Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/28461 holds various files of this Leiden University dissertation.

Author: Brink, Marloes Hendrika ten
Title: Individualized therapeutics in allogeneic stem cell transplantation
Issue Date: 2014-09-03

Chapter 4

Exploratory analysis of 1936 SNPs in ADME genes for association with busulfan clearance in adulthematopoietic stem cell recipients

Marloes H. ten Brink

Jesse J. Swen

Stefan Böhringer

Judith A.M. Wessels

Tahar van der Straaten

Erik W.A. Marijt

Peter A. von dem Borne

Juliëtte Zwaveling

Henk-Jan Guchelaar

Pharmacogenetics and Genomics, 23; 675–683 (2013)

ABSTRACT

Busulfan is used in preparative regimens before stem cell transplantation. There is significant interpatient variability in busulfan pharmacokinetics (PK) and exposure is related to outcome. Polymorphisms in genes encoding glutathione-S-transferases have been associated with busulfan PK but only explain a limited portion of the observed variability. The aim of this study is to identify additional genetic variants associated with busulfan PK by interrogating 1,936 variants in 225 genes involved in drug absorption, distribution, metabolism, and excretion (ADME). In an exploratory cohort (n = 65), patients who received busulfan were genotyped with the DMET array. Top SNPs and haplotypes associated with busulfan clearance were validated in an independent validation cohort (n = 78). In the exploratory cohort, seven variants were identified to be associated with busulfan clearance ($p < 0.001$). In the validation cohort, only *GSTA5* (rs4715354 and rs7746993) remained significantly associated with busulfan clearance ($p = 0.025$). This is the first study using an exploratory pharmacogenetic approach to explain the interindividual variability in busulfan PK. The role of glutathione-Stransferases was confirmed, but no additional genetic markers involved in drug ADME appear to be associated with busulfan PK.

INTRODUCTION

Busulfan is an alkylating agent used widely in high doses in conditioning regimens beforeallogenic hematopoietic stem cell transplantation (allo-HSCT) in adult and pediatric patients. Interpatient variability is significant in busulfan pharmacokinetics (PK); even when busulfan is administered intravenously and body weight adjusted dosing is applied, interpatient variability in exposure or clearance remains between 20 and 30%.[1,2] Furthermore, busulfan exposure is related to outcome. High busulfan exposure appears to result in a higher incidence of toxicities such as sinusoidal obstruction syndrome and mucositis, whereas low busulfan exposure apparently leads to higher rates of graft failure and rejection.[3–5] Therefore, often, therapeutic drug monitoring is applied to individualize busulfan therapy and optimize clinical outcome. Many efforts have been made to develop a population PK model to further characterize busulfan PK.[6,7] The facilities needed for applying therapeutic drug monitoring in a clinical setting are, to date, a barrier to universal acceptance. Furthermore, up-front-dose adaption is not possible. In addition to targeting busulfan doses on the basis of PK-models, pharmacogenetic profiling of patients could play a more important role.

Multiple studies have been carried out to search for a pharmacogenetic explanation for the high variability in busulfan PK. To date, mainly polymorphisms in genes encoding for glutathione-S-transferases (GSTs) have been studied.[8–12] Busulfan is metabolized by conjugation to glutathione. This reaction is catalyzed by different GSTs enzymes, of which GSTA1 is the predominant enzyme responsible for 46% of busulfan conjugation.8 After conjugation, the complex is oxidized in the liver and metabolites are mainly excreted in the urine. In several studies, a single nucleotide polymorphism (SNP) in GSTA1 (rs3957357) was shown to be related to busulfan PK in adult patients receiving busulfan before their HSCT.[9,10] We confirmed this in an independent cohort of adult patients.[11] In this study, three SNPs in three different GST genes were tested for association with busulfan clearance. Only rs3957357 in *GSTA1* was found to be associated with busulfan clearance; however, this SNP could only explain 14% of its variability. Most of the available studies investigating variability in busulfan PK have applied the candidate gene method. Although this method has proven to be successful in the past, its major limitation is that it cannot identify genetic markers in genes not previously associated with busulfan PK. We hypothesized that other absorption, distribution, metabolism, and excretion (ADME) genes involved in drug metabolism or transport of busulfan could add to the variability of PK, meaning that a set of genetic variants may be more important to explain interindividual differences in busulfan exposure than a single gene. Genetic profiling with the Affymetrix drug metabolizing enzymes and transporters genotyping array offers the ability to determine 1,936 variants in 225 genes involved in drug metabolism and disposition.[12] In this

report, we present the results of the exploration of genetic factors associated with busulfan PK using this genotyping array.

PATIENTS AND METHODS

Patient and disease characteristic

Patients receiving allo-HSCT between 2004 and 2011 in the Leiden University Medical Center were included on the basis of the following selection criteria: age between 18 and 70 years, receiving busulfan as conditioning before their allo-HSCT with measured busulfan serum samples, and presence of DNA for genotyping. The Institutional Ethics Committee approved the study protocol. Written informed consent was obtained from all study participants according to the Helsinki Declaration. Diagnoses that required allo-HSCT were malignant: acute myeloid leukemia, chronic myeloid leukemia, chronic lymphoid leukemia, acute lymphoblastic leukemia, chronic myelomonocytic leukemia, non-Hodgkin's lymphoma or multiple myeloma or non-malignant: aplastic anemia, β-thalassemia or sickle cell anemia. Patients receiving busulfan between 2004 and 2008 were included in the exploratory cohort. This cohort was genotyped using de DMET plus arrays (Affymetrix UK Ltd, High Wycombe, UK). Patients receiving busulfan between 2008 and 2011 were included in the validation cohort. In this cohort, the positive findings of the exploratory cohort were tested.

Treatment regimens

Patients received either myeloablative conditioning or non-myeloablative conditioning. The non-myeloablative conditioning consisted of intravenous busulfan 0.8 mg/kg four times daily on days -6 and -5 before allo-HSCT, oral fludarabine 50 mg/m^2 on days -10 to -5, and alemtuzumab 15 mg for 2 days. In case of a matched unrelated donor, antithymocyte globulin 1–3 mg/kg was added. Alternatively, busulfan on days -6 and -5 before allo-HSCT was combined with oral fludarabine 50 mg/m^2 on day -10 to -5, intravenous cyclophosphamide 750 mg/m^2 on days -4 and -3, and alemtuzumab 15 mg during 2 days. The myeloablative conditioning consisted of intravenous busulfan 0.8 mg/kg four times daily on days -9 to -6 before allo-HSCT, intravenous cyclophosphamide 60 mg/kg on days -4 and -3, and alemtuzumab 15 mg during 2 days. Dosing of the different drugs, including busulfan, is based on the actual body weight. All patients received 92 mg phenytoin three times daily for seizure prophylaxis starting one day before busulfan administration. Busulfan was administered intravenously in a 2-hour infusion.

