Supplementary Material

Supplementary Methods

Variant annotation and filtering

After read alignment to the human reference genome (build GRCh37) and variant calling by Illumina, single nucleotide variants (SNVs) and insertion/deletions (indels) were annotated using Ensembl's Variant Effect Predictor (VEP). They were then filtered for population frequency and for consequence (all coding variants plus non-coding variants previously reported as potentially pathogenic in the variant databases ClinVar and the Human Gene Mutation Database) [1-3]. Inheritance filtering was performed for *de novo*, homozygous, compound heterozygous and X-linked variants. Mitochondrial genome variants were analysed using the MToolBox annotation and filtering software [4]. Structural and copy number variants were filtered for overlapping a protein-coding gene and population frequency. Details of filtering parameters are below:

SNVs and indels

Global filter for all variants:

- gnomAD [1] allele frequency ≤ 0.01
- ExAC [1] allele frequency ≤ 0.01
- WGS10k* allele frequency ≤ 0.01
- Internal cohort allele frequency ≤ 0.05
- Moderate or high impact on protein function [2] or found in ClinVar [3] or in HGMD (Human Gene Mutation Database)

Frequency filtering for each inheritance pattern:

- Autosomal dominant *de novo*
  - Allele count ≤ 10 in gnomAD, ExAC, WGS10k, and internal cohort
- Autosomal recessive homozygous (including X for females)
- Autosomal recessive compound heterozygous (including X for females and *de novo* + inherited)
  - Allele frequency ≤ 0.01 in gnomAD, ExAC, and WGS10k
  - Allele frequency ≤ 0.02 internal cohort
  - Homozygous count ≤ 10 in gnomAD and ExAC
- X-linked (male only)
  - Hemizygous count ≤ 10 in gnomAD, ExAC

Mitochondrial genome variants

Frequency filter:

- WGS10k allele frequency < 0.01 and internal cohort allele frequency < 0.05

Heteroplasmy filter:

- homoplasmic OR
- >80% heteroplasmic OR
- >5% heteroplasmic AND read count >1000

Consequence filter:

- Non-synonymous change OR
- MitoMap associated disease OR
- In Clinvar

*WGS10k: ~13,000 internal controls from the NIHR-BioResource Rare Diseases research study

SVs and CNVs

Global filter for all variants:

Pass standard Illumina quality filters
Healthy control cohort [5] overlap‡ = 0

Frequency filtering for each inheritance pattern:

Dominant filter:

WGS10k** overlap frequency < 0.001
Internal cohort overlap count < 10
Overlap protein-coding gene
Heterozygous

Recessive filter:

WGS10k overlap frequency < 0.01
Internal cohort overlap frequency < 0.02
Overlap protein-coding gene
Homozygous OR hemizygous OR compound heterozygous with SV, CNV, or SNV/indel

‡ Overlap is defined by 0.5 reciprocal overlap
**WGS10k: ~9,500 internal controls from the NIHR-BioResource Rare Diseases research study

Phenotype comparison analysis

Gene-associated sets of HPO terms were taken from the following file in the 2018-10-09 HPO release [6]: ALL_SOURCES_FREQUENT_FEATURES_diseases_to_genes_to_phenotypes.txt. Both proband- and gene-associated sets of HPO terms were reduced to non-redundancy with the 'minimal_set' function in the ontologyX R package [7]. All sets of terms were filtered for only those belonging the HPO class "Phenotypic abnormality" and the following terms and their descendants were filtered out: "Abnormal delivery", "Premature birth", "Prenatal maternal abnormality".

For the clustering analysis only, the term "Neonatal respiratory distress" and its descendent were also removed from the proband-associated term sets because it is very common in the NICU patients (64%). Pairwise HPO term similarity scores between the probands were calculated with the 'get_sim_grid' function (ontologyX), transformed into distances with the 'dist' function (method=maximim) and then the probands were hierarchically clustered with the 'hclust' function (fastcluster [8] R package) and split into groups with the 'cutree' function (h=0.75).

Enrichment of HPO terms in both the clustered proband groups and the diagnosed set of probands (compared to the entire cohort) was calculated with Fisher's Exact test 'fisher.test' R function (alternative=greater). HPO term counts for each group also included the counts for the ancestors of all the terms in order to account for variable specificity of the reported terms. A multiple hypothesis testing correction was applied to each group to correct for the large number of HPO terms tested with the 'p.adjust' function (method=BH) and significant enrichment assessed at 0.1 false discovery rate. The HPO terms that were enriched in the set of diagnosed probands ($p<0.05$ but no multiple hypothesis correction), and their ancestors, were plotted with the 'onto_plot' function (ontologyX).

Pairwise HPO term similarity scores between each proband and each gene for which there were HPO terms were also calculated with the 'get_sim_grid' function (ontologyX). For each diagnosed proband, the similarity scores of all the genes were sorted and the rank of the diagnosed gene was determined.

References

1. Lek M, Karczewski KJ, Minikel E V, et al (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291. https://doi.org/10.1038/nature19057
2. Calculated consequences. https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html. Accessed 16 Aug 2018

3. Landrum MJ, Lee JM, Benson M, et al (2018) ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 46:D1062–D1067. https://doi.org/10.1093/nar/gkx1153
4. Calabrese C, Simone D, Diroma MA, et al (2014) MToolBox: A highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. Bioinformatics 30:3115–3117. https://doi.org/10.1093/bioinformatics/btu483
5. Zarrei M, MacDonald JR, Merico D, Scherer SW (2015) A copy number variation map of the human genome. Nat Rev Genet 16:172–183. https://doi.org/10.1038/nrg3871
6. Köhler S, Vasilevsky NA, Engelstad M, et al (2017) The Human Phenotype Ontology in 2017. Nucleic Acids Res 45:D865–D876. https://doi.org/10.1093/nar/gkw1039
7. Greene D, Richardson S, Turro E (2017) OntologyX: A suite of R packages for working with ontological data. Bioinformatics 33:1104–1106. https://doi.org/10.1093/bioinformatics/btw763
8. Müllner D (2013) **fastcluster** : Fast Hierarchical, Agglomerative Clustering Routines for *R* and *Python*. J Stat Softw 53:1–18. https://doi.org/10.18637/jss.v053.i09

SUPPLEMENTARY TABLES

Supplementary Table 1. Recruitment criteria

Criteria	Inclusion		Exclusion	
	NICU	PICU	NICU	PICU
Congenital anomalies	X	X		
Neurological symptoms	X	X		
Suspected metabolic disease	X	X		
Surgical necrotizing enterocolitis (NEC)	X			
Extreme IUGR or failure to thrive	X			
Unexplained critical illness/clinician's request	X	X		
Prematurity without additional features			X	X
Known trisomies or other genetic diagnosis			X	X
Trauma				X
Haematological malignancies and Oncology				X
Bronchiolitis/respiratory tract infections				X

NICU: neonatal intensive care unit
PICU: paediatric intensive care unit
IUGR: intra-uterine growth retardation

Supplementary Table 2. Patient data

Sample ID	Family size	Sex	Recruitment age (postnatal days)	Age at death (postnatal days)	Recruitment ward
NGC00001_01	3	M	38	118	NICU
NGC00002_01	3	M	98	160	NICU
NGC00003_01	3	M	9	NA	NICU
NGC00004_01	3	F	11	NA	NICU
NGC00005_01	3	F	1	1	NICU
NGC00006_01	3	M	58	NA	NICU
NGC00007_01	1	M	29	NA	NICU
NGC00008_01	3	F	33	NA	NICU
NGC00009_01	3	F	6	NA	NICU
NGC00010_01	2	M	27	NA	NICU
NGC00011_01	3	F	30	NA	NICU
NGC00012_01	3	F	4	NA	NICU
NGC00013_01	3	M	4	NA	NICU
NGC00014_01	3	F	8	NA	NICU
NGC00015_01	3	F	23	NA	NICU
NGC00016_01	3	F	4	NA	NICU
NGC00017_01	3	F	7	NA	NICU
NGC00018_01	3	M	6	NA	NICU
NGC00019_01	3	F	12	NA	NICU
NGC00020_01	3	F	25	NA	NICU
NGC00021_01	3	F	7	NA	NICU
NGC00022_01	3	M	205	NA	PICU
NGC00023_01	4	F	5211	NA	PICU
NGC00023_04	4	F	4144	NA	NA
NGC00024_01	3	M	184	190	NICU
NGC00025_01	3	F	4	NA	NICU
NGC00026_01	3	M	13	NA	NICU
NGC00027_01	3	F	40	NA	NICU
NGC00028_01	3	F	2	NA	Postnatal Ward
NGC00029_01	3	F	4038	NA	PICU
NGC00030_01	3	M	1273	NA	PICU
NGC00031_01	3	M	5	43	NICU
NGC00032_01	2	F	246	NA	PICU
NGC00033_01	3	F	4767	NA	PICU
NGC00034_01	3	F	221	NA	PICU
NGC00035_01	3	M	15	NA	NICU

