nature genetics

Precision oncology for acute myeloid leukemia using a knowledge bank approach

Moritz Gerstung[1,2,9], Elli Papaemmanuil[1,3,9], Inigo Martincorena[1], Lars Bullinger[4], Verena I Gaidzik[4], Peter Paschka[4], Michael Heuser[5], Felicitas Thol[5], Niccolo Bolli[1,6], Peter Ganly[7], Arnold Ganser[5], Ultan McDermott[1], Konstanze Döhner[4], Richard F Schlenk[4], Hartmut Döhner[4,10] & Peter J Campbell[1,8,10]

Underpinning the vision of precision medicine is the concept that causative mutations in a patient's cancer drive its biology and, by extension, its clinical features and treatment response. However, considerable between-patient heterogeneity in driver mutations complicates evidence-based personalization of cancer care. Here, by reanalyzing data from 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, which were validated using data from independent patients in The Cancer Genome Atlas. Comparison of long-term survival probabilities under different treatments enables therapeutic decision support, which is available in exploratory form online. Personally tailored management decisions could reduce the number of hematopoietic cell transplants in patients with AML by 20–25% while maintaining overall survival rates. Power calculations show that databases require information from thousands of patients for accurate decision support. Knowledge banks facilitate personally tailored therapeutic decisions but require sustainable updating, inclusive cohorts and large sample sizes.

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 genes involved in cancer (hereafter referred to as cancer genes) fundamentally drive a 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: only a few cancer genes are straightforward therapeutic targets; many cancer genes are only rarely mutated in a given tumor type; and each patient's tumor typically has several driver mutations. In addition to all of the other complications is the challenge that, for most tumor types, there are hundreds to thousands of different combinations of driver mutations that have been observed across patients[5–7].

The promise of precision medicine has triggered considerable funding commitments, such as the Precision Medicine Initiative in the United States, Genomics England in the UK and similar efforts in several other countries[8,9]. Among the 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 that are fine-tuned to the unique genome of an individual cancer, whereas evidence-based medicine requires therapeutic decisions that are grounded on documented, verified data.

Here we explore these questions by reanalyzing genetic data from 111 cancer genes, cytogenetic profiles and clinical data from 1,540 patients with AML who were undergoing intensive treatment[10], and we validated this analysis using data from an independent cohort of patients with AML from The Cancer Genome Atlas (TCGA)[11]. In our previous study[10], we identified 11 genomic subcategories of AML, each of which had a distinctive constellation of clinical features. However, even within individual molecular subgroups, there remains considerable patient-to-patient variability in treatment response and clinical outcomes, which could partially be explained by cooperating driver mutations and other diagnostic

[1]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton, UK. [2]European Bioinformatics Institute EMBL-EBI, Hinxton, UK. [3]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [4]Department of Internal Medicine III, Ulm University, Ulm, Germany. [5]Department of Hematology, Hemostasis, Oncology and Stem Cell Transplantation, Hannover Medical School, Hannover, Germany. [6]Division of Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori, University of Milan, Milan, Italy. [7]Department of Pathology (UOC), University of Otago, Christchurch, New Zealand. [8]Department of Haematology, University of Cambridge, Cambridge, UK. [9]These authors contributed equally to this work. [10]These authors jointly directed this work. Correspondence should be addressed to P.J.C. (pc8@sanger.ac.uk) or H.D. (hartmut.doehner@uniklinik-ulm.de).

Received 24 June 2016; accepted 30 November 2016; published online 16 January 2017; doi:10.1038/ng.3756

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 who is faced with a difficult therapeutic decision, however, it is not at all clear how such genomic complexity affects 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—the question of 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 CR1 undoubtedly decrease relapse rates for most patients, but this comes at the cost of higher treatment-related mortality, which can be as high as 20–25% at 3 months after transplantation versus around 5% with conventional consolidation chemotherapy[14], with a further 30% risk of debilitating chronic morbidity[15]. Furthermore, even though more patients relapse after chemotherapy in first remission, up to one-fifth of these 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, treatments or clinical conundrums.

RESULTS

Predicting complex patient outcomes from genomic and clinical variables

We recently sequenced[10] all of the coding exons from 111 myeloid cancer genes in diagnostic leukemia samples from 1,540 patients with AML who were undergoing intensive treatment in three prospective clinical trials of the German–Austrian AML Study Group (AMLSG). We identified driver point mutations and combined these data with the clinical trials database to generate a comprehensive knowledge bank. Here we focused on evaluating the utility of the knowledge bank for generating predictions that were personally tailored to the individual patient, as well as how these predictions could be used to compare the likelihoods of various clinical outcomes under different treatment strategies. The full knowledge bank, together with all of the analysis code used here, is documented in the **Supplementary Note** and is available as a git repository (see "Data availability" in the Online Methods).

Throughout, we used overall survival as the primary end-point of these analyses, as the aim of intensive therapy in young patients with AML is a cure. The full data set consists of 231 predictor variables—which span the seven broad categories of fusion genes, copy number alterations, point mutations, gene–gene interactions, demographic features, clinical risk factors and treatment received—across 1,540 patients. To assess the accuracy of our predictions, we used the following validation strategies: (i) random cross-validation on this data set; (ii) building models from any two clinical trials here and testing on the third; and (iii) testing the model built from all three AMLSG trials on an independent AML cohort from the United States (TCGA)[11]. All predictions for individual patients reported here were made using models that excluded data from that patient.

We tested a range of regularized regression methods for predicting survival, and we also implemented novel random effects and multistage statistical models, for deriving detailed associations between genomic and clinical end-points (**Fig. 1a** and **Supplementary Note**). 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 as compared to current prognostic criteria[13] (**Fig. 1b**,**c** and **Supplementary Note**). 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, to control for overfitting. In contrast, conventional methods typically choose reduced subsets of 5–20 variables, seemingly at the cost of discarding prognostically relevant information (for more discussion, see the **Supplementary Note**).

