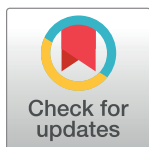


RESEARCH ARTICLE

A novel integrative multi-omics approach to unravel the genetic determinants of rare diseases with application in sinusoidal obstruction syndrome

Nicolas Waespe^{1,2,3}, Simona Jurkovic Mlakar², Isabelle Dupanloup⁴, Mohamed Aziz Rezgui⁵, Henrique Bittencourt^{5,6,7}, Maja Krajcinovic^{5,6,7}, Claudia E. Kuehni^{1,8}, Tiago Nava^{2,9}, Marc Ansari^{2,9*}



1 Institute of Social and Preventive Medicine, Childhood Cancer Research Group, University of Bern, Bern, Switzerland, **2** Faculty of Medicine, Department of Pediatrics, Gynecology and Obstetrics, CANSEARCH Research Platform in Pediatric Oncology and Hematology, University of Geneva, Geneva, Switzerland, **3** Graduate School for Cellular and Biomedical Sciences (GCB), University of Bern, Bern, Switzerland, **4** Swiss Institute of Bioinformatics, Lausanne, Switzerland, **5** Department of Pediatrics, Charles-Bruneau Cancer Center, CHU Sainte-Justine Research Center, Montreal, QC, Canada, **6** Department of Pediatrics, Clinical Pharmacology Unit, CHU Sainte-Justine, Montreal, QC, Canada, **7** Faculty of Medicine, Department of Pharmacology, University of Montreal, Montreal, QC, Canada, **8** Department of Pediatrics, Division of Pediatric Oncology and Hematology, University Hospital of Bern, Bern, Switzerland, **9** Department of Women, Child and Adolescent, Division of Pediatric Oncology and Hematology, Geneva University Hospitals and University of Geneva, Geneva, Switzerland

☞ These authors contributed equally to this work.

* marc.ansari@hcuge.ch

OPEN ACCESS

Citation: Waespe N, Mlakar SJ, Dupanloup I, Rezgui MA, Bittencourt H, Krajcinovic M, et al. (2023) A novel integrative multi-omics approach to unravel the genetic determinants of rare diseases with application in sinusoidal obstruction syndrome. PLoS ONE 18(4): e0281892. <https://doi.org/10.1371/journal.pone.0281892>

Editor: Gary S. Stein, University of Vermont, UNITED STATES

Received: May 14, 2022

Accepted: February 2, 2023

Published: April 5, 2023

Copyright: © 2023 Waespe et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results presented in the study were included in a previously published article and its supplementary information files: Ansari M, Petrykey K, Rezgui MA, Del Vecchio V, Cortyl J, Ralph R-O et al. Genetic Susceptibility to Hepatic Sinusoidal Obstruction Syndrome in Pediatric Patients Undergoing Hematopoietic Stem Cell Transplantation. Biol Blood Marrow Transplant 2020; 26: 920–927. The raw data cannot be shared publicly because of information that could

Abstract

Background

Genotype-phenotype analyses of rare diseases often suffer from a lack of power, due to small sample size, which makes identifying significant associations difficult. Sinusoidal obstruction syndrome (SOS) of the liver is a rare but life-threatening complication of hematopoietic stem cell transplantation (HSCT). The alkylating agent busulfan is commonly used in HSCT and known to trigger SOS. We developed a novel pipeline to identify genetic determinants in rare diseases by combining *in vitro* information with clinical whole-exome sequencing (WES) data and applied it in SOS patients and controls.

Methods

First, we analysed differential gene expression in six lymphoblastoid cell lines (LCLs) before and after incubation with busulfan. Second, we used WES data from 87 HSCT patients and estimated the association with SOS at the SNP and the gene levels. We then combined the results of the expression and the association analyses into an association statistic at the gene level. We used an over-representation analysis to functionally characterize the genes that were associated with a significant combined test statistic.

Results

After treatment of LCLs with busulfan, 1708 genes were significantly up-, and 1385 down-regulated. The combination of the expression experiment and the association analysis of

compromise the privacy of research participants. Data are available from the CANSEARCH Research platform in pediatric oncology and hematology, University of Geneva, Geneva, Switzerland (contact via e-mail: research@cansearch.ch) for researchers who meet the criteria for access to confidential data.

Funding: L.552ff: “CK and MA were supported by grants from the Swiss National Science Foundation (<https://www.snf.ch/en>, grant number 320030_153389 for CK and 31BL30_185396 for both); NW, SJM, ID, TN, MA, MK were supported by the CANSEARCH Foundation (<https://cansearch.ch/en>); MAR, HB, MK were supported by the OAK Foundation (<https://oakfnd.org>) and Charles-Bruneau Foundation (<https://www.charlesbruneau.qc.ca/en>). There was no additional external funding received for this study. The funding agencies had no role in the design, execution, interpretation, or writing of the study.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: ALL, Acute lymphoblastic leukemia; CI, Confidence interval; CNS, Central nervous system; DNA, Deoxyribonucleic acid; FDR, False discovery rate; HSCT, Hematopoietic stem cell transplantation; HWE, Hardy–Weinberg equilibrium; LCL, Lymphoblastoid cell line; LD, Linkage disequilibrium; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAF, Minor allele frequency; ORA, Over-representation analysis; PRS, Polygenic risk scores; RNA, Ribonucleic acid; SNP, Single nucleotide polymorphism; SOS, Sinusoidal obstruction syndrome; UTR, Untranslated region; VEGAS2, Versatile Gene-based Association Study-2; WES, Whole-exome sequencing.

WES data into a single test statistic revealed 35 genes associated with the outcome. These genes are involved in various biological functions and processes, such as “*Cell growth and death*”, “*Signalling molecules and interaction*”, “*Cancer*”, and “*Infectious disease*”.

Conclusions

This novel data analysis pipeline integrates two independent omics datasets and increases statistical power for identifying genotype-phenotype associations. The analysis of the transcriptomics profile of cell lines treated with busulfan and WES data from HSCT patients allowed us to identify potential genetic contributors to SOS. Our pipeline could be useful for identifying genetic contributors to other rare diseases where limited power renders genome-wide analyses unpromising.

Trial registration

For the clinical dataset: Clinicaltrials.gov: [NCT01257854](https://clinicaltrials.gov/ct2/history/NCT01257854). <https://clinicaltrials.gov/ct2/history/NCT01257854>.

1. Introduction

Sinusoidal obstruction syndrome (SOS) of the liver is a serious, sometimes life-threatening complication of chemotherapies and hematopoietic stem cell transplantation (HSCT) [1]. SOS commonly arises from endothelial damage and hepatic injury which are triggered by the conditioning regimens given prior to HSCT [2]. Known risk factors for SOS are treatment with multiple alkylating antineoplastic agents, younger age at HSCT, underlying disease, previous liver damage, iron overload, HSCT-specific factors such as myeloablative versus reduced-intensity conditioning regimens, donor type, number of previous HSCTs, and concomitant hepatotoxic treatments [3–5]. Particularly the alkylating agent busulfan which is used in a large proportion of patients undergoing HSCT shows a strong and dose-dependent association with SOS impacting survival [6–8]. Mouse models identified the exposure to busulfan as initiating factor leading to endothelial cell injury and later SOS [9]. The further mechanisms involved in SOS are complex. Initial endothelial cell damage caused by the conditioning regimen leads to inflammation resulting in activation of the coagulation cascade leading to micro-thrombi in the liver microvasculature [10,11] and obstruction of the centrilobular vein of the liver [12]. The outflow obstruction in the liver causes additional damage to the hepatic cells and leads to painful hepatomegaly, jaundice, and liver dysfunction. In severe cases, patients may develop multiple organ failure with a mortality rate of over 80% [5]. Defibrotide, an anti-coagulant and anti-inflammatory drug, is an effective treatment if administered early in the course of the disease [13,14]. Also, prophylactic treatment with defibrotide starting simultaneously with the conditioning regimen in high-risk populations is feasible and effective but not without side effects [15]. To target the right patients for prophylaxis, we need to identify the risk factors associated with SOS, including genetic contributors.

