

## ARTICLE OPEN



# Structural variation of the coding and non-coding human pharmacogenome

Roman Tremmel<sup>1,2,5</sup>, Yitian Zhou<sup>3,5</sup>, Matthias Schwab<sup>1,2,4</sup> and Volker M. Lauschke<sup>1,2,3</sup>✉

Genetic variants in drug targets and genes encoding factors involved in drug absorption, distribution, metabolism and excretion (ADME) can have pronounced impacts on drug pharmacokinetics, response, and toxicity. While the landscape of genetic variability at the level of single nucleotide variants (SNVs) has been extensively studied in these pharmacogenetic loci, their structural variation is only poorly understood. Thus, we systematically analyzed the genetic structural variability across 908 pharmacogenes (344 ADME genes and 564 drug targets) based on publicly available whole genome sequencing data from 10,847 unrelated individuals. Overall, we extracted 14,984 distinct structural variants (SVs) ranging in size from 50 bp to 106 Mb. Each individual harbored on average 10.3 and 1.5 SVs with putative functional effects that affected the coding regions of ADME genes and drug targets, respectively. In addition, by cross-referencing pharmacogenomic SVs with experimentally determined binding data of 224 transcription factors across 130 cell types, we identified 1276 non-coding SVs that overlapped with gene regulatory elements. Based on these data, we estimate that non-coding structural variants account for 22% of the genetically encoded pharmacogenomic variability. Combined, these analyses provide the first comprehensive map of structural variability across pharmacogenes, derive estimates for the functional impact of non-coding SVs and incentivize the incorporation of structural genomic data into personalized drug response predictions.

*npj Genomic Medicine* (2023)8:24; <https://doi.org/10.1038/s41525-023-00371-y>

## INTRODUCTION

Inter-individual variability in drug response has long been recognized as a major problem in pharmacological treatment. Overall, it is estimated that around 50% of patients experience a lack of efficacy or adverse drug reactions (ADRs), contributing to considerable patient morbidity and mortality<sup>1</sup>. In addition to posing a significant burden on the healthcare system, lack of drug efficacy and ADRs are major hurdles to drug development. More than 80% of candidate drugs fail in clinical trials and around 32% of FDA-approved therapeutics are affected by post-market safety events<sup>2,3</sup>. Mechanistically, variable drug responses can stem from variability in drug disposition or altered pharmacodynamics.

Heritable factors play an important role in differential drug response and genetic variability, including variations in genes modulating drug pharmacokinetics as well as drug targets, explain approximately 20–30% of inter-individual phenotypic differences<sup>4</sup>. Among these, single nucleotide variants (SNVs) have been extensively studied as biomarkers to predict drug efficacy and ADRs. A multitude of such variants in genes involved in drug absorption, distribution, metabolism and excretion (ADME) has been included in the pharmacogenomic guidelines to individualize pharmacological treatment based on patient genotypes<sup>5–7</sup>. Comparatively less is known about the functional effects of pharmacogenetic drug target variability. While the landscape of SNVs in drug targets has been systematically analyzed<sup>8</sup> and elegant recent studies demonstrated striking effects of SNVs on intracellular signal transduction and drug action<sup>9,10</sup>, more evidence is required to enable the translation of such variations into clinical recommendations.

In contrast to SNVs, structural variations (SVs), defined as genomic deletions, duplications, insertions, inversions and other

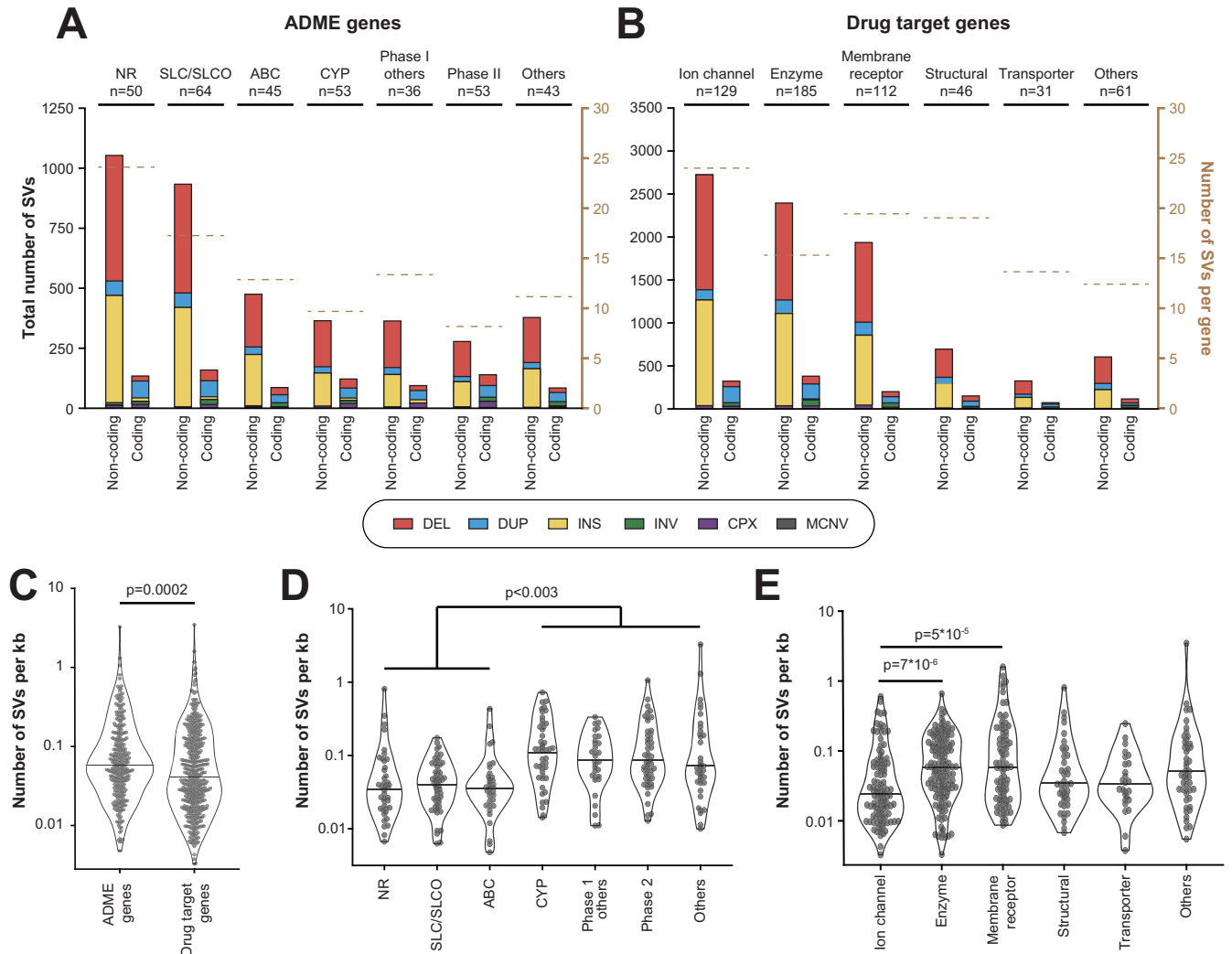
complex rearrangements that affect >50 bp, are substantially less studied<sup>11,12</sup>. While the total number of SVs per human genome is around two orders of magnitude lower than for SNVs (34,000 SVs compared to 3 million SNVs), SVs affect 3.4 times more nucleotides in both coding and non-coding regions of the genome<sup>13</sup> and constitute important contributors to human phenotypes<sup>14–16</sup>. Copy number variations (CNVs) in some ADME genes are well described<sup>17,18</sup>, whereas the structural variability of human drug targets has not been systematically analyzed. Furthermore, comprehensive analyses of non-coding structural variability in pharmacogenes have not been presented. Here, we systematically profiled the landscape of structural variability across 908 pharmacogenes (344 ADME genes and 564 drug targets) based on whole genome sequencing (WGS) data from 10,847 unrelated individuals<sup>19</sup>. Our analyses refine previous SV frequency estimates and, by integrating structural data with experimentally determined transcription factor binding site (TFBS) information, identify a catalog of 1276 SVs that impact pharmacogenetic regulatory elements.

