The American Journal of Human Genetics, Volume *105*

Supplemental Data

Phenome-wide Burden of Copy-Number Variation in the UK Biobank

Matthew Aguirre, Manuel A. Rivas, and James Priest

Supplementary Figures and Tables:

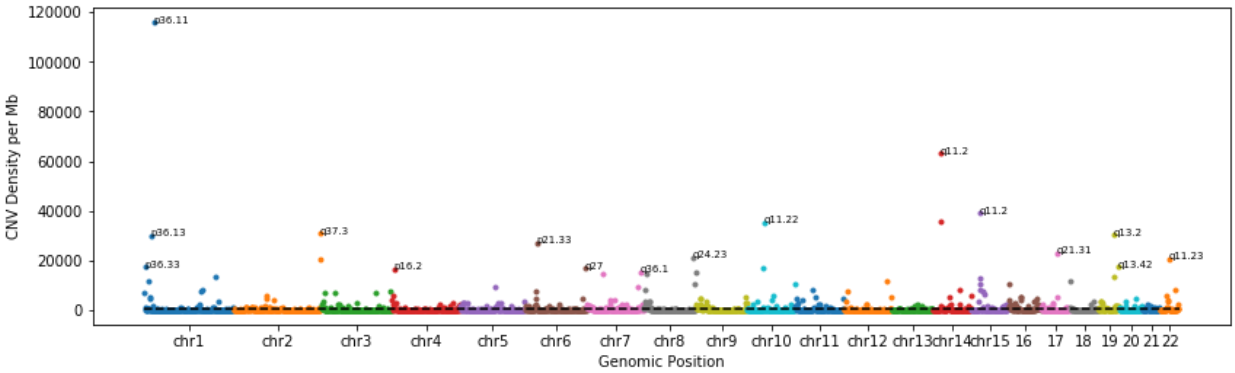


Figure S1: CNV density weighted by allele count in UK Biobank. Per-megabase genomic density of CNV, weighted by number of observations across all samples in UK Biobank. Variants are counted by whether the CNV has any overlap with 10 megabase (Mb) windows tiling each chromosome. Selected hotspots of structural variation are labeled by the region's corresponding cytogenic band.

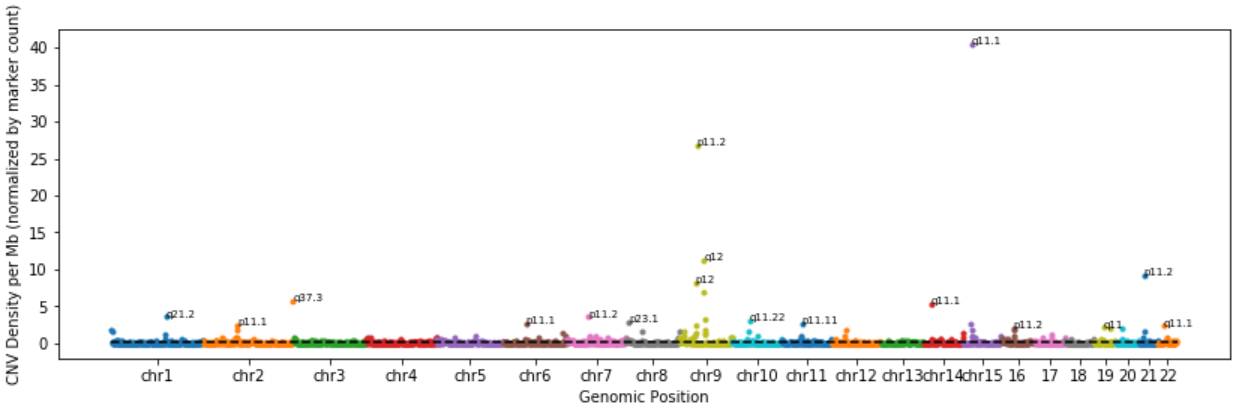


Figure S2: CNV density normalized by array marker density in UK Biobank. Variants are counted by whether the CNV has any overlap with 10 megabase (Mb) windows tiling each chromosome, then divided by the number of markers in the window. Regions with no array markers are defined to have density of zero. Selected hotspots of structural variation are labeled by the region's corresponding cytogenic band.

