 1$

$\alpha_i \in [-0.1, 0.1]\ hours^{-1}$

$b_i \in [0, 1]\ hours^{-1}$

$n_i \in [0.01, 10]$

$S \in [0, 5]$

The E parameter ranges were:

$E_i \in [10^{-6}, 2]\ \mu M$ for patient MM2108.

$E_i \in [10^{-6}, 50]$ μM for patient MM720.

$E_i \in [10^{-6}, 5]$ μM for patient MM195.

$E_i \in [10^{-6}, 5]$ μM for patient MM36.

$E_i \in [10^{-6}, 100]$ μM for patient MM1420.

The σ parameter ranges were:

$\sigma_L, \sigma_H \in [10^{-6}, 50,000]$ for patient MM2108.

$\sigma_L, \sigma_H \in [10^{-6}, 1,000,000]$ for patient MM720.

$\sigma_L, \sigma_H \in [10^{-6}, 150,000]$ for patient MM195.

$\sigma_L, \sigma_H \in [10^{-6}, 250,000]$ for patient MM36.

$\sigma_L, \sigma_H \in [10^{-6}, 150,000]$ for patient MM1420.

Ba/F3 cell line experiments

Preparation of sensitive and resistant cell lines

BCR-Abl-T315I expressing plasmid was established by site-directed mutagenesis of p210 BCR-Abl (Addgene 27481) using QuickChange II XL (Agilent Technologies) with the forward primer 5' GGGAGCCCCCGTTCTATATCATCATTGAGTTCATGACCTACG 3' and the reverse primer 5' CGTAGGTCATGAACTCAATGATGATATAGAACGGGGGCT CCC 3' for T315I. To generate cells stably expressing BCR-Abl (imatinib-sensitive) and BCR-Abl-T315I (imatinib-resistant), parental Ba/F3 cells were transfected with the appropriate plasmids by electroporation using Amaxa biosystems nucleofecor II and stable cells were established by selecting with medium containing 500µg/ml Geneticin (Gibco, UK) and lacking the growth factor IL3 (BCR-ABL activity can overcome the requirement for IL3 of untransformed parental cells for survival/proliferation [20]). Furthermore, Ba/F3 cells expressing BCR-Abl were stably transfected with GFP expression, pRNAT-H1.1/Hygro plasmid from Genscript (Piscataway NJ, USA). The resulting subpopulations exhibited distinctive phenotypic differences upon treatment with Imatinib.

Cell cultures

Parental Ba/F3 cells were maintained in RPMI-1640 supplemented with 10% heat-inactivated Fetal Bovine Serum (FBS), 7.5 ng/ml IL3 and 1% penicillin and streptomycin at 37°C under a humidified atmosphere containing 5% CO2. Ba/F3 cells stably expressing BCR-Abl and BCR-Abl-T315I were maintained in medium lacking IL3.

Experimental procedures

Cells were harvested at 70-80% confluence, stained with trypan blue (ThermoFisher, UK),

and counted with a Countess 3 Automated Cell Counter (Life Technologies). Mono- and co-cultures were seeded at different initial ratios in 384 well microplates (Greiner Bio-One) that contained different concentrations of imatinib (Cayman, USA). Imatinib ranging from (0 - 5μM) was dispensed using an Echo acoustic liquid dispenser (Labcyte, San Jose, CA, USA) in seven replicates per condition. Then time-lapse microscopy images were obtained for bright field and GFP using IncuCyte (Essen BioScience, UK) every 3 hours over the course of 72 hours.

Image Processing

Images were processed with the open-source software ImageJ [35] Images were background subtracted, converted to 8-bit, bandpass filtered, sharpened, contrast enhanced, and thresholded. Then images were converted to binary images, watershed segmentation was performed, and raw cell numbers were extracted.

NSCLC cell line experiments

Cell Cultures

HCC827 and H1975 cell lines were maintained in RPMI-1640 media supplemented with 10% Fetal Bovine Serum and 1% penicillin and streptomycin under standard cell culture growth conditions (37oC and 5% CO2.

Experimental Growth Assay

Tumor cells were seeded in 96-well black walled plates at 5,000 cells per well. The following day, the cells were treated with erlotinib at various concentrations (0, 0.1, 1, 10uM). Cell counts were determined at 0, 24, and 48 hours post drug treatment using the Cellomics Arrayscan High Content Screening Platform. Briefly, cells were stained with 5 g/mL Hoechst 33342 (nuclear marker to determine total cell count) and 5 g/mL Propidium Iodide (PI - vital dye to determine dead cells) for 45 minutes prior to imaging. The average intensity for Hoechst and PI was determined for each cell to classify as live or dead. Each condition was performed in replicates of four. For admixture experiments, each cell line was labeled with a different CellTracker dye (CellTracker orange CMTMR and H1975 labeled with CellTracker green CMFDA). The cells were mixed at the specified ratios (total 5,000 cells/well) and imaged following the procedures outlined above.

Drug screen of Multiple Myeloma patient samples

Patient samples

The multiple myeloma (MM) patients enrolled in this study were recruited from the Oslo Myeloma Center at Ullevål Oslo University Hospital under the Regional Committee approval for Medical and Health Research Ethics of South-Eastern Norway (REC-2016/947)

The MM samples were obtained following written informed consent in compliance with the Declaration of Helsinki.

Primary MM cells processing

Bone marrow samples from 5 relapsed myeloma patients were collected in ACD tubes. Details about patient ID, treatment lines and refractory status are provided in Supplementary Table 1. A Lymphoprep TM (Stemcell Technologies) density gradient centrifugation method was used to obtain bone marrow mononuclear cells (BMMCs) from patient samples. As described in [36], after CD8 T cell depletion by Dynabeads (Life Technologies), BMMCs were then stimulated by activated T helper cells in the presence of Human T-activator CD3/CD28 Dynabeads (Life Technologies) and 100U/ml human interleukin-2 (hIL-2, Roche, Mannheim, Germany). After 24h, CD138+ MM cell enrichment was performed from the BMMC fraction by immune-magnetic microbeads CD138+ (Milteny Biotec, Bergisch Gladbach, Germany).

Drug sensitivity assay

CD138+ MM cells (200,000 cells/ml) derived from activation assays were treated with drugs at 9 concentrations using a drug customized concentration range (0,1-10,000), as described in [24]. The drug panel included clinically relevant anti-myeloma drugs, Dexamethasone (0,1-10,000), ixazomib (1-10,000), thalidomide (0,1-10,000), selinexor (0,1-1000), melflufen (0,1-1000) and venetoclax (0,1-10,000). Pre-printed drug plates were made by an acoustic dispenser (Echo 550, LabCyte Inc., San Jose, CA, USA), by the Chemical Biology Platform, NCMM, University of Oslo. Control agents included a negative control, 0,1% solvent solution dimethyl sulfoxide (DMSO), and a positive control 100 uM benzethonium chloride (BzCl). In brief, MM cells were diluted in culture medium (RPMI 1640 medium supplemented with 10% fetal bovine serum, 2mM L-glutamine, penicillin (100U/ml), streptomycin (100 $\mu g/ml$), and 25 μl of cell suspension was transferred to 384-well plates using a Certus Flex liquid dispenser (Fritz Gyger, Switzerland). Afterward, plates were incubated at 37°C and 5% CO2 humidified environment. Cell viability was measured at 4 different time points (0h-96h), using the CellTiterGlo (Promega, Madison, WI, USA) ATP assay according to manufacturer's instructions and with an Envision Xcite plate reader (Perkin Elmer, Shelton, CT, USA) to measure luminescence.

Data and code availability

All data and code used in this article are publicly available at the online repository of the Oslo Center for Biostatistics and Epidemiology.

References

[1] Michalina Janiszewska et al. "The impact of tumor epithelial and microenvironmental heterogeneity on treatment responses in HER2+ breast cancer". In: *JCI insight* 6.11 (2021).

[2] Andriy Marusyk, Michalina Janiszewska, and Kornelia Polyak. "Intratumor heterogeneity: the rosetta stone of therapy resistance". In: *Cancer cell* 37.4 (2020), pp. 471–484.

[3] Dongya Jia et al. "Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways". In: *Proceedings of the National Academy of Sciences* 116.9 (2019), pp. 3909–3918.

[4] Nurbubu T Moldogazieva, Innokenty M Mokhosoev, and Alexander A Terentiev. "Metabolic heterogeneity of cancer cells: an interplay between HIF-1, GLUTs, and AMPK". In: *Cancers* 12.4 (2020), p. 862.

[5] Charlotte K Y Ng et al. "Predictive performance of microarray gene signatures: impact of tumor heterogeneity and multiple mechanisms of drug resistance". In: *Cancer research* 74.11 (2014), pp. 2946–2961.

[6] Mel Greaves and Carlo C Maley. "Clonal evolution in cancer". In: *Nature* 481.7381 (2012), pp. 306–313.

[7] Carlo C Maley et al. "Genetic clonal diversity predicts progression to esophageal adenocarcinoma". In: *Nature genetics* 38.4 (2006), pp. 468–473.

[8] Daniel L Dexter and John T Leith. "Tumor heterogeneity and drug resistance." In: *Journal of clinical oncology* 4.2 (1986), pp. 244–257.

[9] Sreenath V Sharma et al. "A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations". In: *Cell* 141.1 (2010), pp. 69–80.

[10] Birgit Knoechel et al. "An epigenetic mechanism of resistance to targeted therapy in T cell acute lymphoblastic leukemia". In: *Nature genetics* 46.4 (2014), pp. 364–370.

[11] Marc Hafner, Mario Niepel, and Peter Sorger. "Alternative drug sensitivity metrics improve preclinical cancer pharmacogenomics". In: *Nature Biotechnology* 35 (2017), pp. 500–502.

[12] Nikita Pozdeyev et al. "Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies". In: *Oncotarget* 7.32 (2016), p. 51619.