Serum drug level measurements were performed at 2.5 and 4.0 h after the start of the first infusion on the first day of treatment. Busulfan was analyzed in 200 ml serum using a validated high-performance liquid chromatographic assay involving precolumnderivatization with diethyldithiocarbamate, liquid/liquid extraction, and UV detection, described previously by our group[13] and on the basis of the method of Chow *et al*.[14] The assay is linear between 30 and 8000 mg/L and the limit of quantification is 30 mg/L. Accuracy at 200 and 1500 mg/L was 94 and 98%, respectively. Intraday variability in precision is 3.5 and 0.8% and interday variability in precision is 4.4 and 2.4% at 200 and 1500 mg/L, respectively. A validated limited sampling model was used to minimize the number of blood samples necessary to calculate the busulfanclearance.[13] A one-compartment population PK model for busulfan with linear elimination, developed and validated in MW/Pharm version 3.6 (Mediware, Groningen, the Netherlands), was used.[11,13,15,16] The calculated mean population PK parameters, clearance, volume of distribution, and half-life, were individualized according to the maximum a posteriori Bayesian fitting method.[17] No dose adjustments were made on the basis of the individual PK parameters.

Genotyping methods

DNA was isolated from EDTA blood using Maxwell (Promega, Leiden, the Netherlands) or Magnapure compact (Roche, Almere, the Netherlands). All samples were anonymized according to the instructions in the Codes for Proper Use and Proper Conduct (http:// www.federa.org).

DNA samples of the exploratory cohort were analyzed using DMET plus arrays (Affymetrix UK Ltd) according to the manufacturers' prescription and described in detail by Caldwell *et al*.[18] and Dumaual *et al*.[19]

Genotypes were calculated with DMET console software version 1.1 using the Dynamic Genotype Boundaries algorithm (Affymetrix UK Ltd). Patients with a call rate of less than 90% were excluded from the analyses. All individual SNPs were tested for Hardy–Weinberg equilibrium in gPlink 2.050[20]; SNPs that were not in equilibrium ($p < 0.001$) were excluded from further analysis.

In the validation cohort, SNPs were determined by high-resolution melting (HRM) of small amplicons with the LightScanner (HR-96; Idaho Technology, Salt Lake City, Utah, USA). Oligonucleotides used for small amplicon (40–60 bp) genotyping were chosen adjacent to the SNP. Melting curves were analyzed with LightScanner Software using Call-IT 2.0 (Idaho Technology). As a quality control measure, 10% of samples were genotyped in duplicate. Traditional Sanger sequencing was used to confirm HRM results for each SNPs. An overview of primers used in HRM and sequencing is presented in Table 4.1.

Table 4.1 Patient characteristics

	Exploratory cohort	Validation cohort	p-value
Number of patients	62	78	
Age in years (mean ± s.d.)	53 ± 11	57 ± 10	0.04
Sex			
Male	34	51	0.20
Female	28	27	
Disease			
AML	11	28	
CML	3	2	
CLL	5	7	
ALL	0	1	
MM	22	18	
NHL	10	18	
CMML	2	1	
AA	7	2	
β-thl	0	1	
SCA	2	0	
Busulfan clearance (L/hr/kg)	0.19 ± 0.05	0.17 ± 0.05	0.16
Volume of distribution	47.6 ± 16.6	51.0 ± 15.4	0.21
Conditioning regime			
Flu Bu (ATG or Alemtuzumab)	59	74	
Flu Bu CY	2	3	0.55
Bu Cy Alemtuzumab	1	1	

AML: acute myeloid leukemia, CML: chronic myeloid leukemia, CLL: Chronic lymphocytic leukemia, ALL: acute lymphoblastic leukemia , MM: Multiple Myeloma , NHL: Non-Hodgkin lymphoma , CMML: chronic myelomonocytic leukemia, AA: aplastic anemia, β-thl: β-thalassemia , SCA: sickle cell anemia, Flu: fludarabine, Bu: busulfan, Cy: cyclofosfamide, ATG: anti-thymocyteglobuline.

In both cohorts, more than 93% of the patients were of White ethnicity. Ethnicity was declared by the physician of the patient, but ethnicity was not studied as such in this study. To evaluate the ethnicity and to exclude the possibility of population stratification of the samples in the exploratory cohort, multidimensional scaling (MDS) was used, as implemented in software package plink. Four MDS coordinates were used and plots were created for adjacent MDS components.

Haplotype estimation

In the exploratory cohort, haploblocks were composed of SNPs associated with busulfan clearance within one gene. The haploblocks were tested to detect linkage disequilibrium.

If linkage disequilibrium between SNPs was present, haploblocks (with several haplotypes) were determined. A haplotype was set if the haplotype uncertainty parameter Rh^2 was greater than 0.95. Haplotypes with an Rh^2 of 0.95 or less were not considered for further analysis. Rare haplotypes (frequency < 10%) were pooled into one group in the association analysis. The following SNPs were analyzed in haploblocks: *ABCB1* rs2032588 and rs2235015; *ABCB4* rs45595532, rs2109505, and rs1202283; *ABCC2* rs7899457 and rs8187706; *ABCC6* rs8058694 and rs8058696; *CYP2B6* rs8058694 and rs8058696; *CYP39A1* haploblock 1 rs2277119 and rs59926524; *CYP39A1* haploblock 2 rs9381468 and rs953062; *CYP3A7* rs45467892 and rs45494802; *CYP4F2* rs2108622 and rs3093106; *FMO1* rs742350 and rs1126692, *GSTA1* rs4715332 and rs4715333; *GSTA5* rs4715354 and rs7746993; *NR3C1* rs6195, rs6190 and rs6189; *PPARD* rs7746988, rs6901410, rs6457815, rs7757196, rs7754530, rs7739752, rs6913026, rs6922548, rs6940722, rs6915115, rs6457816, rs6906237, and rs1053046; and *UGT2B15* rs4148269, rs3100, and rs4148269.

In the validation cohort, a tagging SNP was identified in each haploblock to determine haplotypes. The tagging SNP was selected on the basis of the frequencies of the different haplotypes in the exploratory cohort. In case it was not possible to determine haplotypes in the exploratory cohort on the basis of one tagging SNP, two SNPs were selected for determination of the haplotype. Care was taken such that the same SNPs could be analyzed in the exploratory and validation cohorts. This resulted in the following analysis in the validation cohort: Three haploblocks contained two SNPs: *GSTA5*: two tagging SNPs (rs4715354 and rs7746993), *CYP39A1* block 1: one tagging SNP (rs2277119) and block 2: two tagging SNPs (rs9381468 and rs953062). One haploblock (*ABCB4*) contained three SNPs, of which two were determined (rs2109505 and rs1202283). For each individual patient, haplotypes were estimated and haplotype Rh^2 was calculated using gPlink haplotypes; with Rh^2 more than 0.95 haplotypes were considered present.

In Table 4.2, an overview is presented of the number of SNPs and also the frequencies of the different haplotypes in both the exploratory and the validation cohort.

