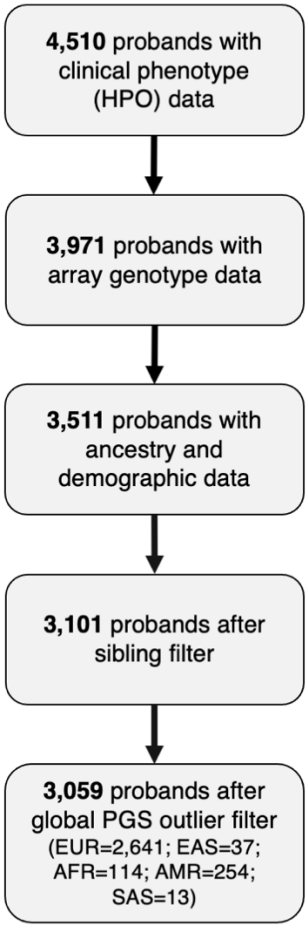
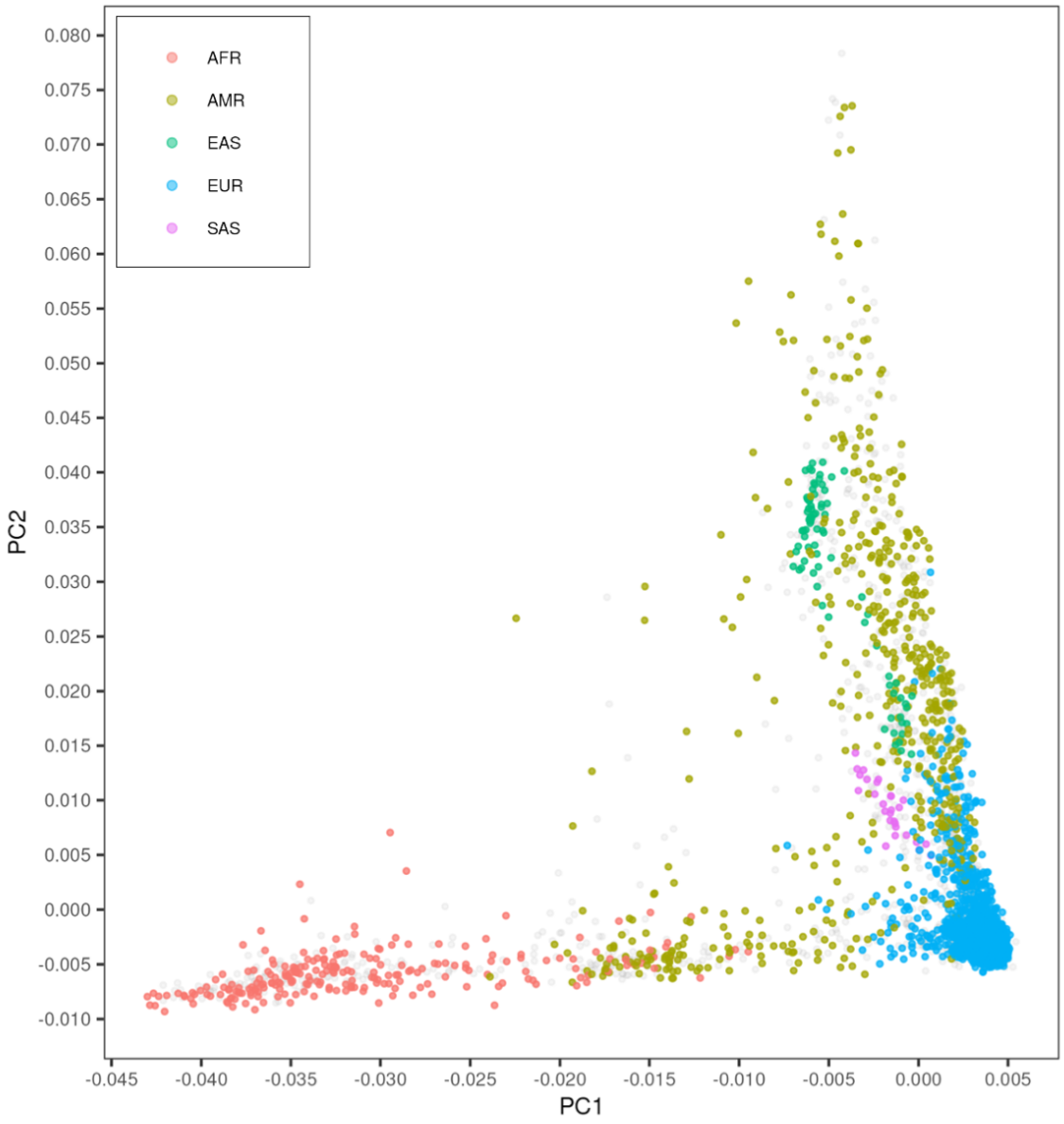
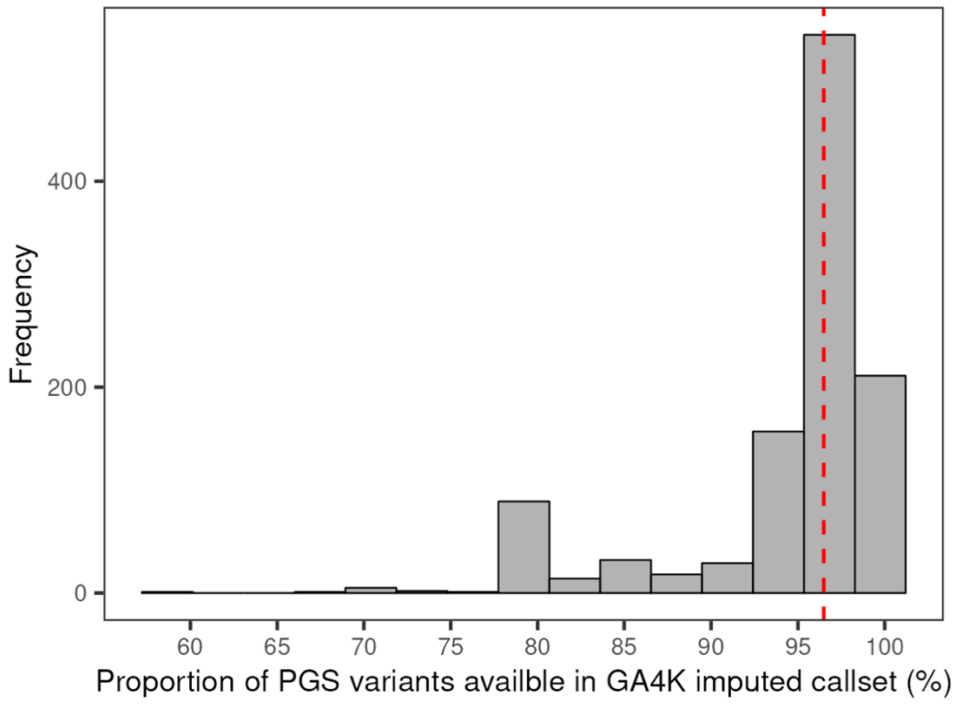