[13] Elizabeth A Brooks et al. "Applicability of drug response metrics for cancer studies using biomaterials". In: *Philosophical Transactions of the Royal Society B* 374.1779 (2019), p. 20180226.

[14] Shannon M Matulis et al. "Functional profiling of venetoclax sensitivity can predict clinical response in multiple myeloma". In: *Leukemia* 33.5 (2019), pp. 1291–1296.

[15] Cecilia Bonolo de Campos et al. ""Direct to Drug" screening as a precision medicine tool in multiple myeloma". In: *Blood cancer journal* 10.5 (2020), pp. 1–16.

[16] Giovanni Y Di Veroli et al. "An automated fitting procedure and software for dose-response curves with multiphasic features". In: *Scientific reports* 5.1 (2015), pp. 1–11.

[17] Marc Hafner et al. "Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs". In: *Nature methods* 13.6 (2016), pp. 521–527.

[18] Daniela Cadena-Herrera et al. "Validation of three viable-cell counting methods: manual, semi-automated, and automated". In: *Biotechnology Reports* 7 (2015), pp. 9–16.

[19] Shannon Mumenthaler et al. "Evolutionary Modeling of Combination Treatment Strategies To Overcome Resistance to Tyrosine Kinase Inhibitors in Non-Small Cell Lung Cancer". In: *Mol. Pharm.* 8.6 (2011), 2069–2079.

[20] Dagim Shiferaw Tadele et al. "A cell competition–based small molecule screen identifies a novel compound that induces dual c-Myc depletion and p53 activation". In: *Journal of Biological Chemistry* 296 (2021).

[21] Shaji Kumar et al. "International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma". In: *The lancet oncology* 17.8 (2016), e328–e346.

[22] Jonathan J Keats et al. "Clonal competition with alternating dominance in multiple myeloma". In: *Blood, The Journal of the American Society of Hematology* 120.5 (2012), pp. 1067–1076.

[23] Jens G Lohr et al. "Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy". In: *Cancer cell* 25.1 (2014), pp. 91–101.

[24] Mariaserena Giliberto et al. "Ex vivo drug sensitivity screening identifies drug combinations which act synergistically against multiple myeloma cells". In: *Submitted, pending revision* (2022).

[25] Jinghua Shi et al. "A survey of optimization models on cancer chemotherapy treatment planning". In: *Annals of Operations Research* 221.1 (2014), pp. 331–356.

[26] Xiaoran Lai et al. "Toward personalized computer simulation of breast cancer treatment: A multiscale pharmacokinetic and pharmacodynamic model informed by multitype patient data". In: *Cancer research* 79.16 (2019), pp. 4293–4304.

[27] George W Swan. "Role of optimal control theory in cancer chemotherapy". In: *Mathematical biosciences* 101.2 (1990), pp. 237–284.

[28] Qie He et al. "Optimized treatment schedules for chronic myeloid leukemia". In: *PLoS computational biology* 12.10 (2016), e1005129.

[29] Pierre Comon and Christian Jutten. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[30] Terry L Riss et al. "Cell viability assays". In: (2016).

[31] Colleen M Garvey et al. "A high-content image-based method for quantitatively studying context-dependent cell population dynamics". In: *Scientific reports* 6 (2016), p. 29752.

[32] *MATLAB Optimization Toolbox*. R2020b.

[33] Min Tang et al. "Myeloma cell dynamics in response to treatment supports a model of hierarchical differentiation and clonal evolution". In: *Clinical Cancer Research* 22.16 (2016), pp. 4206–4214.

[34] Ron Milo et al. "BioNumbers—the database of key numbers in molecular and cell biology, BNID 103994". In: *Nucleic acids research* 38.suppl_1 (2010), pp. D750–D753.

[35] Daniel Sage et al. "MIJ: making interoperability between ImageJ and Matlab possible". In: *ImageJ User & Developer Conference*. Vol. 2426. 2012.

[36] Dong Wang et al. "Autologous bone marrow Th cells can support multiple myeloma cell proliferation in vitro and in xenografted mice". In: *Leukemia* 31.10 (2017), pp. 2114–2121.

Acknowledgements

The numerical computations were performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway.

The project received funding from the UiO:LifeScience initiative through the convergence environment grant PerCaThe.

A.K.L, E.M.M. and A.F. were supported by the center for research-based-innovation BigInsight under project number 237718. J.F. and K.L. were supported by the Fulbright US-Norway Foundation. J.F., K.L., and J.N. were supported by the University of Oslo-University of Minnesota Norwegian Centennial Chair Grant. J.F. was supported by the US National Science Foundation under grant number DMS-2052465. K.L. was supported by the US National Science Foundation under grant number CMMI-1552764.

We acknowledge funding from the Research Council of Norway with project numbers 294916, 261936, 309273 and 262652; the Norwegian Cancer Society with project number 182524; and the Norwegian Health Authority South-East with project number 2019096.

The authors also acknowledge the Centre for Digital Life Norway for supporting the partner projects PerCaThe and PINpOINT.

We thank the Digital Scholarship Center, University of Oslo, for insightful advice on the visual representation and communication of our research findings.

Supplementary Information

Model Extension to Interacting Populations

Our model currently ignores potential interactions between subpopulations. Based on the sample size of our current data sets we were not able to fit a more complex model that allows for interacting populations. For the situation when sufficient data are available, we propose the model below that allows for interactions between the subpopulation. Assuming that there are S subpopulations, for each $i \in \{1, \ldots, S\}$ define the function

$$f_i(\mathbf{X}, d) = \sum_{l=1}^{S} (a_{il} X_j - c_{il} X_i X_l) - \alpha_i(d) X_i,$$

where d is possible drug dose and $\mathbf{X} \in \mathbb{R}_+^S$. In the above equation $\alpha_i(d) = \alpha_i - \log H_i(d)$ represents the death rate, which increases as dose d increases. The parameter a_{il} represent the rate at which type-l cells produce type-i cells, and α_{ii} is simply the birth rate of the type-i cells. We assume that each α_{il} term is non-negative. The term c_{il} represents the effect of population l on population i. If $c_{il} > 0$ then population l inhibits population i, if $c_{il} < 0$ then population l encourages population i to grow, and finally if $c_{il} = 0$ then population l has no direct effect on population i. Note that the term c_{ii} represents the effect of type-i cells on itself and we assume that $c_{ii} > 0$. The parameters α_{il} allow for inter-conversion between cell types, and the parameters c_{il} allow for inhibition or promotion between cell types.

For dose d, and initial population vector $\mathbf{x}^0 = (x_1^0, \ldots, x_S^0)$, define $\{\mathbf{X}(t, d; \mathbf{x}^0); t \geq 0\}$ as the solution to the differential equation

$$\dot{X}_i(t, d) = f_i(\mathbf{X}, d), \qquad \text{for each } i \in \mathcal{S}, \tag{9}$$

with initial condition $X_i(0) = x_i^0$. Define $x_0 = \sum_i x_0^i$ and write $x_i^0 = p_i x_0$. We assume that x_0 is a known quantity, but the proportions $\{p_i\}_{i \in \mathcal{S}}$ are unknown. We denote the model-predicted total population at time t under dose d by $X(t, d)$. Recall that the total population is the observable variable in our model.

In this interacting population model, we have more model parameters, namely the parameter set

$$\mathcal{P} = \{(\{a_{il}\}_{l \in \mathcal{S}}, \{c_{il}\}_{l \in \mathcal{S}}, p_i, \nu_i, b_i, E_i, n_i)\,; i \in \mathcal{S}\}\,.$$

To make clear the dependence on the parameter set $\mathcal{P}$, we denote the predicted total population at time t using d units of drug with parameter set $\mathcal{P}$ by $X(t, d; \mathcal{P})$.

Similar to our main model, we will start by simply using additive Gaussian noise for our measurement error. In particular, we assume that observation at dose d_j and time t_k is given by

$$x_{j,k} = X(d_j, t_k; \mathcal{P}) + Z_{j,k},$$

for i.i.d $N(0, \sigma^2)$ random variables $Z_{j,k}$. We can then implement the same maximum likelihood estimation procedure as for our original model. This will be a more computationally challenging problem because evaluating the likelihood function will require numerically solving the non-linear differential equation (9). In addition, this inference problem is more difficult because we have a higher dimensional parameter space to search over.

Supplementary figures

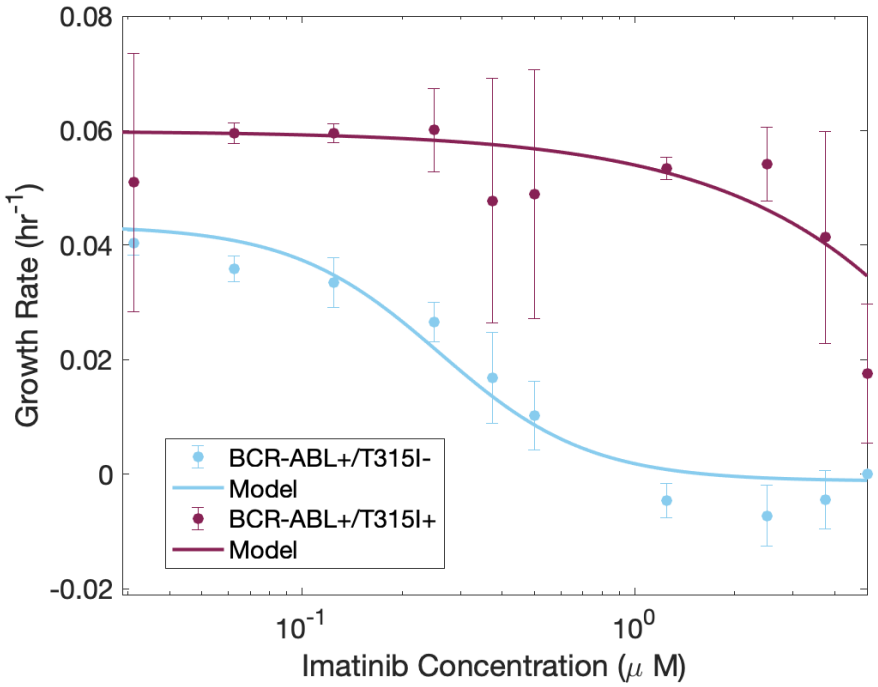

Figure S1: Model from Equation (1) of the Online Methods accurately recapitulates experimental cell viability dependence on drug concentration in two example BCR-ABL positive Ba/F3 cell lines (with and without the T315I mutation) treated with tyrosine kinase inhibitor imatinib.