Reassuringly, we found strong 'out-of-cohort' validation for our models, either when models built using 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 (**Fig. 1a**). Of particular note is the observation that concordance decreased only moderately for predictions from a model trained on younger patients (AMLHD98A and AMLSG0704: age range, 18–65 years) that was evaluated on a trial of older patients (AMLHD98B: age 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 learned from the knowledge bank.

The multistage model offers the advantage of separating long-term outcomes into individual constituents—death without complete remission, death in CR1 without relapse (non-relapse mortality, which is mostly treatment related) and death after relapse, as well as survival during induction therapy, survival in CR1 and survival after relapse (**Fig. 2a–c**). As we demonstrate, 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, as the combined prediction of overall survival in the multistage model yields the same accuracy as predicting overall survival directly (**Fig. 1a**).

Personally tailored prognosis

The models for predicting outcome that were 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 system[13], which defines four categories of disease risk based on six fusion genes, three genes with point mutations and cytogenetic abnormalities. We explored how much more informative our more complex prognostic models are than the ELN system.

We found that individual risk in this AML cohort was continuous, with no obvious cutoff points for stratification, suggesting that grouping patients on the basis of a few predictor variables discards a lot of prognostic information (**Fig. 2d**). Our more detailed survival estimates confirmed the broad trends of known ELN risk groups; however, one-third of the patients had survival predictions that deviated more than 20% from their ELN stratum (**Fig. 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 end-point of treatment (**Fig. 2f** and **Supplementary Tables 1–6**). We found that clinical and demographic factors, such as patient age, performance status and blood counts, exerted the most influence on rates of early death, including death in remission (mostly due to treatment-related mortality). Genomic features, such as 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 emerged was considerable heterogeneity in personal risk profiles across the cohort (**Supplementary Fig. 1**). The heterogeneity of the risk profiles and

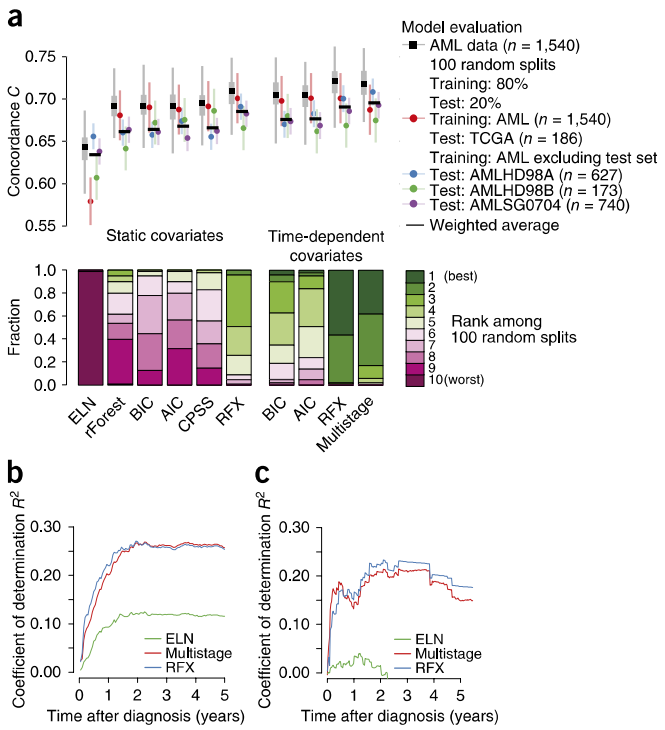

Figure 1 Systematic comparison of models. (**a**) Top, concordance, *C*, of predictions for overall survival from different models. For cross-validation analyses (gray), we generated 100 training and test sets by randomly splitting the full data set. The distribution of concordance values across the 100 random sets is shown as a box-and-whiskers plot. Also shown are point estimates with error bars for predictions evaluated on prespecified splits of the data set, in which the training set represented two of the three trials in the study (AMLHD98A, AMLHD98B and AMLSG0704) and the test set was the third trial (red, blue or green), or in which the training set was the full AML data set with the test set being the TCGA cohort (purple). Predictions for the multistage model are evaluated 3 years after diagnosis. Bottom, using the 100 random cross-validation splits, each of the ten predictive model classes was built on the training set and evaluated on the test set. The ten models were ranked on the basis of their relative performance on the test set and the ranks across the 100 cross-validation splits were aggregated, indicating how often each model scored best (1) to worst (10). Time-dependent models include allogeneic hematopoietic cell transplants, which are treated as a time-dependent covariate to avoid bias. ELN, European LeukemiaNet genetic scoring system[13]; rForest, random survival forest[26]; BIC and AIC, Cox proportional-hazards regression with stepwise variable selection based on the Bayesian information criterion (BIC) or Akaike information criterion (AIC), respectively; CPSS, complementary-pairs stability selection; RFX, Cox proportional-hazards model with random effects. (**b**) Coefficient of determination R^2 for leave-one-out predictions using time-dependent random effects and multistage predictions of the AMLSG cohort (n = 1,540), evaluated at each time point (x axis). (**c**) As in **b**, except for data from TCGA (n = 186).

the variable effect they had on the different AML outcomes combined to generate a kaleidoscope of predictions for the patients' journeys through therapy (**Fig. 3**). Thus, there are distinct groups of patients for whom we can confidently predict long-term survival in first remission, death after relapse or death without achieving remission, which manifested as swathes of purple, green and pink, respectively, in **Figure 3**. Reassuringly, these predictions were consistent with the actual outcomes in the patients (status lines and circles in **Fig. 3**). It is these patients, for whom one color dominates, that are highly likely to have a defined outcome. 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 showed 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 CR1 or after relapse[12,13]. With a transplant-associated mortality rate of 20–25% and substantial rates of chronic graft-versus-host disease, allografts tend to be reserved for high-risk patients. We explored how a knowledge bank could inform the decision to choose 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 (**Fig. 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 can arrive at this probability value through different risk contributions: one patient might be older and more frail but have a leukemia with generally favorable genomic features, whereas 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 option in the latter, fitter patient and prefer standard chemotherapy in the older patient who is at higher risk of treatment-related mortality.