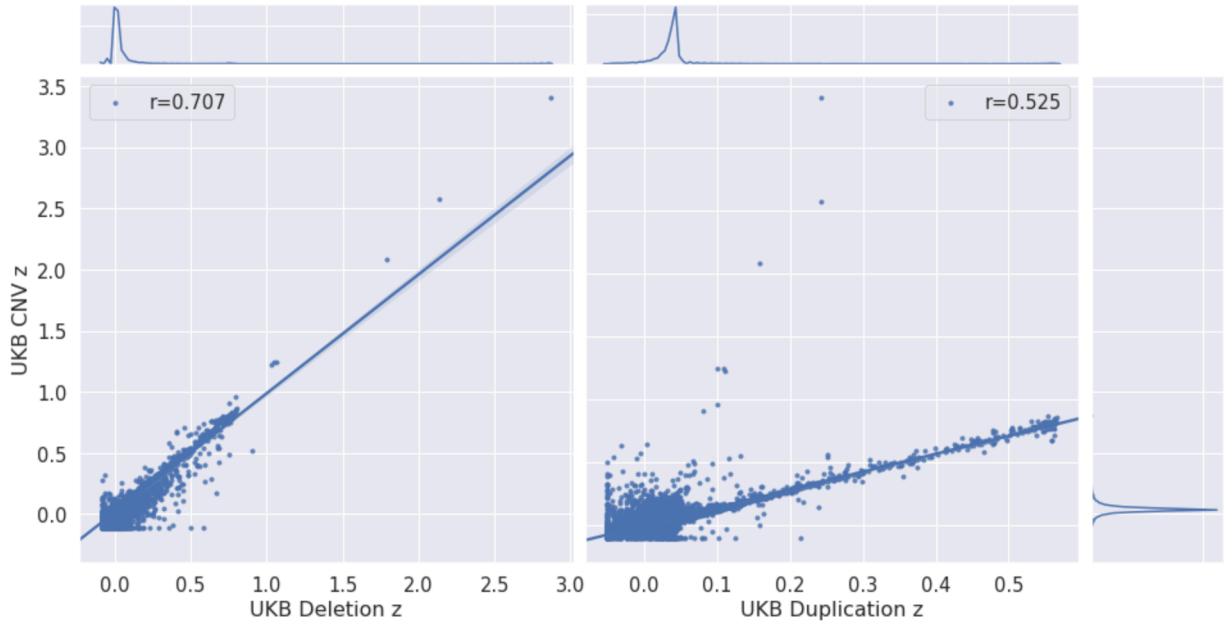


Figure S3: Distribution of deletion- and whole-gene duplication-specific constraint scores from UK Biobank. Correlation between intolerance measures for partial-gene deletion, whole-gene duplication, and CNV burden. The legend for each panel denotes correlation (Spearman's *r*) between burden-constraint and each other measure. Kernel density estimates for each distribution of constraint scores are in the panels opposite their corresponding axis labels.

