Short Communication

For reprint orders, please contact: reprints@futuremedicine.com

Pharmacogenomics

UGT1A1 regulatory variant with potential effect on efficacy of HIV and cancer drugs commonly prescribed in South Africa

Jenny Mary Mathew[1,2], Phelelani Thokozani Mpangase[2], Dhriti Sengupta[2], Stanford Kwenda[3], Demetra Mavri-Damelin[4] & Michèle Ramsay*[,1,2]

[1]Division of Human Genetics, National Health Laboratory Service, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2000, South Africa

[2]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2050, South Africa

[3]Sequencing Core Facility, National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg, 2192, South Africa

[4]School of Molecular & Cell Biology, Faculty of Science, University of the Witwatersrand, Johannesburg, 2050, South Africa

*Author for correspondence: Michele.Ramsay@wits.ac.za

Aim: Despite the high disease burden of human immunodeficiency virus (HIV) infection and colorectal cancer (CRC) in South Africa (SA), treatment-relevant pharmacogenetic variants are understudied. **Materials & methods:** Using publicly available genotype and gene expression data, a bioinformatic pipeline was developed to identify liver expression quantitative trait loci (eQTLs). **Results:** A novel *cis*-eQTL, rs28967009, was identified for *UGT1A1,* which is predicted to upregulate *UGT1A1* expression thereby potentially affecting the metabolism of dolutegravir and irinotecan, which are extensively prescribed in SA for HIV and colorectal cancer treatment, respectively. **Conclusion:** As increased *UGT1A1* expression could affect the clinical outcome of dolutegravir and irinotecan treatment by increasing drug clearance, patients with the rs28967009A variant may require increased drug doses to reach therapeutic levels or should be prescribed alternative drugs.

First draft submitted: 20 May 2021; Accepted for publication: 23 August 2021; Published online: 16 September 2021

Keywords: Dolutegravir • eQTL • Irinotecan • South Africa • *UGT1A1*

Expression quantitative trait loci (eQTLs) are genetic variants that contribute to variation in gene expression levels in specific tissues or cells, thus contributing to the modulation of protein levels in cells [1]. eQTLs have been implicated in susceptibility to disease, and some are lead SNPs in genome-wide association studies for complex diseases. Understanding the effect of sequence variation on gene regulation can thus contribute to understanding human disease [2]. With relevance to this study, eQTLs also have the intrinsic potential to assist in the prediction of the clinical outcome of treatment should they affect the expression of genes that modulate drug absorption, distribution, metabolism and excretion (ADME). eQTLs of ADME genes can be correlated with the level of drug efficacy and an individual's propensity towards developing adverse drug reactions (ADRs) [3]. Therefore, eQTLs play an important role in pharmacogenomics as it is practically more feasible to determine an individual's genotype as a proxy for gene expression, in an effort to ensure that a patient is prescribed drug therapy regimens best suited to their genetic makeup.

The frequencies of eQTLs of ADME genes vary both within and between populations. African ancestry populations show the most inter-population variation for known genetic variants and East Asian ancestry populations generally show the least [4]. There is evidence for significant population sub-structure within SA that likely also applies to variants in ADME genes [5]. As an example of ethnic differences, at the same dose, a twofold higher systemic concentration of rosuvastatin, a drug administered for lowering low-density lipoprotein cholesterol levels, is observed in individuals of East Asian ancestry compared with those of European ancestry. Hence, the recommended dose of rosuvastatin in East Asians is lower, at 5 mg, half the commonly prescribed dose for other

populations [6]. Although this is likely caused by genetic variation, Lee *et al.* detected no associations with two *SLCO1B1* variants (A 388>G and T 521>C). A more recent study, however, showed that rosuvastatin levels were higher in individuals with the *SLCO1B1* 521C allele and the *ABCG2* 421A allele [7] suggesting a possible genetic etiology. Data on genetic variation in ADME genes can therefore help clinicians and researchers to understand individual differences in sensitivity or resistance to certain drugs, thereby avoiding ADRs or ineffective treatment in patients and improving the quality of therapies [8].

While genetic polymorphisms and environmental factors influence how effective a drug is and the individuals' propensity to develop ADRs, 90% of these genetic variants are in non-coding regions of the genome, which could have important regulatory functions [9]. Nevertheless, extensive in-depth knowledge of genetic variations, pre-existing clinical factors and environmental variables only explain a small proportion of variance in drug response and despite advances in understanding drug response variation, there is still a large gap in our knowledge of the causes of inter-individual differences. Changes in gene expression levels have been found to be a major mechanism that underlies drug response variation and susceptibility to disease. eQTLs of ADME genes, could therefore be used to test relevant pharmacogenetic associations with drugs that are metabolized by the corresponding enzyme(s). Since eQTLs modulate gene expression levels, a pipeline that identifies eQTLs for ADME genes in relevant tissues could help to explain some of the variance in drug dose requirements that contribute to the currently unexplainable portion of variability in drug response and propensity to ADRs. The pipeline could be applied to data from any tissue type, but in this study, we focused on the liver since ADME genes are abundantly expressed in this organ.

This study reports on the predicted effect of a novel *cis*-eQTL on drugs metabolized by UGT1A1, a Phase II drug-metabolizing enzyme. UGT1A1 is involved in the metabolism and clearance of a range of drugs and as such has been the focus of recent pharmacogenomics studies since deviation from what is considered normal expression levels, can significantly impact drug response. We focused on dolutegravir and irinotecan as these drugs are commonly prescribed in SA for human immunodeficiency virus (HIV) and colorectal cancer (CRC), respectively. The 16 Southern African Development Community countries have the largest number of people living with HIV/acquired immunodeficiency syndrome (AIDS) [10], and SA has the largest HIV epidemic in the world with 7.2 million HIV-positive individuals [11]. CRC is the fourth most common and sixth most lethal cancer in SA [12,13]. Despite improved survival from antiretroviral therapy (ART) and chemotherapy, HIV and CRC remain serious causes of morbidity and mortality in SA. The potential impact of an African-specific variant of *UGT1A1* on HIV and CRC treatment is explored in this study.