We illustrate these calculations using two patients from the cohort (**Fig. 4**; other representative patients are illustrated in **Supplementary Fig. 2**). The first (PD11104a) was a 29-year-old woman with chromosomal translocation t(8;21) and no other driver mutations, who would have been classified as having 'favorable risk' by ELN criteria[13]. Under a strategy of chemotherapy in CR1 with salvage allograft after relapse, we predicted her chance of 3-year survival to be 86% (95% confidence interval (CI) = 78–91%) (**Fig. 4a**). In contrast, with an allograft in CR1, we estimated her overall cure rate to be 88% (95% CI = 79–93%) (**Fig. 4b**), with the decrease in probability of relapse matched by the increase in non-relapse mortality with a transplant. Hence, there is no indication for an up-front allograft for this patient, with only a two-percentage-point difference in predicted survival (95% CI = −3 to 7%). For this patient, therefore, the treatment recommendation under current clinical standards[13] would remain unchanged using a knowledge bank approach.

The second patient (PD8314a) was a 49-year-old male with mutations in *NPM1* (nucleophosmin), *DNMT3A* (DNA methyltransferase 3α) and *IDH1* (isocitrate dehydrogenase (NADP$^+$) 1), and a normal karyotype. Under ELN criteria, his risk also classifies as favorable, and he would not currently be recommended for allograft in CR1. With standard chemotherapy as first-line therapy, we estimated his 3-year survival probability to be 55% (95% CI = 41–67%), as compared to 68% (95% CI = 55–77%) for allograft in CR1 (**Fig. 4c,d**). Thus, his disease was not especially of favorable risk when all of the predictive information was considered. Furthermore, the absolute risk reduction associated with an up-front allograft was estimated at 13 percentage points (95% CI = 3–24%). Thus, the number needed to treat (NNT) is 7 patients (95% CI = 4–26): that is, 1 additional person would be cured for every 7 equivalent patients 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 in comparison to the 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 could accurately reflect what actually happened to the patients.

On average, we found that patients who were predicted to have poor prognosis, defined as having more than 60–70% chance of mortality at 3 years, were most likely to benefit from allogeneic transplantation