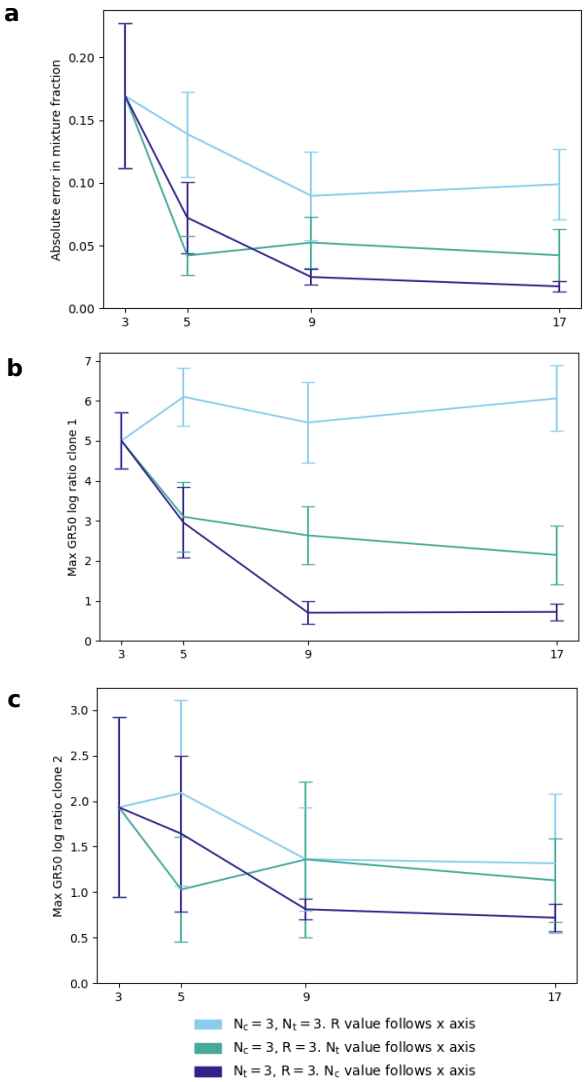

Figure S2: Comparison of accuracy gain in mixture fraction and GR50 values when increasing either the number of replicates (R), the number time points (N_t) or the number of concentrations (N_c) while keeping the other two equal to 3. The inference was carried out on 30 datasets generated from a mixture of 40% sensitive and 60% resistant cells. The standard deviation of the observation noise was equal to 10% of the initial cell count. The random seed for the noise was the only parameter varying between the 30 datasets. In a), the accuracy metric is absolute error in inferred mixture parameter; in b) and c) the metric is $max(GR50_{\text{inferred}}/GR50_{\text{true}}, GR50_{\text{true}}/GR50_{\text{inferred}})$, chosen to address the logarithmic scale of the concentrations. The plots show mean accuracy metrics with 95% confidence intervals for the mean (t-distribution with 29 degrees of freedom). The number of subpopulations (2) was assumed known, and model selection was not performed.

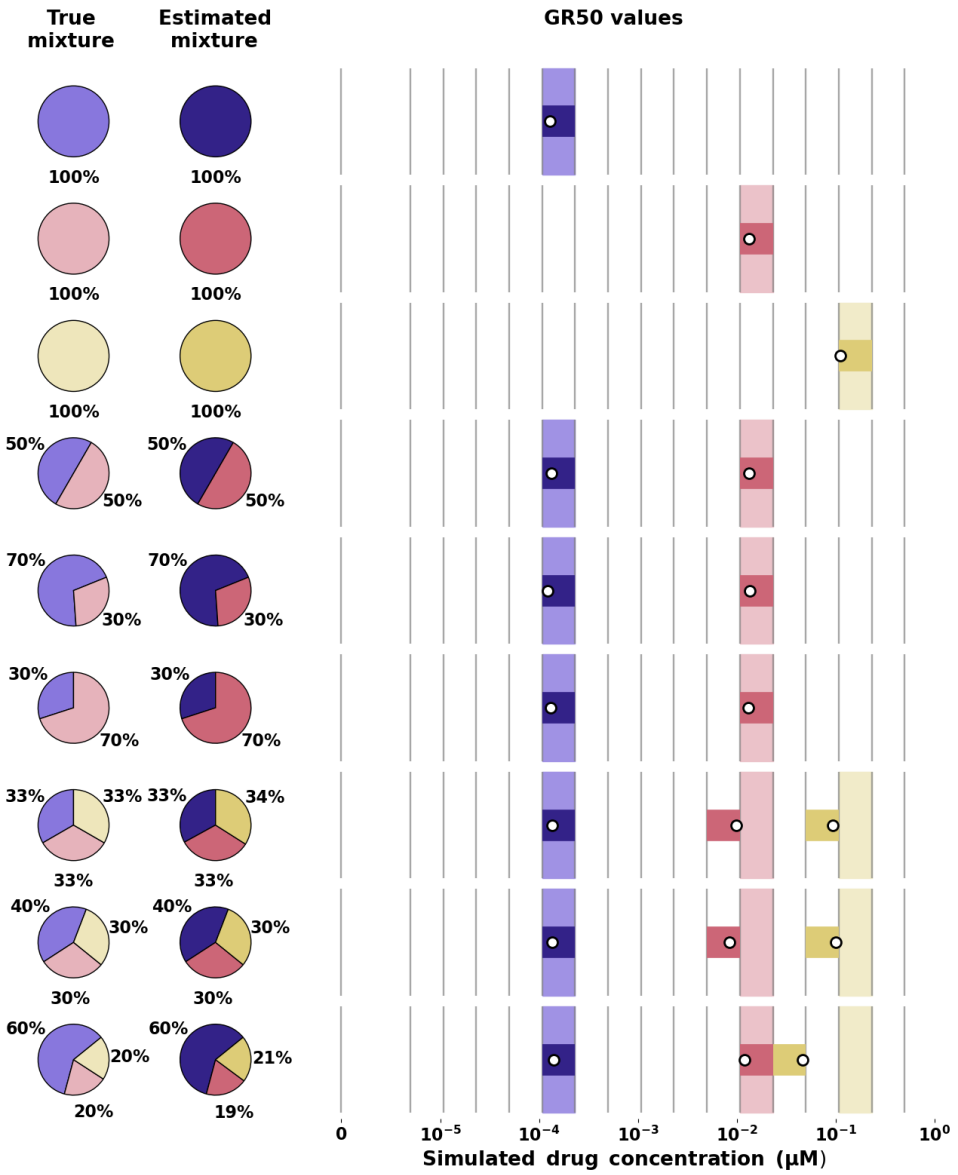

Figure S3: True and estimated mixture fractions and GR50 values for synthetic data with observation noise with standard deviation equal to 1% of the initial cell count. For each row, "True mixture" pie charts show mixture fractions used in the data generation; "Estimated mixture" pie charts show estimated mixture fractions; vertical grey lines show observed concentrations on a logarithmic scale. In the "GR50 values" panel, the region between the two observed concentrations closest to each true GR50 value is given the same color as that subpopulation has in the "True mixtures" pie chart; White dots represent estimated GR50 values, with the region between the closest observed concentrations colored in the same color as the inferred subpopulation in the "Estimated mixture" pie chart.