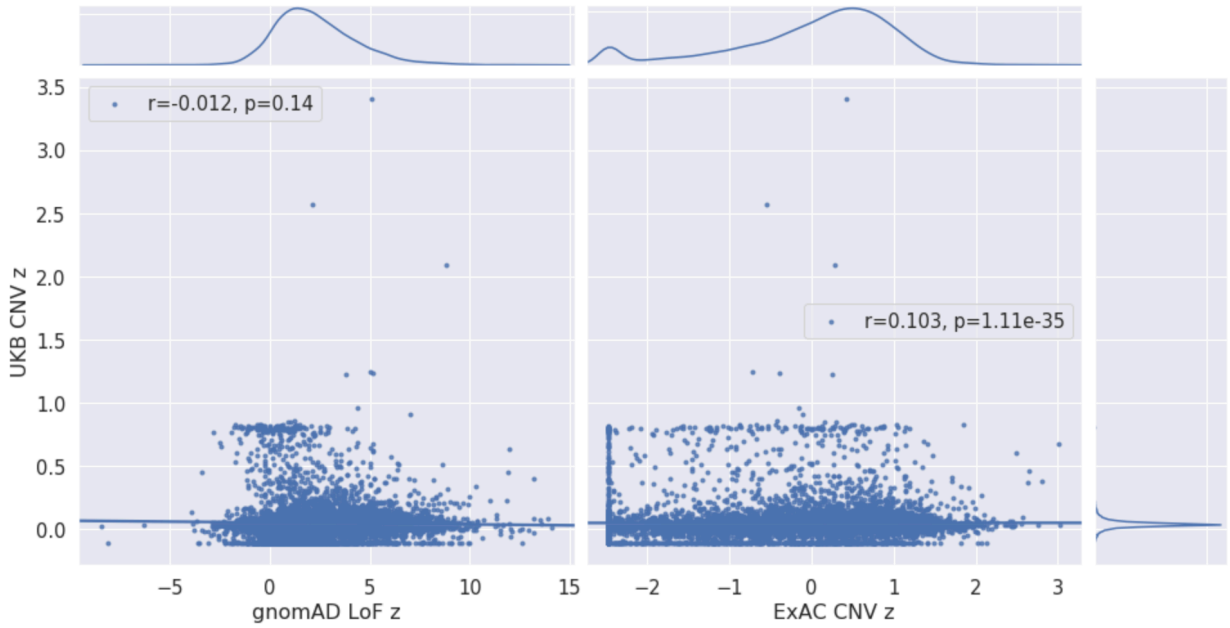
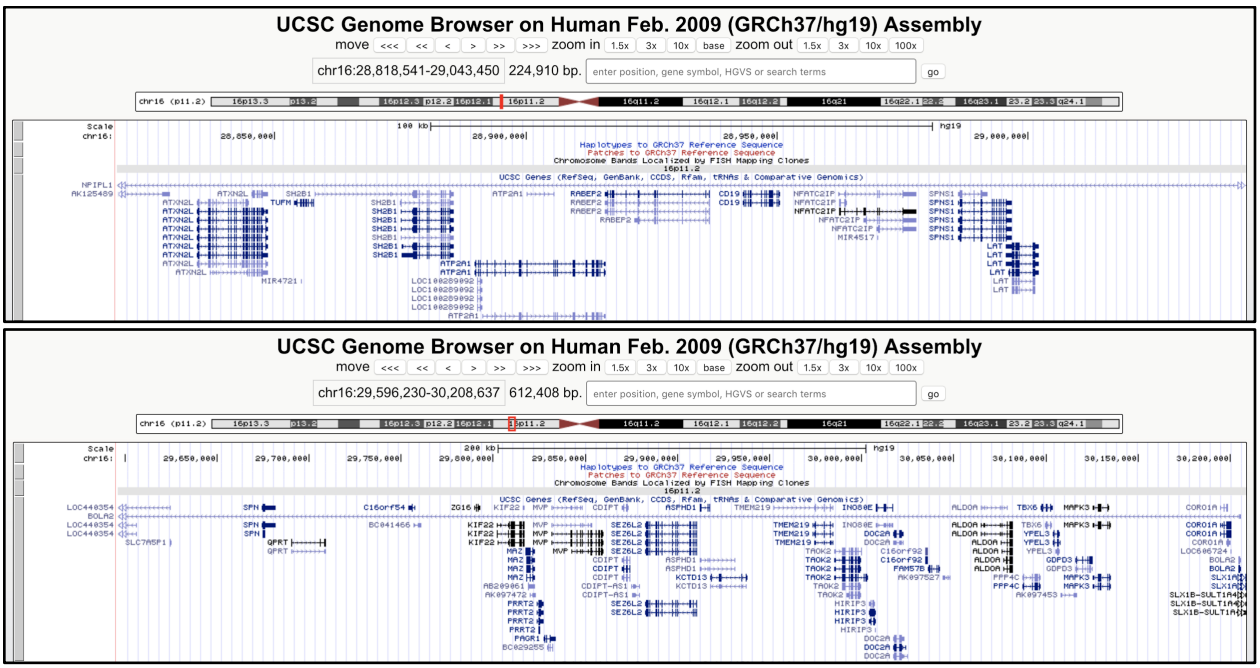
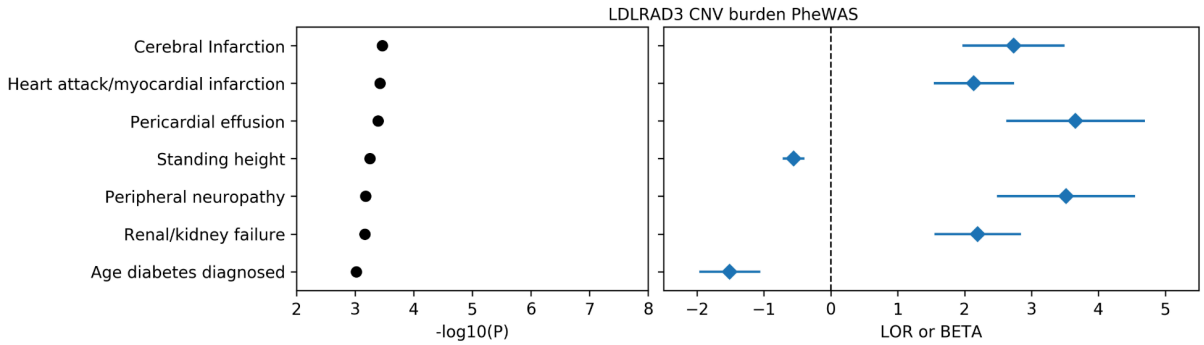


