Exome-wide association study reveals 7 functional variants associated with ex-vivo drug response in acute myeloid leukaemia patients

Anil K Giri*[1,2], Jake Lin*[1], Konstantinos Kyriakidis[1], Henrikki Almusa[1], Matti Pirinen[1], Mark Daly[1,3,6,7]

[1)]Institute for Molecular Medicine Finland (FIMM), HiLIFE-Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland.

[2)]Foundation for the Finnish Cancer Institute (FCI), Tukholmankatu 8, 00290 Helsinki, Finland.

[3)]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

[6)]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.

[7)]Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.

*Equal author

Correspondence to:

Mark J Daly, PhD

Institute for Molecular Medicine Finland (FIMM)

Biomedicum Helsinki

PO Box 20 (Haartmaninkatu 8)

FI-00014 University of Helsinki, Finland

Tel: +358 2941 2583

Email: mark.daly@helsinki.fi

Abstract

Acute myeloid leukemia (AML) is a rare aggressive blood cancer without any long-term cure. Further, due to the extreme molecular heterogeneity of the disease, drug treatment response varies from patient to patient. The variability of drug response can cause unnecessary treatment in more than half of the patients with no or partial therapy responses leading to severe side effects, economic as well as time loss. Understanding the genetic risk factors underlying the drug response in AML can help with improved prediction of treatment responses and identification of biomarkers in addition to mechanistic insights to monitor treatment response.

Here, we report the results of the largest exome-wide association study (EWAS) of ex-vivo drug response performed to date with 175 AML cases and 47 drugs. We used information for 55423 exonic SNPs to perform the analysis. We identified exome-wide significant ($p<9.02 \times 10^{-7}$) associations for rs113985677 in *CCIN* with tamoxifen response, rs115400838 in *TRMT5* with idelalisib response, rs11878277 in *HDGFL2* with entinostat, and rs2229092 in *LTA* associated with vorinostat response.

Further, using multivariate genome-wide association analysis, we identified the association of rs11556165 in *ATRAID*, and rs11236938 in *TSKU* with the combined response of all 47 drugs and 29 nonchemotherapy drugs at the genome-wide significance level ($p<5x10^{-8}$). Additionally, a significant association of rs35704242 in *NIBAN1* was associated with the combined response of nonchemotherapy drugs(p=2.51x10-8) and BI.2536, gefitinib, and belinostat were identified as the central traits. Our study represents the largest EWAS study to date on ex-vivo drug response in AML and reports 7 new associated loci that help to understand the anticancer drug response in AML patients.

Introduction

Acute myeloid leukemia (AML) is a biologically heterogeneous aggressive hematological malignancy[1,2]. The inherent heterogeneity not only influences the clinical presentation of the disease but also affects the response and toxicity towards a therapy[3,4]. Hence, huge interindividual variability exists for drug response in AML patients[5,6,7].

The variability in the treatment response of AML patients is driven by a broad spectrum of genetic (e.g., somatic and germline variants) and nongenetic alterations (e.g., age) present in different patient groups.[8] Although the somatic mutations affecting treatment response in AML have been extensively studied in AML[9,10], the germline variants contributing to the drug response largely remain under-explored mostly because of a lack of well-characterized cohorts with both genetic and drug treatment data[11-13] . Further, the number of drug-treatment that can be explored in a clinical cohort is limited because patients are usually treated with one or two common drugs approved in the clinics, hence neglecting the other drugs. Furthermore, the collection of clinical responses to drugs in a large patient cohort will require long follow-up time, resources, and a network that could be beyond the capability of most academic laboratories.

The drug sensitivity and resistance testing (DSRT) platform can measure the ex-vivo response for hundreds of drugs at multiple doses in patient-derived primary cells[14-16]. The ex-vivo drug response can provide a reasonable estimate for the clinical response of the patient towards the drug[3,9] and has been increasingly used in guiding the treatment for refractory/relapse patients with no available standardized treatment. Further, the ex-vivo drug response measured in patient-derived-primary samples has been used as a surrogate for clinical response to identify genetic variants affecting treatment response in a patient.

Here, using the germline information from the exome-sequence and ex-vivo response data from 47 drugs measured in 75 AML patients, we performed an exome-wide association study to identify functional variants affecting drug response in AML. The result identifies 4 novel loci near *CCIN*, *TRMT5*, *HDGFL2*, and *LTA* genes affecting the ex-vivo response of tamoxifen, idelalisib, entinostat, and vorinostat in AML at exome-wide significance levels. Further, we also used a novel and efficient statistical framework "MetaPhat" to conduct and decomposed multivariate GWAS signals affecting ex-vivo drug response in AML patients. The analysis identified the association of loci in *NIBAN1/*, *ATRAID*, and *TSKU* with combined drug response in AML.

Methods

AML Patient's cohort

One seventy-five bone marrow aspirates (leukemic cells) or peripheral blood samples from adult patients were collected with signed informed consent with protocols in accordance with the Declaration of Helsinki [study acronym HRUHLAB2, Helsinki University Hospital Ethical Committee Statement 303/13/03/01/2011 (original), latest amendment 7 dated June 15, 2016. Latest Helsinki University Hospital study permit HUS/395/2018 dated February 13, 2018].

Mononuclear cells (MNC) were isolated by Ficoll-Paque centrifugation from freshly collected bone marrow and peripheral blood specimens of 108 diagnosed, and 67 relapsed/refractory stage patients. Skin biopsies were collected from all patients for germline DNA analysis.

Exome-sequencing of the samples

Genomic DNA was isolated from the skin biopsies samples using the DNeasy Blood & Tissue Kit (Qiagen) and was used for exome sequencing. For exome sequencing, 3 μg DNA was fragmented and processed according to the NEBNext DNA Sample Prep Master Mix protocol. Exome capture was performed using the Nimblegen SeqCap EZ v2 capture Kit (Roche NimbleGen). Sequencing of exomes was done using HiSeq1500, 2000, or 2500 instruments (Illumina), and approximately 4×10^7 bp paired-end reads were sequenced per sample. The data was curated and analyzed as described before[9, 17.]

Briefly, the sequenced raw exome reads were trimmed using Trimmomatic (version 0.36) to remove adapters and low-quality segments. The trimmed data were aligned to the human reference genome (GRCh build 37) using Burrows-Wheeler Aligner (bwa version 0.7.12, Wellcome Trust Sanger Institute, Cambridge, UK). After alignment, the potential PCR duplicates were removed using Picard (Broad Institute of Harvard and MIT, Cambridge, MA, USA), and BAM files were sorted, and indexed using SAMtools (Wellcome Trust Sanger Institute, Cambridge, UK). From the aligned reads, GATK (version 4.1.3.0) was used to joint call the samples for genetic variants. Variant annotation was performed using ANNOVAR and Ensembl version 6851.

We performed a stringent quality control analysis of the genetic data (Supplementary Figure 1). We removed 1 sample with a genomic call rate of less than 95%, and 9 samples with missing information for >10 of the drug response. We also removed variants without proper annotation (without rsid), call rate (<90%), MAF <0.01, and Hardy Weinberg equilibrium p-value < 10−5) were removed.

Drug resistance and sensitivity testing: We measured ex-vivo response for a library of 47 commercially available chemotherapeutic and targeted oncology compounds consisting of 34 approved drugs, and 13 investigation compounds (Supplementary Table S2) using the DSRT platform at the high throughput biomedicine unit at FIMM as described before[14,18]. Briefly, drugs were plated in a 10-fold dilution series of five concentrations covering a 1,000-fold concentration range on clear-bottomed 384-well plates (Corning #3712), using an Echo 550 Liquid Handler (Labcyte). The concentration ranges were selected for each compound separately to investigate their full dynamic range of the dose-response relationships.

Next, 20 µl of AML cells (approximately10,000) were added to a pre-drugged plate containing the mononuclear cell medium (MCM; PromoCell) supplemented with 0.5 µg/mL gentamicin and 2.5 µg/mL amphotericin B[3,18]. The cell was maintained in the culture media at 37 °C with 5% CO2 in a humidified incubator. After 72h incubation, cell viability was measured using CellTiter-Glo (Promega) reagent in duplicate using a PHERAstar FS (BMG Labtech) plate reader. As positive (total killing) and negative (non-effective) controls, we used 100µM benzethonium chloride and 0.1% dimethyl sulfoxide (DMSO), respectively, when calculating the relative inhibition percent.

