Leukemia (2019) 33:826–843
https://doi.org/10.1038/s41375-019-0387-y

REVIEW ARTICLE

Acute myeloid leukemia

Data mining for mutation-specific targets in acute myeloid leukemia

Brooks Benard[1] · Andrew J. Gentles[2] · Thomas Köhnke[1] · Ravindra Majeti[1] · Daniel Thomas[1]

Received: 29 August 2018 / Revised: 6 October 2018 / Accepted: 24 October 2018 / Published online: 6 February 2019

Abstract

Three mutation-specific targeted therapies have recently been approved by the FDA for the treatment of acute myeloid leukemia (AML): midostaurin for *FLT3* mutations, enasidenib for relapsed or refractory cases with *IDH2* mutations, and ivosidenib for cases with an *IDH1* mutation. Together, these agents offer a mutation-directed treatment approach for up to 45% of de novo adult AML cases, a welcome deluge after a prolonged drought. At the same time, a number of computational tools have recently been developed that promise to further accelerate progress in mutation-specific therapy for AML and other cancers. Technical advances together with comprehensively annotated AML tissue banks have resulted in the availability of large and complex data sets for exploration by the end-user, including (i) microarray gene expression, (ii) exome sequencing, (iii) deep sequencing data of sub-clone heterogeneity, (iv) RNA sequencing of gene expression (bulk and single cell), (v) DNA methylation and chromatin, (vi) and germline quantitative trait loci. Yet few clinicians or experimental hematologists have the time or the training to access or analyze these repositories. This review summarizes the data sets and bioinformatic tools currently available to further the discovery of mutation-specific targets with an emphasis on web-based applications that are open, accessible, user-friendly, and do not require coding experience to navigate. We show examples of how available data can be mined to identify potential targets using synthetic lethality, drug repurposing, epigenetic sub-grouping, and proteomic networks while also highlighting strengths and limitations and the need for superior models for validation.

Introduction

Acute myeloid leukemia (AML) is a blood cancer characterized by the accumulation of clonal myeloid precursor cells arrested in their ability to mature into normal blood elements accompanied by varying degrees of anemia, thrombocytopenia, and leukopenia [1]. While reductions in leukemic blasts can be achieved initially with cytarabine and anthracycline combinations in the majority of patients, long-term outcomes have yet to improve significantly, with 5-year survival rates for elderly patients (>60 years) ranging from 5–15% and median overall survival ~1 year [1]. Despite a relatively low mutational burden compared to other cancers [2], the management of AML is complicated by its molecular and biological heterogeneity: one targeted therapy is unlikely to be effective in all patients. Since the groundbreaking success of all-trans retinoic acid combined with arsenic trioxide in acute promyelocytic leukemia (APML) with *PML/RARA* fusion, until only recently, few other targeted approaches have demonstrated clinical responses in non-APML AML. Most agree that modest improvements in outcome observed in the last 2 decades have been primarily due to dose escalation of chemotherapy and better supportive care [1]. The recent approval of three mutation-specific targeted therapies for AML by the United States Food and Drug Administration (FDA) (midostaurin for *FLT3* mutations, enasidenib for relapsed or refractory cases with *IDH2* mutations, and ivosidenib for cases with an *IDH1* mutation) has revitalized interest in mutation-directed approaches. However, assigning a targeted therapy to a given patient's molecular profile is not trivial and requires ongoing, carefully designed pre-clinical and clinical studies. To aid with this, an ever-increasing body of

✉ Ravindra Majeti
rmajeti@stanford.edu

✉ Daniel Thomas
dthomas3@stanford.edu

[1] Department of Medicine, Division of Hematology, Cancer Institute, and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA

[2] Departments of Medicine (Biomedical Informatics), and Biomedical Data Sciences, Stanford University School of Medicine, Stanford, CA, USA

patient data is available to interrogate AML heterogeneity and fast-track newer agents for successful clinical development. This review summarizes the different data types, repositories, and recently developed computational-based methods to assist in analyzing big data in AML, with a major focus on finding novel mutation-specific patterns and potential targets.

Cytogenetic, epigenetic, and mutational heterogeneity of AML

Clinically, AML with recurrent genetic abnormalities is partitioned into seven major cytogenetic sub-groups based largely on chromosomal translocations according to the most recent classification of the World Health Organization (WHO) for blood diseases [3]. Pre-treatment cytogenetic findings (confirmed by fluorescence in situ hybridization (FISH) of breakpoint specific probes) are generally considered the most important independent prognostic factors in AML. However, screening for recurrent somatic mutations has also become routine practice at the time of diagnosis. Two particular mutations, *NPM1* and biallelic *CEBPA*, are now included in the classification of AML, with recurrent mutations in *RUNX1* and BCR-ABL1 as a provisional entity, although mutation-specific therapies are not yet available for these particular sub-groups. In the context of genetic variants, several studies have uncovered a high degree of intra- and interpatient genetic heterogeneity in AML. In addition to the gene involved, the type of mutation (location in the gene, frame-shift vs. missense, charge reversal vs. modest substitution), and dominant vs. sub-clonal mutation burden (termed the variant allele frequency or VAF) are important considerations for mutation-directed therapy. For example, biallelic *CEBPA* mutations, most commonly involving one amino-terminal and one carboxy-terminal mutation, are associated with favorable outcomes, but monoallelic *CEBPA* mutations are not as favorable [4].

In addition to genetic heterogeneity, AML also shows heterogeneity at the levels of gene expression and epigenetics, which can be exploited therapeutically. RNA sequencing analysis of 179 clinically annotated adult cases of de novo AML from The Cancer Genome Atlas (TCGA) [2] revealed seven major mRNA gene expression groups, which were highly concordant with previously generated microarray data [5] and closely associated with cell morphology. For example, RNA sequencing group 4 was associated with subtype M1 in the French–American–British morphological classification. A number of novel gene fusions were also identified by RNA sequencing, highlighting the potential to uncover further genetic/cytogenetic heterogeneity. Finally, DNA methylation profiling identified 16 distinct groups [6], including four epigenetically distinct groups of AML with *NPM1* mutations and specific methylation profiles for t(8;21), inv16, and t(15;17) leukemia. As described below, analyses of DNA methylation and other epigenetic marks can correlate with RNA expression profiles and might therefore also assist in the identification of epigenetic-directed therapy and present additional targets for mutation-specific therapy.

Additionally, whole-exome or custom-capture sequencing studies for individual AML patients is increasing. The clinical interpretation of uncommon sequence variants and attributing disease causality from this data is still in its infancy with a few consensus guidelines available [7]. Notably, variant-calling pipelines can show poor concordance, especially for indel mutations [8], and the choice of reference genome can influence interpretation. Prediction tools such as SIFT, PolyPhen, SNAP, SNPs&Go, PhyloP, and Mutation Taster can assist in predicting pathogenicity based on protein sequence, inter-species conservation, splicing, and protein structure, but these are based on a priori assumptions and perform poorly with non-coding variants.

Data types for data mining primary AML samples

Historically, data mining in cancer was performed on data types such as cytogenetics, microarray profiling, immunophenotype, and patient survival. However, advances in diverse molecular methods have dramatically expanded the types of data available. Notably, the rapid development of multiple next-generation sequencing (NGS) techniques has produced vast amounts of genome, exome, and RNA sequencing data, which can be used to measure features such as cancer ploidy, structural variants, translocations, focal and regional copy number variations (CNVs), single nucleotide variants, DNA methylation, and gene expression. Additionally, mass spectrometry has been applied to profile the proteomic landscape of cells and tissues. Using patient cancer-derived cell lines, several small molecule and shRNA screens have been performed generating drug sensitivity and synthetic lethality (SL) correlations with molecular profiling [9–14]. Patient demographics (i.e., age, white blood cell count, ethnicity, or gender) are also routinely collected during clinical trials and can be correlated with molecular tumor profiles [15]. A summary of notable databases and the types of data they contain, relevant to AML, is outlined in Table 1. With the advent of this increased diversity and scale of molecular data, there has been a growing appreciation for the applications of machine learning, statistical methodologies, and algorithm development to mine this data for new biological insights (Table 2).

