Pharmacogene variation in the UK 100,000 Genomes Project: an analysis of 60,221 individuals

R M Turner[1], Genomics England Research Consortium, *et* M Pirmohamed[1]

[1] Department of Molecular & Clinical Pharmacology, University of Liverpool, UK

Introduction

- Pharmacogenomics is the study and application of the genetic determinants of drug response.
- To date, clinical guidelines providing treatment recommendations are available for ~70 drugs whose response is associated with variation in one or more genes.
- However importantly, this variation is almost exclusively limited to common variants (minor allele frequency (MAF) ≥ 1%).
- The world-leading UK 100,000 Genomes Project[1] offers the opportunity to further investigate variation in genes involved in drug pharmacokinetics (PK) or pharmacodynamics (PD) – hereafter termed 'pharmacogenes'.

Aim

- The aim of this study was to comprehensively identify and describe variation in pharmacogenes observed within the 100,000 Genomes Project, and to estimate the proportion of functional genomic variation attributable to rare variation.

Figure 1. Study overview

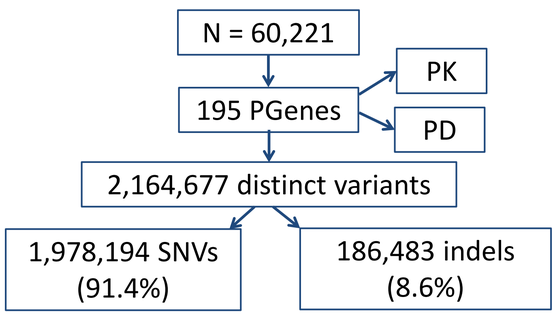


Figure 2. Overall frequency of common & rare variants

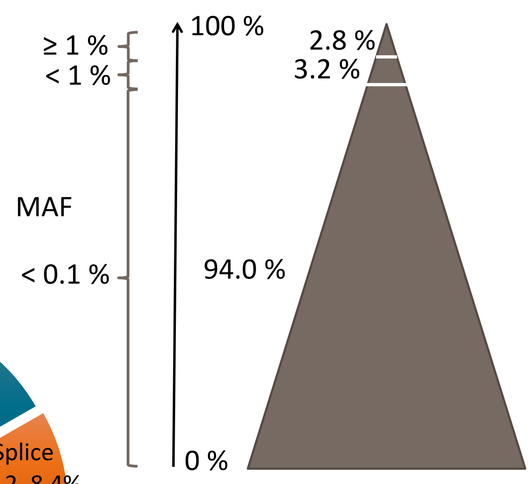


Figure 3. An overview of variant types

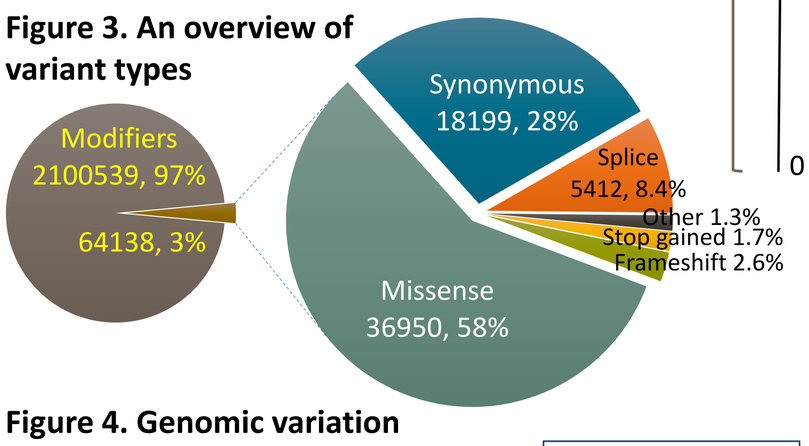
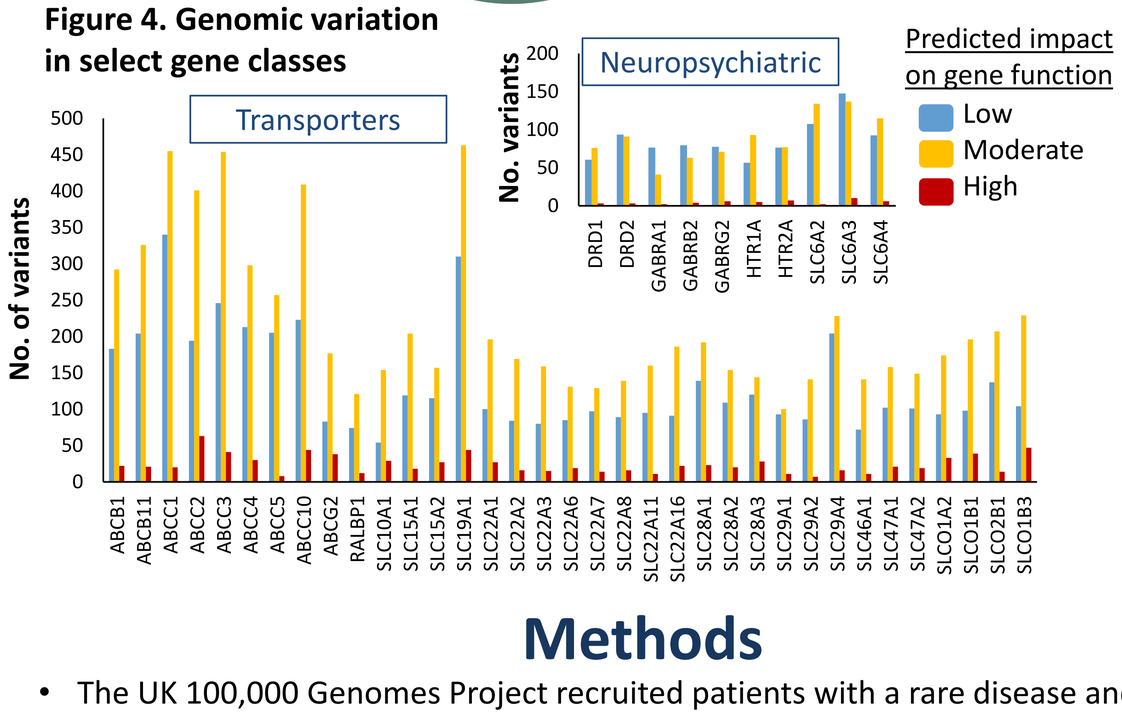


