



Published in final edited form as:

Nat Genet. 2017 March ; 49(3): 332–340. doi:10.1038/ng.3756.

Precision oncology for acute myeloid leukemia using a knowledge bank approach

Moritz Gerstung^{1,2,*}, Elli Papaemmanuil^{1,3,*}, Inigo Martincorena¹, Lars Bullinger⁴, Verena I Gaidzik⁴, Peter Paschka⁴, Michael Heuser⁵, Felicitas Thol⁵, Niccolo Bolli^{1,6}, Peter Ganly⁷, Arnold Ganser⁵, Ultan McDermott¹, Konstanze Döhner⁴, Richard F Schlenk⁴, Hartmut Döhner^{4,§}, and Peter J Campbell^{1,8,§}

¹Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, UK ²European Bioinformatics Institute EMBL-EBI, Hinxton, UK ³Computational Oncology, Epidemiology and Biostatistics Memorial Sloan Kettering Cancer Institute ⁴Department of Internal Medicine III, Ulm University, Ulm, Germany ⁵Department of Hematology, Hemostasis, Oncology, and Stem Cell Transplantation, Hannover Medical School, Hannover, Germany ⁶Division of Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori, University of Milano, Italy ⁷Department of Pathology (UOC), University of Otago, Christchurch, New Zealand ⁸Department of Haematology, University of Cambridge, Cambridge, UK

Abstract

Causative mutations in a patient's cancer drive its biology and, by extension, its clinical features and treatment response, a concept underpinning the vision of precision medicine. However, considerable between-patient heterogeneity in driver mutations complicates evidence-based personalization of cancer care. Here, re-analysing 1,540 patients with acute myeloid leukemia (AML), we explore how large knowledge banks of matched genomic-clinical data can support clinical decision-making. Inclusive, multistage statistical models accurately predicted likelihoods of remission, relapse and mortality, validated on independent TCGA patients. Comparison of long-term survival probabilities under different treatments enables therapeutic decision support,

Addresses for correspondence: Dr Peter J Campbell, Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, Cambridgeshire, UK. pc8@sanger.ac.uk, Dr Hartmut Döhner, Department of Internal Medicine III, Ulm University, Ulm, Germany. Hartmut.Doechner@uniklinik-ulm.de.

*Joint first authors

§Joint corresponding authors

DATA AVAILABILITY STATEMENT

Sequencing data that support the findings of this study have been deposited in the European Genome-Phenome Archive (<http://www.ebi.ac.uk/ega>) under accession number EGAS00001000275. The clinical data and summarized driver mutation calls are available in a github repository (<http://www.github.com/gerstung-lab/aml-multistage>), together with all code used to generate the figures and conclusions of the manuscript.

URLs

The exploratory web application to visualize and explore the data is found at: <http://cancer.sanger.ac.uk/aml-multistage>. The git repository containing all clinical data, summarized mutation calls and code is available at: <http://www.github.com/gerstung-lab/aml-multistage>.

Author contributions: MG developed the statistical methods, analyzed data and wrote the manuscript and supporting information, with input from EP and PJC. EP prepared and curated genetic and clinical data. IM analyzed TCGA data. RS, HD, KD, LB, VIG, PP, MH, FT, AG alongside all contributing institutions to the study group (AMLSG) recruited patients in this study, collated and contributed clinical data. NB, PG and UM provided input into analyses and interpretation of results. EP, KD, HD, RFS and PJC initiated the study. PJC and HD wrote the manuscript, and are joint corresponding authors.

available in exploratory form online. Personally tailored management decisions could reduce the number of hematopoietic cell transplants in AML by 20–25%, while maintaining overall survival rates. Power calculations show that databases require thousands of patients for accurate decision support. Knowledge banks facilitate personally tailored therapeutic decisions, but require sustainable updating, inclusive cohorts and large sample sizes.

INTRODUCTION

Led by a small number of high-profile successes, there has been considerable enthusiasm for the concept of personally tailoring cancer management based on individual genomic profiles^{1,2}. Mutations in cancer genes fundamentally drive the tumor's growth, giving strong rationale for the belief that therapeutic choices made on the basis of these causative events will be biologically sound. Applications of genomics in cancer medicine include enhanced diagnostic accuracy through molecular characterization, personalized forecasts of a patient's prognosis and support for choosing among different therapeutic options^{3,4}. There are, however, complications to this narrative: surprisingly few cancer genes are straightforward therapeutic targets; many cancer genes are only rarely mutated in a given tumor type; each patient's tumor typically has several driver mutations. Above all other complications, though, is the challenge that, for most tumor types, there are hundreds to thousands of different combinations of driver mutations observed across patients^{5–7}.

The promise of precision medicine has triggered considerable funding commitments, such as the Precision Medicine Initiative in USA, Genomics England in UK and similar efforts in several other countries^{8,9}. Amongst other aims, these initiatives will build large banks of patients' genomic data matched to clinical variables, treatments and outcomes. Despite these investments reaching hundreds of millions of dollars in scale, there has been little formal evaluation of the potential utility of knowledge banks. In particular, it is unclear whether accurate predictions about cancer outcomes can be made from a large genomic-clinical database; what improvements in survival at the population level might be achieved from personally tailored therapeutic choices; and what sample sizes knowledge banks need to accrue before predictions are sufficiently accurate to underpin decision support for the individual patient. *Precision medicine* requires therapeutic decisions fine-tuned to the unique genome of an individual cancer; *evidence-based medicine* requires therapeutic decisions grounded on documented, verified data.

Here, we explore these questions by re-analyzing genetic data from 111 cancer genes, cytogenetic profiles, and clinical data from 1,540 patients with acute myeloid leukemia (AML) undergoing intensive treatment¹⁰, validated on an independent AML cohort from the TCGA¹¹. In our previous study¹⁰, we identified 11 genomic subcategories of AML, each with distinctive constellations of clinical features. However, even within individual molecular subgroups, there remains considerable patient-to-patient variability in treatment response and clinical outcomes, partially explained by co-operating driver mutations and other diagnostic clinical variables. At the population level, then, we can make strong statements about overall patterns of long-term survival from such data. At the level of a patient in the clinic faced with a difficult therapeutic decision, however, it is not at all clear

how such genomic complexity impacts on the accuracy or relevance of predictions about potential clinical outcomes for that patient.

AML presents an interesting exemplar for evaluating the potential of precision medicine because of a real, current therapeutic dilemma – who should be offered an allogeneic hematopoietic cell transplant (allograft) in first complete remission (CR1)^{12,13}? The equations are not straightforward. Allogeneic hematopoietic cell transplants in first complete remission undoubtedly decrease relapse rates for most patients, but this comes at the cost of higher treatment-related mortality, as high as 20–25% at 3 months¹⁴, with a further 30% risk of debilitating chronic morbidity¹⁵. Furthermore, even though more patients relapse after chemotherapy in first remission, up to a fifth can then be successfully salvaged with allografts or more intensive chemotherapy^{16,17}. We use this particular therapeutic dilemma to illustrate how a knowledge bank approach can inform therapeutic decisions tuned to the specifics of an individual patient, a concept that could be extended to other cancers, other treatments, other clinical conundrums.

RESULTS

Predicting complex patient outcomes from genomic and clinical variables

We recently sequenced¹⁰ all coding exons of 111 myeloid cancer genes in diagnostic leukemia samples from 1,540 AML patients undergoing intensive treatment within three prospective clinical trials of the German-Austrian AML Study Group (AMLSSG). We identified driver point mutations and combined these data with the clinical trials database to generate a comprehensive knowledge bank. Here, we focus on evaluating the utility of the knowledge bank for generating predictions personally tailored to the individual patient, and how this can be used to compare likelihoods of various clinical outcomes under different treatment strategies. The full knowledge bank, together with all analysis code used here, is documented in the Supplementary Note and is available as a git repository (see URLs section below).