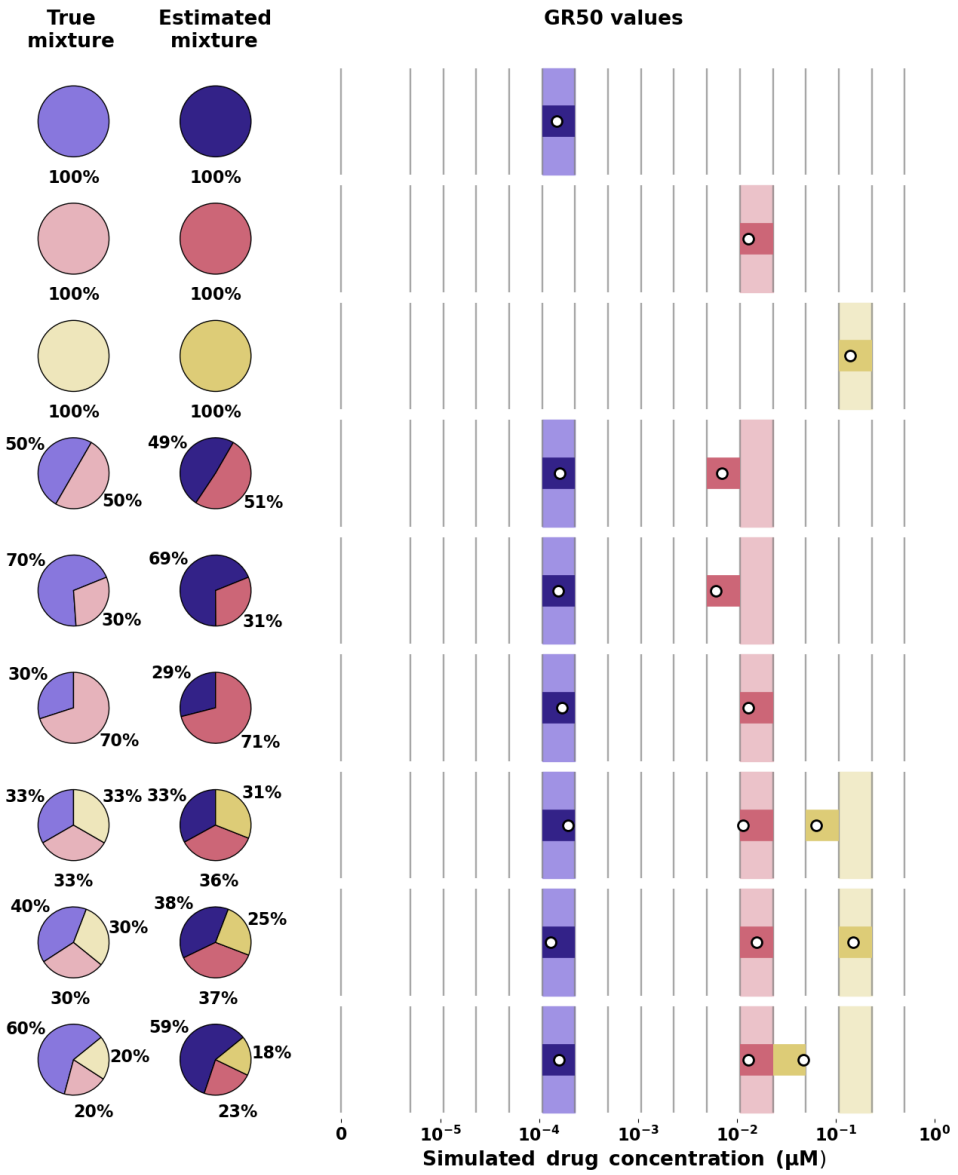

Figure S4: True and estimated mixture fractions and GR50 values for synthetic data with observation noise with standard deviation equal to 10% of the initial cell count. For each row, "True mixture" pie charts show mixture fractions used in the data generation; "Estimated mixture" pie charts show mixture fractions estimated by the model; vertical grey lines show observed concentrations on a logarithmic scale. In the "GR50 values" panel, the region between the two observed concentrations closest to each true GR50 value is given the same color as that subpopulation has in the "True mixtures" pie chart; White dots represent estimated GR50 values, with the region between the closest observed concentrations colored in the same color as the inferred subpopulation in the "Estimated mixture" pie chart.

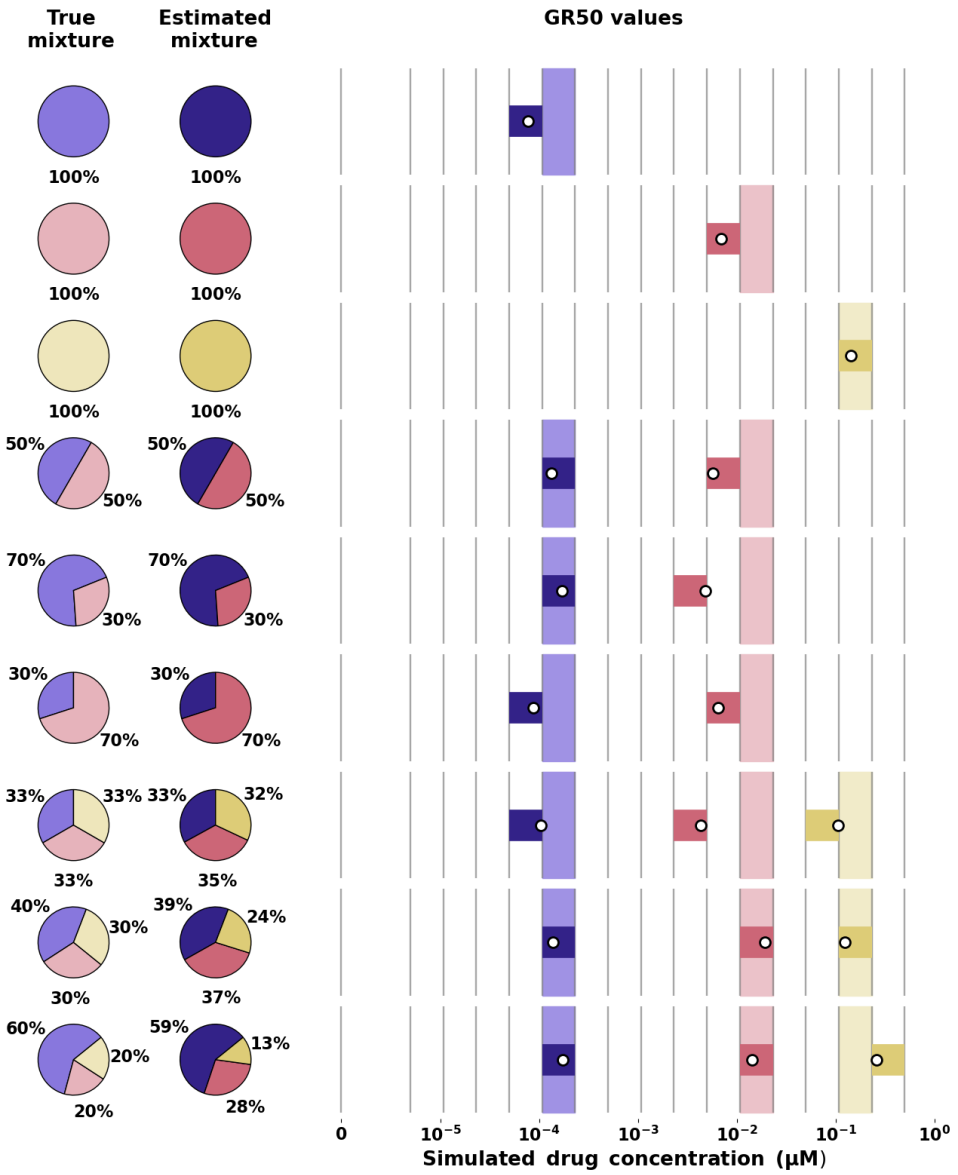

Figure S5: True and estimated mixture fractions and GR50 values for synthetic data with observation noise with standard deviation equal to 20% of the initial cell count. For each row, "True mixture" pie charts show mixture fractions used in the data generation; "Estimated mixture" pie charts show mixture fractions estimated by the model; vertical grey lines show observed concentrations on a logarithmic scale. In the "GR50 values" panel, the region between the two observed concentrations closest to each true GR50 value is given the same color as that subpopulation has in the "True mixtures" pie chart; White dots represent estimated GR50 values, with the region between the closest observed concentrations colored in the same color as the inferred subpopulation in the "Estimated mixture" pie chart.

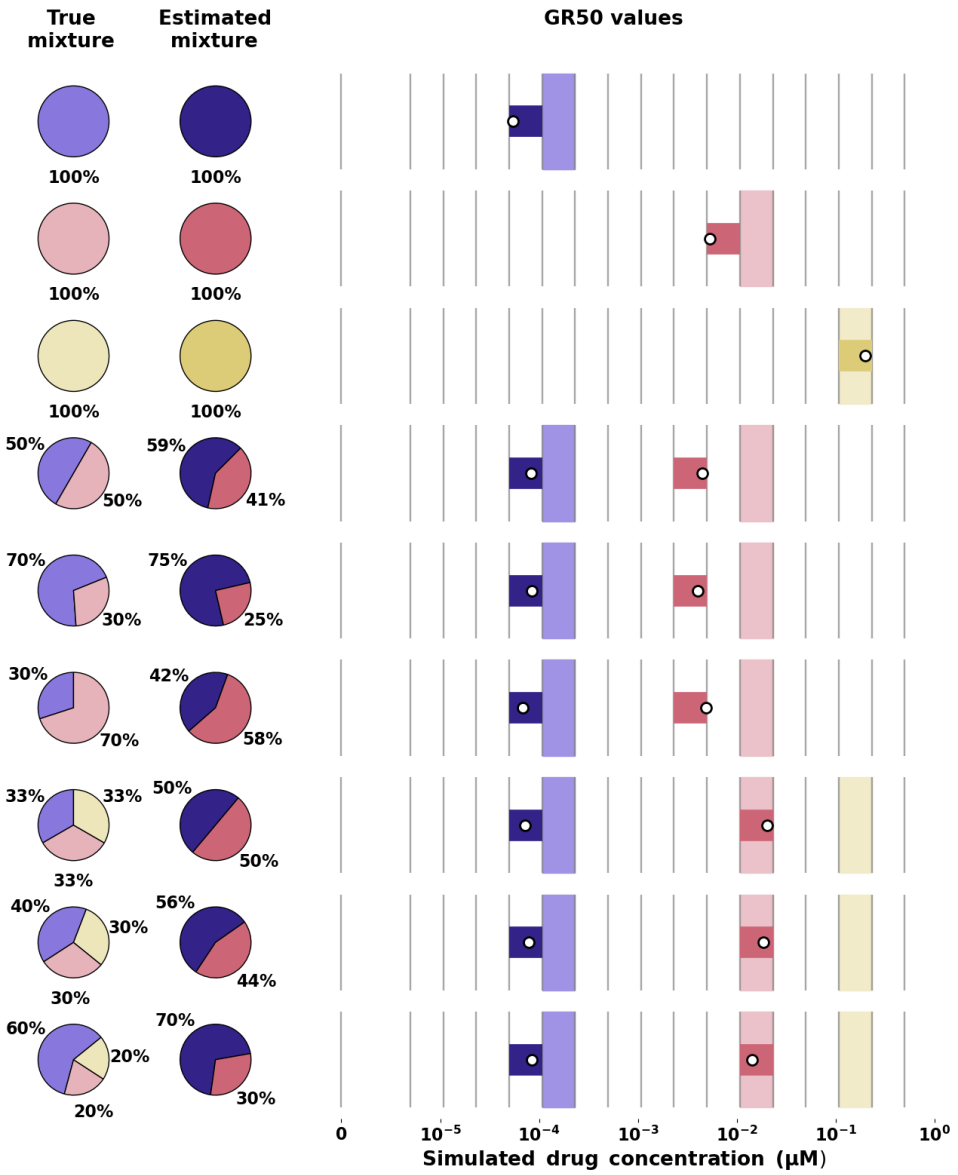

Figure S6: True and estimated mixture fractions and GR50 values for synthetic data with observation noise with standard deviation equal to 50% of the initial cell count. For each row, "True mixture" pie charts show mixture fractions used in the data generation; "Estimated mixture" pie charts show mixture fractions estimated by the model; vertical grey lines show observed concentrations on a logarithmic scale. In the "GR50 values" panel, the region between the two observed concentrations closest to each true GR50 value is given the same color as that subpopulation has in the "True mixtures" pie chart; White dots represent estimated GR50 values, with the region between the closest observed concentrations colored in the same color as the inferred subpopulation in the "Estimated mixture" pie chart.