NGC00036_01	3	F	85	85	Genetics
NGC00037_01	2	M	5963	NA	PICU
NGC00038_01	3	M	6	NA	NICU
NGC00039_01	3	F	68	NA	NICU
NGC00041_01	3	M	9	NA	NICU
NGC00042_01	3	F	3	NA	NICU
NGC00043_01	2	M	1341	NA	PICU
NGC00044_01	2	F	3629	NA	PICU
NGC00045_01	1	M	5534	NA	PICU
NGC00047_01	3	M	319	NA	PICU
NGC00048_01	2	M	5574	NA	PICU
NGC00049_01	3	M	8	NA	NICU
NGC00050_01	3	F	4	NA	NICU
NGC00051_01	3	F	42	NA	NICU
NGC00052_01	3	M	5	NA	NICU
NGC00053_01	3	F	6	NA	NICU
NGC00054_01	3	F	300	NA	PICU
NGC00055_01	3	M	4	NA	NICU
NGC00056_01	3	F	2030	NA	PICU
NGC00057_01	3	F	4228	NA	PICU
NGC00058_01	2	M	21	NA	NICU
NGC00059_01	3	M	720	NA	PICU
NGC00060_01	1	F	5237	NA	PICU
NGC00061_01	4	M	190	NA	PICU
NGC00061_04	4	M	190	NA	NA
NGC00062_01	3	F	78	NA	NICU
NGC00063_01	3	F	3	NA	NICU
NGC00064_01	2	F	4693	NA	PICU
NGC00065_01	3	M	313	NA	PICU
NGC00066_01	3	M	5	NA	NICU
NGC00067_01	2	M	4164	NA	PICU
NGC00068_01	3	M	17	NA	NICU
NGC00069_01	3	F	65	NA	PICU
NGC00070_01	3	M	501	NA	PICU
NGC00071_01	3	F	1478	NA	PICU
NGC00073_01	3	M	8	NA	NICU
NGC00075_01	3	F	154	NA	PICU
NGC00076_01	3	F	22	53	NICU
NGC00077_01	3	M	79	NA	NICU

NGC00078_01	3	F	14	NA	NICU
NGC00079_01	3	F	4	NA	NICU
NGC00080_01	3	F	4	4	NICU
NGC00081_01	3	M	141	NA	NICU
NGC00082_01	3	F	9	NA	NICU
NGC00083_01	3	F	4	8	NICU
NGC00084_01	2	F	18	NA	NICU
NGC00085_01	3	F	9	NA	NICU
NGC00086_01	3	F	105	NA	PICU
NGC00087_01	3	F	1793	NA	PICU
NGC00088_01	3	M	6135	NA	PICU
NGC00089_01	2	M	6	NA	NICU
NGC00090_01	3	F	167	NA	PICU
NGC00091_01	3	M	13	NA	NICU
NGC00092_01	3	M	25	NA	NICU
NGC00093_01	2	M	13	NA	NICU
NGC00094_01	3	M	1974	NA	PICU
NGC00095_01	3	F	2309	NA	PICU
NGC00096_01	3	M	386	NA	PICU
NGC00097_01	2	F	8	NA	NICU
NGC00098_01	3	M	4870	NA	PICU
NGC00099_01	3	F	5550	5595	PICU
NGC00101_01	4	F	1	1	NICU
NGC00101_04	4	M	0	NA	NA
NGC00102_01	3	M	42	NA	NICU
NGC00103_01	3	F	744	NA	PICU
NGC00104_01	3	M	745	NA	PICU
NGC00105_01	3	M	2047	NA	PICU
NGC00106_01	2	F	7	NA	NICU
NGC00107_01	3	M	1716	NA	PICU
NGC00108_01	3	F	86	NA	NICU
NGC00109_01	3	F	275	275	Genetics
NGC00110_01	3	M	116	NA	NICU
NGC00111_01	3	F	3	NA	NICU
NGC00112_01	3	F	7	NA	NICU
NGC00113_01	4	M	51	NA	NICU
NGC00113_04	4	F	1953	NA	NA
NGC00114_01	3	F	89	89	PICU
NGC00115_01	3	M	23	NA	NICU

NGC00116_01	3	M	359	NA	PICU
NGC00117_01	5	F	11028	NA	NA
NGC00117_04	5	F	4	NA	NICU
NGC00118_01	3	M	14	NA	NICU
NGC00119_01	3	F	8	NA	NICU
NGC00120_01	3	F	0	0	Delivery Unit
NGC00121_01	3	F	6	NA	NICU
NGC00122_01	3	F	627	NA	PICU
NGC00123_01	3	M	414	NA	PICU
NGC00124_01	3	F	8	NA	NICU
NGC00125_01	3	M	191	NA	PICU
NGC00126_01	3	F	166	NA	PICU
NGC00127_01	3	M	587	NA	Neuro
NGC00128_01	3	M	220	NA	PICU
NGC00129_01	3	F	81	NA	Neuro
NGC00130_01	3	M	474	NA	Neuro
NGC00131_01	3	F	131	NA	Neuro
NGC00132_01	3	M	28	NA	NICU
NGC00133_01	3	M	1227	NA	Neuro
NGC00134_01	3	M	49	NA	NICU
NGC00135_01	3	F	332	NA	Neuro
NGC00136_01	3	M	17	NA	NICU
NGC00137_01	3	F	24	NA	NICU
NGC00138_01	3	M	11	NA	NICU
NGC00139_01	3	F	962	NA	Neuro
NGC00140_01	3	M	41	NA	PICU
NGC00141_01	3	M	576	NA	PICU
NGC00142_01	3	M	14	NA	NICU
NGC00143_01	3	M	2197	NA	PICU
NGC00144_01	3	M	4	NA	NICU
NGC00145_01	3	F	4042	NA	PICU
NGC00146_01	3	M	19	NA	NICU
NGC00147_01	3	F	2225	NA	Neuro
NGC00148_01	3	F	12	NA	NICU
NGC00149_01	3	M	761	NA	Neuro
NGC00150_01	3	M	941	NA	Neuro
NGC00151_01	3	M	4	NA	NICU
NGC00152_01	3	F	8465	NA	Neuro
NGC00153_01	3	F	13	NA	NICU

NGC00154_01	3	F	19	NA	NICU
NGC00155_01	3	M	106	NA	PICU
NGC00156_01	3	M	72	NA	Genetics
NGC00157_01	3	M	14	NA	NICU
NGC00158_01	2	M	31	NA	NICU
NGC00159_01	3	M	8	NA	NICU
NGC00160_01	2	M	2516	NA	PICU
NGC00161_01	3	F	191	NA	Neuro
NGC00162_01	3	M	53	NA	NICU
NGC00164_01	3	F	11	NA	NICU
NGC00165_01	2	M	280	NA	Genetics
NGC00166_01	3	M	327	NA	Neuro
NGC00167_01	3	M	5552	NA	Neuro
NGC00169_01	3	M	6	NA	NICU
NGC00170_01	3	F	7	NA	NICU
NGC00171_01	3	F	39	NA	NICU
NGC00172_01	3	F	15	NA	NICU
NGC00173_01	3	M	5707	NA	PICU
NGC00174_01	3	F	127	NA	Neuro
NGC00175_01	3	F	5	NA	NICU
NGC00176_01	3	F	3203	NA	Neuro
NGC00177_01	3	M	529	NA	PICU
NGC00178_01	3	M	112	NA	Neuro
NGC00179_01	3	F	18	NA	PICU
NGC00180_01	3	M	58	NA	NICU
NGC00181_01	3	M	25	NA	NICU
NGC00182_01	3	M	56	NA	NICU
NGC00183_01	3	F	899	NA	PICU
NGC00184_01	3	M	40	NA	PICU
NGC00185_01	3	F	6	NA	NICU
NGC00186_01	3	M	8	NA	PICU
NGC00187_01	3	F	10	NA	NICU
NGC00188_01	3	F	6115	NA	Neuro
NGC00189_01	3	M	1608	NA	Neuro
NGC00190_01	3	M	7	NA	NICU
NGC00191_01	3	M	1963	NA	Neuro
NGC00192_01	3	M	564	NA	PICU
NGC00193_01	3	F	137	NA	Neuro
NGC00194_01	3	M	66	NA	NICU