Statistical analysis

PK parameters are presented as means (± SD) (normal distribution) or geometric mean in case of log-normal distribution (± SD) of the parameter. In the exploratory cohort, associations with busulfan clearance and SNPs were tested initially by linear regression analysis with the SNPs in the additive model using gPlink. The effect of the number of copies of the haplotypes was tested in the same manner and the effect of gene copy number variation on busulfan clearance was tested using Student's t-test, both in PASW statistics, version 17.0.01 (SPSS Inc., Chicago,

Table 4.2 Statistical data of univariate and multivariate analysis in the exploratory and validation cohort

		Exploratory, univariate				Multivariate	Validation, univariate			
Gene	Haplotype / SNP	MAF or haplotype frequency	Explained variance (R^2, %)	p-value	Log OR	Log OR	MAF or haplotype frequency	Explained variance (R^2, %)	p-value	Log OR
GSTA5	CG	0.54	24.7	0.00004	0.038	0.016	0.52	0.065	0.026	0.017
SLC22A4	46011C>T	0.43	13.5	0.004	-0.026	-0.021	0.22	0.005	0.529	0.005
ABCB4	GAT	0.58	11.8	0.006	-0.024	-0.025	0.59	0.002	0.727	-0.0001
SLC7A8	-1065T>G	0.40	14.7	0.002	0.027	0.013	0.37	0.004	0.594	-0.004
CYP39A1	TC	0.27	21.3	0.00018	-0.036	-0.018	0.32	0.001	0.769	-0.02
CYP2C19	-806C/T	0.23	6.6	0.046	0.023	0.023	0.22	0.031	0.125	0.015
CYP39A1	GC	0.80	15.0	0.002	0.033	0.018	0.74	0.018	0.245	0.01

Seven markers in six genes identified in the multivariate analysis in the exploratory cohort.
MAF: minor allele frequency, OR: odds ratio.

Illinois, USA). Candidate markers with a p-value less than 0.05 (SNPs or haplotypes) and SNPs with a minor allele frequency more than 10%, to maintain a high enough power, were included in a multivariate linear regression analysis using a stepwise forward conditional approach with busulfan clearance as a dependent variable. The combination of SNPs and haplotypes explaining the largest portion of total variability in clearance was selected (highest R^2). The selection of these most informative SNPs and haplotypes was used to create a pharmacogenetic model predicting busulfan clearance (Cl_{PG}). This model was tested in the validation cohort.

All results from the multivariate analyses with a p-value less than 0.05 were considered significant. As this was an exploratory study, no correction for multiple testing was performed. In addition, this exploratory study used an uninformative approach, meaning that no adjustments for confounding factors and assumptions of polymorphism effects were made.

In the validation cohort, a predicted busulfan clearance was calculated for each patient with the regression model developed in the exploratory cohort (Cl_{PG}). The algorithm to calculate Cl_{PG} was based on the βs of each genetic marker from the multivariate analyses (Table 4.2) and a constant of 0.211; an additive model was assumed, according to Eq 4.1.

$$Cl_{PG} = 0.016*GSTA5 - 0.021*SLC22A4 - 0.025*ABCB4 + 0.013*SLC7A8 - 0.018*CYP39A1\_TC + 0.023*CYP2C19 + 0.018*CYP39A1\_GC \quad \text{(Eq 4.1)}$$

The value of each genetic marker is 0, 1, or 2 depending on the number of variant alleles. The busulfan clearances calculated using the pharmacogenetic model (Cl_{PG}) and the busulfan clearances calculated using the population pharmacokinetic model (Cl_{PK}) of each patient were compared on a case-by-case base with Pearson correlation. Furthermore, the association of the seven genetic markers with busulfan clearance was tested individually by linear regression analysis with the marker in the additive model in gPlink.

RESULTS

Description of the patient population

We carried out a two-stage approach, an exploratory analysis, and an independent validation study. Sixty-five adult patients were included in the exploratory cohort. Three patients were excluded because of a genotype call rate of less than 90%. The validation cohort included 78 adult patients. Patient characteristics of both cohorts are shown in Table 4.3. In the exploratory and validation cohort, 55 and 65% of patients were men, respectively (p = 0.20). The mean age

Table 4.3 Primers for high-resolution melting and Sanger sequencing of the 10 SNPs

Gene	SNP	Orientation	HRM primers 5'->3'	Sequence primers 5'->3'
GSTA5	rs4715354	Forward	GAGCTTTGTTCAGTCAACAC	TTCAGCAGAAAGAAAGGGAG
		Reverse	TCCTTGAGCTCTCAGGTTC	AACAACATAATCAGACAGGAG
GSTA5	rs7746993	Forward	TGTGGCATCTACACCACCC	GGTTCTCCATCACATGTGC
		Reverse	ATGTGGAAGAAGCAAGCTGG	GGGAATGAGGAGAGCAGAG
SLC22A4	rs1050152	Forward	TCTGACTGTCCTGATTGGAAT	CAATTACCTCCACCTTAAGAG
		Reverse	CTTTCAGGGAAAAAAAGGGTG	CTGATAGAGCTAGTCTCTAC
ABCB4	rs2109505	Forward	CTTTGTCACTAAATGCCGAG	AGCTACCCTCCTGCCTTG
		Reverse	CTGTTTCTTTTCTGTCCAG	CAGTAGATGTGGAACTTGAC
ABCB4	rs1202283	Forward	GTATTGAGTTCAGTGGTGTC	ATGGCATAGGCTATAGATGC
		Reverse	GAAATAGGATGGTTTGACATC	GATGAATAAAGGATGGTAGGG
SLC7A8	rs7141505	Forward	ATAAATCAGGGAACAGTTGTG	CTCCATCCTCAGTCCCGTC
		Reverse	TTCTGATGGCATTAAAAGTATC	GCGATTTTGTTTCCTCTGTAC
CYP39A1	rs9381468	Forward	TGGCAAAAATGTTAGCAGTT	TGTCATGAAATCTGTAAACCAC
		Reverse	CTCTATCTCTACCATTGAGA	GTGTTTAAAGTTCTTAAGTGAAG
CYP39A1	rs953062	Forward	GCTCATGGTCTGATGGAAAG	GCTCTCTATACAAATACCATGG
		Reverse	ATCATATTGATTAGAAATTAACTC	CAACCTCTCGGAGATGTTC
CYP2C19	rs12248560	Forward	TGTGTCTTCTGTTCTCAAAG	TGTCTGGAGGAGACCAGG
		Reverse	GCATTATCTCTTACATCAGAG	ATCCTATATCGAAGATTAGGAG
CYP39A1	rs2277119	Forward	GCAGTGCAAAATATCGTTTATC	TTTCTACTTATTTCATAGTACAAG
		Reverse	CTATCTGAAAATATTCTTTACCTG	CTGAGTAACATGTAACCAGTC

of the patient population was 53 ± 10 and 57 ± 11 years for the respective cohorts (p = 0.04). The mean busulfan clearance (Cl_{PK}) was similar in the exploratory cohort (0.19 ± 0.05 L/h/kg) and in the validation cohort (0.17 ± 0.04 L/h/kg) (p = 0.16).