## RESULTS

### The structural variome in genes involved in drug disposition and drug targets

We first analyzed the structural variability of 344 genes involved in ADME processes. The highest number of SVs was found in nuclear receptors ( $n = 1207$ ; average of 24 SVs per gene) and SLC/SLCO transporters ( $n = 1112$ ; average of 17 SVs per gene), whereas SV numbers in phase II enzymes were around 3-fold lower ( $n = 437$ ; 8 SVs per gene; Fig. 1A). Additionally, we analyzed the structural variome in 564 genes encoding the therapeutic targets of 1578

<sup>1</sup>Dr Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart, Germany. <sup>2</sup>University Tübingen, Tübingen, Germany. <sup>3</sup>Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden. <sup>4</sup>Departments of Clinical Pharmacology and Pharmacy and Biochemistry, University Tübingen, Tübingen, Germany. <sup>5</sup>These authors contributed equally: Roman Tremmel, Yitian Zhou. ✉email: [volker.lauschke@ikp-stuttgart.de](mailto:volker.lauschke@ikp-stuttgart.de)



**Fig. 1 Overview of structural variability in the human pharmacogenome.** Number of structural variants (SVs) in different classes of genes that are involved in drug absorption, distribution, metabolism and excretion (ADME; **A**) as well as drug target genes (**B**). **C** The number of SVs per kilo base (kb) gene length differ significantly between ADME genes and drug targets. **D** Among ADME genes, CYPs, as well as phase 1 and phase 2 enzymes harbor significantly more SVs than nuclear receptors and transporters. **E** Among drug target genes, ion channels were significantly less variable than enzymes and membrane receptors. NR nuclear receptors, CYP cytochrome P450s, SLC solute carrier transporters, ABC ATP binding cassette transporters, CPX complex rearrangement, DEL deletion, DUP duplication, INS insertion, INV inversion, MCNV multi-copy number variations.

clinically approved drugs. Most SVs were identified in ion channels ( $n = 3112$ ; 24 SVs per gene) and membrane receptors ( $n = 2840$ ; 19 SVs per gene), whereas the variability in transporter targets was markedly lower ( $n = 427$ ; 14 SVs per gene; Fig. 1B). *PTGS2* ( $n = 189$ ), *GPD2* ( $n = 150$ ), *HCN1* ( $n = 145$ ) and *KCND2* ( $n = 145$ ) featured the most SVs whereas 41 pharmacogenes did not harbor any structural variations (Supplementary Table 3). When normalizing for gene length, ADME genes carried significantly more SVs per kilo base than drug targets (Fig. 1C). The higher variability was primarily driven by genes encoding drug metabolizing enzymes (CYPs, as well as other phase 1 and phase 2 enzymes), whereas transporter genes and nuclear receptors were significantly less variable and harbored similar numbers than drug target genes (Fig. 1D, E).

SVs range in size from 50 bp to 106 Mb with a median size of 312 bp (Supplementary Fig. 1A). Drug target SVs were overall significantly shorter than SVs in ADME genes (281 bp vs 321 bp;  $p < 0.0001$ ). The overall largest SVs (106 Mb) was a singleton complex rearrangement of duplications and inversions that affected almost the complete chromosome 10 covering a total

of 589 genes, as well as a rare duplication on chromosome 5 that affected the target genes *IL6ST*, *GHR*, *HCN1*, *NDUFA2*, *NDUFS4*, *PDE4D*, *PTGER4* (28 Mb). The longest deletions affected the GABA receptor cluster encoding *GABRA1*, *GABRA6*, and *GABRG2* on chromosome 5 (6.5 Mb) and the ADME gene *COMT* on chromosome 22 (2.5 Mb). Insertions and deletions had median sizes of 208–618 bp, whereas the average inversions were more than 10,000 times larger with a median size of 30.2 Mb (Supplementary Fig. 1B–G). Furthermore, both ADME and drug target SVs were significantly smaller than SVs in olfactory genes ( $p < 0.0001$ ), which were selected as one of the most polymorphic human gene families due to low selective pressure<sup>20</sup>.

#### Functional consequences of coding pharmacogenomic structural variability

Of all 14,984 pharmacogenomic SVs, 2198 impacted gene exons, whereas the remainder affected introns, or non-coding regions up- and downstream of the gene body (Fig. 2A). To interpret SV functionality, we classified deletions spanning coding regions as



**Fig. 2 The landscape of functional SVs across the pharmacogenome. A** Across all identified SVs in pharmacogenes, 15% affected exons (blue) and 85% were non-coding (gray). Of the exonic variations, functional consequences could be inferred for one third (dark blue), whereas the functional consequences of the remainder were unknown (mostly SVs in UTRs or in-frame deletions/duplications). **B** Schematic showing the different SV classes causing loss-of-function (LOF) or increased gene dosage (IGD) of the corresponding gene. **C** Fraction of rare and common SVs for the different ADME gene families. **D** Allele frequencies for structural LOF and IGD variants across ADME genes. The sizes of dots indicate the length of the corresponding SVs. Multi CNVs are indicated in red. **E** Bar plots show the average number of LOF and IGD SVs in ADME genes per individual for the entire dataset (global) and for individual ethnogeographic groups. **F** Fraction of rare and common SVs in the different drug target gene classes. **G** Allele frequencies are shown for structural LOF and IGD variants across drug targets. The sizes of dots indicate the length of the corresponding SVs. Multi CNVs are indicated in red. **H** Bar plots show the number of LOF and IGD SVs in drug target genes per individual for the entire dataset (global) and for individual ethnogeographic groups. CPX complex rearrangement, DEL deletion, DUP duplication, INS insertion, INV inversion, AFR African, AMR admixed Americans, EAS East Asians, EUR Europeans.

**Table 1.** Common functional coding SVs in pharmacogenes with minor allele frequencies above 1%.

Gene	Type	Function	Length (bp)	Frequency	Drug associations
<i>ADME genes</i>					
<i>GSTM1</i>	MCNV (DEL)	LOF	19000	84.5%	Cisplatin, nevirapine, azathioprine
<i>GSTT1</i>	MCNV (DEL)	LOF	54150	71.8%	Cisplatin, thalidomide, clozapine
<i>UGT2B17</i>	MCNV (DEL)	LOF	117000	56%	Ibuprofen, exemestane
<i>GSTT2B</i>	DEL	LOF	37153	46.2%	Cisplatin
<i>SULT1A1</i>	MCNV (DUP)	IGD	13000	45.1%	Acetaminophen, minoxidil, estrogens
<i>SULT1A4</i>	MCNV (DUP)	IGD	50000	37.2%	Salbutamol, acetaminophen, morphine
<i>CES1</i>	MCNV (DUP)	IGD	17500	25.6%	Clopidogrel, irinotecan, methylphenidate
<i>UGT2B28</i>	DEL	LOF	109294	21.5%	Androgens
<i>CYP2D6</i>	MCNV (DUP)	IGD	12202	18.8%	Antidepressants, antipsychotics, opioids, tamoxifen
<i>CES1</i>	DUP	LOF	6300	14.4%	Clopidogrel, irinotecan
<i>CYP3A43</i>	DEL	LOF	2171	9.9%	Erythromycin, olanzapine
<i>CYP2D6</i>	MCNV (DEL)	LOF	12202	7.8%	Antidepressants, antipsychotics, opioids, tamoxifen
<i>CYP2E1</i>	DUP	IGD	91684	5%	Acetaminophen, ethanol, anti-tuberculosis drugs
<i>CYP2A6</i>	DEL	LOF	30617	3.6%	Nicotine
<i>CYP4F12</i>	DUP	IGD	53300	2.8%	Astemizole
<i>CYP2E1</i>	DUP	LOF	9600	2.6%	Acetaminophen, ethanol, anti-tuberculosis drugs
<i>SULT1A1</i>	MCNV (DEL)	LOF	13000	2.6%	Acetaminophen, minoxidil, estrogens
<i>CYP2A6</i>	DUP	IGD	30939	1.1%	Nicotine
<i>CYB5R3</i>	DEL	LOF	511	1%	Metoclopramide, lidocaine, dapsone
<i>UGT2B10</i>	DUP	IGD	37000	1%	Amitriptyline, olanzapine, nicotine
<i>Drug target genes</i>					
<i>LAMB4</i>	DEL	LOF	254	30.6%	Ocriplasmin
<i>AMY2A</i>	MCNV (DUP)	IGD	165900	12.6%	Acarbose
<i>MGAM</i>	DEL	LOF	25985	11.8%	Miglitol, acarbose, voglibose
<i>LAMA2</i>	DEL	LOF	835	11.5%	Ocriplasmin
<i>AMY2A</i>	MCNV (DEL)	LOF	165900	6.2%	Acarbose

