



P15.002.B

Benefits and risks of gnomAD v4 in a diagnostic setting

Stephan Drukewitz, Maximilian Radtke, Amica C. Mueller-Nedebock, Konrad Platzer, Rami Abou Jamra

Introduction

The Genome Aggregation Database (gnomAD) is the largest database of genetic variations and is one of the most essential reference for variant interpretation. In November 2023, version 4.0 of this database was released, including whole exome (WES) and whole genome (WGS) sequencing data from 807,162 individuals. Especially the increase from 125,748 to 730,947 WES, compared to gnomAD v.2.1.1, should cover a broader genetic diversity of the coding regions. A more complete database for human genetic variations is beneficial when analyzing sequencing data in a diagnostic setting.

Collecting more samples from different sources also increases the risk to include variants responsible for undiagnosed genetic disorders. However, adding previously uncovered variants to this heavily used reference database can also lead to the need of reclassification of variants initially described as pathogenic due to their absence in gnomAD.

Results

Filtering for 'potentially pathogenic' ClinVar variants newly present in gnomAD v.4.0.0. but missing in gnomAD resulted in 7498 variants. Looking at developmental delay alone, we identified 252 pathogenic variants in 58 of 129 genes associated with this phenotype. 120 of those variants are exclusively present in gnomAD v.4.0.0. For 15 of those genes, no pathogenic ClinVar variants were identified in gnomAD v.2.1.1. Analyzing our in-house variant dataset, we identified 179 pathogenic and 136 likely-pathogenic variants, present exclusively in v.4.0.0 and thus potentially impacting variant classification and necessitating an adaptation of our in-house filtering steps (Fig.1, Fig. 2). Analyzing 2000 WES datasets resulted in a median reduction of rare variants after filtering from ~35%, when switching from gnomAD v2.1.1 to gnomAD v.4.0.0 (Figure 3).