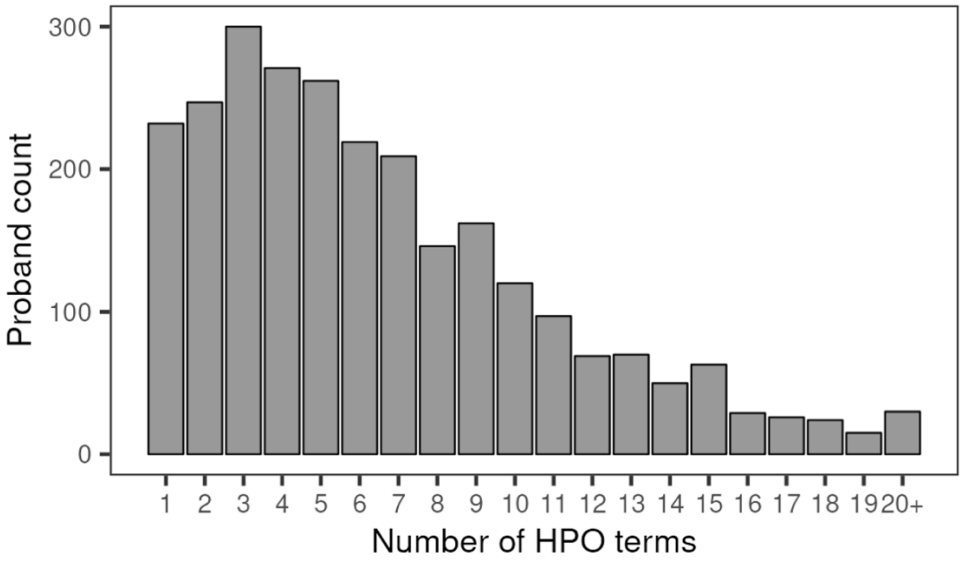
Supplementary Figure 1. Filtering steps used to define the GA4K proband cohort included in this study.



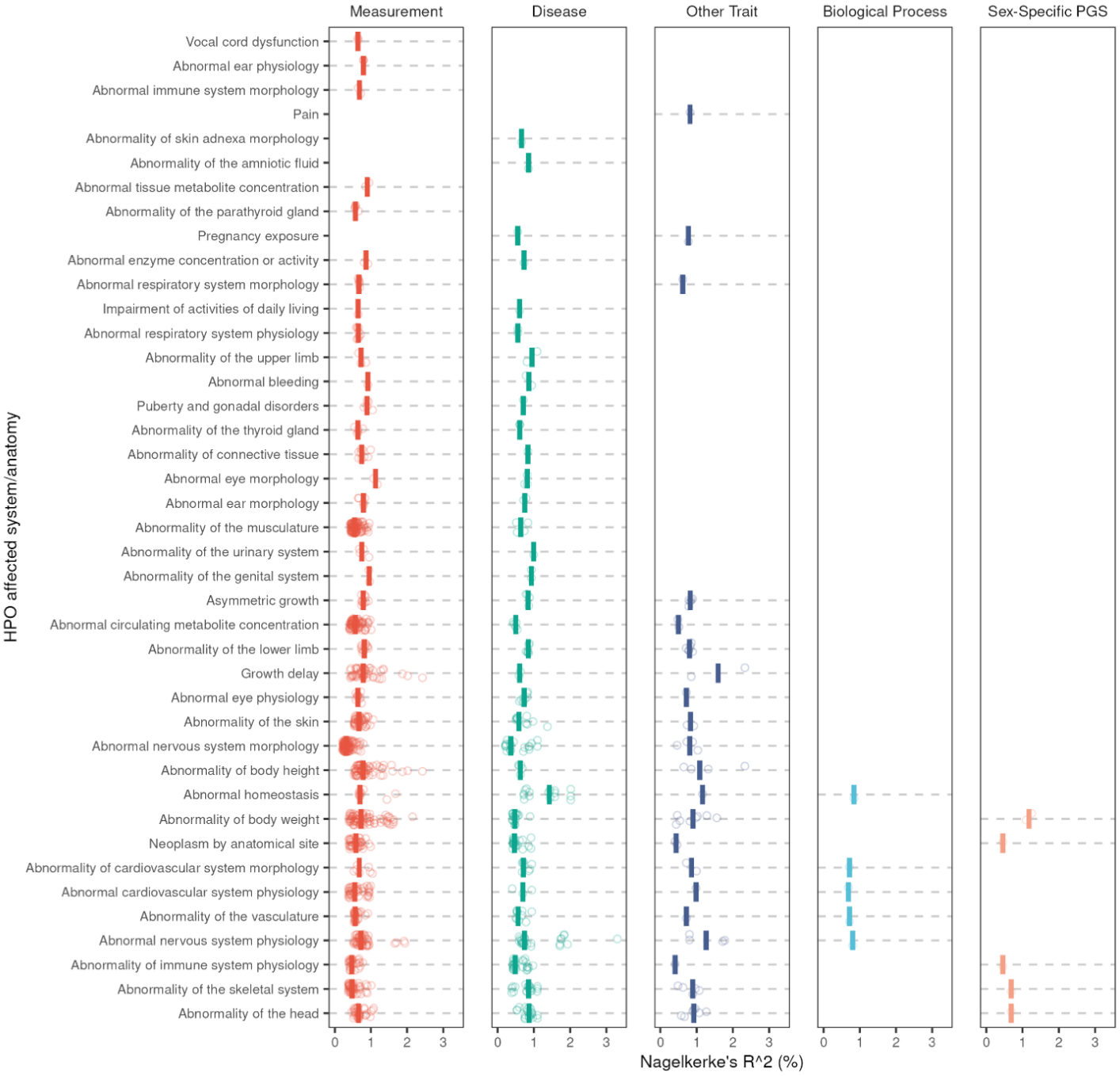
Supplementary Figure 2. Principal components analysis of genotyping data for the full GA4K cohort comprising probands and family members (N = 9,028). Dots in color indicate known ancestry labels. Principal component 1 is displayed on the x-axis, principal component 2 is displayed on the y-axis.



