Chapter 3

Predicting acute myeloid leukaemia risk in the general population

1. Introduction

As discussed in Chapter 1, CH harbouring canonical leukaemia-associated mutations is a risk factor for haematological malignancy, though only a small minority of affected individuals progress (Bowman et al., 2018). Acute myeloid leukaemia (AML) is the commonest acute leukaemia in adults and typically presents suddenly as a fulminant disease with a poor prognosis (Döhner et al., 2015). This chapter describes an experiment to distinguish individuals at high risk of developing *de novo* acute myeloid leukaemia (AML) from those with indolent CH at low risk of malignant transformation. The introduction provides background on AML and reviews existing literature on its pre-clinical evolution and relationship to clonal haematopoiesis.

1.1 Acute myeloid leukaemia

1.1.1 Definition and epidemiology

AML is an aggressive haematopoietic stem cell disorder characterized by clonal proliferation of poorly differentiated myeloid cells (Döhner et al., 2015). It is the commonest acute leukaemia among adults, and comprises around 20% of all paediatric leukaemia (Döhner et al., 2015). The incidence of AML increases dramatically with age, and exceeds 100 cases per 100,000 in those over the age of 60, with a higher risk among men (CRUK, 2018;

SEER, 2018). There are around 3,100 new AML cases and 2,500 AML-related deaths each year in the UK (CRUK, 2018).

1.1.2 Aetiology and risk factors

The dominant AML risk factor is age, though the role ageing plays in the aetiology of AML is incompletely understood (Döhner et al., 2015). The somatic mutation burden seen in AML correlates with age at diagnosis and is similar to that observed in normal HSCs from age-matched individuals without a haematological disorder (Welch et al., 2012). Unlike many common adult epithelial cancers, the role of extrinsic mutational processes appears to be minor, with the age-related mutational SBS11 and SBS5 accounting for the vast majority of AML mutations (Alexandrov et al., 2018; Alexandrov et al., 2013).

Environmental or occupational chemical exposures, notably to benzene and other industrial solvents, may play a role in a minority of AML cases, though evidence for a causal link is weak (Austin et al., 1988).

Germline variants in a growing number of genes have been implicated in myeloid malignancies, including *RUNX1, GATA2, TERT, ATG2B, TP53* and *CEBPA* (Hinds et al., 2016; Saliba et al., 2015; Smith et al., 2004; Zhang et al., 2015). As discussed in the general introduction, germline and somatic mutations in the same cancer gene generally carry different biological and clinical significance and merit distinction (Arber et al., 2016; Döhner et al., 2015). Furthermore, recent evidence has suggested that the distinction between germline and somatic mutation is less clear than previously thought, with a growing catalogue of highly penetrant germline variants strongly predisposing to acquisition or clonal selection of particular somatic mutations (Hinds et al., 2016; Loh et al., 2018).

Other myeloid neoplasms, most commonly myeloproliferative neoplasms and myelodysplastic syndromes, may transform into AML, termed secondary AML (sAML) (Deininger et al., 2017; Sperling et al., 2017).

The most prevalent extrinsic risk factor for AML is previous exposure to chemotherapy or radiotherapy, in particular alkylating agents and topoisomerase II inhibitors (McNerney et al., 2017). Any AML that arises after cytotoxic treatment is termed therapy-related AML (t-AML) and is discussed further in the introduction to Chapter 5.

AML that presents suddenly with manifestations of bone marrow failure is termed *de novo* AML to distinguish it from sAML and t-AML, although, as discussed later on, these distinctions are not always straight-forward or biologically meaningful.

1.1.3 AML genetics

The genetic diversity of AML was first revealed by cytogenetic analyses in the 1970s (Rowley, 2008), and has since been well characterised by several large genomic studies (Arber et al., 2016; Gerstung et al., 2017; Papaemmanuil et al., 2016; TCGA et al., 2013). According to the classic "two-hit" model of AML leukaemogenesis proposed by Gilliland and Griffin, two types of mutations are required to produce AML: type II mutations that impair differentiation and subsequent apoptosis and are typically initiating events, and type I mutations that endow pre-leukaemic clones with a proliferative advantage (Gilliland and Griffin, 2002). Genomic studies have corroborated the main concepts of this model, providing further evidence that the block in differentiation is the initiating event for *de novo* AML. Many of the commonest mutations in AML founding clones target epigenetic regulators (Kronke et al., 2013; Shlush et al., 2014; Welch, 2014), which play central roles in haematopoietic stem cell differentiation (Abdel-Wahab et al., 2012; Challen et al., 2011; Figueroa et al., 2010a; Figueroa et al., 2010b). Furthermore, leukaemia-associated mutations in epigenetic regulators are common drivers of CH, whereas 'type I' mutations are very rarely observed in association with CH, consistent with this class of genetic events occurring later in leukaemogenesis after differentiation arrest has been established (Genovese et al., 2014; McKerrell et al., 2015; Xie et al., 2014).

Although this model remains conceptually useful, sequencing studies have revealed diverse genetic routes to AML, with recurrent mutations identified in over 70 genes (Papaemmanuil et al., 2016; TCGA et al., 2013). The majority of patients harbour multiple driver events, and both individual mutations and co-occurrence patterns are powerful determinants of clinical outcome (Gerstung et al., 2017; Huet et al., 2018; Papaemmanuil et al., 2016). The most recurrent structural and numerical chromosomal abnormalities include t(8;21), inv(16), t(15;17), 11q (MLL) fusions, inv(3), t(6;9), -7/7q, +8/8q, -5/5q and -17/17p (Papaemmanuil et al., 2016; TCGA et al., 2013). The majority of driver events in adult AML, however, are point mutations (single nucleotide variants and indels)(Papaemmanuil et al., 2016; TCGA et al., 2013). Frequently mutated genes include epigenetic regulators (*DNMT3A,*

TET2, IDH1, IDH2), genes involved in the RNA splicing machinery (*SF3B1, SRSF2, U2AF1, ZRSR2*), chromatin regulators (*ASXL1, BCOR, STAG2, MLL-PTD, EZH2, PHF6*), transcription factors (*RUNX1, GAT2, CEBPA*), *NPM1*, and genes involved in RAS and/or STAT signalling (*NRAS, KRAS, PTPN11, NF1, FLT3, CBL, KIT*)(Papaemmanuil et al., 2016; TCGA et al., 2013).

1.1.4 AML classification schemes

The World Health Organisation (WHO) Classification of Haematopoietic and Lymphoid Tissues subdivides AML into four categories: AML with recurrent genetic abnormalities, AML with myelodysplasia-related changes, therapy-related AML and AML not otherwise specified (NOS)(Arber et al., 2016). The latter group is further subdivided by morphological features. The WHO classification scheme was updated in 2016 to include several new disease categories within the section of AML with recurrent genetic abnormalities (Arber et al., 2016). However, several studies suggest that WHO subgroups still do not adequately capture the molecular heterogeneity of AML, which underpins its biological and prognostic features (Gerstung et al., 2017; Metzeler et al., 2016; Papaemmanuil et al., 2016). The largest genomic study of AML to date included 1540 patients enrolled in three prospective clinical trials and identified eleven prognostically relevant molecular-genetic subgroups (Gerstung et al., 2017; Papaemmanuil et al., 2016). This study added considerable nuance to our understanding of AML biological mechanisms and genetic classification. For example, mutations affecting different loci in the same gene, e.g., *IDH2* p.R140 and *IDH2* p.R172, had divergent co-occurrence patterns and impacts on clinical outcome.

1.1.5 Treatment challenges

Despite much progress in understanding AML genetics and pathogenesis, standard AML therapy has changed very little over the past three decades (Döhner et al., 2015; Yates et al., 1973). The backbone of therapy remains the combination of two drugs developed in the 1950s, namely daunorubicin and cytarabine, compounds serendipitously derived from soil microbes and marine sponges, respectively (Schwartsmann et al., 2001; Stutzman-Engwall and Hutchinson, 1989). Improvements in patient outcomes are primarily attributable to better supportive care during periods of myelosuppression (Döhner et al., 2015). Although most patients capable of tolerating intensive chemotherapy achieve remission, the majority

succumb to relapse (Döhner et al., 2015; Rubnitz et al., 2014). Overall survival rates are 35% to 40% for younger patients and 5% to 15% for patients over the age of 60 (Dohner et al., 2010; Rubnitz et al., 2014). Efforts to target recurrently mutated oncogenes, notably the tyrosine kinases FLT3 and KIT, have been met with rapid emergence of disease resistance and little improvement in overall survival (Döhner et al., 2015; Stein, 2015; Wander et al., 2014).