Various studies have reported germline genetic variants in association with SOS. Most of these studies have used a candidate gene approach. They identified several genetic determinants of SOS in genes that encode for detoxification enzymes, particularly glutathione S transferases such as *GSTA1* [16–18], enzymes which affect glutathione levels (e.g. cystathionine gamma-lyase [*CTH*]) [19], oxidative liver injury (e.g. methylenetetrahydrofolate reductase [*MTHFR*]) [19,20], the iron metabolism (e.g. homeostatic iron regulator [*HFE*]) [21], and urea

cycle (e.g. carbamoyl phosphate synthetase I [*CPS1*]) [21]. We previously performed an exome-wide association analysis focusing on exonic variants with predicted functional impact and those in adjacent untranslated regions (UTRs) in patients undergoing busulfan-based HSCT. We identified genetic variants in the *UGT2B10* and *LNPK* genes, which were replicated in an independent population [22]. The polymorphisms that were found to be associated with SOS in previous studies were not validated in the exome-wide discovery analysis, which likely reflects the coverage of the exome-specific analysis and selection of functional variants. Currently, our knowledge of the genetic predictors for SOS is still very limited, identified associations were only inconsistently replicated [23], and novel approaches should be developed to estimate the genetic contribution to SOS.

Since a majority of studies addressing complications of HSCT and similar other rare conditions lack the power to identify genetic predictors due to the small sample size, we used a strategy that is built on the strong link between the exposure to busulfan and the onset of SOS [24]. We generated transcriptomic data from lymphoblastoid cell lines (LCLs) exposed to busulfan *in vitro* and evaluated gene expression changes following treatment. We used the raw whole-exome sequencing (WES) data published previously [22]. Additional filtering steps on the WES data were applied to reduce genetic heterogeneity in the sample, prior to conducting the association analysis. We performed the association analysis with SOS at the SNP and gene level. Importantly, we combined the results of the association analysis and the transcriptomic analysis into a test statistic which measures the strength of association with the phenotype of interest, at the gene level.

2. Materials and methods

2.1. Design of the study

We developed a data analysis pipeline which includes the following steps:

First, we performed a differential gene expression analysis of LCLs before and after *in vitro* exposure to 100 μ M busulfan (Fig 1). Second, we performed an association analysis with SOS at the SNP level, using the raw whole-exome sequencing (WES) data published previously

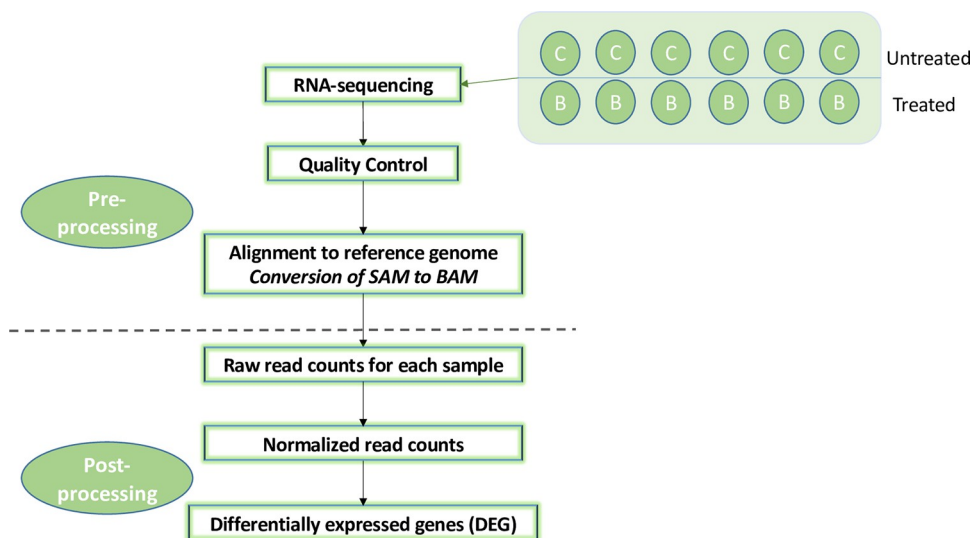


Fig 1. Flowchart of the transcriptomic data analysis pipeline in 6 lymphoblastic cell lines with and without *in vitro* busulfan treatment. Legend: BAM, Binary Alignment Map; DEG, differentially expressed genes; RNA, ribonucleic acid; SAM, Sequence Alignment Map; WES, whole-exome sequencing.

<https://doi.org/10.1371/journal.pone.0281892.g001>

[22], after the application of a series of filtering steps to reduce genetic heterogeneity in the data [25]. Third, we used a gene-based test for association, which allows us to combine the effects of all SNPs in a gene into one test statistic while correcting for linkage disequilibrium (VEGAS2) [26]. A gene-based approach allows achieving greater power for identifying relevant genes by (i) detecting significant associations through the combination of several SNPs showing marginal levels of significance that are often indistinguishable from random noise in the initial WES results [26], and (ii) reducing the multiple-testing problem of WES analysis by considering statistical tests for approximately 20,000 genes per genome as opposed to testing hundreds of thousands of SNPs in a typical exome-wide analysis [26]. Then, we computed a gene-based score, and its associated p-value, by combining the results of the differential gene expression analysis and the gene-based test for association. By integrating the expression and association evidence at the gene level, we increased the power to detect genetic determinants of the clinical outcome of interest, since (i) the resolution of our approach should be higher than a classical exome-wide analysis, as we combine evidence from independent experiments, and (ii) the resolution of our approach should be higher than a candidate gene study, as we don't focus exclusively on a few candidates, selected a priori. This increase in power relies on the following assumption: genetic variants which predispose to SOS are found preferentially in genes which expression level is affected by the treatment. Fourth, we performed a functional analysis, which consists of identifying pathways and gene sets that are over-represented in the list of significantly associated genes. Fifth, we estimated polygenic risk scores (PRS) for the cases and the controls for which the exome data was available [22]. These scores were computed as the sum of the number of risk alleles carried at each variant, weighted by their effect size (Fig 2).

2.2 Transcriptomic analysis of lymphoblastoid cell lines

We collected RNA sequencing data on LCLs (Coriell Cell Repository, Camden, NJ, USA) before and after exposure to busulfan. Briefly, we performed true-seq RNA-sequencing of mRNA extracted from six LCLs (GM7056, GM12239, GM12762, GM12057, GM12489, and

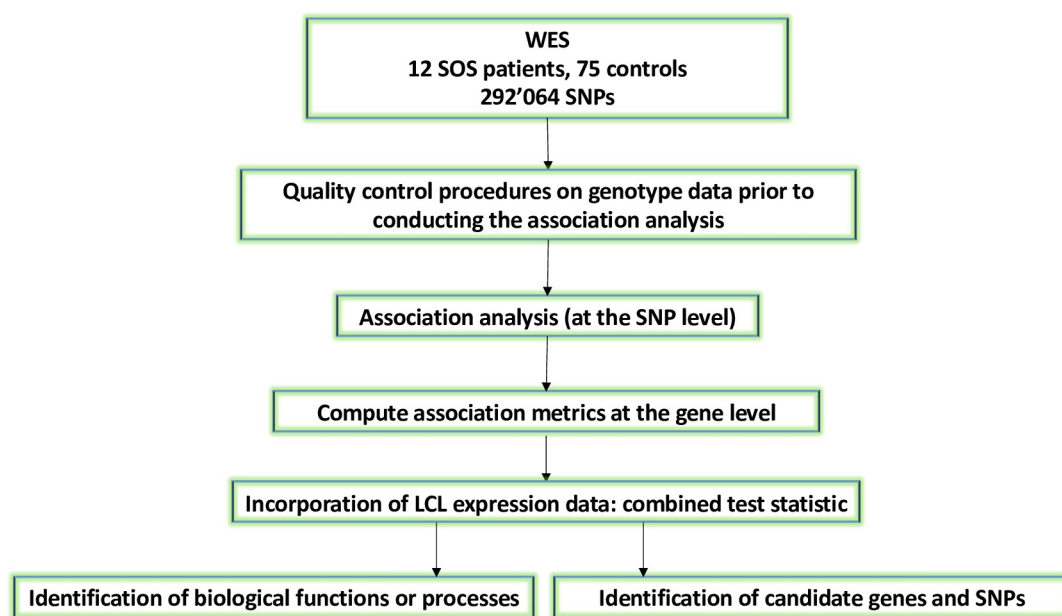


Fig 2. Flowchart of the data analysis pipeline combining whole-exome sequencing of predominantly pediatric patients with *in vitro* transcriptomic data. Legend: LCL, lymphoblastoid cell line; SNP, single nucleotide polymorphism; SOS, sinusoidal obstruction syndrome; WES, whole-exome sequencing.