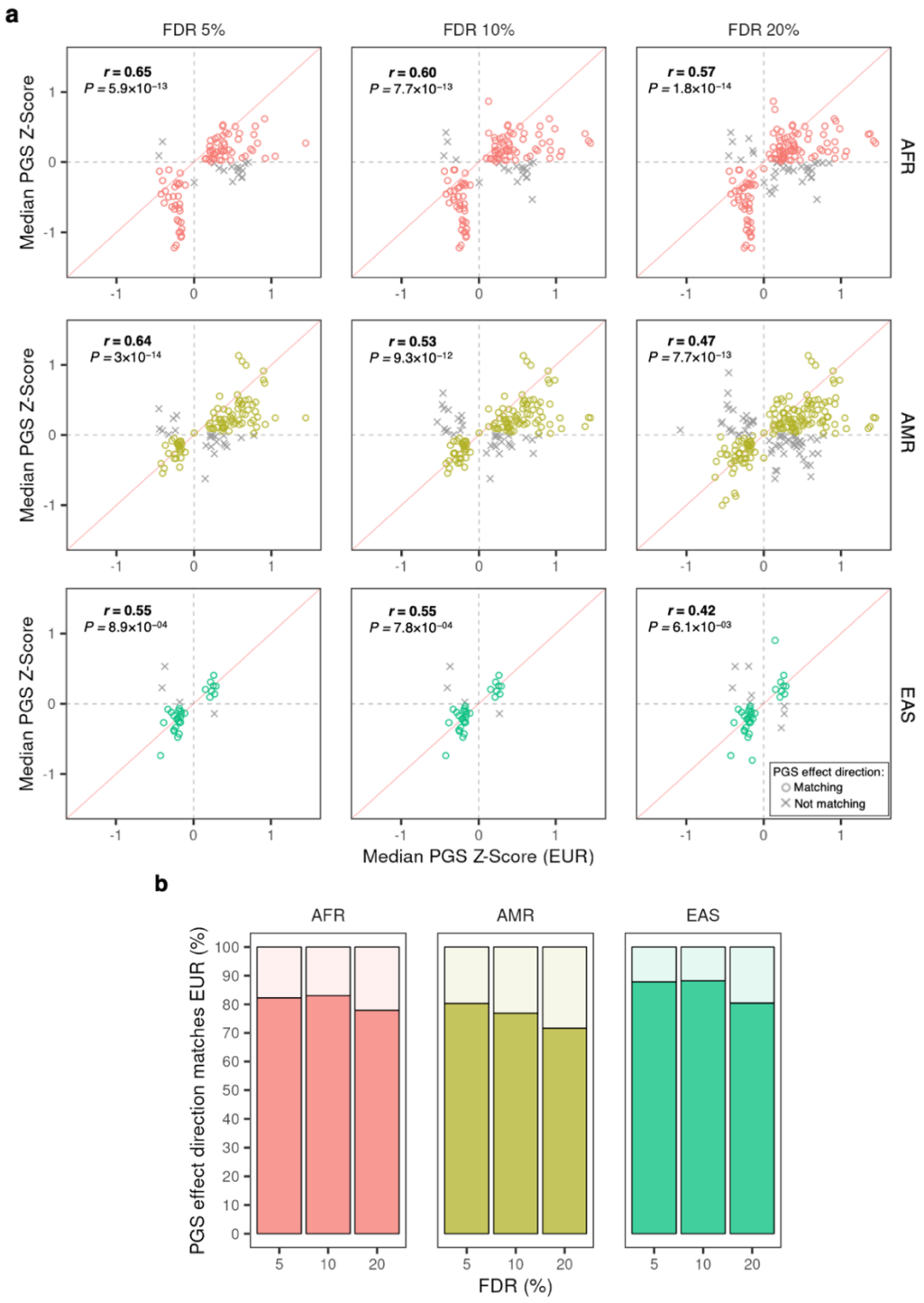
Supplementary Figure 3. Histogram showing the proportion of variants in each PGS also available in the GA4K imputed genotype callset. Red vertical dashed line indicates median proportion across all 1,102 PGS included in the present study.



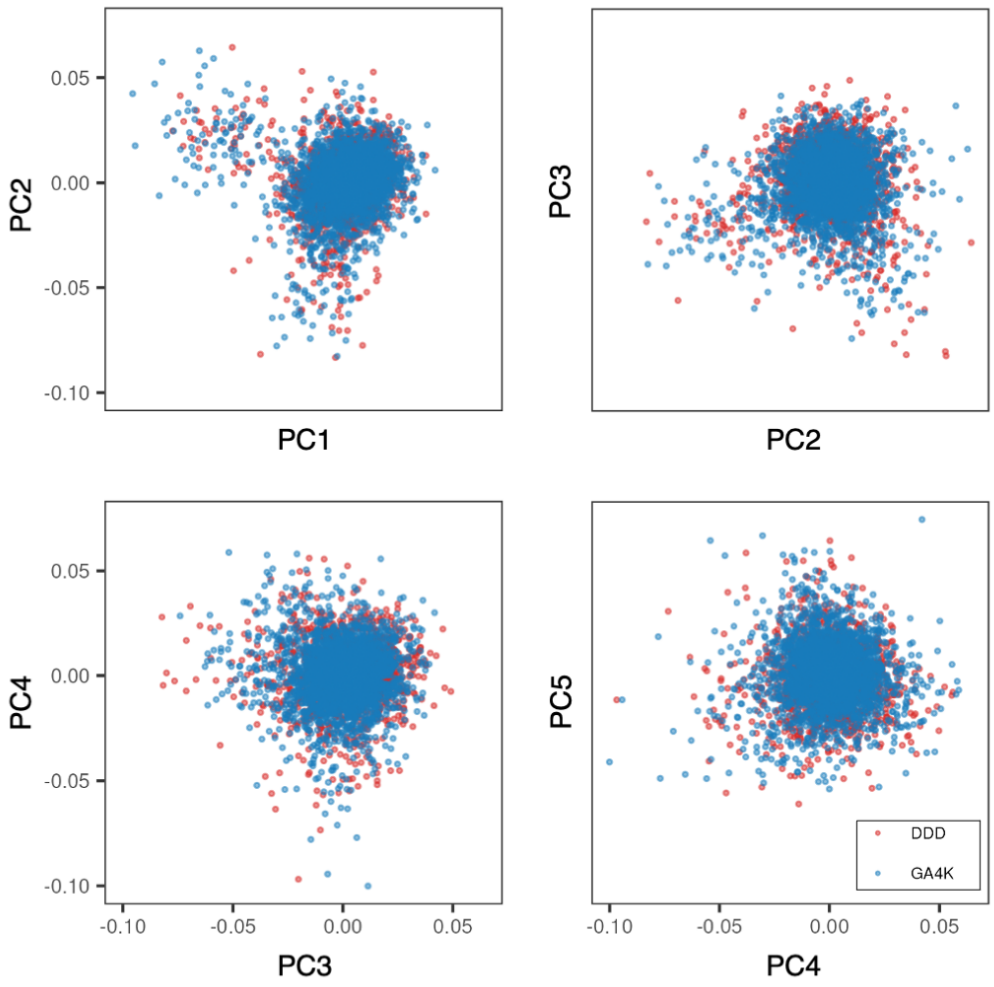
Supplementary Figure 4. Histogram showing the number of HPO terms per proband (N = 2,641 probands) in the EUR proband cohort used for PGS association testing.