Throughout, we use overall survival as the primary end-point of these analyses since the aim of intensive therapy in the young AML patient is cure. The full dataset consists of 231 predictor variables, spanning the seven broad categories of fusion genes, copy number alterations, point mutations, gene-gene interactions, demographic features, clinical risk factors and treatment received, across 1540 patients. To assess the accuracy of our predictions we use the following validation strategies: (1) random cross-validation on this dataset, (2) building models from any two clinical trials here and testing on the third; and (3) testing the model built from all three AMLSSG trials on an independent AML cohort from USA (TCGA)¹¹. All predictions for individual patients reported here were made using models excluding that patient.

We tested a range of regularized regression methods for predicting survival and also implemented more novel approaches, based on random effects and multistage statistical models, for deriving detailed associations between genomic and clinical endpoints (Figure 1a; Supplementary Note sections 2–3). Using a variety of accuracy measures, the random effects models and multistage models typically scored best in predicting overall survival,

roughly doubling the amount of explained variance compared to current prognostic criteria¹³ (Figure 1b–c; Supplementary Note section 4). A key aspect of these approaches is that they include all available variables in the model, but shrink their estimated effects if there is only weak support in the data in order to control overfitting. In contrast, conventional methods typically chose reduced subsets of 5–20 variables, seemingly at the cost of discarding prognostically relevant information (for more discussion, see Supplementary Note section 4).

Reassuringly, we found strong ‘out-of-cohort’ validation for our models, either when models built on this cohort were tested on the TCGA cohort, or when models using two of the three trials in the knowledge bank were tested on the third trial (Figure 1a). Of particular note is the observation that the concordance decreased only moderately for predictions from a model trained on younger patients (AMLHD98A and AMLSG0704: age range, 18–65 years) evaluated on a trial of older patients (AMLHD98B: range, 58–84 years). This implies that many of the differences between age groups in AML outcomes are captured in clinical and genetic variables and can therefore be learnt from the knowledge bank.

The multistage model offers the advantage of separating long-term outcomes into individual constituents – death without complete remission, non-relapse death (mostly treatment-related) and death after relapse; as well as survival in induction, first remission (CR1) and after relapse (Figure 2a–c). As we shall see, understanding which of these constituent outcomes is especially likely for a patient considerably enhances therapeutic decision-making. The added detail does not come at the cost of overfitting, since the combined prediction of overall survival in the multistage model yields the same accuracy as predicting overall survival directly (Figure 1a).

Personally tailored prognosis

The models for predicting outcome developed here are considerably more complex than those currently used in clinical practice. In AML, the current standard is the European LeukemiaNet (ELN) genetic scoring system¹³, which defines four categories of disease risk based on 6 fusion genes, 3 point mutated genes and cytogenetic abnormalities. We explored how much more informative our more complex prognostic models are than the ELN system.

We find that individual risk in this AML cohort was continuous, with no obvious cut-points for stratification, suggesting that grouping patients on the basis of few predictor variables discards much prognostic information (Figure 2d). Our more detailed survival estimates confirm the broad trends of known ELN risk groups, but a third of patients have survival predictions deviating more than 20% from their ELN stratum (Figure 2e).

From the multistage model, we can quantify how much the various classes of predictor variables contribute to explaining patient-to-patient variation in each possible endpoint of treatment (Figure 2f, Supplementary Tables 1–6). We find that clinical and demographic factors, such as patient age, performance status and blood counts, exerted most influence on early death rates, including death in remission (mostly due to treatment-related mortality). Genomic features, be they copy number changes, fusion genes or driver point mutations, most strongly influenced the dynamics of disease remission and relapse.

These estimates represent the contributions of the various categories of predictors to outcomes of treatment *at the population level*. At the individual level, we can score each patient for his or her risk along these dimensions of predictor variables. What emerges is the considerable heterogeneity in personal risk profiles across the cohort (Supplementary Figure 1). The heterogeneity of risk profiles and the variable impact they have on the different AML outcomes combine to generate a kaleidoscope of predictions for patients' journeys through therapy (Figure 3). Thus, there are distinct groups of patients for whom we can confidently predict long-term survival in first remission, or death after relapse, or death without achieving remission, manifesting as swathes of purple, green or pink respectively in Figure 3. Reassuringly, these predictions square well with what actually happened to the patients (status lines and circles in Figure 3). It is these patients for whom the personally tailored predictions have much confidence. There are, however, some patients for whom there is genuine uncertainty about outcomes even with the full model. These patients have predicted survival curves that deviate little from the population average.

Personally tailored therapeutic decision support

The preceding sections show that a knowledge bank can provide meaningful information about a patient's prognosis. The goal of precision medicine is more ambitious than this, however, in seeking to inform the choice of therapy for an individual patient. In AML, a major therapeutic dilemma is deciding which patients should be offered allogeneic hematopoietic cell transplants (allografts), and whether this should be in first complete remission or after relapse^{12,13}. With 20–25% transplant-associated mortality and substantial rates of chronic graft-versus-host disease, allografts tend to be reserved for high-risk patients. We now explore how a knowledge bank can inform the decision on allograft versus chemotherapy in first remission for the individual patient with AML.

Our calculations have shown that using a knowledge bank to model patient outcomes reclassifies the risk estimates of a substantial fraction of patients (Figure 2e). A given patient's risk prediction represents an aggregation across multiple facets of the disease. Thus, two patients can both have an overall intermediate probability of death but arrive at this through different risk contributions: one might be older and more frail but have a leukemia with generally favorable genomic features; the other might be young and fit but with a leukemia carrying adverse driver mutations. Intuitively, a clinician will favor the more intensive allogeneic transplant in the latter, fitter patient while preferring standard chemotherapy in the older patient at higher risk of treatment-related mortality.

We illustrate these calculations using two patients from the cohort (Figure 4; other representative patients illustrated in Supplementary Figure 2). The first was a 29-year woman with t(8;21) and no other driver mutations: favorable risk by ELN criteria¹³. Under a strategy of chemotherapy in CR1 with salvage allograft after relapse, we predict her chance of 3-year survival to be 86% (CI_{95%}: 78–91%) (Figure 4a). In contrast, with allograft in CR1, we estimate her overall cure rate to be 88% (79–93%) (Figure 4b), with the decrease in probability of relapse matched by the increase in non-relapse mortality with transplant. Hence there is no indication for an up-front allograft for this patient, with only 2 percentage points difference in predicted survival (CI_{95%}: –3 to 7). For this patient, the treatment

recommendation under current clinical standards¹³ is unchanged under a knowledge bank approach.

The second patient was a 49-year old male with mutations in *NPM1*, *DNMT3A*, *IDH1* and normal karyotype. Under ELN criteria, his risk also classifies as favorable, and he would not currently be recommended for allograft in first CR. With standard chemotherapy as first-line therapy, we estimate his 3-year survival probability at 55% (41–67%), compared to 68% (55–77%) for allograft in CR1 (Figure 4c–d). Thus, his disease is not especially favorable risk when all predictive information is considered. Furthermore, the absolute risk reduction associated with an up-front allograft is estimated at 13 percentage points (3–24%). This is equivalent to curing 1 additional patient for every 7 (4–26) treated with allograft instead of standard chemotherapy in first remission.

Treatment choices from knowledge banks versus current practice

The cases shown in Figure 4 illustrate that some, but not all, patients would have had their treatments changed using a knowledge bank compared to current recommendations. It is therefore natural to assess how many patients would have had their treatment altered under such an approach, and whether the predictions accurately reflect what actually happened to the patients.