<https://doi.org/10.1371/journal.pone.0281892.g002>

GM12546) pre and post 48h of 100 μ M busulfan treatment. The chosen concentration of busulfan represents 80% of cell viability in our set of LCLs with an average 50% inhibitory concentration of 415,1 \pm 203,3 μ M. To avoid having non-genetic cell-to-cell variability in gene expression which could be introduced by the culturing condition, we fixed a limit on the maximum number of cell culture medium passages of 15 times.

Library construction and sequencing were performed at the iGE3 Genomics Platform—CMU, Geneva, Switzerland, using the Illumina HiSeq 2000 library preparation kit. The differential expression analysis was performed with DESeq2 [27,28] following a standard data analysis pipeline (**Supporting Information: Text and Figures; Text S01 in S1 File**). The p-values of the test for differential expression were adjusted for multiple testing using the procedure of Benjamini and Hochberg [32].

2.3. Data filtering on WES genotype data

We used the unfiltered whole-exome sequencing data, of which analyses were published previously [22]. Briefly, 87 patients were included, of which 12 developed the main outcome SOS as defined by the modified Seattle criteria [3] (**S1 Table**). Prior to conducting the association analysis, we applied a series of filtering steps on genotype data to reduce genetic heterogeneity as previously described [25], using PLINK version 1.9 [29]. (**Supporting Information: Text and Figures; Text S02 in S1 File**). Written informed consent was obtained from every patient or parent/legal guardian by the local research team, who participated in the study and provided genetic material and clinical data. The study was conducted in accordance with the Declaration of Helsinki. The Institutional Review Board of the CHU Sainte-Justine, Montreal, Canada, approved the study and all patients/parents provided an informed consent form (IRB number: 2450, trial registration: Clinicaltrials.gov: NCT01257854).

2.4. Tests of association of WES data with clinical outcome

Our clinical outcome SOS was defined using the modified Seattle criteria, as outlined in the description of the clinical dataset [22]. The test for association between the SNPs and the binary outcome, i.e. SOS, was performed with PLINK version 1.9, using both the standard chi-squared test and the logistic regression model implemented in the software. We included the multidimensional scaling components as covariates in the logistic regression analysis, which have been estimated previously to control for population stratification.

We used the Versatile Gene-based Association Study-2 (VEGAS2) software [26] to combine the effects of all SNPs in a single gene into one test statistic (i.e. a chi-squared statistic) and its associated p-value. The tool accounts for linkage disequilibrium (LD) and gene size (number of SNPs per gene) [26].

2.5. Combined test of transcriptomic and WES data

We computed a gene-based score, and its associated p-value, by combining the results of the differential gene expression analysis (Z_{exp}) and the gene-based test for association (Z_{assoc}), using the weighted Z-method [30]. We converted the p-values of the 2 tests into z-scores and computed the Z_s -combined test statistic, using the following formula:

$$Z_s = \frac{Z_{exp} + Z_{assoc}}{\sqrt{2}} \sim N(0, 1)$$

To assign equal weights to the 2 experiments into the combined statistic, we rescaled the p-values derived from the LCL expression analysis to have the same range as the p-values

calculated from the WES data, before converting them to Z-scores. We used Bonferroni correction for multiple testing.

2.6. Over-representation analysis

The functional characterization of the genes that show a significant difference in expression in LCLs after treatment with busulfan was performed using an over-representation analysis (ORA). We used the Kyoto Encyclopedia of Genes and Genomes (KEGG) [31] release 97 and the Reactome [32] version 75 databases as a collection of annotated gene sets for ORA. We used the same strategy to characterize the genes that were associated with a significant combined test statistic and outcome, after Bonferroni correction for multiple testing.

2.7. Polygenic risk scores

We computed polygenic risk scores (PRSs), for the cases and controls, as the sum of the number of risk alleles carried at each locus ($\sum nb_{effect\ allele}$), weighted by their effect size [33]. The latter was estimated, using PLINK version 1.9, as the logarithm of the odds ratio ($\ln(OR)$).

$$PRS = \frac{\sum nb_{effect\ allele} * \ln(OR)}{nb_{loci}}$$

We used for the estimations of the PRSs the SNPs with the p-value of the association test using the logistic regression model < 0.05 , and located within the boundaries of the genes (including the UTRs) for which the combined test was significant after a Bonferroni correction for multiple testing.

3. Results

3.1. Transcriptomic analysis of lymphoblastoid cell lines

We identified 3093 genes that showed significant differential expression in the LCLs after *in vitro* exposure to busulfan. Among these genes, 1708 genes (55%) were found to be upregulated and 1385 (45%) downregulated in the treated LCLs (S2 and S3 Tables).

Among the gene sets that were enriched in up-regulated genes, we found pathways involved in “Cell growth and death” (p53 signaling pathway, Apoptosis, Necroptosis; Regulation by c-FLIP, Caspase activation via Death Receptors in the presence of ligand, RIPK1-mediated regulated necrosis), “immune system” and “immune disease” (Graft-versus-host disease, NOD-like receptor signaling pathway, Allograft rejection; Interferon alpha/ beta/ gamma, Interleukin-2 / -4/ -10/ -13/ -35 signaling), “Signal transduction” (TNF, NF-kappa B, Phosphatidylinositol, FoxO, JAK-STAT) and “Signaling molecules and interaction” (Cell adhesion molecules, Cytokine-cytokine receptor interaction, ECM-receptor interaction; Signaling by KIT in disease, PECAM1 interactions, RAF-independent MAPK1/3 activation, Activated NOTCH1 Transmits Signal to the Nucleus, Signaling by VEGF/ PDGF).

Among the down-regulated gene sets, we identified pathways involved in “Cell growth and death” (cell cycle, cell cycle checkpoints), “Replication and Repair” (DNA replication, DNA repair, base excision repair, mismatch repair, nucleotide excision repair, chromosome maintenance, cellular responses to stress, activation of ATR in response to replication stress).

3.2. Data filtering on WES genotype data

We used the unfiltered whole-exome sequencing data from our previous publication [19] which contains genotype information for 87 individuals (12 SOS cases and 75 controls). We identified 292,064 bi-allelic SNPs with $MAF \geq 0.05$. We filtered out SNPs and individuals in the

following series of steps: (1) we excluded 21,138 SNPs with a proportion of missing genotypes exceeding 20%; no individuals had a proportion of missing genotypes larger than 20%; (2) 11 individuals (10 women, 1 man) showed inconsistencies in their assigned and genetic sex and were discarded from the dataset; (3) we excluded 4,898 SNPs with $MAF < 0.05$ in the remaining sample; (4) we excluded 8,744 markers which deviated from Hardy–Weinberg equilibrium; (5) we discarded 1 individual from Sub-Saharan African ancestry which showed a heterozygosity rate below the chosen threshold; (6) we excluded 4 individuals that showed cryptic relatedness; (7) we removed 14 ethnic outliers, from Asia and Africa. After these filtering steps, our dataset contained genotype information for 57 individuals (11 cases and 46 controls): 1 individual of American origin, 6 individuals from North Africa and 50 individuals of European ancestry) and 252,817 autosomal SNPs.

3.3. Association tests in the clinical WES dataset

The analysis at the SNP level using a standard chi-squared test identified 60 variants associated with SOS after Benjamini and Hochberg (1995) step-up false discovery rate (FDR) control [34], and 7 variants after Bonferroni adjustment (S4 Table) [35]. Using the logistic regression model, no SNPs reached significance after any of the above adjustments for multiple testing.

At the gene-level, we identified 12 genes that reached significance after Benjamini and Hochberg FDR control (*NACAP1*, *E2F4*, *SHCBP1*, *MIR34C*, *PF4V1*, *CYP7B1*, *TMEM208*, *CXCL6*, *RAB3B*, *LOC101928744*, *LOC101928773*, *CFHR3*), and 4 genes after Bonferroni correction (*NACAP1*, *E2F4*, *SHCBP1*, *MIR34C*) (S5 Table).