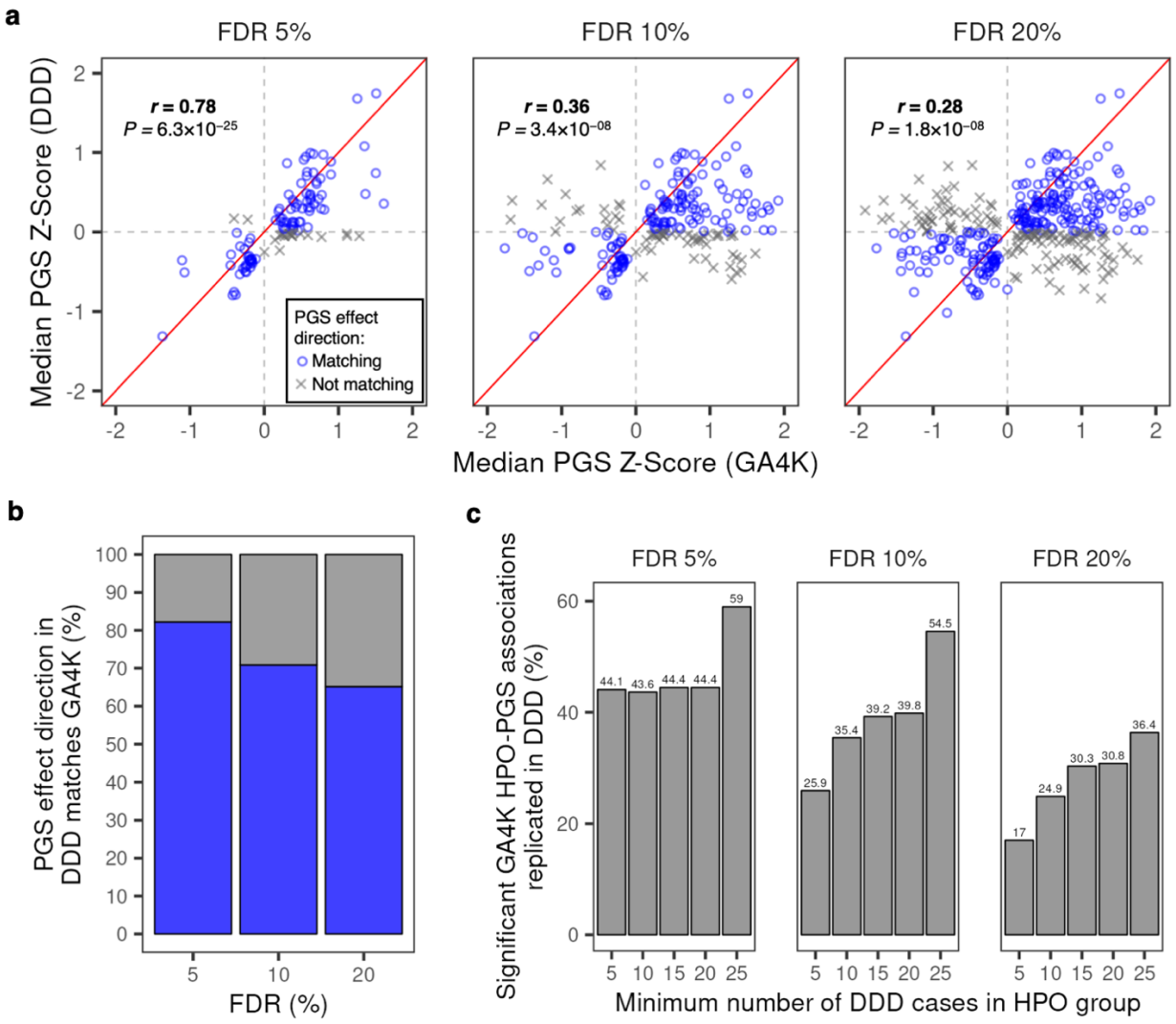
Supplementary Figure 5. Overview of logistic regression model fit (Nagelkerke's R^2) for significant HPO-PGS pairs (FDR < 20%), summarized by parent HPO and PGS categories. Vertical bars indicate median Nagelkerke's R^2.