On average, we find that patients who are predicted to have poor prognosis, more than 60–70% chance of mortality at 3 years, are most likely to benefit from allogeneic transplantation in first remission (Figure 5a), a finding captured in current clinical recommendations. However, there is considerable spread of patient estimates around the population average. This variance around the average is critically important for precision medicine because it suggests that population-based criteria for treatment choices only poorly capture the predictive information available for the individual patient.

Overall, we estimate that 12% (124/995) of patients in CR1 aged 18–60 years would gain more than 10 percentage points improvement in survival at 3 years with an allograft in CR1 compared to standard chemotherapy (number needed to treat <10; Figure 5b). Only 29 of these 124 patients are identified as adverse risk by current criteria, with most being intermediate and some even favorable risk. Furthermore, 57% (302/534) patients classified as adverse or intermediate risk by ELN criteria, and therefore strongly considered for allograft in CR1 under current clinical recommendations¹³, are predicted to derive <5 percentage points improvement in survival from up-front allografts. Similarly, 15% (58/386) ELN favorable patients are predicted to benefit >5% from a bone marrow transplant in CR1. Clearly, then, a knowledge bank approach might change management in up to 1/3 of patients compared to current practice recommendations.

We next compared the therapeutic predictions made by our model with what actually happened to the patients under the two different treatment strategies (Figure 5c; Supplementary Figure 3). We split the cohort into two groups depending on whether they were predicted to derive more or less than 10 percentage points improvement in survival with allograft in CR1 compared to chemotherapy in CR1, allograft after relapse. If our model were correctly identifying those patients most likely to benefit from transplant, then

the survival curves in this group should show distinctly better outcomes for allograft in first remission than for chemotherapy. This is indeed what we observe (blue lines, Figure 5c). For the patients our model predicts minimal or no benefit from an up-front transplant strategy, we do indeed find that there was little meaningful difference in the survival curves between those receiving transplant and those receiving chemotherapy in first remission (grey lines, Figure 5c).

Taken together, then, these data demonstrate that up to a third of patients would have their treatment altered using a knowledge bank approach compared to current practice recommendations¹³. Furthermore, predictions made from the knowledge bank match well with the actual outcomes observed under the two different treatment philosophies, confirming the accuracy of the decision support.

Population impact of a knowledge bank approach

Knowledge banks would be costly to build and maintain, and it is therefore important to evaluate whether the overall impact of improved treatment choices at the population level would justify this outlay. The impact in AML could be expressed as either the improvement in expected survival for a fixed number of allografts in CR1 or the reduction in the number of allografts in CR1 needed to achieve the same overall survival (Figure 5d). In USA, ~30% of patients with AML receive an allograft¹⁸. If the 30% to receive an allograft in CR1 were chosen using an optimal knowledge bank rather than current recommendations, we estimate survival rates across the cohort would increase ~1.3 percentage points (60% to 61.3%).

Alternatively, personally tailored management decisions could reduce the number of hematopoietic cell transplants in AML by 20–25%, while maintaining overall survival rates. Under current practice, 44% young adult patients would receive a transplant, broken down as 30% in CR1 plus 14% post-relapse. In contrast, using a knowledge bank approach to choose when and whom to transplant, 35% patients overall would receive an allograft, as 16% in CR1 plus 19% post-relapse, achieving the same overall survival rate of 60%. Similar overall gains from a knowledge bank approach were found across a range of assumptions for risks and benefits of transplant (Supplementary Figure 4).

We can express the impact of a knowledge bank at the population level in terms of quality-adjusted life years (QALYs). Health utilities for AML survival with and without stem cell transplants have previously been estimated as 0.74 and 0.83, respectively¹⁹, and the cost of an allograft as US\$100,000–200,000²⁰. Thus an increase of 1.3 percentage points in long-term survival while maintaining a 30% allograft rate in CR1 corresponds to ~0.1 QALYs gained per patient over ten years. Alternatively, reducing the number of allografts by better resource allocation, while maintaining overall survival rates, would gain ~0.05 QALYs per patient over ten years as well as saving approximately US\$10,000 per patient.

Portals for exploring decision support predictions

The preceding sections demonstrate that the complex and multifactorial inter-relationships among genomic variables, clinical predictors and cancer outcomes can be learnt with a sufficiently comprehensive knowledge bank. Since the underlying survival models are

complex, diagnostic laboratories may need to provide personalized portals into a given patient's cancer genome.

Our dataset is not appropriate for direct clinical use, as the algorithm has not yet been prospectively validated and sequencing was performed using a research pipeline. Nonetheless, as a research tool, we have created a prototype portal within our website²¹ (see URLs section below) that allows outcome predictions to be generated based on this dataset for user-defined constellations of genomic features, clinical variables and treatment strategies (Supplementary Figure 5). The underlying algorithm is capable of imputing missing variables and computes confidence intervals for each prediction.

The knowledge bank

We explored how both the breadth of genomic profiling and the sample size of the knowledge bank impact on the accuracy of outcome predictions for individual patients. The explained risk grows linearly with the *average number of driver mutations present in each patient* (Supplementary Figure 6a), a relationship underpinned by theoretical arguments (Supplementary Note S5.3.2). Some genes, by virtue of their frequency and/or the magnitude of their prognostic effect, are more informative than others. We have ranked AML genes by their predictive utility (Supplementary Figure 6b) to address the question of how much improvement in prognostic information comes from increasing the number of genes interrogated. The effects of missing mutation data on confidence intervals of patient prediction can be explored in the web portal.

The other critical factor to accurate risk profiling is the sample size of the knowledge bank. Using subsampling analyses and simulations from the AML data, we found that prognostic accuracy steadily increases with larger sample sizes, albeit following a law of diminishing returns (Figure 6a). As a rule of thumb, to detect a moderate-sized prognostic effect of a given cancer gene, say an increase of 50% in relative risk, the knowledge bank needs ~50–100 patients with that mutation (Figure 6b; Supplementary Figure 7a). Thus, for a gene mutated in 10% of patients, a training set of 500–1000 patients would suffice, but for a 1% gene, a cohort of 5,000–10,000 would be needed. These simulations match theoretical expectations^{22,23} (Supplementary Note S5.4.2; Supplementary Figure 7b).

The standard error of individual survival predictions 3 years after CR is about 6%. When using predictions for supporting therapeutic decisions about a specific patient, this uncertainty limits the ability to confidently discriminate small differences in survival. With 1,000 cases, we could achieve an average absolute prediction error for an individual patient of approximately 5 percentage points, which could be brought down to 2 percentage points with 10,000 cases (Figure 6c).

DISCUSSION

Here, we have evaluated the promise of precision medicine, building statistical models that can generate personally tailored clinical decision support from all available prognostic information in a knowledge bank. From a database of 1540 patients, we can make considerably more informative and more accurate statements about an individual's likely

journey through AML therapy than the current standards in clinical practice. Our approach enables us to compare the likelihood of favorable outcomes under different treatment scenarios, providing information that can support genuinely personalized decision-making.

While we have focused on AML in this analysis, we believe that the same logic applies to knowledge banks from other cancer types, which will be generated as genomics enters healthcare and healthcare becomes digitized. Most cancer types are lethal, and most currently available treatment options are either invasive or toxic, burdening the patient with severe side-effects. Therefore a quantitative risk assessment is important in any cancer type in order to reserve the most aggressive treatments for the patients at the highest risk of dying from the disease. All cancers are caused by genetic changes, with considerable heterogeneity among patients, and it is therefore likely that these genetic differences also correlate with differences in outcome, although the details of the logic and strength of association may vary among cancer types. Once knowledge banks are established and ideally populated with information about different treatment options, whether these be chemotherapy, targeted inhibitors or immunotherapies, one can apply the logic outlined here to assess the benefit of these treatments, contrasted with the patient's baseline risk.