3.4. Combined test, Over-representation analysis and polygenic risk score

We identified 35 genes that were found to be significant for the test which combined the results of the differential gene expression analysis in LCLs and the gene-based test for an association in the WES dataset after Bonferroni correction for multiple testing (Fig 3; S6 Table). There was no overlap between the genes being significant for the combined test and the genes identified from the WES analysis only. But all genes that were significant for the combined test were found to be differentially expressed before and after exposure to busulfan (Supporting Information: Text and Figures: Figure S01 and S02 in S1 File).

The functional characterization of these genes using the Reactome database identified 14 significantly enriched pathways (S7 Table). These pathways are mostly involved in “apoptosis and regulated necrosis” (dimerization of procaspase-8, regulation by c-FLIP, inhibition of CASP8 activity, regulation of necroptotic cell death, TP53-regulated transcription of death genes, receptors and ligands, ligand-dependent caspase activation, RIPK1-mediated regulated necrosis, caspase activation via extrinsic apoptotic signalling pathway). Among the enriched gene sets, we found a pathway implicated in “cell proliferation and differentiation” (i.e. transcriptional regulation by RUNX3). “Cell surface proteins” important in cell interaction (integrin cell surface interactions, non-integrin membrane-ECM interactions, extracellular matrix organization).

The KEGG analysis identified 16 enriched pathways. These pathways were involved in “cell growth and death” (p53 signaling pathway, cellular senescence), “signaling molecules and interaction” (cytokine-cytokine receptor interaction), cancer, and infection (Fig 4).

The polygenic risk score computed using 209 SNPs from the 35 genes identified through combined expression and association analyses at the gene level showed a significantly different distribution between the cases and the controls ($p\text{-value} = 8.506\text{e-}07$; Supplementary Figure S02 in S1 File).

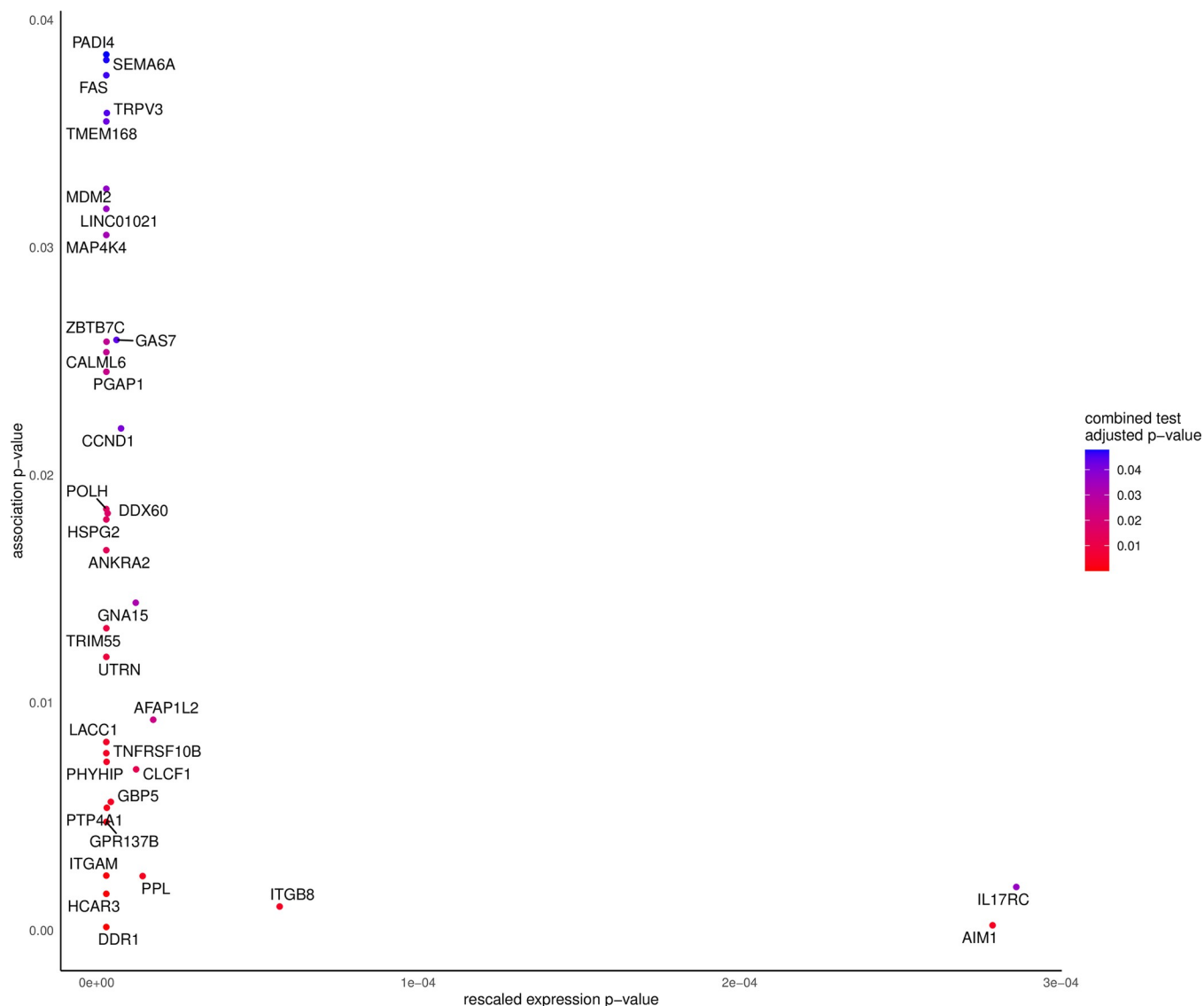


Fig 3. Graphical representation of the relationship between the p-values of the combined test (color coded), the lymphoblastoid cell line expression test (x-axis), and the whole-exome sequencing association test (y axis): We see that the majority of the genes that are associated with a significant combined test statistic show very low p values for the expression test and that the p-value of the combined test increases with the p-value of the association test.

<https://doi.org/10.1371/journal.pone.0281892.g003>

4. Discussion

In this study, we identified new potential genetic determinants of SOS at the SNP, gene, and pathway levels using a new multi-omics data analysis pipeline integrating data from two different experiments. We analyzed *in vitro* transcriptomics data of LCLs after treatment with busulfan and WES data of predominantly pediatric HSCT patients with and without SOS. We combined the results of these analyses into a test statistic which measures the strength of association to SOS at the gene level. Our pipeline addresses the lack of power of small datasets by prioritizing genes from a differential gene expression analysis in LCLs, and the existence of genetic confounders by using filtering steps to increase genetic homogeneity in WES data. Our pipeline allowed the identification of new potential genetic determinants of SOS that can be validated in a follow-up study using a replication cohort.