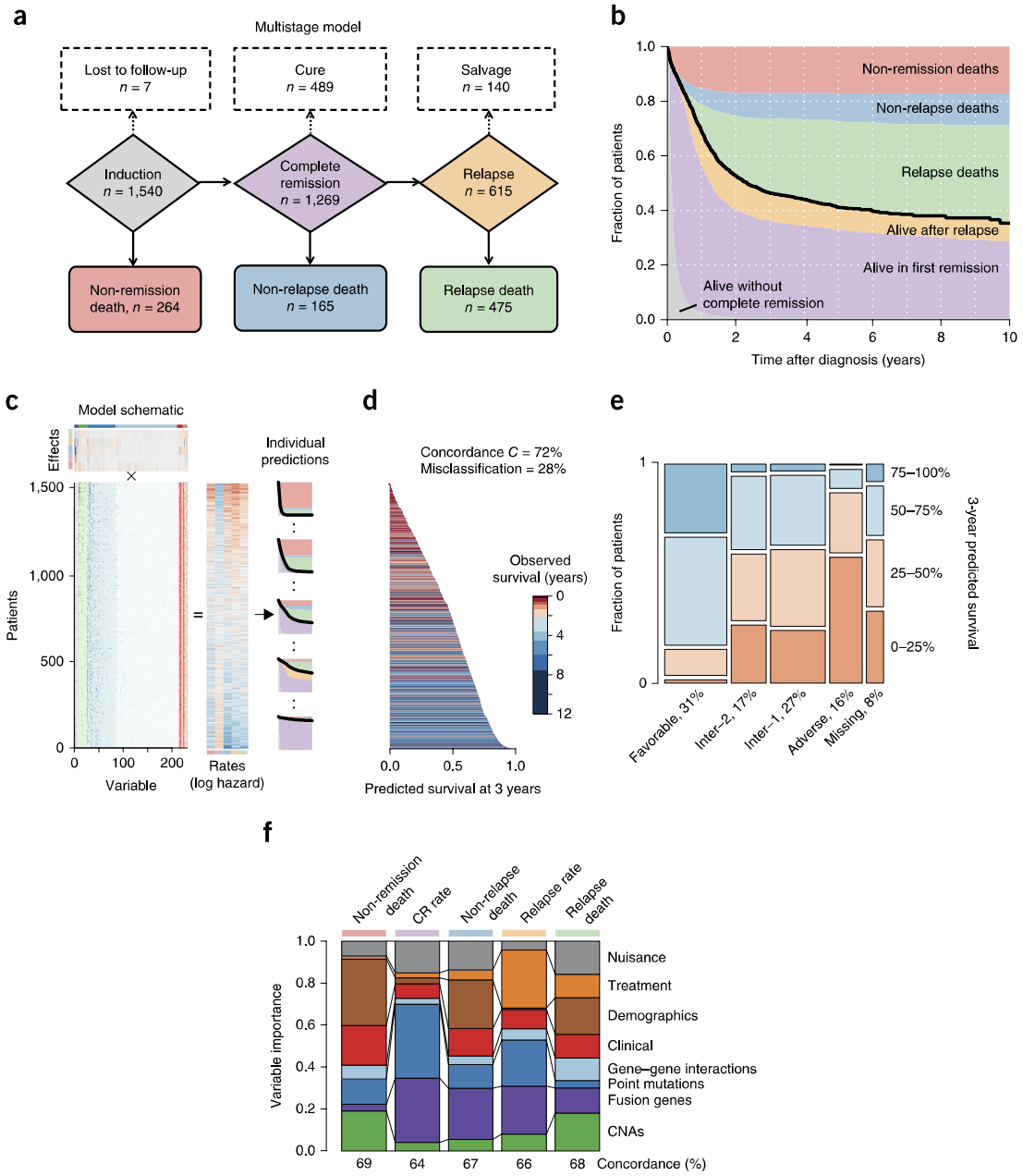

Figure 2 Multistage modeling of patient fate. (**a**) Multistage model of patient trajectories. The six colored shapes correspond to different stages during treatment, with five possible transitions (solid arrows). The number in each shape corresponds to the total number of patients who entered a given stage 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 in patients without complete remission (red), with non-relapse-associated mortality (blue) and with mortality after relapse (green). (**c**) Schematic of multistage regression. The model estimates the log-additive effect of each of 231 prognostic variables on the transition rates for all five possible time-dependent transitions shown in **a**. Rate changes were modeled by Cox proportional-hazards models with random effects. (**d**) Cross-validated model performance, with concordance, *C*, indicating that the survival times at 3 years after diagnosis were correctly ranked by the model. Similarly, at 3 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. Inter-1 and inter-2 denote subgroups with intermediate outcomes in the ELN classification. The number of patients in each group was as follows: favorable, 473; inter-2, 268; inter-1, 417; adverse, 253; missing, 129. (**f**) Relative importance of risk factors for different transitions. The concordance, *C*, is shown as a percentage across the bottom of the bar chart. The number of patients in each group is the same as in **a**. CNAs, copy number alterations.

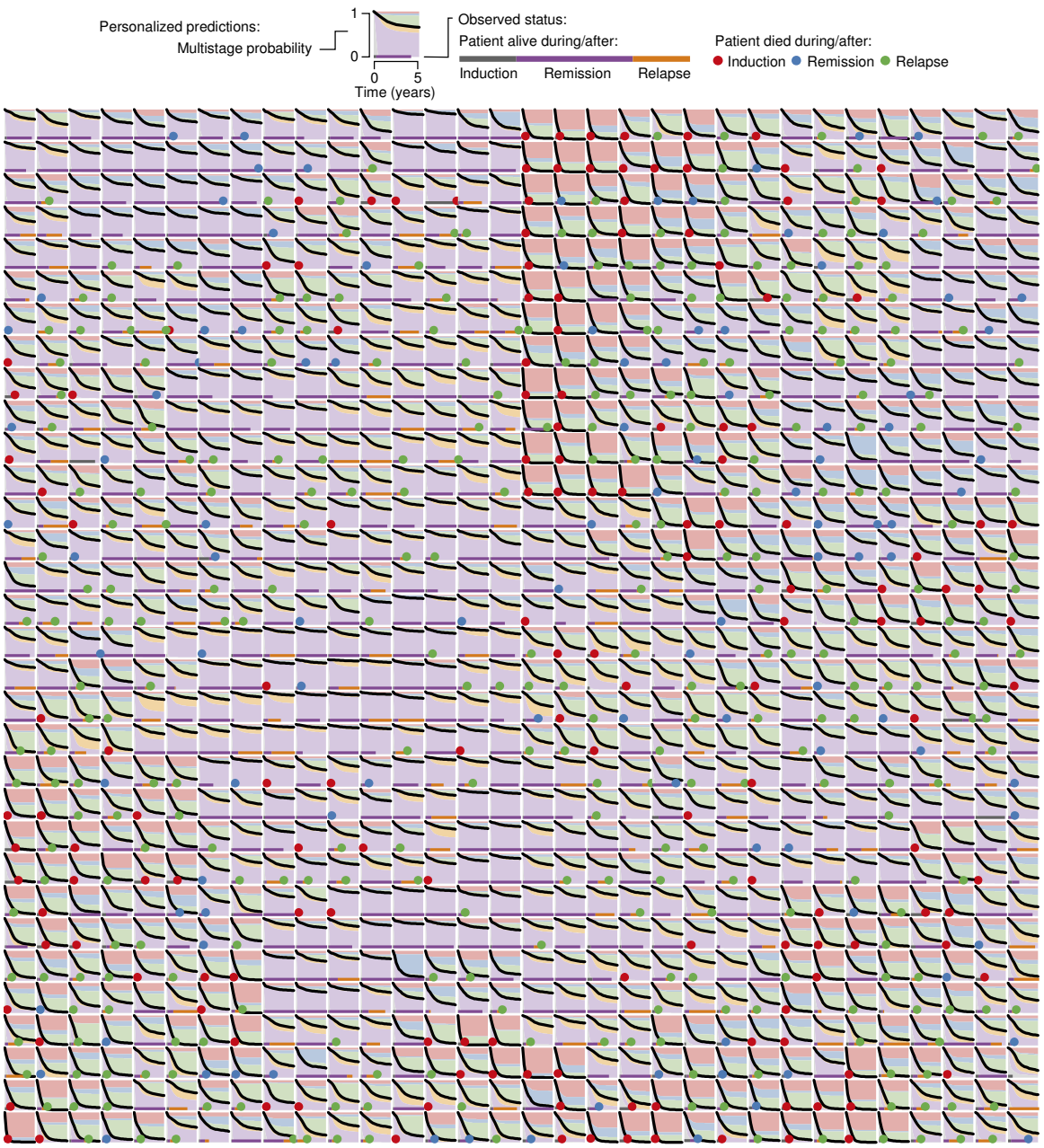

Figure 3 Multistage outcome predictions for 1,024 patients. Cross-validated risk predictions and observed status for 1,024 patients are arranged along a Hilbert curve. This visualization has the property that patients with similar AML subtypes and risk constellations are grouped together in 2D space (compare to **Supplementary Fig. 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 that the patient died (color denotes 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.

in first remission (**Fig. 5a**), a finding that was captured in current clinical recommendations. However, there was a considerable spread in patient estimates around the population average. This variance around the average is critically important for precision medicine because it suggested that population-based criteria for treatment choices only poorly capture the predictive information available for the individual patient.