Building and maintaining clinical-genomic knowledge banks is a formidable challenge, especially for solid tumors where the genome can be considerably more complex than AML. Initially, knowledge banks could be seeded from clinical trials cohorts, as we did here, since these will have high quality clinical data and state-of-the-art therapies. Our power calculations suggest, however, that most clinical trials would not be powered to detect gene-drug interactions involving genes mutated in <20% patients. Additionally, knowledge banks will need to include patients who are representative of the wider cancer population to enable meaningful extrapolation to real-world clinical practice. This suggests that building systems to incorporate data from patients undergoing routine clinical care into the knowledge bank will be important.

Whether the returns justify this investment will be contentious. Here, we have illustrated that a reallocation of allografts could increase survival by 1.3 percentage points. We should not be surprised at how modest the gain is – for the bulk of patients, we predict only small improvements in survival with early allografts (Figure 5b). What may be more important is the more accurate use of a precious resource, since we can potentially reduce the number of allografts performed in AML by 20–25% while maintaining the same overall survival as for current treatment recommendations. Hence a knowledge bank would not only increase the quality of life by reducing morbidity from chronic graft-versus-host disease; at US \$100,000–200,000 per allograft²⁰, the potential monetary savings would far outstrip the costs of the genomic screens. Moreover, the utility of a knowledge bank likely goes beyond this, informing potential drug targets, identifying patients not benefitting from current treatments and providing insights into the relationships between genetic and clinical features.

There is a tension between maintaining the precepts of *evidence-based medicine* while sharpening the focus on the individual with *precision medicine*²⁴. Here, we have demonstrated how knowledge banks can resolve this tension, using the evidence base of

thousands of patients to inform outcomes for the individual. The therapeutic choice we exemplified is binary: transplant versus chemotherapy in AML. The success of FLT3 inhibitors²⁵ potentially squares the number of available treatment options, and other novel agents will add further complexity. Knowledge banks could be a useful tool for clinicians navigating this complexity, but must remain evergreen as the therapeutic armamentarium expands and as our molecular understanding of cancer deepens. The logistic and regulatory hurdles, the scale needed and the costs of such an undertaking are daunting but not insurmountable.

ONLINE METHODS

Patient cohort

Here we reanalyze data first reported and described in detail in Ref.¹⁰. Briefly, we performed targeted gene sequencing of 111 myeloid cancer genes^{11,26–28} on DNA from leukemic cells in a cohort of 1,540 adults with AML who were treated with intensive therapies in three clinical trials run by the German-Austrian AML Study Group^{29–31}. In AML-HD98A, patients aged 18–61 years received induction chemotherapy with idarubicin, cytarabine and etoposide (ICE), followed by allogeneic transplants for intermediate-risk patients with matched related donor and high-risk patients; intensive consolidation chemotherapy for the remainder. Treatments were similar in AMLSG-07-04 but included randomization for all-*trans* retinoic acid therapy (ATRA) or not in induction. In AML-HD98B, patients > 61 years received ICE±ATRA, with further therapy dictated by response. Median follow-up was 5.94 years. All patients gave written informed consent for enrolment in the multicentre trials, which were approved by the local research ethics committees in participating sites ([ClinicalTrials.gov](https://clinicaltrials.gov) number: NCT00146120).

Statistical Methods

We explored a range of statistical methods to build models of overall survival^{32,33}, including random survival forest regression; stepwise Cox proportional hazards model selection with either AIC or BIC penalty; complementary pairs stability selection based on LASSO penalized Cox proportional hazards models; random effects models with Gaussian random effects/ridge penalties; and random effects multistage models (Supplementary Methods S2–4). We found little prognostic significance to whether mutations were subclonal or clonal (Figure S8), and therefore do not consider this information in the multivariate models. All predictions shown are based on a leave-one-out basis; it is therefore informative to compare each prediction with the observed outcome in a given patient. All predictions for individual patients reported here were made using models excluding that patient.

For estimating the population-level impact of the knowledge bank approach, we divide patients into two groups based on whether they are expected to derive more or less than 10 percentage points improvement in survival with allograft in CR1 compared to chemotherapy in CR1, allograft at relapse. In each group, the observed outcomes are then determined separately for those who actually received an allograft in CR1 and those who proceeded with standard chemotherapy in CR1. In the ideal knowledge bank, the treatments used would be randomized, since this ensures they are not confounded with the predictor variables we use.

Here, 711/1540 (46%) patients received an allograft, with the decision to perform a transplant in intermediate-risk patients based on whether a matched related donor was available³¹. This introduces a quasi-randomization, since HLA-matching between siblings derives from Mendelian assortment of parental alleles, but this cannot substitute for prospective validation of the decision support tools we develop.

All predictions for individual patients reported here were made using models excluding that patient. To maximize reproducibility, details of statistical methods and all analysis code used are provided in Supplementary Methods and as a git repository online.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Chris Holmes for stimulating discussions. This work was supported by grants from the Wellcome Trust (077012/Z/05/Z), the Bloodwise charity and the Leukemia-Lymphoma Society. PJC has a Wellcome Trust Senior Clinical Research Fellowship (WT088340MA). EP is supported by an EHA early career fellowship. Supported in part by grants 01G19981 and 01KG0605 from the German Bundesministerium für Bildung und Forschung (BMBF), grant 109675 from the Deutsche Krebshilfe, and by project B3 and B4, Sonderforschungsbereich (SFB) 1074 funded the Deutsche Forschungsgemeinschaft; HD is coordinating investigator of SFB 1074; LB is a Heisenberg Professor of the DFG (BU 1339/3-1). We gratefully acknowledge Daniela Weber for clinical data managing, Veronica Teleanu for assistance in cytogenetics data classification, and Dr. Sabine Kayser for assistance in morphologic evaluation. We are grateful to all members of the German-Austrian AML Study Group (AMLSG) for their participation in this study and providing patient samples; a list of participating institutions and investigators appears in the Appendix to the companion paper. AMLSG treatment trials were in part supported by Amgen and DKH Grant: 109675.

References

1. Van Allen EM, et al. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat. Med.* 2014; 20:682–8. [PubMed: 24836576]
2. Mardis ER. Genome sequencing and cancer. *Curr. Opin. Genet. Dev.* 2012; 22:245–50. [PubMed: 22534183]
3. Garraway LA, Lander ES. Lessons from the cancer genome. *Cell.* 2013; 153:17–37. [PubMed: 23540688]
4. McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. *N Engl J Med.* 2011; 364:340–350. [PubMed: 21268726]
5. Papaemmanuil E, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood.* 2013; 122:3616–3627. [PubMed: 24030381]
6. The Cancer Genome Atlas Research Network. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012; 490:61–70. [PubMed: 23000897]
7. Imielinski M, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell.* 2012; 150:1107–1120. [PubMed: 22980975]
8. Manolio TA, et al. Global implementation of genomic medicine : We are not alone. *Sci. Transl. Med.* 2015; 7:1–8.
9. Collins FS, Varmus H. A New Initiative on Precision Medicine. *N Engl J Med.* 2015; 372:793–795. [PubMed: 25635347]
10. Papaemmanuil E, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* 2016; 374:2209–2221. [PubMed: 27276561]
11. The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013; 368:2059–2074. [PubMed: 23634996]