DEL deletion, DUP duplication, IGD increased gene dosage, LOF loss of function, MCNV multi-allelic copy number variation.

other phase I ( $n = 2$ ) or phase II enzymes ( $n = 11$ ) were common (Fig. 2C). LOF SVs with high frequency were identified in *GSTM1* (84.5% deletion frequency), *GSTT1* (71.8% deletion frequency), *UGT2B17* (56% deletion frequency), *UGT2B28* (21.5% deletion frequency) and *CYP2D6* (7.8% deletion frequency; Fig. 2D and Table 1). Similarly, common IGD SVs were found in *SULT1A1* (45.1% duplication frequency), *SULT1A4* (37.2% duplication frequency), *CES1* (25.6% duplication frequency) and *CYP2D6* (18.8% duplication frequency). In aggregate, each individual harbored on average 7.9 LOF and 2.4 IGD SVs in ADME genes, which might contribute to inter-individual differences in response to medications metabolized or transported by the respective gene products (Fig. 2E). Notably, East Asians harbored most (11.7 per individual) and Europeans the least (9.4 per individual) functional coding SVs in ADME genes.

For pharmacodynamic drug targets, more than 95% of all coding SVs were rare with the only exceptions being found in structural genes (laminins) and enzymes (alpha glucosidases; Fig. 2F, G and Table 1). The laminins *LAMA2* and *LAMB4* are targets in the treatment of ocriplasmin vitreomacular adhesion, whereas the amylases *AMY2A* and *MGAM* are targeted by acarbose, voglibose and miglitol for the improvement of postprandial hyperglycemia. Overall, the number of drug target SVs is 5–10 times lower than in ADME genes with each individual harboring a total of 1.2 LOF and 0.3 IGD SVs (Fig. 2H). In contrast to SVs in ADME genes, aggregated SV frequencies differed almost 2-fold between ethnogeographic groups with the lowest numbers of functional SVs across drug targets in East Asians (0.88 per individual) and the highest number in individuals of African ancestry (1.64 per individual).

### Interpreting the functionality of non-coding SVs

While the consequences of SVs in coding regions have been studied extensively, interpretation of the functional effects of non-coding structural variability, which account for >85% of all pharmacogenomic structural variation, has not yet been presented. Here, we inferred functional effects by analyzing the overlap of structural variation with experimentally determined transcription factor binding site (TFBS) data of 224 transcription factors and their expression across 130 cell types and tissues. Of all 12,786 non-coding SVs identified in ADME genes and drug targets, 2958 (23.1%) overlapped with at least one TFBS (Fig. 3A). The most commonly affected binding motifs corresponded to transcription factors with globally important functions, such as CTCF (impacted by 481 SVs), which plays critical roles in genome partitioning and maintenance of the chromosomal architecture, RAD21 (291 SVs), a member of the cohesin complex, and FOS (272 SVs) and JUND (232 SVs), which dimerize to form the AP-1 transcription complex that plays pleiotropic roles in the activation of gene expression (Fig. 3B). Further, various binding sites of key tissue-specific transcription factors were impacted, including HNF4A (affected by 197 SVs), a transcription factor of central importance for hepatopancreatic development and xenobiotic response<sup>24</sup>, and RXRA (affected by 169 SVs), a combinatorial partner that dimerizes with approximately one third of nuclear receptors in human liver<sup>25</sup>.

Since most TFs are not ubiquitously expressed, SVs in their respective TFBSs can only impact the target gene expression in tissues where the respective transcription factor is expressed. We thus analyzed the expression overlap of pharmacogenes (both ADME and drug targets) that harbor SVs affecting TFBSs with the respective transcription factors across nine tissues of major pharmacokinetic or pharmacodynamic importance (Fig. 3C). In total, we identified 1276 non-coding SVs where the affected gene and the respective transcription factor were co-expressed in at least one tissue with each individual carrying an estimated average of 21.7 putatively functional pharmacogenomic SVs (Supplementary Table 4).

Deletions of TFBSs ablate TF activity for the associated gene, which would entail reduced or increased expression in the case of transcriptional activators or repressors, respectively. Inversely, duplication of TFBSs can be expected to have opposite effects. In ADME genes, the highest frequency of such non-coding deletions affecting TFBSs was found in *SLC10A2* (encoding the intestinal transporter ASBT; MAF = 25.9%) where it affected the binding sites of the co-expressed transcription factors CTCF (Table 2). Similarly, deletion of TFBSs of CTCF, RAD21 and SP1 in *SLC28A1* encoding the renal transporter CNT1 was identified in 20% of alleles, and the most common deletion of an hepatic gene was found in hepatic sulfotransferase *SULT2A1* (MAF = 5.4%), affecting TFBSs of CTCF, CHAMP1, ATF2 and CREB1. When normalizing for gene length, we observed a similar number of TFBS SVs in ADME genes and drug targets ( $p = 0.52$  for Wilcoxon Rank Sum test based on the 1276 non-coding putatively functional SVs) with deletion and insertions being the most common variant types.

In addition to ADME genes, we also discovered a multitude of SVs that impacted transcription factors co-expressed with drug targets (Table 3). For instance, the upstream region of *GABRP* encoding the  $\pi$  subunit of the GABA<sub>A</sub> receptor that constitutes the target of a multitude of mostly anesthetic and anxiolytic drugs, contains a frequent insertion polymorphism (MAF = 62.4%) that impacts the TFBS of the neuronal transcription factors MAFK, which could modulate *GABRP* expression in the central and enteric nervous system. Similarly, expression of the prostaglandin receptor *PTGER4* in the lung might be impacted by common deletions of JUND and SP1 binding sites (MAF = 14.2%), which might have important roles in the modulation of prostaglandins in allergic pulmonary inflammation and asthma. These analyses

constitute to our knowledge the first systematic evaluation of the impact of structural pharmacogenomic variation on experimentally validated transcription factor binding motifs and will provide an important resource for future biological validation efforts.