The percentage inhibition was calculated by normalizing the cell viability to negative control wells containing only 0.1% dimethyl sulfoxide (DMSO) and positive control wells containing 100 µM cell-killing benzethonium chloride (BzCl). The drug responses passing the data quality assessment were included in further analysis. Drug sensitivity score (DSS), a measure of drug efficacy for each drug was calculated for each patient as described before[19]. Next, selective DSS (sDSS) was calculated by normalizing drug responses against 17 healthy controls.

Single SNP exome-wide association analysis: We only selected 47 drugs for an analysis whose ex-vivo response was measured in all the patients. The sDSS scores for each drug were inverse normalized using the inbuilt command in R (http://www.r-project.org/). We performed a linear regression model using PLINK[20] to test the association of SNPs with inverse normalized sDSS. Cell culture medium for the drug assay, disease stage (relapse/diagnosis), and first five principal components were used as covariates in the model. All other statistical analyses and plotting were performed using R programming. We used the Bonferroni correction for multiple tests (n = 55423 tests) and set the statistical significance and suggestive threshold to P-values less than 9.02×10^{-7} and 1×10^{-4}, respectively.

Multivariate exome-wide association analysis using MetaPhat: Several of the drugs measured in the experiment have a similar mechanism of action and act through the same pathway. As these functionally similar drugs have a correlated response in the studied samples, we used our novel method to conduct multivariate EWAS of a linear combination of correlated traits using univariate

GWAS for 55423 SNPs results as implemented in MetaPhat[21] (integrating metaCCA[22]). Further, we decomposed the multivariate associations into sets of central biomarker traits to understand the possible mechanism of action for the identified variant. This analysis tested each SNP separately for its simultaneous associations with all 47 phenotypes, 18 chemotherapies, and 29 nonchemotherapies drugs. A multivariate approach had the dual advantages of achieving data reduction and increasing statistical power compared with running 47 separate univariates EWAS.

Since we did not have a replication cohort, we performed a statistical experiment to validate the true nature of identified association in multivariate EWAS by estimating the chance of observing such an association in a random dataset. For this, we assigned random patient identity to each row of the phenotype file while keeping the correlation between the drug response the same and performed the univariate and multivariate association analysis with the genotypes using the same parameters.

Results

Single SNP association with the ex-vivo drug response identified association of 4 locus at exome-wide significance level

After stringent quality control, we analyzed a total of 55423 genetic markers for their association with single-drug response in 175 AML patients using linear regression adjusting for gender, age (at diagnosis for cases), cell culture medium, and the first 10 principal components. We detected an exome-wide significance level association for rs113985677 in *CCIN* with Tamoxifen (p=3.68x10-07), and rs115400838 in *TRMT5* with the idelalisib response (p=4.20x10-07). We also observed an EWAS level significance for rs11878277 in *HDGFL2* with entinostat (p=4.94x10-07), and rs2229092 in *LTA* with vorinostat response (p=8.28x10-07) (Table 1). The distribution of the association p-value across different chromosomes have been shown in Figure 1.

A good agreement was observed between the theoretical p-value distribution and calculated p-values for these drugs as shown in Supplementary Figure 2. The genomic inflation factor (λ) for the fitted model ranges between 0.98-1.01 which indicates the homogeneity of analyzed samples.

Table1: SNPs showing association with ex-vivo drug response at exome-wide significance levels in the analysis.

CHR	POS	SNP	Position	Gene	A1/A2	N	Effect size	P	Phenotype	Mechanism	Approval status
9	36170290	rs113985677	Missense R (Arg) > Q/p (Gln/Pro)	*CCIN*	A/G	175	4.45 (0.84)	3.68x10-07	Tamoxifen	Selective estrogen receptor modulator	Approved
14	61446147	rs115400838	Missence R (Ser) > P (Cys)	*TRMT5*	A/T	175	7.58(1.44)	4.20x10-07	Idelalisib	PI3K inhibitor	Approved
19	4498000	rs11878277	Synonymous S (Ser) > S (Ser)	*HDGFL2*	A/G	175	10.23(1.96)	4.94x10-07	Entinostat	HDAC inhibitor	Phase 2
6	31572980	rs2229092	Missense H (His) > P (Pro)	*LTA*	C/A	173	-4.9(1)	8.28x10-07	Vorinostat	HDAC inhibitor	Approved

Chromosomal positions of SNPs are based on National Center for Biotechnology Information genome build 38. The effect size was calculated with respect to the minor alleles. The association results presented were obtained from genotyped data in 175 patients.

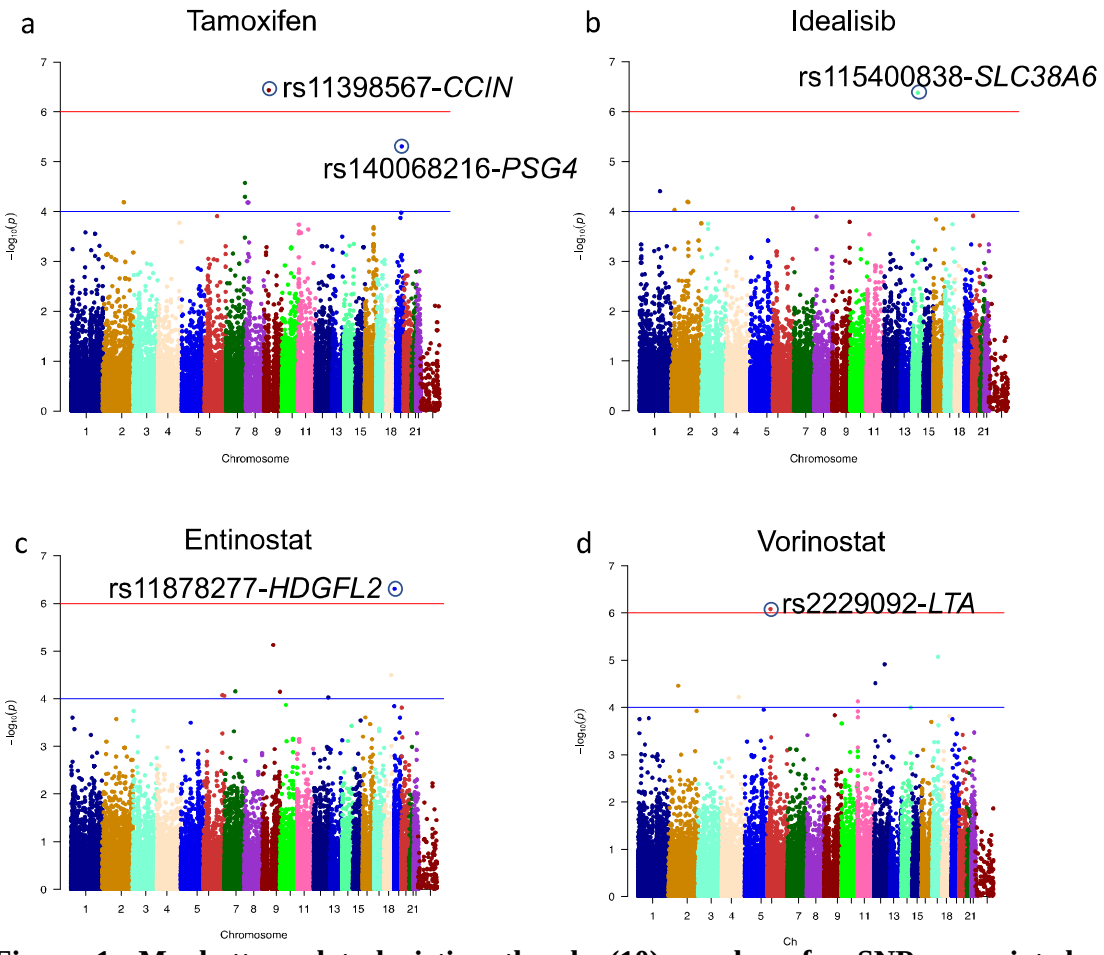


Figure 1: Manhattan plot depicting the -log(10) p-values for SNPs associated with (a) tamoxifen (b) idelalisib, (c) entinostat, and (d) vorinostat across different chromosomes *The p-value has been obtained using linear regression adjusted for disease stage (diagnosis/relapse), cell culture medium and first 5 principal components from genotype data.*

Suggestive association in exome-wide association analysis

Further, we observed significant associations of SNPs with the response of multiple drugs with a similar mechanism of action at the sub-EWAS level of significance suggesting a true nature of the association. For example, we observed the association of SNPs in the same genes with drugs having similar mechanisms of action. For example, rs2394516 and rs2394516 in *OR2J1* (Olfactory receptor 2J1) were associated with both cladribine (purine analog) and Clofarabine (another purine analog) suggesting a similar mechanism for influencing the drug response by identified variants.