The two different cohorts were created sequentially on the basis of the date of HSCT and busulfan administration. The two cohorts were similar in terms of sex distribution, average busulfan clearance, indication, and conditioning regimen. There was a significant difference in the average age between the two cohorts. However, age is not associated with busulfan clearance in adult patients.

Genotyping

In the exploratory cohort, of the 1936 SNPs determined, 46 SNPs were X-chromosomal. As there was no sex difference in busulfan clearance levels (p = 0.19), these SNPs were excluded from the analysis. All 62 patients passed QC metrics and produced useable genotypes.

The average call rate of the remaining 1890 SNP assays was more than 98%. One SNP (RS7436963/*UGT2B17*) failed to meet Hardy–Weinberg equilibrium (p < 0.001) and was excluded from further analysis. Of the remaining 1889 SNPs, 851 SNPs were polymorphic in the study population and were used in the association study. An overview is presented in Figure 4.1.

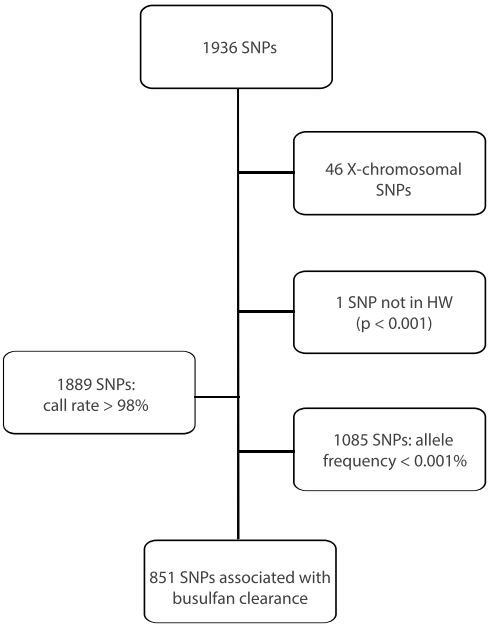


Figure 4.1 Flow charts of genotyping results in the exploratory cohort. HW: Hardy-Weinberg, SNP: single nucleotide polymorphism.

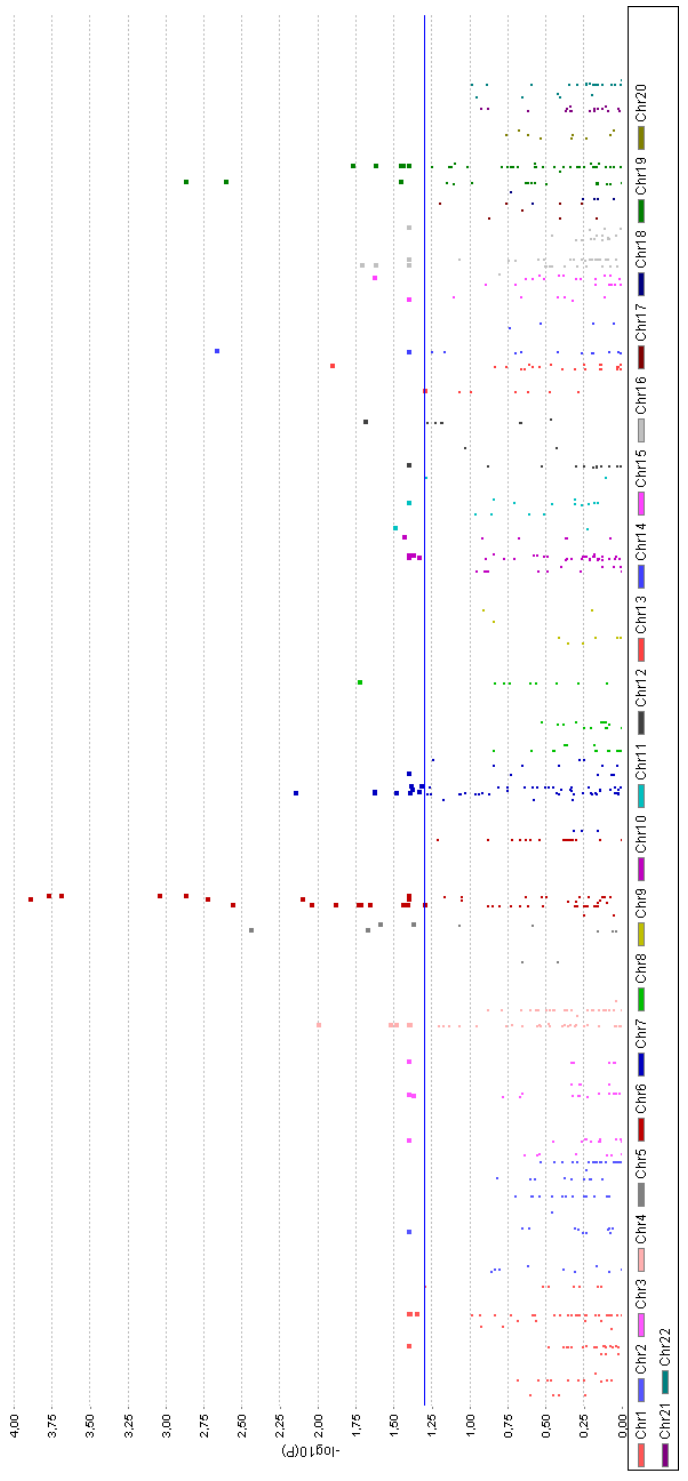


Figure 4.2 Manhattan plot of –Log (p-values) from linear regression analysis of 851 individual genetic variants and busulfan clearance. Genetic variants are organized per chromosome. Blue line indicating a p-value of 0.05.

To exclude population stratification, MDS plots were created (Supplemental Figure 4.1). The plots do not indicate the presence of strong population stratification. If population stratification would be present, clear separated clusters would be observed in the graphs, indicating a difference in allele frequencies between subpopulations.

In the validation cohort, 10 SNPs, which were selected in the exploratory cohort, were determined. The average call rate was more than 99%; all variants showed Hardy–Weinberg equilibrium ($p > 0.05$).

Association of genetic variants with busulfan clearance in the exploratory cohort

Of the 851 SNPs, 86 SNPs in 52 different genes showed a significant association with busulfan clearance ($p < 0.05$; see also Figure 4.2). Of the 45 SNPs, 15 haploblocks, containing 29 haplotypes, located in 14 different genes, were formed. In the next step of association analysis, SNPs and haplotypes that fulfilled the selection criteria ($p < 0.05$, minor allele frequency > 10%) were applied in the multivariate analyses. Seven variants (three SNPs and four haplotypes) in six genes were selected as the top genetic markers, explaining 64% (adjusted R^2) of variance in busulfan clearance in the exploratory cohort ($p < 0.001$) (Table 4.2). To be more specific, three genes of the model are involved in drug metabolism processes in general: *GSTA5* (rs4715354 and rs7746993) and two cytochrome P450 genes: *CYP2C19* (rs12248560) and *CYP39A1* (rs2277119 and rs9381468 and rs953062) and three genes were found to be involved in drug transport: *ABCB4* (rs2109505 and rs1202283), *SLC22A4* (rs1050152), and *SLC7A8* (rs7141505). The effects of each individual genetic marker on busulfan clearance are presented in individual graphs in Figure 4.3.