### Impact of SVs on pharmacogene expression

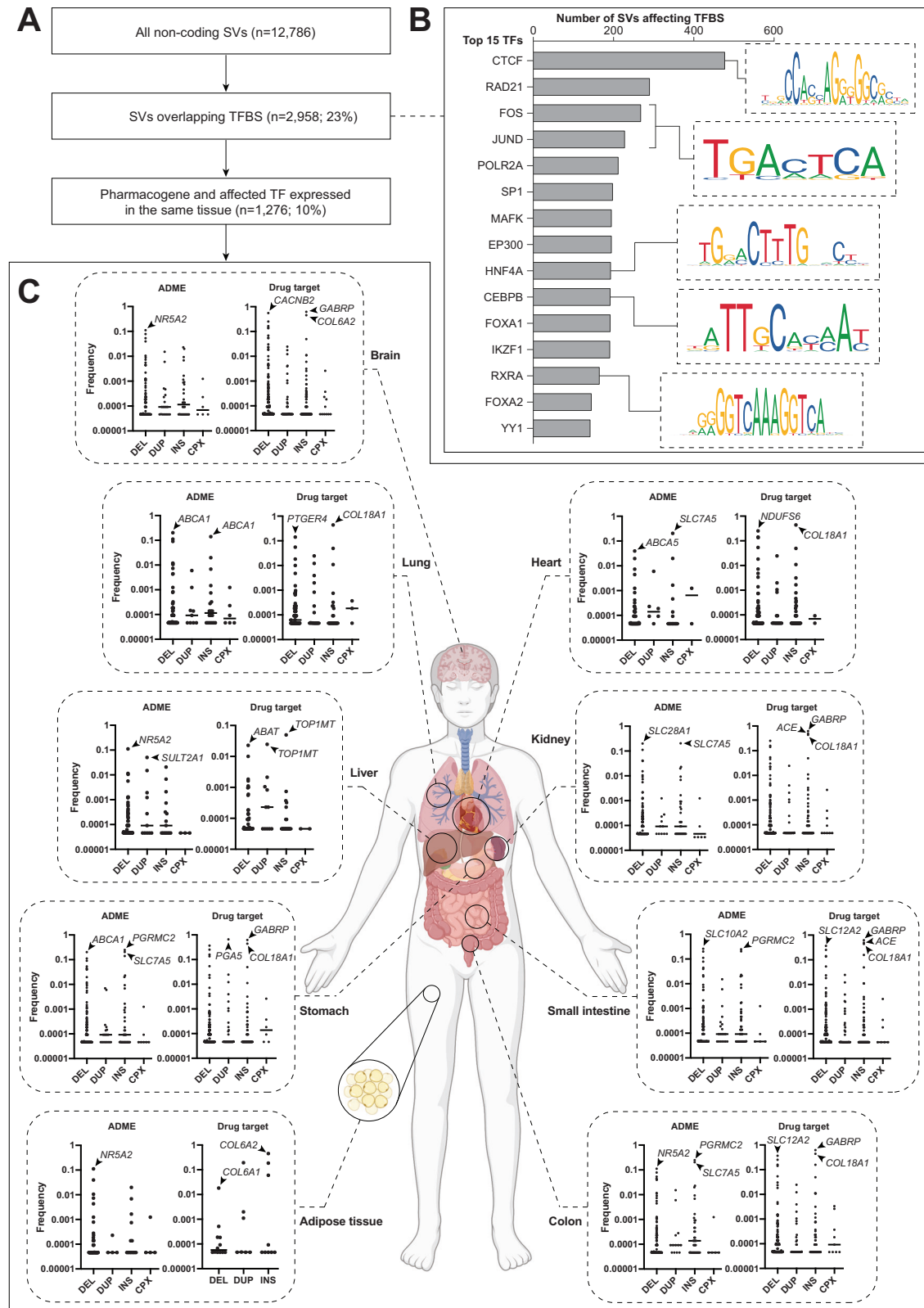
To systematically interrogate the functional impact of PGx and drug target SVs, we mapped the profile of pharmacogenomic SVs to published multi-tissue eQTL data from the GTEx project<sup>26</sup>. Because of different detection workflows and cohort sizes between the eQTL study and gnomAD, the number of detected SVs differed more than 7-fold between both studies (approx. 61k to 433k) and only 23% of SVs mapped within 100 bp in both data sets. In total, we found 21 common SVs of ADME and drug targets (15 coding, 6 non-coding) that were significantly associated with mRNA expression (Table 4). As expected, well-known functional SVs of *AMY2A*, *CYP2A6*, and its corresponding pseudogene *CYP2A7*, *CYP21A2*, *GSTM1*, *GSTT1*, *SULT1A1*, and *UGT2B17* are significantly associated with mRNA expression in various tissues (Table 4, Fig. 4A). Of note, *CYP2D6* SVs, which are known to improve phenotypic predictions<sup>27</sup>, are not included in the GTEx dataset, likely due to issues with appropriately calling variations in this complex locus<sup>28</sup>.

A very frequent partial deletion within the *S1PR4* locus (combined MAF = 0.64) were significantly correlated with its expression in lymphocytes (Benjamini-Hochberg [BH]  $p < 0.005$ ). This finding is interesting as reduced expression of *S1PR4* has been associated with protection from diet-induced non-alcoholic steatohepatitis and hepatic fibrosis<sup>29</sup>. Interestingly, almost one in five individuals carried homozygous *S1PR4* deletions and there was a population difference in SV frequency from 53% in East Asians, Latinos (65%), Africans (88%) to European subjects (90%). Similarly, a previously described intronic deletion (MAF = 2%) of *CYP4F12*, which covers several TFBSs<sup>30</sup>, was associated with decreased expression in thyroid and heart tissue (BH  $p < 0.004$ ). Furthermore, depending on the transcript reference, a 1.2 kb upstream or partial coding duplication of *ALDH1A2* was associated with higher expression in blood, while a non-coding deletion (covering TFBS) of *INSIG2* was associated with decreased expression in adipose and artery tissues.

Overall, each individual carried on average one structural eQTL that impacted the expression of drug targets and 3–5 variations affecting ADME gene expression (Fig. 4B). Interestingly, the distribution of eQTL-SVs per individual were overall similar between Europeans, Africans and admixed Americans, whereas the number of ADME SVs was considerably higher in East Asians. Based on these data, we carefully estimated the functional impact of non-coding structural variations (see Eq. (1) in the Methods section for details). Specifically, by cross-referencing the number of functional non-coding SVs in ADME genes and drug targets (21.7 per individual), as well as the number of functional exonic SVs in ADME genes (10.3 per individual) and drug targets (1.5 per individual) calculated in this study with data about the functional impact of with available information about the number of functional SNVs in ADME genes (40.6 per individual) and drug targets (26 per individual) from the literature<sup>10,31</sup>, we calculated that non-coding structural variants account for approximately 22% of the overall genetically encoded pharmacogenomic variability. As such, both coding and non-coding SVs constitute a considerable source of pharmacogenomic variability, the latter of which is not commonly considered by studies into heritable factors of drug response and safety.

### DISCUSSION

SVs are important mutational forces that shape genomic organization and biological functions<sup>32</sup>. Compared to SNVs, SVs are substantially understudied, at least in part due to the



**Fig. 3 Non-coding SVs overlap with transcription factor binding sites.** **A** All identified non-coding SVs were first filtered based on their overlap with transcription factor binding sites (TFBS). Subsequently, those SVs where the impacted pharmacogene and transcription factor (TF) were expressed in the same tissue were taken forward. **B** 15 TFs that are affected by the highest number of SVs. Binding motifs were indicated for selected TFs. **C** Frequencies of SVs for which expression of the corresponding pharmacogene and transcription factor overlap in nine human tissues. SVs with top frequencies were highlighted in both ADME genes and drug targets.

**Table 2.** Putative effects of common non-coding structural variants in ADME genes.

Tissue	Gene	Type	Carrier frequency	Affected transcription factors	Potential effect	PGx evidence
Small intestine	<i>SLC10A2</i>	DEL	1 in 2 individuals	CTCF	Docetaxel toxicity	57
Stomach, small intestine, colon	<i>SLC7A5</i>	DEL, INS	1 in 3 individuals	POLR2A, HNF4A	Gastrointestinal melphalan toxicity	58
Kidney, Small intestine	<i>SLC28A1</i>	DEL	1 in 3 individuals	CTCF, RAD21, TRIM22	Gemcitabine clearance	59
Brain, Stomach, Colon	<i>SLC16A1</i>	DEL	1 in 10 individuals	SP1	Valproic acid exposure	60
Liver	<i>SULT2A1</i>	DUP	1 in 10 individuals	CTCF, CHAMP1, CREB1	Acetaminophen induced birth defect	61
Stomach, Small intestine, Colon	<i>SLC29A1</i>	DEL	1 in 11 individuals	SP1	Ribavirin exposure	62
Liver, Small intestine	<i>CYP4F1</i>	INS	1 in 24 individuals	HNRNPk	Warfarin does requirement	63
Kidney	<i>UGT1A6</i>	DEL	1 in 33 individuals	RXRA, JUND, HNF4A, CEBPB, RAD21, HDAC2, NR3C1	Valproic acid exposure	63
Lung, Liver	<i>GSTP1</i>	DEL	1 in 44 individuals	HNF4A, RXRA, JUND, FOS, KDM1A, JUN, CEBPB	Cisplatin response	64

Note that not all transcription factors affected by the respective SVs are shown. For the comprehensive list, we refer to Supplementary Table 4.  
DEL deletion, INS insertion.

difficulties associated with their identification via commonly used short-read sequencing technologies. While over 500,000 SVs have been described across the human genome<sup>19</sup>, only a small minority of those are functionally understood. In ADME genes, information about structural variability has long been limited to CNVs and complex rearrangements in few selected loci, such as *CYP2A6*, *CYP2D6*, *SULT1A1*, and various *GSTs*<sup>33</sup>. Even less information was available about the structural variability in drug targets where analyses were largely limited to the *AMY1/2* locus<sup>34</sup>. While CNVs in other drug target genes, such as *PGA5*, have been described in genome-wide studies<sup>35,36</sup>, their precise architecture and functional effects on drug response have not been analyzed. Building on these findings, we here compiled an overview of the structural pharmacogenomic variome across 908 ADME and drug target genes based on publicly available SV data. These data provide a comprehensive map of structural variability in human pharmacogenes and constitute the basis for the first functional interpretation of both coding and non-coding pharmacogenomic structural variation.