Similarly, associations of rs201274224 in *PIK3CD* were observed with daunorubicin (topoisomerase II inhibitor), idarubicin (topoisomerase II inhibitor), and fludarabine (a purine analog) response suggesting a common mechanism of action probably through DNA damage and repair pathways. In total, we observed 295 suggestive associations within 248 unique SNP associated with the 47-drug response (Supplementary Figure 3).

Multi-variate association analysis identified an association of 4 additional SNPS at the exome-wide significance level

In the multivariate analysis, first, we checked the association of 55423 SNPs with a combined response of 47 drugs (Supplementary Figure 4) and observed a GWAS level association for rs11236938-G/A (*TSKU*) with the combined response (p=2.51x10-9, Figure 2a). Further, the decomposition of multivariate association identified tamoxifen, BI.2536, and belinostat as the central traits, as defined by optimal Bayesian Information Criteria (BIC)[23] subset and union with traits iteratively driving significant p-value for the multivariate association. Further, the analysis also identified an association of rs11556165-*ATRAID* (p=4.26x10-8) with the combined response of belinostat, fingolimod, and alvocidib as the central traits at the genome-wide significance level.

Next, we also performed multivariate analysis to check the association of SNPs with only a group of drugs with specific mechanisms of action (e.g., kinase inhibitors). The analysis of response from the 29 non-chemotherapy drugs identified an association of rs11236938-G/A in *TSKU* with combined drug response and tamoxifen, BI.2536, and belinostat (p=3.8x10-11, Figure 2b, Table 2) as the central trait. Further, rs11556165-*ATRAID* was also associated with EWAS significance level (p=3.98x10-9) with the combined response of non-chemo drugs with belinostat, fingolimod, alvocidib ass the major trait. Similarly, rs35704242-C/T in *NIBAN1* was associated with the combined nonchemotherapy drug response and, BI.2536, gefitinib, and belinostat as the main traits (p=2.51x10-8). The observed inflation factor for the p-values were 1.3 for all drug analysis and 1.17 for only nonchemotherapy drug response analysis (Supplementary Figure 5)

Table 2: SNPs showing association with ex-vivo response of nonchemotherapy drugs as group at geome-wide significance levels in the multivariate analysis.

CHR	BP	Variant (Gene)	Allele	MAF	Pval	Central traits	Drug group
11	76507403	rs11236938-G/A (*TSKU*)	S (Ser) > N (Asn)	0.03	3.8x10-11	Tamoxifen, BI.2536, Belinostat	Non chemotherapy
2	27438615	rs11556165-G/C(*ATRAID*)	D (Asp) > H (His)	0.04	3.98x10-9	Belinostat, Fingolimod, Alvocidib	Non chemotherapy

1	184764824	rs35704242-C/T (*NIBAN1)*	D (Asp) > N (Asn)	0.04	2.51x10-8	BI.2536, Gefitinib, Belinostat	Non chemotherapy

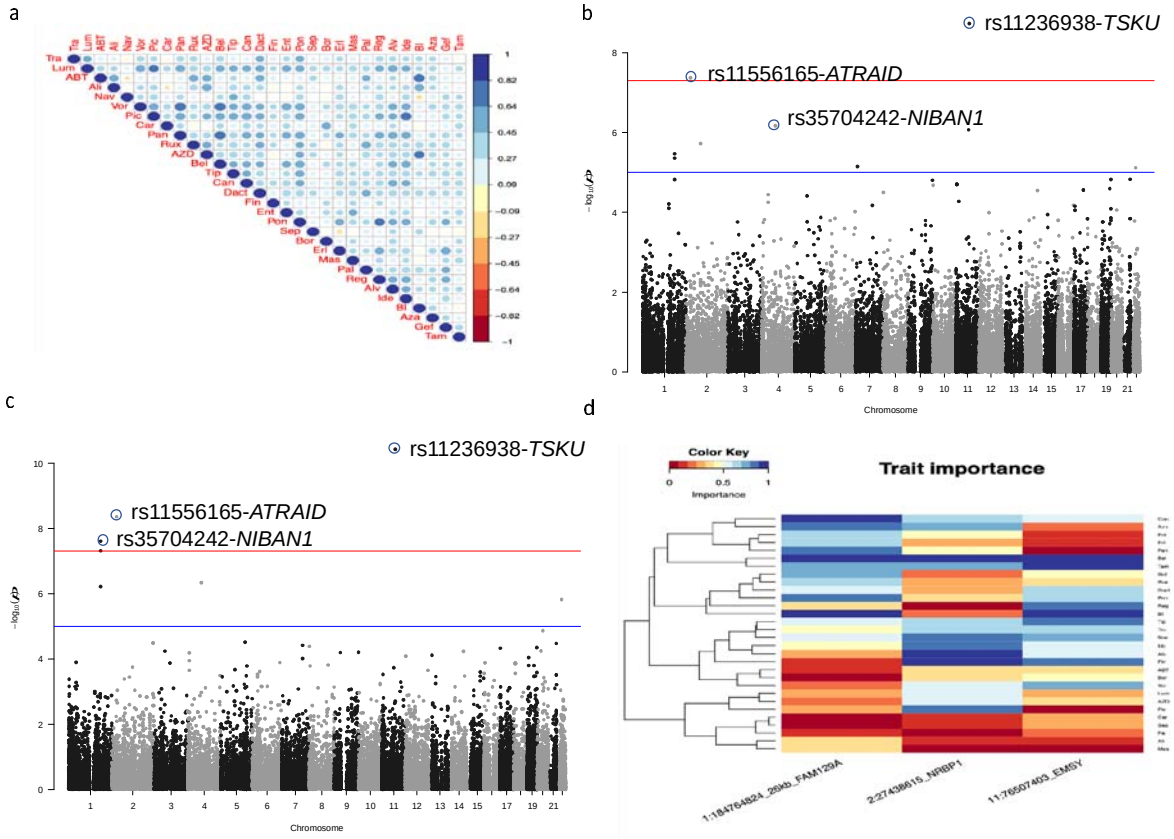


Figure 2: Multivariate association analysis of exonic SNPs with the linear combination of drug response *(a) Heatmap showing the correlation between the analyzed drugs based on the response from the 175 patients. (b) Manhattan plot depicting the -log (10) p-values for SNPs associated with all 47 drugs across different chromosomes in multivariate analysis.* ***(c)*** *Manhattan plot depicting the -log (10) p-values for SNPs associated with only kinase inhibitors* (d)*Heatmap showing the importance of each kinase inhibitor response on the association of identified signals (The increase of importance has been shown from red to blue).*

Validation of the identified association: To check the validity of the association and significance level of the p-value, we also performed an association analysis by creating 200 sets of randomized phenotypes (200x47 phenotypes) where we kept the correlation between responses of the 47 drugs the same. The univariate association analysis did not identify any signals at the exome-wide significance level at any of the random phenotypes.

However, only 8 out of 200 times any exome-wide significant association was observed when the response of all 47 drugs was analysed in the multivariate analysis. Similarly, only 5 EWAS-level significant associations were observed out of 200 times when only the response of 29 non-chemo drugs were analysed together. These analyses firmly ascertain the true nature of identified associations and revealed a maximum of 4% chance of the identified signals as false positive.

Discussion

Genetic risk factors affecting the response of most cancer drugs are poorly defined[24,25]. Most of the large-scale genomics study in AML has explored the effect of somatic mutations on drug response undermining the influence of germline variant except for a few drugs[26,27]. Further, these studies exploring the effect of the germ-line variants on drug response have only analyzed the polymorphisms in genes encoding transporters, metabolizers, or molecular targets of chemotherapy agents neglecting most parts of the genomes[12,28]. Although specific somatic mutations have been associated with the drug response for some of the targeted therapy, such mutations account for only a minority of cases and the great majority of patients does not have such clear marker[27]. The large amount of time and resources required to collect the clinical response data has been one of the major obstacles to performing large, genome-wide association studies to identify pharmacogenetic variants affecting drug response. Our study, which comprised over 175 AML cases profiled for an ex-vivo response of 47 drugs, represents one of the largest such studies performed to date.