Figure 4. Genomic variation in select gene classes



Methods

- The UK 100,000 Genomes Project recruited patients with a rare disease and their unaffected relatives, and patients with cancer, between 2014 and 2018[1].
- DNA from all participants underwent whole genome sequencing, which was completed in December 2018.
- Single nucleotide variants (SNVs) and indels from pharmacogenes were extracted from all quality-controlled germline whole genome sequence data available on 1st February 2019 – this included both rare disease & cancer participants.
- Cellbase variant annotations were used, whilst SIFT and Polyphen2 algorithms were used to predict *in silico* the functional consequences of nonsynonymous SNVs.
- Functional variants were defined as those predicted to have a high or moderate impact on gene function according to the Sequence Ontology consequence hierarchy; missense variants were only considered functional if they were predicted deleterious/at least possible damaging by SIFT/Polyphen2.
- Data analysis was undertaken within the Genomics England research environment, following project approval by Genomics England.

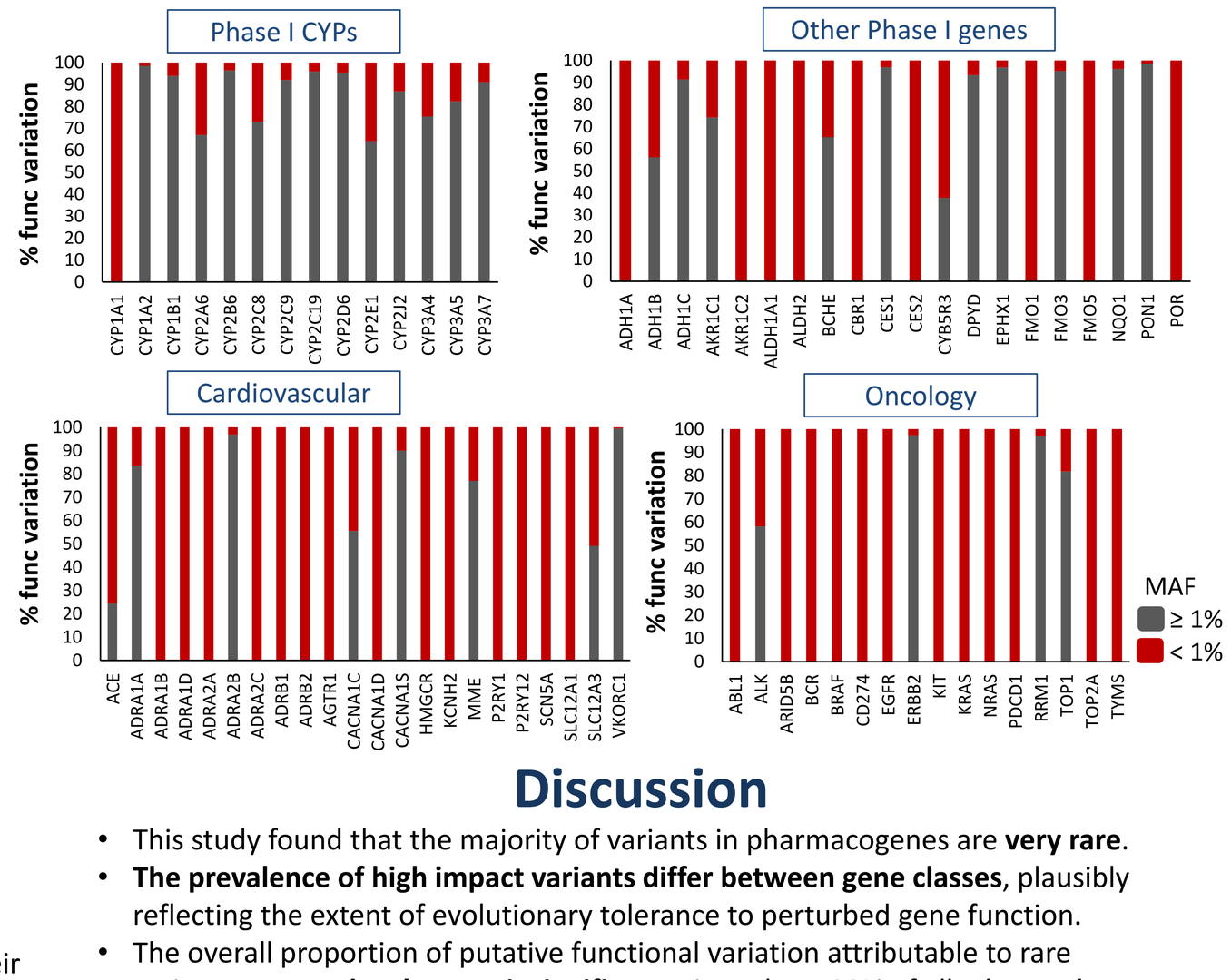
Results

- In total, **195** pharmacogenes from **60,221** individuals were analysed (Figure 1).
- **2,164,677** variants were identified: 1,978,194 SNVs (91.4%) & 186,483 indels (8.6%).
- Only **2.8%** of identified variants were common; **94.0%** of identified variants were very rare (MAF < 0.1%), and **3.2%** rare (MAF < 1% but ≥ 0.1%) (Figure 2).
- The majority of identified variants were intronic and classed as modifiers; 3% of identified variants were predicted to have a high (e.g. frameshift), moderate (e.g. missense) or low (e.g. synonymous) impact on gene function (Figure 3).
- Of this 3% of variants, **58% were missense**, **28% synonymous**, and the majority of the remainder were high impact variants (Figure 3).
- Importantly, the mean number of distinct high impact variants identified per gene varied across gene classes, ranging from **24.2 variants in transporter genes** to **4.8 variants in genes related to neuropsychiatric drug targets** (Figure 4, Table 1).