Predication of busulfan clearance in validation cohort

In the validation cohort, busulfan clearance was calculated using the prediction algorithm and the individual genotypes of the seven markers. The average predicted clearance (Cl_{PG}) was 0.19 ± 0.04 L/h/kg. A correlation graph and a Bland–Altman plot are provided in Supplemental Figure 4.2 showing a poor correlation between Cl_{PG} and clearance calculated using the PK model (Cl_{PK}), $R^2 = 0.024$, but no systematic bias between the two methods.

When the seven genetic markers were tested separately, one haplotype in *GSTA5* (rs4715354 and rs7746993) remained statistically significant ($p = 0.025$) for correlation with busulfan clearance. Busulfan clearance of patients with one variant allele was decreased by 11% and 18% in patients with two variant alleles as compared with *GSTA5* wild-type patients. In the validation cohort, the *GSTA5* haplotype could explain 6.5% of variability in busulfan clearance.

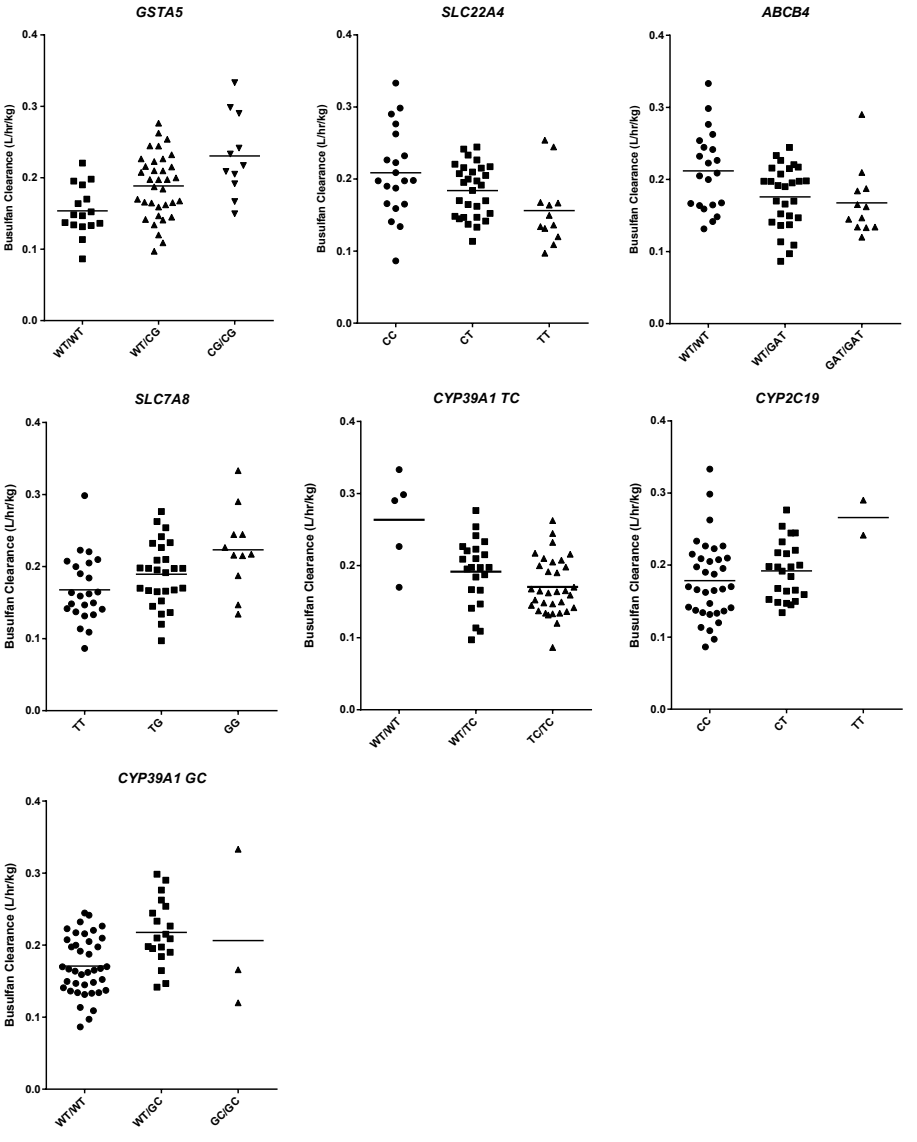


Figure 4.3 Effects of individual genotypes associated with busulfan clearance in the exploratory cohort.

DISCUSSION

This is the first study using an exploratory pharmacogenetic approach including a wide range of genes involved in drug ADME to investigate the interindividual variability in busulfan clearance in adults. The *GSTA5* haplotype was found to be associated significantly with busulfan

clearance both in the exploratory and in the validation cohort. No additional genetic markers involved in drug metabolism and transport appeared to be associated with busulfan clearance. Busulfan is metabolized primarily by conjugation with glutathione, catalyzed by GSTs. Different GST enzymes are involved in busulfan metabolism, GSTA1 being the predominant enzyme.[8] The effect of SNPs in genes encoding for three GST enzymes (GSTA1, GSTM1, and GSTP1) on busulfan PK has been studied before, leading to ambiguous results. Our group[11] and Kim *et al*.[10] showed an effect of a *GSTA1* SNP (rs3957357) on busulfan clearance in adult patients receiving busulfan intravenously; the other two SNPs showed no effects. Our group[21] and Ansari *et al*.[22] studied the same SNPs in a pediatric population. Both studies showed no association between the *GSTA1* SNP and busulfan PK.

Moreover, Hassan *et al*.[23] suggested the involvement of transporter enzymes in busulfan metabolism, transporting the glutathione–busulfan conjugate out of the cell. Only one study investigated the effect of a transporter on busulfan PK. In this study, the effect of three *GST* SNPs and two *ABCB1* SNPs (rs1045642 and rs2032582) was investigated and it was found that combined polymorphisms in *GSTM1* and *ABCB1* were associated with busulfan PK.[24]

In the exploratory cohort, genetic markers in different transporters (two solute carriers and one ABC transporter) were identified to be potentially related to variance in busulfan PK. These findings could not be confirmed in the validation cohort, thus suggesting that the involvement of transporters in busulfan PK is not likely.