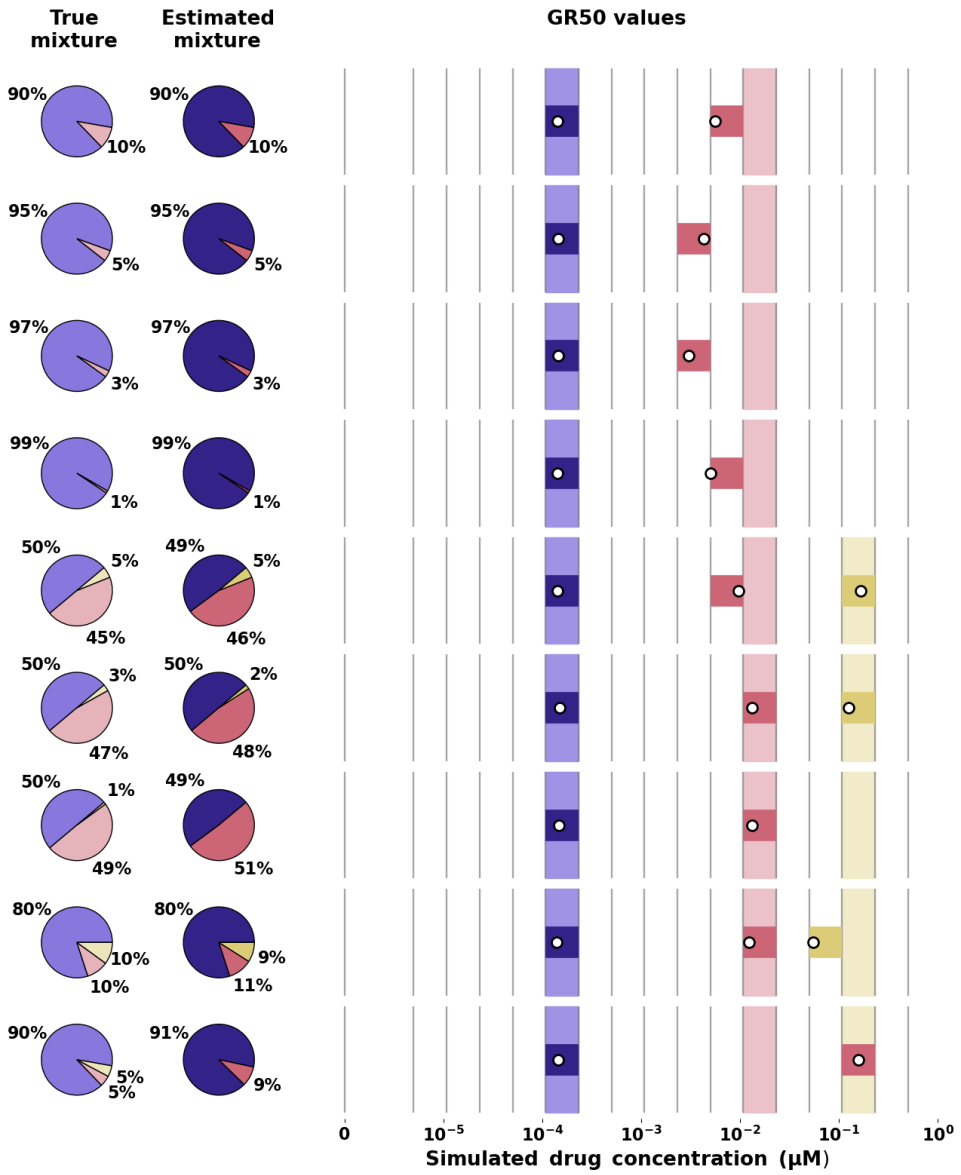

Figure S7: True and estimated mixture fractions and GR50 values for synthetic data with observation noise with standard deviation equal to 5% of the initial cell count. For each row, "True mixture" pie charts show mixture fractions used in the data generation; "Estimated mixture" pie charts show mixture fractions estimated by the model; vertical grey lines show observed concentrations on a logarithmic scale. In the "GR50 values" panel, the region between the two observed concentrations closest to each true GR50 value is given the same color as that subpopulation has in the "True mixtures" pie chart; White dots represent estimated GR50 values, with the region between the closest observed concentrations colored in the same color as the inferred subpopulation in the "Estimated mixture" pie chart.

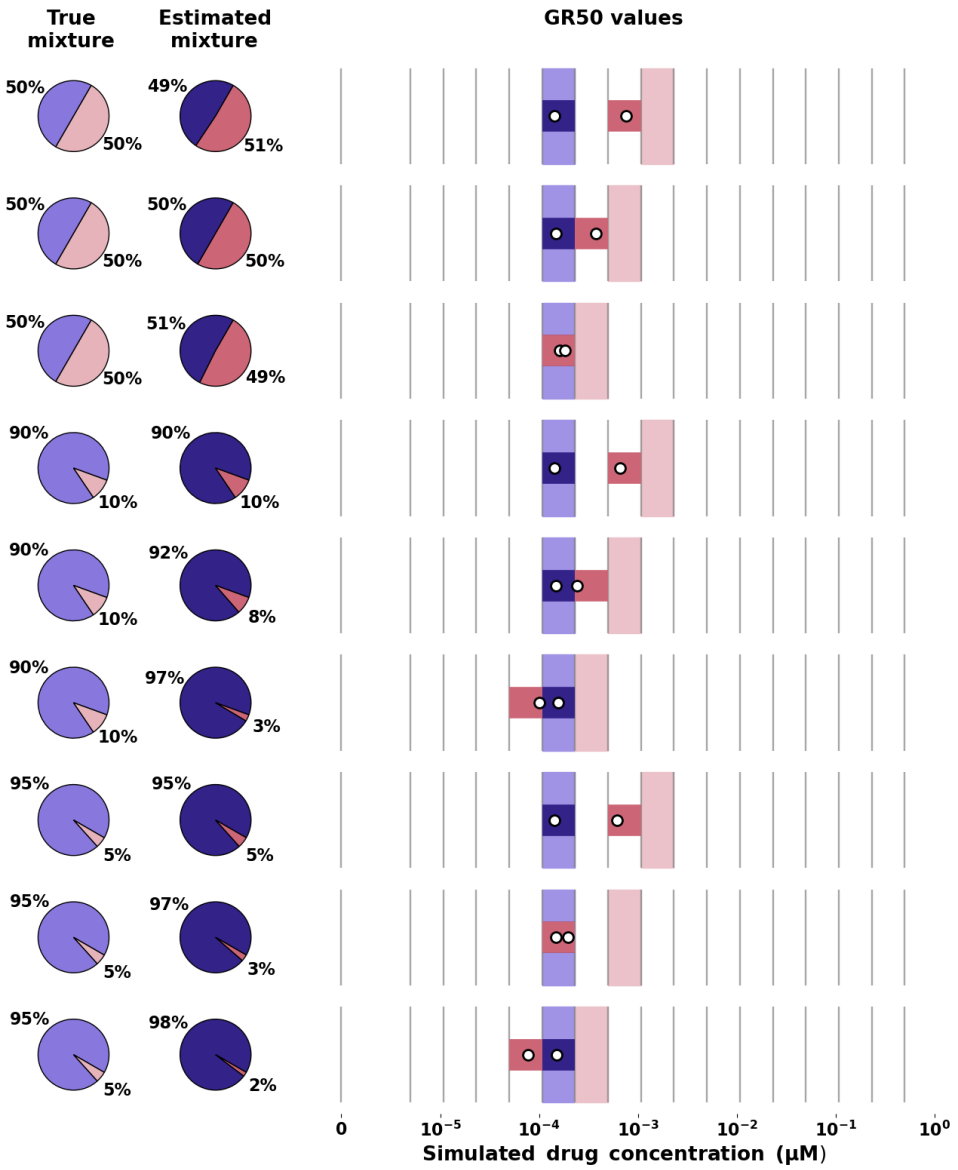

Figure S8: True and estimated mixture fractions and GR50 values for synthetic data with observation noise with standard deviation equal to 5% of the initial cell count. For each row, "True mixture" pie charts show mixture fractions used in the data generation; "Estimated mixture" pie charts show mixture fractions estimated by the model; vertical grey lines show observed concentrations on a logarithmic scale. In the "GR50 values" panel, the region between the two observed concentrations closest to each true GR50 value is given the same color as that subpopulation has in the "True mixtures" pie chart; White dots represent estimated GR50 values, with the region between the closest observed concentrations colored in the same color as the inferred subpopulation in the "Estimated mixture" pie chart.