Structural variability is of considerable importance for determining the molecular phenotype of cells with 18% of total detected genetic variation in gene expression being attributed to CNVs<sup>37</sup>. Of all pharmacogenomic SVs identified, 775 (5.2%) were annotated as putatively causing functional consequences (Supplementary Table 1). Examples include common SVs in multiple *CYPs*, *GSTs* and *UGTs*, as well as in a few drug target genes, primarily those encoding laminins and amylases (Table 1). Furthermore, our data corroborated previous findings of *SULT1A1* duplications<sup>38</sup>, which can translate into enhanced phase II metabolism of multiple drugs (e.g. acetaminophen and tamoxifen) and hormones (e.g. estrogen)<sup>39</sup>. However, the functional consequences of the remaining 14,209 SVs, consisting primarily of those that were located up- and downstream of the gene or that affected UTRs or intronic regions, had not been assigned using current annotation guidelines.

In non-coding regions of the genome, SVs can affect regulatory sequences, such as TFBS, and such variation has been shown to impact gene expression, biological functions and disease risk<sup>40–42</sup>. However, associations of non-coding SVs with drug-related effects have been lacking. We thus integrated structural genomics data with transcription factor binding signatures and expression data across key tissues involved in drug action and drug disposition to pinpoint potential impacts of such non-coding structural variability on drug-related phenotypes. Our analyses identified 1276 SVs that impact experimentally validated TFBS in pharmacogenetic regulatory elements. In ADME genes, multiple common SVs were identified that impact TFBS upstream of the SLC transporters *SLC7A5* (encoding LAT1), *SLC16A1* (MCT1), *SLC28A1* (CNT1), and *SLC29A1* (ENT1), implicated in the disposition of melphalan, valproic acid, gemcitabine or ribavirin, respectively. Notably, while genes encoding *CYP* enzymes or transporters of the SLC and ABC superfamilies have previously been identified as highly variable at the level of single nucleotide polymorphisms<sup>43–45</sup>, these results show that, surprisingly, common structural variants affecting TFBS are predominantly found in *SLC* genes.

Examples of non-coding SVs with putative relevance for drug response include the deletion of a regulatory element upstream of the drug target gene *ABAT* that is found in 1 in 20 individuals. *ABAT* encodes GABA transaminase, one of the key pharmacodynamic targets of valproic acid. While SNVs in *ABAT* had previously been associated with valproic acid response<sup>46</sup>, the impacts of structural variation in this gene have to our knowledge not yet been addressed. Our results suggest that structural variants alter the recruitment of HDAC2, a histone deacetylase expressed in the CNS that controls chromatin accessibility<sup>47</sup>, which in turn might impact *ABAT1* gene expression. Further examples are copy number variants of binding sites for the lysine demethylase KDM1A in the locus encoding the serotonin receptor *HTR2A*. Previous studies

**Table 3.** Tissue-specific drug response that might be affected by putatively functional non-coding SVs in drug target genes.

Tissue	Gene	Type	Carrier frequency	Affected transcription factors	Potential effect on	PGx evidence
Brain	<i>GABRP</i>	INS	1 per individual	MAFK	Response to propofol	
Brain	<i>GRIN2A</i>	DEL	1 in 3 individuals	FO XK2	Ketamine exposure and response	65
Colon	<i>TOP1</i>	DEL, INS	1 in 13 individuals	GATA3	Irinotecan response	66
Adipose tissue	<i>PPARA</i>	DEL, INS, CPX	1 in 17 individuals	POLR2A	Fenofibrate response	67
Brain	<i>ABAT</i>	DEL	1 in 20 individuals	HDAC2	Valproic acid response	46
Heart	<i>PDE3A</i>	DEL, INS, CPX	1 in 197 individuals	RXRA	Amrinone response	
Brain	<i>HTR2A</i>	DEL, INS	1 in 1548 individuals	KDM1A	Response to antidepressants	48
Brain	<i>MTNR1B</i>	DEL	1 in 1706 individuals	RAD21	Response to tasimelteon	
Brain	<i>HTR1A</i>	DEL, INS	1 in 2717 individuals	CTCF	Response to antipsychotics	68
Heart	<i>PDE3A</i>	DUP	1 in 2717 individuals	CTCF	Amrinone response	
Stomach	<i>ATP4A</i>	DUP	1 in 3623 individuals	NR3C1	Omeprazole response	
Brain	<i>MAOA</i>	DEL	1 in 3994 individuals	FOS	Isocarboxazid response	
Brain	<i>MTNR1B</i>	DUP	1 in 5435 individuals	RAD21	Response to tasimelteon	
Brain	<i>GABRA1</i>	DUP	1 in 10870 individuals	RAD21	Response to GABA ligands	69
Brain	<i>GABRA1</i>	INS	1 in 10870 individuals	ATF2	Diazepam response	69

Note that not all transcription factors affected by the respective SVs are shown. For the comprehensive list, we refer to Supplementary Table 4. CPX complex rearrangement, DEL deletion, DUP duplication, INS insertion.