Table 1 Publicly available databases and web portals with current URL links for data mining in AML

Database	AML	Primary tumor derived	Cell line derived	Mutation	Expression	CNV	Methylation	shRNA	Drug efficacy	Drug-target	Survival	URL
Beat AML	√	√		√	√	√			√		√	http://www.vizome.org/aml/
TCGA	√	√		√	√	√	√				√	https://cancergenome.nih.gov
TARGET-AML	√	√		√	√	√	√				√	https://ocg.cancer.gov/programs/target/data-matrix/
ICGC	√	√		√	√	√	√				√	https://icgc.org
Leucegene	√	√		√	√						√	https://leucegene.ca
AML-Multistage	√	√		√		√					√	https://cancer.sanger.ac.uk/aml-multistage/
Gene Expression Commons	√	√			√							https://gexc.riken.jp/
cBioPortal	√	√	√	√	√	√					√	http://www.cbioportal.org/index.do
COSMIC	√	√	√	√					√			https://cancer.sanger.ac.uk/cosmic
Leukemia Gene Atlas	√	√	√	√	√		√				√	http://www.leukemia-gene-atlas.org
BloodSpot	√	√	√	√	√	√						http://servers.binf.ku.dk/bloodspot/
ArrayExpress	√	√	√	√	√		√					https://www.ebi.ac.uk/arrayexpress/
SynLethDB	√	√	√	√	√	√		√	√	√		http://histone.sce.ntu.edu.sg/SynLethDB/index.php
Expression Atlas	√	√	√	√	√				√			https://www.ebi.ac.uk/gxa/about.html
CCLE	√		√	√	√				√			https://portals.broadinstitute.org/ccle
GEO	√		√		√							https://www.ncbi.nlm.nih.gov/geo/
Project Achilles			√	√	√	√		√	√			https://portals.broadinstitute.org/achilles
LINCS			√					√	√			http://lincs.hms.harvard.edu/db/
Genomics of Drug Sensitivity in Cancer			√	√		√			√			https://www.cancerrxgene.org
Cancer Therapeutics Response Portal			√	√	√	√				√		https://portals.broadinstitute.org/ctrp/
ChEMBL			√						√	√		https://www.ebi.ac.uk/chembl/
Comparative Toxicogenomic Database (CTD)			√							√		http://ctdbase.org
TARGET	√	√		√	√	√	√			√	√	https://software.broadinstitute.org/cancer/cga/target

Table 2 Computational tools for drug repurposing and synthetic lethal discovery applicable to AML

Method	Input data	Advantages/strengths	Disadvantages/limitations	Used in AML	Reference
DAISY	Mutation, Copy Number, Expression, shRNA	• Identifies synthetic "dosage" lethals • Broad in application with diverse inputs • Validated in prognostic prediction	• Does not input all somatic mutations • Partially dependent on experimental shRNA cell line data • Correlated gene expression is a major assumption of synthetic lethality • Does not incorporate DNA methylation data	√	[16]
DeepWalk	DrugBank, Diseasome	• Outperforms similar methods • Publicly available code	• May not predict novel drugs or targets that are not in a known network • Does not include mutation or expression data		[99]
IMPACT	Whole Exome Sequencing	• Integrates mutation and copy number from whole-exome sequencing (WES) data	• Does not include expression data for functional validation of loss of function (LOF) • Drug database pulled from available data in 2016 and has not been updated		[41]
ksREPO	Methylation, Comparative Toxicogenomic Database	• Only known method to predict drug repurposing in AML based on DNA methylation data	• Lack of data directionality and significant compression may result in dropout of potential candidates • Requires gene-specific methylation annotations	√	[36]
MERGE	Mutation, Copy Number, Expression, Methylation	• Based on primary patient tumor data • Integrates five different data types	• Does not account for cytogenetic risk category or genotype of cases	√	[33]
MiSL	Mutation, Copy Number, Expression	• Based on primary patient tumor data • Uses all types of mutations to predict LOF	• Does not include a pipeline for streamlining drug repurposing • Does not take into account variant allele frequency • Does not identify synthetic dosage lethal interactions • Druggable targets that do not undergo CNV in multiple cancers are missed	√	[17]
MiSTIC	Mutation, Expression	• User-friendly web interface • Interactive plots • Best for gene correlation analysis	• Does not utilize the survival data from the TCGA patients in the correlated gene expression • Not directly linked to synthetic lethality	√	[25]
MutExSL	Copy Number, Expression, siRNA	• Synthetic lethals validated in vitro	• Limited cancer types analyzed and examples • Synthetic lethal interactions based only on a handful of DNA damage response genes		[55]
PDOD	KEGG, CTD, GEO, DrugBank	• Web interface for hypothesis generation • Uses network directionality	• Did not include binding/association, phosphorylation, or DNA methylation data in the modeling (limited to four effect types) • Only analyzes the shortest KEGG paths to find perturbations and drug targets	√	[100]

Analysis of these large data sets is already generating new general and mutation-specific targets and insights for drug repurposing to optimize therapeutic intervention [9, 16–18].

Notable repositories and searchable databases for AML

As data generated by modern biomedical research has exponentially increased, individual groups and consortiums have begun to develop static data repositories, searchable databases, and even data portals with user-friendly analytical and visualization tools (Table 1). The first major online centralized database "Netscape" was developed in 2001 and is no longer active. Netscape predominantly collated clinical data, cytogenetic information, molecular mutation, and microarray gene expression [19]. Other public databases containing gene expression or copy number data from microarray studies include ArrayExpress and Gene Expression Omnibus (GEO) [20, 21]. With the advent of NGS and high-throughput methods, public repositories are also now populated with data types such as whole-genome/exome sequencing, RNA sequencing (RNA-Seq), DNA methylation profiling, proteomic profiles, chromatin immunoprecipitation (ChIP) sequencing, shRNA-mediated gene knockdown results, and small molecule sensitivity screens (data types including primary AML sample data for each repository summarized in Table 1). Additionally, several clinical trials have provided patient demographics, response, and survival data types to online repositories [15, 22]. The most recent public data banks for AML include the Leukemia Gene Atlas (LGA) [23], TCGA [15], the TARGET (Therapeutically Applicable Research to Generate Effective Treatments) database [24], the Leucegene dataset [25], and the BloodSpot database [26], all of which have web interfaces for data download, visualization, and analysis. Multiple groups have used these public databases to report novel drug–gene and potential synthetic lethal relationships in AML.

Leukemia Gene Atlas (LGA)

The LGA is a web interface that contains results from 25 leukemia studies, the majority of which are AML. While comprised primarily of microarray gene expression profiles, it does include ChIP sequencing, DNA methylation, single nucleotide variant, and patient survival data. At the time of inception (2011), the LGA was the only central repository for accessing and analyzing datasets specific to AML and incorporates several data analysis and visualization tools. A user can perform differential gene expression *t*-tests, survival analysis, and hierarchical clustering, and can visualize data in the form of bar graphs, data distributions, expression heatmaps, principal component analysis, and Kaplan–Meier survival curves. As a proof-of-principle, Hebestreit et al. used LGA to analyze the role of differential RUNX1 DNA binding in leukemias [22]. Using previously published chromatin sequencing results from Tijssen et al., they computed the 33 most differentially bound sites between non-progenitor and progenitor cells from Novershtern et al., visualized the distribution of RUNX1 expression across different leukemias, performed hierarchical clustering of gene expression between the most differentially bound RUNX1 binding sites, and performed survival analysis in AML patients factored by RUNX1 expression, thus demonstrating the versatility of the platform. Other groups have utilized the LGA to show RUNX1 cooperates with FLT3-ITD to drive leukemia [27]. Lack of ongoing updates and RNA-seq expression data is a limitation.

The Cancer Genome Atlas (TCGA) and TARGET

The first major multi-omic dataset in adult AML is the TCGA effort, first established 2013 [2]. This dataset includes 200 AML patients with clinical and biospecimen annotations: 150 have whole exome sequencing data, 50 have whole genomes, 179 have expression data, and 192 have DNA methylation profiling. Additionally, most patients have germline sequencing performed. Although it is not the largest AML cohort, the TCGA-AML represents the largest adult AML dataset with the most diverse molecular assays performed on the same patients. Similarly, the TARGET AML study provides a comparable level of diverse molecular assays and clinical annotation for 993 pediatric cases of AML [24]. These studies facilitate the integration of multiple molecular features for the dissection of disease-causing biology and clinical correlations. The TCGA web portal provides some simple analysis tools for data visualization (e.g., Kaplan–Meier plots, OncoPrints, frequency distributions). However, it does not contain user-friendly tools for dissecting and representing integrated analysis of data types. The TARGET data portal does not include any analysis or data visualization tools, but simple analysis (e.g., pathway enrichment and survival analysis) can be performed through the International Cancer Genomics Consortium's (ICGC) [28] data portal (https://dcc.icgc.org/analysis). Fortunately, platforms such as cBioPortal have been developed to provide a wide array of intuitive, aesthetic, and informative analysis and visualizations of TCGA and TARGET data [29].

Leucegene and the MiSTIC gene correlation tool

Unlike repositories such as the LGA and TCGA, the Leucegene dataset and web interface are still being expanded. The Leucegene project (https://leucegene.ca/)

aims to improve the prognostic classification of AML using RNA sequencing (RNA-Seq) of 457 AML samples from diverse cytogenetic subgroups. MiSTIC (Minimum Spanning Trees Inferred Clustering) is an open and available platform designed to assist in visualizing Leucegene RNA-Seq data or other primary cell data for rapid exploration [25]. The major innovation of MiSTIC is to present highly correlated gene sets in a data-set as spikes on a circularized dendrogram, called an icicle plot (Fig. 1e, lower panel). The length of the spike reflects the degree of similarity, transformed using a power-scale to aid in visual comparison of clusters. Interrogation of the spikes by simple clicking gives the names of the highly correlated genes within the cluster and allows rapid comparison with published gene sets. This feature enables the user to determine which clusters are related to chromosomal alterations or are involved in cell-cycle or simply represent contamination with peripheral blood T cells. Currently, Canadian patient and TCGA-AML RNA-Seq data is available with clinical characteristics including somatic mutations. Significantly, MiSTIC was able to show that certain genes such as *AURKA* and *CENPA* are enriched in bone marrow AML samples rather than samples collected from peripheral blood [25]. The marked difference in proliferative genes in these two collection types suggests blasts in peripheral blood are much less proliferative than bone marrow-derived blasts. Similarly, MiSTIC found that *CEBPA*-mutated samples can be clearly distinguished in a plot of *CD34* vs. *HOXA9* gene expression in comparison to other poorer prognosis subgroups of AML [25].

