SUPPLEMENTARY MATERIAL

Supplementary Table S1: Datasets included in the Phenotypic Linkage Network (PLN).

Data type	Source
Biological Process	Gene Ontology
Cellular Location	Gene Ontology
Molecular Function	Gene Ontology
Protein-Protein Interactions	BioGRID, IntAct, Corum, DICS, Reactome (Mouse & Human)
Gene Coexpression	GNF2, GSE3594 (Gene Expression Omnibus), MTAB-62 (Gene Expression Atlas) , 5 SMD sets (Stanford Microarray Database)
Protein Domains	InterPro
Pathways, Reactions	KEGG & Reactome
Genetic Interactions	Yeast
Co-citation	STRING
Phenotype	Mouse Genome Informatics

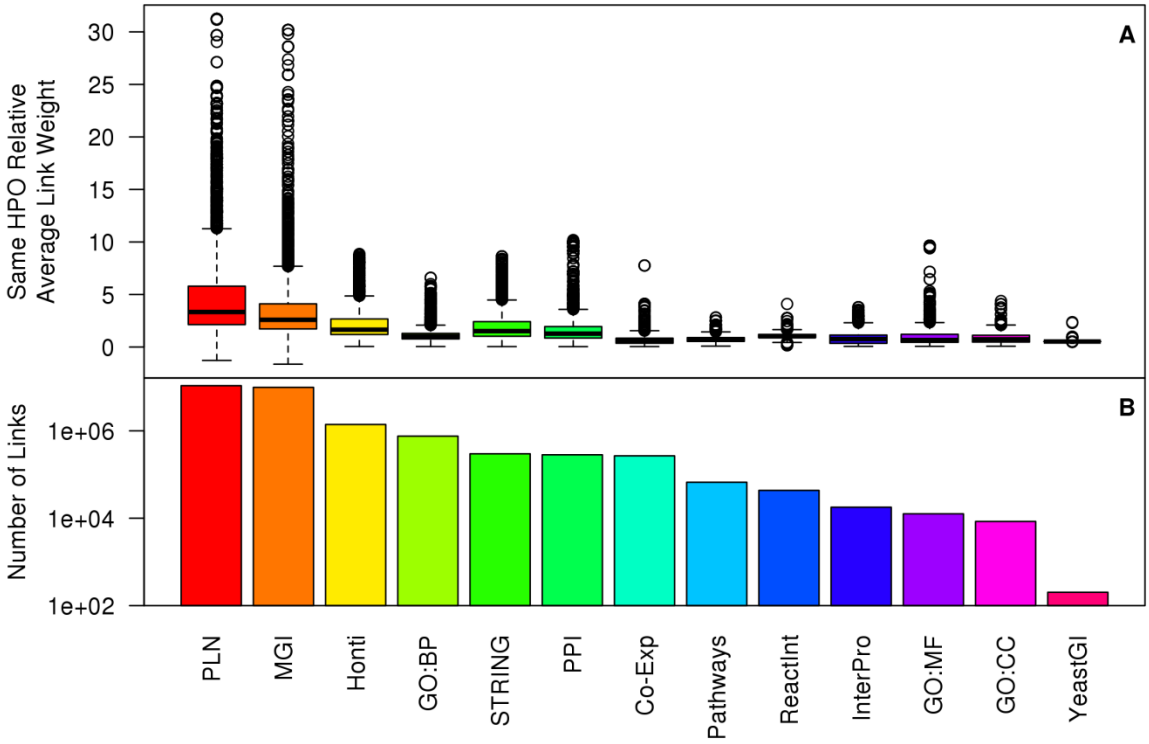
Supplementary Table S2: Functional similarity networks used to identify clusters of functionally-related genes.

Source	PLN	HumanNet	COXPRESdb
Genes	17,039	16,243	13,236
Direct edges	10,792,987	476,399	1,771,841
Filtering	None	None	Pearson correlation >= 0.5
Shortest paths	142,864,287 (98.4%)	129,146,568 (97.9%)	81,046,264 (92.5%)
1% threshold value	0.1	0.6	0.5
Types of data used	PPI, co-expression, pathways, co-citation, domains, functional annotations, mouse phenotypes	PPI, co-citation, co-expression from human, worm, fly, and yeast	>100 human expression experiments
Reference	(Honti et al. 2014)	(Lee et al. 2011)	(Obayashi et al. 2008)

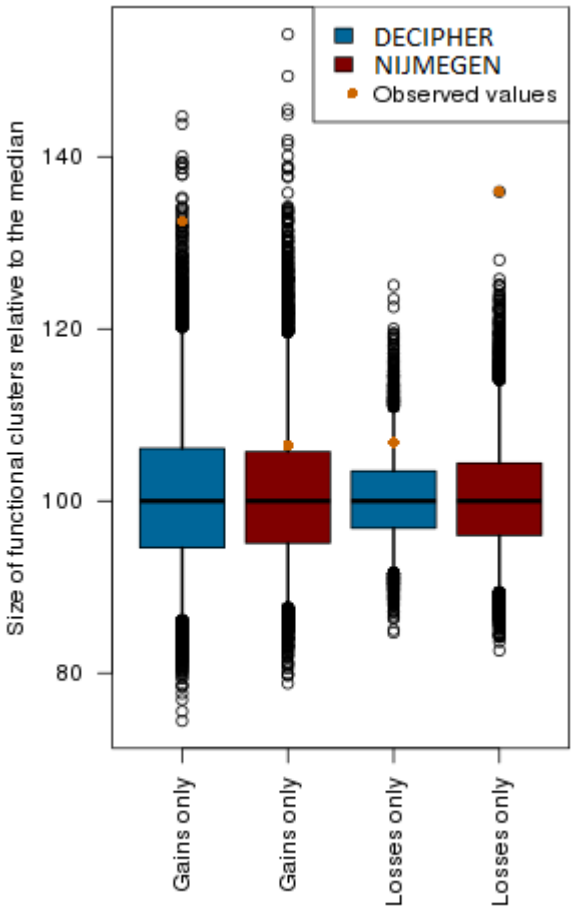
Supplementary Table S3: Logistic regression using a published case-control set of CNVs from patients with developmental disorders. Predictor are: the presence of functional clusters (Cluster), the presence of known haploinsufficient genes (HIS), the presence of known disease genes (OMIM), and the number of genes affected by the CNV (#genes). The lowest size threshold considered is >= 3 genes due to the fact that 15,000 control CNVs affected exactly 2 genes which was dominating the results at that threshold.