**Table 4.** Common eQTL SVs located in ADME genes and drug targets.

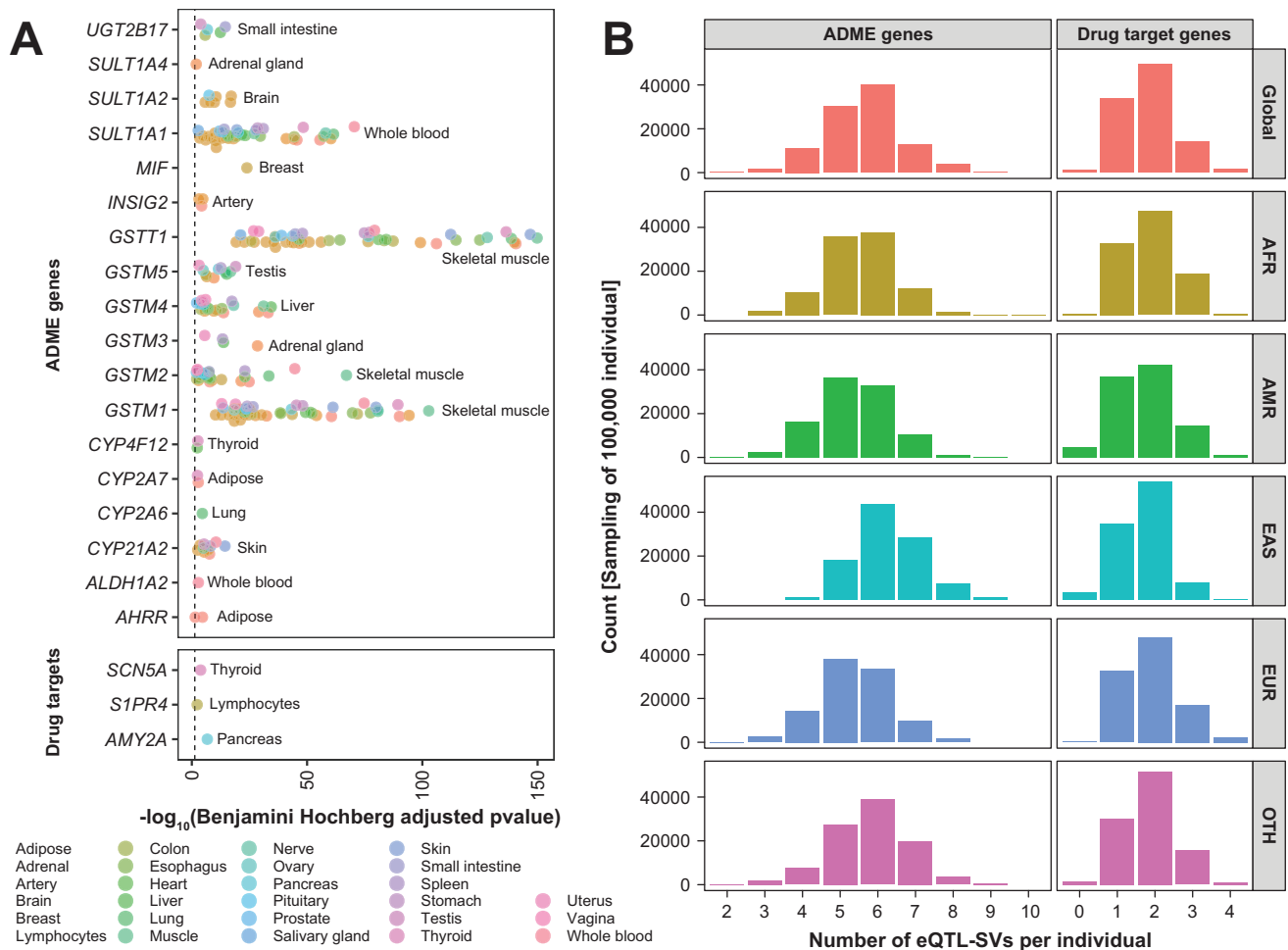
Gene	SV type	SV Function	Affected TFBSs	Tissues with eQTL	$-\log_{10}(BH)$ [range]	Beta value [range] <sup>a</sup>	Matched SVs <sup>b</sup>	GnomAD AF [range]
<i>AHRR</i>	DUP	non-coding	EGR1, NRF1, POLR2A, RNF2	2	1.48, 4.61	0.97, 1.48	0	NA
<i>ALDH1A2</i>	DUP	coding		1	2.84	0.97	1	0.08
<i>AMY2A</i>	MCNV	coding		1	6.72	0.16	2	0.14–0.19
<i>CYP2A6</i>	MCNV	coding		1	4.56	0.53	4	0.01–0.04
<i>CYP2A7</i>	MCNV	coding		2	2.46, 2.81	−0.63, −0.53	3	0.01–0.04
<i>CYP4F12</i>	DEL	non-coding	MAX, EP300, SP1, POLR2A, REST, JUND, NR3C1, CTCF, HNF4A, FOXA1, HDAC2, FOXA2, RXRA, RAD21, ZNF24	2	2.36	−1.06, −0.84	1	0.02
<i>CYP21A2</i>	MCNV	coding		10	2.56, 14.39	−0.44, −0.21	0	NA
<i>GSTM1</i>	MCNV	coding		25	10.29, 102.8	0.52, 0.64	1	0.85
<i>GSTM2</i>	MCNV	coding		15	2.02, 67.09	0.08, 0.41	1	0.85
<i>GSTM3</i>	MCNV	coding		3	5.59, 28.48	0.39, 0.49	0	NA
<i>GSTM4</i>	MCNV	coding		15	2.09, 34.44	0.14, 0.46	0	NA
<i>GSTM5</i>	MCNV	coding		8	3.14, 19.03	0.17, 0.42	0	NA
<i>GSTT1</i>	MCNV	coding		40	19.24, 149.8	0.55, 0.67	1	0.72
<i>INSIG2</i>	DEL	non-coding	ATF2, ATF7, CTCF, FOS, POLR2A, STAT3	3	3.13, 4.76	−1.08, −0.9	1	0.03
<i>MIF</i>	DEL	non-coding	CTCF, HNF4A, RAD21, RXRA, SMC3, TRIM22	1	23.9	0.82	0	NA
<i>S1PR4</i>	DEL	coding		1	2.33	0.51	2	0.82
<i>SCN5A</i>	INS	non-coding		1	3.82	−0.38	1	0.31
<i>SULT1A1</i>	MCNV	coding		26	2.8, 70.48	0.19, 0.49	1	0.48
<i>SULT1A2</i>	MCNV	coding		3	5.8, 17.1	−0.51, −0.32	0	NA
<i>SULT1A4</i>	MCNV	coding		1	1.92	0.18	0	NA
<i>UGT2B17</i>	DEL	coding		5	3.87, 14.53	−1.2, −0.7	1	0.56

AF allele frequency, eQTL expression quantitative trait locus, DUP duplication, DEL deletion, INS insertion, MCNV multi-copy number variation, TFBS Transcription factor binding site, BH Benjamini–Hochberg value, SV structural variant.

<sup>a</sup>Effect size of the correlation between SV and gene expression. Values greater than 0 reflect increased expression levels.

<sup>b</sup>A matched eQTL variant is defined as a SV of the same type (e.g., DUP, DEL) that was identified in both gnomAD and GTEx.





**Fig. 4 Impact of structural variation on pharmacogene expression. A** Structural variants in pharmacogenes associated with mRNA expression across multiple GTEx tissues. **B** Distribution of structural variants associated with mRNA expression levels of ADME and drug target genes. Carrier numbers are shown for a simulated population of 100,000 individuals. AFR Africans, EUR Europeans, AMR admixed Americans, EAS East Asians, OTH Others.

suggested that *HTR2A* activity associates with response to antidepressive treatment and remission of depressive symptoms<sup>48</sup>. Moreover, genetic manipulation of lysine methyltransferases in mice was shown to alter *Htr2a* expression and histone methylation has thus been proposed as an epigenetic drug target for anxiety and depression<sup>49</sup>. Our findings thus suggest that structural variability of the *HTR2A* locus might impact epigenetic remodeling and gene expression, thus potentially contributing to serotonergic signaling and response to selective serotonin reuptake inhibitors (SSRIs).

Combined, our results provide the most comprehensive map of coding and non-coding structural variations in the human pharmacome published to date. Furthermore, we provide the first functional interpretation of this structural variability, highlight a multitude of structural variants with putative tissue-specific impacts on drug response or toxicity due to deletion or insertion of regulatory elements for further experimental and epidemiological validations. Our data indicate that non-coding structural variants might present an understudied, but important class of variation, which might account for 22% of genetically encoded pharmacogenomic variability. As such, the presented findings constitute an important resource for variant prioritization and incentivize the incorporation of both coding and non-coding pharmacogenomic variability into personalized drug response predictions.