Materials & methods

Datasets used & quality control (QC) done

The genotype and gene expression datasets, GSE39036 and GSE32504 respectively, used in this analysis were obtained from a study by Schröder *et al.* and downloaded from NCBI's Gene Expression Omnibus [14]. Prior to quality control (QC), this dataset had 149 samples, 318,237 genotyped SNPs and 15,439 gene expression transcripts. The genotype dataset from GSE39036 was imputed to increase the number of SNPs using the Michigan imputation server with the Phase III 1000 Genomes Project reference panel. Post imputation, r^2 value (reflecting imputation quality) greater than 0.8, minor allele count greater than one, SNPs with two alleles and SNPs with missing data of up to 60% were tolerated.

Ethics approval

Ethics approval (M170756) was obtained from the University of the Witwatersrand, Johannesburg, SA. This study was conducted according to the principles of the Declaration of Helsinki.

Pipeline development

A Nextflow (https://www.nextflow.io/) bioinformatic pipeline was developed (https://github.com/phelelani/nf-eqtl) to identify tissue-specific eQTLs from GSE39036 and GSE32504. The pipeline takes as input four files (normalized gene expression file, genotype file, gene expression platform-specific supplementary file with probe information and genotype platform-specific supplementary file with SNP data) and generates PLINK input and binary files (namely PED, MAP, BED, BIM and FAM files). Sample and SNP QC is then performed on the genotype dataset using PLINK (http://zzz.bwh.harvard.edu/plink/). Samples with discordant sex information, missingness of more than 5%, extreme mean heterozygosity and population outliers based on principal component analysis, as well as duplicated or related samples, are removed during QC and excluded from analysis. For SNP

Table 1. *Cis*-eQTLs for core ADME genes using the imputed genotype dataset and gene expression dataset.

eQTL	rsID	Gene	β-value	t-stat	p-value	FDR
10:97236373	rs933765367	*CYP2C19*	3238	5.74	5.64E-08	0.0009
7:100085809	rs143210517	*CYP3A5*	16158	5.82	3.76E-08	0.0006
1:97464825	rs114045355	*DPYD*	325	5.16	8.27E-07	0.0075
16:27948796	rs1000598999	*SULT1A1*	47	5.26	5.33E-07	0.0053
2:234515276	rs28967009	*UGT1A1*	1328	5.24	5.73E-07	0.0056
4:68462245	rs917897719	*UGT2B17*	3944	5.54	1.46E-07	0.0019

β-value: Beta value effect on gene expression; eQTL: Expression quantitative trait loci; FDR: False discovery rate; t-stat: t-statistic.

QC, SNPs with more than 6.5% missing data, minor allele frequency (MAF) lower than 1% and Hardy-Weinberg equilibrium p-value lower than 0.00001, are excluded from the analysis. The pipeline identifies sample outliers using a basic principal component analysis plot based on the normalized gene expression dataset. Since samples in the genotype and gene expression datasets have different 'Sample IDs', the pipeline uses 'sample titles' to map and identify which samples must be removed, thereby ensuring that only individual level matched genotype and gene expression data are included in the analysis. The pipeline was used to generate the five input files namely: gene expression file; gene location file; genotype file; SNP location file and covariate file, which are required to identify eQTLs from the datasets. Once created, the pipeline uses the R-package MatrixEQTL to generate the list of eQTLs [2]. *Cis*-eQTLs are defined as being within 1 megabase (Mb) on either side of a gene's transcription start site (TSS) and *trans*-eQTLs are at least 5 Mb downstream or upstream of a gene's TSS or on a different chromosome. eQTLs were detected first using the post-QC genotype and gene expression data and then using imputed genotype data and gene expression data. An identified *cis*-eQTL was only considered to be significant if it had a p-value < 0.05 and a false discovery rate (FDR) value <0.05.

Bioinformatic analysis

The list of genes with associated eQTLs generated using our pipeline from both the datasets, GSE39036 and GSE32504, was searched for core ADME genes obtained from the PharmaADME consortium (http://pharmaadme.org/joomla/). An in-depth *in silico* analysis was performed for all the significant *cis*-eQTLs identified. The first step was to find whether the chromosome coordinates obtained for each eQTL had rsIDs associated with them using the online tool SNPnexus. The GTEx database was then queried to identify whether the ADME gene associated with each *cis*-eQTL in question, is expressed in the liver and whether the eQTL has been previously identified. Allele frequencies of the eQTLs in different African populations were assessed using whole genome sequence (WGS) data from the 1000 Genomes Project (KGP) African populations and African Genome Variation Project (AGVP) WGS data (n = 320). If present in African populations, LDlink was used to identify if the *cis*-eQTL is in linkage disequilibrium (LD) with variants in the gene whose expression it is associated with. Finally, drug label and variant annotations associated with the *cis*-eQTLs were derived from PharmGKB.

Results

During sample QC of the genotype dataset, no sex discordances, duplicates, related samples or samples with extreme mean heterozygosity were identified. Four samples had missingness of >5% (GSM954409, GSM954410, GSM954415, GSM954416) and three samples were identified as population outliers (GSM954398, GSM954391, GSM954346). These seven samples were excluded from further analysis. From a total of 318,237 SNPs, 20,652 were excluded from the analysis based on call rates, minor allele frequency and departure from Hardy-Weinberg equilibrium. QC of the gene expression dataset led to the removal of an additional six samples (GSM804707, GSM804739, GSM804673, GSM804662, GSM804715 and GSM804689) from further analysis as they were identified as outliers. Therefore, a total 136 samples were included for further analysis.

When the pipeline was run using GSE39036 and GSE32504, no *cis*-eQTLs were identified for the core ADME genes. However, when run using the imputed genotype dataset and GSE32504, six *cis*-eQTLs for core-ADME genes were identified (Table 1). The six core genes for which *cis*-eQTLs were identified, are all expressed in the liver, and none has previously been identified as eQTLs in GTEx, GeneCards and Ensembl.