In our study, genetic variants were tested in relation to busulfan clearance. Unfortunately, our study does not have sufficient statistical power to relate genetic variations to clinical outcome parameters such as toxicity, efficacy (engraftment), or relapse. Instead, we explored relationships with busulfan clearance. However, for busulfan, a clear relationship between PK and clinical outcome has been described.[3–5]

The *GSTA5* haplotype was found to be associated significantly with busulfan clearance. The *GSTA5* gene is a member of the same family as *GSTA1* and *GSTA5* protein is absent in human tissue.[25] However, the previously studied *GSTA1* SNP (rs3957357, -69C/T), a tagging SNP in a haploblock with three other SNPs in the promoter of *GSTA1* (-631T/G, -567T/G, -69C/T, and -52G/A), and the tagging SNP of *GSTA5* (rs4715354) are linked; haploview linkage disequilibrium of these SNPs is $D' = 1$ and $r^2 = 0.69$. Thus, the association with the *GSTA5* SNP is in fact a positive control for earlier findings for the association of *GSTA1* SNP and busulfan PK. The *GSTA1* SNP (rs3957357) is not present on the DMET array and was therefore analyzed separately. The absence of the tagging SNP of the *GSTA1* gene on the DMET array and the relatively small sample size resulted in a slightly larger effect size of *GSTA5* in comparison with other *GSTA1* SNPs on the

DMET array in the exploratory cohort. Separate analyses of *GSTA1* (rs3957357) showed an association with busulfan clearance in the exploratory cohort ($p < 0.01$) and in the validation cohort ($p = 0.02$). Linkage of the two SNPs could be confirmed in both cohorts, resulting in $D'=1$ and $r^2 = 0.57$ in the exploratory cohort and $D'= 1$ and $r^2 = 0.71$ in the validation cohort.

Previous studies indicate that only a small proportion of variability in busulfan PK can be attributed to genetic variants in *GST*s. We hypothesized that several yet unknown metabolic enzymes and drug transporters could be involved in busulfan PK, each with a small influence. Therefore, in this study, we used an unsupervised genetic association approach. The advantage of such an approach is the possibility of discovering new pathways and genetic variants that might contribute toward variability in busulfan exposure. One of the first examples of the possibility of finding a novel unexpected relationship using the DMET array is given by Ross *et al*.[26], showing an association between cisplatin ototoxicity and genetic variants in *TPMT* and *COMT* using the same genotyping platform.

A disadvantage of this exploratory approach is the difficult interpretation of its results as the effects of the identified variants on gene function or expression are not always clear. Hence, the effect of variants found in the exploratory cohort on busulfan metabolism can only be hypothesis-generating and should be confirmed in a validation study, as we performed in the current study.

An overview of the function and possible effects on busulfan clearance of the seven genes that were found in the exploratory study is shown in Supplemental Table 4.1.

Concomitant medications have been associated with busulfan PK. Especially antifungals, but also antiepileptic drugs are known for drug–drug interactions with busulfan. All patients included in our study received 92 mg phenytoin three times daily for seizure prophylaxis, starting one day before busulfan administration. Phenytoin could induce the metabolism of busulfan. However, the effect of phenytoin is not very strong. Furthermore, if this interaction would affect busulfan PK, it will be the same in all patients included in this study. A more significant interaction is the effect of antifungals on busulfan PK. Antifungals such as ketoconazole can increase busulfan exposure, probably because of inhibition of busulfan metabolism in the liver. In our study population, the only antifungal administered to the patients concomitantly with busulfan was oral amphotericin B. This drug is not absorbed and will therefore not affect busulfan metabolism in the liver.

In our study, a large number of SNPs were evaluated in the exploratory cohort, which introduces the potential problem of multiple testing and an increase in the risk of false-positive findings. To enrich our SNP set with predictive SNPs to be used by a prediction model for busulfan clearance,

we applied a mild statistical threshold of p less than 0.05 in the exploratory cohort. P-value thresholds in an exploratory set are meant to enrich for associated SNPs not strictly controlling type I error (following Purcell *et al*.[27]). Evaluation of the association of scores in the exploratory set has descriptive merits and formal evaluation of significance is carried out in the validation set. It is possible to use alternative strategies for the construction of a prediction model by using different model selection techniques (such as penalized regression or ensemble methods); however, by testing external validity in an independent data set, we offer a conservative and robust approach to model evaluation. Construction of a simple model has the advantage of overfitting and making our results more comparable with similar studies (compare Purcell *et al*.[27]).

Evaluation of the associations in the validation cohort resulted in confirming one of the seven genetic markers (e.g. *GSTA5* haplotype). However, the explained variability of *GSTA5* in the validation cohort was relatively small in comparison with the findings in the exploratory cohort. This decrease in explained variability could be because of unmeasured confounders that are different between the two different cohorts. Furthermore, the explained variability of 64% in the exploratory cohort might be an overestimate because of the so-called winners curse. To assess the magnitude of this phenomenon, we performed a data-split by fitting the model in the first half and computing the R^2 of the resulting predictor in the second half. With this analysis, the R^2 decreased from 75 to 37%, indicating the presence of overfitting (aka winner's curse). Still, this analysis confirms that the model has considerable explanatory power. The findings of involvement of *GSTA5* on busulfan PK will not directly have an impact on clinical practice.

This is the first study using an exploratory pharmacogenetic approach in 225 genes involved in ADME to find an explanation for the interindividual variability in busulfan clearance. The *GSTA5* haplotype, and thus *GSTA1*, was significantly associated with busulfan clearance both in the exploratory and in the validation cohort. No additional genetic markers involved in drug metabolism and transport were found to be associated with busulfan clearance.

Acknowledgments

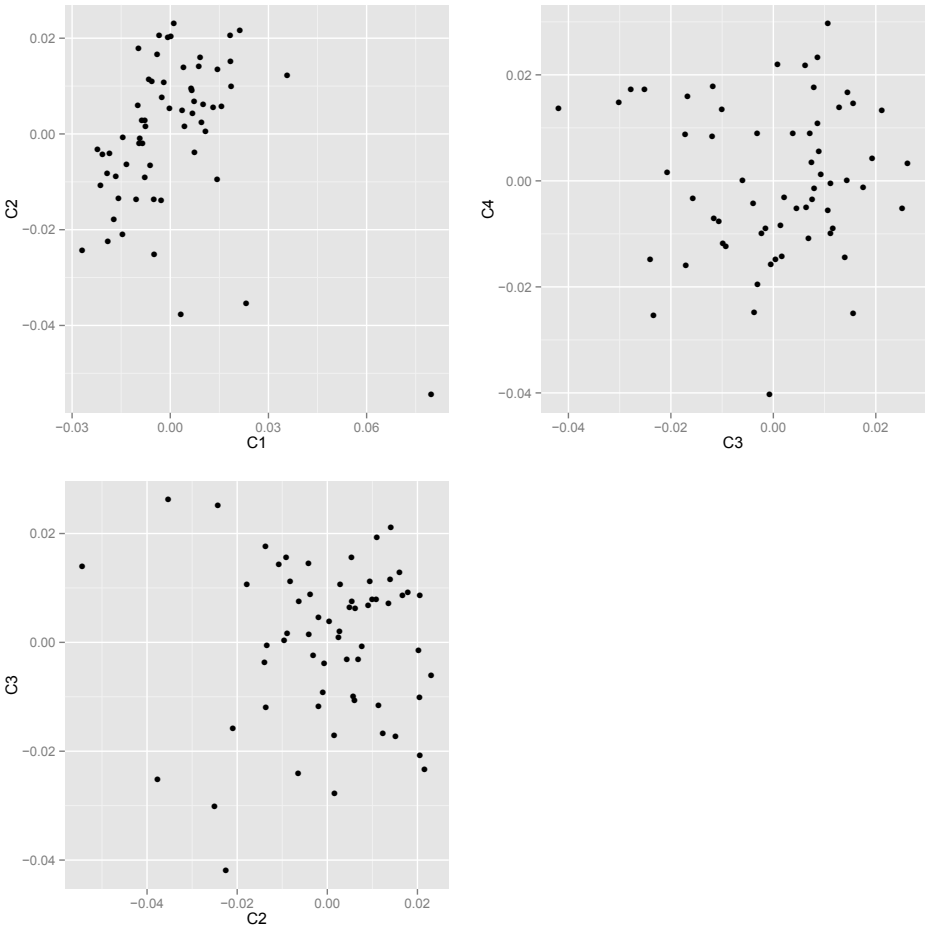
The Central Clinical Hematology Laboratory of the Leiden University Medical Center is acknowledged for providing DNA samples.