NGC00196_01	3	F	2359	NA	Neuro
NGC00197_01	3	M	18	NA	NICU
NGC00198_01	3	F	302	NA	Neuro
NGC00199_01	2	M	15	NA	NICU
NGC00200_01	3	F	1613	NA	PICU
NGC00205_01	3	F	81	NA	NICU
NGC00206_01	3	M	921	NA	PICU
NGC00207_01	3	M	5647	NA	Genetics

NICU: neonatal intensive care unit
PICU: paediatric intensive care unit
Neuro: paediatric neurology department
Genetics: clinical genetics department

Supplementary Table 3. Diagnostic rates by age and ward

Recruited from	Age[a] range at recruitment	Number of probands	Number of diagnoses	Percent diagnosed
NICU	0 - 14 days	62	10	16%
	14 days – 6 months	44	4	9%
PICU	8 days – 2 years	29	7	24%
	2 years – 16 years	32	8	25%
Neuro/Genetics	80 days - 2 years	15	7	47%
	2 years – 23 years	13	4	31%

NICU: neonatal intensive care unit
PICU: paediatric intensive care unit
Neuro: paediatric neurology department
Genetics: clinical genetics department
[a]Postnatal age

Supplementary Table 4. Diagnosed cases

Family ID	Age[a]	Sex	Gene	Diagnosis [OMIM]	Variant
31	5 d	M	*MTM1*	Myotubular myopathy	X-linked recessive, missense
33	13 y	F	*BRAF*	Noonan syndrome	*de novo*, missense
44	10 y	F	*COL2A1*	Spondyloepiphyseal dysplasia congenita[b]	*de novo*, missense
58	21 d	M	*CYP21A2*	Congenital adrenal hyperplasia	Homozygous, splice variant
62	11 w	F	*SATB2*	Glass syndrome	*de novo*, missense
67	11 y	M	*SMC1A*	Cornelia de Lange syndrome	X-linked dominant, missense
69	2 m	F	*ARID1B*	Coffin-Siris syndrome	*de novo*, deletion (16 Mb)
75	5 m	F	*ARID1B*	Coffin-Siris syndrome	*de novo*, frameshift
76	22 d	F	*NIPBL*	Cornelia de Lange syndrome	*de novo*, frameshift
78	14 d	F	*COL2A1*	Kniest dysplasia	*de novo*, splice site
94	5 y	M	*HBB*	Beta thalassemia[b]	Homozygous, frameshift
96	13 m	M	*ASXL3*	Bainbridge-Ropers syndrome	*de novo*, frameshift
99	15 y	F	*SCN2A*	Early infantile epileptic encephalopathy	*de novo*, missense
101	1 d	F	*TTN*	Congenital titinopathy[c]	Comp. het., frameshifts
104	2 y	M	*GK*	Glycerol kinase deficiency	X-linked recessive, missense
109	9 m	F	*TBCD*	Early onset, progressive encephalopathy	Comp. het., splice site + missense
111	3 d	F	5q del	5q15-23 deletion[d]	*de novo*, deletion (24 Mb)
114	3 m	F	*NDUFA6*	Mitochondrial complex I deficiency[e]	Comp. het., frameshifts
125	6 m	M	*TGFBR1*	Loeys-Dietz syndrome	*de novo*, missense
126	5 m	F	*RHOBTB2*	Early infantile epileptic encephalopathy	*de novo*, missense
127	20 m	M	*GFAP*	Alexander disease	*de novo*, missense
130	16 m	M	*PDHA1*	Pyruvate dehydrogenase deficiency	*de novo* X-linked dom., missense
135	11 m	F	*SCN1A*	Dravet syndrome	*de novo*, missense
138	11 d	M	*DNAH11*	Ciliary dyskinesia, primary	Comp. het., stop gain + missense
139	3 y	F	*ALS2*	Juvenile amyotrophic lateral sclerosis	Comp. het., splice site + frameshift
143	6 y	M	*PYGM*	McArdle disease	Comp. het., stop gain + missense
144	4 d	M	*NPHS2*	Nephrotic syndrome, steroid-resistant	Homozygous, missense
147	6 y	F	*UBTF*	Neurodegeneration, with brain atrophy	*de novo,* missense
151	4 d	M	*KAT6B*	Genitopatellar syndrome	*de novo,* frameshift
159	8 d	M	*FLNB*	Larsen syndrome	*de novo,* missense
164	11 d	F	*CHD7*	CHARGE syndrome	*de novo*, splice site
170	7 d	F	7q del	7q36.3 deletion[d]	*de novo*, deletion (16 Mb)
174	4 m	F	*NDUFV1*	Mitochondrial complex I deficiency	Homozygous, missense
176	9 y	F	*SCN1A*	Dravet syndrome	*de novo,* missense
177	18 m	M	*PPP2R5D*	Mental retardation	*de novo,* missense
178	4 m	M	*DNM1*	Early infantile epileptic encephalopathy	*de novo,* missense
191	5 y	M	*GNAO1*	Neurodev. disorder, involun. movements	*de novo,* in frame deletion
198	10 m	F	*CDKL5*	Early infantile epileptic encephalopathy	*de novo* X-linked dom., missense
205	3 m	F	*MAP2K1*	Cardiofaciocutaneous syndrome	*de novo,* missense
206	2 y	M	*DMD*	Duchenne muscular dystrophy	*de novo* X-linked rec., stop gain

[a] Age postnatal; d: days, w: weeks, m: months, y: years
[b] Diagnosis only partially explains phenotype
[c] Not yet in OMIM, published in Oates E. C., *et al.* Ann Neurol (2018) [25]
[d] Specific gene not determined, multi-genic
[e] Not yet in OMIM, published in Alston C. L., *et al.* Am J Hum Genet (2018) [24]

Supplementary Table 5. Pathogenic variant information

Sample ID	Gene	Pathogenicity	Variant type	Consequence	Inheritance	HGVSc
NGC00031_01	*MTM1*	likely pathogenic	SNV	missense variant	X-linked recessive	ENST00000370396.2:c.1373T>C
NGC00033_01	*BRAF*	pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000288602.6:c.1513C>T
NGC00044_01	*COL2A1*	likely pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000380518.3:c.1358G>C
NGC00058_01	*CYP21A2*	pathogenic	SNV	splice variant	autosomal recessive homozygous	NM_000500.7:c.293-13C>G
NGC00062_01	*SATB2*	likely pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000417098.1:c.1564C>T
NGC00067_01	*SMC1A*	vus	SNV	missense variant	X-linked dominant	ENST00000322213.4:c.1994G>A
NGC00069_01	*ARID1B*	pathogenic	complex SV	deletion	*de novo* autosomal dominant	chr6:154773915-171115067
NGC00075_01	*ARID1B*	pathogenic	indel	frameshift variant	*de novo* autosomal dominant	ENST00000346085.5:c.5570_5573 delAAGA
NGC00076_01	*NIPBL*	pathogenic	indel	frameshift variant	*de novo* autosomal dominant	ENST00000282516.8:c.6679_6682 delGTCA
NGC00078_01	*COL2A1*	likely pathogenic	SNV	splice donor variant	*de novo* autosomal dominant	ENST00000380518.3:c.1419+1G>A
NGC00094_01	*HBB*	pathogenic	indel	frameshift variant	autosomal recessive homozygous	ENST00000335295.4:c.17_18delCT
NGC00096_01	*ASXL3*	pathogenic	indel	frameshift variant	*de novo* autosomal dominant	ENST00000269197.5:c.3332_3333 delTT
NGC00099_01	*SCN2A*	likely pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000357398.3:c.782T>G
NGC00101_01	*TTN*	likely pathogenic	indel	frameshift variant	autosomal recessive comp. het.	ENST00000589042.1:c.104717delA
NGC00101_01	*TTN*	likely pathogenic	indel	frameshift variant	autosomal recessive comp. het.	ENST00000589042.1:c.56330_56334 delTGAGA
NGC00101_04	*TTN*	likely pathogenic	indel	frameshift variant	autosomal recessive comp. het.	ENST00000589042.1:c.104717delA
NGC00101_04	*TTN*	likely pathogenic	indel	frameshift variant	autosomal recessive comp. het.	ENST00000589042.1:c.56330_56334 delTGAGA
NGC00104_01	*GK*	likely pathogenic	SNV	missense variant	X-linked recessive	ENST00000378943.3:c.1088C>A
NGC00109_01	*TBCD*	pathogenic	indel	splice acceptor variant	autosomal recessive comp. het.	ENST00000539345.2:c.1150_1171 delATCGGTAGGATGGCTGGCAGGC
NGC00109_01	*TBCD*	likely pathogenic	SNV	missense variant	autosomal recessive comp. het.	ENST00000539345.2:c.1589T>C
NGC00111_01	*None*	likely pathogenic	CNV	deletion	*de novo* autosomal dominant	chr5:96267459-120620855del
NGC00114_01	*NDUFA6*	likely pathogenic	indel	frameshift variant	autosomal recessive comp. het.	ENST00000498737.2:c.433delC
NGC00114_01	*NDUFA6*	likely pathogenic	indel	frameshift variant	autosomal recessive comp. het.	ENST00000498737.2:c.387delT