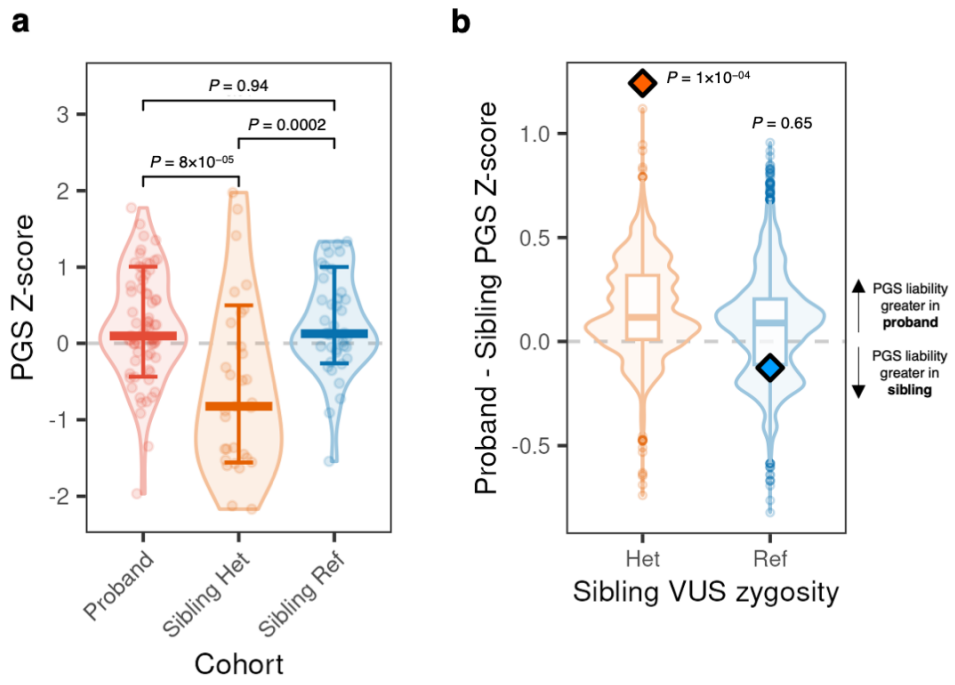
Supplementary Figure 6. Trans-ancestry replication of PGS liability in HPO case cohorts. **A.** X-axis shows median Z-score for each significantly associated PGS at the indicated FDR cutoff for the EUR proband cohort. Y-axis shows median Z-score for each significantly associated PGS at the indicated FDR cutoff for AFR (top; N probands = 114), AMR (middle; N probands = 254) and EAS (bottom; N probands = 37) cohorts. Pearson correlation coefficient tests (two-sided) and associated P-values are indicated in figure. **B.** Proportion of PGSs with matching direction of effect in EUR and AFR (left), EUR and AMR (middle), and EUR and EAS (right) cohorts.



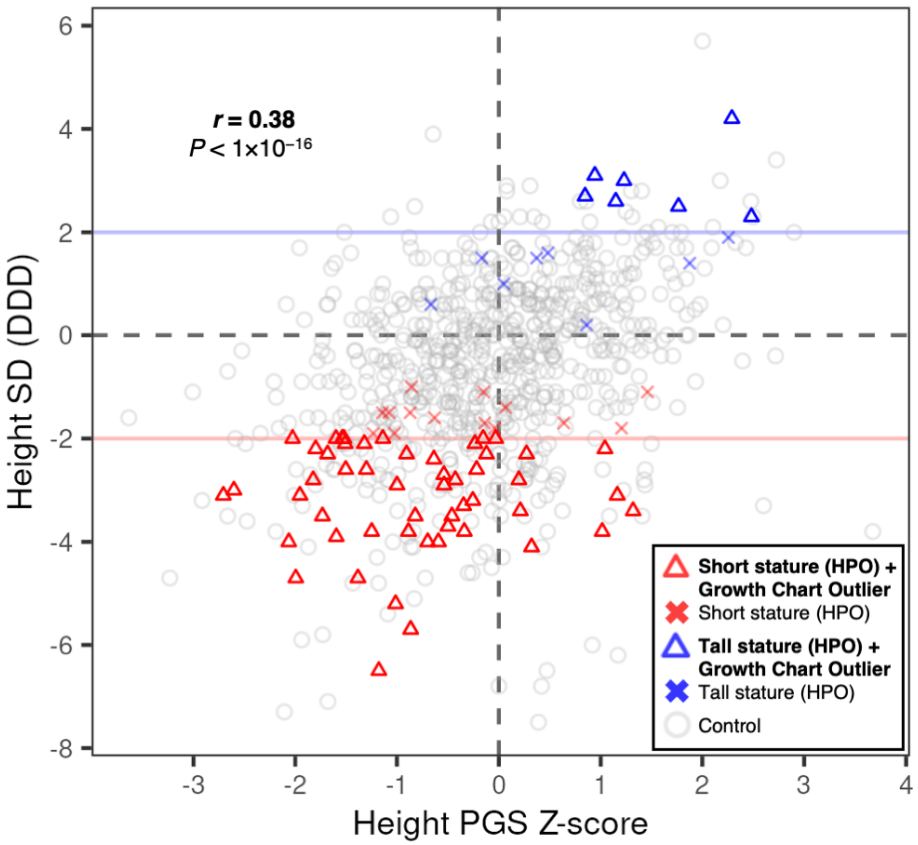
Supplementary Figure 7. Principal components analysis of genotyping data for the GA4K and DDD probands used in replication analysis (N probands: GA4K, N = 2,462; DDD, N = 1,464). Principal component 1 is displayed on the x-axis, principal component 2 is displayed on the y-axis.



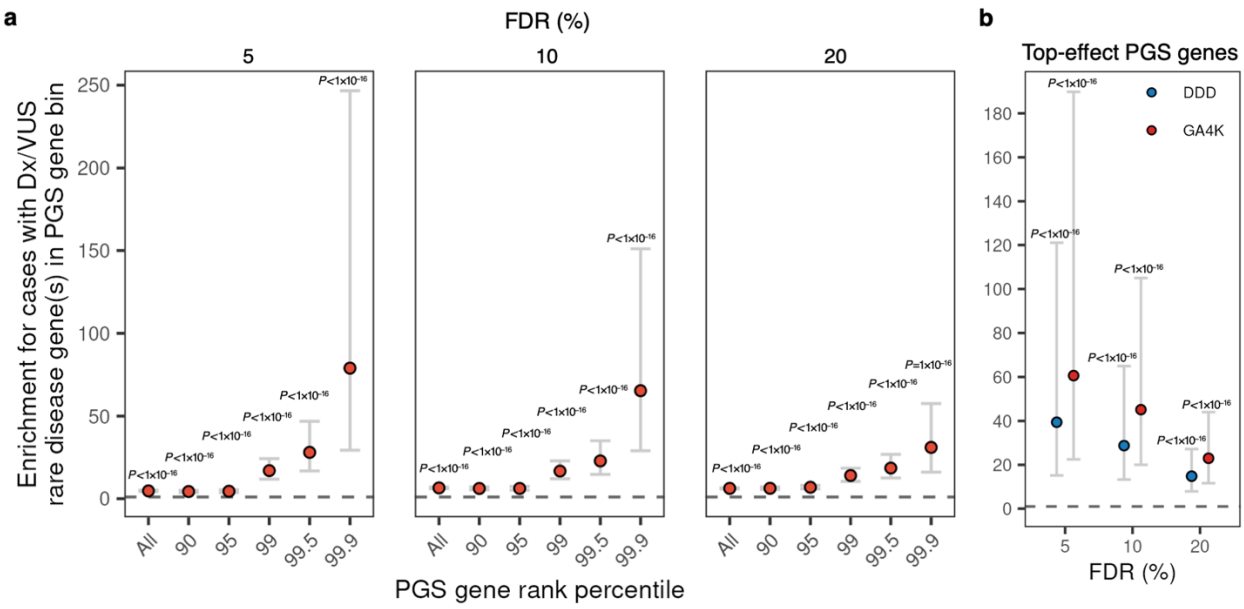
Supplementary Figure 8. Replication of PGS liability in HPO case cohorts in an external rare disease cohort (N probands = 1,464). **A.** Median PGS Z-score for each significantly associated PGS at the indicated FDR in GA4K (x-axis) and DDD (y-axis) for HPO cohorts with >=5 probands available in both GA4K and DDD. Pearson correlation coefficient tests (two-sided) and associated P-values are indicated on figure. **B.** Proportion of PGSs with matching direction of effect in both GA4K and DDD for HPO cohorts with >=5 probands available in both GA4K and DDD. **C.** Rate of replication in DDD for HPO-PGS pair associations from GA4K as a function of DDD sample size and significance threshold (FDR). Replication is defined as a matching coefficient direction of effect in a logistic regression test (two-sided), and nominal significance ($P < 0.05$) in the replication dataset.