## METHODS

### Structural variant analysis

Structural genomic data for 908 pharmacogenes (344 ADME genes and 564 drug targets) from 10,847 unrelated individuals was extracted from gnomAD<sup>19,50</sup>. The ADME genes were selected based on previous work describing a targeted sequencing panel for ADME sequencing<sup>51</sup>. As drug target genes, we considered all genes that encode a target of an FDA-approved drug that was encoded in the nuclear genome<sup>10</sup>. In total 387,477 SVs were identified of which variants with filter status other than "PASS" or "MULTIALLELIC" and type of "unresolved non-reference breakpoint junction" & "reciprocal translocation" were excluded ( $n = 305,149$  after this exclusion). SVs with neighboring intervals were aggregated by gene and SV type using the `bed_cluster` function from the R package `valr`<sup>52</sup>. Specifically, we used `max_dist = 0` to merge overlapping and directly adjacent intervals, resulting in 256,429 unique SVs genome-wide. Subsequently, we filtered for overlap with the 908 pharmacogenes (Gencode v19), yielding a total of 14,984 SVs across the human pharmacome (Supplementary Table 1). SVs spanning more than one pharmacogene were counted for each gene individually. SVs were annotated as coding when they impacted at least one pharmacogenomic exon or as non-coding when the SV affected only intergenic or intronic regions. Non-coding variants were furthermore analyzed for the presence of transcription factor binding sites (TFBS) using the Transcription Factor ChIP-seq Cluster data (338

transcription factors [TFs], 130 cell types) from ENCODE 3<sup>53</sup>. After exclusion of TFBS with peak scores <200 and single study observations (1/1264), 224 TFs were analyzed. SV categories were extracted from the original study<sup>19</sup> and translated into putative functional consequences according to Supplementary Table 2. Information about 440 olfactory-related genes was extracted from the KEGG pathway “hsa04740”. Tissue-dependent expression levels of candidate genes and TFs were evaluated using median gene-level RNA-Seq data from GTEx<sup>26</sup>. Information about significant associations between SVs and RNA-seq expression was obtained from a multi-tissue eQTL study<sup>54</sup>. The data was filtered for SV-eQTLs, and gene information was added using biomart. The overlap between the breakpoints of SV-eQTLs and gnomAD-SVs was assessed using the bed\_closest function from valr<sup>52</sup>. Furthermore, SV-eQTLs that overlapped >99% with gnomAD-SVs were included in the analyses. The carrier frequency or number of total SVs associated with mRNA expression was assessed by simulating 100,000 individual using reported allele frequencies in gnomAD.

### Calculation of the functional impact of non-coding structural variations

The relative functional importance of non-coding SVs was calculated according to Eq. (1) as follows:

$$func_{ncSV} = \frac{n_{ncSV}}{n_{ncSV} + n_{SNV} + n_{cSV}} \quad (1)$$

with  $n_{ncSV}$  defined as the number of functional non-coding SVs in ADME genes and drug targets per individual,  $n_{cSV}$  defined as the number of functional exonic SVs in ADME genes and drug targets per individual and  $n_{SNV}$  defined as the combined number of functional SNVs in ADME genes and drug targets per individual. The number of SNVs in ADME genes per individual was obtained from ref. <sup>31</sup>, while the number of SNVs in drug target genes was calculated from ref. <sup>10</sup> by aggregating all drug target variants with putative functional impacts weighted with the respective frequencies in the entire cohort.

### Statistical analyses

Common variations were defined as variants with a minor allele frequency (MAF)  $\geq 1\%$ , while SVs with frequencies <1% were considered as rare. All analyses including the filtering steps were performed using R version 4.0.1 with the additional packages tidyverse\_1.3.0<sup>55</sup>, valr\_0.6.1<sup>52</sup>, ggsignif\_0.6.0<sup>56</sup>. If not other stated, we used Wilcoxon Rank Sum Tests to compare continuous data between groups. All tests were two-sided and significance was assumed at 0.05.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

SV data is available via gnomAD (<https://gnomad.broadinstitute.org/>), TFBS data is provided by ENCODE (<https://www.encodeproject.org/>) and eQTL information is available via the GTEx Portal (<https://gtexportal.org/home/>). All these repositories are publicly available.

Received: 5 May 2023; Accepted: 29 August 2023;

Published online: 08 September 2023

### REFERENCES

- Spear, B. B., Heath-Chiozzi, M. & Huff, J. Clinical application of pharmacogenetics. *Trends Mol. Med.* **7**, 201–204 (2001).
- Dowden, H. & Munro, J. Trends in clinical success rates and therapeutic focus. *Nat. Rev. Drug Disc.* **18**, 495–496 (2019).
- Downing, N. S. et al. Postmarket safety events among novel therapeutics approved by the US Food and Drug Administration Between 2001 and 2010. *JAMA* **317**, 1854–1863 (2017).
- Lauschke, V. M. & Ingelman-Sundberg, M. Prediction of drug response and adverse drug reactions: from twin studies to Next Generation Sequencing. *Eur. J. Pharm. Sci.* **130**, 65–77 (2019).
- Lauschke, V. M., Zhou, Y. & Ingelman-Sundberg, M. Novel genetic and epigenetic factors of importance for inter-individual differences in drug disposition, response and toxicity. *Pharmacol. Ther.* **197**, 122–152 (2019).
- Russell, L. E. et al. Pharmacogenomics in the Era of Next Generation Sequencing —from Byte to Bedside. *Drug Metab. Rev.* **53**, 253–278 (2021).
- Pirmohamed, M. Pharmacogenomics: current status and future perspectives. *Nat. Rev. Genet.* **24**, 350–362 (2023).
- Schärfe, C. P., Tremmel, R., Schwab, M., Kohlbacher, O. & Marks, D. S. Genetic variation in human drug-related genes. *Genome Med.* **9**, 117 (2017).
- Hauser, A. S. et al. Pharmacogenomics of GPCR drug targets. *Cell* **172**, 41–54 (2018).
- Zhou, Y. et al. Rare genetic variability in human drug target genes modulates drug response and can guide precision medicine. *Sci. Adv.* **7**, eabi6856 (2021).
- Alkan, C., Coe, B. P. & Eichler, E. E. Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12**, 363–376 (2011).
- Lappalainen, T., Scott, A. J., Brandt, M. & Hall, I. M. Genomic analysis in the age of human genome sequencing. *Cell* **177**, 70–84 (2019).
- Huddleston, J. et al. Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res.* **27**, 677–685 (2017).
- Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat. Rev. Genet.* **14**, 125–138 (2013).
- Spielmann, M. & Mundlos, S. Looking beyond the genes: the role of non-coding variants in human disease. *Hum. Mol. Genet.* **25**, R157–R165 (2016).
- Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
- Santos, M. et al. Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. *Genet. Med.* **20**, 622–629 (2018).
- Tremmel, R. et al. Copy number variation profiling in pharmacogenes using panel-based exome resequencing and correlation to human liver expression. *Hum. Genet.* **139**, 137–149 (2020).
- Collins, R. L. et al. A structural variation reference for medical and population genetics. *Nature* **581**, 444–451 (2020).
- Hasin-Brumshtein, Y., Lancet, D. & Olender, T. Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet.* **25**, 178–184 (2009).
- Jarvis, J. P., Peter, A. P. & Shaman, J. A. Consequences of CYP2D6 copy-number variation for pharmacogenomics in psychiatry. *Front Psychiatry* **10**, 432 (2019).
- Hebbring, S. J. et al. Human SULT1A1 gene: copy number differences and functional implications. *Hum. Mol. Genet.* **16**, 463–470 (2007).
- Tremmel, R., Klein, K., Winter, S., Schaeffeler, E. & Zanger, U. M. Gene copy number variation analysis reveals dosage-insensitive expression of CYP2E1. *Pharmacogenomics J.* **16**, 551–558 (2016).
- Tirona, R. G. et al. The orphan nuclear receptor HNF4a determines PXR- and CAR-mediated xenobiotic induction of CYP3A4. *Nat. Med.* **9**, 220–224 (2003).
- Pérez, E., Bourguet, W., Gronemeyer, H. & de Lera, A. R. Modulation of RXR function through ligand design. *Biochim. Biophys. Acta* **1821**, 57–69 (2012).
- GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
- Dalton, R. et al. Interrogation of CYP2D6 structural variant alleles improves the correlation between CYP2D6 genotype and CYP2D6-mediated metabolic activity. *Clin. Transl. Sci.* **13**, 147–156 (2020).
- Nofziger, C. et al. PharmVar GeneFocus: CYP2D6. *Clin. Pharmacol. Ther.* **107**, 154–170 (2020).
- Hong, C. H. et al. Sphingosine 1-phosphate receptor 4 promotes nonalcoholic steatohepatitis by activating NLRP3 inflammasome. *Cell Mol. Gastroenterol. Hepatol.* **13**, 925–947 (2022).
- Cauffiez, C. et al. Functional characterization of genetic polymorphisms identified in the human cytochrome P450 4F12 (CYP4F12) promoter region. *Biochem. Pharmacol.* **67**, 2231–2238 (2004).
- Ingelman-Sundberg, M., Mkrtchian, S., Zhou, Y. & Lauschke, V. M. Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum. Genomics* **12**, 26 (2018).
- Hurles, M. E., Dermitzakis, E. T. & Tyler-Smith, C. The functional impact of structural variation in humans. *Trends Genet.* **24**, 238–245 (2008).
- He, Y., Hoskins, J. M. & McLeod, H. L. Copy number variants in pharmacogenetic genes. *Trends Mol. Med.* **17**, 244–251 (2011).