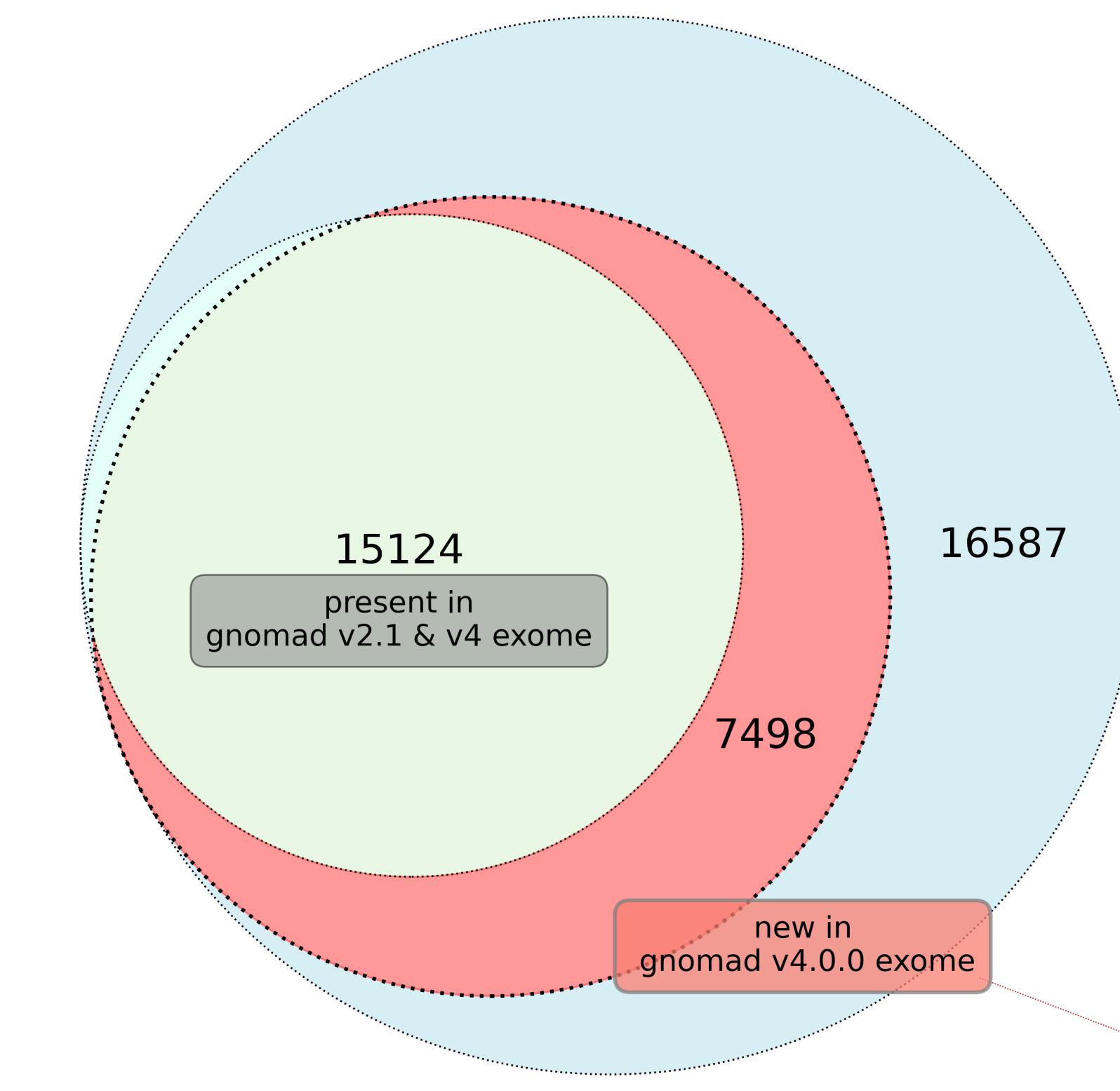


Figure 1. Overlap of potentially pathogenic ClinVar variants with the gnomAD v2.1.1 and v4.0.0 exome dataset. Pathogenic variants present in gnomAD v2.1.1 and v4.0.0 light green, Pathogenic variants present in v4.0.0 but not v2.1.1 light red, Pathogenic variants present in none of the gnomAD datasets light blue.

Discussion

The rapid growth of gnomAD will undoubtedly help to increase the quality of genetic diagnostics, especially for currently underrepresented ethnicities. We showed that using the v.4.0.0 dataset can lead to a drastic reduction of rare variants that need further interpretation (Fig. 3).

However, the nearly five-fold increase in contributing individuals compared to previous releases requires adaptation in the way we work with this data to interpret clinically relevant variants. Here we present a first insight into the challenges and benefits of incorporating gnomAD v.4.0.0 in a contemporary NGS-based diagnostic setting. We defined a variant set of interest and extracted 'potentially pathogenic' variants newly present in the gnomAD database. We are aware that our filtering strategy is rather relaxed and can also include artefacts and/or variants of unclear significance present in the ClinVar database. Nonetheless we provide an overview for variants that are of potential interest when working in a diagnostic setting, furthermore we assessed our own ClinVar submissions. We identified 315 likely pathogenic/pathogenic variants, that we submitted to ClinVar, that are newly present in gnomAD. Those variants harbor important information's for adapting our filter strategies and at the same time are candidates that potentially need a reclassification.

References

- Kaplanis, J., Samocha, K.E., Wiel, L. et al. Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757–762 (2020)
- https://gnomad.broadinstitute.org/
- Karczewski, K.J., Franciolini, L.C., Tiao, G. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020)
- Melissa J. Landrum, Jennifer M. Lee, George R. Riley, Wonhee Jang, Wendy S. Rubinstein, Deanna M. Church, Donna R. Maglott, ClinVar: public archive of relationships among sequence variation and human phenotype, *Nucleic Acids Research*, Volume 42, Issue 1, 1 January 2014, Pages D980-D985,

Methods

To measure the potential effect of gnomAD v4 on variant interpretation, we included all 'potentially pathogenic' ClinVar variants and determined their respective allele frequencies in gnomAD v.2.1.1 versus v.4.0.0. Variants were classified as 'potentially pathogenic' when the following criteria are fulfilled:

- 3 or more submissions with a classification **Pathogenic** or **Likely pathogenic**
- no submissions with the classification **Benign** or **Likely benign**

We assessed the potential effect on our routine diagnostic work by analyzing the following subgroups:

- ~1700 pathogenic and ~1600 likely pathogenic variants, reported in patients from our facility
- filtering for variants on 129 autosomal-dominant genes associated with developmental delay
- filtering for variants absent in v2.1.1 and present in v.4.0.0

To analyze the effect of v4.0.0 on filtering strategies we annotated WES data from 2000 index patients using gnomAD v2.1.1 and v.4.0.0 and compared the variant counts using different population allele frequencies for filtering.