BloodSpot: comparing normal vs. disease states in hematopoiesis

Initially published as HemaExplorer in 2013 [30], the newly redesigned BloodSpot interactive web portal contains gene expression data from 23 human and murine studies, with the goal of providing an accessible platform for hypothesis generation in hematopoiesis [26]. BloodSpot builds on data from sorted normal and leukemic blood cells from HemaExplorer in addition to curating data from six independent studies on AML (>2000 samples) and data from the Differentiation Map [31] and the Immunological Genome Project [32]. To use BloodSpot, one inputs a gene name or a gene signature; outputs include Kaplan–Meier plots (generated from TCGA), jitter plots, and hierarchical expression in the hematopoietic tree. The user can rapidly interrogate gene expression in hematopoietic cells and AML subsets, and whether expression correlates with overall AML survival. Limitations include stratification of patient survival above or below the median expression and lack of patient demographics.

Beat AML: ex vivo drug screening of molecularly annotated primary AML

Recently, Tyner et al. performed whole-exome, RNA-Seq, and ex vivo drug screening (using 122 inhibitors) on 672 tumor samples from 562 patients, 275 of which were de novo cases. This study provides a rich dataset to tease out if a mutation, or a combination of mutations, show differential drug sensitivity in primary patient material (http://www.vizome.org).

Data mining for drug repurposing in AML

Machine learning integration of big data for precision medicine

As multiple-omics data types are being generated, it is important to determine which types are best predictive for identifying therapeutically actionable events in cancer. Additionally, it remains unclear if and what types of data integration will improve our understanding of these disease networks. To address this issue, Lee et al. recently developed MERGE, which integrates multi-omic data to identify statistically correlated gene markers of drug sensitivity in AML (http://merge.cs.washington.edu) [33]. Using mutations, CNVs, and DNA methylation data from the TCGA-AML dataset, gene expression from the Cancer Cell Line Encyclopedia, and regulatory annotations, MERGE learns how much each of these unique driver features successfully predict known drug sensitivity. MERGE identified an association between high *FLT3* expression and sensitivity to midostaurin, ponatinib, sunitinib, and tandutinib (Table 3); all but tandutinib are FDA approved therapies. Although midostaurin is already approved for *FLT3*-mutated AML, MERGE analysis showed that *FLT3* expression was a better predictor of drug sensitivity than *FLT3* mutation status, thus potentially expanding the therapeutic index of midostaurin in AML. Additionally, the authors show that high expression of *L2HGDH*, which was inversely correlated with the *IDH1* mutant oncometabolite R-2-hydroxyglutarate, was predictive of increased sensitivity to cyclin-dependent kinase inhibitors, thus identifying a potential vulnerability in *IDH1* mutant AML. A current limitation of MERGE is that it does not take into account the cytogenetic risk category or genotype of cases, both of which are known to be prognostic. However, compared to other approaches (elastic net and multitask learning), MERGE outperforms in its ability to consistently predict drug sensitivity based on individual gene expression and presents one of the first applications of machine learning for multi-omics-driven therapeutic prediction.

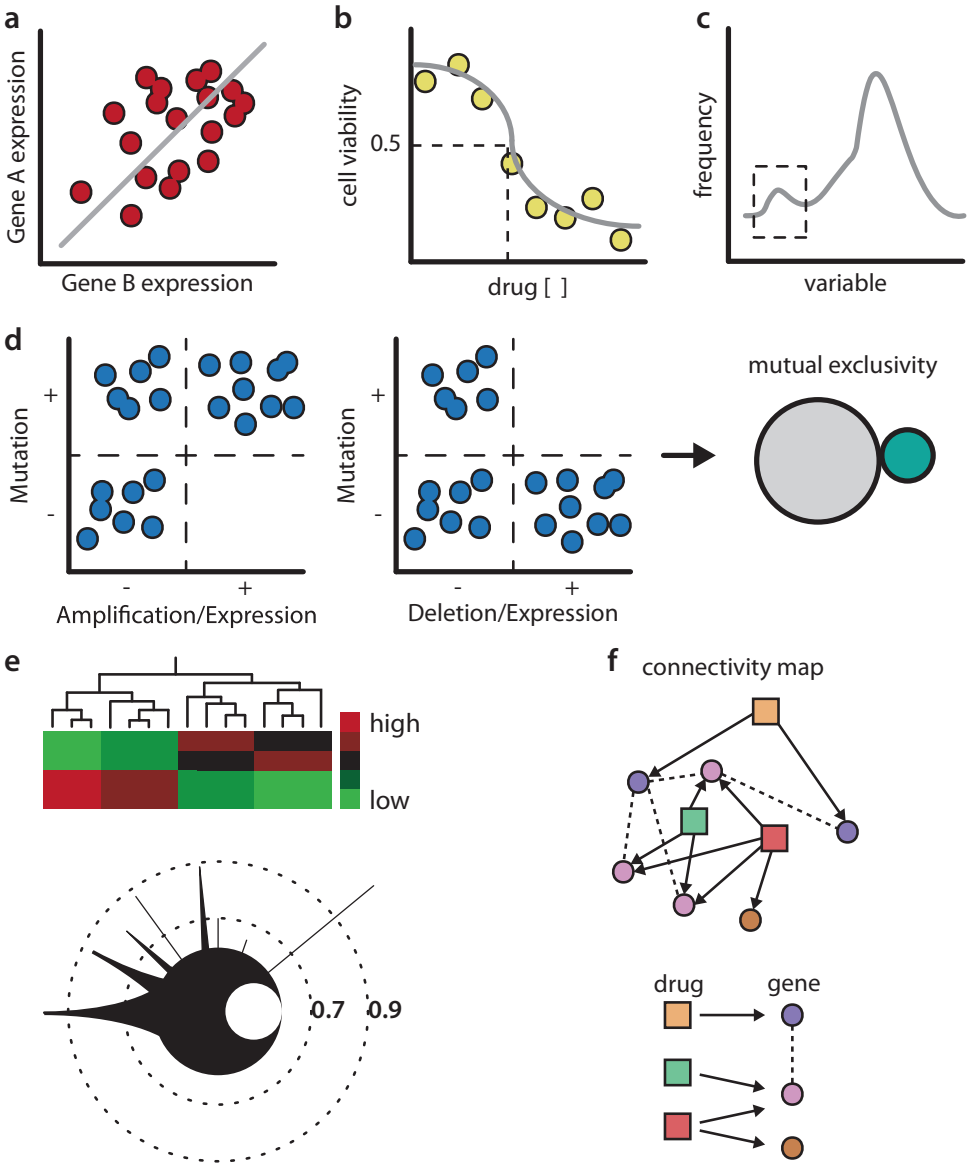


Fig. 1 Mathematical approaches for discovering novel gene–gene and gene–drug therapeutic vulnerabilities in cancer. **a** Pairwise correlation relationships and regression analysis can be used to determine if there is a significant linear relationship between two variables (e.g., expression of gene A vs. gene B may indicate both are in the same cell-type or cellular process). **b** Non-linear regression analysis of biological data, such as inhibitory concentration at 50% of maximal effect (IC_{50}) dose–response curves, can be identified from large drug/small molecule screens; note some effective hits may not show changes in IC_{50}, but may show important differences in other end-points. **c** Unusual non-normal distributions of certain variables such as bi-modality can be used to predict synthetic lethal gene pairs. **d** Boolean implications resembling IF-THEN rules can be inferred from large data sets that represent unique mutual exclusivity or subset relationships. Tools like STEPMiner can be used to binarize complex variables [35]. **e** Unsupervised hierarchical clustering shown as either a dendrogram and heat-map (upper panel) or an icicle plot (as in the MiSTIC interface, lower panel) can show closely related samples not otherwise apparent from clinical or pathological parameters. Variables can be calculated to be closely related based on different distance metrics (e.g., Euclidian, Manhattan) and clinically actionable sub-groups can be recognized by common patterns. **f** Networks of experimentally validated or bioinformatically predicted "nodes" can be used in machine learning and data mining approaches by applying decision tree learning as a predictive modeling to inferring novel interaction networks, critical nodes in a pathway, and drug repurposing

Deciphering DNA methylation patterns for drug repurposing

Adult AML, especially in elderly patients, is characterized by a high frequency of somatic mutations involving genes that regulate DNA methylation. Epigenetic patterns are attractive targets for differentiation therapy in AML. However, consistent methylation signatures that correspond with mutation, prognosis, or targeted therapy have been difficult to identify, partly because DNA methylation data often

Table 3 Mutation-specific drugs currently approved for AML and mutation-specific drugs that are predicted from data mining with their matching genetic lesion

Drug	Mutated gene	Reference
FDA approved mutation-specific drugs		
Ivosidenib	*IDH1*	PMID: 30209701
Enasidenib	*IDH2*	PMID: 28588020
Midostaurin	*FLT3*	PMID: 28546144
Arsenic trioxide All-trans retinoic acid	t(15;17) (*PML-RARA*)	PMID: 23841729
FDA approved drugs with potential mutation-specific sensitivity		
Venetoclax	*IDH1/2*	PMID: 27721426
Gemtuzamab ozogamicin	*FLT3, NPM1*	PMIDs: 24927407, 24300852, 21791474
Decitabine	*TP53, DNMT3A*[a]	PMIDs: 27959731, 22124213
Repurposed drugs with potential mutation-specific sensitivity		
Lenalidomide (approved for myeloma)	Trisomy 13	PMID: 18824593
Ruxolitinib (approved for myelofibrosis)	*JAK2*	PMID: 22422826
Dasatinib (approved for Ph^+ ALL)	*KIT*	PMID: 27566651
Non-approved drugs with potential mutation-specific sensitivity		
Sorafenib, Crenolanib, Quizartinib, Gilteritinib, Ponatinib, Sunitinib, Tandutinib	*FLT3*	PMIDs: 29298978, 26776182, 18230792, 24227820, 29859851
EZH2 inhibitors (e.g., GSK-126), Alitretinoin, Panobinostat, and Progesterone	*WT1*	PMIDs: 25398938, 26860211
Avapritinib	*KIT D816V*	PMID: 29233825
H3B-8800	*SRSF2, U2AF1, ZRSR2, SF3B1*	PMID: 28030373, NCT02841540

[a]Findings from one retrospective study only

follows a bimodal distribution (Fig. 1c) and is prone to batch effects. Using a computational mining approach for mutation-specific Boolean implications of DNA methylation, Sinha et al. found that *WT1* mutations were strongly linked to DNA hypermethylation (Table 3) and to known drivers such as *IDH2* [34]. Boolean implications are IF-THEN rules inferred from large data by looking for unexpected sparse quadrants when data is binarized using tools such as StepMiner [35] (Fig. 1d): for example when mutation X is present, cytosine site Y is always methylated. The pattern of methylation was consistently at the promoters of polycomb target genes, and EZH2 inhibitors exhibited differentiation-promoting activity in this subgroup of AML in vitro.