34. Usher, C. L. et al. Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. *Nat. Genet.* **47**, 921–925 (2015).
35. Gamazon, E. R., Huang, R. S., Dolan, M. E. & Cox, N. J. Copy number polymorphisms and anticancer pharmacogenomics. *Genome Biol.* **12**, R46–12 (2011).
36. Sudmant, P. H. et al. Global diversity, population stratification, and selection of human copy-number variation. *Science* **349**, aab3761 (2015).
37. Stranger, B. E. et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
38. Vijzelaar, R. et al. Multi-ethnic SUL1A1 copy number profiling with multiplex ligation-dependent probe amplification. *Pharmacogenomics* **19**, 761–770 (2018).
39. Tremmel, R. et al. Methylugenol DNA adducts in human liver are associated with SUL1A1 copy number variations and expression levels. *Arch. Toxicol.* **91**, 3329–3339 (2017).
40. Haas, J. et al. Genomic structural variations lead to dysregulation of important coding and non-coding RNA species in dilated cardiomyopathy. *EMBO Mol. Med.* **10**, 107–120 (2018).
41. Han, L. et al. Functional annotation of rare structural variation in the human brain. *Nat. Commun.* **11**, 2990 (2020).
42. D'haene, E. & Vergult, S. Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. *Genet. Med.* **23**, 34–46 (2021).
43. Fujikura, K., Ingelman-Sundberg, M. & Lauschke, V. M. Genetic variation in the human cytochrome P450 supergene family. *Pharmacogenet. Genom.* **25**, 584–594 (2015).
44. Schaller, L. & Lauschke, V. M. The genetic landscape of the human solute carrier (SLC) transporter superfamily. *Hum. Genet.* **138**, 1359–1377 (2019).
45. Xiao, Q., Zhou, Y. & Lauschke, V. M. Ethnogeographic and inter-individual variability of human ABC transporters. *Hum. Genet.* **139**, 623–646 (2020).
46. Li, X. et al. Polymorphisms of ABAT, SCN2A and ALDH5A1 may affect valproic acid responses in the treatment of epilepsy in Chinese. *Pharmacogenomics* **17**, 2007–2014 (2016).
47. Guan, J.-S. et al. HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature* **459**, 55–60 (2009).
48. Horstmann, S. et al. Polymorphisms in GRIK4, HTR2A, and FKBP5 show interactive effects in predicting remission to antidepressant treatment. *Neuropsychopharmacology* **35**, 727–740 (2010).
49. Shen, E. Y. et al. Neuronal deletion of Kmt2a/Mll1 histone methyltransferase in ventral striatum is associated with defective spike-timing-dependent striatal synaptic plasticity, altered response to dopaminergic drugs, and increased anxiety. *Neuropsychopharmacology* **41**, 3103–3113 (2016).
50. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
51. Klein, K. et al. A new panel-based next-generation sequencing method for ADME genes reveals novel associations of common and rare variants with expression in a human liver cohort. *Front. Genet.* **10**, 7 (2019).
52. Riemondy, K. A. et al. valr: Reproducible genome interval analysis in R. *F1000Research* **6**, 1025 (2017).
53. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
54. Scott, A. J., Chiang, C. & Hall, I. M. Structural variants are a major source of gene expression differences in humans and often affect multiple nearby genes. *Genome Res.* **31**, 2249–2257 (2021).
55. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
56. Ahlmann-Eltze, C., Patil I. ggsignif: R Package for Displaying Significance Brackets for 'ggplot2'. *PsyArXiv* <https://doi.org/10.31234/osf.io/7awm6> (2021).
57. Chan, S. L. et al. Genetic diversity of variants involved in drug response and metabolism in Sri Lankan populations. *Pharmacogenet. Genom.* **26**, 28–39 (2016).
58. Giglia, J. L. et al. A single nucleotide polymorphism in SLC7A5 is associated with gastrointestinal toxicity after high-dose melphalan and autologous stem cell transplantation for multiple myeloma. *Biol. Blood Marrow Transplant* **20**, 1014–1020 (2014).
59. Mitra, A. K. et al. Pathway-based pharmacogenomics of gemcitabine pharmacokinetics in patients with solid tumors. *Pharmacogenomics* **13**, 1009–1021 (2012).
60. Adjei, A. A., Gaedigk, A., Simon, S. D., Weinshilboum, R. M. & Leeder, J. S. Inter-individual variability in acetaminophen sulfation by human fetal liver: Implications for pharmacogenetic investigations of drug-induced birth defects. *Birth Defects Res. A: Clin. Mol. Teratol.* **82**, 155–165 (2008).
61. Allegra, S. et al. Role of pharmacogenetic in ribavirin outcome prediction and pharmacokinetics in an Italian cohort of HCV-1 and 4 patients. *Biomed. Pharmacother.* **69**, 47–55 (2015).
62. Zhang, J. E. et al. Effect of genetic variability in the CYP4F2, CYP4F11, and CYP4F12 genes on liver mRNA levels and warfarin response. *Front. Pharmacol.* **8**, 323 (2017).
63. Guo, Y., Hu, C., He, X., Qiu, F. & Zhao, L. Effects of UGT1A6, UGT2B7, and CYP2C9 genotypes on plasma concentrations of valproic acid in Chinese children with epilepsy. *Drug Metab. Pharmacokinet.* **27**, 536–542 (2012).
64. Ye, H. et al. Predictive assessment in pharmacogenetics of Glutathione S-transferases genes on efficacy of platinum-based chemotherapy in non-small cell lung cancer patients. *Sci. Rep.* **7**, 2670 (2017).
65. Chen, M.-H. et al. Treatment response to low-dose ketamine infusion for treatment-resistant depression: A gene-based genome-wide association study. *Genomics* **113**, 507–514 (2021).
66. Paolicchi, E. et al. Topoisomerase 1 promoter variants and benefit from irinotecan in metastatic colorectal cancer patients. *Oncology* **91**, 283–288 (2016).
67. Irvin, M. R. et al. Rare PPARA variants and extreme response to fenofibrate in the Genetics of Lipid-Lowering Drugs and Diet Network Study. *Pharmacogenet. Genomics* **22**, 367–372 (2012).
68. Takekita, Y. et al. HTR1A polymorphisms and clinical efficacy of antipsychotic drug treatment in schizophrenia: a meta-analysis. *Int. J. Neuropsychopharmacol.* **19**, pyv125 (2016).
69. Steudle, F. et al. A novel de novo variant of GABRA1 causes increased sensitivity for GABA in vitro. *Sci. Rep.* **10**, 2379 (2020).

## ACKNOWLEDGEMENTS

The work in the authors' laboratory is funded by the Swedish Research Council [grant agreement numbers: 2019-01837 and 2021-02801], by the EU/EFPIA/OICR/McGill/KTH/Diamond Innovative Medicines Initiative 2 Joint Undertaking (EUOPEN grant number 875510), and by the European Union's Horizon 2020 research and innovation program Ubiquitous Pharmacogenomics (grant agreement number 668353) and by the Robert Bosch Stiftung, Stuttgart, Germany.

## AUTHOR CONTRIBUTIONS

R.T. and Y.Z. collected and analyzed the data. M.S. and V.M.L. designed and supervised the study. All authors contributed to the writing of the manuscript.

## FUNDING

Open access funding provided by Karolinska Institute.

## COMPETING INTERESTS

V.M.L. is CEO and shareholder of HepaPredict AB. The remaining authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41525-023-00371-y>.

**Correspondence** and requests for materials should be addressed to Volker M. Lauschke.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.