1.2 The relationship between CH and AML

As discussed in Chapter 1, the two largest studies of clonal haematopoiesis in the general population demonstrated an increased risk of haematological cancers in general (not specifically AML) in those with CH, which was higher in those with mutations at high VAFs (Genovese et al., 2014; Jaiswal et al., 2014). Genovese et al. identified thirty-one participants diagnosed with a hematologic cancer more than 6 months after DNA sampling, of whom thirteen (42%) had antecedent CH (Genovese et al., 2014). Of these, two developed AML and one developed "acute leukemia of unspecified origin". Of the remaining ten, three developed CLL, two MPN (both *JAK2* V617F mutated), one B-cell lymphoma, one multiple myeloma, one monoclonal gammopathy of unknown significance, one CMML and one MDS (Genovese et al., 2014). Two of the three MDS/AMLs in this paper were diagnosed within two months after DNA sampling (Genovese et al., 2014). Furthermore, Genovese et al. found that CH with putative drivers (CH-PD) afforded the same risk of haematological cancers as CH without known drivers, potentially alluding to indirect risks associated with CH (Jaiswal et al., 2014). Similarly, Jaiswal et al. reported sixteen haematological cancers during a median 95-month follow-up period, of which only five (31%) had CH detected in their pre-diagnosis sample (Jaiswal et al., 2014). Of these, two developed lymphoma, one "cancer of the spleen" (*JAK2* V617F mutated), one "myeloid leukaemia" and one "leukaemia" not otherwise specified (Jaiswal et al., 2014). Together, these two studies captured up to five possible AMLs amongst 29,652 study participants (Genovese et al., 2014; Jaiswal et al., 2014). Collectively, only a minority of blood cancers arising during follow-up were diagnosed in individuals with antecedent CH, and several of these were indolent myeloproliferative or chronic lymphoid conditions. It therefore remained unclear whether or not CH could be used to predict the subsequent development of blood cancers, let alone of *de novo* AML, with any degree of sensitivity or specificity.

2. Results

To investigate whether individuals at high risk of developing *de novo* AML can be distinguished from those with benign CH, genes recurrently mutated in AML or CH were deep-sequenced in peripheral blood cell DNA from a total of 125 individuals sampled before AML diagnosis (pre-AML group), together with 676 unselected age- and gender-matched individuals (control group). To detect somatic mutations with maximum sensitivity, deep error-corrected targeted sequencing was first applied to a discovery cohort of 95 pre-AML cases sampled on average 6.3 years before AML diagnosis and 414 age- and gender-matched controls (Appendix 1). Error-corrected sequencing was performed by Dr Sagi Abelson as detailed in Methods section 2.1. A validation cohort comprising 29 pre-AML cases and 262 controls (Appendix 2) was analysed using conventional deep sequencing with an overlapping gene panel (Methods section 2.2).

2.1 Prevalence of CH-PD in pre-AML versus controls

Taking both cohorts together, CH, defined by the presence of mutations in putative driver genes (CH-PD), was found in 73.4% of the pre-AML cases at a median of 7.6 years before diagnosis (Appendices 8 and 9). By contrast, CH-PD was observed in 36.7% of controls ($P < 2.2 \times 10^{-16}$, two-sided Fisher's exact test; Figure 3.1a). This CH-PD prevalence in the controls is consistent with data from a study of more than 2,000 healthy individuals assayed using a similarly sensitive error-corrected sequencing method (Acuna-Hidalgo et al., 2017). Additionally, 39% of pre-AML cases over age 50 had a driver mutation with a VAF exceeding 10%, compared to only 4% of controls, a prevalence that is in line with the largest studies of CH-PD in the general population (Genovese et al., 2014) ($P < 2.2 \times 10^{-16}$, two-sided Fisher's exact test; Figure 3.1b). The median number of driver mutations per individual increased with age and was significantly higher in the pre-AML group relative to controls ($P < 2.2 \times 10^{-16}$, two-sided Wilcoxon rank-sum test; Figure 3.1c). Furthermore, examination of VAF distribution revealed significantly larger clones among the pre-AML cases ($P = 1.2 \times 10^{-13}$, two-sided Wilcoxon rank-sum test; Figure 3.1d).

Figure 3.1

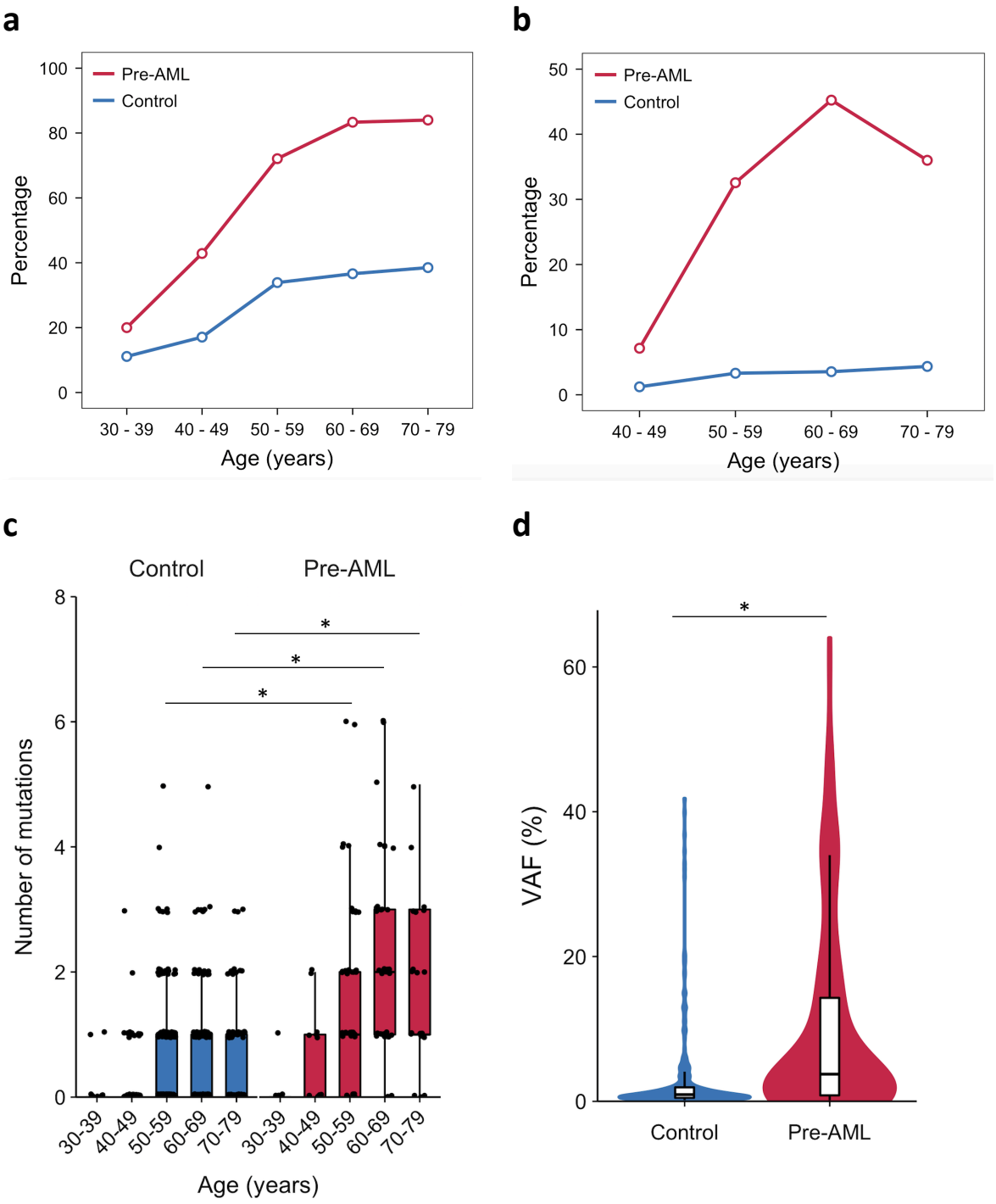

Figure 3.1 | Prevalence of CH-PD, number of mutations and clone size in pre-AML and control cohorts. a, Prevalence of CH-PD among pre-AML cases (red) and controls (blue). **b,** Prevalence of CH-PD clones with VAF > 10% among pre-AML cases (red) and controls (blue). **c**, The number of CH-PD mutations detected in cases and controls according to age. Box plot centres, hinges and whiskers represent the median, first and third quartiles and 1.5× interquartile range, respectively. **d**, VAF of CH-PD mutations. All panels show data for n = 800 biologically independent samples. *P < 0.0005, two-sided Wilcoxon rank-sum test with Bonferroni multiple testing correction.

2.2 Clonal dynamics over time and evolution to AML

In order to explore the mechanisms underpinning the higher mutation burden in pre-AMLs and the relationship between CH-PD and future leukaemia, I sequenced serially collected samples available for a subset of the VC (12 pre-AMLs and 141 controls) as well as three FFPE-fixed bone marrow biopsy samples available from AML diagnosis (PD29962, PD30054, PD30089). Comparison of the pre-AML mutations to the mutations detected in the diagnostic specimen demonstrated that most, though not all, drivers persisted and of these only a subset expanded to become clonal in the future AML (Figure 3.2a-c). The sensitivity of sequencing for the AML diagnostic samples was limited by the low quality of the FFPE-derived DNA and variable sequencing coverage. For PD29962, no putative drivers with VAF exceeding 9% were detected at diagnosis. In this individual, a clone harbouring a *TET2* p.E852* variant persisted for over 14 years, but decreased in size. A *KRAS* p.G12D variant also detected pre-diagnosis became undetectable, though with only 79 reads covering this locus in the diagnosis DNA, it is possible that it persisted at a subclonal level. Both PD30054 and PD30089 show evidence of persistent clones that became clonal in the AML, as well as new drivers present at diagnosis. PD30089 also developed a *JAK2* p.V617F-mutated clone, which persisted but decreased in size. For an additional case (PD29918), a third blood sample was taken very close to AML diagnosis (~1 month prior), demonstrating an *SRSF2* p.P95R mutation detected at all three time points (Figure 3.2d), which almost certainly contributed to the AML, while the second mutation detected (*TET2* p.S354*) persisted at declining VAF. Furthermore, data from individuals for whom blood sampling was done less than a year before AML diagnosis (n=9) show that the majority of these cases have driver mutations at high VAF (Figure 3.2e-f, Appendix 9), again suggesting that the pre-AML clones detected are likely to include those that later evolved into AML in most cases. Collectively these findings suggest that the driver mutations identified in pre-AML cases may represent a combination of pre-leukaemic clones as well as additional 'bystander' clones which do not transform. Several studies suggest that such independent clones may be common in AML patients at diagnosis (Parkin et al., 2017; Wong et al., 2015a). For example, a recent study of patients undergoing induction therapy found that five out of fifteen had marked expansion of clones unrelated to the founding AML clone but detectable in diagnostic specimens using error-corrected sequencing (Wong et al., 2015a).