NGC00125_01	*TGFBR1*	pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000552516.1:c.734C>T
NGC00126_01	*RHOBTB2*	pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000519685.1:c.1448G>A
NGC00127_01	*GFAP*	pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000586793.1:c.235C>T
NGC00130_01	*PDHA1*	pathogenic	SNV	missense variant	*de novo* X-linked dominant	ENST00000379806.5:c.901C>G
NGC00135_01	*SCN1A*	pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000303395.4:c.1178G>A
NGC00138_01	*DNAH11*	pathogenic	SNV	stop gained	autosomal recessive comp. het.	ENST00000328843.6:c.4348C>T
NGC00138_01	*DNAH11*	vus	SNV	missense variant	autosomal recessive comp. het.	ENST00000328843.6:c.10493G>A
NGC00139_01	*ALS2*	pathogenic	indel	frameshift variant	autosomal recessive comp. het.	ENST00000264276.6:c.1867_1868delCT
NGC00139_01	*ALS2*	pathogenic	SNV	splice acceptor variant	autosomal recessive comp. het.	ENST00000264276.6:c.1738-1G>C
NGC00143_01	*PYGM*	likely pathogenic	SNV	missense variant	autosomal recessive comp. het.	ENST00000164139.3:c.1160G>A
NGC00143_01	*PYGM*	pathogenic	SNV	stop gained	autosomal recessive comp. het.	ENST00000164139.3:c.148C>T
NGC00144_01	*NPHS2*	pathogenic	SNV	missense variant	autosomal recessive homozygous	ENST00000367615.4:c.413G>A
NGC00147_01	*UBTF*	pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000302904.4:c.628G>A
NGC00151_01	*KAT6B*	pathogenic	indel	frameshift variant	*de novo* autosomal dominant	ENST00000287239.4:c.3752_3759 delGAACAAAG
NGC00159_01	*FLNB*	pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000490882.1:c.5164G>A
NGC00164_01	*CHD7*	pathogenic	SNV	splice acceptor variant	*de novo* autosomal dominant	ENST00000423902.2:c.3990-1G>T
NGC00170_01	*None*	pathogenic	translocation	deletion	*de novo* autosomal dominant	chr7:155207859-159125030del
NGC00174_01	*NDUFV1*	likely pathogenic	SNV	missense variant	autosomal recessive homozygous	ENST00000322776.6:c.1157G>A
NGC00176_01	*SCN1A*	likely pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000303395.4:c.4489G>A
NGC00177_01	*PPP2R5D*	likely pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000485511.1:c.752A>T
NGC00178_01	*DNM1*	likely pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000372923.3:c.1076G>A
NGC00191_01	*GNAO1*	likely pathogenic	indel	in-frame deletion	*de novo* autosomal dominant	ENST00000262493.6:c.1021_1023 delGAC
NGC00198_01	*CDKL5*	pathogenic	SNV	missense variant	*de novo* X-linked dominant	ENST00000379989.3:c.380A>T
NGC00205_01	*MAP2K1*	likely pathogenic	SNV	missense variant	*de novo* autosomal dominant	ENST00000307102.5:c.389A>G
NGC00206_01	*DMD*	pathogenic	SNV	stop gained	*de novo* X-linked recessive	ENST00000357033.4:c.2137C>T

Supplementary Table 6. Probands from the NICU grouped by phenotype similarity

Group	Diagnostic rate	Number of probands	Enriched phenotypes [# probands][a]	Diagnosed genes
1	100%	2	Hypertelorism [2], Talipes [2]	*COL2A1, FLNB*
2	50%	4	None significant. Top hits[b]: Abnormal renal physiology [4], Abnormal urine output [3]	*CYP21A2, NPHS2*
3	50%	2	Abnormal heart valve morphology [2]	*KAT6B*
4	33%	3	Cleft palate [3], Micrognathia [2]	16Mb deletion
5	25%	4	None significant. Top hits[b]: Polyhydramnios [3], Abnormal vascular physiology [3]	*TTN*
6	20%	15	None significant. Top hits[b]: Abnormal morphology of the great vessels [15], Abnormal cardiac ventricle morphology [8]	*NIPBL, CHD7,* 24Mb deletion
7	20%	5	Abnormality of nervous system physiology [5], Seizures [3], Feeding difficulties [3], Abnormality of muscle physiology [3]	*DNAH11*
8	14%	7	None significant. Top hits[b]: Tachypnea [4], Tracheoesophageal fistula [4]	*MAP2K1*
9	13%	8	Decreased body weight [8], Microcephaly [5]	*MTM1*
10	10%	10	Sepsis [10]	*SATB2*
11	0%	14	Abnormality of nervous system physiology [14], Abnormality of cation homeostasis [12], Encephalopathy [11], Seizures [11], Hyponatremia [8]. Hypotension [8], Hypoxic ischaemic encephalopathy[c] [11]	--
12	0%	13	None significant. Top hit[b]: Hyperglycemia [7]	--
13	0%	10	Abnormality of prenatal development or birth [10], Abnormality of the amniotic fluid [9]	--
14	0%	5	Meconium stained amniotic fluid [5], Aspiration [4]	--
15	0%	4	Ventriculomegaly [4], Absent septum pellucidum [4], Agenesis of corpus callosum [3], Abnormal cerebellum morphology [3]	--

[a] Phenotypes significantly enriched in group over cohort (Fisher's Exact test, FDR<0.1) and found in over half of the probands in the group. Simplified for ontology redundancy. [b] Most enriched term(s) in a group with no significantly enriched terms. [c] Hypoxic ischaemic encephalopathy (HIE) is not a term in HPO, but cases with suspected HIE are significantly enriched Group 11.