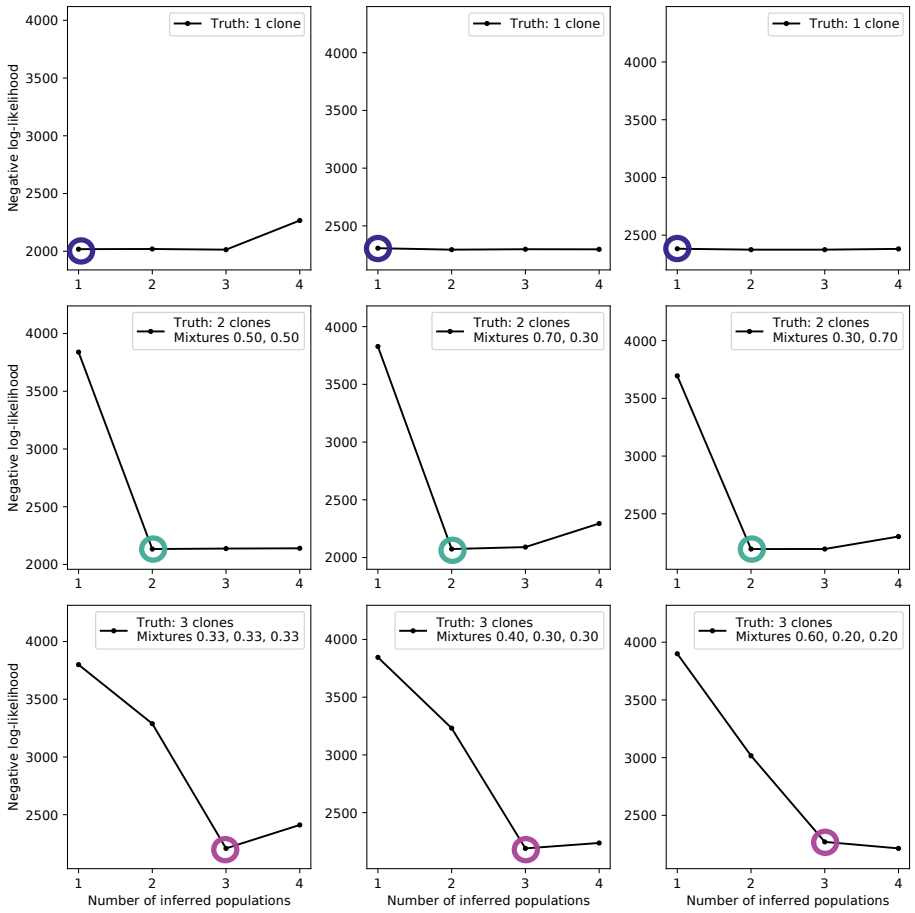

Figure S9: Elbow plots showing the negative log-likelihood for simulated data with observation noise with standard deviation equal to 1% of the initial cell count (Figure S3), with the selected model marked by a circle. The color of the circle also indicates the selected model: blue for 1 population, teal for 2, dark magenta for 3.

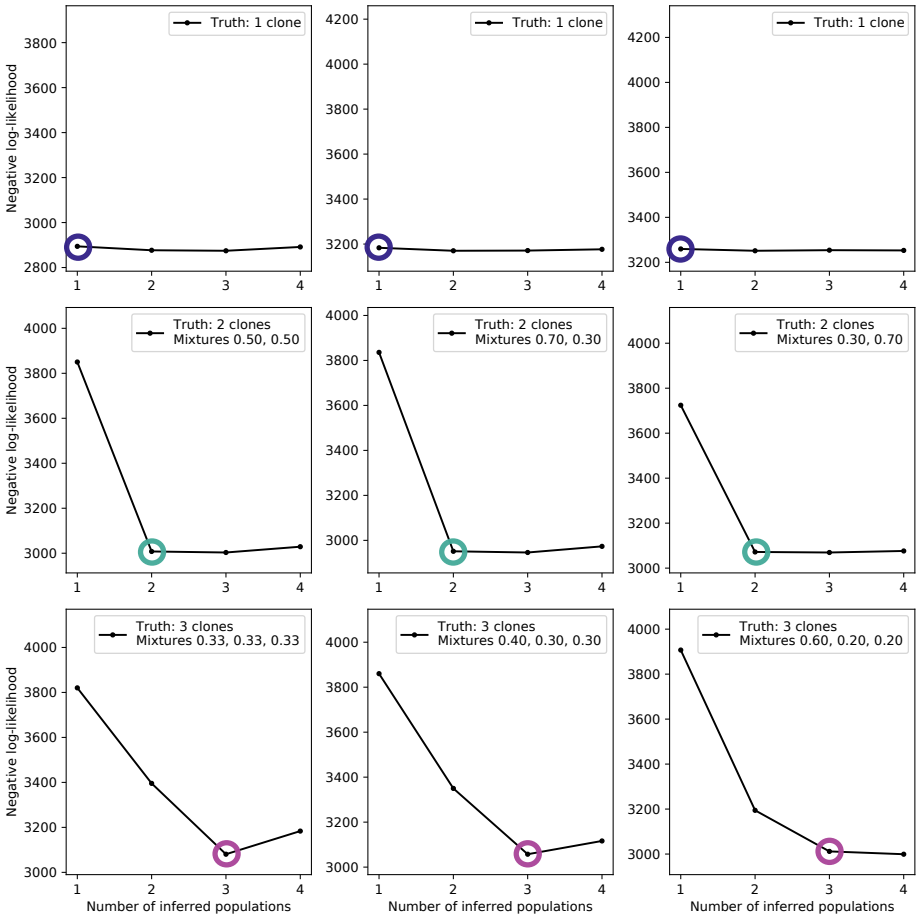

Figure S10: Elbow plots showing the negative log-likelihood for simulated cases with observation noise standard deviation equal to 5% of the initial cell count (Figure 2), with the selected model marked by a circle. The color of the circle also indicates the selected model: blue for 1 population, teal for 2, dark magenta for 3.

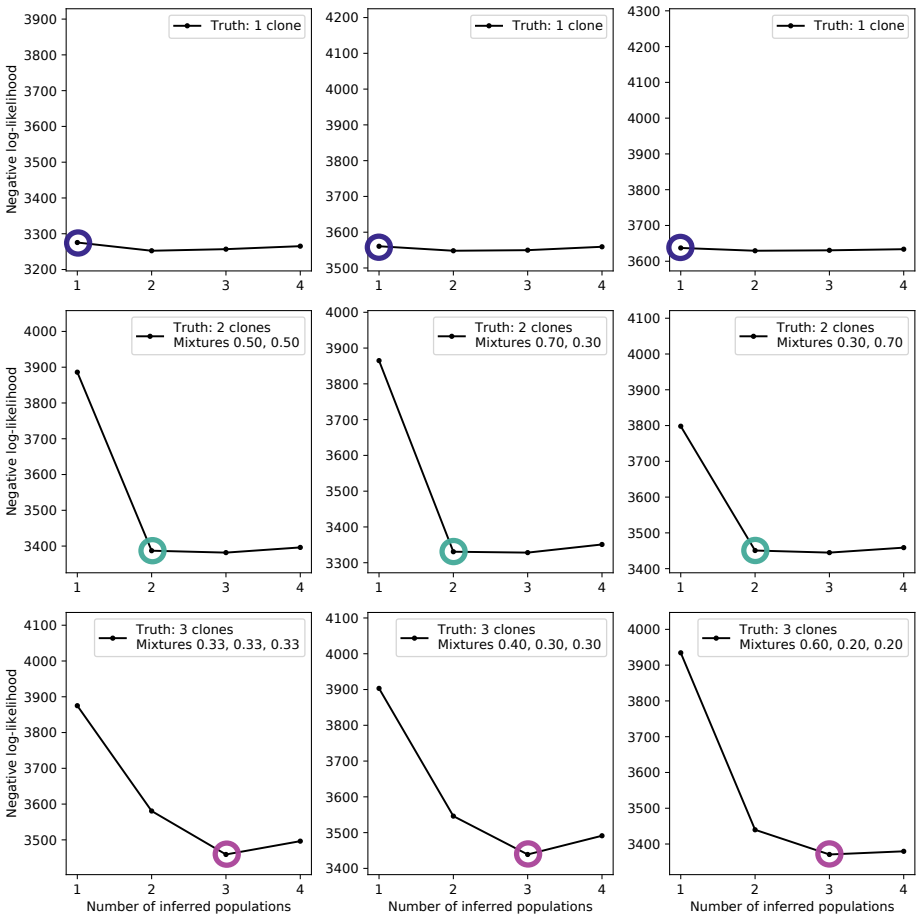

Figure S11: Elbow plots showing the negative log-likelihood for simulated cases with observation noise with standard deviation equal to 10% of the initial cell count (Figure S4), with the selected model marked by a circle. The color of the circle also indicates the selected model: blue for 1 population, teal for 2, dark magenta for 3.

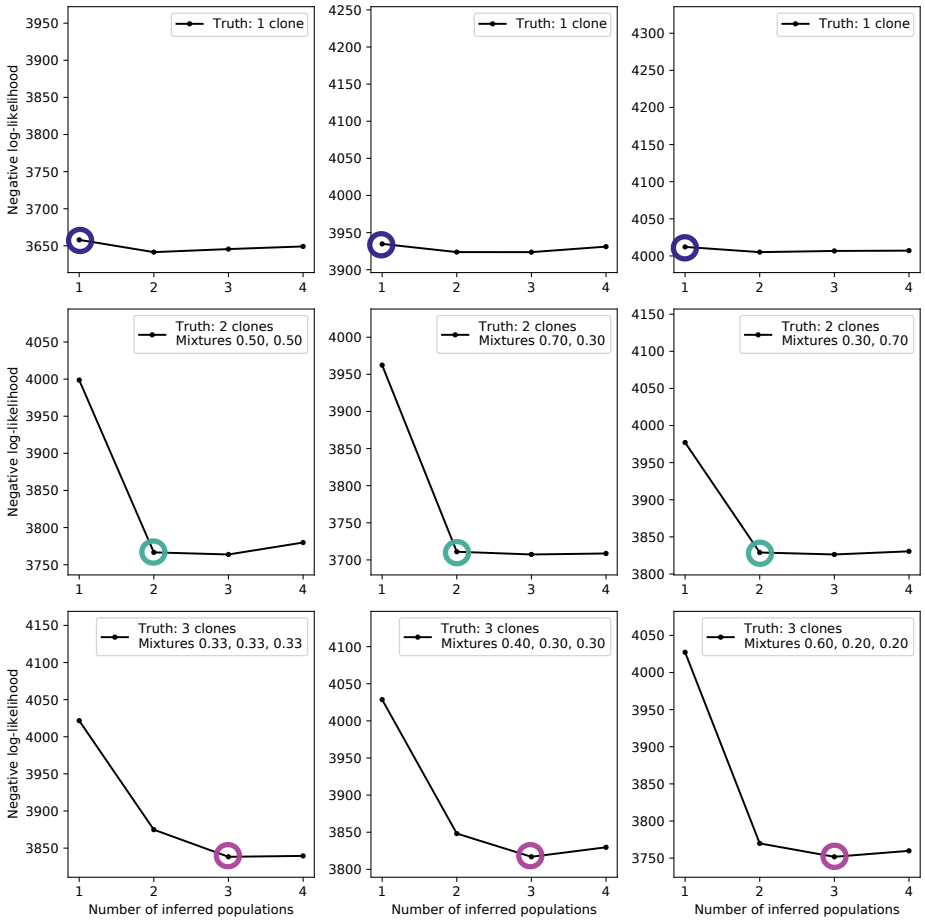

Figure S12: Elbow plots showing the negative log-likelihood for simulated cases with observation noise with standard deviation equal to 20% of the initial cell count (Figure S5), with the selected model marked by a circle. The color of the circle also indicates the selected model: blue for 1 population, teal for 2, dark magenta for 3.

