context-specific gene regulatory mechanisms underlying complex human diseases.

**References:**.

**Grants:**.

**Conflict of Interest:** Andrew Lu: None declared, Mike Thompson: None declared, Gracie Gordon: None declared, Andy Dahl: None declared, Chun Jimmie Ye C.J.Y. has received research support from Chan Zuckerberg Initiative,437Chan Zuckerberg Biohub, and Genentech, C.J.Y. is a Scientific Advisory Board member for and hold equity in435Related Sciences and ImmunAI, a consultant for and hold equity in Maze Therapeutics, and a436consultant for TReX Bio., Noah Zaitlen: None declared, Brunilda Balliu: None declared.

### P18.067.C DGH-GO: Dissecting the Genetic Heterogeneity of complex diseases using Gene Ontology, an interactive and user-friendly web application

***Muhammad Asif***[1;2], *Hugo Martiniano*[3], *Francisco Couto*[1]

[1]*LASIGE, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal, Lisbon, Portugal;* [2]*CENTURI Multi Engineering Platform, Aix Marseille University France, Marseille, France;* [3]*Instituto Nacional de Sau´de Doutor Ricardo Jorge, Avenida Padre Cruz, Lisboa, Portuga, Lisbon, Portugal.*

**Background/Objectives:** Neurodevelopmental disorders (NDDs) are phenotypically heterogeneous and difficult to diagnose at early-age. The genetic heterogeneity of NDDs matches to their clinical-variability. The different NDDs share biological mechanisms that further complex the patient's stratification, thus, limiting the applications of personalized medicine for NDDs. Existing studies have employed biological networks and machine-learning methods to dissect the genetic heterogeneity. Such methods suffer from many parameter tuning and lack biological interpretations, resulting in the reduced generalizability of the proposed method.

**Methods:** Here, we presented an interactive and user-friendly application, DGH-GO that allows biologists to dissect the genetic heterogeneity of complex diseases by stratifying the genes, disrupted by any type of genetic variants (SNV, CNVs). The application can also be used to study the shared etiology of complex-diseases.

**Results:** DGH-GO creates a functional similarity matrix of putative disease-causing genes or known-disease genes for multiple disorders using Gene Ontology (GO). The resultant matrix can be visualized in a 2D space using different dimension reduction methods (T-SNE and Principal-Coordinate-Analysis). Functional similarities from GO and projected space coordinates from dimension reduction methods can be used to identify clusters by employing four different clustering methods (K-means, Hierarchical, Fuzzy and PAM). The user may change the clustering parameters and see their effect on stratification results immediately.

**Conclusion:** In summary, functional-similarities, dimension-reduction and clustering, coupled with interactive-visualization and control over analysis allows biologists to explore and analyze their datasets, without knowing the execution of complex methods. The proposed application and its application to NDDs is available at https://github.com/Muh-Asif/DGH-GO.

**References:** Asif et al. 2018, 2019, 2020.

**Grants:**.

**Conflict of Interest:** None declared.

### P18.068.D Whole-Genome Sequencing: The Long and the Short of It

*Sylvan Caspar*[1], *Patricia Stoll*[1], *Siro Fritzmann*[1], *Gilles Gut*[1], *Daniel Salerno*[1], ***Janine Meienberg***[1], *Gabor Matyas*[1;2]

[1]*Foundation for People with Rare Diseases, Center for Cardiovascular Genetics and Gene Diagnostics, Schlieren, Switzerland;* [2]*Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland.*

**Background/Objectives:** Whole-exome sequencing (WES), short-read whole-genome sequencing (SR-WGS) and long-read WGS (LR-WGS) enable the detection of sequence variants at unprecedented scale, leading to new challenges in variant calling. Indeed, the calling of all clinically-relevant ClinVar/HGMD variants is challenging due to the limitations of sequencing and data analysis platforms. Here, we provide new insights into the performance of the most recent sequencing, alignment, and variant-calling pipelines in the detection of ClinVar/HGMD variants.

**Methods:** We used raw data of SR-WGS (PE150, ~60×) of >50 in-house samples as well as SR-WGS (PE150, ~60×) and LR-WGS (PacBio HiFi, ~30×) of 4 publicly available samples (HG001-HG004). We implemented 12 state-of-the-art analysis pipelines for SR-WGS, LR-WGS, or the combination of both as well as developed a workflow to assess the pipelines' performance in the detection of ClinVar/HGMD variants.

**Results:** For a ~60× genome, accelerated pipelines decreased the runtime of BWA/GATK from ~2.5 d to ~2-5 h, enabling the analysis of multiple samples. LR-WGS outperformed SR-WGS, particularly in regions with mappability <1, while WES failed to detect variants in non-exonic or GC-rich regions. However, no pipeline alone detected all ClinVar/HGMD variants. By analyzing read depth, strand bias, variant allele fraction, and population-based allele frequency, we identified a substantial number of false-positive calls and ClinVar/HGMD entries.

**Conclusion:** Sequencing, alignment, and variant-calling pipelines can significantly influence the detection of all ClinVar/HGMD variants, leading to both false-negative and false-positive results. Owing to its inherent advantages in variant detection/calling, LR-WGS should be implemented in clinical practice as soon as it is affordable.

**References:**.

**Grants:**.

**Conflict of Interest:** None declared.

### P18.069.A Hybrid semi-automated approach for neonatal screening using whole exome sequencing

***JEKATERINA SHUBINA***[1], *Ekaterina Lozier*[1], *Vladislav Tkachuk*[2], *Igor Sadelov*[1], *Dmitry Trofimov*[1]

[1]*FSBI «National medical research center for obstetrics, gynecology and perinatology named after academician V.I.Kulakov» ministry of healthcare of the Russian federation, Moscow, Russian Federation;* [2]*DNA-technology, Moscow, Russian Federation.*

**Background/Objectives:** The advances in high-throughput sequencing technologies have provided powerful tools for the analysis of genomic variants and allowed for newborn screening that included a huge amount of rare hereditary disorders that would not be detected using conventional screening methods. However the step of variant interpretation remains time and labor consuming and requires a highly qualified specialist. Therefore, a manual analysis of every sample is impossible for routine screening. At the same time, a fully automated approach is not the best decision due to the huge amount of variant data with ambiguous clinical relevance. Therefore, we decided to apply a combined approach that allows for automated analysis of samples with no "suspicious" variants followed by manual analysis of those samples that carry variants of interest.

**Methods:** The variants were called using GATK best practices pipeline. The automated variant prioritization was performed for