CNV size	predictor	OR	Pval	No. CNVs (% Patho)
>=3 genes	Cluster	1.4	0.00563	30755(56%)
	HIS	1.6	$<2x10^{-16}$	
	OMIM	1.3	$<2x10^{-16}$	
	#genes	1.1	$<2x10^{-16}$	
>=5 genes	Cluster	1.4	0.000861	16273(69%)
	HIS	1.6	$1.62x10^{-13}$	

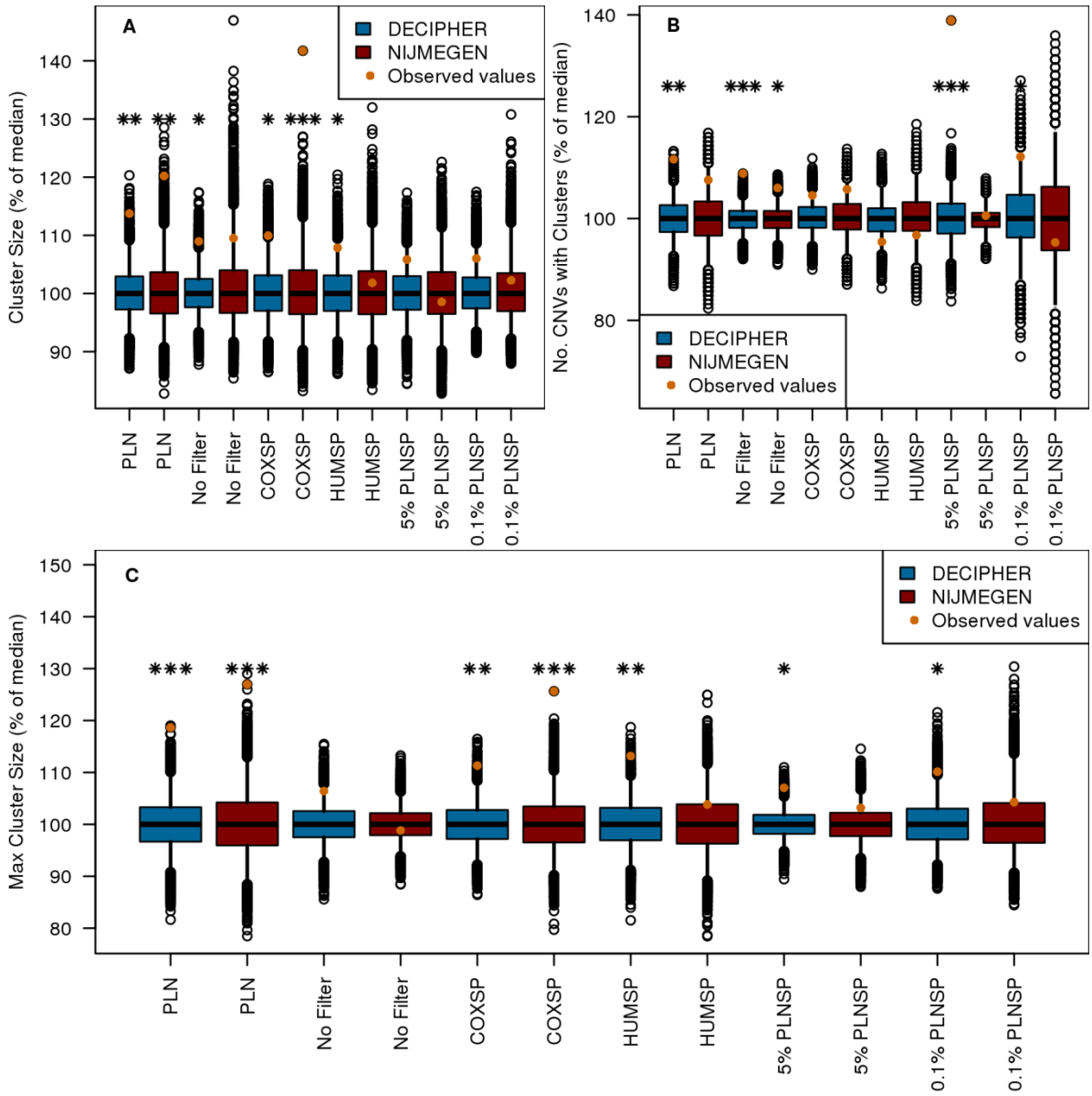
	OMIM	1.3	$1.41x10^{-9}$	
	#genes	1.1	$<2x10^{-16}$	
>=10 genes	Cluster	2.1	$7.51x10^{-9}$	8096(86%)
	HIS	1.6	$5.21x10^{-9}$	
	OMIM	0.9	0.504	
	#genes	1.1	$<2x10^{-16}$	
>=15 genes	Cluster	2.1	$2.39x10^{-6}$	5573(93%)
	HIS	1.5	0.00189	
	OMIM	1.2	0.102	
	#genes	1.0	$8.65x10^{-15}$	