12. Gale RP, Wiernik PH, Lazarus HM. Should persons with acute myeloid leukemia have a transplant in first remission? *Leukemia*. 2014; 28:1949–52. [PubMed: 24727674]
13. Döhner H, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*. 2010; 115:453–74. [PubMed: 19880497]
14. Koreth J, et al. Allogeneic stem cell transplantation for acute myeloid leukemia in first complete remission: systematic review and meta-analysis of prospective clinical trials. *JAMA*. 2009; 301:2349–61. [PubMed: 19509382]
15. Flowers MED, Martin PJ. How I treat chronic graft-versus-host disease. *Blood*. 2015; 125:606–615. [PubMed: 25398933]
16. Burnett AK, et al. Curability of patients with acute myeloid leukemia who did not undergo transplantation in first remission. *J. Clin. Oncol.* 2013; 31:1293–301. [PubMed: 23439754]
17. Schlenk RF, et al. The value of allogeneic and autologous hematopoietic stem cell transplantation in prognostically favorable acute myeloid leukemia with double mutant CEBPA. *Blood*. 2013; 122:1576–82. [PubMed: 23863898]
18. Doria-Rose VP, Harlan LC, Stevens J, Little RF. Treatment of *de novo* acute myeloid leukemia in the United States: a report from the Patterns of Care program. *Leuk. Lymphoma*. 2014; 55:2549–2555. [PubMed: 24467221]
19. Cressman S, et al. Economic impact of genomic diagnostics for intermediate-risk acute myeloid leukaemia. *Br. J. Haematol.* 2016; 174:526–35. [PubMed: 27098559]
20. Khera N, Zeliadt SB, Lee SJ. Economics of hematopoietic cell transplantation. *Blood*. 2012; 120:1545–1551. [PubMed: 22700725]
21. Forbes SA, et al. COSMIC: exploring the world’s knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* 2014; 43:D805–11. [PubMed: 25355519]
22. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics*. 1983; 39:499–503. [PubMed: 6354290]
23. Schmoor C, Sauerbrei W, Schumacher M. Sample size considerations for the evaluation of prognostic factors in survival analysis. *Stat. Med.* 2000; 19:441–52. [PubMed: 10694729]
24. Jameson JL, Longo DL. Precision medicine—personalized, problematic, and promising. *N. Engl. J. Med.* 2015; 372:2229–34. [PubMed: 26014593]
25. Stone RM, et al. The Multi-Kinase Inhibitor Midostaurin (M) Prolongs Survival Compared with Placebo (P) in Combination with Daunorubicin (D)/Cytarabine (C) Induction (ind), High-Dose C Consolidation (consol), and As Maintenance (maint) Therapy in Newly Diagnosed Acute Mye. *Am. Soc. Hematol. Annu. Meet.* 2015:A6.
26. Ley TJ, et al. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med.* 2010; 363:2424–2433. [PubMed: 21067377]
27. Welch JS, et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell*. 2012; 150:264–278. [PubMed: 22817890]
28. Delhommeau F, et al. Mutation in TET2 in myeloid cancers. *N Engl J Med.* 2009; 360:2289–2301. [PubMed: 19474426]
29. Schlenk RF, et al. All-trans retinoic acid improves outcome in younger adult patients with nucleophosmin-1 mutated acute myeloid leukemia – results of the AMLSG 07-04 randomized treatment trial [abstract]. *Blood*. 2011; 118 Abstract 80.
30. Schlenk RF, et al. Phase III study of all-trans retinoic acid in previously untreated patients 61 years or older with acute myeloid leukemia. *Leukemia*. 2004; 18:1798–803. [PubMed: 15385923]
31. Schlenk RF, et al. Prospective evaluation of allogeneic hematopoietic stem-cell transplantation from matched related and matched unrelated donors in younger adults with high-risk acute myeloid leukemia: German-Austrian trial AMLHD98A. *J. Clin. Oncol.* 2010; 28:4642–4648. [PubMed: 20805454]
32. Therneau TM, Grambsch PM, Pankratz VS. Penalized Survival Models and Frailty. *J. Comput. Graph. Stat.* 2003; 12:156–175.
33. Shah RD, Samworth RJ. Variable selection with error control: Another look at stability selection. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 2013; 75:55–80.

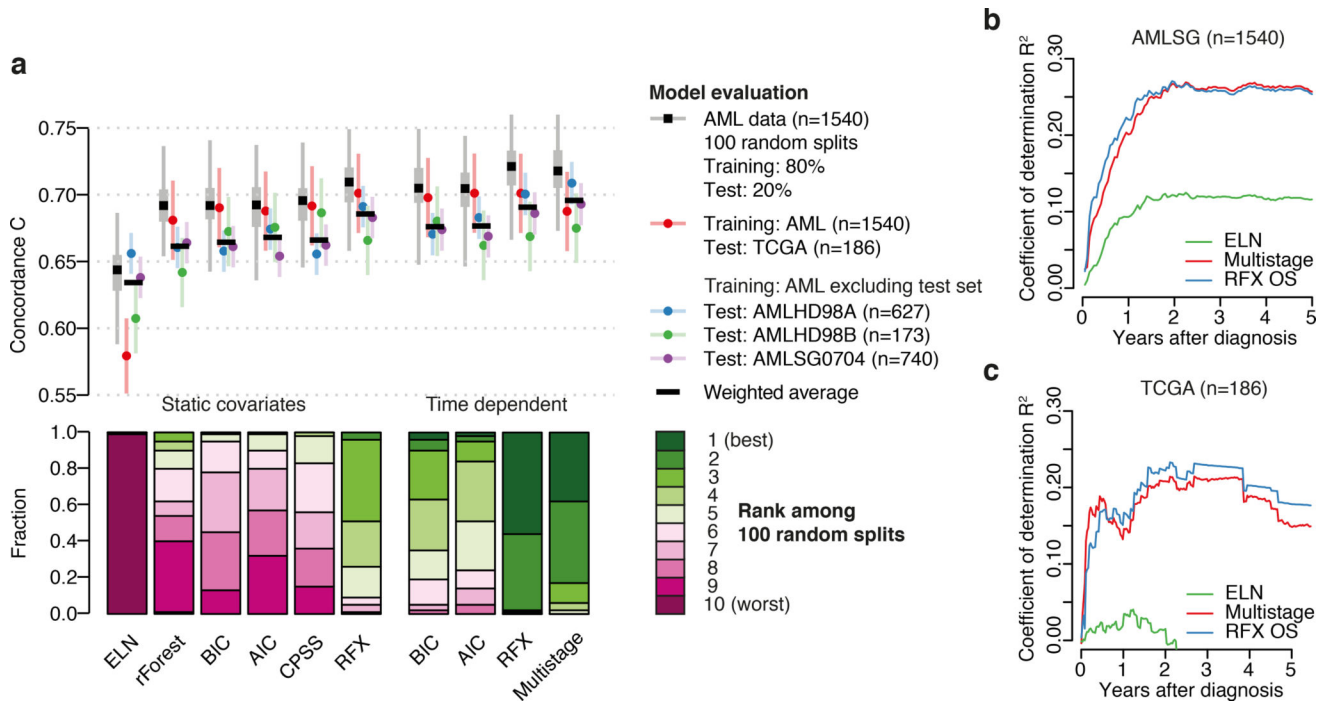


Figure 1. Systematic model comparison

a. Top panel: Concordance C of different model predictions for overall survival. For cross-validation analyses (grey), we generated 100 training and test sets by randomly splitting the full dataset. The distribution of concordance values across the 100 random sets is shown as a box-and-whisker plot. Also shown are point estimates with error bars for predictions evaluated on pre-specified splits of the dataset, where the training set represented 2 of the 3 trials in the study and the test set was the third trial (red, blue, green) or where the training set was the full AMLSG dataset with the test set being the TCGA cohort (purple). Predictions for the multistage model are evaluated 3yrs after diagnosis.