Supplementary Table 7. Probands from the PICU grouped by phenotype similarity

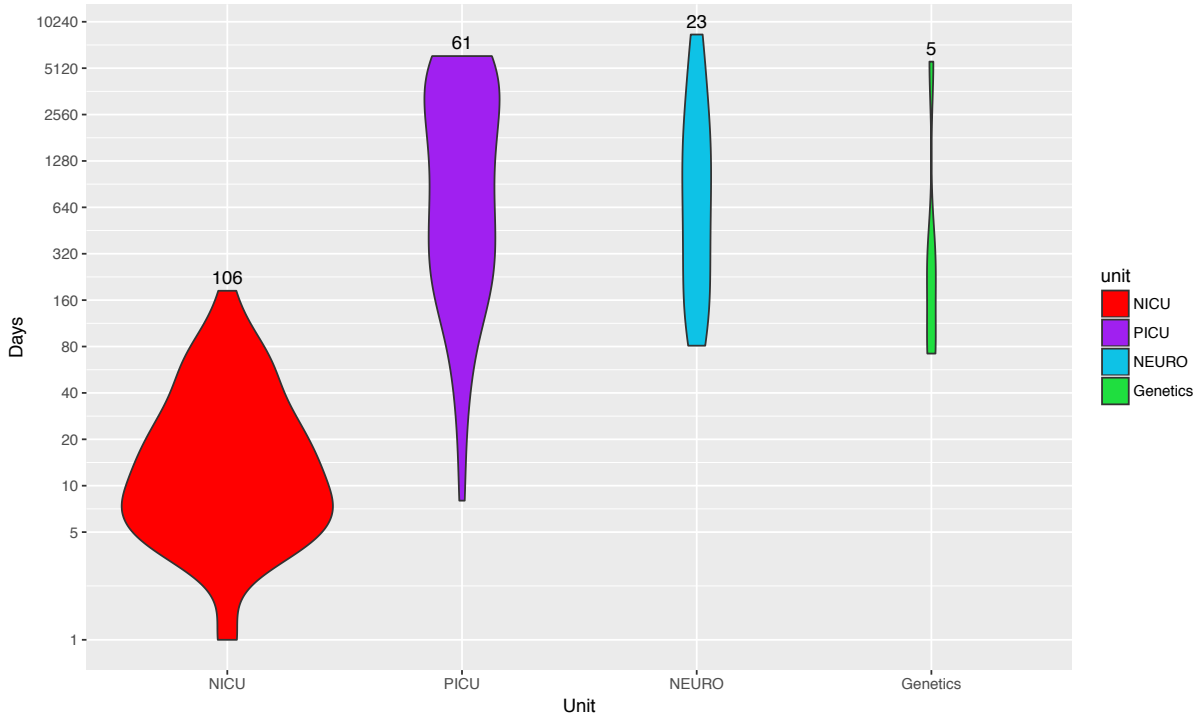
Group	Diagnostic rate	Number of probands	Enriched phenotypes [# probands][a]	Diagnosed genes
1	50%	4	None significant. Top hit[b]: Abnormality of the amniotic fluid [4]	*BRAF, RHOBTB2*
2	50%	2	None significant. Top hit[b]: Lactic acidosis [2]	*NDUFA6*
3	40%	5	Cerebral palsy [3]	*SCN2A, COL2A1*[c]
4	36%	11	Abnormality of body height [10], Abnormality of skull size [10], Short stature [8]	*SMC1A, ARID1B, PPP2R5D, HBB*[c]
5	33%	3	Ventilator dependence with inability to wean [2]	*PYGM*
6	25%	4	Congenital malformation of the great arteries [3]	*TGFBR1*
7	22%	9	None significant. Top hits[b]: Feeding difficulties [7], Abnormal facial shape [5], Abnormality of eye movement [5]	*ASXL3, ARID1B*
8	13%	15	Abnormality of immune system physiology [14], Sepsis [10], Respiratory tract infection [10]	*DMD, GK*
9	0%	4	None significant. Top hit[b]: Functional respiratory abnormality [4]	--
10	0%	2	Joint hypermobility [2]	--
11	0%	2	Prolonged neonatal jaundice [2]	--

[a] Phenotypes significantly enriched in group over cohort (Fisher's Exact test, FDR<0.1) and found in over half of the probands in the group. Simplified for ontology redundancy. [b] Most enriched term(s) in a group with no significantly enriched terms. [c] Gene partially explains phenotype.

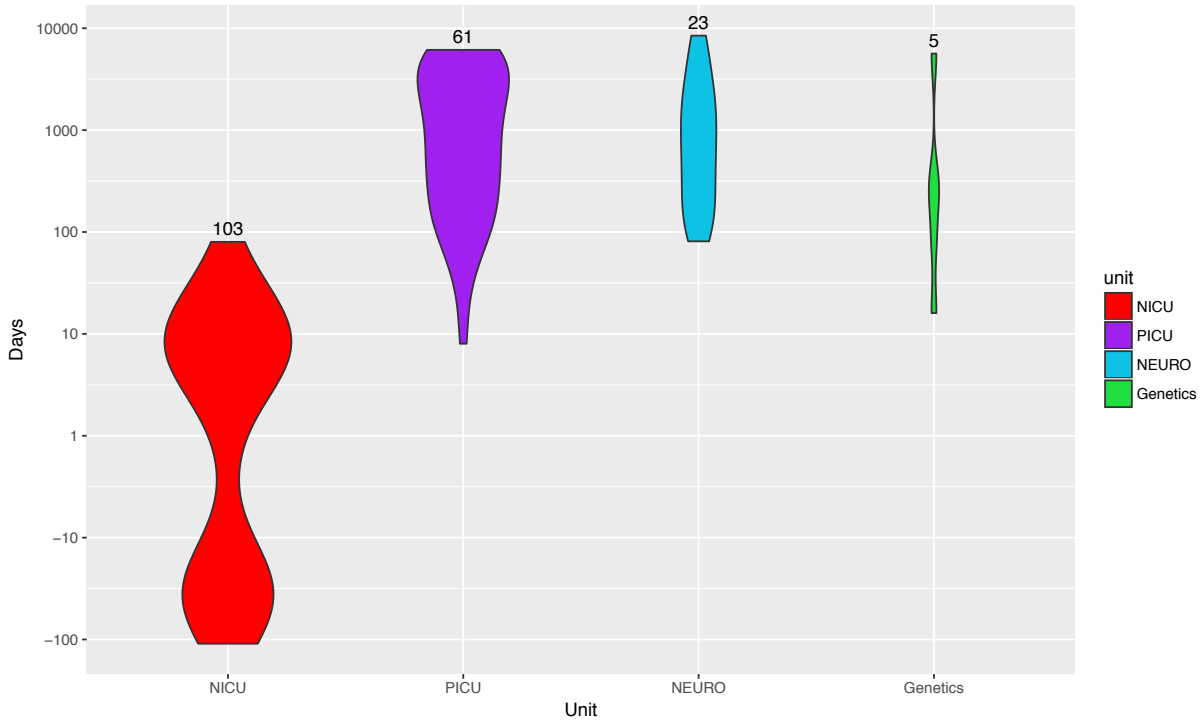
SUPPLEMENTARY FIGURES

Supplementary Figure 1. Recruitment age and ward

a.

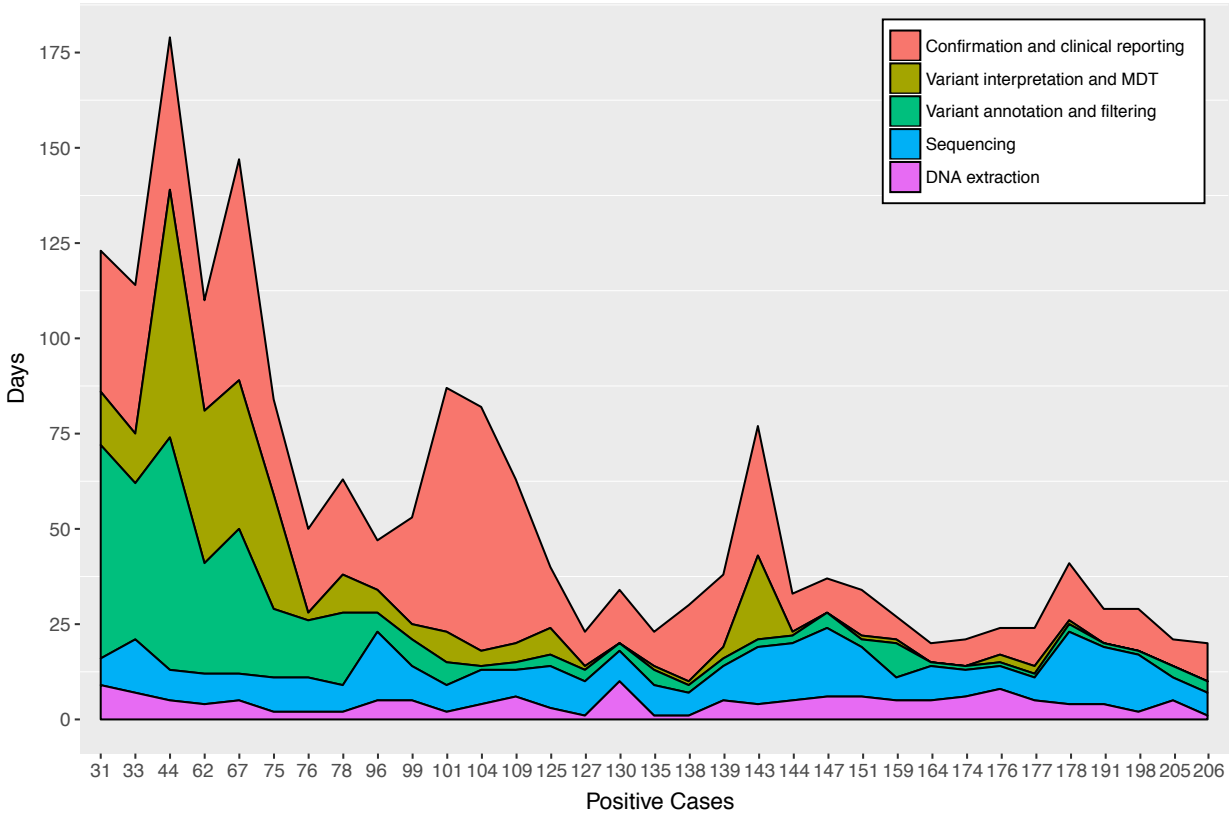


b.