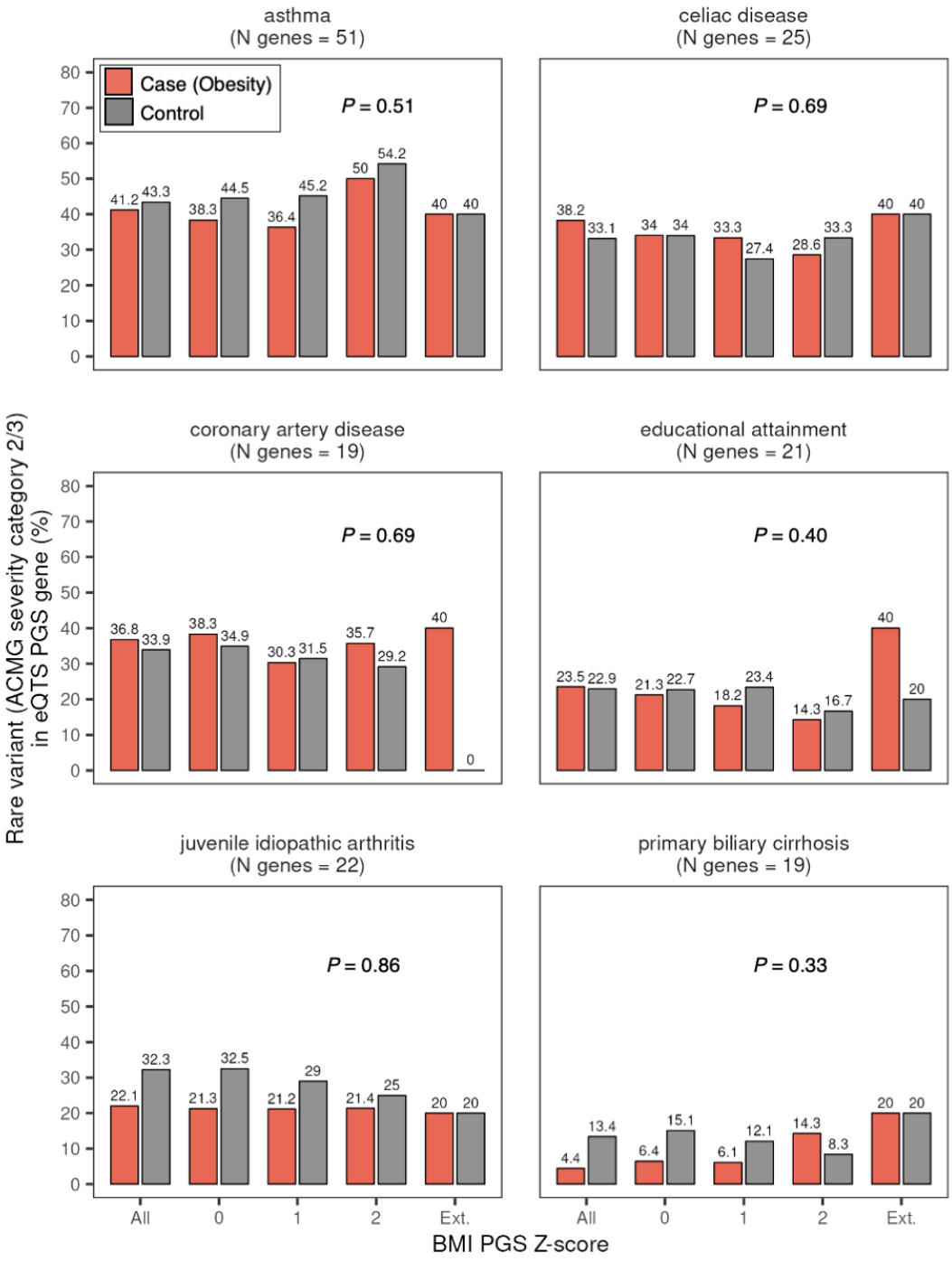
Supplementary Figure 9. Differences in PGS for probands with a clinical variant of unknown significance (VUS) and unaffected carrier sibling(s). **A.** Distribution of PGS for probands, unaffected carrier (Het) siblings, and unaffected non-carrier (Ref) siblings. PGS with association direction of effect less than zero are inverted to enable visualization. Cross bars indicate median, error bars indicate standard deviation. P-values are from a Wilcoxon Rank Sum test (two-sided). No adjustments were made for multiple comparisons. Results are from 72 HPO:PGS associations across three proband/sibling pairs. **B.** Proband minus sibling PGS Z-score/standard deviation for significantly associated PGS in probands with a clinical VUS compared to unaffected carrier (Het) and non-carrier (Ref) siblings. Points above zero indicate PGS liability is greater in the proband and points below zero indicate PGS liability is greater in the sibling. Pairwise differences for PGS with association direction of effect less than zero were inverted to enable visualization. Diamonds denote median PGS for PGS linked with VUS HPO(s) for carrier (orange) and non-carrier (blue) siblings. Box and violin plots denote distribution of median PGS for randomly selected PGS not associated with VUS HPO(s), where – for boxplots - the middle line corresponds to the median, the lower and upper edges of the box correspond to the first and third quartiles, the whiskers represent the interquartile range (IQR) ×1.5 and beyond the whiskers are outlier points. P-values are derived from the empirical distribution of these background PGS (N permutations = 10,000). No adjustments were made for multiple comparisons. Results are from 72 HPO:PGS associations across three proband/sibling pairs.