Epigenome-based drug repositioning

Another approach to identify the therapeutic target is to infer gene expression from epigenetic signatures and then match these signatures to a drug–gene database. Brown et al. developed ksRepo to leverage DNA methylation data for the prediction of drug repurposing in AML [36]. Using differential CpG methylation data from two AML GEO [21] studies, gene set enrichment results were fed into ksRepo and matched to the Comparative Toxicogenomics Database dataset [37]. ksRepo identified alitretinoin, panobinostat, and progesterone as novel therapeutics in *WT1* mutated AML, but further validation work is required (Table 3).

Prediction of drug repurposing based on "effect type" network directionality

Yu et al. explored drug repurposing through a network-based approach that identified drugs which have the opposite effect of a disease gene. Kyoto Encyclopedia of Genes and Genome pathways (KEGG) [38], drug–gene relationships from DrugBank, and GEO gene expression data were used to create a directed network which predicted actionable nodes (genes) for 898 drugs across nine diseases. Compared to a previous approach which lacked a directional component to the network [39], their method outperformed in almost all diseases, indicating the importance of including "effect-type" directionality in these networks. Interestingly, the network predicted mecasermin (recombinant IGF-1) as a potential drug with activity in *IGF1*-deficient AML with uncertain clinical implications (Table 4). Future experimental validation of novel and established targets across a

Table 4 Potential drug-mutation associations predicted from computational methods

In silico pathway-predicted drugs		
Mutated gene	Drug(s)	Method
EGFR	Osimertinib, Afatinib	Rubio-Perez et al.
KIT	Dasatinib, Sunitinib, Imatinib, Sorafinib,	Rubio-Perez et al.
NF1	Selumetinib	Rubio-Perez et al.
IGF1	Mecasermin	PDOD
PTEN	Erlotinib, Temsirolimus	TARGET
FLT3	PIM inhibitor (AZD1208)	Chebouba et al.

In silico synthetic lethality-predicted drugs		
Mutated gene	Synthetic lethal partner	Drug (of SL partner)
IDH1	*ACACA*	TOFA
IDH1	*PI4KA*	LY294002
Cohesin complex genes	*TP53*	Nutlin
KRAS	*PIK3R1*	PI3K/AKT/MTOR inhibitors
RUNX1	*JAK2*	Ruxolitinib
TP53	*MLL*	HDACi

Some predictions have been clinically tested in other cancers but not AML specifically. Kinase inhibitors are under active investigation for *KIT* mutations in AML. Generally, drug-mutation associations are based on a (i) pathway prediction wherein a given mutation is known to activate a druggable pathway or (ii) SL relationships where the SL partner of a mutated gene is druggable

number of cancers is required to endorse this and similar approaches.

Precision medicine in AML using a knowledge bank approach

One of the key limitations in data mining for mutation-specific therapies is the lack of patient-matched genomic and clinical datasets. This makes it difficult to investigate how differences in therapeutic intervention between patients with similar underlying mutations alter the overall survival of patients. Gerstung et al. recently sequenced 111 myeloid cancer genes in 1540 diagnostic AMLs to generate a knowledge bank of driver mutations and clinical outcomes [18]. Using a multistage model and knowledge bank approach, they showed that up to one-third of AML patients could be treated differently with an allograft compared to current clinical guidelines. As a tool for the larger research community, they developed a web portal (https://cancer.sanger.ac.uk/aml-multistage/) that displays Kaplan–Meier outcome predictions based on an assortment of user-defined clinical, genomic, and therapeutic variables. A major limitation of this tool is that the treatment variable only includes allogenic transplants. In the future, incorporating data for additional therapeutic interventions would greatly increase the potential of this tool to model treatments best correlated with favorable survival in mutation-specific contexts.

Extending druggable targets in AML by integrating data types

In silico prescription of anticancer drugs

A common limitation of many studies is that they only contain one data type, thus limiting the integration of different data modalities for additional correlations and insights. Using mutation, copy number, and gene fusion data from the TCGA AML study, along with established drug-target data, Rubio-Perez et al. demonstrate an "in silico prescription approach" to inform drug repurposing to increase the number of clinically-actionable cases across several cancers [40]. For AML-specific genes with actionable mutations targeted by approved FDA drugs, only three patients were identified. However, they identified 26 AML samples with oncogenic mutations in *EGFR*, *NF1*, *FLT3*, *MLL*, *KIT*, and *ABL1* as therapeutically actionable if other FDA-approved drugs were repurposed (Table 4). They estimate that although few cancers (5.9%) are treatable by approved agents according to current guidelines, up to 40.2% could benefit from different repurposing options, and up to 73.3% considering treatments currently under clinical investigation. The study highlighted the potential of identifying actionable events when multiple data types are integrated (mutation plus copy number), but further cancer-type specific validation is required.

Integrating Molecular Profiles with Actionable Therapeutics (IMPACT): matching whole exome sequencing data with druggable events

VAFs derived from whole exome sequencing can generate both tumor copy number information and mutation calls, which can then be integrated to infer complete or partial loss-of-function in a gene. Integrating mutation and copy number data from three lung cancer cohorts (IMPACT, http://tanlab.ucdenver.edu/IMPACT/about.html), Hintzsche et al. predicted tumor-specific loss of function in several genes and matched these to approved and investigational drugs [41]. We found that IMPACT correctly predicted enasidenib for *IDH2* and midostaurin for *FLT3* mutations. A major limitation of this tool is the lack of a cancer-specific gene expression filter, hence increasing the number of false-positive predictions. Additionally, the database has not been updated with new compounds and FDA approvals since its creation.

Data mining gene expression for novel therapeutic insights

Gene expression analyses of AML samples have revealed a number of subgroups that differ according to the over-expression or decreased expression of correlated gene sets. Some of these subgroups can be directly linked to cytogenetic risk group and prognosis, while others appear to reflect the biological properties and differentiation state of blasts. Despite the heterogeneity of AML, it is notable that in contrast to other cancer types, results obtained in specific gene expression studies have often validated in separate cohorts. For example, the various signatures that have been developed for risk prognostication are robust provided that specific features of datasets such as distribution of cytogenetic risk groups is accounted for. This has been facilitated by a number of studies comprising several thousand patients with well-annotated clinical data that have generated high-quality expression data on microarrays [42–44]. The availability of RNA sequencing has greatly improved the dynamic range of gene expression, resolution, and isoform specificity compared with microarrays, and a number of RNA-Seq data sets linked to clinical outcomes and genotyping are now available for hypothesis generation [15, 22, 23, 25]. Exploration of these data sets is ongoing, especially for mutation-specific pathways or SL targets.

While cytogenetic aberrations and common prognosis groups correlate strongly with gene expression profile, not all recurrent somatic mutations can be predicted based on gene-expression. Certain mutations appear to be enriched in some clusters, perhaps partly depending on the degree of clonal dominance in the sample. New platforms that facilitate the rapid examination of gene expression data sets are now being established. Moreover, novel computational methods are constantly being developed that attempt to distill mechanistic insights from these large datasets [45]. Despite the known limitations of expression profiling, it remains one of the most powerful molecular techniques for prognostication and prediction, particularly when integrated with other data types [9].