Supp. Fig. 1 Violin plots showing distribution of age at recruitment in each ward for 195 families. The number above the plots is the number recruited from each ward. NICU: neonatal intensive care unit and 2 cases from delivery wards. PICU: paediatric intensive care unit. NEURO: paediatric neurology department. Genetics: clinical genetics department. **a.** days after birth, **b.** days from 40 week full term birth (3 probands were removed from NICU because gestational age information wasn't available)

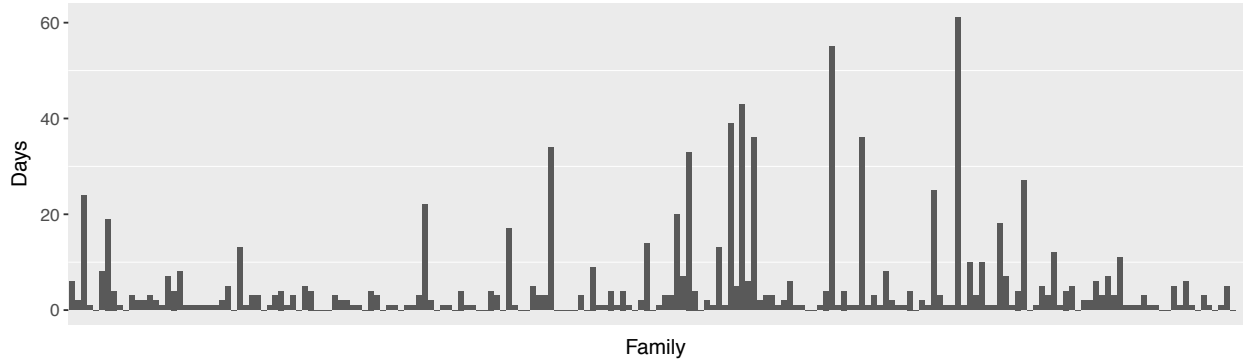
Supplementary Figure 2. Time to clinical diagnostic report



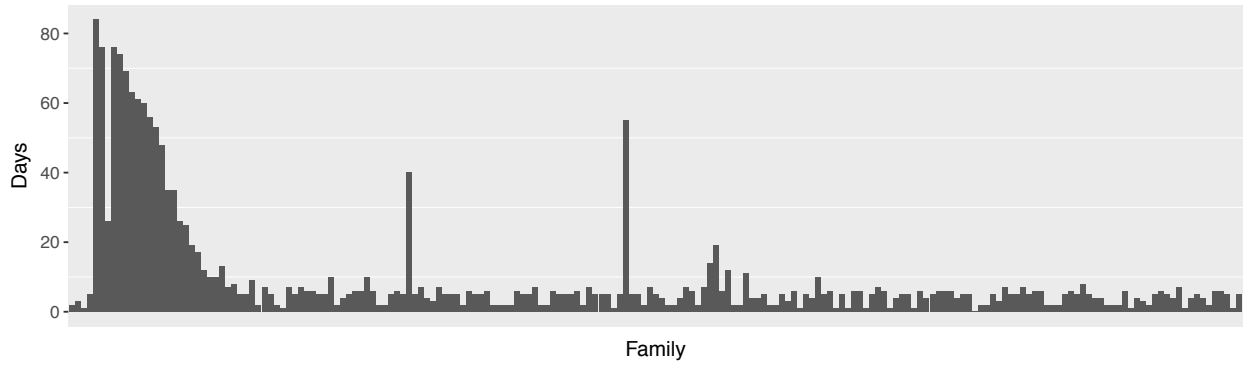
Supp. Fig. 2 Stacked area chart showing the length of time in days for each step for positive cases. Note that confirmation and clinical reporting (red) and variant interpretation (gold) occasionally overlapped for difficult cases (e.g. 101). Not included are five cases where the variant was not originally reported through our study and two cases where diagnoses were made after subsequent re-analysis

Supplementary Figure 3. Turnaround times

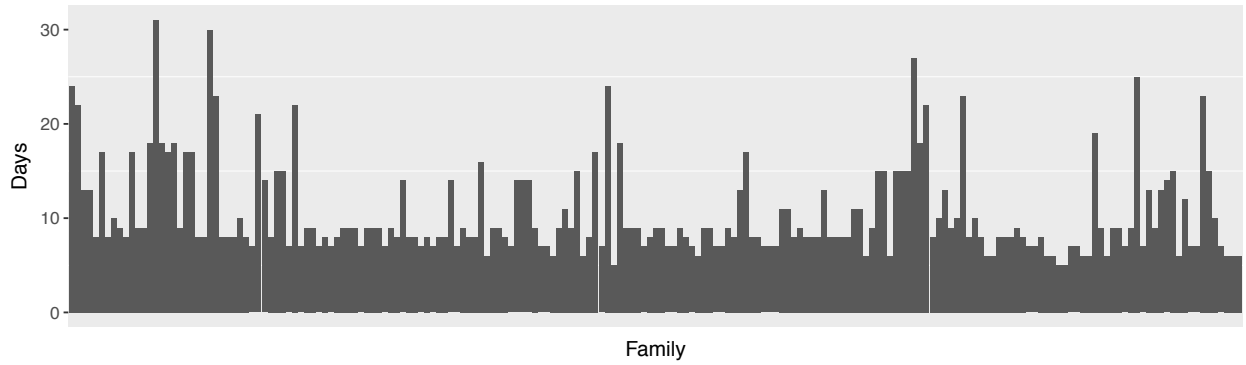
a. Consent date to all blood samples received