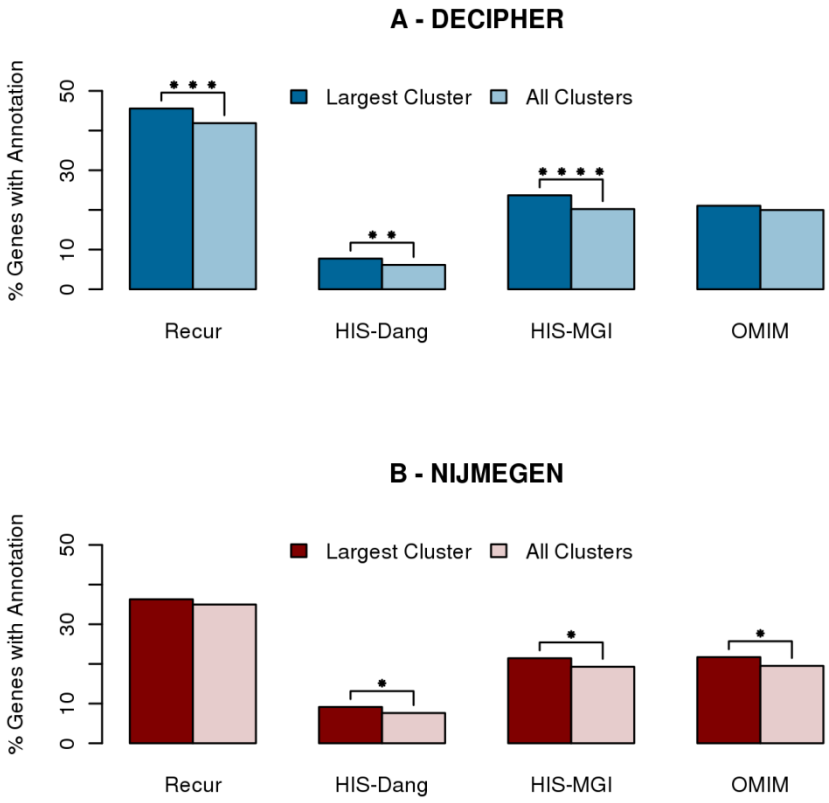
Supplementary Figure S1: Incorporating the mouse knockout phenotypes increases the specificity (A) and coverage (B) of the Phenotypic Linkage Network (PLN). Top: the average link weight between genes annotated with a given Human Phenotype Ontology (HPO) term for each network divided by the average link weight between all genes with any HPO term. All HPO terms with frequency < 0.5 were considered. Bottom: Total number of links in the whole network. PLN = the Phenotypic Linkage Network. MGI = semantic similarity between mouse phenotypes only network. Honti = the integrated functional network from (Honti et al. 2014) which combines all datasets except the mouse phenotypes. GO:BP = Gene Ontology Biological Process. PPI = combined protein-protein interaction network from all databases considered (**Supplementary Table S2**). Co-Exp = gene co-expression. Pathways = Reactome + KEGG. ReactInt = Reactome Interactions. GO:MF = Gene Ontology Molecular Function. GO:CC = Gene Ontology Cellular Component. YeastGI = Yeast genetic interactions.



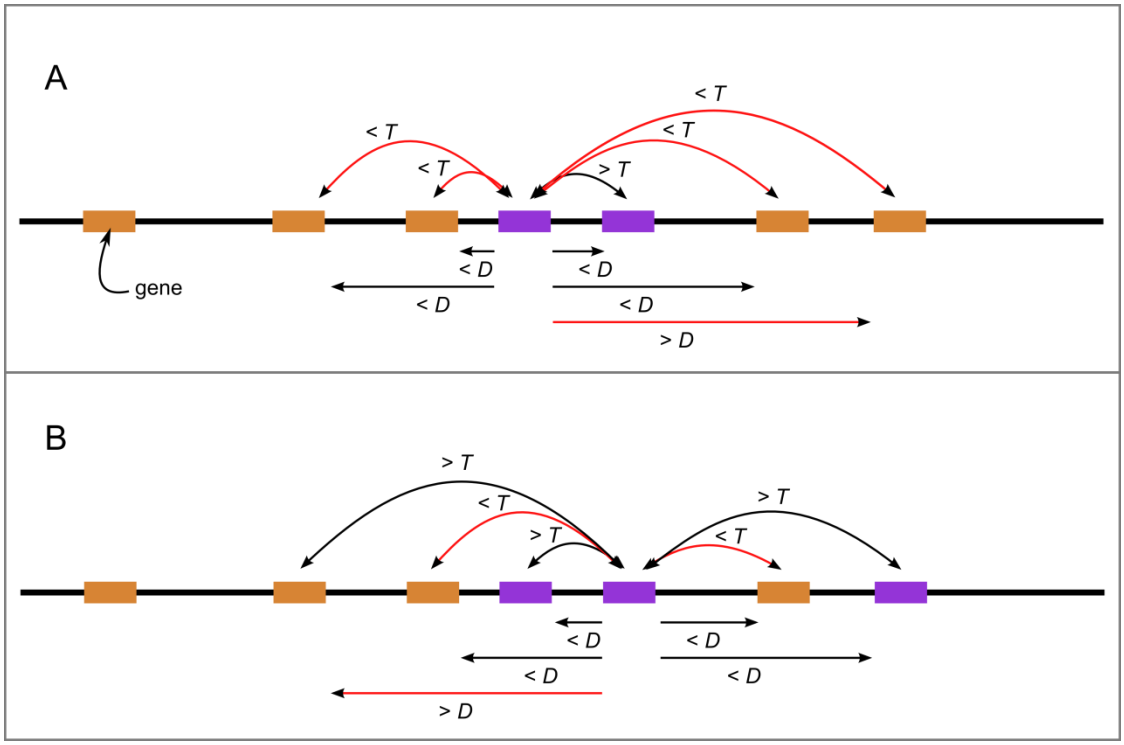
Supplementary Figure S2: Both deletions (Losses) and duplications (Gains) contain large functional clusters. Observed and randomized values were divided by the median of all randomizations to facilitate comparisons after collapsing paralogs. Whiskers contain 95% of randomizations. Orange dots indicate observed values. Stars indicate significant results, * = $p < 0.05$, ** = $p < 0.005$, *** = $p < 0.0005$.