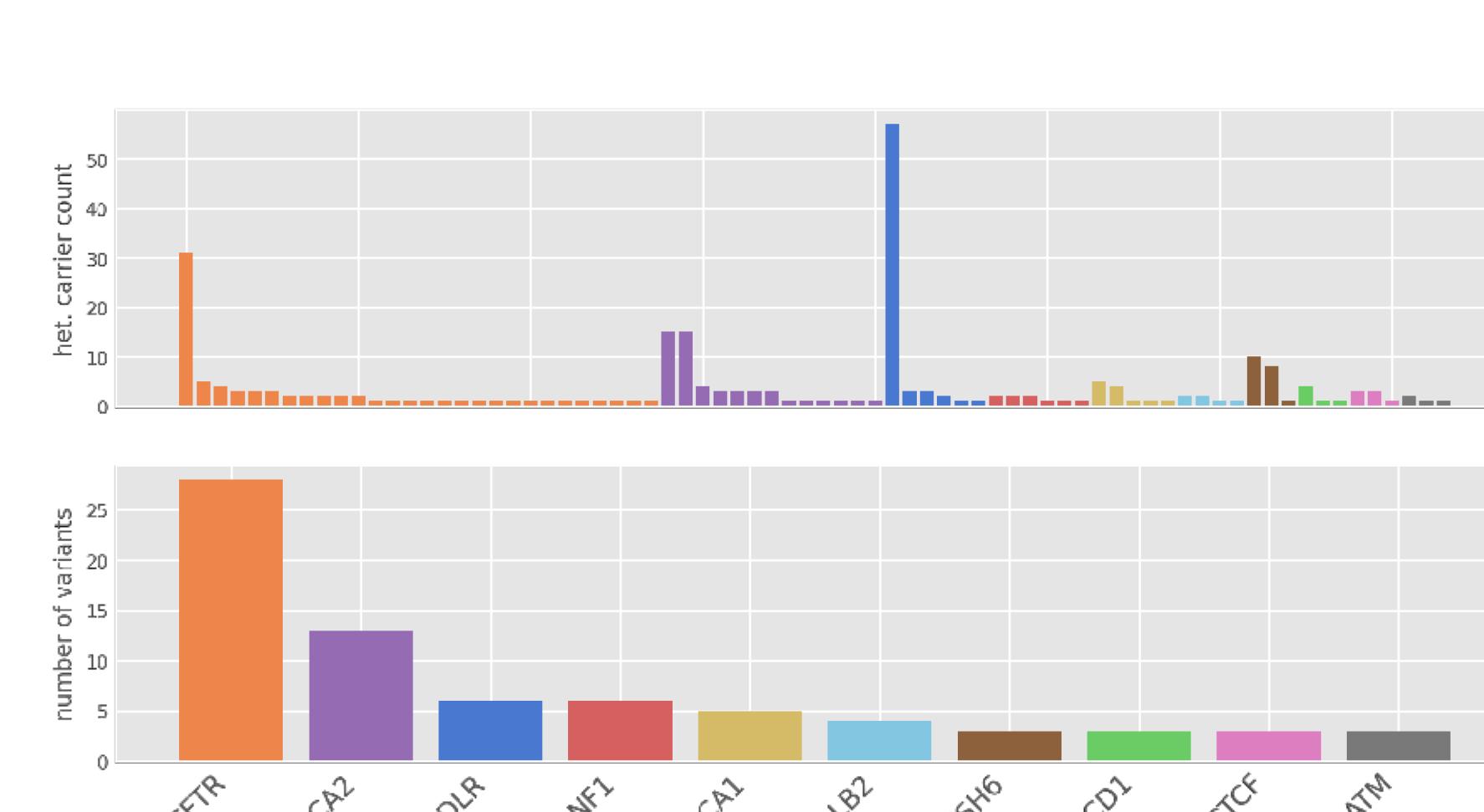


Figure 2. Overview of 'Pathogenic' variants submitted to ClinVar by the "Institute of Human Genetics, University of Leipzig Medical Center". Only variants newly present in the gnomAD v.4.0.0 exome database were included in the figure. Only the ten genes with the highest variant count are included in this figure.