Lower panel: Using the 100 random cross-validation splits, each of the 10 classes of predictive model was built on the training set and evaluated on the test set. The 10 models were ranked based on their relative performance on the test set and the ranks across the 100 cross-validation splits aggregated, indicating how often each model scored best (1st) to worst (10th). Time-dependent models include allogeneic hematopoietic stem cell transplants, which is treated as a time-dependent covariate to avoid bias.

b. Coefficient of determination R^2 for leave-one-out predictions using time-dependent random effects and multistage predictions of the AMLSG cohort, evaluated at each time (x-axis).

c. Same as b, evaluated on TCGA data.

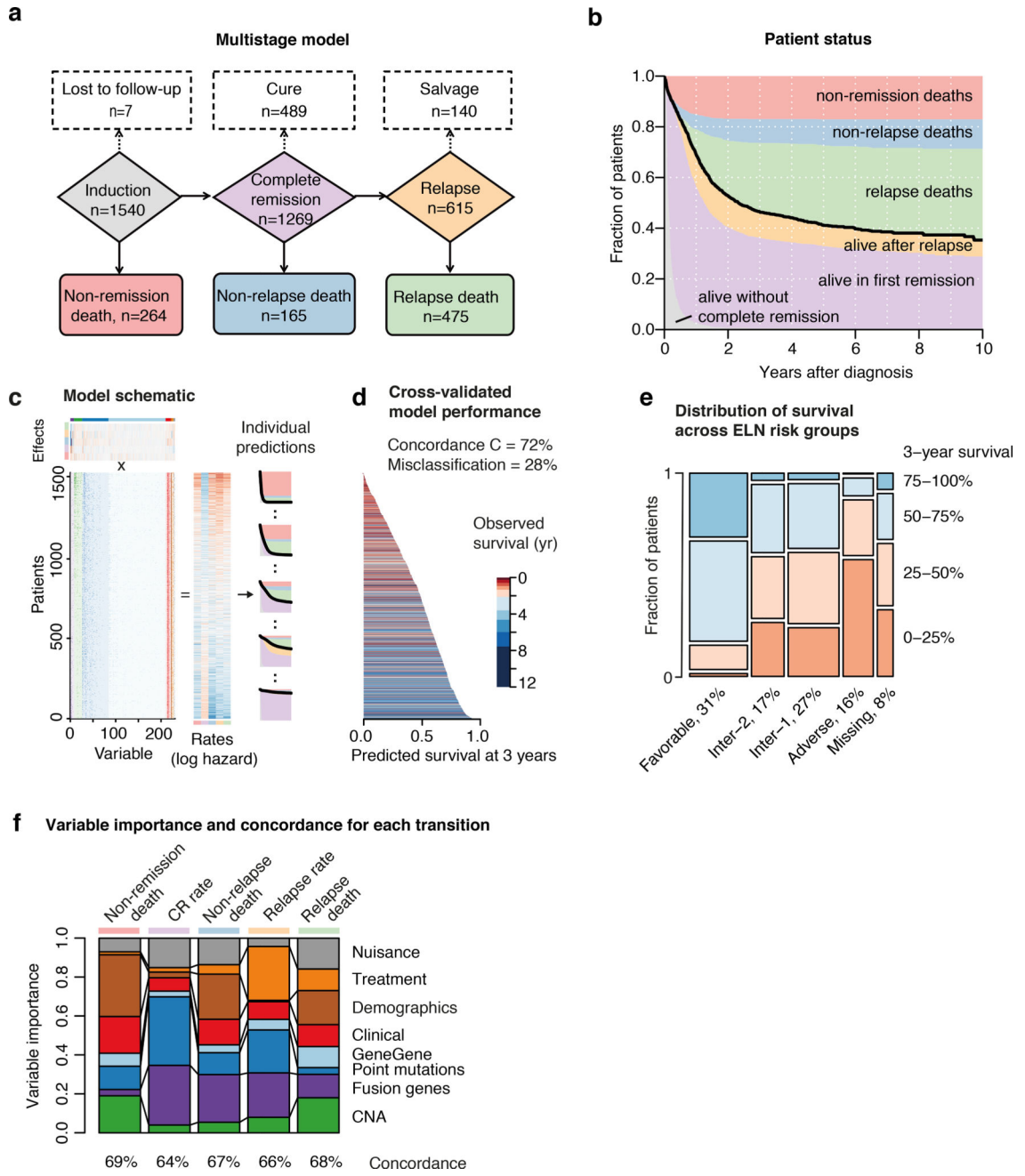


Figure 2. Multistage modeling of patient fate

a. Multistage model of patient trajectories. The six colored boxes indicate different stages during treatment, with five possible transitions indicated by solid arrows. Numbers in each box indicate the total number of patients that have entered a given stage in during follow-up.

b. Sediment plot showing the fraction of patients in a given stage at a given time after diagnosis. The thick black line denotes overall survival, which is the sum of the deaths without complete remission (red), non-relapse mortality (blue) and mortality after relapse (green).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- c. Schematic overview of multistage regression. The model estimates the log-additive effect of each of 231 prognostic variables on the transition rates for all 5 possible time-dependent transitions shown in (a). Rate changes are modelled by Cox proportional hazards models with random effects.
- d. Concordance, C , indicates the survival times at 3 years after diagnosis were correctly ranked by the model. Similarly, at three years after diagnosis only 28% of patients were incorrectly predicted to be alive or dead.
- e. Mosaic plot of predicted 3-year survival across ELN categories. The height of each bar denotes the fraction of patients in each quarter of survival for each ELN group, and the width of each bar is proportional to the percentage of patients in each ELN group.
- f. Relative importance of risk factors for different transitions. The concordance C , is shown as percentages across the top of the bar chart.

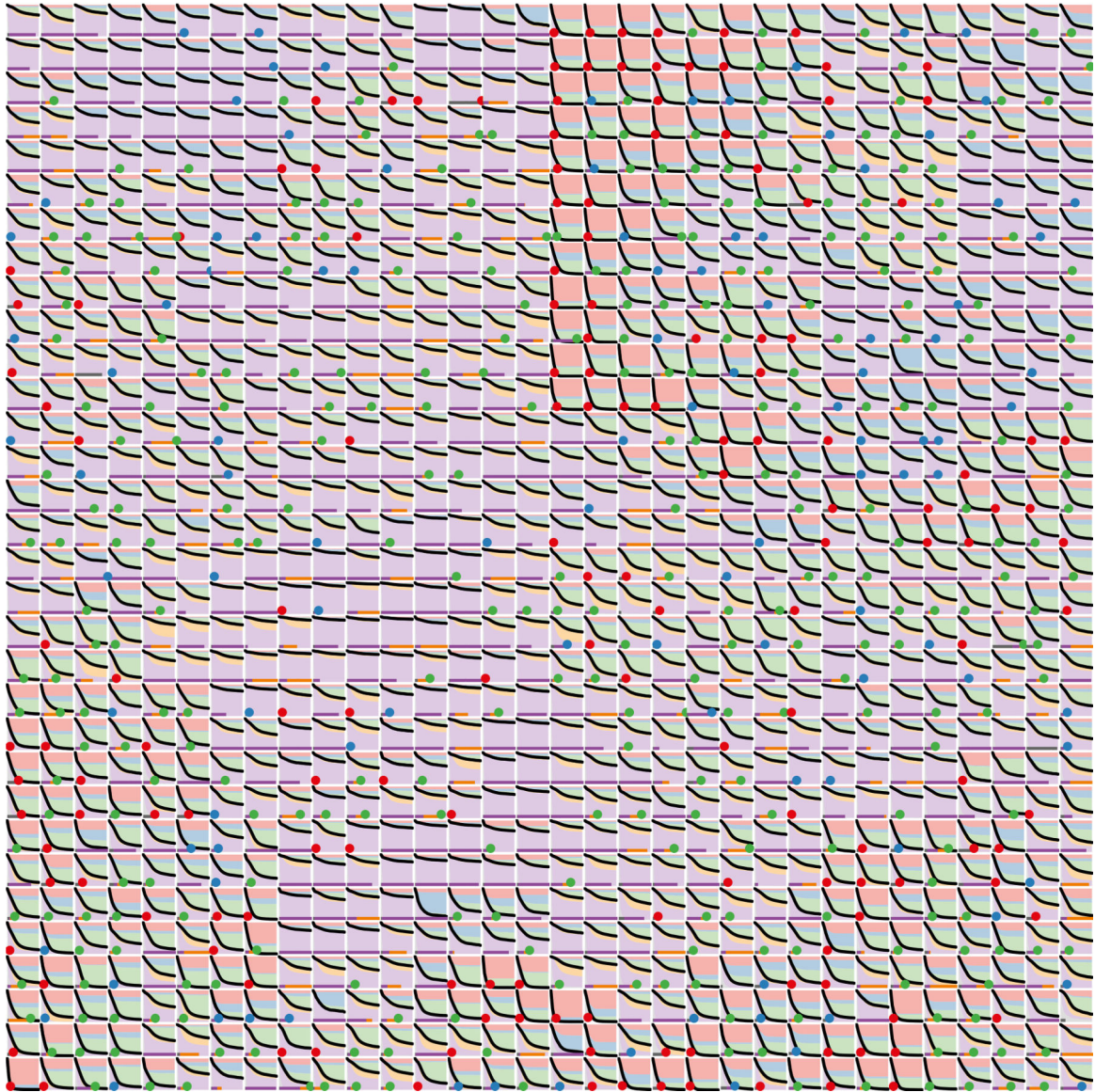
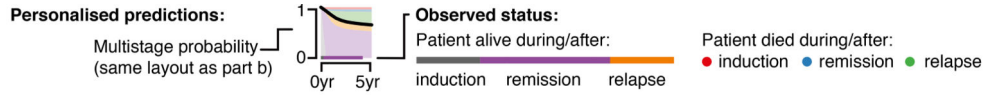