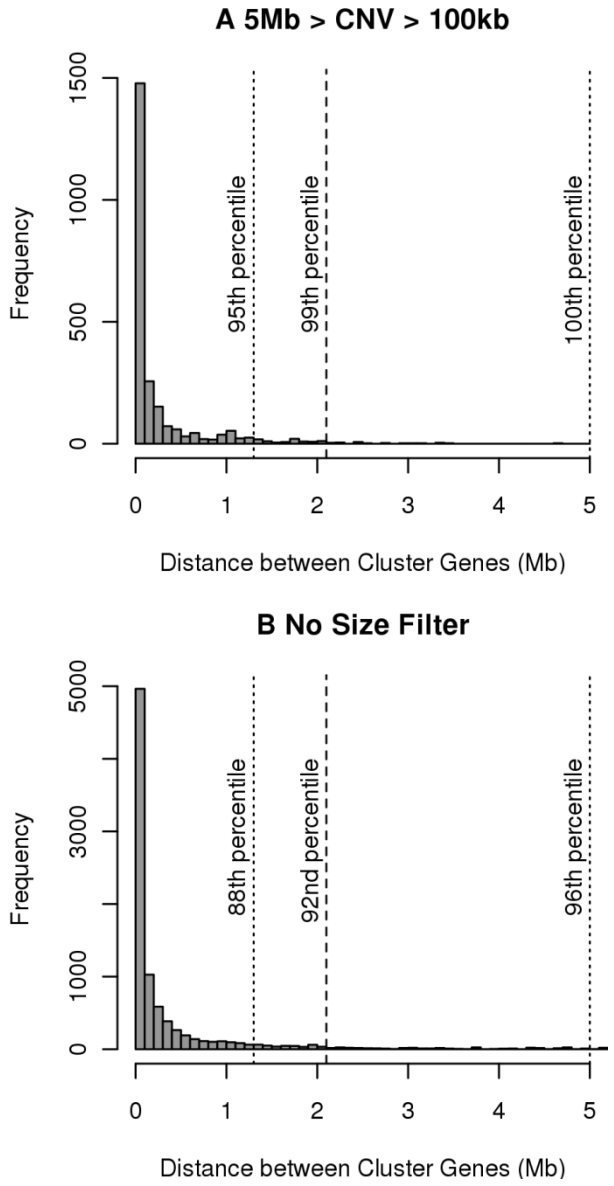
Supplementary Figure S3: Functional clusters are larger than expected across many different networks. Since average size functional clusters and number of CNVs with a functional cluster depend on the number of genes in the network and is sensitive to the threshold used, observed and randomized values were divided by the median of all randomizations to facilitate comparisons. Whiskers contain 95% of randomizations. PLN = direct edges only in the phenotypic linkage network, No Filter = all CNVs without imposing a size filter, COXSP = shortest paths in COXPRESdb (Obayashi et al. 2008), HUMSP = shortest paths in HumanNet (Lee et al. 2011), 5% PLNSP = top 5% shortest paths threshold using the phenotypic linkage network (at this level the majority of genes with any mouse phenotype information would be linked together, unlike the 1% and 0.1% thresholds). 0.1% PLNSP = top 0.1% shortest paths threshold using the phenotypic linkage network. Orange dots indicate observed values. Stars indicate significant results, * = $p < 0.05$, ** = $p < 0.005$, *** = $p < 0.0005$. **(A)** Average cluster size was consistently high in both datasets reaching significance in 4/6 variations for DECIPHER and 2/6 variations for NIMEGEN. **(B)** Number of CNVs containing functional clusters was less consistently high, with HumanNet showing the reverse non-significant trend in both datasets. **(C)** Average size of the largest cluster in each CNV was very consistently high, reaching significance for 5/6 variations for DECIPHER and tending in the same direction in 5/6 variations for NIJMEGEN.