Overall, we estimated that 12% (124/995) of patients in CR1 aged 18–60 years would have more than a 10-percentage-point improvement in survival at 3 years with an allograft in CR1 as compared to standard chemotherapy (NNT < 10; **Fig. 5b**). Only 29 of these 124 patients were identified as having 'adverse risk' by current criteria, with most being of 'intermediate risk' and some even of 'favorable risk'. Furthermore, 57% (302/534) of patients who were classified as having adverse or intermediate risk by ELN criteria, and therefore strongly considered for allograft in CR1 under current clinical recommendations[13], were predicted to derive less than a 5-percentage-point improvement in survival from up-front allografts. Similarly, 15% (58/386) of patients with

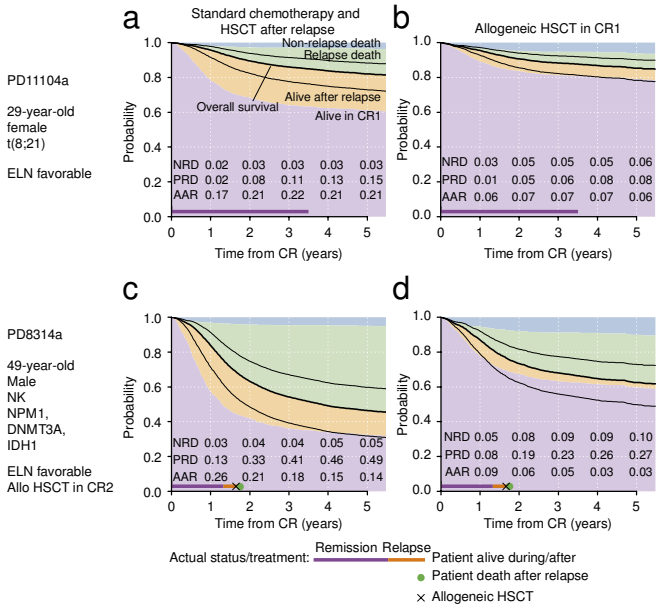

Figure 4 Individualized risk exemplified for two patients. (**a**,**b**) Sediment plots showing predicted multistage probabilities after remission for patient PD11104a under a management strategy of standard chemotherapy in CR1 with intended allograft after relapse (**a**) or allograft in CR1 (**b**). The patient was alive at the last follow-up 3.5 years after achieving CR1. (**c**,**d**) Sediment plots showing predicted multistage probabilities after remission for patient PD8314a under a management strategy of standard chemotherapy in CR1 with intended allograft after relapse (**c**) or allograft in CR1 (**d**). The patient relapsed 1.2 years after CR1 and died soon after. Predictions shown are based on models in which the given patients were excluded from the training data set; the bar at the bottom denotes the observed outcome (as for **Fig. 3**). Numbers at the bottom correspond to the probability of non-relapse death (NRD), post-relapse death (PRD) and being alive after relapse (AAR) at years 1 to 5 after achieving complete remission. Details of these calculations are presented in the **Supplementary Note**; data for additional patients are shown in **Supplementary Figure 2**. Allo-HSCT, allogenic hematopoietic stem cell transplant; NK, normal karyotype.

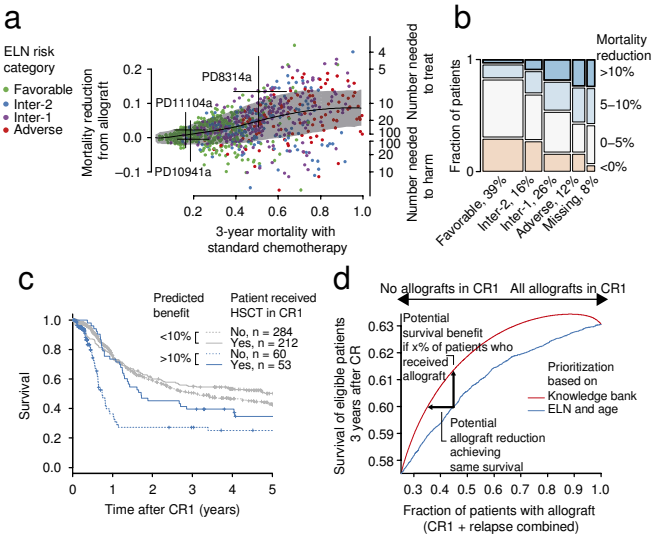

Figure 5 Benefit of allograft in CR1 versus allograft after relapse. (**a**) Predicted 3-year absolute mortality reduction in patients who received allografts in CR1 versus those who received standard chemotherapy in CR1 and allografts after relapse (*y* axis). Calculations are based on patients who were <60 years old in CR1 ($n = 995$) and who would be eligible for allogeneic transplant. The black curve represents the population average, with the 95% confidence interval in gray. Points denote individual patients in the cohort, colored by risk category as determined by ELN criteria. (**b**) Mosaic plot of 3-year survival benefit from an allograft in CR1 relative to standard chemotherapy after CR1 versus ELN risk category. The predicted benefit was discretized into four groups, indicated by color, with intervals of 5%. The number of patients in each ELN group was as follows: favorable, 386; inter-2, 160; inter-1, 257; adverse, 117; missing, 75. (**c**) Kaplan–Meier curves for patients with high (>10%; blue) and low (<10%; gray) predicted benefit from receiving an allograft in CR1 (cross-validated), with and without receipt of an allograft in CR1. Patients with favorable ELN risk were excluded. (**d**) Predicted overall survival at 3 years after CR1 as a function of the total number of allografts performed (in CR1 + after relapse). Patients were first ranked from those most likely to benefit from transplant to those least likely to benefit, as judged by current guidelines (blue line) or our current knowledge bank (red line). The curves show expected survival if the fraction of patients receiving allografts in CR1 increases from 0% to 100%, starting with the patient with the greatest predicted benefit and ending with the patient with the lowest predicted benefit. The *x* axis starts at ~0.25, as about 50% of patients will relapse without an allograft in CR1, of whom about 50% manage to undergo post-relapse transplantation.

favorable risk by ELN criteria were predicted to have a >5-percentage-point improvement in expected survival with a bone marrow transplant in CR1. Clearly, then, treatment management in up to one-third of patients might be changed with a knowledge bank approach relative to the recommendations provided by current practice guidelines.