From the six eQTLs identified, we found that only the *cis*-eQTL for *UGT1A1*, rs28967009, located 153,640 bps upstream from the TSS, is present in the African populations from the two WGS datasets (Table 2). The

Table 2. Allele frequencies of the eQTL for *UGT1A1* in different African populations as reported in the KGP and the AGVP.

eQTL for *UGT1A1*: rs28967009 (2:234515276)			
Dataset	**Population**	**MAF**	**Sample size**
KGP	GWD	0.031	113
	MSL	0.018	85
	ESN	0.061	99
	YRI	0.060	108
	LWK	0.081	99
	ASW	0.033	61
	ACB	0.042	96
AGVP	ZUL	0.080	100
	BAG	0.125	100
	ETH	0.025	120

ACB and ASW are admixed African populations.
ACB: African–Caribbean in Barbados (admixed); AGVP: African Genome Variation Project; ASW: American's of African Ancestry in SW USA (admixed); BAG: Baganda; ETH: Ethiopia; ESN: Esan in Nigeria; eQTL: Expression quantitative trait loci; GWD: Gambian in Western Divisions in the Gambia; LWK: Luhya in Webuye, Kenya; MAF: Minor allele frequency; MSL: Mende in Sierra Leone; YRI: Yoruba in Ibadan, Nigeria; ZUL: Zulu.

Table 3. Allele frequencies of the eQTL for *UGT1A1* in non-African populations.

eQTL for *UGT1A1*: rs28967009 (2:234515276)			
Dataset	**Population**	**Sub-population**	**MAF**
KGP	SAS	BEB	0.029
		ITU	0.039
		STU	0.044
		PJL	0.068
		GIH	0.029
	EUR	CEU	0
		IBS	0.009
		FIN	0.030
		TSI	0.009
		GBR	0
	AMR	CLM	0.037
		PEL	0.176
		MXL	0.062
		PUR	0.019
	EAS	CHB	0.010
		JPT	0.005
		CHS	0.038
		CDX	0.059
		KHV	0.035

AMR: American; BEB: Bengali in Bangladesh; CDX: Chinese Dai in Xishuangbanna, China; CEU: Utah residents with Northern and Western European ancestry; CHB: Han Chinese in Beijing, China; CHS: Southern Han Chinese, China; CLM: Colombian in Medellin, Colombia; EAS: East Asian; eQTL: Expression quantitative trait loci; EUR: European; FIN: Finnish in Finland; GBR: British in England and Scotland; GIH: Gujarati Indian in Houston, Texas; IBS: Iberian populations in Spain; ITU: Indian Telugu in the UK; JPT: Japanese in Tokyo, Japan; KHV: Kinh in Ho Chi Minh City, Vietnam; MAF: Minor allele frequency; MXL: Mexican Ancestry in Los Angeles, California; PEL: Peruvian in Lima, Peru; PJL: Punjabi in Lahore, Pakistan; PUR: Puerto Rican in Puerto Rico; SAS: South Asian; STU: Sri Lankan Tamil in the UK; TSI: Toscani in Italy.

allele frequencies of the eQTL for *UGT1A1* in the different non-African populations from the KGP dataset are shown in Table 3. This *cis*-eQTL was present in ten different African populations in varying allele frequencies (ranging from 0.025 to 0.125) and is generally lower in non-African populations. It is almost absent in European populations (except for the Finnish population at 0.030), and low in Asian populations (ranging from 0.005 to 0.059). The *cis*-eQTL rs28967009 was not in LD with the *UGT1A1* TA repeat allele variants (**28*, **37* and **36*) in African populations from the KGP (Yoruba in Nigeria [YRI], Luhya in Kenya [LWK], Gambian in Western Gambia [GWD], Mende in Sierra Leone [MSL], Esan in Nigeria [ESN], Americans of African ancestry