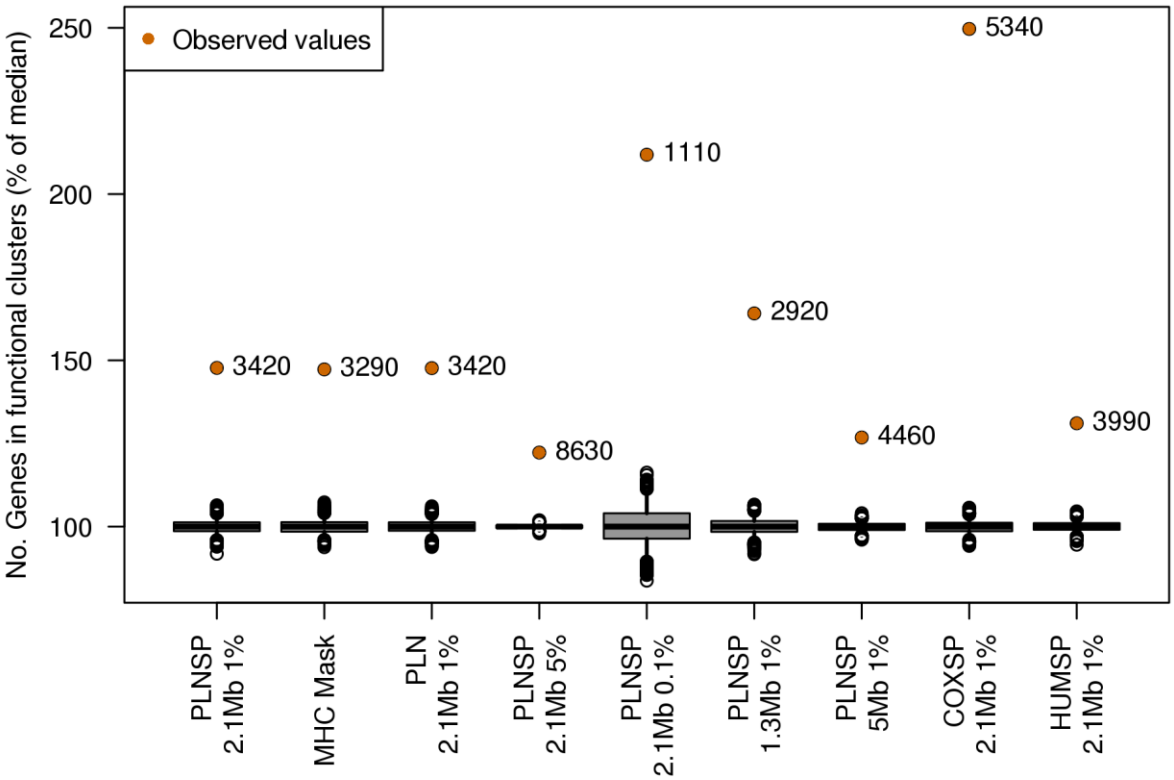
Supplementary Figure S4: The largest cluster per CNV is most enriched in disease genes.
Enrichment of disease genes in DECIPHER (A) and NIJMEGEN (B) largest functional cluster per CNV vs all functional clusters. Recur indicates genes found in more than one *de novo* CNV in the same dataset, Dang-HIS are haplo-insufficient genes identified in (Dang et al. 2008), OMIM are genes causally related to a disease in the Online Mendelian Inheritance in Man database (Online Mendelian Inheritance in Man, OMIM 2012). Significance was calculated using a one-sided hypergeometric test: * = $p < 0.05$, ** = $p < 0.005$, *** = $p < 0.0005$, etc.. up to a maximum of 5 stars.



