Genomic analysis using 455,000 UK Biobank exome sequences to identify candidate drug targets

S.Das[1], K. Mégy[1], A. O'Neill[1], G. Alamgir[1], J. Harrow[1], O. Burren[1], Q. Wang[2], K. Carss[1], E. Wheeler[1], R. S. Dhindsa[2], A. Harper[1], A. Nag[1], I. Tachmazidou[1], D. Vitsios[1], S. Deevi[1], S. Wasilewski[1], K. R. Smith[1], S. Petrovski[1], AstraZeneca Genomics Initiative.

Shikta.das@astrazeneca.com

1) Centre for Genomics Research, Discovery Sciences, Biopharmaceuticals R&D, AstraZeneca, Cambridge, UK.

2) Centre for Genomics Research, Discovery Sciences, Biopharmaceuticals R&D, AstraZeneca, Waltham, MA, USA.

1085-P

Abstract

The Centre for Genomics Research at AstraZeneca implements our ambitious aim to process up to 2M genomes before 2026 in order to expand our understanding of disease biology; to lead the identification for new targets for medicines; and to support selection of patients for clinical trails and inform indication expansion opportunities for therapeutics.

Our strategy, beyond sequencing samples from own clinical trials, is to establish a network of academic and industry partnerships. We have been using the UK Biobank (UKB), a cohort of 500k participants with deep phenotypic linked to whole exome and genome sequences.

Introduction

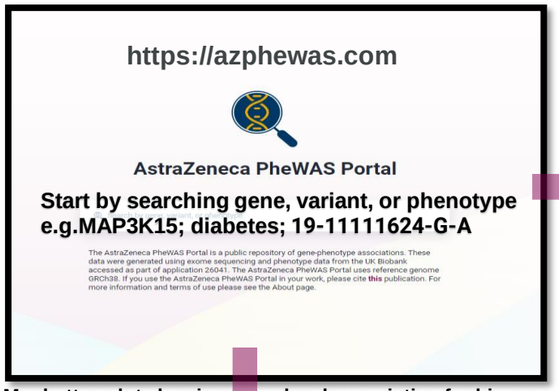
- We performed both gene and variant-level phenome-wide association statistics (PheWAS) using the exome sequences of the UK Biobank participants and considered ~17K binary and ~1.4K quantitative phenotypes.
- We derived phenotypes from UKB electronic health records, questionnaire data, and continuous traits
- We identified 2,818 significant (p<2x10-9) gene-phenotype relationships and ~40K significant variant-phenotype relationships.

Methods

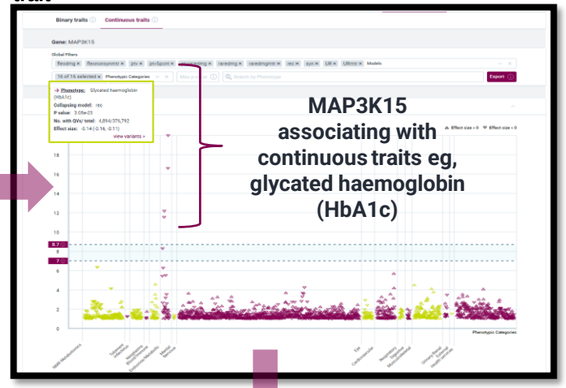
- The study uses UKB, a large-scale biomedical database and research resource, with approximately 500,000 participants 40–69 years of age at recruitment.
- We studied binary and quantitative traits taken from the February 2020 data release that was accessed on 27 March 2020 as part of UKB application 26041.
- The PEACOK R package. (https://github.com/astrazeneca-cgr-publications/PEACOK) implementation focuses on separating phenotype matrix generation from statistical association tests.