We next compared the therapeutic predictions made by our model with what actually happened to the patients under the two different treatment strategies (**Fig. 5c** and **Supplementary Fig. 3**). We split the cohort into two groups depending on whether the patients were predicted to derive more or less than a 10-percentage-point improvement in survival with allograft in CR1 as compared to chemotherapy in CR1 and allograft after relapse. If our model was correctly identifying those patients who were most likely to benefit from a transplant, then the survival curves in this group should show distinctly better outcomes for allograft in first remission than those for chemotherapy. This was indeed what we observed (**Fig. 5c**, blue lines). For the patients for whom our model predicted minimal or no benefit from an up-front transplant strategy, we did indeed find that there was little meaningful difference in the survival curves between those patients who received a transplant and those who received chemotherapy in first remission (**Fig. 5c**, gray lines).

Taken together, these data demonstrate that up to one-third of patients would have had their treatment altered using a knowledge bank approach as compared to the treatment they received using current practice recommendations[13]. Furthermore, the predictions made using the knowledge bank matched well the actual outcomes observed under the two different treatment philosophies, confirming the accuracy of the decision support.

Population effect 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 effect in patients with 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 (**Fig. 5d**). In the United States, ~30% of patients with AML receive an allograft[18]. If the 30% of patients to receive an allograft in CR1 were chosen using an optimal

knowledge bank rather than current recommendations, we estimate that survival rates across the cohort would increase ~1.3 percentage points (from 60% to 61.3%).

Alternatively, personally tailored management decisions could reduce the number of hematopoietic cell transplants in patients with AML by 20–25% while maintaining the same overall survival rates. Under current practice, 44% of young adult patients would receive a transplant (30% in CR1 plus 14% after relapse). In contrast, by using a knowledge bank approach to choose when and whom to give a transplant, 35% of patients overall would receive an allograft (16% in CR1 plus 19% after relapse) to achieve 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 receiving a transplant (**Supplementary Fig. 4**).

We can express the effect of a knowledge bank approach at the population level in terms of quality-adjusted life years (QALYs). Health utilities, which measure the quality adjustment factors for survival, have previously been estimated to be 0.74 for the survival of patients with AML who received an allograft and 0.83 for the survival of those who did not receive a transplant[19], and the cost of an allograft has been estimated to be $100,000–200,000 (ref. 20). 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 10 years. Alternatively, reducing the number of allografts by better resource allocation while maintaining overall survival rates would result in a gain of ~0.05 QALYs per patient over 10 years, as well as saving approximately $10,000 per patient.

Portals for exploring decision-support predictions

The preceding sections demonstrated that the complex and multifactorial inter-relationships among genomic variables, clinical predictors and cancer outcomes can be learned with a sufficiently comprehensive knowledge bank. Because the underlying survival models are complex, diagnostic laboratories may need to provide personalized portals into a given patient's cancer genome.

Our data set 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 website[21] (http://cancer.sanger.ac.uk/aml-multistage) that allows outcome predictions to be generated based on this data set for user-defined constellations of genomic features, clinical variables and treatment strategies (**Supplementary Fig. 5**). The underlying algorithm is capable of imputing missing variables and computing 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 affect 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 Fig. 6a**), a relationship underpinned by theoretical arguments (**Supplementary Note**). Some genes, by virtue of their frequency and/or the magnitude of their prognostic effect, are more informative than others. We ranked genes involved in AML by their predictive utility (**Supplementary Fig. 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 the confidence intervals of patient prediction can be explored in the web portal.

The other critical factor for 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 increased with larger sample sizes, although it followed a law of diminishing returns (**Fig. 6a**). As a rule of thumb, to detect a moderate-sized prognostic effect for a given cancer gene, for example, an increase of 50% in relative risk, the knowledge bank needed ~50–100 patients with that mutation (**Fig. 6b** and **Supplementary Fig. 7a**). Thus, for a gene mutated in 10% of patients, a training set of 500–1,000 patients would suffice, but for a gene mutated in 1% of patients a cohort of 5,000–10,000 would be needed. These simulations matched theoretical expectations[22,23] (**Supplementary Fig. 7b** and **Supplementary Note**).

The standard error of individual survival predictions 3 years after CR1 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 (**Fig. 6c**).