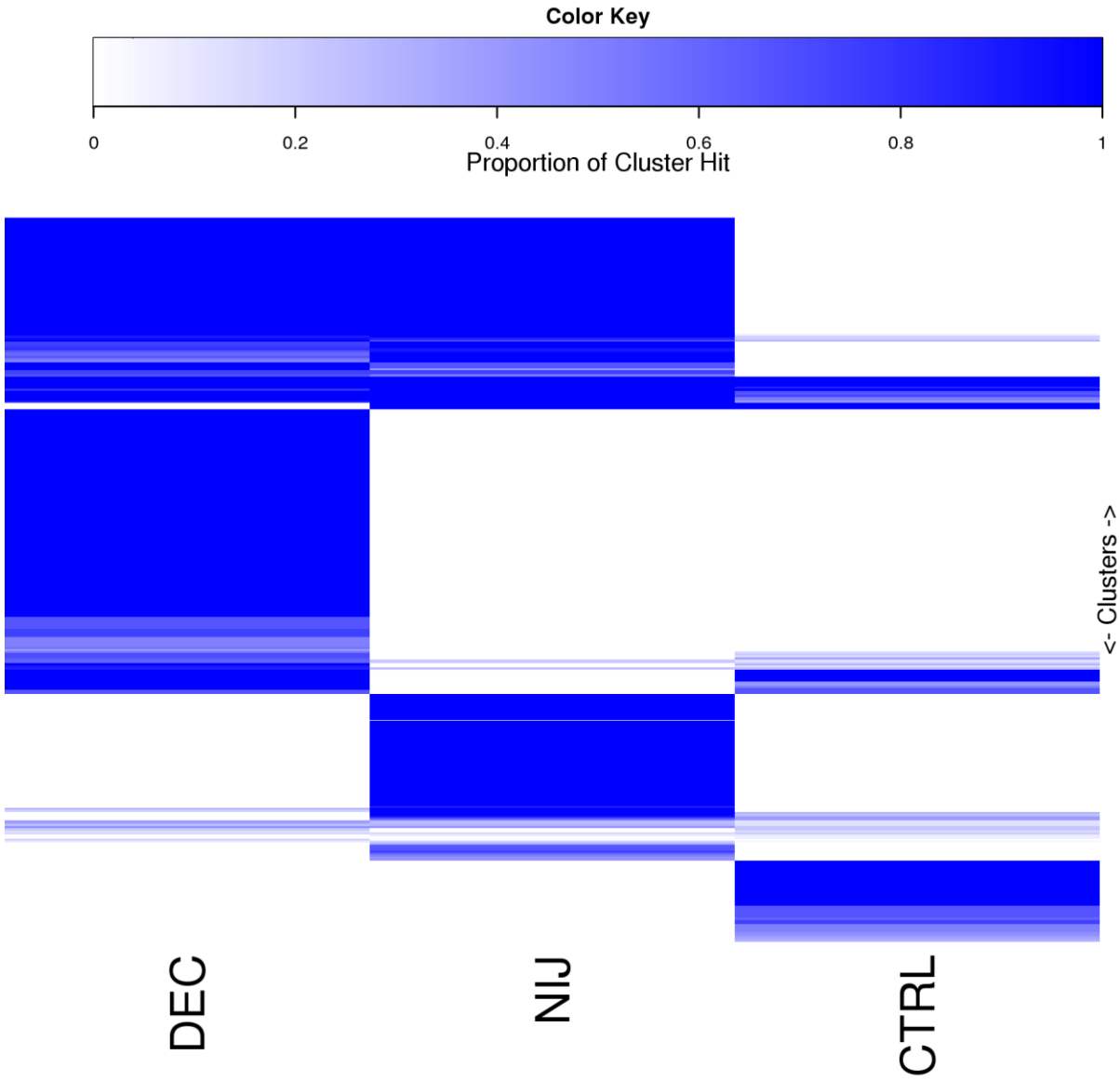
Supplementary Figure S5: The Genome-wide clustering of functionally-related genes algorithm. Genomic clusters of functionally-related genes were identified using a growing algorithm where a gene is added to the cluster if it is within distance threshold *D* and above similarity threshold *T* of another gene in the cluster. Purple indicates genes in the cluster, red arrows indicate relationships which do not pass the relevant threshold. B is the next iterative step after that shown in A.



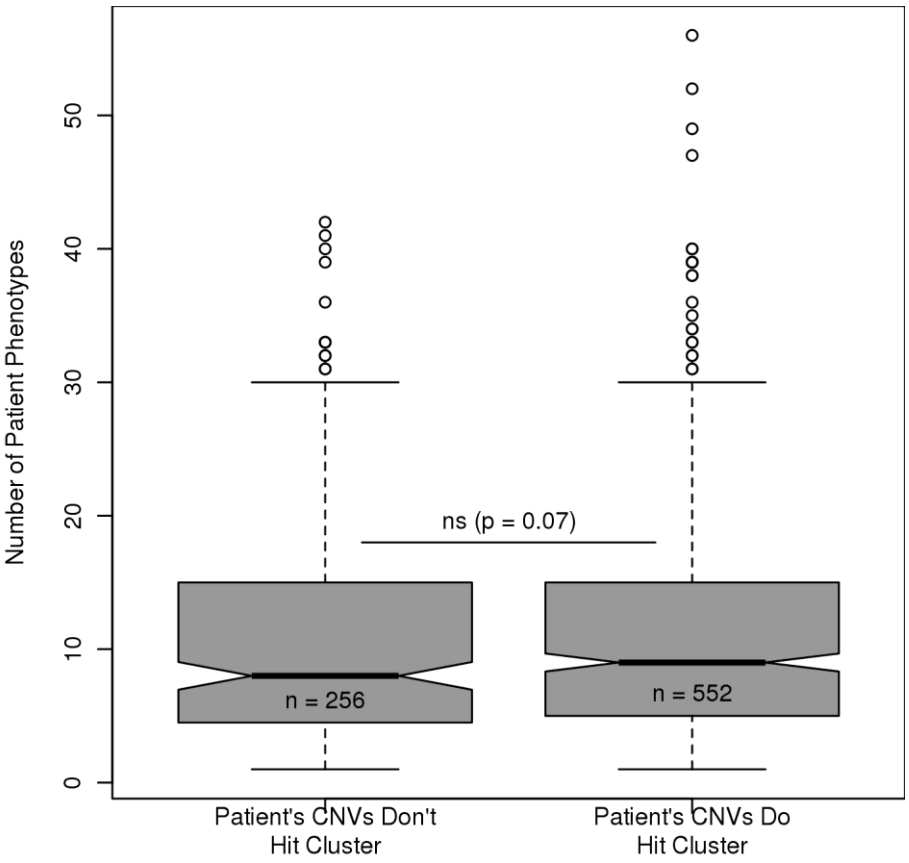
Supplementary Figure S6: Distribution of distances between the cluster genes and their nearest neighbouring cluster gene. The distance threshold used to identify genome-wide clusters of functionally-related genes was based on the distribution of distances between nearest-neighbours within the clusters of functionally-related genes we identified in the *de novo* CNVs. (A) Includes functional clusters identified in both DECIPHER and NIJMEGEN *de novo* CNVs between 100kb and 5Mb in size. 95th percentile corresponds to 1.3 Mb (dotted-line, used to check robustness), 99th percentile corresponds to 2.1 Mb (dashed-line, main threshold), and 5Mb was used for the 100th percentile (dotted-line, used to check robustness). (B) Includes functional clusters identified in all DECIPHER and NIJMEGEN *de novo* CNVs irrespective of length.