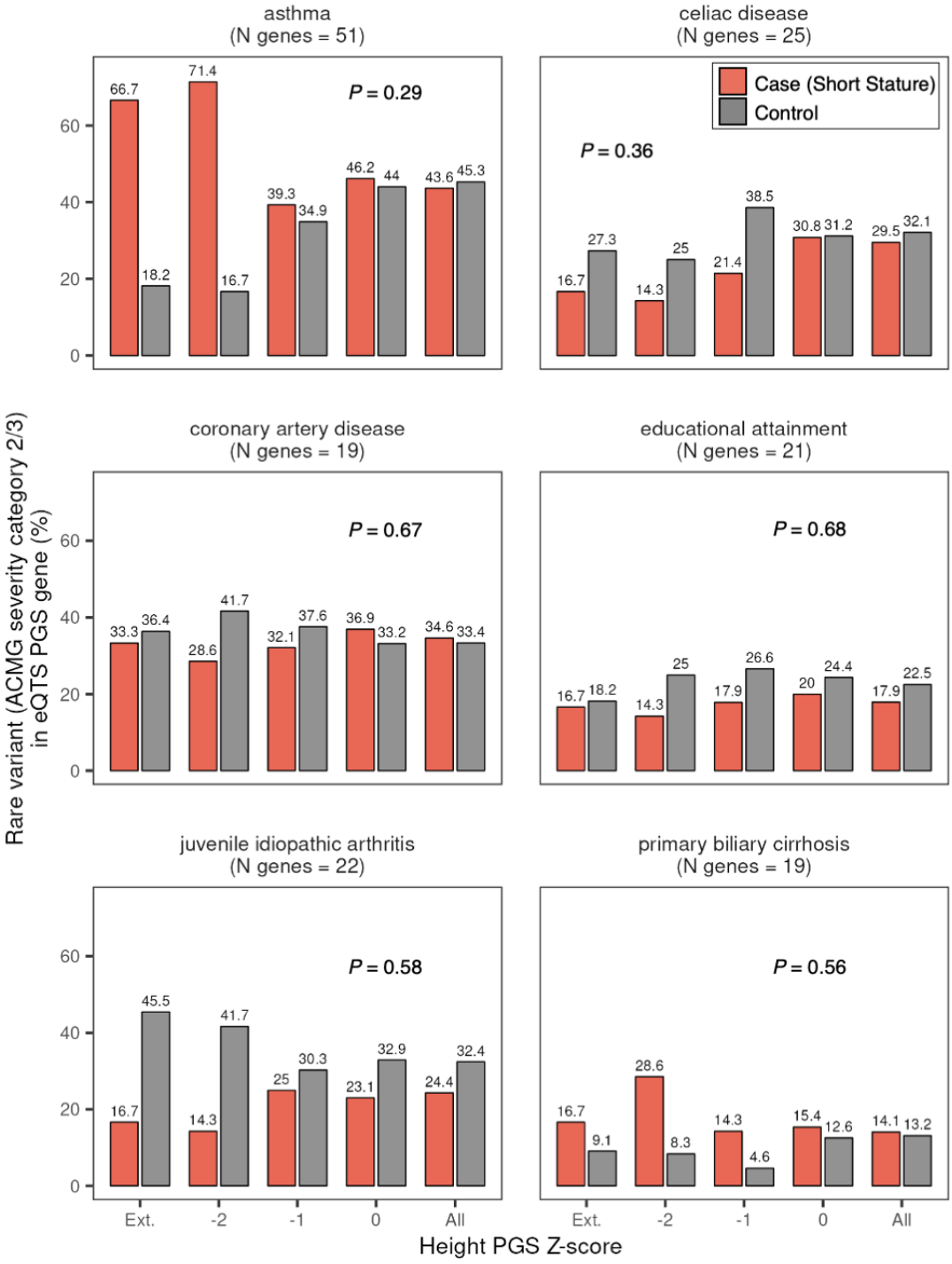
Supplementary Figure 10. Integrating PGS and clinical observation data in DDD (N probands = 766). Individual height observations for DDD probands indicate height Z-scores and are displayed on the y-axis. Individual scores for a significantly associated height PGS are displayed on the x-axis. Red horizontal line indicates short stature threshold (CDC). Blue horizontal line indicates tall stature threshold (CDC). Probands with a short stature HPO in DDD are indicated by a red triangle if observed height value passes CDC threshold for short stature, or a red cross if not passing this threshold. Probands with tall stature HPO are indicated by a blue triangle if observed height value passes CDC threshold for tall stature, or a blue cross if not passing this threshold. Gray indicates control (no short or tall stature HPO terms present in DDD). Pearson correlation coefficient test (two-sided) and associated P-value is indicated in figure. Pearson correlation coefficient $P = 3\times10^{-26}$.



Supplementary Figure 11. Enrichment for diagnostic or candidate diagnostic rare disease variants in associated PGS genes. **A.** Enrichment for HPO cases versus controls (proband N = 538) for probands with a diagnostic or candidate (VUS) rare disease variant in an associated PGS gene with indicated effect rank percentile in PGS, across significance (FDR) thresholds. Dots indicate the odds ratio from logistic regression (two-sided), error bars indicate 95% confidence interval. P-values for PGS gene rank percentile: FDR 20%: All, $P = 0$; 90, $P = 3\times10^{-127}$; 95, $P = 5\times10^{-89}$; 99, $P = 1\times10^{-45}$; 99.5, $P = 1\times10^{-31}$; 99.9, $P = 1\times10^{-17}$; FDR 10%: All, $P = 0$; 90, $P = 2\times10^{-95}$; 95, $P = 7\times10^{-62}$; 99, $P = 1\times10^{-42}$; 99.5, $P = 2\times10^{-30}$; 99.9, $P = 1\times10^{-20}$; FDR 5%: All, $P = 1\times10^{-227}$; 90, $P = 2\times10^{-58}$; 95, $P = 2\times10^{-40}$; 99, $P = 6\times10^{-39}$; 99.5, $P = 6\times10^{-30}$; 99.9, $P = 1\times10^{-19}$. No adjustments were made for multiple comparisons. **B.** Enrichment as in **(A.)** for large effect size PGS genes (gene rank percentile = top 0.01% genes) in GA4K (proband N = 438) and DDD (proband N = 836) probands for subset of HPO-PGS pairs available in both cohorts at indicated significance (FDR) threshold. Dots indicate the odds ratio from logistic regression (two-sided), error bars indicate 95% confidence interval. P-values: FDR 20%: GA4K, $P = 3\times10^{-15}$; DDD $P = 9\times10^{-15}$; FDR 10%: GA4K, $P = 4\times10^{-18}$; DDD $P = 3\times10^{-17}$; FDR 5%: GA4K, $P = 6\times10^{-18}$; DDD $P = 4\times10^{-17}$. No adjustments were made for multiple comparisons.