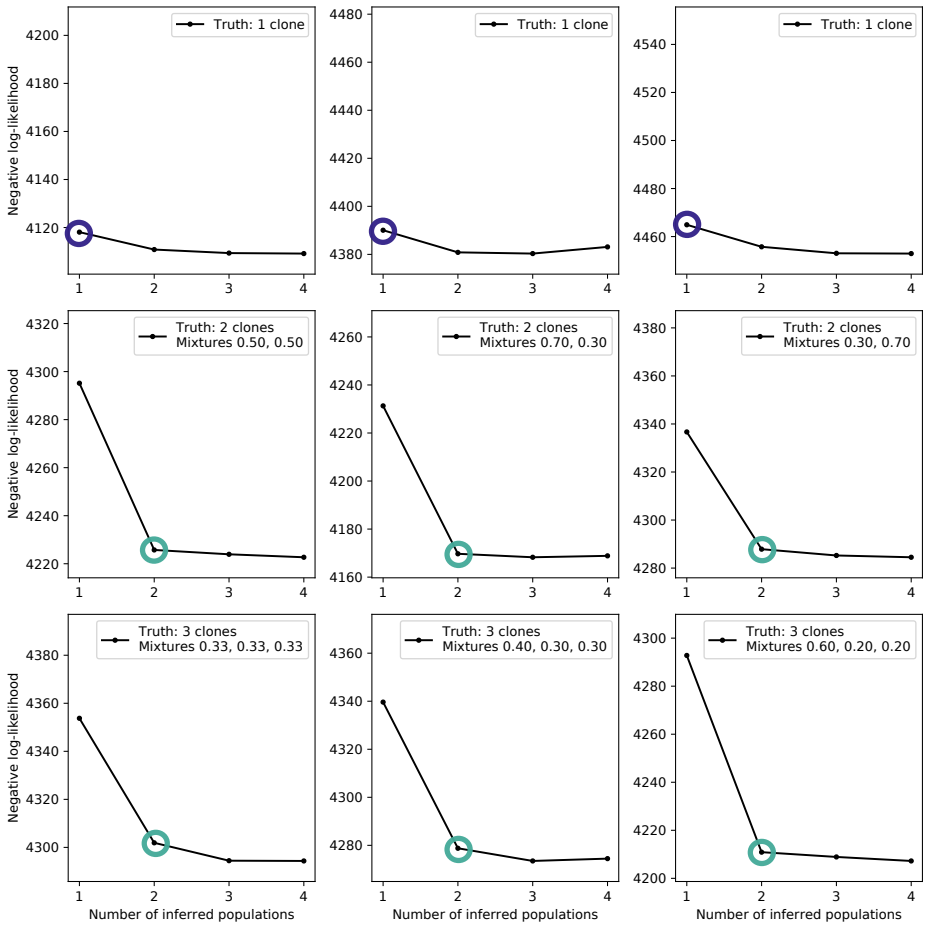

Figure S13: Elbow plots showing the negative log-likelihood for simulated cases with observation noise with standard deviation equal to 50% of the initial cell count (Figure S6), with the selected model marked by a circle. The color of the circle also indicates the selected model: blue for 1 population, teal for 2, dark magenta for 3.

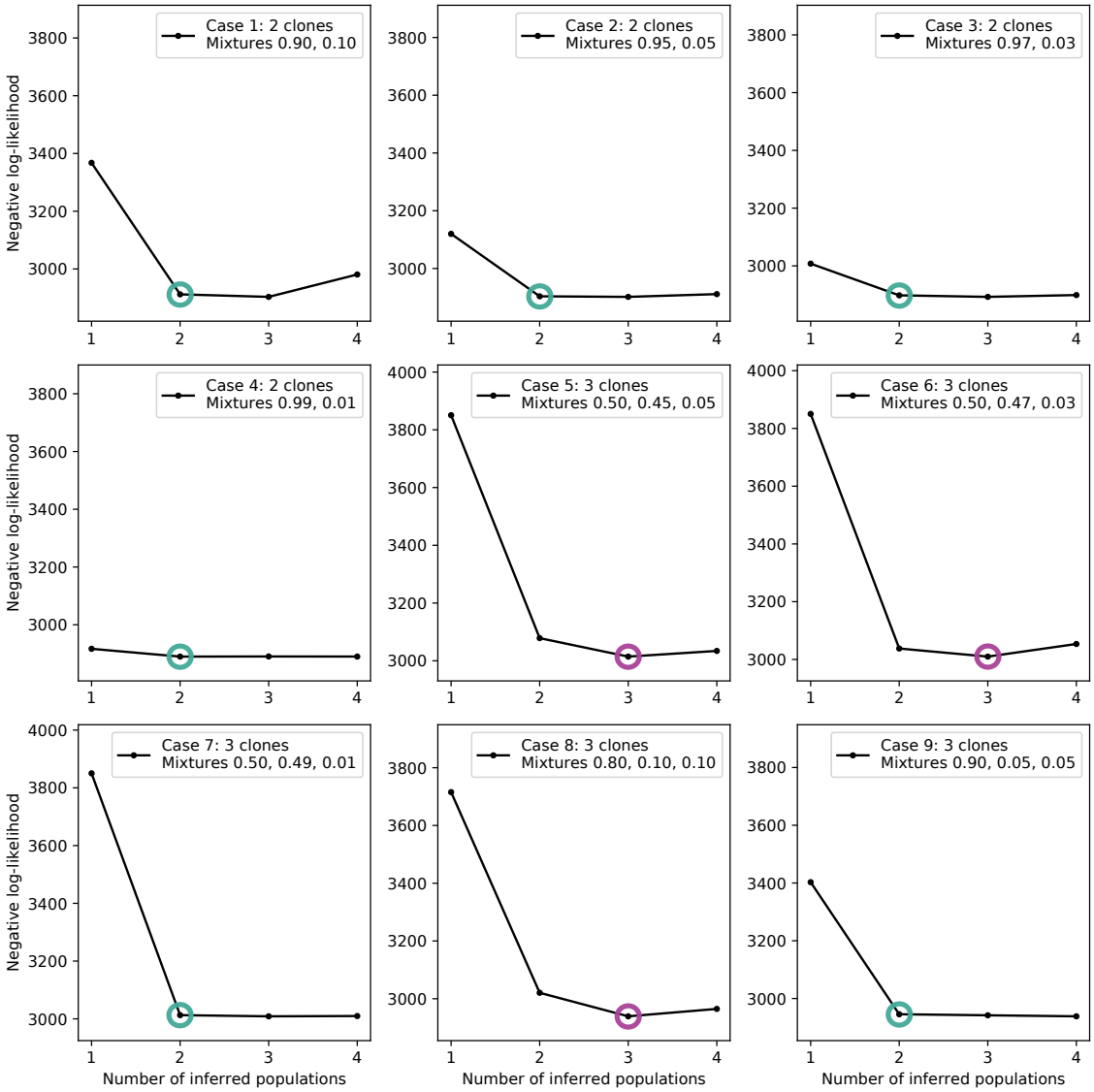

Figure S14: Elbow plots showing the negative log-likelihood for all cases in Figure S7, with the selected model marked by a circle. The color of the circle also indicates the selected model: blue for 1 population, teal for 2, dark magenta for 3.