We identified 4 EWAS significant variants affecting the ex-vivo drug response in our cohort overall using the univariate association analysis. However, we could not validate these signals because of the lack of other cohorts having this kind of data. We identified the association of rs113985677, a missense variant (changes arginine to glutamine or proline) in *CCIN* (calcin) with Tamoxifen ex-vivo response. Calcin is a cytoskeleton protein and has been associated with lateral sinus thrombosis, spermatogenic Failure 9, globozoospermia[29,30] and is required for the formation of the sperm head and maintain the nuclear structure in mice[31]. Downregulation of the cytoskeleton genes, connexin 43, induces tamoxifen-resistance in breast cancer[32] cells via activation of c-Src/PI3K/Akt signaling and Epithelial to mesenchymal transition (EMT) of cells. The identified variant in the calcin gene may alter the tamoxifen response by altering the cytoskeleton dynamics during EMT.
Further, the variant is an expression quantitative trait locus for *GLIPR2* gene in the esophagus and salivary gland (P=6.5x10-12, and 7.9x10-8) where the presence of alternate allele is associated with increased expression of *GLIPR2*. *GLIPR2* is particularly expressed in leukocytes and has a role in EMT transition[33], and autophagy[34], suggesting a possible role in drug response via these mechanisms. However, a functional study will be needed to understand the exact role of genes and variants.

We also found a significant association between rs115400838, an upstream variant in *SLC38A6*, and a missense variant in *TRMT5* with an idelalisib response. *SLC38A6* is a glutamine and glutamate transporter[35] and has been associated with angiotensin-converting enzyme inhibitor-induced cough[36]. The available evidence suggests that SLC38A6 may be a direct transporter of idelalisib itself or its substrates. We also observed a significant association of rs11878277 (*HDGFL2*) with entinostat in the univariate analysis. Entinostat is a histone deacetylase inhibitor and is under clinical trial for multiple cancers. *HDGFL2* (Hepatoma-Derived Growth Factor-Related Protein 2) is dispensable for Mixed Lineage Leukemia (MLL)-rearranged (MLL-R) leukemogenesis[37] and regulates gene expression via HRP2-DPF3a-BAF complex formation[38].

Similarly, a significant association was observed for a missense variant rs2229092 (H (His) > P (Pro) in Lymphotoxin Alpha (*LTA*) gene with vorinostat response, an HDAC inhibitor. LTA,

formerly known as tumor necrosis factor-beta (TNF-β) is a toxin produced from lymphocytes (e.g., T and B) that can kill other cell types. It acts as a causative factor for autocrine and paracrine activation of canonical and noncanonical NF-κB signaling as well as promotes JAK2/STAT6 signaling in Hodgkin lymphoma[39] resulting in prevention of apoptosis. Furthermore, LTA shapes the expression of lymphoid and myeloid-specific genes in blood. Given the wide role of *LTA* in immunity and tumor survival[40], LTA may influence vorinostat response by modulating cancer survival-related gene expression in synergy with vorinostat.

The multivariate analysis identified rs11556165-*ATRAID* as one of the significant SNPs at EWAS level. ATRAID (all-trans retinoic acid-induced differentiation factor), is a genetic target of nitrogen-containing-bisphosphonates (N-BP), a commonly used drug to treat osteoporosis and other bone-related diseases[41],[33]. Further, ATRAID binds with SLC37A3 (solute carrier family 37 member A3) which is a lysosomal membrane protein with 12 transmembrane segments and helps in NBP transport.

The multivariate analysis identified rs11556165-*CAD* (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase) as one of the significant SNPs at EWAS level (Table 2). CAD protein is a multifunctional multi-domain enzyme involved in the initial three speed-limiting steps of pyrimidine nucleotide synthesis. CAD is hyperactivated in most of the cancer and targeting it inhibits cancer cell survival, and self-renewal[41].

Genetic target of nitrogen-containing-bisphosphonates (N-BP), a commonly used drug to treat osteoporosis and other bone-related diseases[41,42]. Further, ATRAID binds with *SLC37A3* (solute carrier family 37 member A3) which is a lysosomal membrane protein with 12 transmembrane segments and helps in NBP transport[41].

We also identified the association of rs35704242 in niban apoptosis regulator 1 (NIBAN1) multi-drug response with BI 2536, gefitinib, and belinostat as the central trait set. *NIBAN1* protein is highly expressed in multiple cancers such as such as renal, colorectal, and thyroid cancer[43,44] and may play an important role in cell survival during cellular stress such as ultraviolet irradiation cell migration and proliferation[43], cell autophagy[45] and immune microenvironment[46]. The association of BI 2536, gefitinib, and belinostat central responses which induce induces cell cycle, and apoptosis of cancer cells with *NIBAN1* align with the versatile function of NIBAN1 in cellular homeostasis seems justifiable.

We also observed a significant association of tsukushi (*TSKU*), a small leucine-rich proteoglycan, with overall kinase (p=2.51 x10-8) with Tamoxifen, BI.2536, Belinostat as the central traits. Tsukushin is an extracellular matrix protein and is also a hepatokine. TSKU is involved in different signaling pathways including BMP pathway, notch pathway, transforming growth factor-β (TGF-β) pathway, mitogen-activated protein kinase (MAPK)/fibroblast growth factor (FGF) pathway, and β-catenin/Wnt pathway[47]. Absence of TSKU increases cleaved poly ADP-ribose polymerase (PARP) protein levels and caspase 3/7 activity thereby increasing carcinogenicity in neuroblastoma cells[48]. Further, *TSKU* expression is strongly associated with poor overall survival in lung cancer[49]. Given its role in carcinogenicity and tumor survival, TSKU may affect drug response in cancer. However, further functional study is needed to reveal the exact mechanism by which TSKU affects drug response in AML.

In summary, our exome-wide analysis identified a potential genetic locus affecting the ex-vivo drug response of some of the common cancer therapy. These loci are in proximity to genes with a possible role in the transport of the respective drugs inside the cell. Our results provide a basis for

initial exploration of the role of genes associated with this locus. However, replication studies to confirm the observed associations are needed in larger patient-group.

Limitation

The identification of only a single associated SNP without other associated confirming SNPs in the identified locus region in the univariate analysis suggests this finding may be spurious and should be interpreted with caution. These SNPs are not located on genes implicated in drug transport pathways gives literature support for the finding. However, the lack of another cohort for validation is a major limitation of the study. Though being a novel and significantly large cohort for a rare disease like AML, our study is still smaller than genome-wide association studies reported in other malignancies and is potentially limited in its power to detect associations. Additionally, in light of the increasing evidence that AML tumors are genetically heterogeneous, genetic predisposing factors likely vary across patients. Although we attempted to address this potential weakness by performing correlation based MetaPhat analysis, these analyses are based on drugs with different targets and mechanisms of action, resulting in the potential to miss associations due to the lack of statistical power, as well as the potential to identify spurious associations.

Reference

1. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013 May 30; **368**(22)**:** 2059-2074.
2. Quek L, David MD, Kennedy A, Metzner M, Amatangelo M, Shih A, *et al.* Clonal heterogeneity of acute myeloid leukemia treated with the IDH2 inhibitor enasidenib. *Nat Med* 2018 Aug; **24**(8)**:** 1167-1177.
3. Ianevski A, Lahtela J, Javarappa KK, Sergeev P, Ghimire BR, Gautam P, Vähä-Koskela M, Turunen L, Linnavirta N, Kuusanmäki H, Kontro M, Porkka K, Heckman CA, Mattila P, Wennerberg K, Giri AK, Aittokallio T. Patient-tailored design for selective co-inhibition of leukemic cell subpopulations. *Sci Adv*., **7**(8):eabe4038 (2021).
4. Giri, A.K., Ianevski, A. & Aittokallio, T. Genome-wide off-targets of drugs: risks and opportunities. *Cell Biol Toxicol* **35**, 485-487 (2019).
5. Daver N, Garcia-Manero G, Basu S, Boddu PC, Alfayez M, Cortes JE, *et al.* Efficacy, Safety, and Biomarkers of Response to Azacitidine and Nivolumab in Relapsed/Refractory Acute Myeloid Leukemia: A Nonrandomized, Open-Label, Phase II Study. *Cancer Discov* **9,** 370-383 (2019).
6. Giri A.K., Ianevski, A. High-throughput screening for drug discovery targeting the cancer cell-microenvironment interactions in hematological cancers. *Expert Opin Drug Discov.* **17**, 181-190 (2021)
7. Schroeder T, Rautenberg C, Kruger W, Platzbecker U, Bug G, Steinmann J, *et al.* Treatment of relapsed AML and MDS after allogeneic stem cell transplantation with decitabine and DLI-a retrospective multicenter analysis on behalf of the German Cooperative Transplant Study Group. *Ann Hematol* **97,** 335-342 (2018).
8. Bottomly, D., *et al.* Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. *Cancer Cell* **40**, 850-864 e859 (2022).
9. Malani, D., *et al.* Implementing a Functional Precision Medicine Tumor Board for Acute Myeloid Leukemia. *Cancer Discov* **12**, 388-401 (2021).

10. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, *et al.* Functional genomic landscape of acute myeloid leukaemia. *Nature* **562,** 526-531 (2018).
11. Williams MS, Amaral FM, Simeoni F, Somervaille TC. A stress-responsive enhancer induces dynamic drug resistance in acute myeloid leukemia. *J Clin Invest*; **130:** 1217-1232 (2020).
12. Lamba JK. Genetic factors influencing cytarabine therapy. *Pharmacogenomics* **10,** 1657-1674 (2009).
13. Zhu KW, Chen P, Zhang DY, Yan H, Liu H, Cen LN, *et al.* Association of genetic polymorphisms in genes involved in Ara-C and dNTP metabolism pathway with chemosensitivity and prognosis of adult acute myeloid leukemia (AML). *J Transl Med,* **16:** 90 (2018).
14. Ianevski, A., Giri, A.K., Gautam, P. *et al.* Prediction of drug combination effects with a minimal set of experiments. *Nat Mach Intell* **1**, 568–577 (2019).
15. Ianevski, A., Giri, A.K. & Aittokallio, T. SynergyFinder 3.0: an interactive analysis and consensus interpretation of multi-drug synergies across multiple samples. *Nucleic Acids Res* **50**, W739–43 (2022).
16. Ianevski, A., Giri, A.K. & Aittokallio, T. SynergyFinder 2.0: visual analytics of multi-drug combination synergies. *Nucleic Acids Res* **48**, W488-W493 (2020).
17. Wahnschaffe L, Braun T, Timonen S, Giri AK, Schrader A, Wagle P, *et al.* JAK/STAT-Activating Genomic Alterations Are a Hallmark of T-PLL. *Cancers (Basel)* 2019 Nov 21; **11**(12).
18. Ianevski A, Timonen S, Kononov A, Aittokallio T, Giri AK. SynToxProfiler: An interactive analysis of drug combination synergy, toxicity and efficacy. *PLoS Comput Biol* 2020 Feb; **16**(2)**:** e1007604.
19. Yadav B, Pemovska T, Szwajda A, Kulesskiy E, Kontro M, Karjalainen R, *et al.* Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Sci Rep* 2014 Jun 5; **4:** 5193.
20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007 Sep; **81**(3)**:** 559-575.
21. Lin J, Tabassum R, Ripatti S, Pirinen M. MetaPhat: Detecting and Decomposing Multivariate Associations From Univariate Genome-Wide Association Statistics. *Front Genet* 2020; **11:** 431.
22. Cichonska, A., *et al.* metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. *Bioinformatics* **32**, 1981-1989 (2016).
23. Schwarz G. Estimating the Dimension of a Model. *Ann. Statist.* **6**(2): 461-464 (1978).
24. Menden MP, Casale FP, Stephan J, Bignell GR, Iorio F, McDermott U, *et al.* The germline genetic component of drug sensitivity in cancer cell lines. *Nat Commun* 2018 Aug 23; **9**(1)**:** 3385.
25. Hariprakash JM, Vellarikkal SK, Keechilat P, Verma A, Jayarajan R, Dixit V, *et al.* Pharmacogenetic landscape of DPYD variants in south Asian populations by integration of genome-scale data. *Pharmacogenomics* 2018 Feb;**19**(3):227-241.
26. Little P, Lin DY, Sun W. Associating somatic mutations to clinical outcomes: a pan-cancer study of survival time. *Genome Med* 2019 May 28; **11**(1)**:** 37.

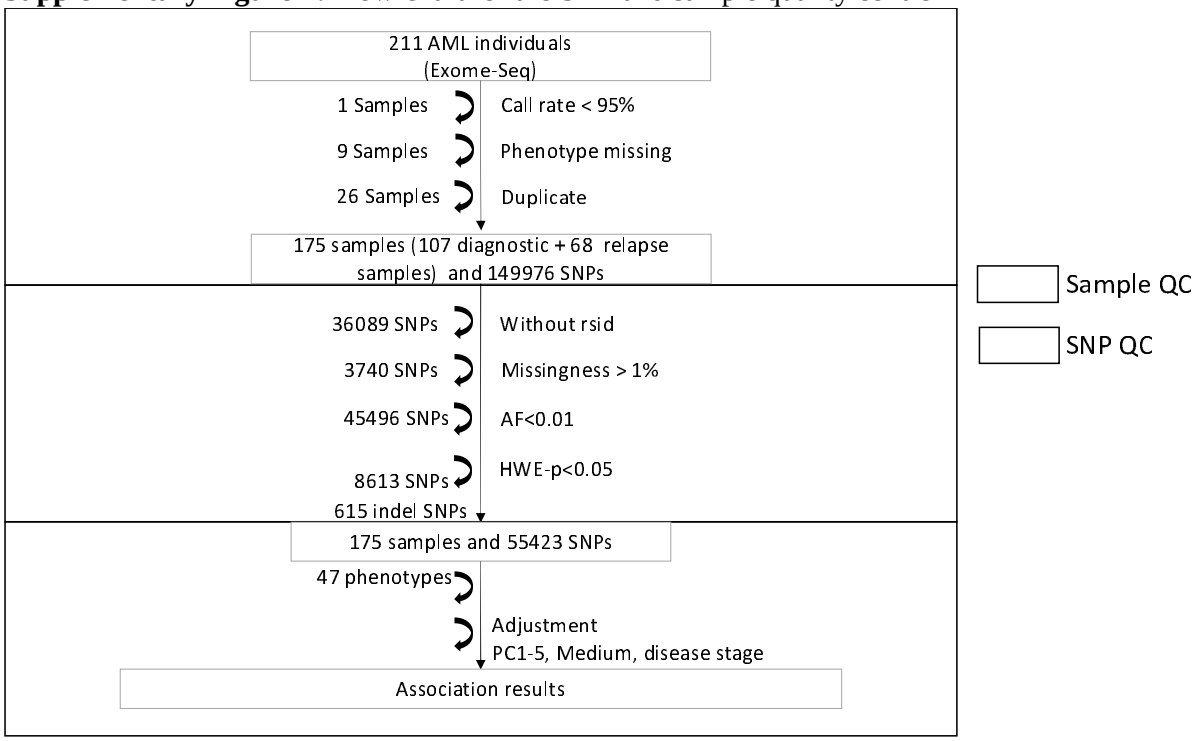
27. Levatic, J., Salvadores, M., Fuster-Tormo, F. & Supek, F. Mutational signatures are markers of drug sensitivity of cancer cells. *Nat Commun* **13**, 2926 (2022).