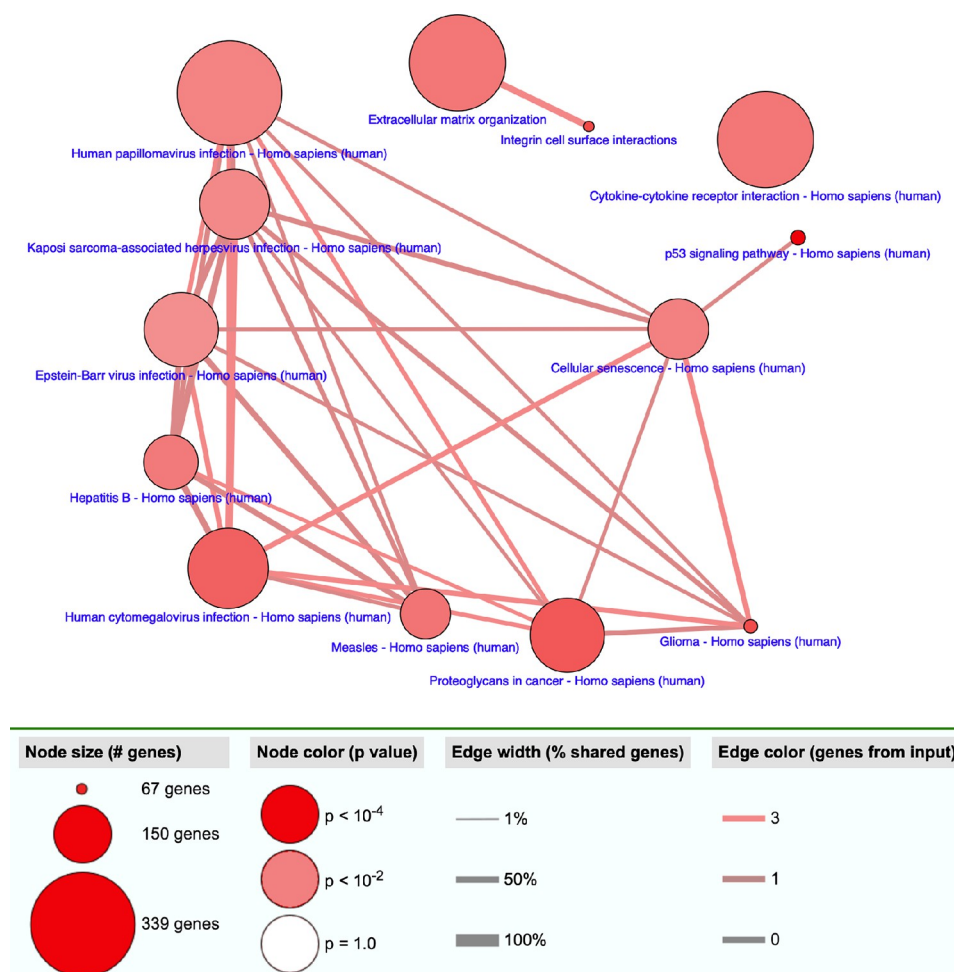


Fig 4. Over representation analysis using CPDB.org incorporating 35 genes associated with sinusoidal obstruction syndrome identified through a combined test statistic.

<https://doi.org/10.1371/journal.pone.0281892.g004>

In the LCLs treated with busulfan, we found several upregulated pathways, which are mainly involved in the inflammatory response, signaling by interleukins and interferon-gamma, *TP53* signaling, and apoptosis. Interleukins-1 and -2, tumor necrosis factor-alpha, and transforming growth factor-beta were identified previously to be elevated in patients with SOS [36,37] or after busulfan exposure *in vitro* in endothelial cells [38]. The expression change we measured is in agreement with these findings, supporting the assumption that LCLs might be a valid cell model for our study. Among the downregulated pathways, we identified cell cycle and mitochondrial translation. Previous *in vitro* analyses have demonstrated a cell cycle arrest after busulfan exposure which supports our findings [39].

We identified new genetic markers in significant association with SOS, using the WES data, after the application of filtering steps to increase the homogeneity of the genetic dataset. Our data workflow reduced the number of individuals from the unfiltered WES dataset by 34.5% and the number of SNPs by 12.4%. Our pipeline thus generated a more homogenous sample at the cost of a smaller sample size. This workflow allowed us to identify through the analyses at the gene level 12 genes in significant association with SOS. None of these genes was identified in the previously published study, where a different strategy was used with a focus on single SNPs and functional variants [22].

Existing data suggest that the association of the currently identified genes with SOS could be meaningful. *NACAP1* is a pseudogene which was recently associated with pre-eclampsia, a disorder of pregnant women with arterial hypertension, neurological symptoms, and fetal growth restriction [40]. This disease is understood as being mediated by endothelial cell dysfunction [41], which is also a major driver in SOS. *PF4V1* is involved in platelet disorders and clot formation [42,43]; interaction with the coagulation system was long suggested to play a role in SOS [44]. *SHCBP1* is involved in various cell signalling processes, cell proliferation, and upregulated in various neoplasms [45]. *CXCL6* is a chemokine with angiogenic properties and is involved in TNF signaling pathway. The *CYP7B1* gene from the cytochrome P450 gene family is involved in cholesterol and lipid metabolism and is associated with neonatal liver disease [46]. Additional genes identified here are associated with apoptosis (*MIR34C*, *E2F4*) [47,48] and autophagy (*TMEM208*) [49]. Among those genes, only *SHCBP1* was significantly downregulated in LCLs after treatment with busulfan, while *CXCL6*, *NACAP1* and *PF4V1* were not expressed in LCLs.

The combined test statistic allowed to integrate the expression results and the association metrics into a single statistic measure. We identified 35 genes that were found to be associated with SOS after correction for multiple testing. None of these genes was previously investigated for an association with SOS. The functional analysis of these genes, using the KEGG database revealed that many are involved in the *TP53* signaling pathway and cellular senescence (*CCND1*, *FAS*, *MDM2*, *TNFRSF10B*, *CALML6*). This result can be explained by the response of the cells to the toxic effect of busulfan and its metabolites which might also play a role in SOS [50]. In addition, *TP53* regulated genes were significantly upregulated in our gene expression analysis such as *MDM2*, and the pro-apoptotic genes *BBC3* and *GADD45A*. *CALML6* plays an important role in the regulation of the innate immune response through the NF- κ B signaling pathway. Previous reports have shown a similar signal of cell cycle gene dysregulation after busulfan exposure [38]. *TNFRSF10B* is a member of the TNF-receptor superfamily, that is activated by TNF-related apoptosis-inducing ligand (*TNFSF10*), and transduces an apoptosis signal [51]. Additionally, the cytokine-cytokine receptor interaction gene set could be explained by inflammatory processes.

The functional analysis using the Reactome database identified pathways mostly involved in programmed cell death. Additionally, gene sets linked to the integrin cell surface interactions were found (*ITGB8*, *HSPG2*, *ITGAM*). Cellular responses to external stimuli are mainly a stress response pathway and regulate cell adaptations when exposed to reactive oxygen species, heat, or other stimuli. Integrins are the receptors that interact with the extracellular matrix and are important for cell adhesion. Two upregulated genes after the treatment with busulfan in this gene set, *ITGAM* and *ITGB8*, are also involved in the complement and coagulation pathway [52], a relevant mechanism involved in SOS [44,53–55], and endothelial cell function [38]. Other genes of interest but not identified in the ORA analysis were *GBP5* which is expressed in endothelial cells and promotes the inflammasome [56], *DDR1* which is associated with endothelial cell senescence [57], *HSPG2* that regulates vascular response to injury and may inhibit thrombosis [58], *LACC1* and *HCAR3* which are proteins from fatty-acid oxidation, lipogenesis, and lipid metabolism, while *POLH* is implicated in DNA damage repair [59].

We estimated polygenic risk scores for the cases and the controls that were included in this study, using the SNPs that were found within the boundaries of the 35 genes that were associated with a significant combined test statistic. The difference in the distributions of the PRSs between the cases and the controls is highly significant. Classically, PRSs are estimated in another cohort than the one that was used to estimate the association metrics with the phenotype of interest. Here, our goal was to estimate the magnitude of differences in scores between our cases and our controls. The difference in scores between cases and controls was highly

significant and underscores the notion that the SNPs we used to estimate those scores are good candidates for further validation. In a future study, selecting those SNPs that contributed most to the PRS should be tested or some of the SNPs could be further prioritized, choosing some specific functional categories, i.e. non-synonymous variants.

The pipeline we developed and used in this study allowed us to identify genes and pathways in association with a rare clinical outcome with a limited number of patient samples, where machine learning-based approaches are inapplicable [60]. Classical approaches are based on candidate gene studies that rely on specific genes which are selected based on previous knowledge. Some of these candidate-gene studies have led to relevant findings with an impact on clinical practice (e.g. *TPMT* polymorphism-associated mercaptopurine intolerance in leukemia) [61]. But often, candidate-gene studies suffered from inconsistent replication efforts [62,63] such as was witnessed for SOS [23]. In contrast to a classical candidate gene approach, our method does not rely on the selection of genes a priori. By combining two independent tests, we enable the identification of novel associations that were not previously investigated or hypothesized.

There are some limitations to our approach that are associated with the assumptions that are made in the setup of this study. First, the hypothesis that the genes that are dysregulated in LCLs after busulfan exposure are more likely to carry variants which are modifiers of busulfan toxicity might not be true. The importance of transcriptomic changes after drug exposure was shown previously to correlate with adverse events after drug exposure [64]. Based on the data on busulfan exposure as initiating factor of SOS, we hypothesized that transcriptomic changes after treatment with busulfan play a pivotal role in the development of this complication. The disadvantage is that we focus with this approach on an upstream event and not on the outcome directly. Second, the magnitude of dysregulation might not be linked to the outcome of interest. Third, LCLs have been extensively used in pharmacogenomic research [65], but some genes which might be relevant for SOS are not expressed in LCLs, which limits their applicability to model endothelial or liver cells. This inherent weakness of our model could be addressed in a future study by choosing cell types that reflect more closely the effector cells of the target disease. Conversely, peripheral blood cells such as the used lymphoid cells might harbor the advantage of being more easily accessible in patients and transcriptomic data could be measured as a potential biomarker of SOS in future studies as previously illustrated [66]. Fourth, to further assess the impact of the identified genes, additional analyses will be needed such as inhibition or overexpression analyses. As this manuscript covers the methodological aspects of combining *in vitro* data with clinical information and due to the complexity of mechanisms leading to SOS, we will perform these analyses in a future study. Fifth, we did not include clinical confounders (such as patient age, conditioning regimen, underlying disease among others) in our pipeline due to the low number of patients with SOS in our sample. We will include these confounders in a future replication analysis.