Figure 3.2

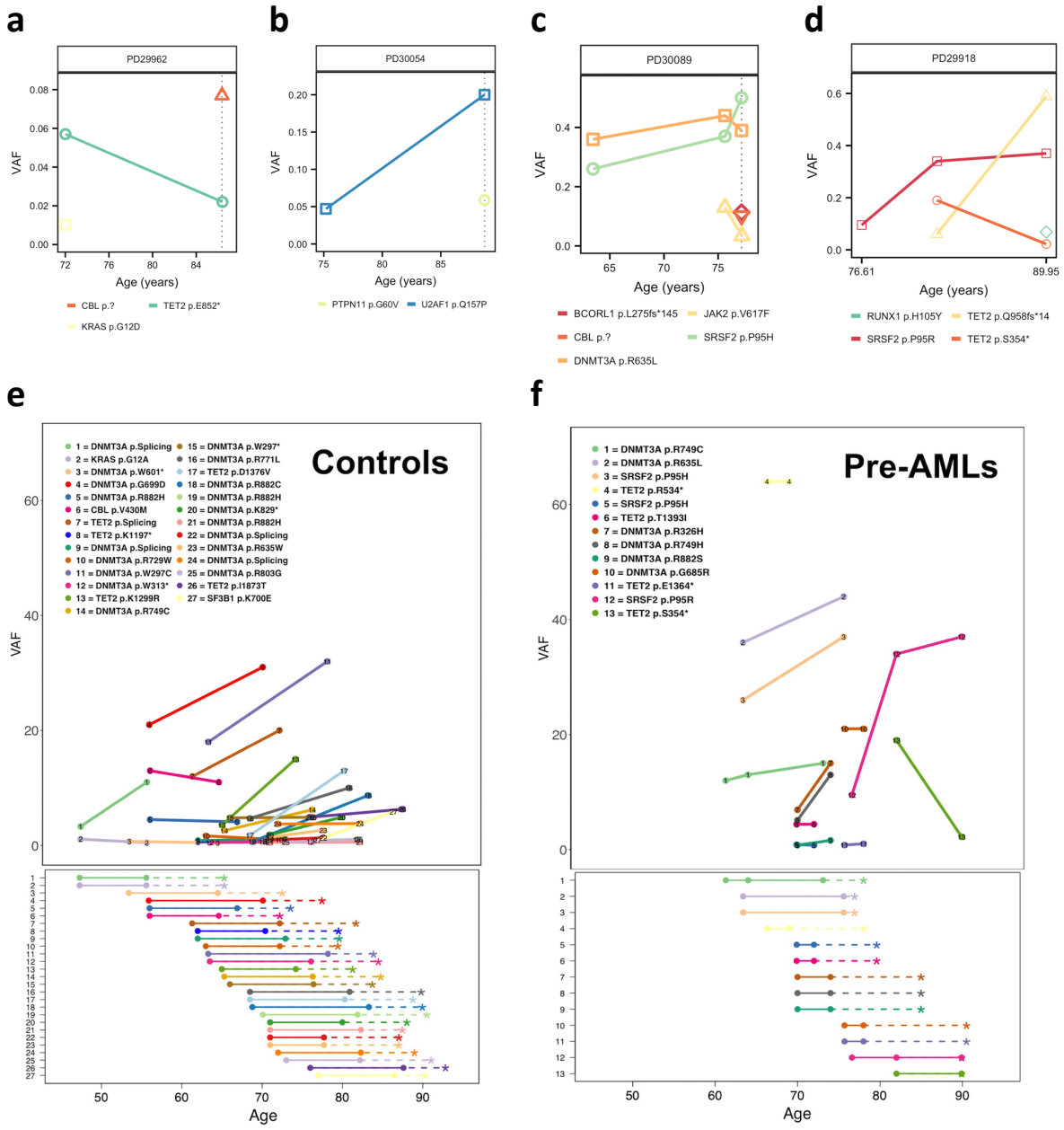

Figure 3.2 | Evolution of clonal haematopoiesis and relationship with future AML. **a-c,** VAF trajectories of putative driver mutations in three individuals for whom bone marrow biopsy specimens taken at time of AML diagnosis (dashed black vertical line) were available for sequencing. Note that coverage for the diagnostic sample of PD30089 was insufficient to meaningfully compare the relative VAFs of the drivers in *DNMT3A* and *SRSF2*. **d,** VAF trajectories of driver mutations in an individual sampled three times, with last sample taken one month before AML diagnosis.
e,f, VAF trajectory of persistent clones carrying putative driver mutations in controls (**e**) and pre-AML cases (**f**). Upper plots: Circles denote individual serial samples and solid lines representing the growth trajectory between serial samples. Lower plots: dashed lines indicate the time interval between the last sampling and the end of follow-up (controls) or AML diagnosis (cases). Code for panels e and f by Dr Sagi Abelson.

We sought to formally assess whether the clonal expansion rate was significantly different for the serial samples taken from controls versus pre-AMLs. However, this measurement is confounded by multiple factors, not least the inability to determine whether or not co-occurring mutations reside in the same clone. Hence, this experiment is inadequate to draw any conclusions. Studying the impact of mutation on AML development at the clonal level, for example by culturing and sequencing single-cell derived colonies, would help to address this question (Nangalia et al., 2019).

2.3 The genetic landscape of pre-AML versus CH

In line with previous studies of CH in the general population (Jaiswal et al., 2014; Xie et al., 2014), *DNMT3A* and *TET2* were the most commonly mutated genes in both groups (Figure 3.3a). No canonical *NPM1* mutations nor any *FLT3*-internal tandem duplication mutations were detectable, consistent with these arising late in leukaemogenesis (Kronke et al., 2013; McKerrell et al., 2015). Recurrent *CEBPA* mutations, which are implicated in around 10% of *de novo* AML (Papaemmanuil et al., 2016), were also absent, suggesting that driver events in this gene may also be late events in *de novo* AML evolution, despite their involvement in familial AML. Notably, mutations in splicing factor genes (*SF3B1, SRSF2* and *U2AF1*) were significantly enriched among the pre-AML cases relative to the controls (odds ratio, 17.5; 95% confidence interval, 8.1–40.4; $P = 5.2 \times 10^{-16}$, two-sided Fisher's exact test) and were present in significantly younger individuals (median age 60.3 compared to 77.3 years, $P = 1.7 \times 10{-}4$, two-sided Wilcoxon rank-sum test; Figure 3.3b). Screening all SNPs for potential pathogenic germline variants relevant to cancer or blood disorders (Methods section 3.4) identified only one likely pathogenic lesion, *MPL* p.Q186K (ClinVar accession RCV000015217.22). This SNP has been implicated in congenital amegakaryocytic thrombocytopenia (Ihara et al., 1999), though the participant carrying it (PD30060) had normal pre-diagnosis blood counts and developed AML aged 91.

Figure 3.3

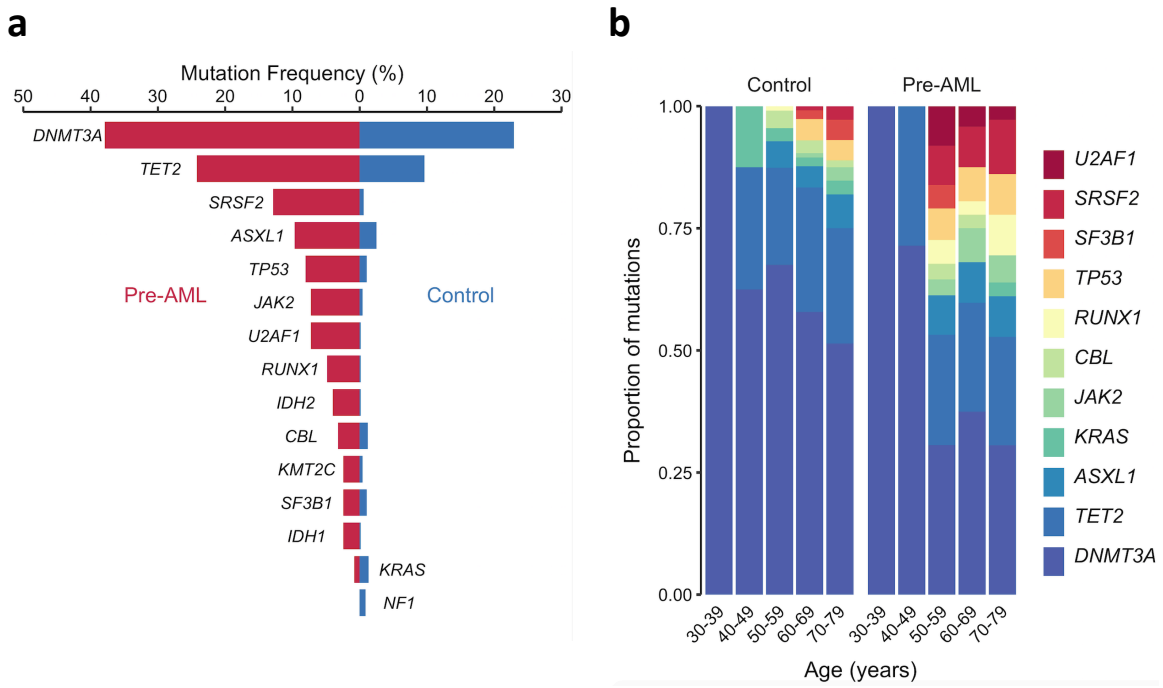

Figure 3.3 | The mutational landscape of clonal haematopoiesis in pre-AML and controls. a, Proportion of pre-AML cases (red) and controls (blue) who had CH-PD mutations in recurrently mutated genes. **b**, Relative frequency of mutations in the indicated genes according to age group for pre-AML cases and controls. **P* < 0.05, Fisher's exact test with Bonferroni multiple testing correction.