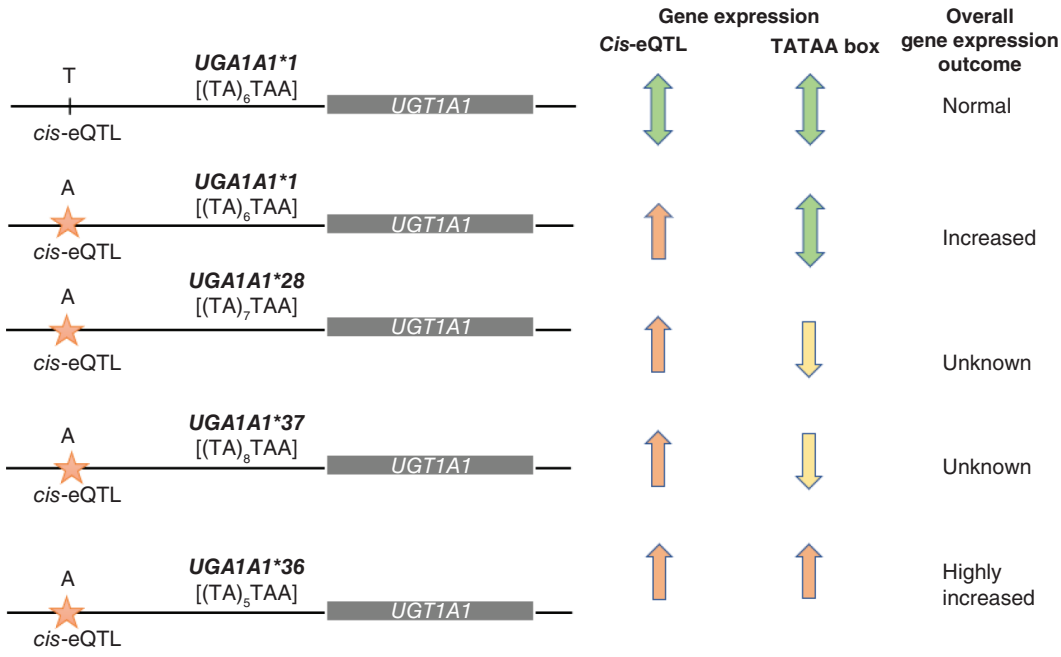

Figure 1. The effect of the variant allele A of the *cis*-eQTL, rs28967009, on *UGT1A1* gene expression in the presence of different star alleles on the same chromosome. The star allele nomenclature is based on biochemical assays of enzyme activity. The star alleles shown here, *UGT1A1*1, *28, *37* and **36* are well studied and occur in the regulatory TATAA Box, affecting the binding of transcription factors and the rate of transcription. These are haplotypes and the combinations of haplotypes in an individual will determine their metabolizer status.
*UGT1A1*1* ($[TA]_6TAA$): Normal enzyme activity; *UGT1A1*36* ($[TA]_5TAA$): Increased activity; *UGT1A1*28* ($[TA]_7TAA$]): Reduced activity; *UGT1A1*37* ($[TA]_8TAA$): Reduced activity.

in south-west United States of Africa [ASW] and African–Caribbeans in Barbados [ACB]). The prediction of this SNP as a *cis*-eQTL is significant with a FDR of 0.006, a p-value of 5.73E-07 and a beta value of 1328. A beta value of 1328 indicates that the variant allele (allele A), is predicted to increase the expression of *UGT1A1* by 1328-times on average. A comprehensive list of variant annotations for *UGT1A1* is available in PharmGKB (https://www.pharmgkb.org/gene/PA420/variantAnnotation).

Discussion

This study identified rs28967009 as a *cis*-eQTL for *UGT1A1*, a core ADME gene located at 2q37 that encodes an enzyme which plays a major role in the conjugation and subsequent elimination of xenobiotics, carcinogens and pharmaceutical drugs [15,16]. This *cis*-eQTL may have important consequences for the metabolism of dolutegravir and irinotecan, which are extensively prescribed in SA for HIV and CRC, respectively. Since the variant allele A, of the *cis*-eQTL is associated with increased *UGT1A1* expression, it may affect the clinical outcome of dolutegravir and irinotecan treatment. Figure 1 illustrates the potential effect the variant allele A could have on *UGT1A1* expression and phenotypic outcome, in the presence of additional alleles that may affect the functional activity of the enzyme.

Irinotecan is widely prescribed in combination with other drugs to treat advanced/metastatic and/or recurrent CRC [17]. It is a prodrug that is metabolized by carboxylesterases in the liver to the active form SN-38, which is 100–1000-times more cytotoxic than irinotecan [18]. UGT1A1 is involved in the glucuronidation of SN-38 and results in the formation of SN-38 glucuronide, the excretable form of the drug [19]. Since rs28967009A, the *UGT1A1* variant allele of the *cis*-eQTL, correlates with greater *UGT1A1* expression levels, this variant is predicted to result in faster/increased glucuronidation of SN-38 to SN-38 glucuronide and this suggests that it is cleared from the body before SN-38 reaches therapeutic levels. Therefore, determining an optimal dose for irinotecan in individuals carrying this variant allele may be more challenging since it has a very narrow therapeutic range. Irinotecan treatment is generally limited by the high incidence of drug-toxicity due to high levels of SN-38 but would not be effective if the circulation levels were below the therapeutic range. The common ADRs of irinotecan include severe neutropenia, fever, asthenia and diarrhea, caused by prolonged exposure to SN-38, and can be severe

enough to lead to the prescription of a reduced drug dose or discontinuation of drug use [16,20]. The rate of severe ADRs associated with irinotecan treatment is around 20–25% [21] and approximately 7% of patients with severe neutropenia and fever following irinotecan treatment die from these complications [22,23]. In general, it is predicted that individuals with the *cis*-eQTL rs28967009A, will have a lower tendency to develop ADRs as the predicted increased expression of *UGT1A1* will result in the drug being cleared from the system faster, but it may also lead to treatment failure.

Dolutegravir is an antiretroviral drug indicated for the treatment of HIV infection in combination with other ARTs. It is primarily metabolized in the liver by UGT1A1 to form dolutegravir glucuronide, which is the excretable form of the drug [24,25]. Dolutegravir is widely used in SA and other developing countries in combination with tenofovir and lamivudine as first-line ART [26–28]. It is highly effective in reducing the HIV viral load thereby lowering the chance of HIV transmission and developing AIDS and HIV-related illnesses or complications like infections and cancer [29]. Since the *cis*-eQTL rs28967009A allele is predicted to increase *UGT1A1* expression, it may likely result in faster clearance of the drug from the body. Dose modification for dolutegravir could be recommended to avoid ineffective doses in individuals with the rs28967009A allele. Common ADRs associated with dolutegravir are neuropsychiatric adverse events like mood changes, depression and anxiety, and gastrointestinal adverse events like nausea and vomiting [30–33]. The presence of the *cis*-eQTL rs28967009A, is hypothesized to lower the chance of developing ADRs as the drug is likely to be cleared faster from the system because of the predicted increase in *UGT1A1* expression and/or may result in treatment failure because therapeutic levels of the drug are not reached.