In this paper, we proposed a novel approach to unravel the genetic determinants of SOS by combining *in vitro* gene expression information after treatment exposure with clinical genomic data into a multi-omics model. Our approach could be used in other research contexts to increase the power to identify the genetic contribution to other rare diseases.

Supporting information

S1 Table. Clinical data of 87 patients included in this study (adapted from Ansari et al. BBMT, 2020, with permission).
(PDF)

S2 Table. Differential gene expression analysis in lymphoblastoid cell lines; sorted by expression change.

(PDF)

S3 Table. Over-representation analysis of differentially expressed genes in lymphoblastoid cell lines, sorted by strength of association.

(PDF)

S4 Table. List of SNPs that are found to be in significant association with VOD, after correction for multiple testing. The SNPs are sorted by chromosome and position. Functional annotations of the variants were retrieved from ENSEMBL (Genome assembly: GRCh37.p13). The r2 column shows the r2 measure of linkage disequilibrium between each SNP and the following SNP in the table, in the CEU population from the 1,000 Genomes phase 3 data, as reported in ENSEMBL.

(PDF)

S5 Table. Whole-exome sequencing association analysis of 57 individuals (11 cases with sinusoidal obstruction syndrome and 46 controls), genes with the strongest association, sorted by adjusted association metric.

(PDF)

S6 Table. Combined association analysis of differential gene expression in lymphoblastoid cell lines and whole-exome sequencing in 57 individuals; genes sorted by adjusted association metric.

(PDF)

S7 Table. Over-representation analysis of differentially expressed genes in a combined test statistic integrating differential expression data of lymphoblastoid cell lines after exposure to busulfan and a whole-exome association analysis.

(PDF)

S1 File. Supporting Information: Text and Figures. • **Text S01.** Description of lymphoblastic cell models, RNA sequencing, and post-processing. • **Text S02.** Description of data filtering steps used in the whole-exome sequencing dataset. • **Figure S01.** Venn diagram showing the number of genes identified in (a) the whole-exome sequencing analysis using the VEGAS2 tool (VEGAS2), (b) the differential gene expression (DGE) from the in vitro analysis before and after treatment of lymphoblastoid cell lines with busulfan, and (c) the combined test aggregating data from both previously mentioned tests into one test statistic. • **Figure S02.** Polygenic risk score including 209 SNPs as output of the data analysis pipeline combining LCL expression and WES clinical association data ($MAF > 0.05$, $p < 0.05$), stratified by presence (cases) or absence (controls) of sinusoidal obstruction syndrome.

(DOCX)

Acknowledgments

The authors thank all patients and their parents for their participation in the clinical study, as well as all study collaborators for their valuable contribution.

Author Contributions

Conceptualization: Nicolas Waespe, Simona Jurkovic Mlakar, Isabelle Dupanloup, Tiago Nava, Marc Ansari.

Data curation: Simona Jurkovic Mlakar, Isabelle Dupanloup, Maja Krajinovic.

Formal analysis: Nicolas Waespe, Simona Jurkovic Mlakar, Isabelle Dupanloup.

Funding acquisition: Claudia E. Kuehni, Marc Ansari.

Investigation: Nicolas Waespe, Simona Jurkovic Mlakar, Isabelle Dupanloup.

Methodology: Nicolas Waespe, Simona Jurkovic Mlakar, Isabelle Dupanloup, Maja Krajinovic.

Project administration: Nicolas Waespe, Marc Ansari.

Resources: Nicolas Waespe, Simona Jurkovic Mlakar, Isabelle Dupanloup, Maja Krajinovic, Claudia E. Kuehni, Marc Ansari.

Software: Isabelle Dupanloup.

Supervision: Marc Ansari.

Validation: Simona Jurkovic Mlakar.

Visualization: Nicolas Waespe, Simona Jurkovic Mlakar, Isabelle Dupanloup.

Writing – original draft: Nicolas Waespe, Simona Jurkovic Mlakar, Isabelle Dupanloup.

Writing – review & editing: Nicolas Waespe, Simona Jurkovic Mlakar, Isabelle Dupanloup, Mohamed Aziz Rezgui, Henrique Bittencourt, Maja Krajinovic, Claudia E. Kuehni, Tiago Nava, Marc Ansari.

References

1. Mahadeo KM, Bajwa R, Abdel-Azim H, Lehmann LE, Duncan C, Zantek N, et al. Diagnosis, grading, and treatment recommendations for children, adolescents, and young adults with sinusoidal obstructive syndrome: an international expert position statement. *Lancet Haematol*. 2020; 7: e61–e72. [https://doi.org/10.1016/S2352-3026\(19\)30201-7](https://doi.org/10.1016/S2352-3026(19)30201-7) PMID: 31818728
2. Fan CQ, Crawford JM. Sinusoidal Obstruction Syndrome (Hepatic Veno-Occlusive Disease). *J Clin Exp Hepatol*. 2014; 4: 332–346. <https://doi.org/10.1016/j.jceh.2014.10.002> PMID: 25755580
3. McDonald GB, Hinds MS, Fisher LD, Schoch HG, Wolford JL, Banaji M, et al. Veno-occlusive disease of the liver and multiorgan failure after bone marrow transplantation: a cohort study of 355 patients. *Ann Intern Med*. 1993; 118: 255–267. <https://doi.org/10.7326/0003-4819-118-4-199302150-00003> PMID: 8420443
4. Corbacioglu S, Carreras E, Ansari M, Balduzzi A, Cesaro S, Dalle J-H, et al. Diagnosis and severity criteria for sinusoidal obstruction syndrome/veno-occlusive disease in pediatric patients: a new classification from the European society for blood and marrow transplantation. *Bone Marrow Transplant*. 2018; 53: 138–145. <https://doi.org/10.1038/bmt.2017.161> PMID: 28759025
5. Dalle J-H, Giralot SA. Hepatic Veno-Occlusive Disease after Hematopoietic Stem Cell Transplantation: Risk Factors and Stratification, Prophylaxis, and Treatment. *Biology of Blood and Marrow Transplantation*. 2016; 22: 400–409. <https://doi.org/10.1016/j.bbmt.2015.09.024> PMID: 26431626
6. Feng X, Wu Y, Zhang J, Li J, Zhu G, Fan D, et al. Busulfan systemic exposure and its relationship with efficacy and safety in hematopoietic stem cell transplantation in children: a meta-analysis. *BMC Pediatr*. 2020; 20: 176. <https://doi.org/10.1186/s12887-020-02028-6> PMID: 32312247
7. Bognar T, Bartelink IH, Egberts TCG, Rademaker CMA, Versluys AB, Slatter MA, et al. Association Between the Magnitude of Intravenous Busulfan Exposure and Development of Hepatic Veno-Occlusive Disease in Children and Young Adults Undergoing Myeloablative Allogeneic Hematopoietic Cell Transplantation. *Transplantation and Cellular Therapy*. 2022; 28: 196–202. <https://doi.org/10.1016/j.jtct.2022.01.013> PMID: 35065280
8. Bartelink IH, Lalmohamed A, Reij EML van, Dvorak CC, Savic RM, Zwaveling J, et al. Association of busulfan exposure with survival and toxicity after haemopoietic cell transplantation in children and young adults: a multicentre, retrospective cohort analysis. *The Lancet Haematology*. 2016; 3: e526–e536. [https://doi.org/10.1016/S2352-3026\(16\)30114-4](https://doi.org/10.1016/S2352-3026(16)30114-4) PMID: 27746112