REFERENCES

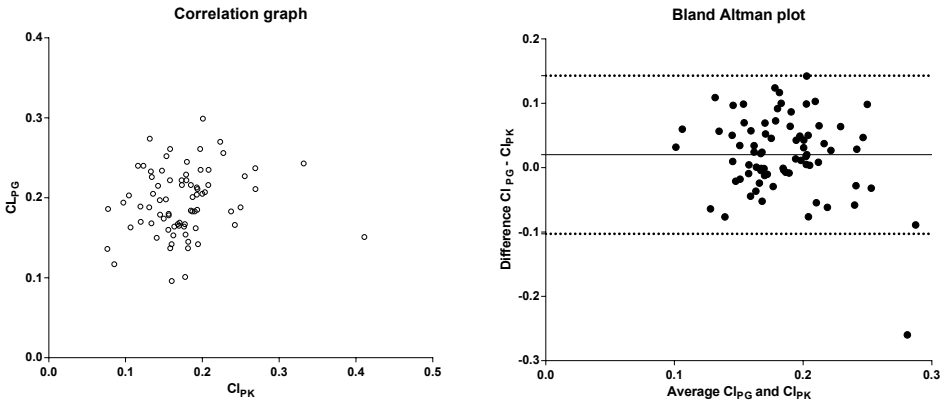
1. Madden, T., Lima, M. de, Thapar, N., Nguyen, J., Roberson, S., Couriel, D., *et al.* Pharmacokinetics of once-daily IV busulfan as part of pretransplantation preparative regimens: a comparison with an every 6-hour dosing schedule. *Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant.* **13**, 56–64 (2007).
2. Lee, J. W., Kang, H. J., Lee, S. H., Yu, K. S., Kim, N. H., Yuk, Y. J., *et al.* Highly variable pharmacokinetics of once-daily intravenous busulfan when combined with fludarabine in pediatric patients: phase I clinical study for determination of optimal once-daily busulfan dose using pharmacokinetic modeling. *BiolBlood Marrow Transpl.* **18**, 944–950 (2012).
3. McCune, J. S., Gibbs, J. P. & Slattery, J. T. Plasma concentration monitoring of busulfan: does it improve clinical outcome? *Clin. Pharmacokinet.* **39**, 155–165 (2000).
4. Grochow, L. B., Jones, R. J., Brundrett, R. B., Braine, H. G., Chen, T. L., Saral, R., *et al.* Pharmacokinetics of busulfan: correlation with veno-occlusive disease in patients undergoing bone marrow transplantation. *Cancer Chemother. Pharmacol.* **25**, 55–61 (1989).
5. Slattery, J. T., Clift, R. A., Buckner, C. D., Radich, J., Storer, B., Bensinger, W. I., *et al.* Marrow transplantation for chronic myeloid leukemia: the influence of plasma busulfan levels on the outcome of transplantation. *Blood* **89**, 3055–3060 (1997).
6. Salinger, D. H., Vicini, P., Blough, D. K., O'Donnell, P. V., Pawlikowski, M. A. & McCune, J. S. Development of a Population Pharmacokinetics-Based Sampling Schedule to Target Daily Intravenous Busulfan for Outpatient Clinic Administration. *J. Clin. Pharmacol.* **50**, 1292–1300 (2010).
7. Booth, B. P., Rahman, A., Dagher, R., Griebel, D., Lennon, S., Fuller, D., *et al.* Population pharmacokinetic-based dosing of intravenous busulfan in pediatric patients. *J.Clin.Pharmacol.* **47**, 101–111 (2007).
8. Czerwinski, M., Gibbs, J. P. & Slattery, J. T. Busulfan conjugation by glutathione S-transferases alpha, mu, and pi. *Drug Metab Dispos* **24**, 1015–1019 (1996).
9. Kusama, M., Kubota, T., Matsukura, Y., Matsuno, K., Ogawa, S., Kanda, Y., *et al.* Influence of glutathione S-transferase A1 polymorphism on the pharmacokinetics of busulfan. *Clin. Chim. Acta Int. J. Clin. Chem.* **368**, 93–98 (2006).
10. Kim, S. D., Lee, J. H., Hur, E. H., Lee, J. H., Kim, D. Y., Lim, S. N., *et al.* Influence of GST gene polymorphisms on the clearance of intravenous busulfan in adult patients undergoing hematopoietic cell transplantation. *BiolBlood Marrow Transpl.* **17**, 1222–1230 (2011).
11. Brink, M. H. ten, Wessels, J. A., Hartigh, J. den, Straaten, T. van der, Borne, P. A. von dem, Guchelaar, H.-J., *et al.* Effect of genetic polymorphisms in genes encoding GST isoenzymes on BU pharmacokinetics in adult patients undergoing hematopoietic SCT. *Bone Marrow Transplant.* **47**, 190–195 (2012).
12. Sissung, T. M., English, B. C., Venzon, D., Figg, W. D. & Deeken, J. F. Clinical pharmacology and pharmacogenetics in a genomics era: the DMET platform. *Pharmacogenomics* **11**, 89–103 (2010).
13. Cremers, S., Schoemaker, R., Bredius, R., Hartigh, J. den, Ball, L., Twiss, I., *et al.* Pharmacokinetics of intravenous busulfan in children prior to stem cell transplantation. *Br. J. Clin. Pharmacol.* **53**, 386–389 (2002).
14. Chow, D. S., Bhagwatwar, H. P., Phadungpojna, S. & Andersson, B. S. Stability-indicating high-performance liquid chromatographic assay of busulfan in aqueous and plasma samples. *JChromatogrB BiomedSciAppl* **704**, 277–288 (1997).
15. Proost, J. H. & Meijer, D. K. MW/Pharm, an integrated software package for drug dosage regimen calculation and therapeutic drug monitoring. *Comput. Biol. Med.* **22**, 155–163 (1992).
16. Zwaveling, J., Bredius, R. G. M., Cremers, S. C. L. M., Ball, L. M., Lankester, A. C., Teepe-Twiss, I. M., *et al.* Intravenous busulfan in children prior to stem cell transplantation: study of pharmacokinetics in association with early clinical outcome and toxicity. *Bone Marrow Transplant.* **35**, 17–23 (2005).
17. Proost, J. H. Adaptive control of drug dosage regimens using maximum a posteriori probability Bayesian fitting. *Int. J. Clin. Pharmacol. Ther.* **33**, 531–536 (1995).