As variants in the promoter and coding regions of *UGT1A1* are known to affect the expression level and activity of the enzyme, respectively, genetic testing of *UGT1A1* variants is recommended prior to administering any drugs metabolized by *UGT1A1*. Commonly known variants of *UGT1A1* are *UGT1A1*1*, **6*, **2*, **28*, *36, **37* and **60,* with **1, *36*, **28* and **37* occurring in the gene promoter region. In enzyme assays, the *UGT1A1*1* and **60* variants are associated with normal UGT1A1 activity; *UGT1A1 *28* and **37* variants are associated with decreased UGT1A1 activity; and *UGT1A1*36* with increased UGT1A1 activity [34]. *UGT1A1*6* (glycine-71 to arginine) results in reduced enzyme activity and is predominantly found in people of East Asian ancestry [35] and is absent in European and African populations [36–39]. *UGT1A1*27* is a cytosine to adenine change in codon 229 (686C>A) that is in LD with *UGT1A1*28* which results in reduced enzyme activity [40]. *UGT1A1*28* is commonly found in Africans with an allele frequency of 0.39–0.40 and in African–Americans with an allele frequency of 0.42–0.45 [41], with a predicted frequency of homozygotes of 0.16. Similarly, the frequency of *UGT1A1**28 homozygosity in Europeans is found to be in the range 0.09–0.16 [42]. The *UGT1A1*36* (increased UGT1A1 activity) and **37* (decreased UGT1A1 activity) alleles occur almost exclusively in populations of African origin, with estimated allele frequencies of 0.03–0.10 and 0.02–0.07, respectively [38,39,43]. The joint presence of these variants on the same haplotype has not been investigated and the potential outcomes on enzyme activity would need to be investigated (Figure 1).

From a phenotypic viewpoint, depending on which two alleles of *UGT1A1* a person has, they are classified as extensive metabolizer (EM), intermediate metabolizer (IM) or poor metabolizer (PM). An EM has two functional *UGT1A1* alleles, two increased function alleles or a functional allele and an increased function allele. An IM has a functional allele or an increased function allele with a decreased function allele. A PM has two decreased function alleles [34]. Individuals with increased UGT1A1 enzyme activity have lower circulating drug concentrations and may only reach sub-therapeutic levels at standard dosage, requiring higher doses to achieve efficacy. UGT1A1 IMs and PMs are at higher risk of side effects as the drug remains in circulation for extended periods of time because of poor/reduced UGT1A1 enzyme activity. Drug dosage must be lowered in such individuals to prevent the occurrence of ADRs. With the aforementioned in mind, individuals with one or two copies of the rs28967009A allele identified in our study, are predicted to be EMs, either requiring increased drug doses or alternative treatment due to an inability to achieve therapeutic levels. PharmGKB documents that the European Medicines Agency, Health Canada/Santé Canada, Swiss Agency of Therapeutic Products and United States Food and Drug Administration indicate that the starting dose of irinotecan must be reduced by at least one level in patients homozygous for the low function allele, *UGT1A1*28,* and in patients who have experienced hematologic toxicity with previous treatment to avoid/reduce risk for hematologic toxicity.

The effect of the variant allele A of the *cis*-eQTL rs28967009 on UGT1A1 PMs, IMs and EMs is difficult to predict as the presence of multiple and possibly opposing effects of regulatory variants on the same haplotype is unknown (Figure 1). In UGT1A1 PMs and IMs with at least one rs28967009A allele, it is expected that ADRs

would be low as the predicted increase in *UGT1A1* expression would clear the toxic active metabolites of drugs like irinotecan and dolutegravir faster, however this may reduce the therapeutic effect. In UGT1A1 EMs with at least one rs28967009A allele, dosage levels of irinotecan and dolutegravir may need to be increased to reach therapeutic levels as a result of the rapid clearance of the drugs likely making them ineffective. If increased doses are still likely to lead to a failure of drug efficacy, an alternate treatment strategy to irinotecan and dolutegravir should be recommended.

In addition to dolutegravir and irinotecan, which currently have drug label annotations for *UGT1A1* variants, but not the one identified by our study, the efficacy of various other drugs is also affected by variants in *UGT1A1.* These include atazanavir and ritonavir that are used as combination ART for the treatment of HIV in SA [44]. Atazanavir inhibits UGT1A1, thereby preventing the glucuronidation and elimination of bilirubin. This causes hyperbilirubinemia and eventually jaundice and can culminate in discontinuation of atazanavir treatment. Risk for bilirubin-related discontinuation of atazanavir is highest among UGT1A1 PMs [20,34,45]. Variants in *UGT1A1* are associated with increased risk of nephrolithiasis (rs8330G, rs1042640G, rs10929303T), increased likelihood of developing hyperbilirubinemia (rs887829T), increased chance of discontinuation of atazanavir treatment (rs887829T) and increased metabolism of atorvastatin (rs887829T), a drug used in the treatment of cardiovascular disease [46–49]. rs4148323A is associated with decreased metabolism of SN-38 in people with neoplasms, decreased overall survival and progression free survival [50–52]. *UGT1A1*28* is associated with slower elimination of raloxifene (prescribed to reduce the risk of invasive breast cancer) and development of hyperbilirubinemia when treated with tranilast, an antiallergic medication [38,53,54]. *UGT1A1*6* allele is associated with lower clearance of etoposide (an anticancer agent) in people of African ancestry [55]. Since the novel *cis*-eQTL described in this paper, rs28967009A, is predicted to increase *UGT1A1* expression, its presence along with existing variants of *UGT1A1* that increase susceptibility to ADRs, may lower the risk of developing these ADRs. For example, as rs887829T is associated with increased metabolism of atorvastatin, an alternate drug may need to be prescribed to people who have both rs887829T and the A allele of the *cis*-eQTL as the drug is postulated to be cleared much faster due to increased *UGT1A1* expression.

A limitation of this study is that it is based on one genotype-gene expression dataset pair because of unavailability of other datasets in the public domain. Therefore, the findings of this analysis must be replicated by additional studies using more datasets. In addition, the effect of the variant allele A of the *cis*-eQTL rs28967009 needs to be clinically and functionally validated before being included in drug label annotations. If clinically validated, changes in drug dose or the use of alternate drugs could be recommended for this variant.