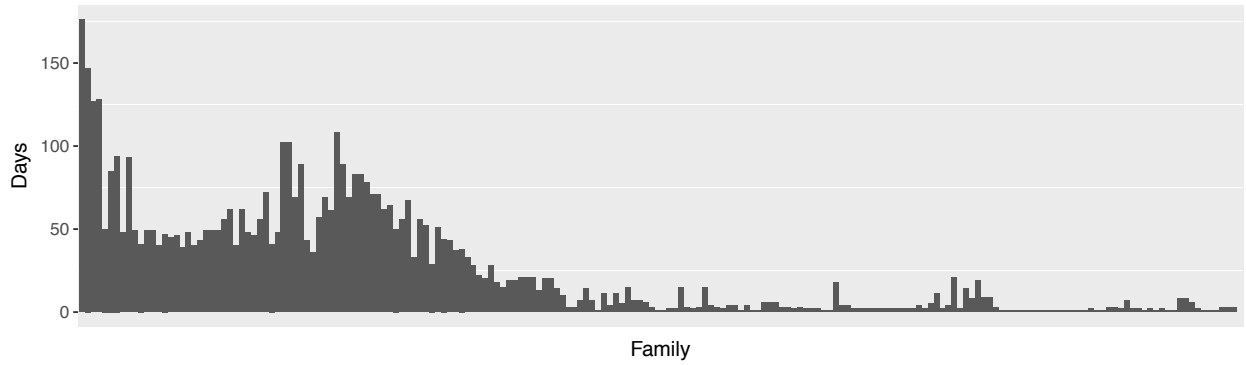
b. DNA extraction and processing



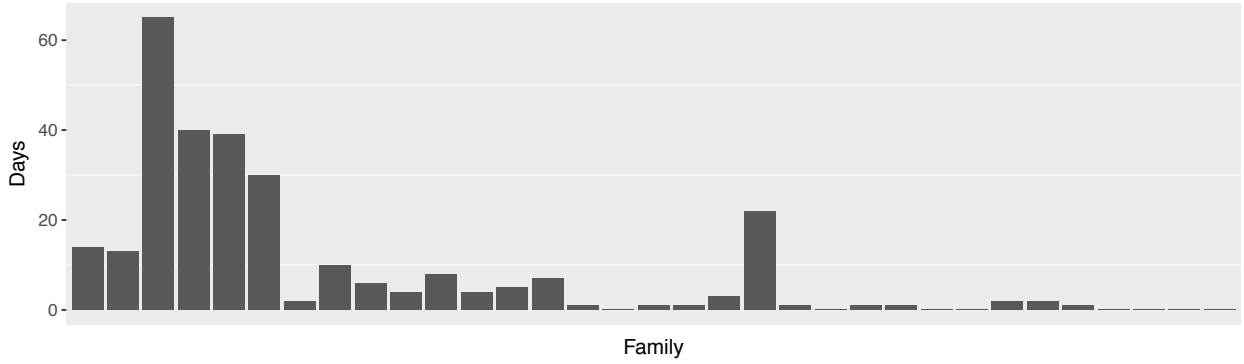
c. Sequencing and variant calling



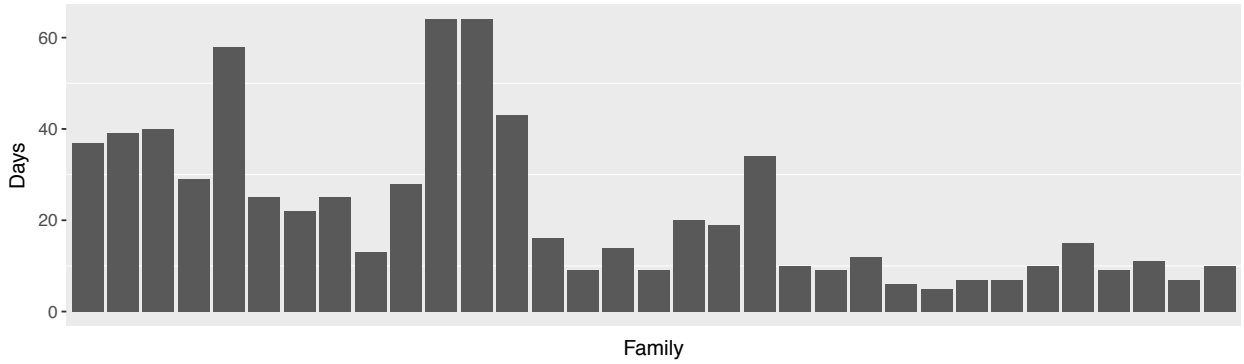
d. Data analysis



e. Variant interpretation

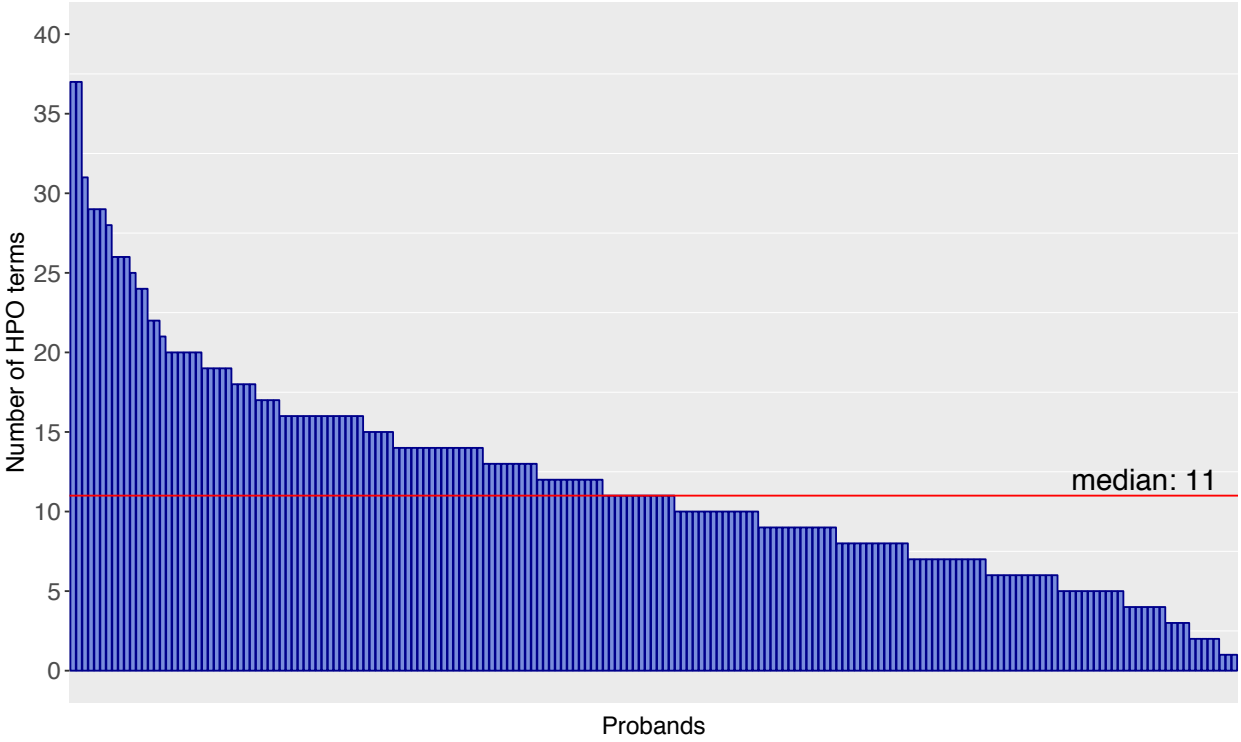


f. Confirmation and clinical reporting



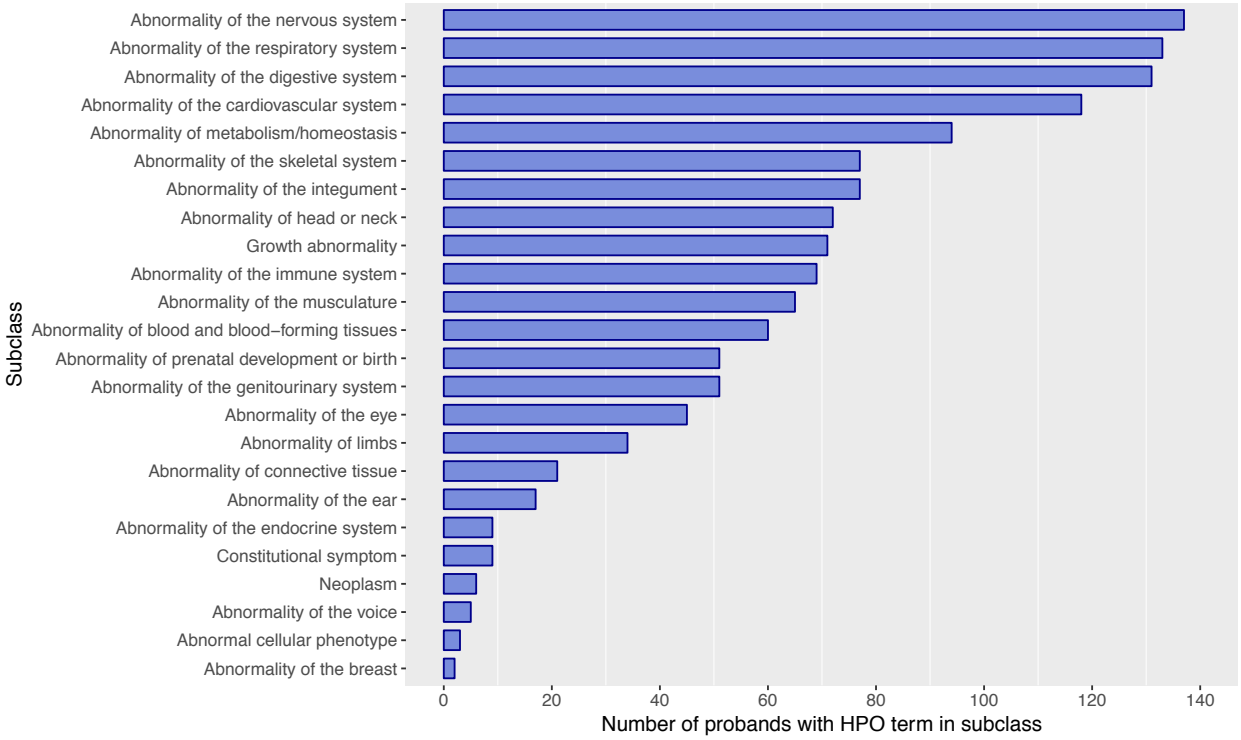
Supp. Fig. 3 Bar plots of the turnaround times for each step for each family. a) Sample collection for all members of family. b) Sample processing, DNA extraction, quality control, and delivery to sequencing center. c) Sequencing and variant calling. d) Variant annotation, filtering, and review. e) Variant interpretation in multi-disciplinary team meetings (positive cases only). f) Variant confirmation, continuing interpretation, and diagnostic reporting (positive cases only). Note that variant interpretation (e) and confirmation and clinical reporting (f) occasionally overlapped for difficult cases