Figure 3. Multistage outcome predictions for 1024 patients

Cross-validated risk predictions and observed statuses for 1024 patients, arranged along a Hilbert curve. This has the property that patients with similar AML subtype and risk constellation are grouped together in the 2-dimensional space (compare Supplementary Figure 1 for constellations of risk factors). For each individual patient, the survival curves predicted by the multistage model are shown, with the competing outcomes colored as in the legend and Figure 2b. What actually happened to the patient is shown as a line across the base of the graph, with a filled circle indicating the patient died, its color indicating the mode of death. Note that there are many patients for whom one color dominates the

diagram, indicating that the probability that a particular event occurs is very high. Reassuringly, for such patients the observed outcomes are highly concordant with the cross-validated predictions and occur at frequencies matching the predicted probabilities.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

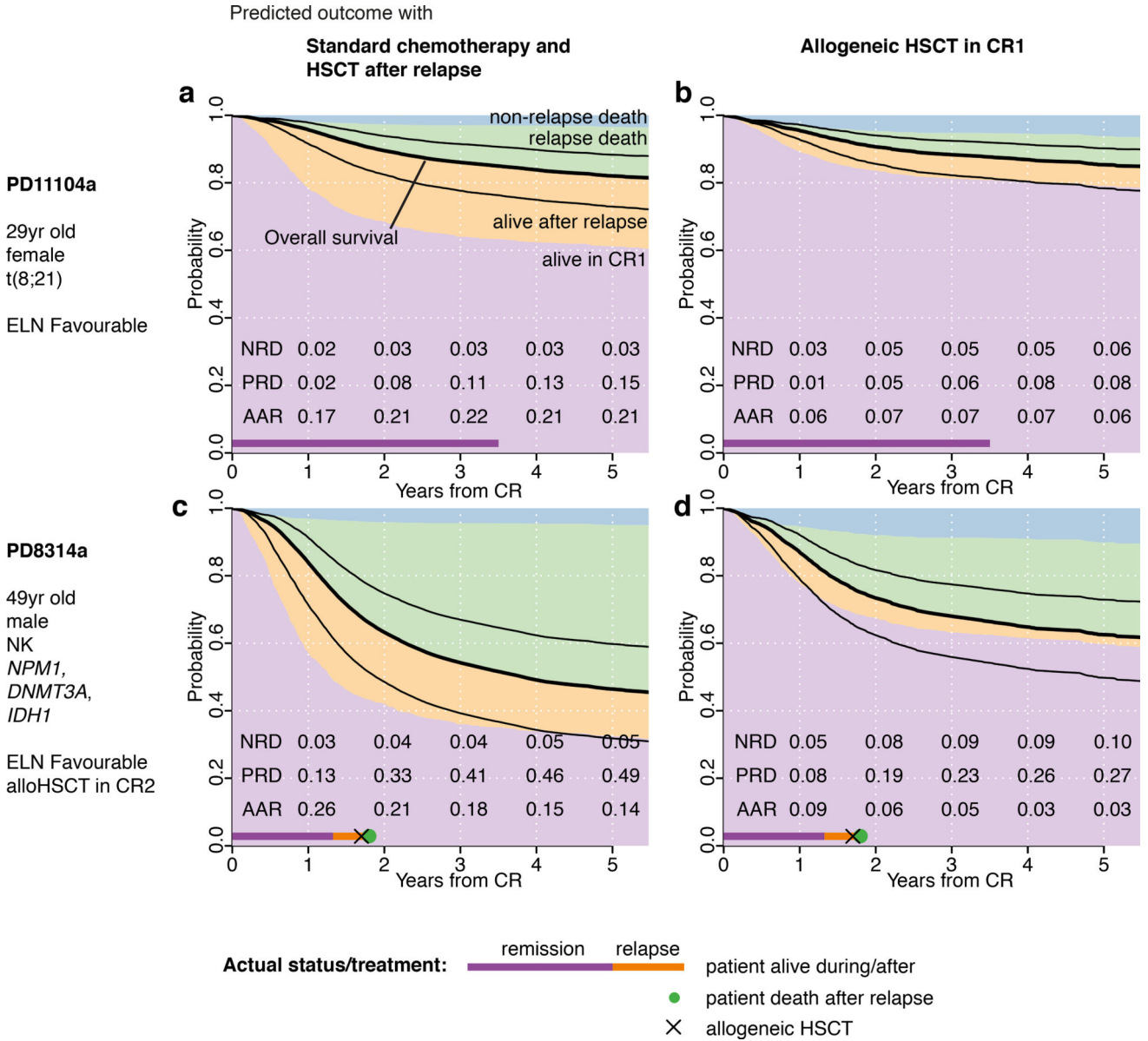


Figure 4. Individualized risk exemplified for 2 patients

- a. Sediment plot showing predicted multistage probability after remission for patient PD11104a under a management strategy of standard chemotherapy in CR1 with intended allograft after relapse. Predictions shown are based on models where the given patients were excluded for training; the bar at the bottom denotes the observed outcome (as for Figure 3). The patient was alive at the last follow-up 3.5 years after achieving first complete remission. Numbers at the bottom indicate the probabilities of non-relapse death (NRD), post-relapse death (PRD) and being alive after relapse (AAR) at years 1 to 5 from achieving complete remission.
- b. Multistage probability for PD11104a in the scenario of an allograft in first complete remission.
- c. Same as **a** for patient PD8314a. The patient relapsed after 1.2 years and died soon after.

d. Same as **b** for patient PD8314a.