28. Huang RS, Duan S, Kistner EO, Bleibel WK, Delaney SM, Fackenthal DL, *et al.* Genetic variants contributing to daunorubicin-induced cytotoxicity. *Cancer Res* 2008 May 1; **68**(9)**:** 3161-3168.
29. Celse T, Cazin C, Mietton F, Martinez G, Martinez D, Thierry-Mieg N, *et al.* Genetic analyses of a large cohort of infertile patients with globozoospermia, DPY19L2 still the main actor, GGN confirmed as a guest player. *Hum Genet* 2021 Jan; **140**(1)**:** 43-57.
30. Li YZ, Wu RF, Zhu XS, Liu WS, Ye YY, Lu ZX, *et al.* Identification of a novel deletion mutation in DPY19L2 from an infertile patient with globozoospermia: a case report. *Mol Cytogenet* 2020; **13:** 24.
31. Zhang, X.Z., Wei, L.L., Jin, H.J., Zhang, X.H. & Chen, S.R. The perinuclear theca protein Calicin helps shape the sperm head and maintain the nuclear structure in mice. *Cell Rep* **40**, 111049 (2022).
32. Wu, D.P., *et al.* Cx43 deficiency confers EMT-mediated tamoxifen resistance to breast cancer via c-Src/PI3K/Akt pathway. *Int J Biol Sci* **17**, 2380-2398 (2021).
33. Huang, S., *et al.* GLIPR-2 overexpression in HK-2 cells promotes cell EMT and migration through ERK1/2 activation. *PLoS One* **8**, e58574 (2013).
34. Aalberts, M., *et al.* Identification of distinct populations of prostasomes that differentially express prostate stem cell antigen, annexin A1, and GLIPR2 in humans. *Biol Reprod* **86**, 82 (2012).
35. Gandasi NR, Arapi V, Mickael ME, Belekar PA, Granlund L, Kothegala L, *et al.* Glutamine Uptake via SNAT6 and Caveolin Regulates Glutamine-Glutamate Cycle. *Int J Mol Sci* 2021 Jan 25; **22**(3).
36. Hallberg P, Persson M, Axelsson T, Cavalli M, Norling P, Johansson HE,et al . Genetic variants associated with angiotensin-converting enzyme inhibitor-induced cough: a genome-wide association study in a Swedish population. Pharmacogenomics. 2017 Feb;**18**(3):201-213..
37. Van Belle S, El Ashkar S, Cermakova K, Matthijssens F, Goossens S, Canella A, *et al.* Unlike Its Paralog LEDGF/p75, HRP-2 Is Dispensable for MLL-R Leukemogenesis but Important for Leukemic Cell Survival. *Cells* 2021 Jan 19; **10**(1).
38. Zhu X, Lan B, Yi X, He C, Dang L, Zhou X, *et al.* HRP2-DPF3a-BAF complex coordinates histone modification and chromatin remodeling to regulate myogenic gene transcription. *Nucleic Acids Res*; **48**(12)**:** 6563-6582.
39. von Hoff L, Kärgel E, Franke V, McShane E, Schulz-Beiss KW, Patone G, *et al.* Autocrine LTA signaling drives NF-κB and JAK-STAT activity and myeloid gene expression in Hodgkin lymphoma. Blood.**133**(13):1489-1494 (2019).
40. Fhu, C.W., *et al.* Reed-Sternberg cell–derived lymphotoxin-α activates endothelial cells to enhance T-cell recruitment in classical Hodgkin lymphoma. *Blood* **124**, 2973-2982 (2014).
41. Yu Z, Surface LE, Park CY, Horlbeck MA, Wyant GA, Abu-Remaileh M, *et al.* Identification of a transporter complex responsible for the cytosolic entry of nitrogen-containing bisphosphonates. *Elife* 10; **7 (**2018).
42. Surface LE, Burrow DT, Li J, Park J, Kumar S, Lyu C, et al. *ATRAID* regulates the action of nitrogen-containing bisphosphonates on bone. *Sci Transl Med*, **12**(544):eaav9166 (2020).

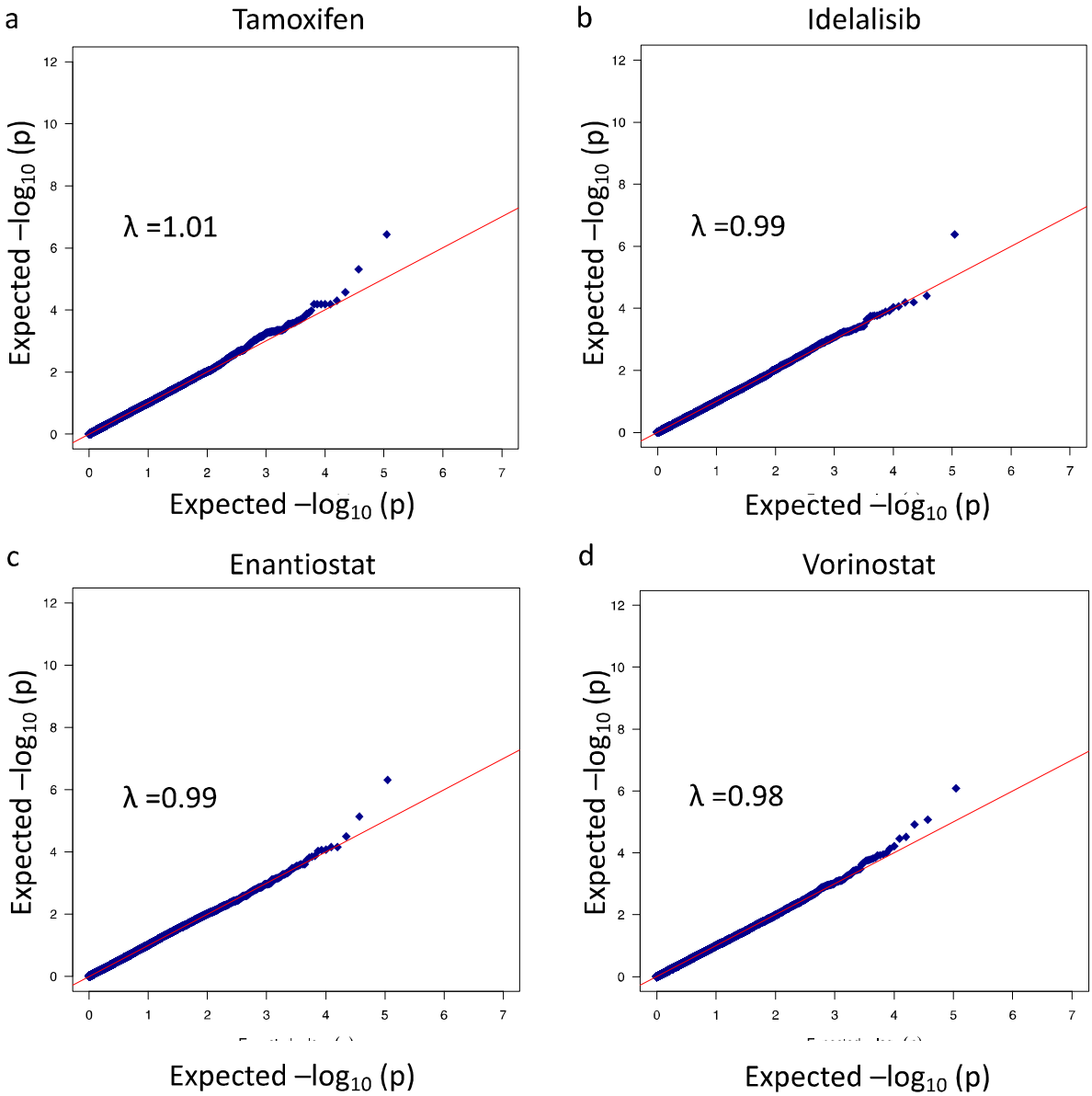
43. Feng, X., *et al.* miR-4521-FAM129A axial regulation on ccRCC progression through TIMP-1/MMP2/MMP9 and MDM2/p53/Bcl2/Bax pathways. *Cell Death Discov* **5**, 89 (2019).
44. Diana, P. & Carvalheira, G.M.G. NIBAN1, Exploring its Roles in Cell Survival Under Stress Context. *Front Cell Dev Biol* **10**, 867003 (2022).
45. Ji, H., *et al.* AKT-dependent phosphorylation of Niban regulates nucleophosmin- and MDM2-mediated p53 stability and cell apoptosis. *EMBO Rep* **13**, 554-560 (2012).
46. Zhang, R., *et al.* Tissue Treg Secretomes and Transcription Factors Shared With Stem Cells Contribute to a Treg Niche to Maintain Treg-Ness With 80% Innate Immune Pathways, and Functions of Immunosuppression and Tissue Repair. *Front Immunol* **11**, 632239 (2020).
47. Deng, X., Li, Y., Guo, C., Zhao, Z. & Yuan, G. Novel roles of Tsukushi in signaling pathways and multiple disease processes. *Biofactors* **47**, 512-521 (2021).
48. Ichikawa, T., Horie-Inoue, K., Ikeda, K., Blumberg, B. & Inoue, S. Steroid and xenobiotic receptor SXR mediates vitamin K2-activated transcription of extracellular matrix-related genes and collagen accumulation in osteoblastic cells. *J Biol Chem* **281**, 16927-16934 (2006).
49. Huang H, Zhang D, Fu J, Zhao L, Li D, Sun H, *et al*. Tsukushi is a novel prognostic biomarker and correlates with tumor-infiltrating B cells in non-small cell lung cancer. *Aging* (Albany NY). 10;13(3):4428-4451 (2021).