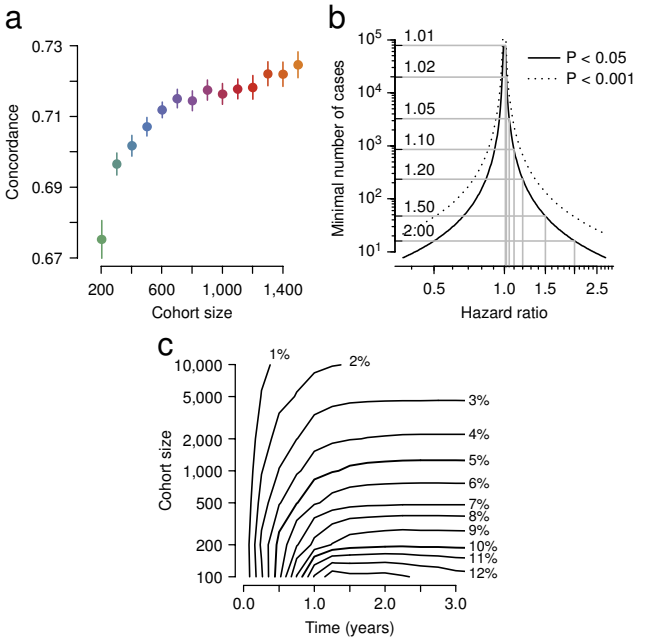

Figure 6 Extrapolations and power calculations. (**a**) Subsampling to a lower number of patients (*x* axis) shows a steady but saturating increase in prognostic concordance, *C* (*y* axis), with increasing sample size for a random-effects model of overall survival. Error bars correspond to the 95% confidence intervals for the concordance obtained from multiple independent subsamples of the data set. (**b**) Graph relating the effect size (hazard ratio) of a prognostic variable to the absolute number of patients with the given factor that were required to reach significance in a random-effects model for overall survival (solid line, $P < 0.05$; dashed line, $P < 0.001$). (**c**) Average prediction error between simulated and estimated survival in a random-effects model for overall survival as a function of survival time (*x* axis) and size of the training cohort (*y* axis).

DISCUSSION

Here we have evaluated the promise of precision medicine, building statistical models that can generate personally tailored clinical decision support from all of the available prognostic information in a knowledge bank. From a database of 1,540 patients, we were able to make considerably more informative and accurate statements about an individual's likely journey through therapy for AML than we were

with the current standards in clinical practice. Our approach enabled us to compare the likelihood of favorable outcomes under different treatment scenarios, providing information that could support genuinely personalized decision-making.

Although we have focused on AML in this analysis, we believe that the same logic will apply to knowledge banks from other cancer types, which will be generated as genomics enters healthcare and as 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, quantitative risk assessment is important in any cancer type to reserve the most aggressive treatments for the patients who are 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. After 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, which can be contrasted with the patient's baseline risk.

Building and maintaining clinical–genomic knowledge banks is a formidable challenge, especially for solid tumors in which the genome can be considerably more complex than that of AML. Initially, knowledge banks could be seeded from clinical trial cohorts, as we did here, as these will have high-quality clinical data and state-of-the-art therapies. Our power calculations suggested, however, that most clinical trials would not be powered to detect gene–drug interactions involving genes that are 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 a contentious issue. 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 predicted only small improvements in survival with early allograft (**Fig. 5b**). What may be more important is the more accurate use of a precious resource, as we could potentially reduce the number of allografts performed in patients with AML by 20–25% while maintaining the same overall rate of survival as for the current treatment recommendations. Hence, the availability of a knowledge bank would increase quality of life by reducing morbidity from chronic graft-versus-host disease, and, at a cost of $100,000–200,000 per allograft[20], the potential monetary savings would far outstrip the costs of the genomic screens. Moreover, the utility of a knowledge bank likely goes beyond these reasons by informing potential drug targets, identifying patients who are 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 and 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 from thousands of patients to inform outcomes for the individual. The therapeutic choice we exemplified is binary, constituting transplant versus chemotherapy in patients with AML. The success of inhibitors of the tyrosine kinase FLT3, whose gene is frequently mutated in patients with AML, will increase the number of available treatment options[25], and the discovery or clinical development of other agents will add further complexity. Knowledge banks could be a useful tool for clinicians navigating this complexity, but they must remain dynamic and up to date as the therapeutic armamentarium expands and as molecular understanding of cancer deepens. The logistical and regulatory hurdles, the scale needed and the costs of such an undertaking are daunting but not insurmountable challenges.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank C. Holmes for stimulating discussions. We gratefully acknowledge D. Weber for clinical data managing, V. Teleanu for assistance in cytogenetics data classification and S. Kayser for assistance in morphological evaluation. This work was supported by grants from the Wellcome Trust (077012/Z/05/Z; P.J.C.), the Bloodwise charity (P.J.C.), the Leukemia and Lymphoma Society (P.J.C.) and the Deutsche Krebshilfe (109675; K.D.), in part by grants from the German Bundesministerium für Bildung und Forschung (BMBF) (01GI9981 (H.D.) and 01KG0605 (R.F.S. and H.D.)), by a Wellcome Trust Senior Clinical Research Fellowship (WT088340MA; P.J.C.), by an EHA early career fellowship (E.P.), and by the Deutsche Forschungsgemeinschaft (project B3, Sonderforschungsbereich (SFB) 1074; K.D. and L.B.); H.D. is coordinating investigator. L.B. is a Heisenberg Professor of the DFG (BU 1339/3-1). We are grateful to all members of the German–Austrian AML Study Group (AMLSG) for their participation in this study and for providing patient samples; a list of participating institutions and investigators appears in the Appendix of Papaemmanuil *et al.*[5]. AMLSG treatment trials were in part supported by Amgen and DKH grant 109675.

AUTHOR CONTRIBUTIONS

M.G. developed the statistical methods, analyzed data and wrote the manuscript and supporting information, with input from E.P. and P.J.C. E.P. prepared and curated the genetic and clinical data. I.M. analyzed TCGA data. R.F.S., H.D., K.D., L.B., V.I.G., P.P., M.H., F.T. and A.G., along with all of the institutions contributing to the study group (AMLSG), recruited patients in this study, and collated and contributed clinical data. N.B., P.G. and U.M. provided input into analyses and interpretation of results. E.P., K.D., H.D., R.F.S. and P.J.C. initiated the study. P.J.C. and H.D. wrote the manuscript and are joint corresponding authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Van Allen, E.M. *et al.* Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat. Med.* **20**, 682–688 (2014).
2. Mardis, E.R. Genome sequencing and cancer. *Curr. Opin. Genet. Dev.* **22**, 245–250 (2012).
3. Garraway, L.A. & Lander, E.S. Lessons from the cancer genome. *Cell* **153**, 17–37 (2013).
4. McDermott, U., Downing, J.R. & Stratton, M.R. Genomics and the continuum of cancer care. *N. Engl. J. Med.* **364**, 340–350 (2011).
5. Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616–3627, quiz 3699 (2013).
6. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumors. *Nature* **490**, 61–70 (2012).
7. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107–1120 (2012).
8. Manolio, T.A. *et al.* Global implementation of genomic medicine: we are not alone. *Sci. Transl. Med.* **7**, 290ps13 (2015).
9. Collins, F.S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).
10. Papaemmanuil, E. *et al.* Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.* **374**, 2209–2221 (2016).
11. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
12. Gale, R.P., Wiernik, P.H. & Lazarus, H.M. Should persons with acute myeloid leukemia have a transplant in first remission? *Leukemia* **28**, 1949–1952 (2014).