Supplementary Figure 12. Overlap in non-disease matched putative core/key PGS genes (eQTS genes) for cases compared to controls for rare variants prioritized in to ACMG severity categories 2 or 3. P-values are derived from logistic regression (two-sided), testing for an interaction between case/control cohort and PGS Z-score on rare variant burden in indicated eQTS gene set. Ext. denotes maximum outlier PGS cutoff that could be assessed where the cohort contained at least one case with a selected rare variant in an eQTS gene.



Supplementary Figure 13. Overlap in non-disease matched putative core/key PGS genes (eQTS genes) for cases compared to controls for rare variants prioritized in to ACMG severity categories 2 or 3. P-values are derived from logistic regression (two-sided), testing for an interaction between case/control cohort and PGS Z-score on rare variant burden in indicated eQTS gene set. Ext. denotes maximum outlier PGS cutoff that could be assessed where the cohort contained at least one case with a selected rare variant in an eQTS gene.

Genomic Answers for Kids (GA4K) Consortium Author List

Ana S.A. Cohen[1,2,3]
Emily G. Farrow[1,3,4]
Ahmed T. Abdelmoity[4]
Joseph T. Alaimo[2,3]
Shivarajan M. Amudhavalli[3,5]
John T. Anderson[6]
Lalit Bansal[4]
Lauren Bartik[3,5]
Primo Baybayan[7]
Bradley Belden[1]
Courtney D. Berrios[1]
Rebecca L. Biswell[1]
Pawel Buczkowicz[8]
Orion Buske[8]
Shreyasee Chakraborty[7]
Warren A. Cheung[1]
Keith A. Coffman[4]
Ashley M. Cooper[4]
Laura A. Cross5
Tom Curran[9]
Thuy Tien T. Dang[4]
Mary M. Elfrink[1]
Kendra L. Engleman[5]
Erin D. Fecske[4]
Cynthia Fieser[4]
Keely Fitzgerald[4]
Emily A. Fleming[5]
Randi N. Gadea[5]
Jennifer L. Gannon[5]
Rose N. Gelineau-Morel[3,4]
Margaret Gibson[1]
Jeffrey Goldstein[4]
Elin Grundberg[1]
Kelsee Halpin[3,4]
Brian S. Harvey[6]
Bryce A. Heese[5]
Wendy Hein[4]
Suzanne M. Herd[1]
Susan S. Hughes[5]
Mohammed Ilyas[3,4]
Jill Jacobson[3,4]
Janda L. Jenkins[5]

Shao Jiang[10]
Jeffrey J. Johnston[1]
Kathryn Keeler[6]
Jonas Korlach[7]
Jennifer Kussmann[5]
Christine Lambert[7]
Caitlin Lawson[5]
Jean-Baptiste Le Pichon[4]
James Steven Leeder[1]
Vicki C. Little[4]
Daniel A. Louiselle[1]
Michael Lypka[10]
Brittany D. McDonald[1]
Neil Miller[1,3,11]
Ann Modrcin[4]
Annapoorna Nair[1]
Shelby H. Neal[1]
Christopher M. Oermann[4]
Donna M. Pacicca[6]
Kailash Pawar[4]
Nyshele L. Posey[1]
Nigel Price[6]
Laura M.B. Puckett[1]
Julio F. Quezada[3,4]
Nikita Raje[3,12]
William J. Rowell[7]
Eric T. Rush[3,5,13]
Venkatesh Sampath[14]
Carol J. Saunders[1,2,3]
Caitlin Schwager[5]
Richard M. Schwend[6]
Elizabeth Shaffer[4]
Craig Smail[1]
Sarah Soden[4]
Meghan E. Strenk[5]
Bonnie R. Sullivan[5]
Brooke R. Sweeney[3,4]
Jade B. Tam-Williams[4]
Adam M. Walter[1]
Holly Welsh[5]
Aaron M. Wenger[7]
Laurel K. Willig[4]
Yun Yan[3,4]
Scott T. Younger[1]

Dihong Zhou[5]
Tricia N. Zion[1,3,4,5]
Isabelle Thiffault[1,2,3]
Tomi Pastinen[1,3]

Affiliations:

1. Genomic Medicine Center, Children's Mercy Research Institute and Children's Mercy Kansas City, Kansas City, MO, USA
2. Department of Pathology and Laboratory Medicine, Children's Mercy Kansas City, Kansas City, MO, USA
3. UKMC School of Medicine, University of Missouri Kansas City, Kansas City, MO, USA
4. Department of Pediatrics, Children's Mercy Kansas City, Kansas City, MO, USA
5. Division of Genetics, Children's Mercy Kansas City, Kansas City, MO, USA
6. Department of Orthopedic Surgery, Children's Mercy Kansas City, Kansas City, MO, USA
7. Pacific Biosciences of California, Inc, Menlo Park, CA, USA
8. PhenoTips, Toronto, Canada
9. Children's Mercy Research Institute, Kansas City, MO, USA
10. Bionano Genomics, Inc, San Diego, CA, USA
11. Division of Allergy Immunology Pulmonary and Sleep Medicine, Children's Mercy Kansas City, Kansas City, MO, USA
12. Division of Neonatology, Children's Mercy Kansas City, Kansas City, MO, USA
13. Department of Internal Medicine, University of Kansas School of Medicine, Kansas City, MO, USA
14. Division of Neonatology, Children's Mercy Hospital Kansas City, Kansas City, MO, USA