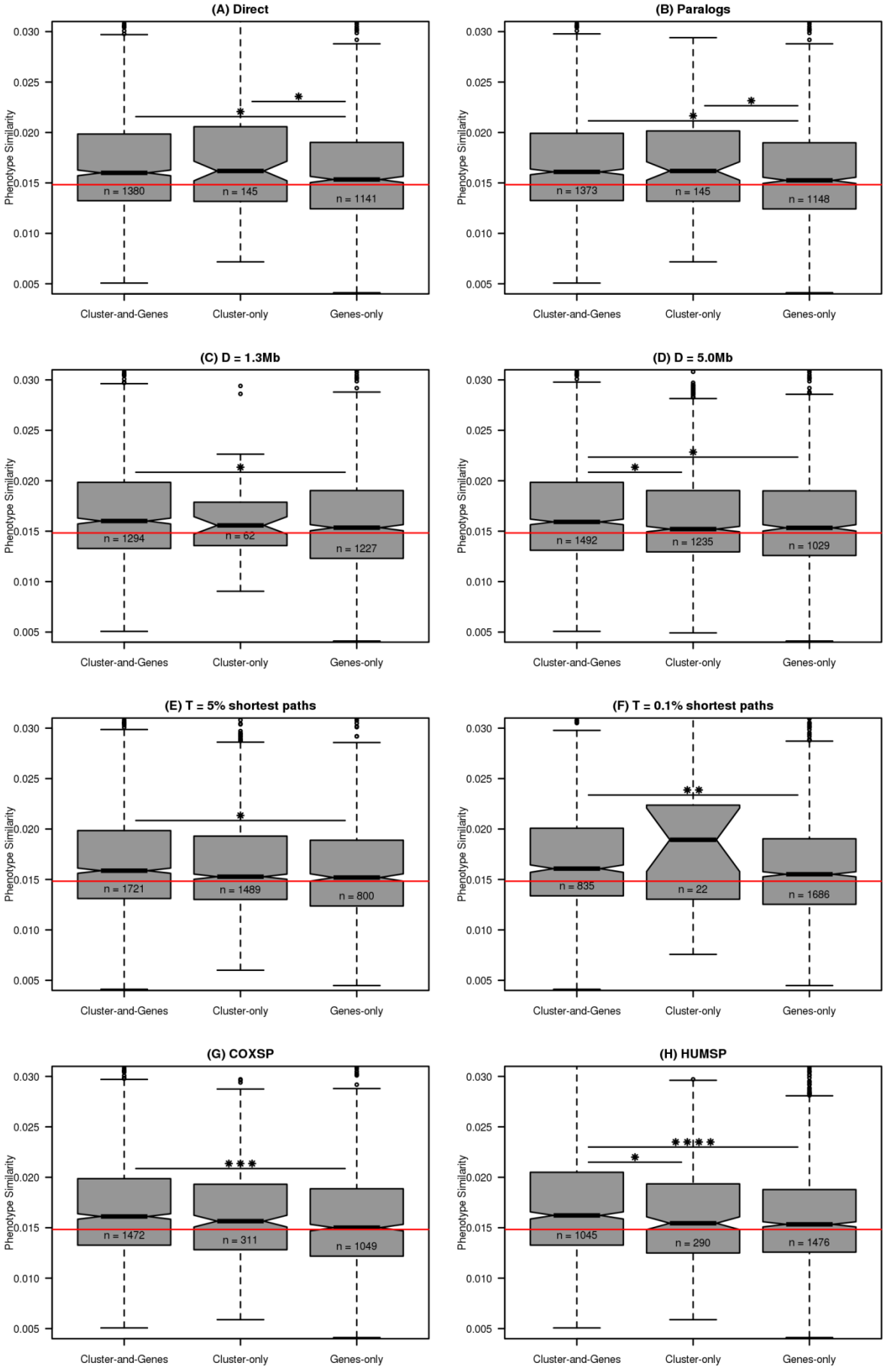
Supplementary Figure S7: The human genome is significantly functionally clustered regardless of the network or parameters used. The degree of clustering of functionally-related genes observed in the genome under various definitions of functional similarity (orange dots) compared to 1,000 node-label permutations (grey distribution). PLNSP: similarity calculated as the shortest-paths in the phenotypic linkage network, PLN: similarity calculated as the direct edges only in the phenotypic linkage network, COXSP: similarity calculated as the shortest paths in COXPRESdb (Obayashi et al. 2008), HUMSP: similarity calculated as the shortest paths in HumanNet (Lee et al. 2011), 2.1Mb [1.3 Mb, 5 Mb]: the distance threshold *D* = 2.1 Mb [1.3 Mb, 5 Mb]. 1% [5%, 0.1%]: the similarity threshold *T* = top 1% [5%, 0.1%] of shortest paths, MHC Mask = PLNSP 2.1Mb 1% but excluding the entire short arm of chromosome 6 to exclude the MHC region. Orange dots indicate observed values and are labelled with the total number of genes involved in functional clusters across the genome.