Supplementary Information

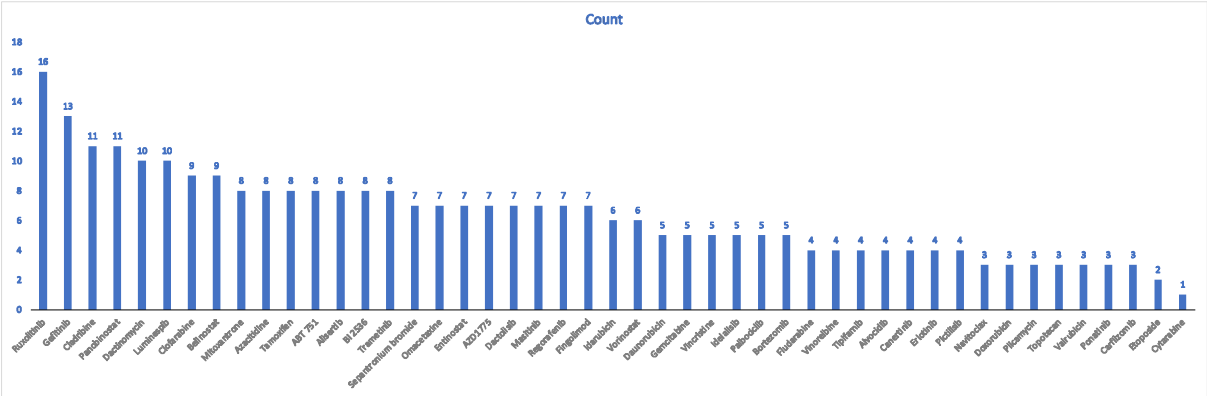
Supplementary Figure 1: Flow-chart for the SNP and sample quality control



Supplementary Figure 2: QQ plot for the univariate GWAS associations

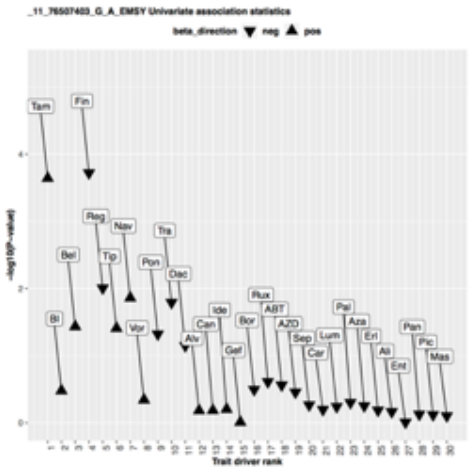


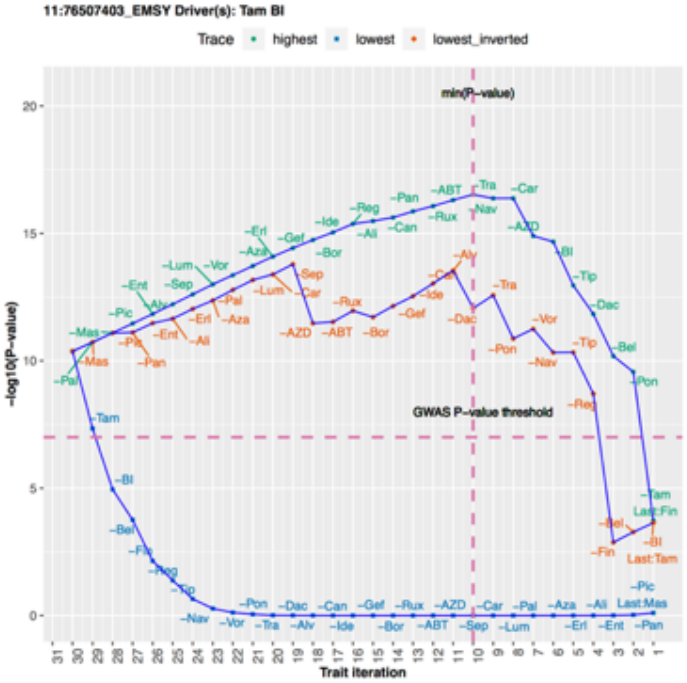
Supplementary Figure 3: The number of associations observed for each trait at sub exome wide significance level



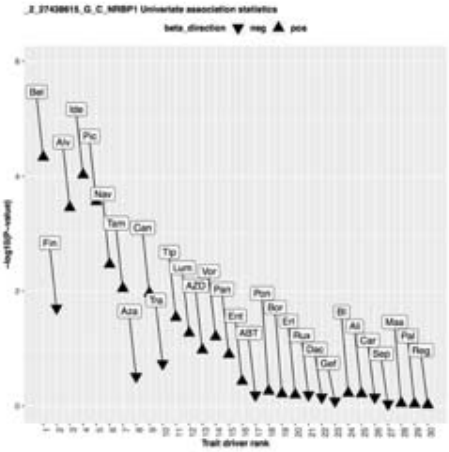
Supplementary Figure 4: QQ plot for the multivariate analysis