2.4 Genetic AML risk prediction model

These findings demonstrate marked differences in both mutation burden and driver landscape between CH-PD observed in controls and pre-AML. Moreover, these results, in conjunction with recent insights into the origins of AML relapse (Shlush et al., 2017), suggests that AML progression typically occurs over many years through clonal evolution of pre-leukaemic haematopoietic stem and progenitor cells (HSPCs) before acquisition of late mutations leads to overt malignant transformation. In order to quantify the relative contributions of driver mutations and clone sizes to the risk of progressing to AML, we applied a Cox proportional hazards regression approach, which achieved similar performance in both the discovery cohort (concordance (C) = 0.77 ± 0.03) and the validation cohort (C = 0.84 ± 0.05; Figure 3.4a-f and Table 3.1). A ridge regularised logistic regression model trained using the same variables produced very similar results (Table 3.2) As discussed in Methods section 4.1, we used weighting to minimise the biases introduced by the artificial case-control ratio (Antoniou et al., 2005; Therneau and Grambsch, 2000) and calculated hazard ratios relative to the (approximate) true cumulative incidence of about 1-3/1,000 in the given age range over a follow up of 10-20 years. The observed driver mutation frequency and VAF in pre-malignant samples closely resembled values expected based on the estimated risks, indicating that risk model and driver prevalence are well aligned (Figure 3.4g-h).

Table 3.1 Cox proportional hazard model performance

Cox proportional hazards model	Concordance	Standard error	Time-dependent AUC
VC data and fit	0.84	0.05	0.74
DC data and fit	0.77	0.03	0.78
VC fit DC data	0.72	0.03	0.7
DC fit VC data	0.82	0.05	0.79
Combined cohorts	0.77	0.05	0.79*

*Derived from 100 bootstraps out-of-bag validation

DC, discovery cohort; VC, validation cohort

Figure 3.4

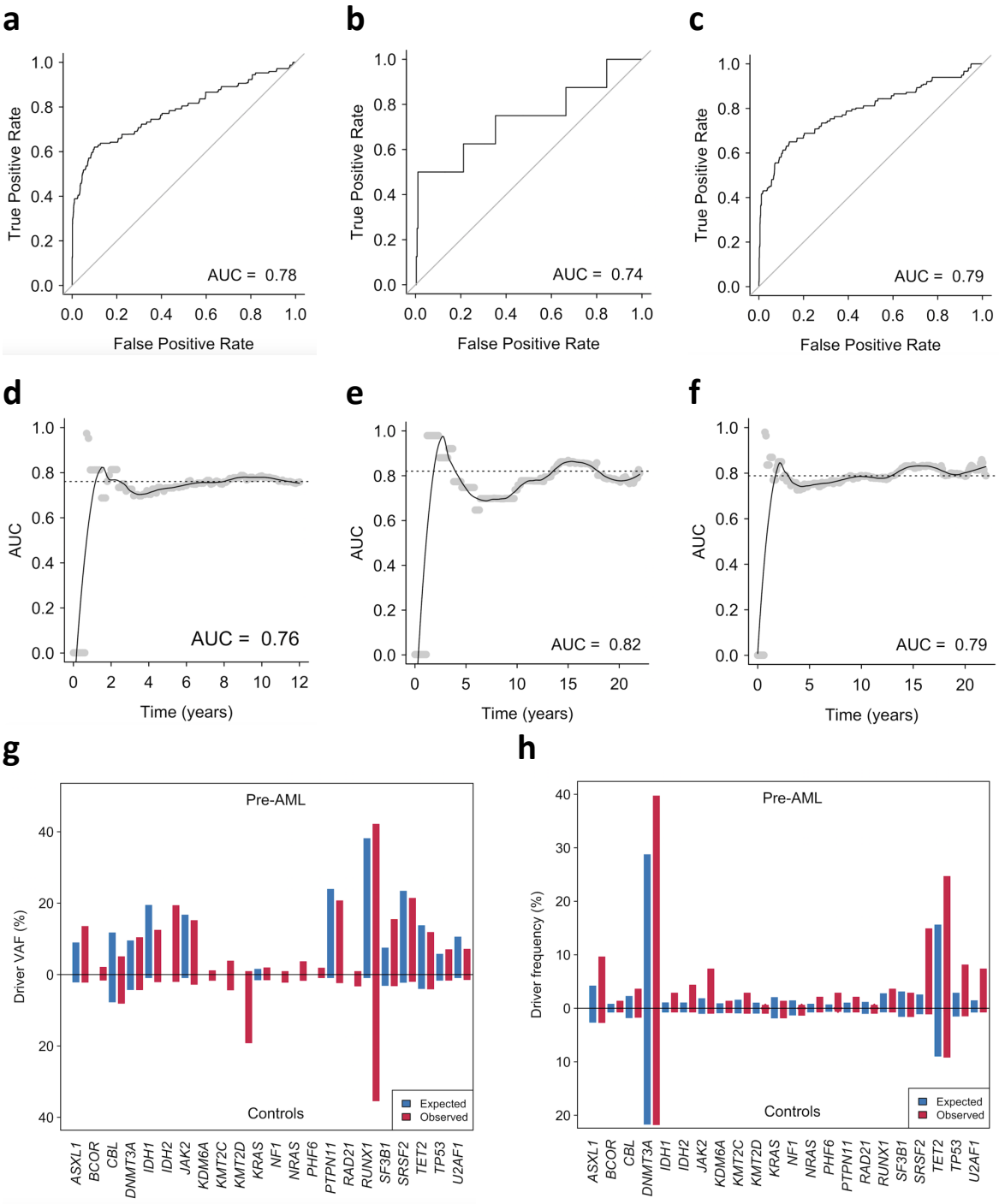

Figure 3.4 | AML predictive model performance. a–c, Time-dependent receiver operating characteristic curve for Cox proportional hazards model of AML-free survival trained on the discovery cohort (n = 505 unique individuals, 91 pre-AML and 414 controls) (**a**), validation cohort (n = 291 unique individuals, 29 pre-AML and 262 controls) (**b**) and combined cohorts (**c**). **d–f**, Dynamic AUC for Cox proportional hazards models trained on the discovery cohort (**d**), validation cohort (**e**) or combined cohort (**f**). **g, h**, Red and blue bars indicate the observed and expected VAF (**g**) and driver frequency (**h**) of pre-AML cases and controls for each gene indicated on the *x* axis. One can speculate that the discrepancies between expected and observed driver VAF for RUNX1 and KMT2D relate to the relatively high prevalence of pathogenic germline mutations seen in these genes and the challenge in distinguishing the latter from somatic drivers.

Table 3.2 Ridge regularised logistic regression model performance

Ridge regularised logistic regression	AUC
VC data and fit	0.85
DC data and fit	0.76
VC fit DC data	0.69
DC fit VC data	0.81
Combined	0.81*

*Derived from 100 bootstraps out-of-bag validation
DC, discovery cohort; VC, validation cohort

Models that were only trained on data from the discovery or validation cohort had similar coefficients (Figure 3.5, Appendix 10). We therefore combined the datasets for a more accurate analysis of the contributions of mutations in individual genes to risk (C = 0.77 ± 0.05; area under curve, 0.79; Figure 3.4c,f and Table 3.1).

Quantitatively, we found that driver mutations in most genes conferred an approximately twofold increased risk of developing AML per 5% increase in clone size (Figure 3.5). Notable exceptions to this trend were the most frequently mutated CH genes, *DNMT3A* and *TET2*, which conferred a relatively lower risk of progression to AML (Figure 3.5, Fig 3.6a,c,e). By contrast, a larger effect size was apparent for *TP53* (hazard ratio, 12.5; 95% confidence interval, 5.0–160.5) and *U2AF1* (hazard ratio, 7.9; 95% confidence interval, 4.1–192.2) mutations (Figure 3.5, Figure 3.6a,b,d). However, other CH-PD genes, such as *SRSF2*, contributed a similar relative risk owing to their presence at a higher VAF in pre-AML cases (Figure 3.5, Figure 3.6a). Because the effect of each driver mutation is deleterious and the effect of multiple mutations that are present in the same individual is multiplicative, a higher number of mutations is predicted to increase the risk of progression to AML (Figure 3.7a). Similarly, the size of the largest driver clone was also strongly associated with the risk of progression to AML, in agreement with the risk of individual mutations generally being proportional to VAF (Figure 3.7b).