9. Zeng L, Jia L, Xu S, Yan Z, Ding S, Xu K. Vascular endothelium changes after conditioning in hematopoietic stem cell transplantation: role of cyclophosphamide and busulfan. *Transplant Proc.* 2010; 42: 2720–2724. <https://doi.org/10.1016/j.transproceed.2010.04.024> PMID: 20832576
10. Park YD, Yasui M, Yoshimoto T, Chayama K, Shimono T, Okamura T, et al. Changes in hemostatic parameters in hepatic veno-occlusive disease following bone marrow transplantation. *Bone Marrow Transplant.* 1997; 19: 915–920. <https://doi.org/10.1038/sj.bmt.1700760> PMID: 9156266
11. Tanikawa S, Mori S, Ohhashi K, Akiyama H, Sasaki T, Kaku H, et al. Predictive markers for hepatic veno-occlusive disease after hematopoietic stem cell transplantation in adults: a prospective single-center study. *Bone Marrow Transplant.* 2000; 26: 881–886. <https://doi.org/10.1038/sj.bmt.1702624> PMID: 11081389
12. Carreras E. Veno-occlusive disease of the liver after hemopoietic cell transplantation. *Eur J Haematol.* 2000; 64: 281–291. <https://doi.org/10.1034/j.1600-0609.2000.9r200.x> PMID: 10863974
13. Richardson P, Aggarwal S, Topaloglu O, Villa KF, Corbacioglu S. Systematic review of defibrotide studies in the treatment of veno-occlusive disease/sinusoidal obstruction syndrome (VOD/SOS). *Bone Marrow Transplant.* 2019; 54: 1951–1962. <https://doi.org/10.1038/s41409-019-0474-8> PMID: 30804485
14. Richardson PG, Smith AR, Kernan NA, Lehmann L, Soiffer RJ, Ryan RJ, et al. Pooled analysis of Day 100 survival for defibrotide-treated patients with hepatic veno-occlusive disease/sinusoidal obstruction syndrome and ventilator or dialysis dependence following hematopoietic cell transplantation. *Br J Haematol.* 2020; 190: 583–587. <https://doi.org/10.1111/bjh.16552> PMID: 32157682
15. Corbacioglu S, Cesaro S, Faraci M, Valteau-Couanet D, Gruhn B, Rovelli A, et al. Defibrotide for prophylaxis of hepatic veno-occlusive disease in paediatric haemopoietic stem-cell transplantation: an open-label, phase 3, randomised controlled trial. *Lancet.* 2012; 379: 1301–1309. [https://doi.org/10.1016/S0140-6736\(11\)61938-7](https://doi.org/10.1016/S0140-6736(11)61938-7) PMID: 22364685
16. Ansari M, Curtis PH-D, Uppugunduri CRS, Rezgui MA, Nava T, Mlakar V, et al. GSTA1 diplotypes affect busulfan clearance and toxicity in children undergoing allogeneic hematopoietic stem cell transplantation: a multicenter study. *Oncotarget.* 2017; 8: 90852–90867. <https://doi.org/10.18632/oncotarget.20310> PMID: 29207608
17. Srivastava A, Poonkuzhali B, Shaji RV, George B, Mathews V, Chandy M, et al. Glutathione S-transferase M1 polymorphism: a risk factor for hepatic venoocclusive disease in bone marrow transplantation. *Blood.* 2004; 104: 1574–1577. <https://doi.org/10.1182/blood-2003-11-3778> PMID: 15142875
18. Ansari M, Rezgui MA, Théoret Y, Uppugunduri CRS, Mezziani S, Vachon M-F, et al. Glutathione S-transferase gene variations influence BU pharmacokinetics and outcome of hematopoietic SCT in pediatric patients. *Bone Marrow Transplant.* 2013; 48: 939–946. <https://doi.org/10.1038/bmt.2012.265> PMID: 23292236
19. Huezo-Diaz Curtis P, Uppugunduri CRS, Muthukumaran J, Rezgui MA, Peters C, Bader P, et al. Association of CTH variant with sinusoidal obstruction syndrome in children receiving intravenous busulfan and cyclophosphamide before hematopoietic stem cell transplantation. *Pharmacogenomics J.* 2016. <https://doi.org/10.1038/tpj.2016.65> PMID: 27779248
20. Efrati E, Zuckerman T, Ben-Ami E, Krivoy N. MTHFR C677T/A1298C genotype: a possible risk factor for liver sinusoidal obstruction syndrome. *Bone Marrow Transplant.* 2014; 49: 726–727. <https://doi.org/10.1038/bmt.2014.16> PMID: 24583625
21. Kallianpur AR, Hall LD, Yadav M, Byrne DW, Speroff T, Dittus RS, et al. The hemochromatosis C282Y allele: a risk factor for hepatic veno-occlusive disease after hematopoietic stem cell transplantation. *Bone Marrow Transplant.* 2005; 35: 1155–1164. <https://doi.org/10.1038/sj.bmt.1704943> PMID: 15834437
22. Ansari M, Petrykey K, Rezgui MA, Del Vecchio V, Cortyl J, Ralph R-O, et al. Genetic Susceptibility to Hepatic Sinusoidal Obstruction Syndrome in Pediatric Patients Undergoing Hematopoietic Stem Cell Transplantation. *Biol Blood Marrow Transplant.* 2020; 26: 920–927. <https://doi.org/10.1016/j.bbmt.2019.11.026> PMID: 31790828
23. Waespe N, Strebel S, Jurkovic Mlakar S, Krajnovic M, Kuehni CE, Nava T, et al. Genetic Predictors for Sinusoidal Obstruction Syndrome—A Systematic Review. *Journal of Personalized Medicine.* 2021; 11: 347. <https://doi.org/10.3390/jpm11050347> PMID: 33925809
24. McCune JS, Holmberg LA. Busulfan in hematopoietic stem cell transplant setting. *Expert Opin Drug Metab Toxicol.* 2009; 5: 957–969. <https://doi.org/10.1517/17425250903107764> PMID: 19611402
25. Marees AT, Kluiver H de, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research.* 2018; 27: e1608. <https://doi.org/10.1002/mpr.1608> PMID: 29484742
26. Mishra A, Macgregor S. VEGAS2: Software for More Flexible Gene-Based Testing. *Twin Res Hum Genet.* 2015; 18: 86–91. <https://doi.org/10.1017/thg.2014.79> PMID: 25518859