Conclusion

UGT1A1 is a core ADME gene that plays a major role in the metabolism of drugs like dolutegravir and irinotecan that are prescribed extensively in SA to treat HIV and CRC, respectively. Therapeutic levels of the drugs irinotecan and dolutegravir may not be reached in EMs with the rs28967009A variant in *UGT1A1*, which is present at appreciable frequencies in different African populations, including those in SA. Furthermore, UGT1A1 PMs and IMs who have at least one rs28967009A allele may have a reduced likelihood of developing ADRs but may not reach therapeutic drug levels of the active compound of the drugs. Since dolutegravir is an ART that is used extensively in SA because of the high disease burden of HIV and because CRC is the fourth most common cancer in SA, understanding the effect of the novel *cis*-eQTL on the metabolism of drugs commonly used for these disorders is of high importance to attain optimal drug efficacy levels. Pharmacogenomic testing of this variant prior to administering dolutegravir or irinotecan could present a novel strategy to identify individuals who may be poor responders due to elevated drug clearance. As such, this study identifies a sub-population of Africans that may benefit from alternative therapies.

Financial & competing interests disclosure

JM Mathew was supported by a grant holder linked bursary from the South African National Research Foundation. M Ramsay is a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology, and administered by the National Research Foundation. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Summary points

Background

- Despite the high disease burden of human immunodeficiency virus infection and colorectal cancer in South Africa, treatment-relevant pharmacogenetic variants are understudied.
- Expression quantitative trait loci (eQTLs) are genetic variants that contribute to variation in gene expression levels.
- eQTLs can assist in the prediction of the clinical outcome of treatment should they affect the expression of genes that modulate drug absorption, distribution, metabolism and excretion (ADME).
- The frequencies of eQTLs of ADME genes vary both within and between populations. African ancestry populations show the most inter-population variation for known genetic variants and East Asian ancestry populations generally show the least.
- eQTLs of ADME genes can be correlated with the level of drug efficacy and an individual's propensity toward developing adverse drug reactions.
- eQTLs of ADME genes could therefore be used to test relevant pharmacogenetic associations with drugs that are metabolized by the corresponding enzyme(s).

Materials & methods

- A bioinformatic pipeline was built that identifies eQTLs for ADME genes from any tissue type using publicly available genotype and gene expression data (https://github.com/phelelani/nf-eqtl). In this study, the focus was on liver eQTLs since ADME genes are abundantly expressed in this organ.

Results

- A novel *cis*-eQTL, rs28967009, was identified for *UGT1A1,* with a false discovery rate of 0.006, a p-value of 5.73E-07 and a beta value of 1328. The variant allele A of the *cis*-eQTL is predicted to upregulate *UGT1A1* expression by 1328-times on average.
- This *cis*-eQTL was present in 10 different African populations in varying allele frequencies (ranging from 0.025 to 0.125) and is found in lower frequencies in non-African populations. It is almost absent in European populations (except for the Finnish population at 0.030), and low in Asian populations (ranging from 0.005 to 0.059).
- The *cis*-eQTL rs28967009 was not in linkage disequilibrium with the *UGT1A1* TA repeat allele variants (**28*, **37* and **36*) in African populations from the 1000 Genomes Project.

Conclusion

- Irinotecan is a prodrug prescribed for colorectal cancer treatment that is metabolized by carboxylesterases in the liver to SN-38, its active form. UGT1A1 glucuronidates SN-38 and forms SN-38 glucuronide, the excretable form of the drug. Since rs28967009A correlates with greater *UGT1A1* expression levels, it is predicted to result in increased glucuronidation of SN-38 to SN-38 glucuronide thereby suggesting that it is cleared from the body before SN-38 reaches therapeutic levels.
- Dolutegravir is an antiretroviral drug, used for HIV treatment, that is metabolized in the liver by UGT1A1 to form dolutegravir glucuronide, the excretable form of the drug. Since rs28967009A is predicted to increase *UGT1A1* expression, it may likely result in faster clearance of the drug from the body.
- In UGT1A1 poor and intermediate metabolizers with at least one rs28967009A allele, the predicted increase in *UGT1A1* expression is expected to lower adverse drug reaction occurrence as the toxic metabolites of irinotecan and dolutegravir would be cleared faster. However, this may reduce the therapeutic effect.
- In UGT1A1 extensive metabolizers with at least one rs28967009A allele, dosage levels of irinotecan and dolutegravir may need to be increased to reach therapeutic levels as the rapid clearance of the drugs might likely make them ineffective. If increased doses are still likely to lead to a failure of drug efficacy, an alternate treatment strategy to irinotecan and dolutegravir should be recommended.
- As increased *UGT1A1* expression could affect the clinical outcome of dolutegravir and irinotecan treatment by increasing drug clearance, patients with the rs28967009A variant may require increased drug doses to reach therapeutic levels or should be prescribed alternative drugs.

References

1. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos. Trans. RSoc. B Biol. Sci.* 368(1620), (2013).
2. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28(10), 1353–1358 (2012).
3. Hovelson DH, Xue Z, Zawistowski M *et al.* Characterization of ADME gene variation in 21 populations by exome sequencing. *Pharmacogenet. Genomics* 27(3), 89–100 (2017).
4. Ramos E, Doumatey A, Elkahloun AG *et al.* Pharmacogenomics, ancestry and clinical decision making for global populations. *Pharmacogenomics J.* 14(3), 217–222 (2014).
5. Sengupta D, Choudhury A, Fortes-Lima C *et al.* Genetic-substructure and complex demographic history of South African Bantu speakers. *Nat. Comm.* 12, 2080 (2021).