Estimates of model sensitivity and specificity necessitate arbitrary age-cut-offs which dramatically impact the interpretation of predictions. Is it most relevant to know whether or not an individual will develop AML before age 100 or before age 60 and which estimate should sensitivity/specificity be determined for? The Cox proportional hazards model illustrated in

figure 3.5 facilitate a more tangible interpretation of excess risk on an individual level, harnessing the genomic snapshot from a blood sample to estimate the risk of developing AML over the next 10 years in a manner which accounts both for a person's age and the incidence of AML in their given age bracket.

Comparing AML risk prediction models based on the VAF of mutations in individual genes versus mutation burden alone demonstrated that the gene-level model performed best (Figure 3.7c,d). Concordance and AUC were both 3-4% improved for the models incorporating gene-level risk, which is a considerable margin, particularly for a rare disease. Moreover, the disparities in gene-level hazard ratios (HR) were significant (Figure 3.5), despite the fact that the genes with the highest HR are not mutated frequently enough to have a very dramatic effect on overall model AUC. Collectively, although the VAF and the number of mutations confer much of the predictive value, the gene-level analysis (Figure 3.5) does demonstrate distinct gene-level risks, and is able to quantify the cumulative impact of multiple mutations and clonal size on the likelihood of progression to AML. Furthermore, in order to examine whether the genetic model can distinguish between CH-PD and pre-AML even when individuals without mutations were excluded, we retrained the model using only cases and controls with CH-PD. We found that performance was if anything marginally improved by this manoeuvre (Concordance > 0.8 on both discovery and validation cohorts, Appendix 7).

Figure 3.5

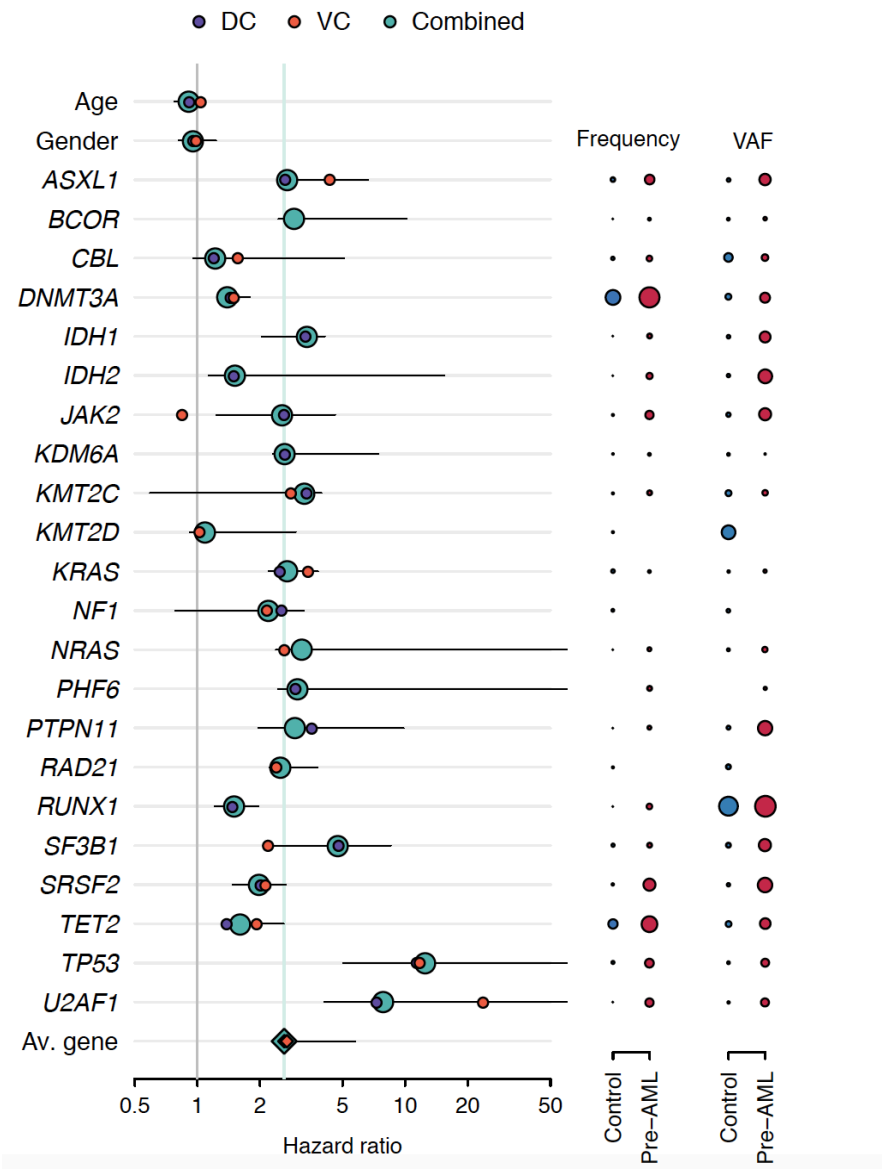

Figure 3.5 | Forest plot indicating gene-level hazard ratios for risk of developing AML. Purple, orange and green circles indicate hazard ratios (HR) for the discovery (DC), validation (VC) and combined cohort, respectively. Horizontal lines denote 95% confidence intervals for the combined cohort. For each gene, the indicated HR applies to the 10-year risk of AML conferred by each 5% increase in mutation VAF. The green vertical line indicates the mean HR across all genes. The HR for *RUNX1* must be interpreted with caution owing to the relatively high prevalence of deleterious germline variants in this gene, which may not be readily distinguishable from somatic mutations in unmatched sequencing assays. The proportion of individuals with mutations in each gene and the average VAF are indicated to the right of the forest plot.

Figure 3.6

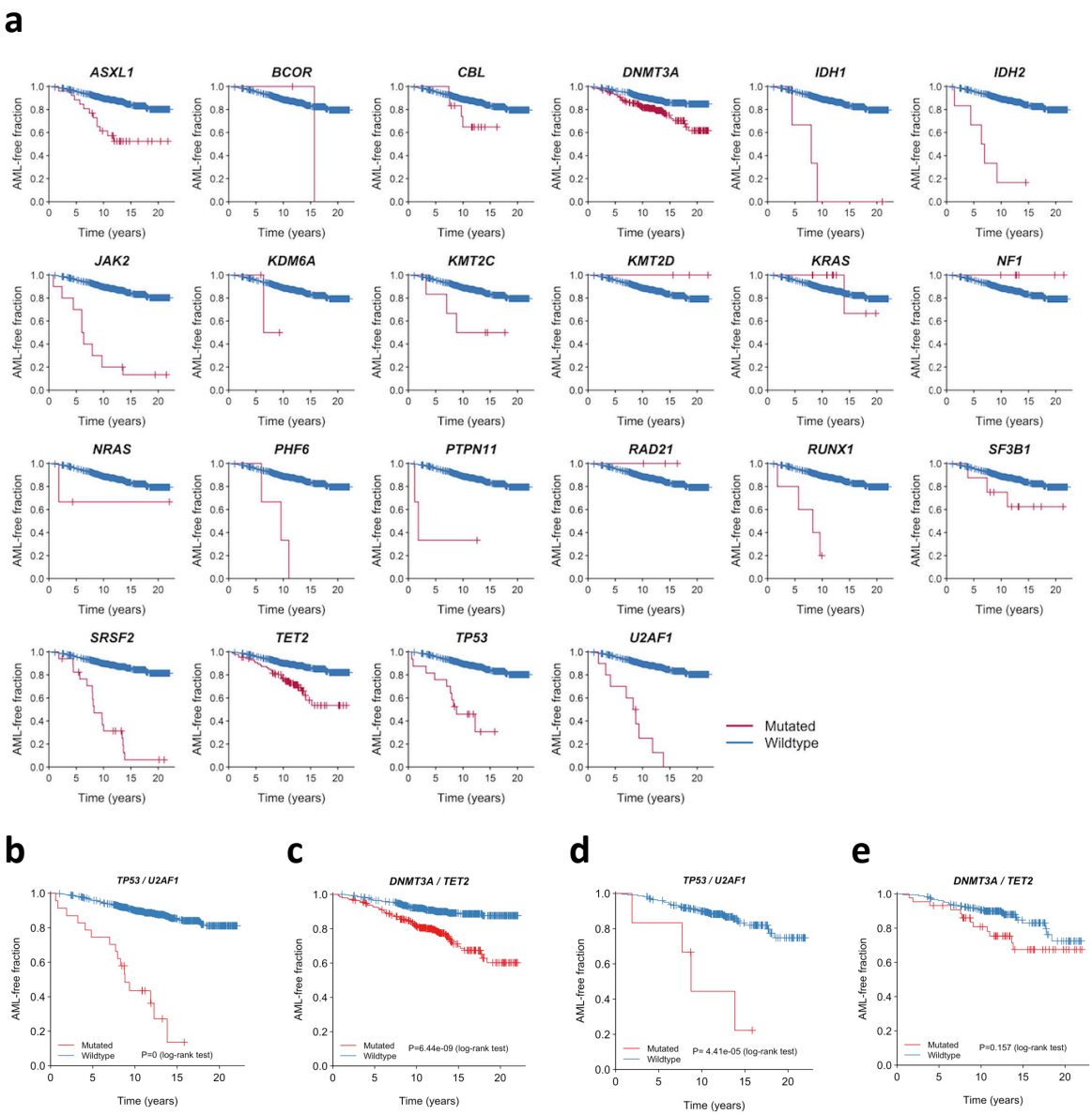

Figure 3.6 | Gene-level impact on AML-free survival. a, Kaplan–Meier (KM) curves of AML-free survival, defined as the time between sample collection and AML diagnosis, death or last follow-up. Survival curves are stratified according to mutation status in genes mutated in at least three samples across the combined validation and discovery cohorts. *n* = 796 unique individuals. **b-c** For illustrative purposes, KM curves according to co-mutation status in *DNMT3A/TET2* and *TP32/U2AF1* are shown. All patients harbouring any mutation in *TP53* or *U2AF1* (**b**) or *DNMT3A* or *TET2* (**c**). **d,e** The same relationship between mutation status and AML-free survival persists when considering only individuals with a total of one driver mutation. KM curves for participants with their only driver mutation in either *DNMT3A* or *TET2* (**d**) or *U2AF1* or *TP53* (**e**). Red and blue lines indicate mutated and wildtype, respectively. *P*-values for significance of survival differences by mutation status calculated by the log-rank test. AML, acute myeloid leukaemia; KM, Kaplan-Meier.