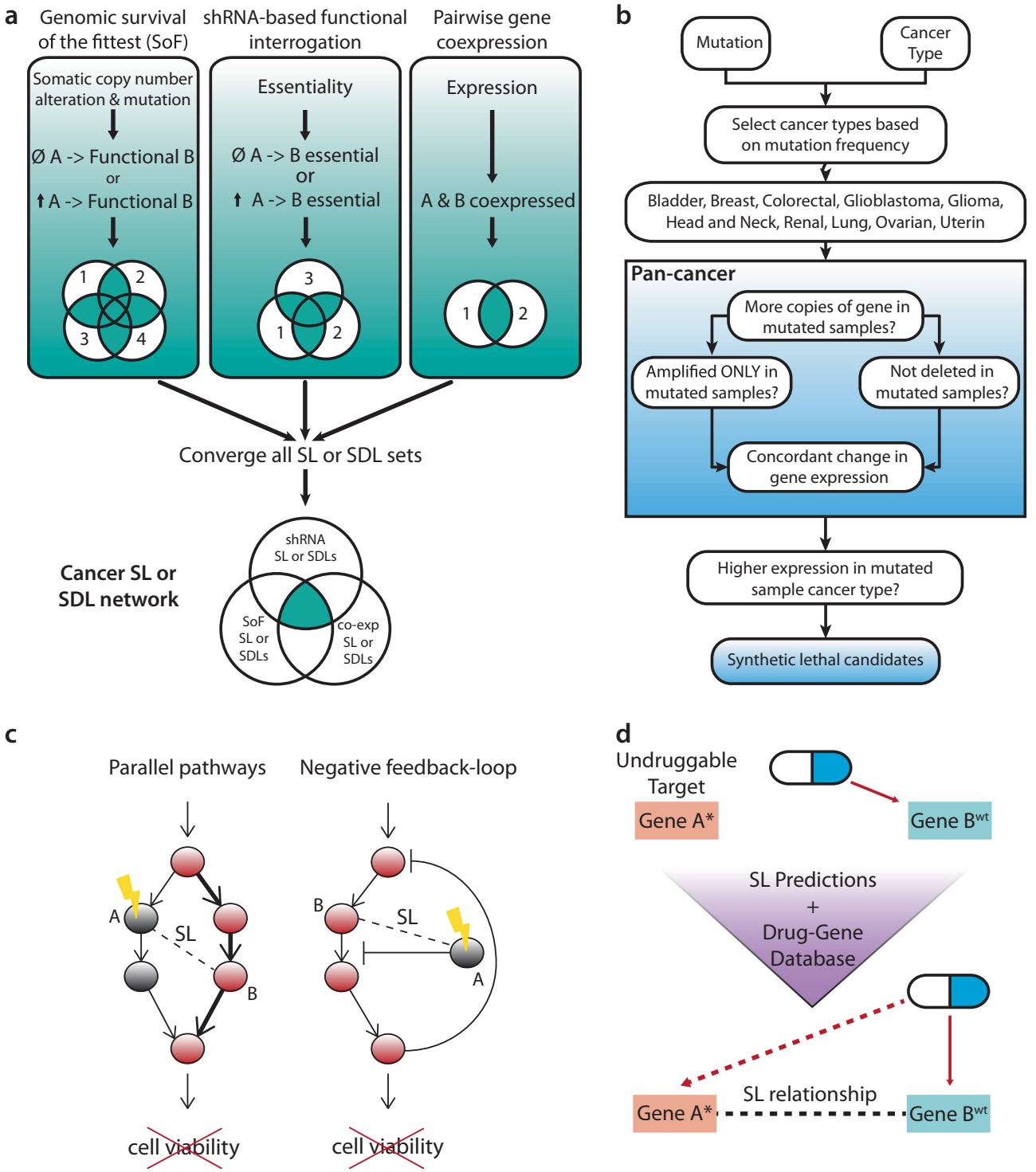
Data mining using protein data

Applying Boolean networks to discover unique pathways in primary resistant AML

The principles of Boolean logic can be used with other forms of data beyond gene expression, DNA methylation, and genetic mutations. Chebouba et al. used Boolean networks to predict treatment outcomes—responsive or resistant—using proteomics data from 191 AML patients [46]. Although the number of proteins and the number of patients in the final learning phase was modest (33 proteins and 26 patients), the network could predict clinical remission from proteomics data with an accuracy of 64.7%. The strength of this approach is that signaling pathways can be directly inferred from the network, resulting in new protein pathway associations. A similar method is described by Shnaps et al. using a network of protein–protein interaction data to assign a rank of relevance of disease-causing genes that may then be used to infer drug targets for each patient individually, given a set of mutations in AML [47]. Using previously published data for validation, their algorithm implicated Pim-1 kinase inhibition rather than PI3K for *FLT3* mutated AML (Table 4).

Data mining for synthetic lethal relationships in AML

The identification of mutation-specific therapies remains a major challenge for precision medicine, precisely because the majority are probably “undruggable”. A particularly promising approach is to identify alternative therapies that do not target the mutation directly, but may be deleterious when a mutation is present (Fig. 2d). In this case, the mutation and the second gene are called a synthetic lethal (SL) pair, since a defect in either gene is compatible with viability, but defects together are lethal to the cell. Large-scale functional screens in cell lines using shRNA, CRISPR, or small molecule libraries have been used for high-throughput identification of SL interactions [48–50]. However, since most large scale experimental screens must be performed in cell lines, they can be negatively impacted by limited coverage of certain mutations and the artificiality



of in vitro screening conditions, which cannot fully capture in vivo leukemia evolution in the patient. Screens can also be affected by false positive hits due to off-target effects and false negatives due to limited coverage of the library. New computational methods are therefore needed to complement the limitations of existing cell line screening methods and to reduce experimentation costs and time. Recently developed computational tools to predict SL include human orthologs of yeast SL interactions [51, 52], protein–protein networks [53], or metabolic network analysis [54], and multi-omics approaches centered on copy number changes and gene expression.

◀ **Fig. 2** Algorithm schematics for recent computational tools to predict mutation-specific synthetic lethal interactions. **a** DAISY uses three different data inputs in parallel to predict synthetic lethal and synthetic dosage lethal (SDL) interactions: (i) overlap of somatic copy number alterations and somatic mutation, (ii) results of essentiality from large scale cell-line shRNA screen results, and (iii) pairwise gene co-expression based on Pearson's correlation score. The overlap between these three orthogonal methods constructs a network of gene pairs for a given cancer that can then be tested in experimental systems. **b** MiSL uses data exclusively from primary cell patient samples across multiple cancer types to find a set of synthetic lethal pairs for a given mutation and a given cancer. The mutation must be recurrent in some pan-cancer samples. Data inputs are (i) copy number, (ii) somatic mutation (all mutation types), and (iii) RNA-seq expression data to infer mutation-specific Boolean implications for a given cancer type and a given mutation. Expression data is used as a final filter to ensure that the predicted gene pair has concordant gene expression changes with copy number alterations across cancer types and is over-expressed in the cancer type of interest in the presence of the mutation. **c** MutExSL utilizes copy number and expression to predict SL interactions in two different pathway-based approaches. In the parallel pathway model, deletions or reduced expression in gene A create a susceptibility for targeting the synthetic lethal target, gene B, and thus inhibit both pro-survival networks. Additionally, using a negative feedback-loop model, MutExSL identifies interactions where loss of function of gene A is predicted to increase pro-survival signaling through gene B, thus identifying gene B as a targetable node. **d** Combining computational synthetic lethal tools with drug–gene databases. Many recurrent mutations in cancer occur in undruggable genes and present a major clinical challenge. Mining results from SL tools and drug–gene databases can be used to identify SL relationships involving an undruggable gene where the SL partner is druggable, resulting in a possible mutation-specific sensitivity in a traditionally undruggable situation. Asterisk: mutated

Inferring synthetic lethality from mutually exclusive relationships in cancer

Validation for computational predictions can be inferred from retrospective analysis of primary outcome patient data and in vitro experimentation, but the gold standard is ultimately prospective validation in animal models and clinical trials. Srihari et al. recently developed MutExSL, an approach to infer SL events from copy number and gene expression data [55]. Unlike previous methods, which were based on yeast or in vitro models [16, 51, 53], their inferences were based entirely on primary patient tumor data and validated in tumor-derived cell lines and patient outcome data. MutExSL is unique in that it introduced a negative feedback-loop model of inferring SL by including gain-of-function events (amplification/overexpression) as potential essentiality targets for cancer-specific interactions (Fig. 2c). Correlating loss-of-function mutations with synthetic lethal candidates could assist in predicting druggable targets from sequencing data.

Multi-omic measurement of mutually exclusive LOF enriches for potential SL relationships

Wappett et al. developed a novel method that enriches for potential SL gene pairs based on assessing bimodality of gene expression [56]. The algorithm is available as the BiSEp CRAN package, but has not been applied to AML datasets. They note that for certain clinically established SL gene pairs such as *BRCA1* and *PARP1*, both *PARP1* or *PARP2* genes do not show a genetic loss in tumors but display markedly low mRNA level in a number of patient samples. Identifying bimodality and other non-normal distributions of gene expression can infer loss-of-function through unexpected loss of gene expression, and can supplement gene deletion and somatic mutation data (Fig. 1c, d). Their method was able to enrich for human homologs of yeast SL interactions and correctly identified *PARP3* and *BRCA1* as known SL pairs.

DAISY—method to enrich for synthetic lethals using shRNA, correlated expression, and genomic data

DAISY uses tumor genomic data and shRNA data from cell lines to predict SL interactions (Fig. 2a) [16]. DAISY uses three different statistical inference strategies: (i) detection of co-inactivation events from somatic copy alterations and somatic mutation data that occur significantly less than otherwise expected; (ii) shRNA-identified gene pairs where individual knockdown induced essentiality when the other gene is under-expressed or at low copy; and (iii) pairwise-gene expression using Pearson's coefficient based on the concept that SL pairs tend to participate in closely related processes and are therefore likely to be co-expressed. DAISY predicted a global network of potential SL interactions in human cells and has been well validated in predicting prognosis in breast cancer. However, DAISY primarily utilizes shRNA data from existing cell lines as part of its inference strategy, which means it will generate false negatives caused either by incomplete genetic knockdown or by inadequate representation of mutations in existing cell lines. An updated method has been recently published using TCGA data to identify clinically relevant interactions from a given set of inhibitors [57].