6. Lee H-K, Hu M, Lui SS, Ho C-S, Wong C-K, Tomlinson B. Effects of polymorphisms in ABCG2, SLCO1B1, SLC10A1 and CYP2C9/19 on plasma concentrations of rosuvastatin and lipid response in Chinese patients. *Pharmacogenomics* 14(11), 1283–1294 (2013).
7. Birmingham BK, Bujac SR, Elsby R *et al.* Rosuvastatin pharmacokinetics and pharmacogenetics in Caucasian and Asian subjects residing in the United States. *Eur. J. Clin. Pharmacol.* 71(3), 329–340 (2015).
8. Iskakova AN, Romanova AA, Aitkulova AM, Sikhayeva NS, Zholdybayeva EV, Ramanculov EM. Polymorphisms in genes involved in the absorption, distribution, metabolism, and excretion of drugs in the Kazakhs of Kazakhstan. *BMC Genet.* 17(1), (2016).
9. Hrdlickova B, de Almeida RC, Borek Z, Withoff S. Genetic variation in the non-coding genome: involvement of micro-RNAs and long non-coding RNAs in disease. *Biochim. Biophys. Acta – Mol. Basis Dis.* 1842(10), 1910–1922 (2014).
10. Gona PN, Gona CM, Ballout S *et al.* Burden and changes in HIV/AIDS morbidity and mortality in Southern Africa Development Community Countries, 1990–2017. *BMC Public Health* 20(1), 867 (2020).
11. Satoh S, Boyer E. HIV in South Africa. *Lancet* 394(10197), 467 (2019).
12. Brand M, Gaylard P, Ramos J. Colorectal cancer in South Africa: an assessment of disease presentation, treatment pathways and 5-year survival. *South African Med. J.* 108(2), 118 (2018).
13. Herbst C, Miot JK, Moch SL, Ruff P. Colorectal Cancer (CRC) treatment and associated costs in the public sector compared to the private sector in Johannesburg, South Africa. *BMC Health Serv. Res.* 20(1), 290 (2020).
14. Schröder A, Klein K, Winter S *et al.* Genomics of ADME gene expression: mapping expression quantitative trait loci relevant for absorption, distribution, metabolism and excretion of drugs in human liver. *Pharmacogenomics J.* 13(1), 12–20 (2013).
15. Schulz C, Boeck S, Heinemann V, Stemmler H-J. UGT1A1 genotyping: a predictor of irinotecan-associated side effects and drug efficacy? *Anticancer Drugs* 20(10), 867–879 (2009).
16. Barbarino JM, Haidar CE, Klein TE, Altman RB. PharmGKB summary: very important pharmacogene information for UGT1A1. *Pharmacogenet. Genomics* 24(3), 177–183 (2014).
17. Ramesh M, Ahlawat P, Srinivas NR. Irinotecan and its active metabolite, SN-38: review of bioanalytical methods and recent update from clinical pharmacology perspectives. *Biomed. Chromatogr.* 24(1), 104–123 (2010).
18. Mathijssen RHJ, van Alphen RJ, Verweij J *et al.* Clinical pharmacokinetics and metabolism of irinotecan (CPT-11). *Clin. Cancer Res.* 7(8), 2182 LP–2194 (2001).
19. Dias MM, McKinnon RA, Sorich MJ. Impact of the UGT1A1*28 allele on response to irinotecan: a systematic review and meta-analysis. *Pharmacogenomics* 13(8), 889–899 (2012).
20. Lankisch TO, Schulz C, Zwingers T *et al.* Gilbert's syndrome and irinotecan toxicity: combination with UDP-glucuronosyltransferase 1A7 variants increases risk. *Cancer Epidemiol. Biomarkers Prev.* 17(3), 695 LP–701 (2008).
21. Tam VC, Rask S, Koru-Sengul T, Dhesy-Thind S. Generalizability of toxicity data from oncology clinical trials to clinical practice: toxicity of irinotecan-based regimens in patients with metastatic colorectal cancer. *Curr. Oncol.* 16(6), 13–20 (2009).
22. Obradovic M, Mrhar A, Kos M. Cost–effectiveness of UGT1A1 genotyping in second-line, high-dose, once every 3 weeks irinotecan monotherapy treatment of colorectal cancer. *Pharmacogenomics* 9(5), 539–549 (2008).
23. Liu X, Cheng D, Kuang Q, Liu G, Xu W. Association of UGT1A1*28 polymorphisms with irinotecan-induced toxicities in colorectal cancer: a meta-analysis in Caucasians. *Pharmacogenomics J.* 14(2), 120–129 (2014.)
24. Castellino S, Moss L, Wagner D *et al.* Metabolism, excretion, and mass balance of the HIV-1 integrase inhibitor dolutegravir in humans. *Antimicrob. Agents Chemother.* 57(8), 3536–3546 (2013).
25. Yagura H, Watanabe D, Kushida H *et al.* Impact of UGT1A1 gene polymorphisms on plasma dolutegravir trough concentrations and neuropsychiatric adverse events in Japanese individuals infected with HIV-1. *BMC Infect. Dis.* 17(1), 622 (2017).
26. Walmsley SL, Antela A, Clumeck N *et al.* Dolutegravir plus abacavir–lamivudine for the treatment of HIV-1 infection. *N. Engl. J. Med.* 369(19), 1807–1818 (2013).
27. Moorhouse MA, Carmona S, Davies N *et al.* Southern African HIV Clinicians Society Guidance on the use of dolutegravir in first-line antiretroviral therapy. *South African J. HIV Med.* 19(1), (2018).
28. Venter WDF, Moorhouse M, Sokhela S *et al.* Dolutegravir plus two different prodrugs of tenofovir to treat HIV. *N. Engl. J. Med.* 381(9), 803–815 (2019).
29. Spano J-P, Poizot-Martin I, Costagliola D *et al.* Non-AIDS-related malignancies: expert consensus review and practical applications from the multidisciplinary CANCERVIH Working Group. *Ann. Oncol.* 27(3), 397–408 (2016).
30. Elzi L, Erb S, Furrer H *et al.* Adverse events of raltegravir and dolutegravir. *AIDS* 31(13), 1853–1858 (2017).
31. Lepik KJ, Yip B, Ulloa AC *et al.* Adverse drug reactions to integrase strand transfer inhibitors. *AIDS* 32(7), (2018).
32. Cuzin L, Pugliese P, Katlama C *et al.* Integrase strand transfer inhibitors and neuropsychiatric adverse events in a large prospective cohort. *J. Antimicrob. Chemother.* 74(3), 754–760 (2019).