18. Caldwell, M. D., Awad, T., Johnson, J. A., Gage, B. F., Falkowski, M., Gardina, P., *et al.* CYP4F2 genetic variant alters required warfarin dose. *Blood* **111**, 4106–4112 (2008).
19. Dumaual, C., Miao, X., Daly, T. M., Bruckner, C., Njau, R., Fu, D.-J., *et al.* Comprehensive assessment of metabolic enzyme and transporter genes using the Affymetrix Targeted Genotyping System. *Pharmacogenomics* **8**, 293–305 (2007).
20. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
21. Zwaveling, J., Press, R. R., Bredius, R. G. M., Derstraaten, T. R. J. H. M. van, Hartigh, J. den, Bartelink, I. H., *et al.* Glutathione S-transferase polymorphisms are not associated with population pharmacokinetic parameters of busulfan in pediatric patients. *Ther. Drug Monit.* **30**, 504–510 (2008).
22. Ansari, M., Lauzon-Joset, J.-F., Vachon, M.-F., Duval, M., Théoret, Y., Champagne, M. A., *et al.* Influence of GST gene polymorphisms on busulfan pharmacokinetics in children. *Bone Marrow Transplant.* **45**, 261–267 (2010).
23. Hassan, M. & Andersson, B. S. Role of pharmacogenetics in busulfan/cyclophosphamide conditioning therapy prior to hematopoietic stem cell transplantation. *Pharmacogenomics* **14**, 75–87 (2013).
24. Krivoy, N., Hoffer, E., Lurie, Y., Bentur, Y. & Rowe, J. M. Busulfan use in hematopoietic stem cell transplantation: pharmacology, dose adjustment, safety and efficacy in adults and children. *CurrDrug Saf* **3**, 60–66 (2008).
25. Singh, S. P., Zimniak, L. & Zimniak, P. The human hGSTA5 gene encodes an enzymatically active protein. *Biochim. Biophys. Acta* **1800**, 16–22 (2010).
26. Ross, C. J. D., Katzov-Eckert, H., Dubé, M.-P., Brooks, B., Rassekh, S. R., Barhdadi, A., *et al.* Genetic variants in TPMT and COMT are associated with hearing loss in children receiving cisplatin chemotherapy. *Nat. Genet.* **41**, 1345–1349 (2009).
27. International Schizophrenia Consortium, Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
28. Hassan, M. & Andersson, B. S. Role of pharmacogenetics in busulfan/cyclophosphamide conditioning therapy prior to hematopoietic stem cell transplantation. *Pharmacogenomics* **14**, 75–87 (2013).

SUPPLEMENTAL MATERIAL



Supplemental Figure 4.1 Multidimensional scaling (MDS) plots of adjacent MDS components The plots do not indicate the presence of strong population stratification. If stratification would be present, clear clusters would be depicted in the graphs, demonstrating a difference in allele frequencies between subpopulations.



Supplemental Figure 4.2 Correlation graph and Bland Altman Plot. Correlation graph showing busulfan clearance based on PK data (Cl_{PK}) on the x-axis and clearance predicted with genetic markers (Cl_{PG}) on the y-axis, $R^2 = 0.024$. Bland Altman plot: on the x-axis: average of Cl_{PG} and Cl_{PK} per individual and on the y-axis the difference between Cl_{PG} and Cl_{PK} per individual. Middle line: the average difference and the two dashed lines: upper and lower limits of agreement (-2SD and +2SD).

Supplemental Table 4.1 An overview of the function and possible effects on busulfan clearance of the 7 genes that were found in the exploratory cohort

Gene	Role gene[1]	Postulated effect
GSTA5	Part of the alpha class of gluthationetransferases, catalyze conjugation with gluthation.	No functional protein identified in human.[2]
SLC22A4	Influx transporter also known as carnitine/organic cation transporter (OCTN1), present in liver, kidney, intestine and other organs. Critical for elimination of many small organic cations.[3]	
ABCB4	ATP-binding cassette, sub-family B (MDR/TAP). This gene encodes a full transporter and member of the p-glycoprotein family; it may involve transport of phospholipids from liver hepatocytes into bile.	p-GP is a transporter of gluthatione conjugates out of the cell and may transport busulfan conjugate.[4]
SLC7A8	Cationic amino acid transporter, also known as *LAT2* gene.	SNP located in haploblock in promoter region of gene; no evidence for effect of SNP on expression of gene.[5]
CYP39A1 Block A & B	A member of the cytochrome P450 superfamily of enzymes. This endoplasmic reticulum protein is involved in the conversion of cholesterol to bile acids.	
CYP2C19	This protein localizes to the endoplasmic reticulum and is known to metabolize many xenobiotics including omeprazol and anticonvulsive drugs.	*17 allele: increased transcription and ultrarapid allele.[6] Busulfan conjugate is oxidized in the liver, which may be facilitated by *CYP2C19*.

References

1 Home - Gene - NCBI. 2013. http://www.ncbi.nlm.nih.gov/gene/ (accessed 8 Jul 2013).
2 Singh SP, Zimniak L, Zimniak P. The human hGSTA5 gene encodes an enzymatically active protein. *Biochim Biophys Acta.* **1800**, 16–22 (2010).
3 Tamai I. Pharmacological and pathophysiological roles of carnitine/organic cation transporters (OCTNs: SLC22A4, SLC22A5 and Slc22a21). *Biopharm Drug Dispos.* **34**, 29–44 (2013).
4 Hassan M, Andersson BS. Role of pharmacogenetics in busulfan/cyclophosphamide conditioning therapy prior to hematopoietic stem cell transplantation. *Pharmacogenomics.* **14**, 75–87 (2013).
5 Kühne A, Kaiser R, Schirmer M, Heider U, Muhlke S, Niere W, et al.Genetic polymorphisms in the amino acid transporters LAT1 and LAT2 in relation to the pharmacokinetics and side effects of melphalan. *Pharmacogenet Genomics.* **17**, 505–517 (2007).
6 CYP2C19. http://www.cypalleles.ki.se/cyp2c19.htm (accessed 8 Jul 2013).