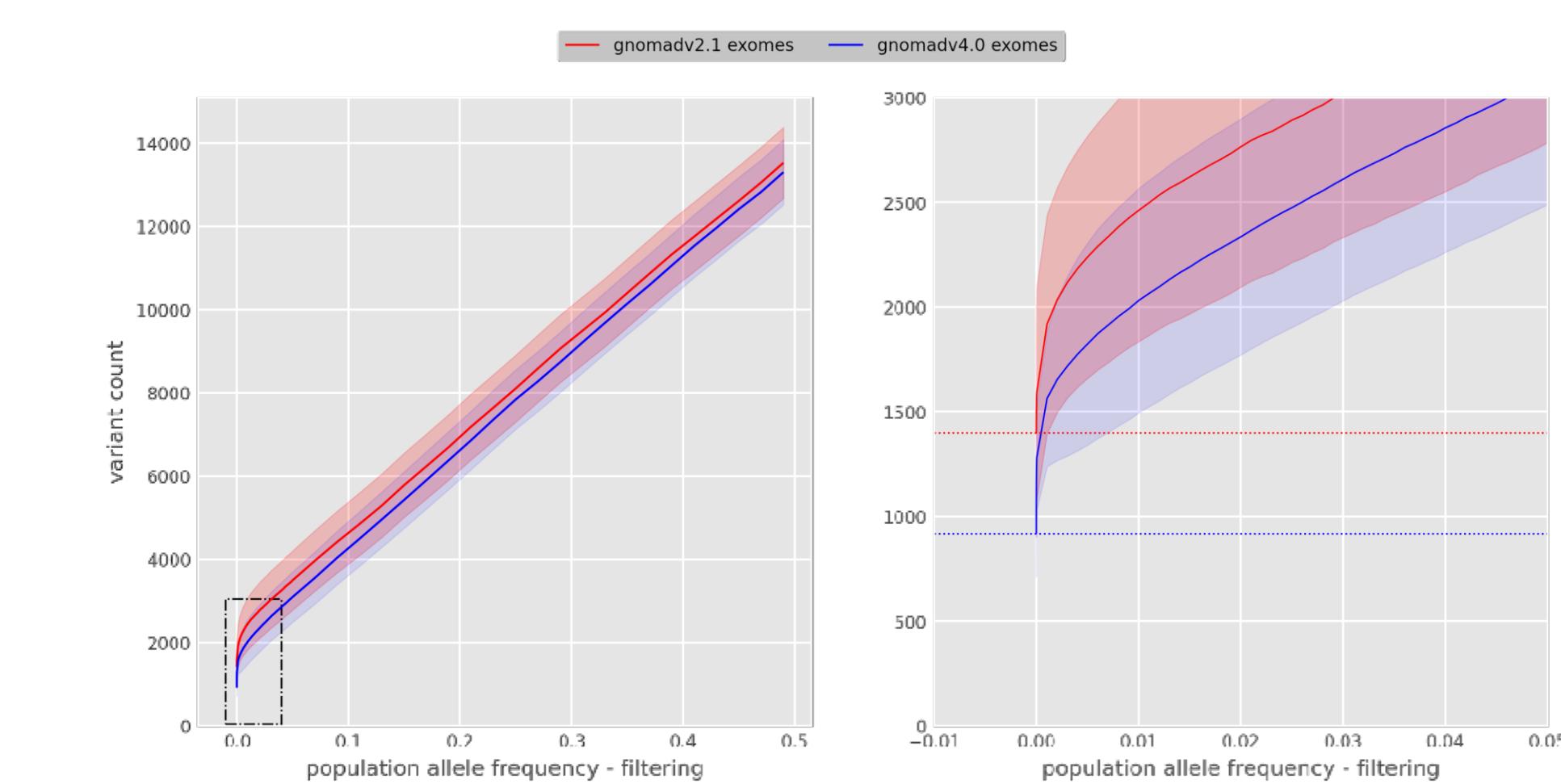


Figure 3. Filtering 2000 WES datasets using gnomAD v.2.1.1 and gnomAD v.4.0.0 using solely different population allele frequencies. Filtering for variants not present in gnomAD v.2.1.1 results in a median of 1399 variants. Filtering the same datasets for variants not present in gnomAD v.4.0.0 results in a median of 917 variants.