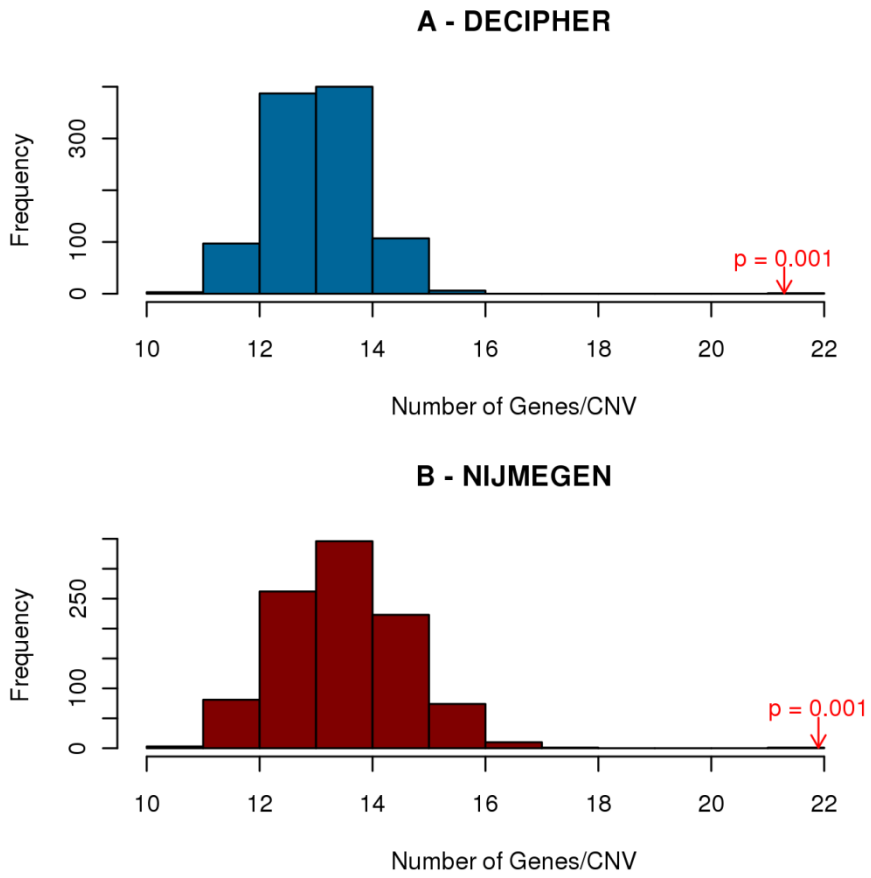
Supplementary Figure S8 : Proportion of genome-wide clusters affected by CNVs in DECIPHER (DEC), NIJMEGEN (NIJ), and the control (CTRL) datasets. Darkness of blue indicates the largest proportion of genes in the cluster hit by a single CNV in each dataset; each row represents a single genome-wide cluster. Only the largest cluster affected by each CNV was considered hit by that CNV. DECIPHER *de novo* CNVs hit 242 different functional clusters and 67% of the time the CNV contains the whole cluster. NIMEGEN *de novo* CNVs hit 175 different functional clusters and 72% of the time the CNV contains the whole cluster. In contrast, control CNVs only hit 95 different clusters and only 29% of the time does the CNV affect the whole cluster. The 586 of 942 genome-wide clusters not hit by any CNV are not included in the figure.



Supplementary Figure S9: *de novo* CNVs which affect >= 2 genes in a cluster of functionally-related genes do not come from patients with more phenotypes. *de novo* CNVs affecting < 2 genes were excluded. Phenotypes unique to LND or HPO were excluded. CNVs not filtered by size.


(A) Direct
(B) Paralogs
(C) D = 1.3Mb
(D) D = 5.0Mb
(E) T = 5% shortest paths
(F) T = 0.1% shortest paths
(G) COXSP
(H) HUMSP
Phenotype Similarity
0.030
0.025
0.020
0.015
0.010
0.005
n = 1380
n = 145
n = 1141
n = 1373
n = 145
n = 1148
n = 1294
n = 62
n = 1227
n = 1492
n = 1235
n = 1029
n = 1721
n = 1489
n = 800
n = 835
n = 22
n = 1686
n = 1472
n = 311
n = 1049
n = 1045
n = 290
n = 1476
Cluster-and-Genes
Cluster-only
Genes-only

Supplementary Figure S10: Clusters of functionally-related genes were better indicators of phenotypic similarity than genes regardless of the parameters used, and using a different functional network. Phenotype similarity as measured by the *Goodall3* index (Boriah et al. 2008) between pairs of patients in each category shown in (**Figure 5 A**): Cluster-and-Genes affect the same functional cluster and the same genes, Cluster-and-OMIM affect the same functional cluster and the same OMIM genes, Cluster-only affect the same functional cluster but different genes, Genes-only affect the same genes but not the same functional cluster, OMIM-only affect the same OMIM genes but not the same functional cluster. Stars indicate significance, calculated using a Wilcox-rank sum test. One star = $p < 0.05$. Two stars = $p < 0.005$. Three stars = $p < 0.0005$ etc… upto a maximum of five stars. Red line indicates the median phenotypic similarity over all patient pairs.



Supplementary Figure S11: *de novo* CNVs contained more genes than expected. DECIPHER (A) & NIJMEGEN (B) *de novo* CNVs between 100kb and 5Mb in size compared to 1,000 random genomic segments of equal length. Genes were mapped such that the CNV must affect at least one exon in every transcript of the gene. The red arrows indicate observed average number of genes per CNV and the associated p-value. Since both CNV sets were enriched in genes we compared the functional clusters within the CNVs to those in random segments containing an equal number of genes to the CNVs.