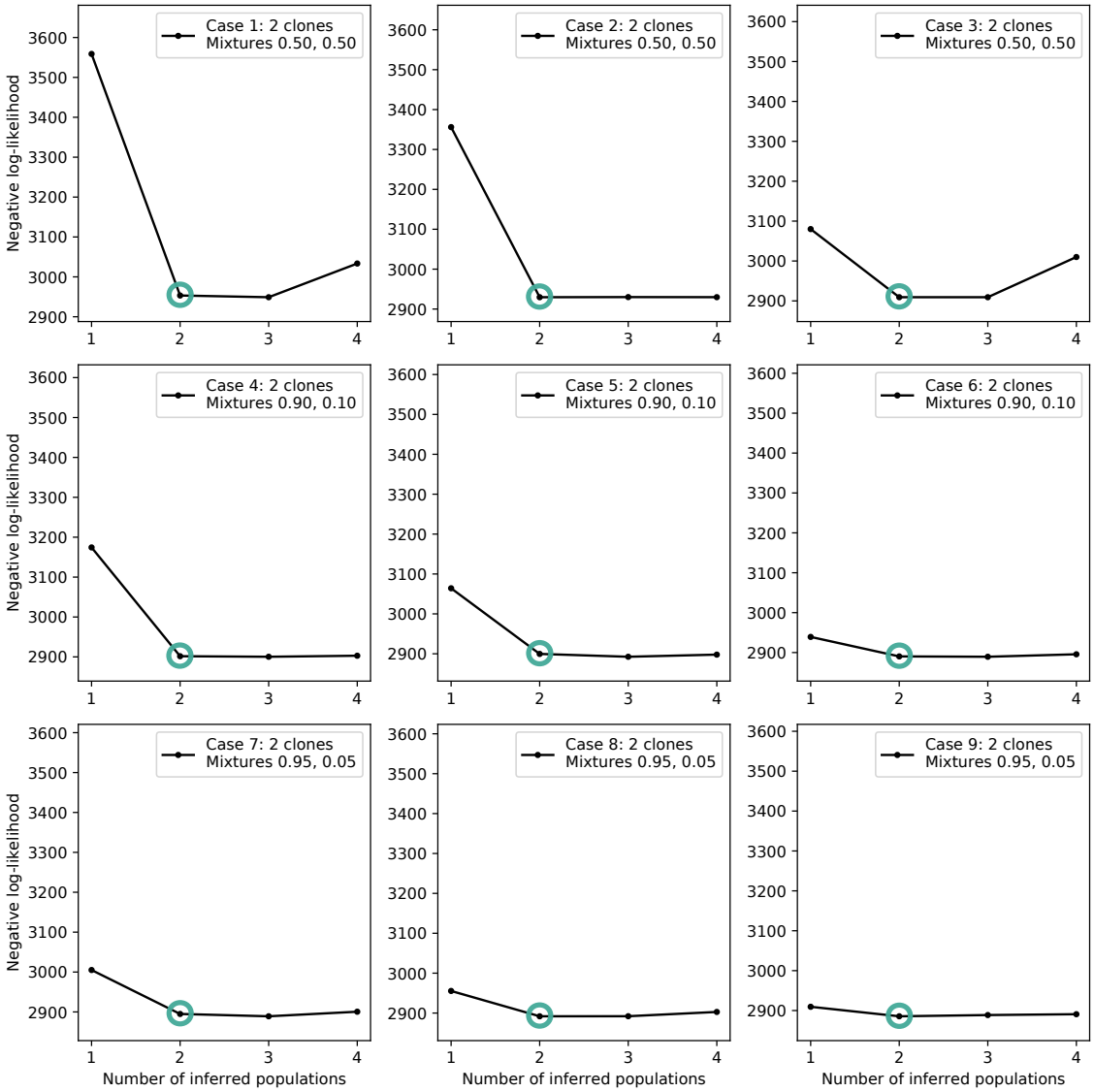

Figure S15: Elbow plots showing the negative log-likelihood for all cases in Figure S8, with the selected model marked by a circle. The color of the circle also indicates the selected model: blue for 1 population, teal for 2, dark magenta for 3.

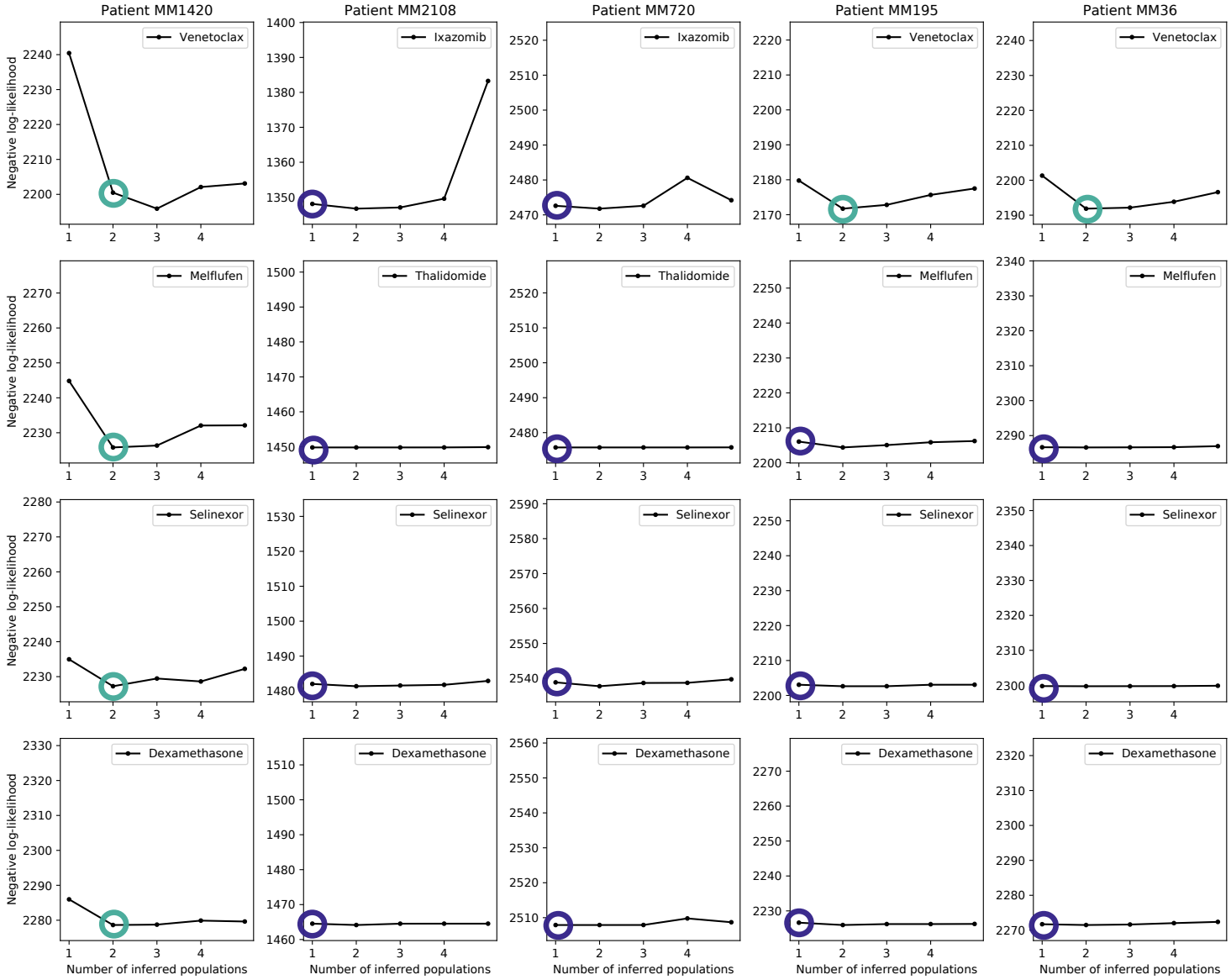

Figure S16: Elbow plots showing the negative log-likelihood for all drugs for all multiple myeloma patients (Figure 3), with the selected model marked by a circle. The color of the circle also indicates the selected model: blue for 1 population, teal for 2, dark magenta for 3.

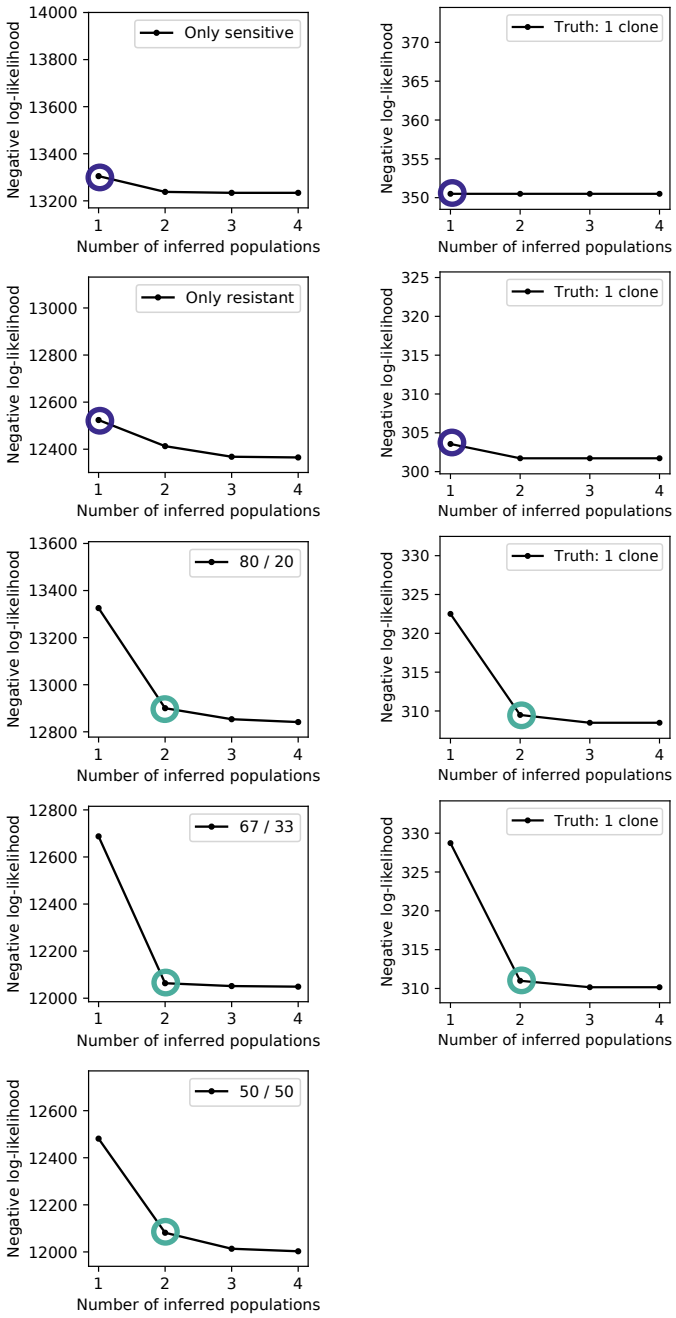

Figure S17: Elbow plots showing the negative log-likelihood values for all cases in Figure 2b (left) and 2c (right), with the selected model marked by a circle. The color of the circle also indicates the selected model: blue for 1 population, teal for 2, dark magenta for 3.