Figure S4: Distribution of constraint z-scores from UK Biobank and ExAC/gnomAD. Our measures of gene-level intolerance to structural variation show nominal correlation with gnomAD loss of function constraint *z*-scores (Spearman's *r* = -0.012, left), and modest correlation with CNV-intolerance in ExAC (Spearman's *r* = 0.103, right panel). Gaussian kernel density estimates for each distribution of *z*-scores are opposite their corresponding axes.

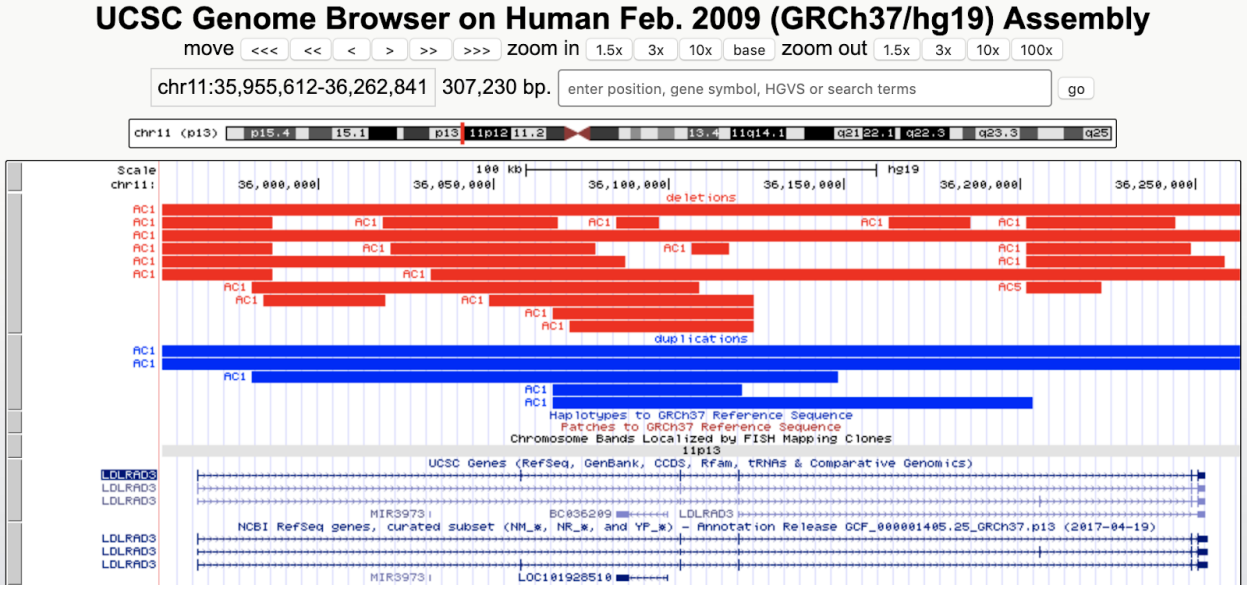
While correlation between constraint measures across datasets is non-random, we suspect cohort-specific effects and varying definitions of genic burden of variation drive these departures. As a cohort of predominantly healthy adults, intolerance to variation in UK Biobank constraint is driven by severe early onset disease, while the same measures in ExAC/gnomAD, whose samples have a more diverse age range and relatively higher of burden of disease, highlight genes involved with fundamental biological processes whose loss of function likely confer phenotypic consequences causing embryonic lethality.



<u>Figure S5:</u> Location of *16p11.2* Deletions. UCSC Genome Browser tracks for *220kb* (top panel) and *593kb* (bottom panel) CNVs at *Chr16q11.2*.



<u>Figure S6:</u> *LDLRAD3* burden test PheWas. Significant ($p < 10^{-3}$) associations between regularized burden tests for *LDLRAD3* CNV and phenotypes. We highlight quantitative traits with $n > 15{,}000$ observations and binary traits with $n > 500$ cases. Traits are grouped by data type then sorted by *p*-value (left). Log-odds ratio and standardized betas (right; for binary and quantitative traits, respectively) align with trait names on the y-axis, with the vertical dashed line separating positive and negative direction of association.



