N. Pontikos[1,2], I. Moghul[3], J. Yu[4], A. Fiorentino[1], G. Arno[1,2], A. J. Hardcastle[1], A. R. Webster[1,2], V. Plagnol[5]

[1]UCL Institute of Ophthalmology, London, United Kingdom, [2]Moorfields Eye Hospital, London, United Kingdom, [3]UCL Cancer Institute, London, United Kingdom, [4]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom, [5]UCL Genetics Institute, London, United Kingdom

Introduction: The molecular diagnosis of rare genetic diseases requires detailed clinical phenotypes and processing of large amounts of genetic data. This motivates large-scale collaborations between clinicians, geneticists and bioinformaticians across multiple sites where patient data are pooled together to boost the chances of solving rare cases, and validating novel genes. This motivated the development of Phenopolis, a platform for harmonisation of genetic and phenotypic data.

Methods: Using Human Phenotype Ontology (HPO) encoded phenotypes, Phenopolis is able to prioritise causative genes using different sources of evidence, such as literature search and model organism phenotype ontology analysis. Additionally, Phenopolis uncovers gene-phenotype relationships within the stored patient data through variant filtering and statistical enrichment of HPO terms. The database is implemented using a graph database for scalability which allows efficient linking of HPO, genes and variants.

Results: Phenopolis is an open-source web server providing an intuitive interface to genetic and phenotypic databases. Phenopolis is also an ideal platform for studying the pleiotropy of genes. The online version available at www.phenopolis.org, includes four example patients with inherited retinal dystrophies, to illustrate our methods.

Discussion: The Phenopolis platform accelerates clinical diagnosis, gene discovery and encourages wider adoption of the HPO in the study of rare genetic diseases. We plan on extending Phenopolis to interface with the Genomics England GenePanel app to retrieve relevant genes and contribute novel disease genes.

N. Pontikos: None. **I. Moghul:** None. **J. Yu:** None. **A. Fiorentino:** None. **G. Arno:** None. **A.J. Hardcastle:** None. **A.R. Webster:** None. **V. Plagnol:** A. Employment (full or part-time); Significant; Inivata. E. Ownership Interest (stock, stock options, patent or other intellectual property); Significant; Inivata.

P16.06B

An integrative approach to correlate clinical presentation patterns in Autism Spectrum Disorder with biological processes disrupted by Copy Number Variants (CNV) in brain genes

M. Asif[1,2,3], H. Martiniano[1,3], C. Rasga[2,3], A. R. Marques[2,3], J. X. Santos[2,3], G. Oliveira[4,5], F. M. Couto[1], A. M. Vicente[2,3,6]

[1]Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal, [2]Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisbon, Portugal, [3]BioISI - Biosystems & Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal, [4]Faculdade de Medicina da Universidade de Coimbra, Coimbra, Portugal, [5]Centro de Investigação e Formação Clinica do HP-CHUC, Coimbra, Portugal, [6]Instituto Gulbenkian de Ciência, Oeiras, Lisbon, Portugal

Autism Spectrum Disorder (ASD) is characterized by highly heterogeneous clinical phenotypes and complex genetic architecture, rendering early diagnosis and prognosis difficult. CNVs in many genes have been implicated in ASD, disrupting specific Biological Processes (BP) eventually associated with distinct clinical phenotypes. Here we sought to identify and predict clinical patterns associated with BP disrupted by specific brain gene CNVs, using a machine learning-based integrative approach. Firstly, clustering analysis of clinical records from 2446 ASD patients from the Autism Genome Project identified two consistent and highly stable clusters, differing in ASD severity, adaptive behaviour and cognitive ability. Secondly, functional enrichment analysis of rare CNVs, disrupting brain-expressed genes in these ASD subjects, identified 15 significant BPs, including nervous system development, cognition, and protein polyubiquitination. Random Forest feature importance analysis showed that all these BP contributed positively to the classification of ASD severity, as defined by the identified clusters. Finally, a Naive Bayes classifier was trained using cluster and BP information from a subset of data, comprising the 325 individuals with highest BP information content scores. A stratified five-fold cross validated model predicted the severity of ASD phenotype from BPs, with precision of 0.82 and recall of 0.39. This study thus shows that severity predictions can be attained from BP putatively disrupted by brain-gene CNVs. However, precise predictions are only achieved in subgroups with high BP information content, and specificity is generally low. A clinical application will thus require further analysis, in much larger datasets that include detailed phenotypic information. FCT-Portugal (SFRH/BD/52485/2014)

M. Asif: None. **H. Martiniano:** None. **C. Rasga:** None. **A.R. Marques:** None. **J.X. Santos:** None. **G. Oliveira:** None. **F.M. Couto:** None. **A.M. Vicente:** None.

P16.07C

BEA: a web tool for BioMark gene expression analysis