Mining Synthetic Lethals (MiSL): a platform for rapid detection of mutation-specific synthetic lethals using Boolean analysis

In collaboration with the Dill group in the Stanford Computer Science Department, we developed MiSL, an algorithm based on Boolean implications mined from large pan-cancer patient datasets to identify SL partners for specific cancer mutations in a given cancer type (Fig. 2b). MiSL demonstrated good concordance between predictions and mutation-specific SL partners identified by existing screens such as Achilles and prospective large-scale shRNA functional screens in our laboratory. It also identified known SL partners in AML and kidney cancer, and demonstrated

same-pathway enrichment of the predicted SL partners, which is consistent with previous work in yeast. MiSL is a first step toward solving two problems that are directly translatable to clinical applications: identifying novel mutation-specific SL interactions in a cancer-specific context, in particular *IDH1* mutation and metabolic genes such as *ACACA* in AML (Table 4), and pinpointing predictive genetic biomarkers that can guide more precise targeting of existing targeted therapies with a well-established binding partner.

Data mining for mutation-specific enrichment for established therapies

Venetoclax (in combination with hypomethylating agents) is an example of an effective agent under rapid clinical development for elderly AML patients that is not mutation-specific [58]. Biomarkers that could predict response to venetoclax are currently being evaluated, and it is likely that such determinants can be predicted through computational analyses (Table 4). For example, MiSL identified mutation of *IDH1* and deletion of *BCLW*, a *BCL2* family member, as SL partners in AML [17]. The *BCL2* family-*IDH1* SL relationship has been validated using AML cell lines expressing mutant *IDH1* and patient-derived xenografts [59], and is consistent with emerging clinical results from relapsed/refractory AML [60].

Gemtuzamab ozogamicin (GO) is another example of a semi-targeted agent with proven efficacy, but is not considered mutation-specific. GO has recently been approved for patients with *CD33*-positive AML at a safer dose regimen [61] and is being investigated as monotherapy in fit elderly patients [62]. Although it was not developed as a mutation-specific agent, increased CD33 expression on blasts by flow cytometry was shown for *NPM1* and *FLT3*-mutant AML [63–65]. Similarly, using both the LGA and cBioPortal online platforms, we confirmed the mutation-specific enrichment of *CD33* expression in *NPM1* and *FLT3*-mutated patients from RNA expression data (Table 3). Bioinformatic analyses may pinpoint minor molecular sub-groups that may show superior responses to approved drugs such as GO.

Subclonal mutations and implications for data mining

The success of precision medicine requires both the identification of unique disease-specific molecular events and the development of therapies targeted to these perturbations. Although there has been excitement in the field of personalized medicine regarding the recent approval of mutation-specific therapies in AML, resistance to enasidenib has already been shown to arise through the acquisition of additional *cis* and *trans* mutations in *IDH2* [66]. The use of targeted therapies has long been known to create selective pressure that can lead to the emergence of secondary mutations which render the drug ineffective [67]. It is also now accepted that resistance to targeted therapy can arise through the emergence of one or more subclones present in the initial tumor [68]. Together, these avenues of resistance discourage the idea that a single agent approach of precision medicine will generate long-lasting remission or cures. With this in mind, future computational approaches designed to identify mutation-specific vulnerabilities will have to include baseline assessments of tumor clonality to predict the emergence of a particular subclone upon mutation-specific treatment. Additionally, sequential monitoring of patient's tumors will be necessary to detect resistance mechanisms of subclonal (outgrowth) and non-subclonal (secondary mutations) expansion resulting from selective pressure of therapeutic intervention.

In addition to the well-established inter-patient heterogeneity observed in AML, several seminal studies have highlighted a significant degree of intra-patient heterogeneity [69–71]. This has been demonstrated on several levels, showing heterogeneity with respect to cell function [72], somatic mutation composition [3], and epigenetic state [73]. Indeed, studies identifying the sequential acquisition of somatic mutations and the clonal composition in AML allowed for the identification of pre-leukemic clones which act as precursors to AML [74, 75], a cellular hierarchy mimicking aspects of benign hematopoiesis with the identification of leukemic stem cells (LSCs) [76]. Moreover, subclones that might be present as a minor fraction at diagnosis may dominate at relapse, suggesting subclone specific responses to therapy and disease evolution [70]. While most studies delineating the clonal composition based on somatic mutations were performed using bulk tissue, biologically relevant epigenetic and transcriptional variability might be masked by bulk tissue analysis [77]. These aspects highlight the relevance to interrogate cell state and function as well as the composition of mutations on a single-cell level.

Single cell transcriptomic data for AML

With the advent of single cell RNA-seq (scRNA-seq) in 2009 [78], and more recent high-throughput methods able to interrogate from several hundred to thousands of individual cells (reviewed by Svennson et al. [79]), single cell transcriptome analysis is now widely available. Indeed, the blood system has been used as a testbed for demonstrating novel technological platforms in this area, probably in part

because both cancer and normal cells are readily available without tissue disaggregation that can distort transcriptional profiles [80]. These methods have successfully been applied to uncover numerous biological processes, including fate decisions in steady-state as well as stress hematopoiesis [81], comprehensive mapping of hematopoietic stem and progenitor differentiation (web interface: http://blood.stemcells.cam.ac.uk/single_cell_atlas.html) [82] and potential biomarkers of leukemia stem cells [83]. To provide a centralized resource of single-cell transcriptome data, the European Bioinformatics Institute (https://bioinfo.uth.edu/scrnaseqdb/, multiple species) facilitate access to data generated by multiple published scRNA-seq experiments. Importantly, scRNA-seq in the context of targeted therapies can give valuable insights into biomarkers predicting response and elucidating mechanisms of resistance [84]. This has already been applied in the context of the highly selective tyrosine kinase inhibitor of *FLT3-ITD* quizartinib [85] and elucidates the potential clinical implications in the near future.

Single-cell epigenomics and "multi-omics"

In addition, several methods interrogating the epigenetic state of single cells have been established, and this continues to be a rapidly changing, but powerful technology. There are four main data-types/techniques: (1) DNA methylation obtained from single-cell bisulfite sequencing [86], (2) DNA–protein interaction in single cells using ChIP [87], (3) single cell DNAse-seq [88] and single-cell ATAC-seq [89, 90] to elucidate the heterogeneity of chromatin accessibility, and (4) single-cell combinational indexed sequencing to map 3-dimensional chromatin structure [91]. In the context of AML, single cell ATAC-seq could be utilized to map the malignant cells at different points of their leukemic evolution compared to normal counterparts [92]. Additionally, emerging multi-omic methodologies have enabled simultaneous investigation of single cell transcriptomes and genomes [93–95]. Studies incorporating such techniques have revealed novel aspects of tumor biology, clonal architecture, and molecular underpinnings of therapeutic resistance [85]. Continued exploration of mutational heterogeneity and clonal/subclonal patterns of acquisition in a single cell manner will likely lead to a greater understanding of disease dynamics and improved clinical intervention.

Conclusions

The recent explosion of large datasets in cancer research has led to the development of multiple data repositories, web-based interaction portals, and a wide array of computational tools and methodologies to extract meaningful and hopefully therapeutically actionable insights relevant to leukemia researchers. Several of these platforms are extremely user-friendly and provide accessible tools for discovery and hypothesis generation. Despite great promise, data mining as a field is still in its infancy and has been minimally exploited in leukemia research. Many datasets are derived from in vitro experiments and only contain one datatype, thus limiting integration and representation. Increasingly, datasets based on human primary tumor and survival data have facilitated clinically-relevant interrogation and will continue to grow. As such resources grow in size and number, the application of artificial intelligence (e.g., machine learning, neural networks, pattern recognition, etc.) will likely become an increasingly routine methodology to discover "hidden" relationships in clinical and biological datasets, either individually or across multiple datasets using online tools such as MetaSignature [96]. Additionally, the emergence of in silico clinical trials using organ-on-a-chip or computational modeling of biological systems will likely help reduce the time and cost of drug screening and provide an avenue of predicting individual response across the heterogeneity of AML patients.

Computational methodologies can likely save valuable time in hypothesis generation, but careful experimental validation (both prospective and retrospective) is critical for their eventual translation prior to clinical testing. The strongest system to validate computational predictions of tumor-specific drug susceptibility is in analyzing primary human tumor data, but for AML, superior models that reflect the mutational diversity and clonal heterogeneity are required [97]. Validated mutation-specific therapies in AML can then be explored more broadly across other cancers in basket clinical trials.

Subclonal heterogeneity will likely be a determinant of resistance to mutation-specific targeted therapy, thus, computational approaches to track and model subclones to integrate with other established data types will be a major step forward in personalized medicine. In addition, the noncoding genome is under-utilized information that can assist in identifying novel mutation-specific vulnerabilities [98]. Because no two cases of AML are ever the same, both experimental biologists and hematologists will need to embrace computational tools to successfully advance mutation-specific medicine. Such advancement will require collaborative partnerships between physicians, molecular biologists, and bioinformaticians. Without the combined expertise of such groups, the optimal prediction and application of mutation-specific therapies in AML will never be fully realized.