<u>Figure S7:</u> *LDLRAD3* burden test CNVs. Chromosomal location of all CNVs considered for the *LDLRAD3* burden test, with respective allele count in the population used for association. Deletions are in red, duplications in blue. CNVs which extend beyond this locus are pruned at the edges of the 10*kb* padded window of *LDLRAD3* used for the burden test.

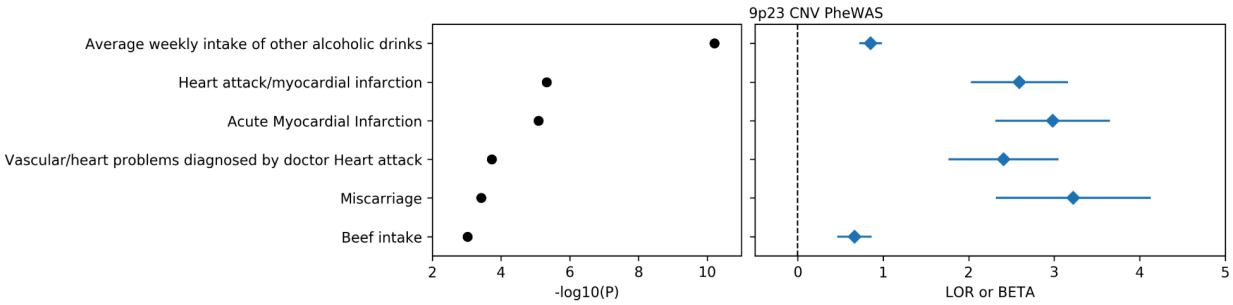


Figure S8: ***9p23*** **CNV PheWas.** Significant ($p < 10^{-3}$) associations between regularized burden tests for *9p23* CNV (top hit from Acute CAD GWAS) and other phenotypes. We highlight quantitative traits with $n > 15{,}000$ observations and binary traits with $n > 500$ cases. Traits are grouped by data type then sorted by *p*-value (left). Log-odds ratio and standardized betas (right; for binary and quantitative traits, respectively) align with trait names on the y-axis, with the vertical dashed line separating positive and negative direction of association.

Gene	Deletion z	Deletion pLI
BRCA2	2.870	0.9834
BRCA1	2.136	0.9578
APC	1.790	0.9463
ATM	1.063	0.9790
MSH2	1.048	0.9843
MLH1	1.033	0.9985
MYH7	0.902	0.8711
PMS2	0.858	0.9027
SBDS	0.800	0.9670
CYP3A4	0.799	0.9962
SPATA31D1	0.799	0.9962
TTN	0.798	0.9669
OTOP1	0.793	0.9962
MSH6	0.792	0.9924
FAM205A	0.790	0.9905

Gene	Duplication z	Duplication pLI
HLA-DRB1	0.566	0.9970
FRG2B	0.565	1.0000
SPATA31D1	0.565	0.9985
SLC35G6	0.565	1.0000
NAT8	0.565	1.0000
TUBB8	0.564	1.0000
CSH2	0.564	1.0000
ZNF302	0.564	1.0000
CSHL1	0.564	1.0000
GH1	0.564	1.0000
CGB2	0.564	1.0000
OR4F17	0.564	1.0000
CGB5	0.564	1.0000
CGB7	0.564	1.0000
GH2	0.564	1.0000

Table S1: Deletion- and whole-gene duplication-specific selective constraint. 15 genes most intolerant to overlapping deletion (left), and whole-gene duplication (right), with respective constraint z-scores.

GO ID	Deletion-intolerant Pathway Name	Delta z	P
GO:0045095	keratin filament	0.229	3.00E-27
GO:0000137	Golgi cis cisterna	0.351	4.14E-25
GO:0005515	protein binding	0.055	3.09E-19
GO:0008194	UDP-glycosyltransferase activity	0.314	4.03E-17
GO:0052697	xenobiotic glucuronidation	0.423	1.22E-16
GO:0000800	lateral element	0.380	3.54E-16
GO:0031424	keratinization	0.147	1.36E-15
GO:0042954	lipoprotein transporter activity	0.397	5.36E-15
GO:0015020	glucuronosyltransferase activity	0.293	1.47E-13
GO:0005131	growth hormone receptor binding	0.492	1.36E-12
GO:0008202	steroid metabolic process	0.231	2.55E-12
GO:0046703	natural killer cell lectin-like receptor binding	0.377	2.69E-12
GO:0008274	gamma-tubulin ring complex	0.308	6.07E-12
GO:0035459	cargo loading into vesicle	0.348	4.19E-11
GO:0070531	BRCA1-A complex	0.692	1.09E-10