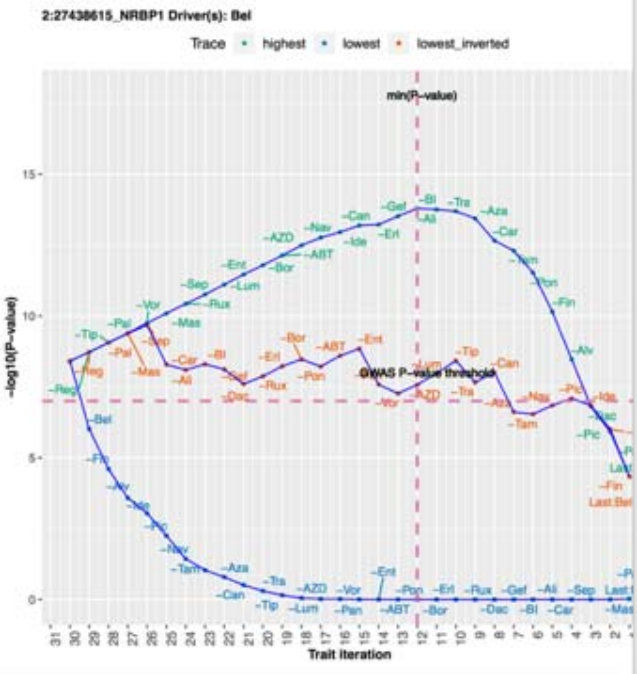
rs35704242

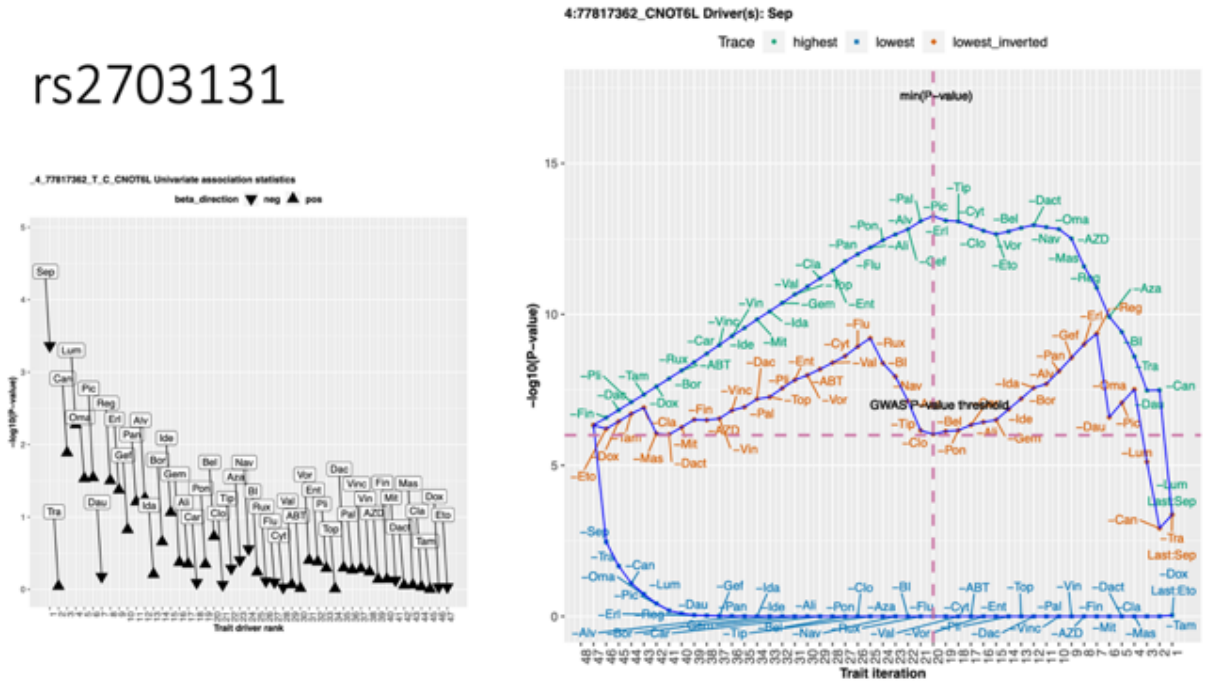




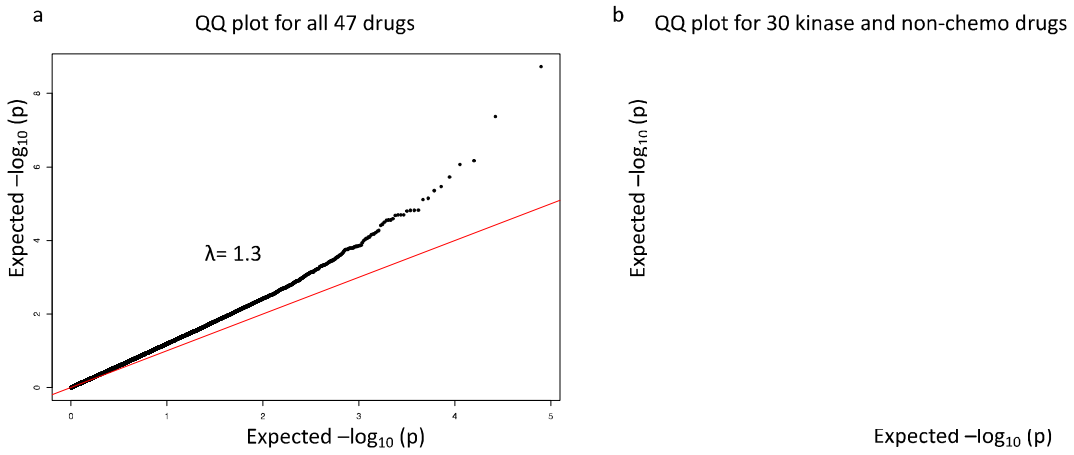
rs11556165







Supplementary Figure 5: QQ plot for the multivariate analysis



Supplementary Table 1: Details of the analyzed drugs and their mechanism of action

Trait	**Diagnosis (N=108)**	**Relapsed/refractory (n=67)**
Culture medium (MCM/CM)	82/26	46/21
Age (median, range)	64(21-81)	60(19-76)

Sex (M/F/NA)	45/55/8	20/23/24

NA: Not available,

Supplementary Table 2: Details of the analyzed drugs and their mechanism of action

SN	Drug	Mechanism/Targets	Class.explained	High.phase/Approval.status	Solvent	High.conc.(nM)
1	Tamoxifen	Estrogen receptor antagonist	F. Hormone therapy	Approved	DMSO	10000
2	Navitoclax	Bcl-2/Bcl-xL inhibitor	G. Apoptotic modulator	Investigational (Ph 2)	DMSO	10000
3	Belinostat	HDAC inhibitor	E. Differentiating/ epigenetic modifier	Approved (US)	DMSO	10000
4	Dactolisib	mTOR/(PI3K) inhibitor	B. Kinase inhibitor	Investigational (Ph 2)	DMSO	1000
5	Cladribine	Antimetabolite; Purine analog	A. Conv. Chemo	Approved	DMSO	1000
6	Idarubicin	Topoisomerase II inhibitor	A. Conv. Chemo	Approved	DMSO	1000
7	Panobinostat	HDAC inhibitor	E. Differentiating/ epigenetic modifier	Approved	DMSO	1000
8	Clofarabine	Antimetabolite; Purine analog	A. Conv. Chemo	Approved	DMSO	
9	Fludarabine	Antimetabolite; Purine analog	A. Conv. Chemo	Approved	DMSO	10000
10	Vorinostat	HDAC inhibitor	E. Differentiating/ epigenetic modifier	Approved	DMSO	10000
11	Bortezomib	Proteasome inhibitor (26S subunit)	L. Protease/proteasome inhibitor	Approved	DMSO	
12	Tipifarnib	Farnesyltransferase inhibitor	E. Differentiating/ epigenetic modifier	Investigational (Ph 3)	DMSO	10000
13	Ponatinib	Broad TK inhibitor	B. Kinase inhibitor	Approved	DMSO	1000
14	Cytarabine	Antimetabolite, interferes with DNA synthesis	A. Conv. Chemo	Approved	DMSO	
15	Carfilzomib	Proteasome inhibitor (20S subunit)	L. Protease/proteasome inhibitor	Approved	DMSO	1000
16	Etoposide	Topoisomerase II inhibitor	A. Conv. Chemo	Approved	DMSO	10000
17	Idelalisib	PI3K inhibitor, p110?-selective	B. Kinase inhibitor	Approved	DMSO	10000
18	Palbociclib	CDK4/6 inhibitor	B. Kinase inhibitor	Approved	AQ	10000
19	Mitoxantrone	Topoisomerase II inhibitor	A. Conv. Chemo	Approved	DMSO	1000
20	Ruxolitinib	JAK1&2 inhibitor	B. Kinase inhibitor	Approved	DMSO	10000

21	Topotecan	Topoisomerase I inhibitor. Camptothecin analog	A. Conv. Chemo	Approved	DMSO	10000
22	Canertinib	pan-HER inhibitor	B. Kinase inhibitor	Investigational (Ph 3)	DMSO	10000
23	Valrubicin	Topoisomerase II inhibitor	A. Conv. Chemo	Approved	DMSO	5000
24	Vincristine	Mitotic inhibitor. Vinca alkaloid microtubule depolymerizer	A. Conv. Chemo	Approved	DMSO	1000
25	Dactinomycin	RNA and DNA synthesis inhibitor	A. Conv. Chemo	Approved	DMSO	1000
26	Daunorubicin	Topoisomerase II inhibitor	A. Conv. Chemo	Approved	DMSO	1000
27	Doxorubicin	Topoisomerase II inhibitor	A. Conv. Chemo	Approved	DMSO	1000
28	Entinostat	HDAC inhibitor	E. Differentiating/ epigenetic modifier	Investigational (Ph 2)	DMSO	10000
29	Luminespib	HSP90 inhibitor	K. HSP inhibitor	Investigational (Ph 2)	DMSO	1000
30	Fingolimod	S1PR antagonist	X. Other	Approved	DMSO	10000
31	Sepantronium bromide	Survivin inhibitor	G. Apoptotic modulator	Investigational (Ph 2)	DMSO	10000
32	Azacitidine	Nucleoside analog DNA methyl transferase inhibitor	E. Differentiating/ epigenetic modifier	Approved	DMSO	10000
33	BI 2536	PLK1 inhibitor	B. Kinase inhibitor	Investigational (Ph 2)	DMSO	1000
34	Masitinib	KIT inhibitor	B. Kinase inhibitor	Investigational (Ph 3)	DMSO	10000
35	Omacetaxine	Protein synthesis inhib (80 S ribosome)	A. Conv. Chemo	Approved	DMSO	10000
36	Pictilisib	PI3K inhibitor, pan-class I	B. Kinase inhibitor	Investigational (Ph 2)	DMSO	10000
37	Alisertib	Aurora A inhibitor	B. Kinase inhibitor	Investigational (Ph 3)	DMSO	10000
38	Erlotinib	EGFR inhibitor	B. Kinase inhibitor	Approved	DMSO	10000
39	Plicamycin	RNA synthesis inhibitor	A. Conv. Chemo	Approved	DMSO	
40	Vinorelbine	Mitotic inhibitor. Vinca alkaloid microtubule depolymerizer	A. Conv. Chemo	Approved	DMSO	
41	Gefitinib	EGFR inhibitor	B. Kinase inhibitor	Approved	DMSO	10000
42	Regorafenib	B-Raf, c-Kit, VEGFR2 inhibitor	B. Kinase inhibitor	Approved	DMSO	10000
43	Trametinib	MEK1/2 inhibitor	B. Kinase inhibitor	Approved	DMSO	250
44	ABT-751	Mitotic inhibitor. Colchicine site binding microtubule depolymerizer.	A. Conv. Chemo	Investigational (Ph 2)	DMSO	10000
45	AZD1775	Wee1 inhibitor	B. Kinase inhibitor	Investigational (Ph 2)	DMSO	10000

46	Gemcitabine	Antimetabolite; Nucleoside analog	A. Conv. Chemo	Approved	DMSO	1000
47	Alvocidib	CDK inhibitor	B. Kinase inhibitor	Investigational (Ph 2)	DMSO	10000

Supplementary Table 3: SNPs showing association with ex-vivo response of all 47 drugs as a group at the genome-wide significance levels in the multivariate analysis.

CHR	BP	Variant (Gene)	Allele	MAF	Pval	Central traits	Drug group
11	76507403	rs11236938-G/A (*TSKU*)	S (Ser) > N (Asn)	0.03	2.51x10-9	Tamoxifen, BI.2536, Belinostat	Non chemotherapy
2	27438615	rs11556165-G/C(*ATRAID*)	D (Asp) > H (His)	0.04	4.26x10-8	Belinostat, Fingolimod, Alvocidib	Non chemotherapy