Figure 3.7

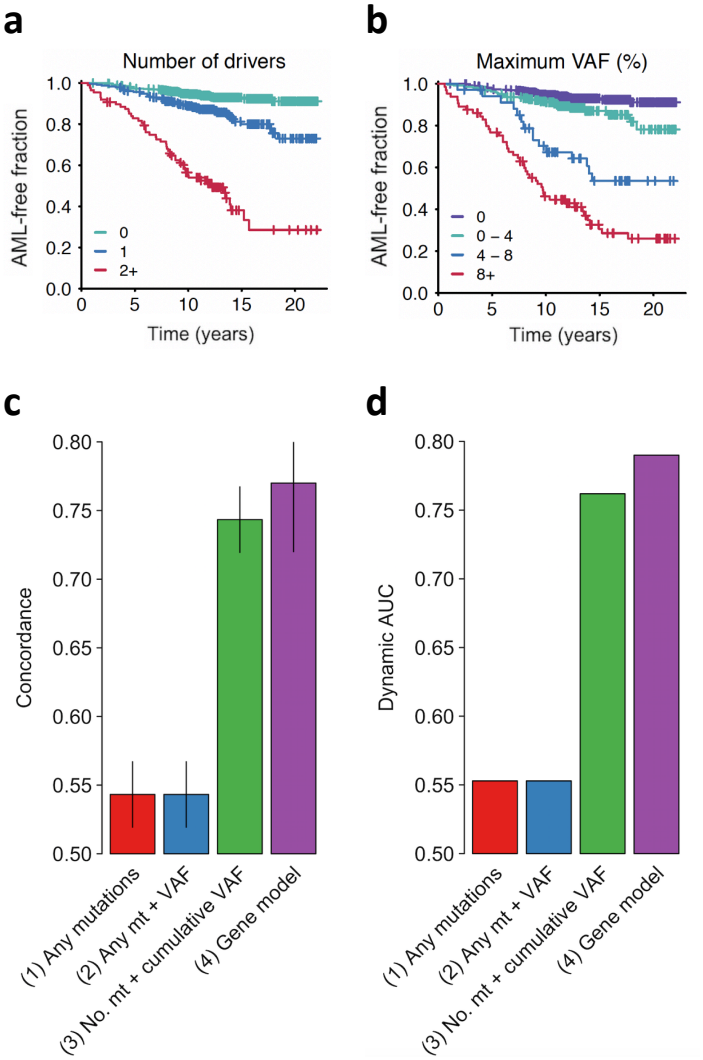

Figure 3.7 | Performance of AML risk prediction models based on gene-level factors versus mutation burden. **a-b**, Kaplan–Meier curves of AML-free survival, defined as the time between sample collection and AML diagnosis, death or last follow-up. Survival curves are stratified according to number of driver mutations per individual (**a**) and largest clone detected (**b**). VAF bins of 4% are shown in (**b**) to illustrate the consistency of the trend towards lower AML-free survival with larger clone size. **c**, Leave-one-out crossvalidated concordance C of different risk models based on (1) the presence of any mutation, (2) the presenced of any mutation and the cumulative VAF of different clones, (3) the number of different driver mutations and cumulative VAF as predictors and (4) a model incorporating the effects of individual genes. **d,** Same models as in (**c**), but using Uno's dynamic AUC as a measure of model performance. VAF, variant allele fraction; mt, mutation; No. mt, number of mutations; AUC, area under the curve.

2.5 Clinical factors associated with AML risk

Although genetic features alone are capable of identifying many individuals at risk of developing AML in these experimental cohorts, AML incidence rates in the general population are low (4:100,000) (Deschler and Lubbert, 2006), and thus millions of individuals would need to be screened to identify the few pre-AML cases, with many false positives. To determine whether routinely available clinical information could improve prediction accuracy or identify a high-risk population for targeted genetic screening, I initially reviewed full blood count and biochemistry data that were available for 37 of the pre-AML cases and 262 controls. These data also permitted a screen for any potentially undiagnosed cases of MDS, a known risk factor for (secondary) AML (Arber et al., 2016). The diagnosis of MDS based on the WHO criteria relies not only on the presence of dysplasia in at least one lineage, but also on the presence of at least one significant cytopenia (haemoglobin (Hb) <10g/dL; platelet count<100 x10^9/L and absolute neutrophil count<1.8 x 10^9/L)(Arber et al., 2016). The latest WHO criteria state verbatim that "Cytopenia is a 'sine qua non' for any MDS diagnosis…", hence enabling exclusion of MDS based on normal blood counts alone (Arber et al., 2016). Out of the 37 pre-AMLs only one had Hb<10g/dL at recruitment (PD30116, Hb 9.8g/dL); however, three years later Hb had normalised to 13.7g/dL, thus excluding MDS. The only other cytopenia in a pre-AML was a sample with platelets of 91 x 10^9/L at baseline (PD30010); however, 3.7 years later the platelet count had risen above the WHO guideline threshold (106 x 10^9/L), suggesting that MDS was not the diagnosis. CH-PD was also overwhelmingly associated with normal blood counts in the controls, even in individuals harbouring multiple mutations at high VAF (e.g., PD35659c, PD35733b and PD35788b with leukaemia-free follow-up of 20.3, 20.4 and 17 years, respectively). The presence of normal blood counts in association with large clones corroborates the findings of previous studies of CH in the general population (Buscarlet et al., 2017; Jaiswal et al., 2014; McKerrell et al., 2015). Overall, full blood count data between controls and pre-AMLs did not differ, with the notable exception of red cell distribution width (RDW) (Figure 3.8a,b) Despite the limited sample size, there was a significant association between higher RDW and risk of progression to AML (P = 0.0016, Wald test with Bonferroni multiple-testing correction). Although traditionally used in the evaluation of anaemias, raised RDW has been correlated with inflammation, ineffective erythropoiesis, CVD and adverse outcomes in several inflammatory and malignant conditions (Hu et al., 2017). The correlation

between RDW and risk of AML development remained highly significant when only controls with CH-PD were compared to pre-AMLs ($P = 3.5 \times 10^{-6}$, Wald test with Bonferroni multiple testing correction). Higher RDW has previously been associated with CH and overall mortality (Jaiswal et al., 2014; Salvagno et al., 2015), but has never been shown to distinguish CH from pre-leukaemia.