33. Llibre JM, Montoliu A, Miró JM *et al.* Discontinuation of dolutegravir, elvitegravir/cobicistat and raltegravir because of toxicity in a prospective cohort. *HIV Med.* 20(3), 237–247 (2019).
34. Gammal RS, Court MH, Haidar CE *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for UGT1A1 and atazanavir prescribing. *Clin. Pharmacol. Ther.* 99(4), 363–369 (2016).
35. Teh LK, Hashim H, Zakaria ZA, Salleh MZ. Polymorphisms of UGT1A1*6, UGT1A1*27 & UGT1A1*28 in three major ethnic groups from Malaysia. *Indian J. Med. Res.* 136(2), 249–259 (2012).
36. Dai X, Wu C, He Y *et al.* A genome-wide association study for serum bilirubin levels and gene-environment interaction in a Chinese Population. *Genet. Epidemiol.* 37(3), 293–300 (2013).
37. Liu JY, Qu K, Sferruzza AD, Bender RA. Distribution of the UGT1A1*28 polymorphism in Caucasian and Asian populations in the US: a genomic analysis of 138 healthy individuals. *Anticancer Drugs* 18(6), 693–696 (2007).
38. Palomaki GE, Bradley LA, Douglas MP, Kolor K, Dotson WD. Can UGT1A1 genotyping reduce morbidity and mortality in patients with metastatic colorectal cancer treated with irinotecan? An evidence-based review. *Genet. Med.* 11(1), 21–34 (2009).
39. Dean L. Irinotecan therapy and UGT1A1 genotype drug. In: *Medical Genetics Summaries [Internet].* Pratt VM, Scott SA, Pirmohamed M *et al.* (Eds). National Center for Biotechnology Information, Bethseda, MA, USA (2018).
40. Fukuda M, Suetsugu T, Shimada M *et al.* Prospective study of the UGT1A1*27 gene polymorphism during irinotecan therapy in patients with lung cancer: results of Lung Oncology Group in Kyusyu (LOGIK1004B). *Thorac. Cancer* 7(4), 467–472 (2016).
41. Sukasem C, Atasilp C, Chansriwong P, Chamnanphon M, Puangpetch A, Sirachainan E. Development of pyrosequencing method for detection of UGT1A1 polymorphisms in Thai colorectal cancers. *J. Clin. Lab. Anal.* 30(1), 84–89 (2016).
42. Marques SC, Ikediobi ON. The clinical application of UGT1A1 pharmacogenetic testing: gene–environment interactions. *Hum. Genomics* 4(4), 238–249 (2010).
43. Beutler E, Gelbart T, Demina A. Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? *Proc. Natl Acad. Sci. USA* 95(14), 8170–8174 (1998).
44. Achenbach CJ, Darin KM, Murphy RL, Katlama C. Atazanavir/ritonavir-based combination antiretroviral therapy for treatment of HIV-1 infection in adults. *Future Virol.* 6(2), 157–177 (2011).
45. Turatti L, Sprinz E, Lazzaretti RK *et al.* Short Communication: UGT1A1*28 variant allele is a predictor of severe hyperbilirubinemia in HIV-infected patients on HAART in Southern Brazil. *AIDS Res. Hum. Retroviruses* 28(9), 1015–1018 (2011).
46. Johnson DH, Venuto C, Ritchie MD *et al.* Genomewide association study of atazanavir pharmacokinetics and hyperbilirubinemia in AIDS Clinical Trials Group protocol A5202. *Pharmacogenet. Genomics* 24(4), 195–203 (2014).
47. Nishijima T, Tsuchiya K, Tanaka N *et al.* Single-nucleotide polymorphisms in the UDP-glucuronosyltransferase 1A-3′ untranslated region are associated with atazanavir-induced nephrolithiasis in patients with HIV-1 infection: a pharmacogenetic study. *J. Antimicrob. Chemother.* 69(12), 3320–3328 (2014).
48. Vardhanabhuti S, Ribaudo HJ, Landovitz RJ *et al.* Screening for UGT1A1 genotype in study A5257 would have markedly reduced premature discontinuation of atazanavir for hyperbilirubinemia. *Open forum Infect. Dis.* 2(3), ofv085–ofv085 (2015).
49. Turner RM, Fontana V, Zhang JE *et al.* A genome-wide association study of circulating levels of atorvastatin and its major metabolites. *Clin. Pharmacol. Ther.* 108(2), 287–297 (2020).
50. Han J-Y, Lim H-S, Shin ES *et al.* Comprehensive analysis of UGT1A polymorphisms predictive for pharmacokinetics and treatment outcome in patients with non-small-cell lung cancer treated with irinotecan and cisplatin. *J. Clin. Oncol.* 24(15), 2237–2244 (2006).
51. Jada SR, Lim R, Wong CI *et al.* Role of UGT1A1*6, UGT1A1*28 and ABCG2 c.421C>A polymorphisms in irinotecan-induced neutropenia in Asian cancer patients. *Cancer Sci.* 98(9), 1461–1467 (2007).
52. Takano M, Kato M, Yoshikawa T *et al.* Clinical significance of UDP-Glucuronosyltransferase 1A1*6 for toxicities of combination chemotherapy with irinotecan and cisplatin in gynecologic cancers. *Oncology* 76(5), 315–321 (2009).
53. Danoff TM, Campbell DA, McCarthy LC *et al.* A Gilbert's syndrome UGT1A1 variant confers susceptibility to tranilast-induced hyperbilirubinemia. *Pharmacogenomics J.* 4(1), 49–53 (2004).
54. Trontelj J, Marc J, Zavratnik A, Bogataj M, Mrhar A. Effects of UGT1A1*28 polymorphism on raloxifene pharmacokinetics and pharmacodynamics. *Br. J. Clin. Pharmacol.* 67(4), 437–444 (2009).
55. Kishi S, Yang W, Boureau B *et al.* Effects of prednisone and genetic polymorphisms on etoposide disposition in children with acute lymphoblastic leukemia. *Blood* 103(1), 67–72 (2004).