Acknowledgements The authors thank Amy Fan for her critical review of this manuscript. BB is supported by NIH training grant 5T32CA9302-40. DT was funded by a Pathway-to-Independence K99 National Institutes of Health, National Cancer Institute Grant Number 5K99CA207731-02. Supported by NIH R01-CA188055 (to RM) and the Ludwig Institute for Cancer Research (to RM). RM is a Leukemia & Lymphoma Society Scholar. Due to citation limits, we apologize for not referencing all notable studies.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Khwaja A, Bjorkholm M, Gale RE, Levine RL, Jordan CT, Ehninger G, et al. Acute myeloid leukaemia. Nat Rev Dis Prim. 2016;2. https://doi.org/10.1038/nrdp.2016.10.
2. Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson G, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368:2059–74.
3. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016. https://doi.org/10.1182/blood-2016-03-643544.
4. Li H-Y, Deng D-H, Huang Y, Ye F-H, Huang L-L, Xiao Q, et al. Favorable prognosis of biallelic CEBPA gene mutations in acute myeloid leukemia patients: a meta-analysis. Eur J Haematol. 2015. https://doi.org/10.1111/ejh.12450.
5. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CAJ, van Doorn-Khosrovani BWW, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med. 2004;350:1617–28.
6. Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. dna methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell. 2010;17:13–27.
7. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014. https://doi.org/10.1038/nature13127.
8. O'Rawe J, Jiang T, Sun G, Wu Y, Wang W, Hu J, et al. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. Genome Med. 2013. https://doi.org/10.1186/gm432.
9. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483:603–7.
10. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2013;41. https://doi.org/10.1093/nar/gks1111.
11. Cowley GS, Weir BA, Vazquez F, Tamayo P, Scott JA, Rusin S, et al. Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. Sci Data. 2014;1. https://doi.org/10.1038/sdata.2014.35.
12. Rees MG, Seashore-Ludlow B, Cheah JH, Adams DJ, Price EV, Gill S, et al. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. Nat Chem Biol. 2016;12:109–16.
13. Seashore-Ludlow B, Rees MG, Cheah JH, Coko M, Price EV, Coletti ME, et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. Cancer Discov. 2015;5:1210–23.
14. Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell. 2013;154:1151–61.
15. Chang K, Creighton CJ, Davis C, Donehower L, Drummond J, Wheeler D, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45:1113–20.
16. Jerby-Arnon L, Pfetzer N, Waldman YY, McGarry L, James D, Shanks E, et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell. 2014;158:1199–209.
17. Sinha S, Thomas D, Chan S, Gao Y, Brunen D, Torabi D, et al. Systematic discovery of mutation-specific synthetic lethals by mining pan-cancer human primary tumor data. Nat Commun. 2017;8. https://doi.org/10.1038/ncomms15580.
18. Gerstung M, Papaemmanuil E, Martincorena I, Bullinger L, Gaidzik VI, Paschka P, et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. Nat Genet. 2017;49:332–40.
19. Dugas M, Schoch C, Schnittger S, Haferlach T, Danhauser-Riedl S, Hiddemann W, et al. A comprehensive leukemia database: integration of cytogenetics, molecular genetics and microarray data with clinical information, cytomorphology and immunophenotyping. Leukemia. 2001;15:1805–10.
20. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, et al. ArrayExpress—a public database of microarray experiments and gene expression profiles. Nucleic Acids Res. 2007;35:D747–50.
21. Edgar R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30:207–10.
22. Verhaak RGW, Wouters BJ, Erpelinck CAJ, Abbas S, Beverloo HB, Lugthart S, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. Haematologica. 2009;94:131–4.
23. Hebestreit K, Gröttrup S, Emden D, Veerkamp J, Ruckert C, Klein HU, et al. Leukemia gene atlas—a public platform for integrative exploration of genome-wide molecular data. PLoS One. 2012;7. https://doi.org/10.1371/journal.pone.0039148.
24. Bolouri H, Farrar JE, Triche T, Ries RE, Lim EL, Alonzo TA, et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med. 2018. https://doi.org/10.1038/nm.4439.
25. Lemieux S, Sargeant T, Laperrière D, Ismail H, Boucher G, Rozendaal M, et al. MiSTIC, an integrated platform for the analysis of heterogeneity in large tumour transcriptome datasets. Nucleic Acids Res. 2017;45. https://doi.org/10.1093/nar/gkx338.
26. Bagger FO, Sasivarevic D, Sohi SH, Laursen LG, Pundhir S, Sønderby CK, et al. BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. Nucleic Acids Res. 2016. https://doi.org/10.1093/nar/gkv1101.
27. Behrens K, Maul K, Tekin N, Kriebitzsch N, Indenbirken D, Prassolov V, et al. RUNX1 cooperates with FLT3-ITD to induce leukemia. J Exp Med. 2017;214:737–52.
28. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, et al. International Cancer Genome Consortium data portal—a one-stop shop for cancer genomics data. Database. 2011. https://doi.org/10.1093/database/bar026.

29. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013. https://doi.org/10.1126/scisignal.2004088.
30. Bagger FO, Rapin N, Theilgaard-Mönch K, Kaczkowski B, Thoren LA, Jendholm J, et al. HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis. Nucleic Acids Res. 2013. https://doi.org/10.1093/nar/gks1021.
31. Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell. 2011. https://doi.org/10.1016/j.cell.2011.01.004.
32. Miller JC, Brown BD, Shay T, Gautier EL, Jojic V, Cohain A, et al. Deciphering the transcriptional network of the dendritic cell lineage. Nat Immunol. 2012. https://doi.org/10.1038/ni.2370.
33. Lee S-I, Celik S, Logsdon BA, Lundberg SM, Martins TJ, Oehler VG, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. Nat Commun. 2018;9:42.
34. Sinha S, Thomas D, Yu L, Gentles AJ, Jung N, Corces-Zimmerman MR, et al. Mutant WT1 is associated with DNA hypermethylation of PRC2 targets in AML and responds to EZH2 inhibition. Blood. 2015;125:316–26.
35. Sahoo D, Dill DL, Tibshirani R, Plevritis SK. Extracting binary signals from microarray time-course data. Nucleic Acids Res. 2007. https://doi.org/10.1093/nar/gkm284.
36. Brown AS, Kong SW, Kohane IS, Patel CJ. ksRepo: a generalized platform for computational drug repositioning. BMC Bioinformatics. 2016;17. https://doi.org/10.1186/s12859-016-0931-y.
37. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, et al. The Comparative Toxicogenomics Database: Update 2017. Nucleic Acids Res. 2017;45:D972–8.
38. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999;27:29–34.
39. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med. 2011;3. https://doi.org/10.1126/scitranslmed.3003215.
40. Rubio-Perez C, Tamborero D, Schroeder MP, Antolín AA, Deu-Pons J, Perez-Llamas C, et al. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer Cell. 2015;27:382–96.
41. Hintzsche J, Kim J, Yadav V, Amato C, Robinson SE, Seelenfreund E, et al. IMPACT: a whole-exome sequencing analysis pipeline for integrating molecular profiles with actionable therapeutics in clinical samples. J Am Med Informatics Assoc. 2016. https://doi.org/10.1093/jamia/ocw022.
42. Metzeler KH, Hummel M, Bloomfield CD, Spiekermann K, Braess J, Sauerland MC, et al. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. Blood. 2008. https://doi.org/10.1182/blood-2008-02-134411.
43. Wouters BJ, Löwenberg B, Erpelinck-Verschueren CAJ, Van Putten WLJ, Valk PJM, Delwel R. Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. Blood. 2009. https://doi.org/10.1182/blood-2008-09-179895.
44. Tomasson MH, Xiang Z, Walgren R, Zhao Y, Kasai Y, Miner T, et al. Somatic mutations and germline sequence variants in the expressed tyrosine kinase genes of patients with de novo acute myeloid leukemia. Blood. 2008. https://doi.org/10.1182/blood-2007-09-113027.
45. Logsdon BA, Gentles AJ, Miller CP, Blau CA, Becker PS, Lee SI. Sparse expression bases in cancer reveal tumor drivers. Nucleic Acids Res. 2015. https://doi.org/10.1093/nar/gku1290.
46. Chebouba L, Miannay B, Boughaci D, Guziolowski C. Discriminate the response of acute myeloid leukemia patients to treatment by using proteomics data and answer set programming. BMC Bioinformatics. 2018;19. https://doi.org/10.1186/s12859-018-2034-4.
47. Shnaps O, Perry E, Silverbush D, Sharan R. Inference of personalized drug targets via network propagation. Pac Symp Biocomput. 2016;21:156–67.
48. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet. 2009;84:524–33.
49. Stark C. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34:D535–9.
50. Gilsdorf M, Horn T, Arziman Z, Pelz O, Kiner E, Boutros M. GenomeRNAi: a database for cell-based RNAi phenotypes. 2009 update. Nucleic Acids Res. 2009;38. https://doi.org/10.1093/nar/gkp1038.
51. Deshpande R, Asiedu MK, Klebig M, Sutor S, Kuzmin E, Nelson J, et al. A comparative genomic approach for identifying synthetic lethal interactions in human cancer. Cancer Res. 2013;73:6128–36.
52. Jacunski A, Dixon SJ, Tatonetti NP. Connectivity homology enables inter-species network models of synthetic lethality. PLoS Comput Biol. 2015;11. https://doi.org/10.1371/journal.pcbi.1004506.
53. Astsaturov I, Ratushny V, Sukhanova A, Einarson MB, Bagnyukova T, Zhou Y, et al. Synthetic lethal screen of an EGFR-centered network to improve targeted therapies. Sci Signal. 2010;3. https://doi.org/10.1126/scisignal.2001083.
54. Megchelenbrink W, Katzir R, Lu X, Ruppin E, Notebaart RA. Synthetic dosage lethality in the human metabolic network is highly predictive of tumor growth and cancer patient survival. Proc Natl Acad Sci USA. 2015;112:12217–22.
55. Srihari S, Singla J, Wong L, Ragan MA. Inferring synthetic lethal interactions from mutual exclusivity of genetic events in cancer. Biol Direct. 2015;10. https://doi.org/10.1186/s13062-015-0086-1.
56. Wappett M, Dulak A, Yang ZR, Al-Watban A, Bradford JR, Dry JR. Multi-omic measurement of mutually exclusive loss-of-function enriches for candidate synthetic lethal gene pairs. BMC Genomics. 2016;17. https://doi.org/10.1186/s12864-016-2375-1.
57. Lee JS, Das A, Jerby-Arnon L, Arafeh R, Auslander N, Davidson M, et al. Harnessing synthetic lethality to predict the response to cancer treatment. Nat Commun. 2018. https://doi.org/10.1038/s41467-018-04647-1.
58. DiNardo CD, Pratz KW, Letai A, Jonas BA, Wei AH, Thirman M, et al. Safety and preliminary efficacy of venetoclax with decitabine or azacitidine in elderly patients with previously untreated acute myeloid leukaemia: a non-randomised, open-label, phase 1b study. Lancet Oncol. 2018. https://doi.org/10.1016/S1470-2045(18)30010-X.
59. Chan SM, Thomas D, Corces-Zimmerman MR, Xavy S, Rastogi S, Hong WJ, et al. Isocitrate dehydrogenase 1 and 2 mutations induce BCL-2 dependence in acute myeloid leukemia. Nat Med. 2015. https://doi.org/10.1038/nm.3788.
60. Konopleva M, Pollyea DA, Potluri J, Chyla B, Hogdal L, Busman T, et al. Efficacy and biological correlates of response in a phase II study of venetoclax monotherapy in patients with acute myelogenous leukemia. Cancer Discov. 2016. https://doi.org/10.1158/2159-8290.CD-16-0313.