Figure 3.8

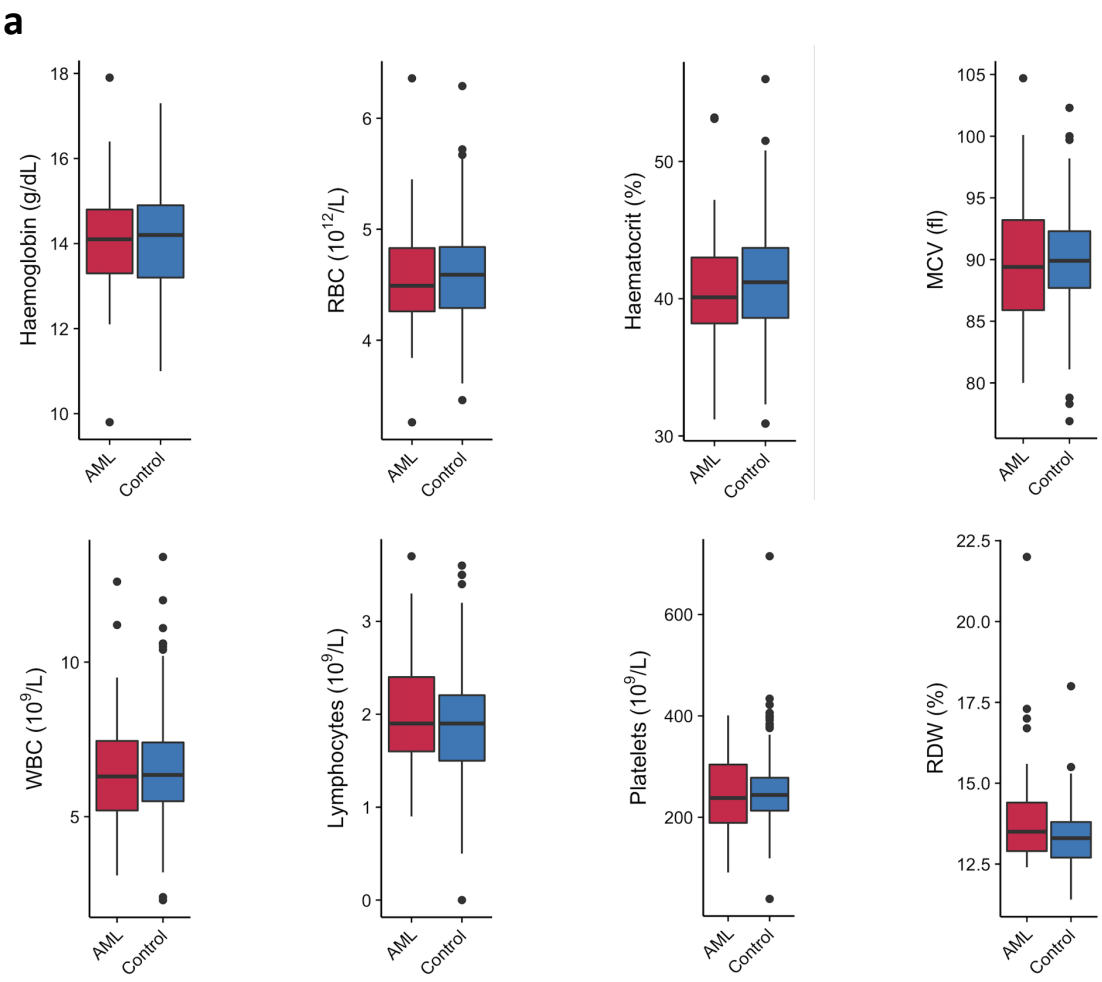

Figure 3.8 | Full blood count indices in pre-AMLs and controls. a, Box plots of full blood count parameters. Box plot centres, hinges and whiskers represent the median, first and third quartiles and 1.5× interquartile range, respectively. **b**, Kaplan–Meier curves of AML-free survival, defined as the time between sample collection and AML diagnosis, death or last follow-up. Survival curve is stratified according to RDW measurement data for n = 299 unique individuals for whom full blood count measurements were available. Among the blood indices shown, only RDW was significantly different between pre-AML cases and controls (P = 0.0016, Wald test with Bonferroni multiple-testing correction).

In order to verify RDW as a predictive factor and determine whether additional clinical parameters are associated with risk of AML development, we collaborated with Dr Netta Mendelson Cohen, Dr Elisabeth Niemeyer and Dr Noam Barda, who analysed the Clalit electronic health record (EHR) database (Balicer and Afek, 2017). This resource contains EHRs for an average of 3.45 million individuals per year collected over a 15-year period. Stringent criteria based on diagnostic codes and treatment records identified 875 AML cases (Appendix 11). Consistent with case ascertainment strategy for the genetic model, all cases of secondary AML following another myeloid malignancy were excluded. Analysis of RDW trends revealed significantly raised measurements several years before AML diagnosis relative to age and sex-matched controls (Figure 3.9a). The most pronounced increase in RDW was observed at 6-12 months before diagnosis, with ~10% of pre-AMLs having RDW values which were greater than the 99th centile of the controls. Many other blood indices, including several full blood count (FBC) parameters, changed six months to a year before diagnosis. Additional parameters that correlated with risk of AML development included reductions in monocyte, platelet, red blood cell and white blood cell counts (Figure 3.9a). However, in the majority of cases measurements did not fall outside the normal reference ranges. Nevertheless, these values were statistically distinct from those seen in large numbers of age and sex-matched controls. This is important, as it shows that these individuals did not have undiagnosed MDS/MPN, and suggests instead that evolving *de novo* AML may sometimes have a considerable prodrome with subtle but discernible clinical manifestations, potentially reflecting large pre-leukaemic clones.

Our collaborators next applied a machine-learning approach to construct an AML prediction model based entirely on variables that are routinely documented in electronic health records (Appendix 11). This model predicted AML 6–12 months before diagnosis with a sensitivity of 25.7% and overall specificity of 98.2%. The model performed consistently across different age groups with an increased relative risk of 28 for males and 24 for females between the age of 60 and 70 years (Figure 3.9b). To our knowledge this represents the first analysis of its kind in AML prediction from routinely collected clinical records. In order to better understand which patients are most likely to be accurately classified by this model, our collaborators compared absolute laboratory values for true positives and false negatives. This revealed that 35.5% of false-negative predictions were for patients for whom infrequent blood count data were available. Some of the true-positive cases had mildly abnormal blood

counts that would not initiate a diagnostic work-up (Figure 3.9c), whilst cytopenias that would be compatible with undiagnosed myelodysplastic syndrome (Arber et al., 2016) were uncommon. Other non-haematological variables associated with progression to AML included higher triglyceride levels and lower high- and low-density lipoprotein levels (Figure 3.9d).

Figure 3.9

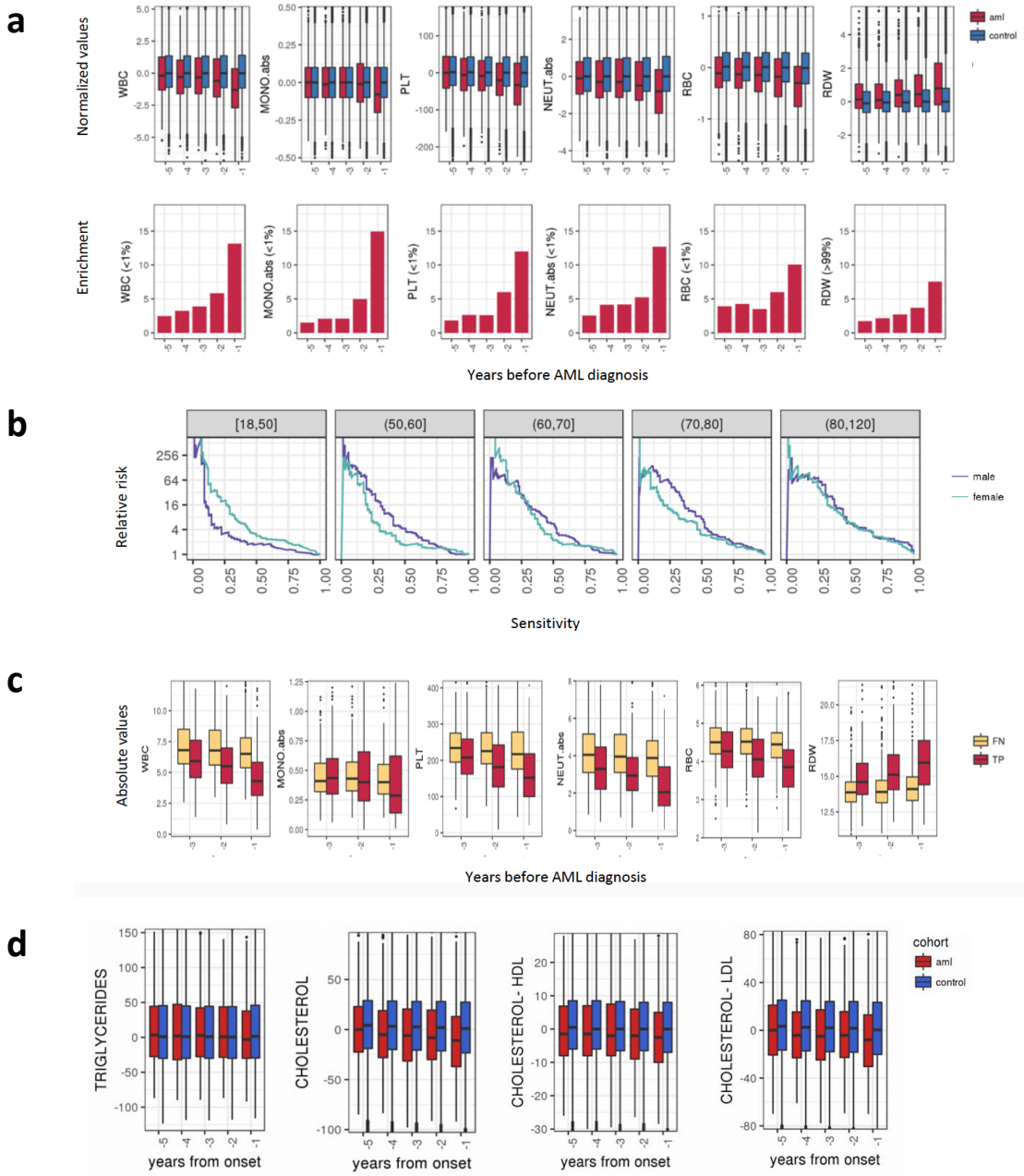

Figure 3.9 | Increased risk of AML development inferred from electronic health records. **a**, Box plots of normalized laboratory measurements. Increased RDW, reduction in monocyte, platelet, red blood cell (RBC) and white blood cell (WBC) counts (top) show a high association (bottom) with a higher risk of AML development and differed at least a year before AML diagnosis. **b**, Model performance stratification by age and gender. Age ranges are indicated above each graph. **c**, Absolute laboratory values for true positive (TP) and false negative (FN) predictions. **d,** Box plots of lipid levels. Box plots indicate median, first and third quartiles and 1.5× interquartile range. WBC, white blood cell count; MONO.abs, absolute monocyte count; PLT, platelet; NEUT, neutrophil; RBC, red blood cell; RDW, red cell distribution width; FN, false positive; TP, true positive; AML, acute myeloid leukaemia; HDL, high-density lipoprotein; LDL, low-density lipoprotein.

3. Discussion

This study sought to explore the natural history and genetic landscape of nascent AML and the extent to which the latter is distinct from CH in the general population. Collectively, these findings provide new insights into the pre-clinical evolution of AML and the feasibility of identifying CH at high risk of malignant transformation.