Details of these calculations are presented in Supplementary Note, section 3.5.5.8; additional patients shown in Supplementary Figure S2.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

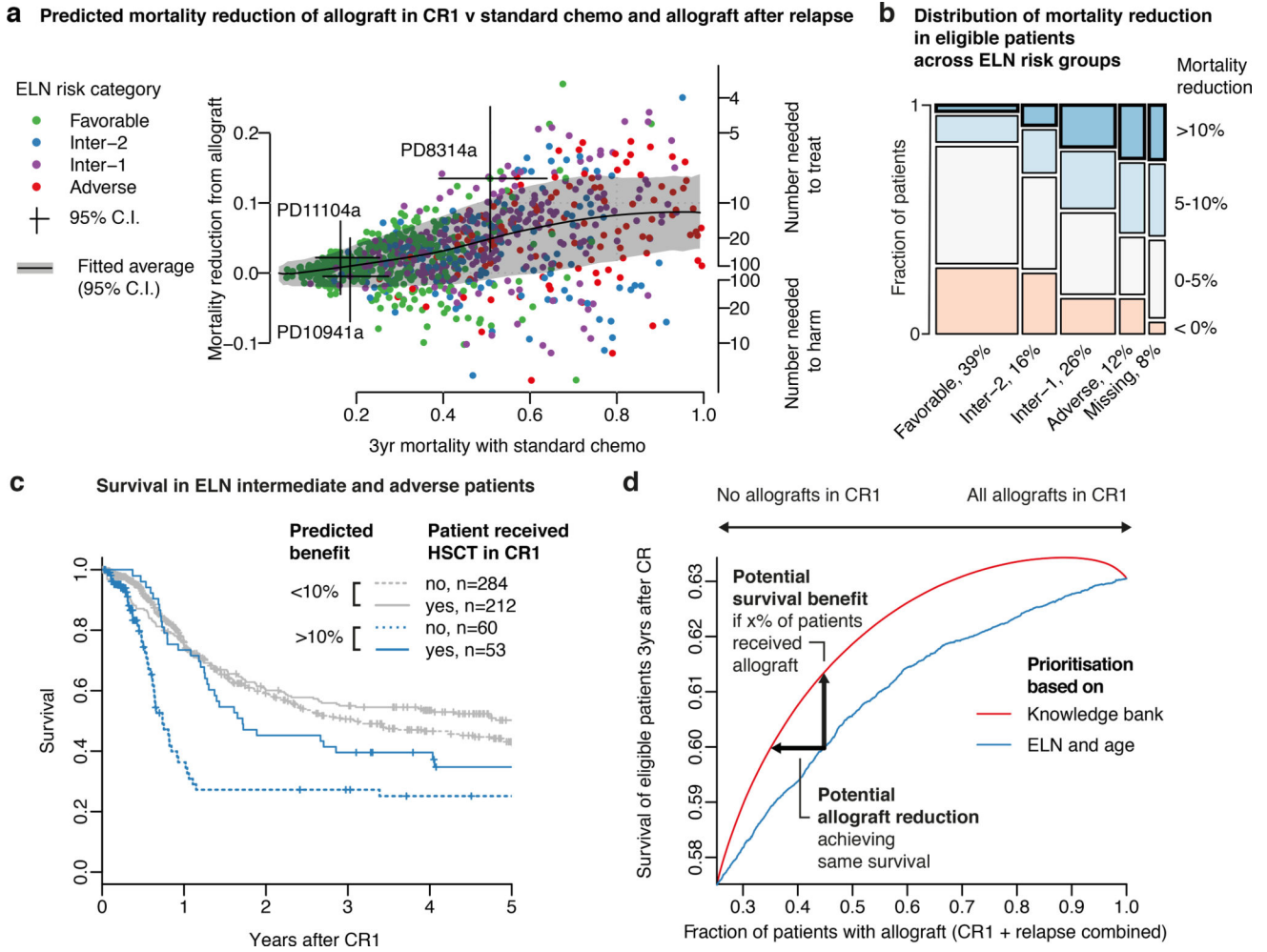


Figure 5. Benefit of allograft in CR1 vs after relapse

a. Predicted three-year absolute mortality reduction by allografts in CR1 over standard chemotherapy in CR1 and allograft after relapse (y-axis). Calculations are based on patients <60yr in CR1 (n=995), who would be eligible for allogeneic transplants. The black curve represents the population average, with 95% confidence intervals in grey. Points denote individual patients in the cohort, colored by ELN risk category.

b. Mosaic plot of absolute survival benefit at 3 years by an allograft in CR1 over standard chemotherapy after CR1 versus ELN risk category. The predicted benefit was discretized into four groups, indicated by colors, with intervals of 5%.

c. Kaplan-Meier curves for patients with high (>10%, blue) and low (<10%, grey) predicted benefit of early allograft (cross-validated), each with and without allograft in CR1. Patients with favorable ELN risk were excluded.

d. Predicted overall survival at 3yrs as a function of total number of allografts (in CR1 + after relapse). Patients are first ranked from those most likely to benefit from transplant to those least likely to benefit, as judged by current guidelines (solid blue line) or our current knowledge bank (solid red line). The curves show expected survival if allografts in CR1 increased from 0% to 100%, starting with the patient with the greatest and ending with the lowest predicted benefit. The x-axis starts at ~0.25, since about half of patients will relapse

without an allograft in CR1, with a further half managing to undergo post-relapse transplantation.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

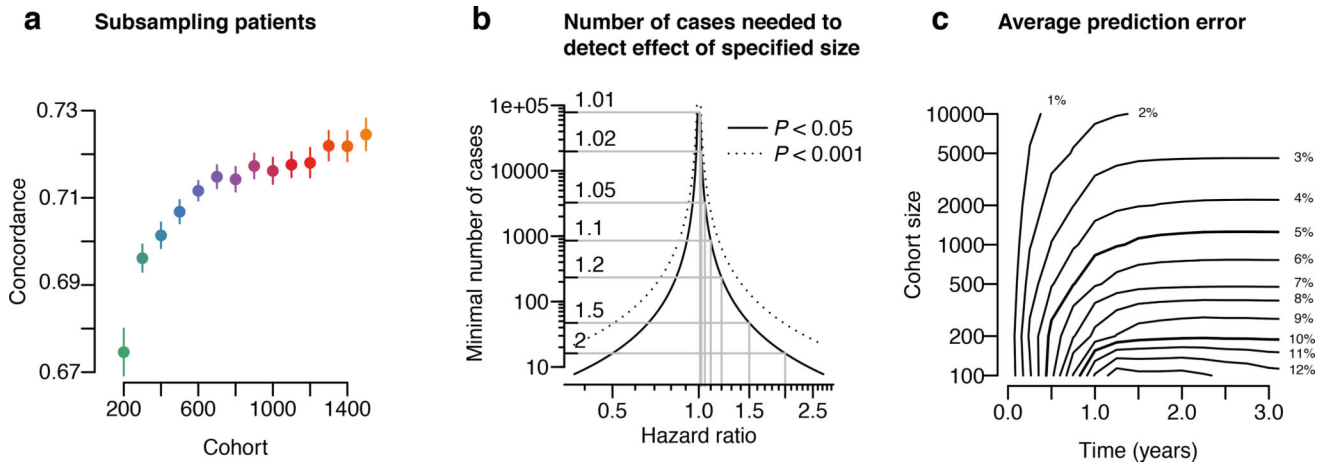


Figure 6. Extrapolations and power calculations

- a. Subsampling the number of patients reveals a steady, but saturating increase in prognostic concordance C for a random effects model for overall survival. Error bars show the 95% confidence intervals for the concordance obtained from multiple independent subsamples of the dataset.
- b. Graph relating the effect size (hazard ratio) of a prognostic variable to the absolute number of patients with the given factor required to reach significance in a random effects model for overall survival (solid line: $P < 0.05$; dotted $P < 0.001$).
- c. Average prediction error between simulated and estimated survival a random effects model for overall survival as a function of survival time (x-axis) and training cohort size (y-axis).