GO ID	Duplication-intolerant Pathway Name	Delta z	P
GO:0000137	Golgi cis cisterna	0.333	3.59E-44
GO:0045095	keratin filament	0.190	3.19E-33
GO:0005515	protein binding	0.049	4.32E-31
GO:0031424	keratinization	0.134	1.49E-22
GO:0008202	steroid metabolic process	0.215	1.99E-21
GO:0005801	cis-Golgi network	0.167	2.46E-18
GO:0046703	natural killer cell lectin-like receptor binding	0.362	4.47E-18
GO:0008194	UDP-glycosyltransferase activity	0.235	4.16E-17
GO:0005132	type I interferon receptor binding	0.244	1.79E-15
GO:0052697	xenobiotic glucuronidation	0.290	8.34E-15
GO:0005131	growth hormone receptor binding	0.375	1.19E-14
GO:0042271	susceptibility to natural killer cell mediated cytotoxicity	0.237	2.17E-14
GO:0042954	lipoprotein transporter activity	0.283	5.69E-14
GO:0002323	natural killer cell activation involved in immune response	0.246	1.30E-13
GO:0008395	steroid hydroxylase activity	0.152	2.98E-13

Table S2: Deletion- and whole-gene duplication-specific pathway constraint. GO pathways most intolerant to overlapping deletion (top), and whole-gene duplication (bottom), with change in constraint z-scores and significance thereof (t-test) relative to other pathways.

HPO ID	Deletion-intolerant HPO Term	Delta z	P
HP:0006725	Pancreatic adenocarcinoma	0.469	1.22E-36
HP:0012432	Chronic fatigue	0.631	2.91E-32
HP:0025318	Ovarian carcinoma	0.576	9.38E-32
HP:0003003	Colon cancer	0.343	2.55E-30
HP:0004389	Intestinal pseudo-obstruction	0.576	2.83E-29
HP:0100273	Neoplasm of the colon	0.291	4.29E-27
HP:0100787	Prostate neoplasm	0.417	6.41E-26
HP:0012125	Prostate cancer	0.417	6.41E-26
HP:0030406	Primary peritoneal carcinoma	0.488	3.38E-24
HP:0100834	Neoplasm of the large intestine	0.241	1.76E-23
HP:0012334	Extrahepatic cholestasis	0.480	2.44E-23
HP:0003002	Breast carcinoma	0.267	2.69E-22
HP:0009592	Astrocytoma	0.449	9.80E-22
HP:0100707	Abnormality of the astrocytes	0.449	9.80E-22
HP:0002885	Medulloblastoma	0.444	3.32E-21

HPO ID	Duplication-intolerant HPO Term	Delta z	P
HP:0000707	Abnormality of the nervous system	0.039649	3.10E-17
HP:0012638	Abnormality of nervous system physiology	0.039371	1.93E-16
HP:0012759	Neurodevelopmental abnormality	0.03731	5.04E-15
HP:0000007	Autosomal recessive inheritance	0.039233	2.08E-14
HP:0012639	Abnormality of nervous system morphology	0.038852	5.67E-14
HP:0003011	Abnormality of the musculature	0.038621	9.58E-14
HP:0012373	Abnormal eye physiology	0.037309	9.81E-14
HP:0000478	Abnormality of the eye	0.039384	1.16E-13
HP:0100022	Abnormality of movement	0.036173	6.27E-13
HP:0012758	Neurodevelopmental delay	0.036116	8.81E-13
HP:0001249	Intellectual disability	0.035805	1.02E-12
HP:0012443	Abnormality of brain morphology	0.037981	1.32E-12
HP:0002011	Morphological abnormality of the central nervous system	0.039044	2.08E-12
HP:0011842	Abnormality of skeletal morphology	0.039995	2.24E-12
HP:0000924	Abnormality of the skeletal system	0.040534	4.56E-12

Table S3: Deletion- and whole-gene duplication-specific medical term constraint. HPO terms most intolerant to overlapping deletion (top), and whole-gene duplication (bottom), with change in constraint z-scores and significance thereof (t-test) relative to other pathways.