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* **115**, 453–474 (2010).
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. *J. Am. Med. Assoc.* **301**, 2349–2361 (2009).
15. Flowers, M.E.D. & Martin, P.J. How we treat chronic graft-versus-host disease. *Blood* **125**, 606–615 (2015).
16. Burnett, A.K. *et al.* Curability of patients with acute myeloid leukemia who did not undergo transplantation in first remission. *J. Clin. Oncol.* **31**, 1293–1301 (2013).
17. Schlenk, R.F. *et al.* The value of allogeneic and autologous hematopoietic stem cell transplantation in prognostically favorable acute myeloid leukemia with double-mutant *CEBPA*. *Blood* **122**, 1576–1582 (2013).
18. Doria-Rose, V.P., Harlan, L.C., Stevens, J. & Little, R.F. Treatment of *de novo* acute myeloid leukemia in the United States: a report from the Patterns of Care program. *Leuk. Lymphoma* **55**, 2549–2555 (2014).
19. Cressman, S. *et al.* Economic impact of genomic diagnostics for intermediate-risk acute myeloid leukemia. *Br. J. Haematol.* **174**, 526–535 (2016).
20. Khera, N., Zeliadt, S.B. & Lee, S.J. Economics of hematopoietic cell transplantation. *Blood* **120**, 1545–1551 (2012).
21. Forbes, S.A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).
22. Schoenfeld, D.A. Sample-size formula for the proportional-hazards regression model. *Biometrics* **39**, 499–503 (1983).
23. Schmoor, C., Sauerbrei, W. & Schumacher, M. Sample-size considerations for the evaluation of prognostic factors in survival analysis. *Stat. Med.* **19**, 441–452 (2000).
24. Jameson, J.L. & Longo, D.L. Precision medicine—personalized, problematic and promising. *N. Engl. J. Med.* **372**, 2229–2234 (2015).
25. Stone, R.M. *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 myeloid leukemia (AML) patients (pts) age 18–60 with *FLT3* mutations (muts): an international prospective randomized (rand) *P*-controlled double-blind trial. *Blood* **126**, 6 (2015).
26. Ishwaran, H., Kogalur, U.B., Blackstone, E.H. & Lauer, M.S. Random survival forests. *Ann. Appl. Stat.* **2**, 841–860 (2008).

ONLINE METHODS

Patient cohort. Here we reanalyzed data that were first reported and described in detail by Papaemmanuil *et al.*[10]. Briefly, we performed targeted gene sequencing of 111 myeloid cancer genes[11,27–29] using 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[30–32]. 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 (ATRA) therapy or not in induction chemotherapy. 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 enrollment in the multicenter trials, which were approved by the local research ethics committee of each participating site (ClinicalTrials.gov number: NCT00146120).

Statistical methods. We explored a range of statistical methods to build models of overall survival[33,34], 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 Note**, subsections 2–4). We found little prognostic significance in whether the mutations were subclonal or clonal (**Supplementary Note**), and we therefore did not consider this information in the multivariate models. All of the predictions shown were based on a leave-one-out basis; it was 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 divided patients into two groups, based on whether they were expected to derive more or less than a 10-percentage-point improvement in survival with allograft in CR1 as compared to that with chemotherapy in CR1 and allograft at relapse. In each group, the observed outcomes were then determined separately for those patients 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, as this would ensure that they were not confounded with the predictor variables that we used. Here, 711/1,540 (46%) of patients received an allograft, with the decision to perform a transplant in intermediate-risk patients based on whether a matched related donor was available[32]. This introduces a quasi-randomization, as human leukocyte antigen (HLA)-matching between siblings derives from Mendelian assortment of parental alleles, but this cannot substitute for the prospective validation of the decision support tools we developed.

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

Data availability. Sequencing data that support the findings of this study have been deposited in the European Genome–phenome Archive under accession EGAS00001000275. The exploratory web application to visualize and explore the data is found at http://cancer.sanger.ac.uk/aml-multistage. 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 of the code used to generate the figures and conclusions of the manuscript.

27. Ley, T.J. *et al. DNMT3A* mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
28. Welch, J.S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
29. Delhommeau, F. *et al.* Mutation in *TET2* in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
30. Schlenk, R.F. *et al*; German-Austrian Acute Myeloid Leukemia Study Group. All-*trans* retinoic acid as adjunct to intensive treatment in younger adult patients with acute myeloid leukemia: results of the randomized AMLSG 07-04 study. *Ann. Hematol.* **95**, 1931–1942 (2016).
31. Schlenk, R.F. *et al.* Phase 3 study of all-*trans* retinoic acid in previously untreated patients 61 years or older with acute myeloid leukemia. *Leukemia* **18**, 1798–1803 (2004).
32. Schlenk, R.F. *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.* **28**, 4642–4648 (2010).
33. Therneau, T.M., Grambsch, P.M. & Pankratz, V.S. Penalized survival models and frailty. *J. Comput. Graph. Stat.* **12**, 156–175 (2003).
34. Shah, R.D. & Samworth, R.J. Variable selection with error control: another look at stability selection. *J. R. Stat. Soc. Series B Stat. Methodol.* **75**, 55–80 (2013).

doi:10.1038/ng.3756