Results

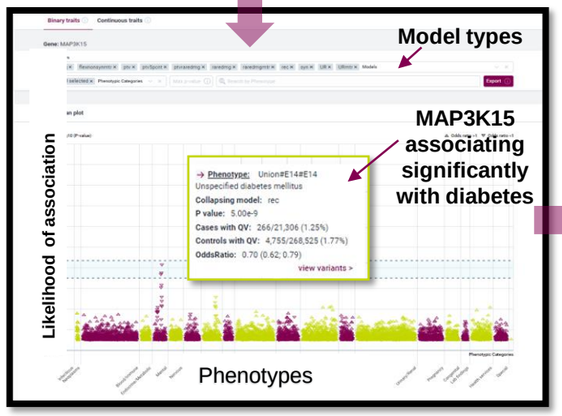
Homepage AstraZeneca's PheWAS Portal



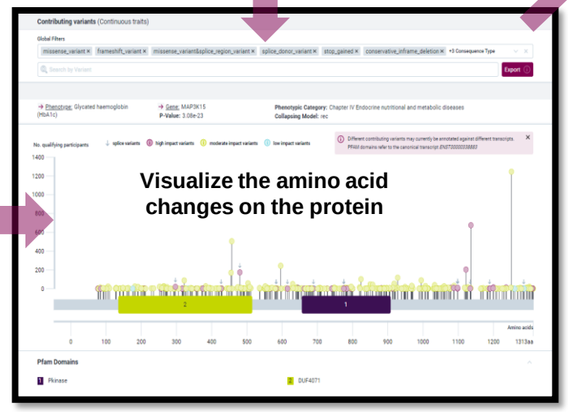
Manhattan plot: gene-level association with continuous trait



Manhattan plot showing gene-level association for binary trait



Lollipop plot for protein domains derived from PFAM database



Missense Tolerance Ratio (MTR) estimates for MAP3K15 gene generated from Exome Aggregation Consortium Database (ExAC), version 2.0

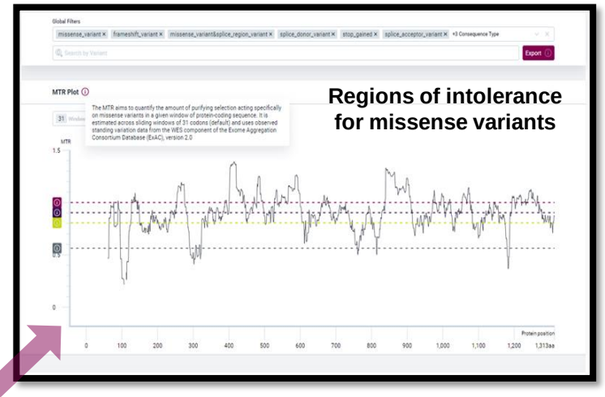


Table 1: Snapshot of table showing recessive genetic association with MAP3K15 gene

UK Biobank Phenotypes	P value	No. cases with QV	No. controls with QV	Odds ratio	Odds ratio LCI	Odds ratio UCI
Union#E14#E14 Unspecified diabetes mellitus	5.00E-09	266	4755	0.70	0.62	0.79
Diabetes diagnosed by doctor	3.05E-08	239	4751	0.70	0.62	0.80
20002#1220#diabetes	4.27E-08	238	4024	0.70	0.62	0.80
Source of report of E14 (unspecified diabetes mellitus)	4.60E-07	231	3137	0.72	0.63	0.82

QV = qualifying variant; LCI = lower confidence interval; UCI = upper confidence interval

COPD-related example:

SCAN ME

Conclusions

- We present AstraZeneca's PheWAS Portal (https://azphewas.com), a repository of gene-phenotype associations.
- Users can search the portal by gene, phenotype, or variant, and the data can be visualised, filtered, or downloaded for further analysis.
- AstraZeneca PheWAS portal is one of the most comprehensive genomic resources of its kind intended to empower a wide research community.

References

King et al. (2019). Nature Genetics. PMID: 31830040
Nelson MR et al (2015). Nature Genetics. PMID:26121088
Szustakiwski et al (2021). Nature Genetics. PMID: 34183854
Nag A. et al (2021). medRxiv, 2021.11.14.21266328
Wang Q. et al (2021). Nature. PMID: 34375979
Petrovski S. et al (2017). Genome Research. PMID: 28864458

Acknowledgements

We thank the participants and investigators in the UK Biobank study who made this work possible.
We thank the UK Biobank Exome Sequencing Consortium (UKB-ESC) members AbbVie, Alnylam Pharmaceuticals, AstraZeneca, Biogen, Bristol-Myers Squibb, Pfizer, Regeneron and Takeda for funding the generation of the data and Regeneron Genetics Center for completing the sequencing and initial quality control of the exome sequencing data.
Finally, we thank the AstraZeneca Centre for Genomics Research Analytics and Informatics team for processing and analysis of sequencing data.