3.1 A long latency period is the rule rather than the exception in AML

This work demonstrates for the first time that pre-leukaemic clones can be detected in the majority of individuals who develop AML 6 or more years before clinical disease manifestations, even when interrogating for point mutations alone. This long latency has now also been reported by Desai et al, who performed a very similar nested case-control study (Desai et al., 2018). Desai and colleagues sequenced 67 AML-associated genes in peripheral blood samples from 212 women diagnosed with AML a median of 9.6 years later alongside the same number of controls (Desai et al., 2018). Consistent with our results, pre-leukaemic clones (VAF>1%) were present in 68.6% and 30.9% of pre-AML cases and controls, respectively (Desai et al., 2018). This long pre-clinical evolution highlights important aspects of AML biology and reveals that the window for potential intervention is measured in years for the majority of individuals who develop AML.

3.2 The distinct driver landscape of pre-AML

This work also reveals that the mutational landscape, and not simply the mutation burden, differs between CH in controls versus pre-AML. The differences in the mutational spectrum observed between pre-AML cases and controls may arise through cell-intrinsic or -extrinsic factors. As discussed in Chapter 1, previous studies of clonal haematopoiesis have demonstrated that clones with particular mutations dominate in the context of specific environmental pressures (Gibson et al., 2017; Hsu et al., 2018; McKerrell et al., 2015; Takahashi et al., 2017; Wong et al., 2015b), suggesting an important role for cell-extrinsic factors in haematopoietic somatic evolution. Although such factors in CH remain poorly understood, it is intriguing that mutations in splicing factor genes and *TP53* were significantly enriched among the pre-AMLs relative to the controls, with the former presenting in

significantly younger individuals than in benign CH. Spliceosome mutations appear to confer a competitive advantage in the context of ageing, and were almost exclusively observed in the general population in individuals over age 70 years (McKerrell et al., 2015). Similarly, clones harbouring *TP53* mutations expand dramatically with exposure to intensive chemo- and/or radiotherapy (Bondar and Medzhitov, 2010; Wong et al., 2015b). However, *TP53*-mutated HSC clones are very common at extremely low VAF in the elderly, but tend to remain stable in size over time, suggesting only a modest selective advantage in the absence of increased genotoxic stress (Wong et al., 2015b). Therefore, it is possible that the significantly higher prevalence of clones with *TP53* and spliceosome gene mutations in pre-AML cases may reflect distinct microenvironmental selection pressures rather than earlier mutation acquisition.

3.3 The significance of the higher mutation burden in pre-AML

The observation of the higher burden of putatively oncogenic mutations (driver mutations) in the pre-AML cases across all age groups raised two main related questions. Firstly, what is the mechanism underpinning the discrepancy in mutation burden between controls and pre-AMLs? Secondly, do driver mutations detected in pre-AML cases reflect the presence of an AML ancestor, or do these mutations behave as surrogate markers of factors predisposing to leukaemogenesis?

Although speculative, several mechanisms may account for the higher mutation burden and clone size observed in the pre-AMLs. It could reflect a higher mutation rate in the pre-AML cases, for example due to higher HSC turnover, potentially secondary to depletion of the functional HSC pool. Alternatively, chance may play a dominant role, with stochastic driver mutation acquisition triggering clonal expansion, thus increasing the odds of further driver events on a pre-malignant background leading to selection for progressively more mutated clones. However, this multistage cancer evolution paradigm does not account for the relationship between the fitness advantage conferred by a driver mutation and the environmental context of the mutated cell (Rozhok et al., 2014). Clones with drivers could be under stronger selective pressure in certain bone marrow environments, as is seen in particular clinical contexts such as aplastic anaemia or after intensive cytotoxic therapy (Hsu et al., 2018; Wong et al., 2015b; Yoshizato et al., 2015). As discussed in the introduction, the presence of selective pressure favouring clonal expansions, rather than mutation acquisition,

may thus be an important determinant of the number of mutations detectable by bulk sample sequencing.

Our time series experiment and sequencing of diagnostic specimens helped partially address the second question, demonstrating that clones in pre-AML cases represent a combination of leukaemia ancestors and 'bystander' clones that likely are not related to the future AML. However, our experiment using bulk cell populations was too small and hindered by confounding factors to enable strong conclusions about clonal growth kinetics or mutation rates. We hope that future experiments using single cell and/or highly purified cell population studies on viable cells at serial time points will shed light on these questions.

3.4 Rationale for AML risk prediction and future directions

Cancer predictive models have enabled successful early detection and intervention programmes for several solid tumours (Vickers, 2011; Wang et al., 2014). However, screening tests are unavailable for the sub-clinical stages of most haematological malignancies. Given that the main cause of mortality in AML is treatment resistance/relapse (Döhner et al., 2015), there is a rationale for identifying and treating a genomically simpler antecedent of the disease. In this context, reduction of clonal size rather than complete clonal extinction may be sufficient to significantly reduce the risk or slow AML progression. Such an approach has proven very effective in CML, which has been transformed by targeted therapy into a chronic condition with a dramatically reduced incidence of progression to CML blast crisis (Kalmanti et al., 2015). Furthermore, CH is associated with and may play a causal role in common non-malignant conditions (Fuster et al., 2017; Jaiswal et al., 2017), which may strengthen the case for screening and intervention.

3.4.1 Further development of genetic AML prediction methods

This study provides proof-of-concept for the feasibility of early detection of healthy individuals at high risk of developing AML. The models presented here demonstrate that somatic genetic features are predictive of AML progression and that the presence of mutations in certain genes confers a greater risk. Desai et al have since identified similar gene-level risk factors (Desai et al., 2018). Consistent with our results, *TP53* mutations conferred the highest odds ratio of progression from CH to AML, followed by drivers in *IDH1/2* and

spliceosome genes (Desai et al., 2018). Although Kaplan-Meier analysis (Figure 3.6) is consistent with a trend towards shorter AML-free survival with *IDH1/2* mutations, we chose not to group functionally-related genes in our analysis in order to reach significance, as their mechanistic consequences may differ (e.g., *IDH2* p.R140 and *IDH2* p.R172 (Papaemmanuil et al., 2016)). In addition to improving model performance, the identification of highly significant disparities in gene-level HR offers compelling biological insights into the determinants of clonal progression, which warrant further investigation.

Given that most of the genetic model's predictive power stems from mutations with VAFs >0.005, our data suggests that conventional deep targeted sequencing, as used for the validation cohort, is adequate for future screens when combined with stringent variant calling and driver mutation curation. Thus, the additional cost of error correcting sequencing is unlikely to be justified. However, it is possible that future studies may show that specific mutations may have predictive value when detected accurately even at low VAF (e.g. *U2AF1* hotspot variants).

As recurrent chromosomal translocations are likely to be initiating events in approximately 20% of AML (Papaemmanuil et al., 2016), incorporating these into the genetic model is likely to further increase predictive accuracy. McKerrell et al. have shown that it is feasible to simultaneously capture several recurrent translocations/inversions with targeted panels only slightly larger than the ones used in the current study (McKerrell et al., 2016). Additionally, expanding this dataset will make it possible to investigate whether co-mutation patterns carry prognostic significance, as is the case in AML (Gerstung et al., 2017; Papaemmanuil et al., 2016).

3.4.2 Combining clinical and genetic information to risk-stratify clonal haematopoiesis

The predictive model based on mutations and demographic features partially overcomes the limitations imposed by the low overall incidence of AML, but does not eliminate them. We have shown that commonly recorded clinical parameters, notably RDW and other FBC indices, may identify a smaller population with higher pre-test AML risk for screening. Although clinical parameters were predictive relatively close to the time of AML diagnosis, pre-AML clones can be of significant size many years before diagnosis and it is entirely plausible that surrogate laboratory markers of their presence may be identifiable

much earlier, as we found for RDW in the validation cohort. Analysis of the 37 individuals for whom both genomic and clinical information were available found that 6% of the relative risk contribution was attributable to clinical variables, suggesting that combining routinely available clinical data with genomic variables may strengthen AML prediction models. Extending this analysis in a large EHR database further revealed that pre-AML has additional subtle clinical manifestations which in themselves had considerable predictive power 6-12 months prior AML diagnosis. This further supports a role for clinical variables in strengthening genomic prediction models and/or in targeting the population most likely to benefit from screening for CH.

Defining the population most likely to benefit from genetic screening will also depend on improved understanding of the role of CH in common non-malignant conditions. If, as several recent studies strongly suggest, some pre-leukaemic clones are pro-inflammatory and actively promote atherosclerosis and cerebro/cardiovascular adverse events (Fuster et al., 2017; Jaiswal et al., 2017), then a significantly larger proportion of the population might benefit from screening for CH and could thus be considered for possible interventions to suppress pre-leukaemic clones and/or mitigate established cardiovascular risk factors (blood pressure, dyslipidaemia, etc). Our analysis of a large EHR database reveals that subtle clinical manifestations, including trends in triglycerides and RDW that are established risk factors for cardio/cerebrovascular disease also correlated with risk of AML. It is conceivable that there are unifying characteristics of high-risk CH emblematic of the emerging links between ageing and dysregulated inflammation or immune senescence (Green et al., 2011; Shaw et al., 2013).

Clearly these findings cannot address the challenging question of how genomic screening methods should be implemented in a real-world setting, and a combined clinical and genetic screening approach requires validation in large prospective cohort studies. Promisingly, the infrastructure for performing such studies is increasingly available, for example the UK Biobank (Bycroft et al., 2018). These resources should help stimulate large prospective studies that take account of all health outcomes associated with CH.