61. Appelbaum FR, Bernstein ID. Gemtuzumab ozogamicin for acute myeloid leukemia. Blood. 2017. https://doi.org/10.1182/blood-2017-09-797712.
62. Amadori S, Suciu S, Selleslag D, Aversa F, Gaidano G, Musso M, et al. Gemtuzumab ozogamicin versus best supportive care in older patients with newly diagnosed acute myeloid leukemia unsuitable for intensive chemotherapy: results of the randomized phase III EORTC-GIMEMA AML-19 trial. J Clin Oncol. 2016. https://doi.org/10.1200/JCO.2015.64.0060.
63. Ehninger A, Kramer M, Röllig C, Thiede C, Bornhäuser M, Von Bonin M, et al. Distribution and levels of cell surface expression of CD33 and CD123 in acute myeloid leukemia. Blood Cancer J. 2014. https://doi.org/10.1038/bcj.2014.39.
64. Krupka C, Kufer P, Kischel R, Zugmaier G, Bögeholz J, Köhnke T, et al. CD33 target validation and sustained depletion of AML blasts in long-term cultures by the bispecific T-cell-engaging antibody AMG 330. Blood. 2014. https://doi.org/10.1182/blood-2013-08-523548.
65. de Propris MS, Raponi S, Diverio D, Milani ML, Meloni G, Falini B, et al. High CD33 expression levels in acute myeloid leukemia cells carrying the nucleophosmin (NPM1) mutation. Haematologica. 2011. https://doi.org/10.3324/haematol.2011.043786.
66. Intlekofer AM, Shih AH, Wang B, Nazir A, Rustenburg AS, Albanese SK, et al. Acquired resistance to IDH inhibition through trans or cis dimer-interface mutations. Nature. 2018. https://doi.org/10.1038/s41586-018-0251-7.
67. Heinrich MC, Corless CL, Blanke CD, Demetri GD, Joensuu H, Roberts PJ, et al. Molecular correlates of imatinib resistance in gastrointestinal stromal tumors. J Clin Oncol. 2006. https://doi.org/10.1200/JCO.2006.06.2265.
68. Shah NP, Nicoll JM, Nagar B, Gorre ME, Paquette RL, Kuriyan J, et al. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. Cancer Cell. 2002. https://doi.org/10.1016/S1535-6108(02)00096-X.
69. Shlush LI, Mitchell A, Heisler L, Abelson S, Ng SWK, Trotman-Grant A, et al. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. Nature. 2017. https://doi.org/10.1038/nature22993.
70. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012. https://doi.org/10.1038/nature10738.
71. Paguirigan AL, Smith J, Meshinchi S, Carroll M, Maley C, Radich JP. Single-cell genotyping demonstrates complex clonal diversity in acute myeloid leukemia. Sci Transl Med. 2015. https://doi.org/10.1126/scitranslmed.aaa0763.
72. Klco JM, Spencer DH, Miller CA, Griffith M, Lamprecht TL, O'Laughlin M, et al. Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. Cancer Cell. 2014. https://doi.org/10.1016/j.ccr.2014.01.031.
73. Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F, et al. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. Nat Med. 2016. https://doi.org/10.1038/nm.4125.
74. Corces-Zimmerman MR, Hong W-J, Weissman IL, Medeiros BC, Majeti R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc Natl Acad Sci USA. 2014. https://doi.org/10.1073/pnas.1324297111.
75. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature. 2014. https://doi.org/10.1038/nature13038.
76. Thomas D, Majeti R. Biology and relevance of human acute myeloid leukemia stem cells. Blood. 2017;129:1577–85.
77. Wills QF, Livak KJ, Tipping AJ, Enver T, Goldson AJ, Sexton DW, et al. Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. Nat Biotechnol. 2013. https://doi.org/10.1038/nbt.2642.
78. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009. https://doi.org/10.1038/nmeth.1315.
79. Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. Nat Protoc. 2018. https://doi.org/10.1038/nprot.2017.149.
80. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017. https://doi.org/10.1038/ncomms14049.
81. Giladi A, Paul F, Herzog Y, Lubling Y, Weiner A, Yofe I, et al. Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. Nat Cell Biol. 2018. https://doi.org/10.1038/s41556-018-0121-4.
82. Nestorowa S, Hamey FK, Pijuan Sala B, Diamanti E, Shepherd M, Laurenti E, et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood. 2016. https://doi.org/10.1182/blood-2016-05-716480.
83. Giustacchini A, Thongjuea S, Barkas N, Woll PS, Povinelli BJ, Booth CAG, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. Nat Med. 2017. https://doi.org/10.1038/nm.4336.
84. Povinelli BJ, Rodriguez-Meira A, Mead AJ. Single cell analysis of normal and leukemic hematopoiesis. Mol Aspects Med. 2018. https://doi.org/10.1016/j.mam.2017.08.006.
85. Smith CC, Paguirigan A, Jeschke GR, Lin KC, Massi E, Tarver T, et al. Heterogeneous resistance to quizartinib in acute myeloid leukemia revealed by single-cell analysis. Blood. 2017. https://doi.org/10.1182/blood-2016-04-711820.
86. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods. 2014. https://doi.org/10.1038/nmeth.3035.
87. Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, Weitz DA, et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat Biotechnol. 2015. https://doi.org/10.1038/nbt.3383.
88. Jin W, Tang Q, Wan M, Cui K, Zhang Y, Ren G, et al. Genome-wide detection of DNase i hypersensitive sites in single cells and FFPE tissue samples. Nature. 2015. https://doi.org/10.1038/nature15740.
89. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. Nature. 2015. https://doi.org/10.1038/nature14590.
90. Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, et al. Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell. 2018. https://doi.org/10.1016/j.cell.2018.03.074.
91. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017. https://doi.org/10.1038/nmeth.4380.
92. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet. 2016. https://doi.org/10.1038/ng.3646.
93. Dey SS, Kester L, Spanjaard B, Bienko M, Van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol. 2015. https://doi.org/10.1038/nbt.3129.

94. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods. 2015. https://doi.org/10.1038/nmeth.3370.
95. Cheow LF, Courtois ET, Tan Y, Viswanathan R, Xing Q, Tan RZ, et al. Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. Nat Methods. 2016. https://doi.org/10.1038/nmeth.3961.
96. Haynes WA, Vallania F, Liu C, Bongen E, Tomczak A, Andres-Terrè M, et al. Empowering multi-cohort gene expression analysis to increase reproducibility. Pac Symp Biocomput. 2016;22:144–53.
97. Reinisch A, Thomas D, Corces MR, Zhang X, Gratzinger D, Hong WJ, et al. A humanized bone marrow ossicle xenotransplantation model enables improved engraftment of healthy and leukemic human hematopoietic cells. Nat Med. 2016. https://doi.org/10.1038/nm.4103.
98. Zhang W, Bojorquez-Gomez A, Velez DO, Xu G, Sanchez KS, Shen JP, et al. A global transcriptional network connecting noncoding mutations to changes in tumor gene expression. Nat Genet. 2018. https://doi.org/10.1038/s41588-018-0091-2.
99. Zong N, Kim H, Ngo V, Harismendy O. Deep mining heterogeneous networks of biomedical linked data to predict novel drug-target associations. Bioinformatics. 2017;33:2337–44.
100. Yu H, Choo S, Park J, Jung J, Kang Y, Lee D. Prediction of drugs having opposite effects on disease genes in a directed network. BMC Syst Biol. 2016;10. https://doi.org/10.1186/s12918-015-0243-2.