- The fraction of predicted functional variation attributable to rare variants **differed extensively between genes** and, **in over half** of analysed pharmacogenes, rare variants accounted for all identified functional variation (Figure 5).
- Overall on average, each participant carried **37** putatively functional variants within the analysed gene set, of which **4 variants (10.9%)** were rare.

Table 1. Variation in the number of high impact variants by gene class

Gene class	Pharmacokinetics				Pharmacodynamics		
	Phase 1	Phase 2	Transporters	Nuclear Rs	CVS	Oncology	Neuropsychiatric
Mean no. high impact variants per gene	20.2	13.9	24.2	9.5	17.6	8.5	4.8

Figure 5. Functional variation attributable to common & rare variants in select gene classes



Discussion

- This study found that the majority of variants in pharmacogenes are **very rare**.
- **The prevalence of high impact variants differ between gene classes**, plausibly reflecting the extent of evolutionary tolerance to perturbed gene function.
- The overall proportion of putative functional variation attributable to rare variants was **modest but not insignificant**, given that ~80% of all adverse drug reactions are 'type A' reactions related to excessive on-target drug action where variation in drug PK and on-target PD are key.
- Importantly, **the impact of rare variation on drug response is predicted to differ notably between drugs**, dependent on the extent of rare variation in genes relevant to each drug.
- This study will be extended to determine the extent of variation in this gene set within the somatic genome.

References

1. Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, et al. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ (Clinical research ed). 2018;361:k1687.

Acknowledgements

Dr Richard Turner is funded by a Postdoctoral Research Fellowship from Health Education England Genomics Education Programme. This research was possible through access to the data and findings generated by the 100,000 Genomes Project; http://www.genomicsengland.co.uk.

UNIVERSITY OF LIVERPOOL

RESEARCH FUNDED BY

Genomics Education Programme

NHS Health Education England