27. Thomas S, Bonchev D. A survey of current software for network analysis in molecular biology. *Hum Genomics*. 2010; 4: 353–360. <https://doi.org/10.1186/1479-7364-4-5-353> PMID: 20650822
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014; 15: 550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
29. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81: 559–575. <https://doi.org/10.1086/519795> PMID: 17701901
30. Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of Evolutionary Biology*. 2005; 18: 1368–1373. <https://doi.org/10.1111/j.1420-9101.2005.00917.x> PMID: 16135132
31. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017; 45: D353–D361. <https://doi.org/10.1093/nar/gkw1092> PMID: 27899662
32. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020; 48: D498–D503. <https://doi.org/10.1093/nar/gkz1031> PMID: 31691815
33. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*. 2020; 12: 44. <https://doi.org/10.1186/s13073-020-00742-5> PMID: 32423490
34. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1995; 57: 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
35. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*. 2001; 29: 1165–1188. <https://doi.org/10.1214/aos/1013699998>
36. Kumar S, Deleve LD, Kamath PS, Tefferi A. Hepatic Veno-occlusive Disease (Sinusoidal Obstruction Syndrome) After Hematopoietic Stem Cell Transplantation. *Mayo Clinic Proceedings*. 2003; 78: 589–598. <https://doi.org/10.4065/78.5.589> PMID: 12744547
37. Schots R, Kaufman L, Van Riet I, Ben Othman T, De Waele M, Van Camp B, et al. Proinflammatory cytokines and their role in the development of major transplant-related complications in the early phase after allogeneic bone marrow transplantation. *Leukemia*. 2003; 17: 1150–1156. <https://doi.org/10.1038/sj.leu.2402946> PMID: 12764383
38. Reimer J, Bien S, Ameling S, Hammer E, Völker U, Hempel G, et al. Antineoplastic agent busulfan regulates a network of genes related to coagulation and fibrinolysis. *Eur J Clin Pharmacol*. 2012; 68: 923–935. <https://doi.org/10.1007/s00228-011-1209-y> PMID: 22286157
39. Ben-Ami E, Krivoy N, Efrati E. Busulfan directly inhibits T cell proliferation through non-anergic cell cycle arrest. *Leuk Lymphoma*. 2014; 55: 932–933. <https://doi.org/10.3109/10428194.2013.810739> PMID: 23822540
40. He J, Liu K, Hou X, Lu J. Identification and validation of key non-coding RNAs and mRNAs using co-expression network analysis in pre-eclampsia. *Medicine (Baltimore)*. 2021; 100: e25294. <https://doi.org/10.1097/MD.00000000000025294> PMID: 33832098
41. Powe CE, Levine RJ, Karumanchi SA. Preeclampsia, a disease of the maternal endothelium: the role of antiangiogenic factors and implications for later cardiovascular disease. *Circulation*. 2011; 123: 2856–2869. <https://doi.org/10.1161/CIRCULATIONAHA.109.853127> PMID: 21690502
42. Sofie Struyf, Burdick Marie D., Proost Paul, Van Damme Jo, Strieter Robert M. Platelets Release CXCL4L1, a Nonallelic Variant of the Chemokine Platelet Factor-4/CXCL4 and Potent Inhibitor of Angiogenesis. *Circulation Research*. 2004; 95: 855–857. <https://doi.org/10.1161/01.RES.0000146674.38319.07> PMID: 15459074
43. Struyf S, Salogni L, Burdick MD, Vandercappellen J, Gouwy M, Noppen S, et al. Angiostatic and chemotactic activities of the CXC chemokine CXCL4L1 (platelet factor-4 variant) are mediated by CXCR3. *Blood*. 2011; 117: 480–488. <https://doi.org/10.1182/blood-2009-11-253591> PMID: 20980681
44. Piccin A, Sartori MT, Bisogno G, Van Schilfgaarde M, Saggiorato G, Pierro AMD, et al. New insights into sinusoidal obstruction syndrome. *Intern Med J*. 2017; 47: 1173–1183. <https://doi.org/10.1111/imj.13550> PMID: 28707749
45. Zhang G-Y, Ma Z-J, Wang L, Sun R-F, Jiang X-Y, Yang X-J, et al. The Role of Shcgp1 in Signaling and Disease. *Curr Cancer Drug Targets*. 2019; 19: 854–862. <https://doi.org/10.2174/1568009619666190620114928> PMID: 31250756
46. Stiles AR, McDonald JG, Bauman DR, Russell DW. CYP7B1: One Cytochrome P450, Two Human Genetic Diseases, and Multiple Physiological Functions. *J Biol Chem*. 2009; 284: 28485–28489. <https://doi.org/10.1074/jbc.R109.042168> PMID: 19687010

47. He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, et al. A microRNA component of the p53 tumour suppressor network. *Nature*. 2007; 447: 1130–1134. <https://doi.org/10.1038/nature05939> PMID: 17554337
48. Leone G, Sears R, Huang E, Rempel R, Nuckolls F, Park CH, et al. Myc requires distinct E2F activities to induce S phase and apoptosis. *Mol Cell*. 2001; 8: 105–113. [https://doi.org/10.1016/s1097-2765\(01\)00275-1](https://doi.org/10.1016/s1097-2765(01)00275-1) PMID: 11511364
49. Zhao Y, Hu J, Miao G, Qu L, Wang Z, Li G, et al. Transmembrane protein 208: a novel ER-localized protein that regulates autophagy and ER stress. *PLoS One*. 2013; 8: e64228. <https://doi.org/10.1371/journal.pone.0064228> PMID: 23691174
50. Wei R, Zhang X, Cai Y, Liu H, Wang B, Zhao X, et al. Busulfan Suppresses Autophagy in Mouse Spermatogonial Progenitor Cells via mTOR of AKT and p53 Signaling Pathways. *Stem Cell Rev Rep*. 2020; 16: 1242–1255. <https://doi.org/10.1007/s12015-020-10027-4> PMID: 32839922
51. Sullivan GP, O'Connor H, Henry CM, Davidovich P, Clancy DM, Albert ML, et al. TRAIL Receptors Serve as Stress-Associated Molecular Patterns to Promote ER-Stress-Induced Inflammation. *Developmental Cell*. 2020; 52: 714–730.e5. <https://doi.org/10.1016/j.devcel.2020.01.031> PMID: 32109381
52. Wen S, Yan W, Wang L. mRNA expression disturbance of complement system related genes in acute arterial thrombotic and paroxysmal atrial fibrillation patients. *Annals of Palliative Medicine*. 2020; 9: 83546–83846. <https://doi.org/10.21037/apm.2020.04.18> PMID: 32389013
53. Mauro M, Saggiorato G, Sartori MT, Gallo G, De Bortoli M, Bonetti E, et al. Venocclusive disease due to chemotherapy for pediatric acute lymphoblastic leukemia is associated with increased levels of plasminogen-activator inhibitor-1. *Pediatr Blood Cancer*. 2018; 65: e26963. <https://doi.org/10.1002/pbc.26963> PMID: 29350496
54. Goldberg SL, Shubert J, Rao AK, Redei I, Klumpp TR, Mangan KF. Treatment of hepatic veno-occlusive disease with low-dose tissue plasminogen activator: impact on coagulation profile. *Bone Marrow Transplant*. 1996; 18: 633–636. PMID: 8879629
55. Ertem M, Akar N. Factor V Leiden mutation as a predisposing factor for veno-occlusive disease following BMT. *Bone Marrow Transplant*. 2000; 25: 1110–1111. <https://doi.org/10.1038/sj.bmt.1702390> PMID: 10828876
56. Tripal P, Bauer M, Naschberger E, Mörtinger T, Hohenadl C, Cornali E, et al. Unique features of different members of the human guanylate-binding protein family. *J Interferon Cytokine Res*. 2007; 27: 44–52. <https://doi.org/10.1089/jir.2007.0086> PMID: 17266443
57. Yan L, Xie X, Niu B-X, Wu M-T, Tong W-Q, He S-Y, et al. Involvement of miR-199a-3p/DDR1 in vascular endothelial cell senescence in diabetes. *European Journal of Pharmacology*. 2021; 908: 174317. <https://doi.org/10.1016/j.ejphar.2021.174317> PMID: 34270989
58. Segev A, Nili N, Strauss BH. The role of perlecan in arterial injury and angiogenesis. *Cardiovascular Research*. 2004; 63: 603–610. <https://doi.org/10.1016/j.cardiores.2004.03.028> PMID: 15306215
59. McIlwraith MJ, Vaisman A, Liu Y, Fanning E, Woodgate R, West SC. Human DNA Polymerase η Promotes DNA Synthesis from Strand Invasion Intermediates of Homologous Recombination. *Molecular Cell*. 2005; 20: 783–792. <https://doi.org/10.1016/j.molcel.2005.10.001> PMID: 16337601
60. MacEachern SJ, Forkert ND. Machine learning for precision medicine. *Genome*. 2021; 64: 416–425. <https://doi.org/10.1139/gen-2020-0131> PMID: 33091314
61. Relling MV, Hancock ML, Rivera GK, Sandlund JT, Ribeiro RC, Krynetski EY, et al. Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. *J Natl Cancer Inst*. 1999; 91: 2001–2008. <https://doi.org/10.1093/jnci/91.23.2001> PMID: 10580024
62. Pasche B, Yi N. Candidate gene association studies: Successes and failures. *Curr Opin Genet Dev*. 2010; 20: 257–261. <https://doi.org/10.1016/j.gde.2010.03.006> PMID: 20417090
63. Sullivan PF. Spurious Genetic Associations. *Biological Psychiatry*. 2007; 61: 1121–1126. <https://doi.org/10.1016/j.biopsych.2006.11.010> PMID: 17346679
64. Piñero J, Gonzalez-Perez A, Guney E, Aguirre-Plans J, Sanz F, Oliva B, et al. Network, Transcriptomic and Genomic Features Differentiate Genes Relevant for Drug Response. *Front Genet*. 2018; 9: 412. <https://doi.org/10.3389/fgene.2018.00412> PMID: 30319692
65. Wheeler HE, Dolan ME. Lymphoblastoid cell lines in pharmacogenomic discovery and clinical translation. *Pharmacogenomics*. 2012; 13: 55–70. <https://doi.org/10.2217/pgs.11.121> PMID: 22176622
66. Liew C-C, Ma J, Tang H-C, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *Journal of Laboratory and Clinical Medicine*. 2006; 147: 126–132. <https://doi.org/10.1016/j.lab.2005.10.005> PMID: 16503242