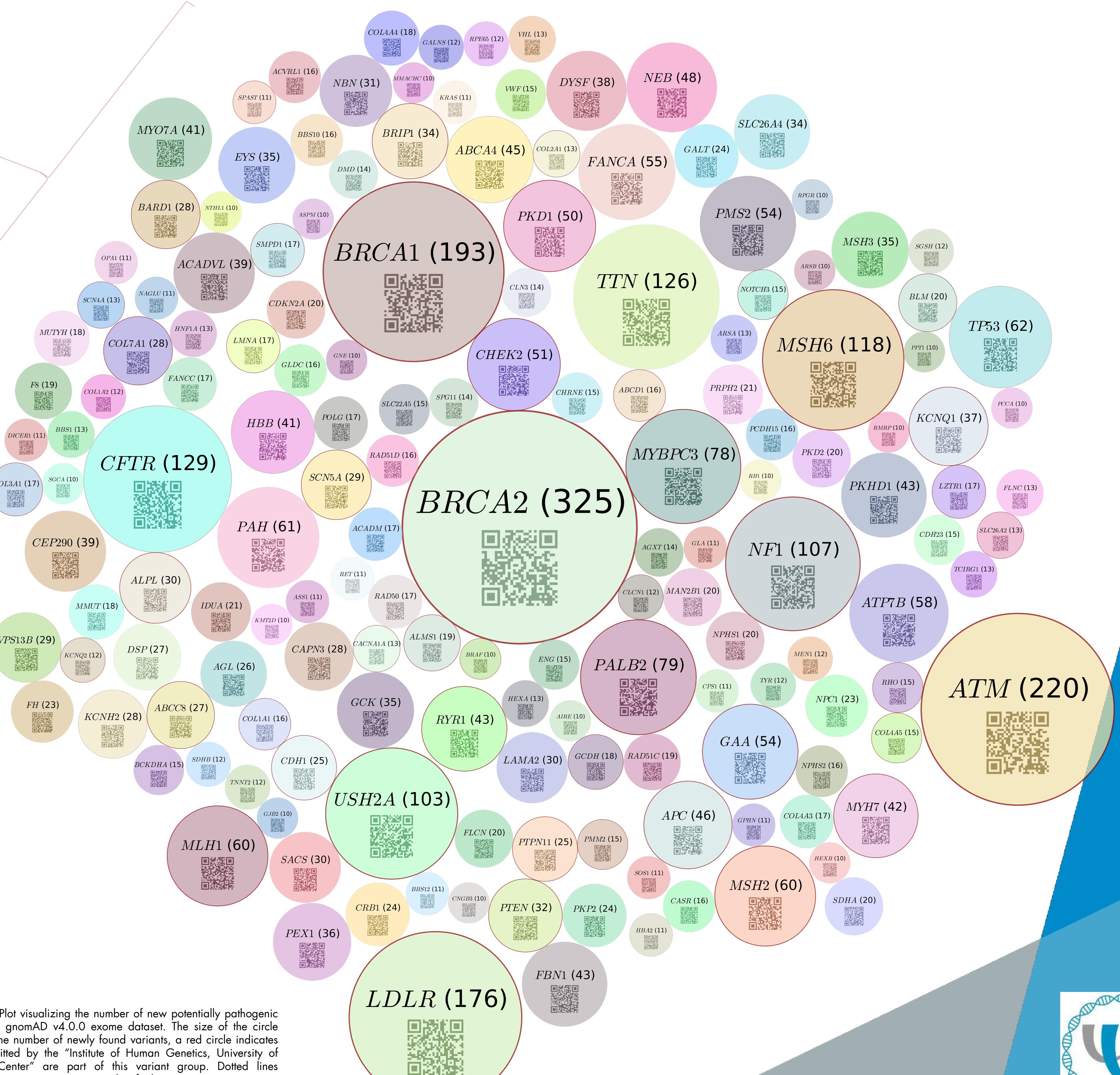


Figure 4. Bubble-Plot visualizing the number of new potentially pathogenic variants present in gnomAD v4.0.0 exome dataset. The size of the circle corresponds with the number of newly found variants, a red circle indicates that variants submitted by the "Institute of Human Genetics, University of Leipzig Medical Center" are part of this variant group. Dotted lines indicating that homozygous carriers were identified. Scanning the QR-Code will lead you to an variant overview with direct links to the ClinVar entries.