Supplementary Figure 4. Number of HPO terms per proband



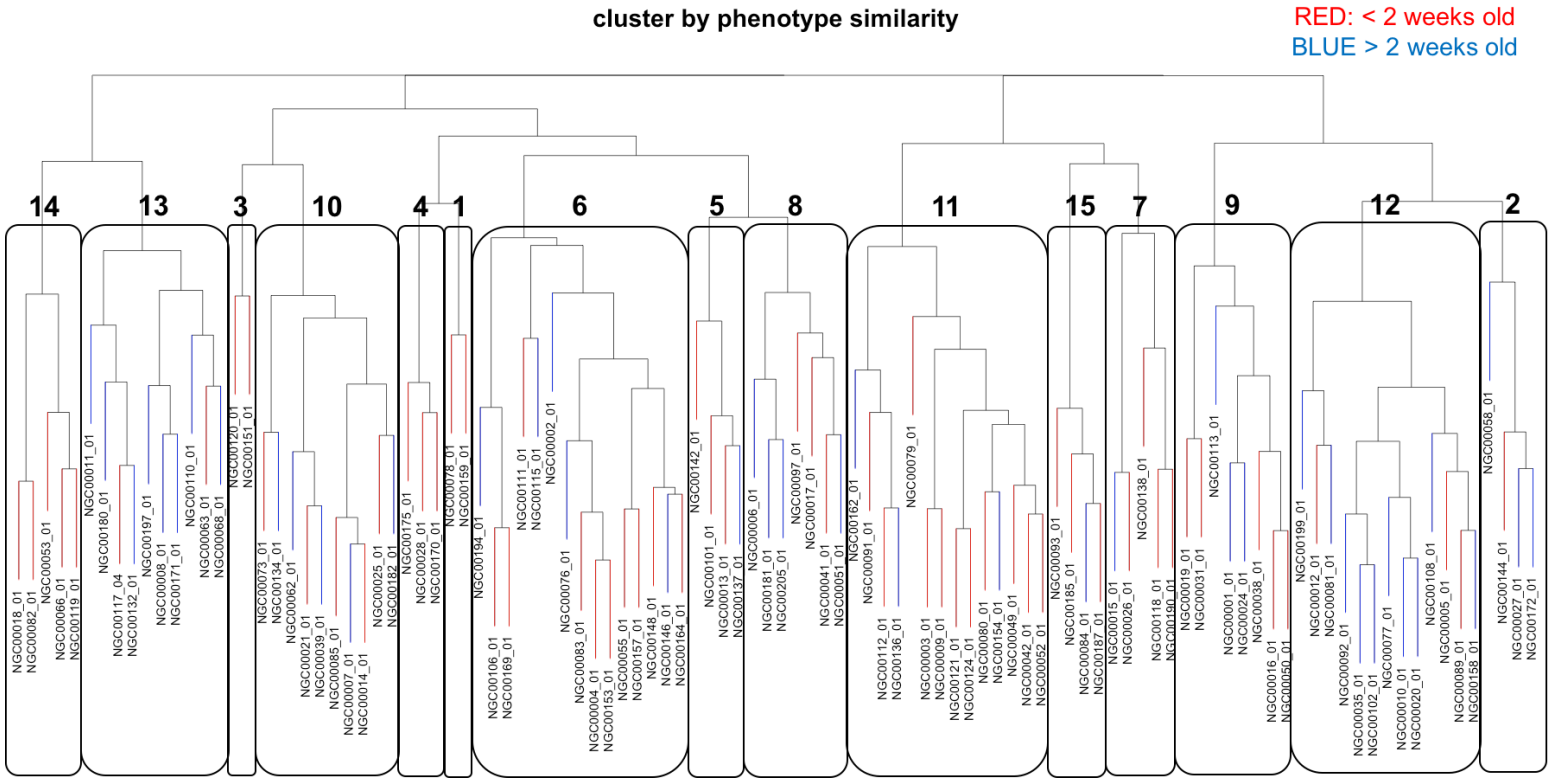
Supp. Fig. 4 Number of non-redundant HPO terms assigned to each proband. These exclude maternal phenotypes, and those relating to birth (e.g. 'Premature birth', 'Abnormal birth'). Probands are ordered by number of terms

Supplementary Figure 5. Number of probands with phenotypes in the major subclasses



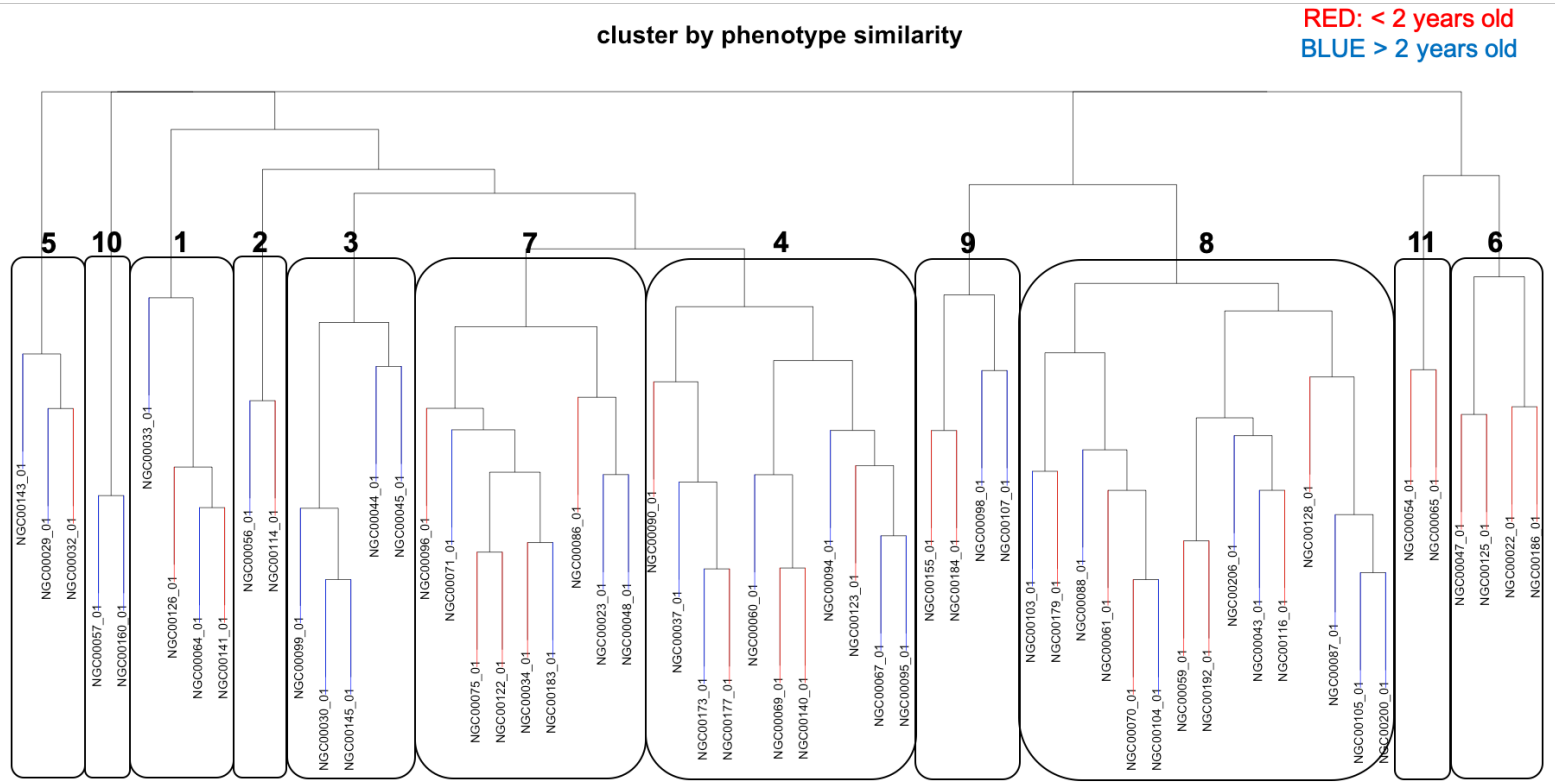
Supp. Fig. 5 Number of probands with at least one HPO term for each class. The classes are the children terms of the HPO term 'Phenotypic abnormality'

Supplementary Figure 6. Probands from the NICU grouped by phenotype similarity



Supp. Fig 6. Probands from the neonatal intensive care unit (NICU) were hierarchically clustered by the semantic similarity score of their HPO term profiles and divided into 15 groups. Group numbers match those in Table 2 and Supplementary Table 6. Leaves are colored by the proband's age at recruitment (red: less than 2 weeks old; blue: over 2 weeks old)

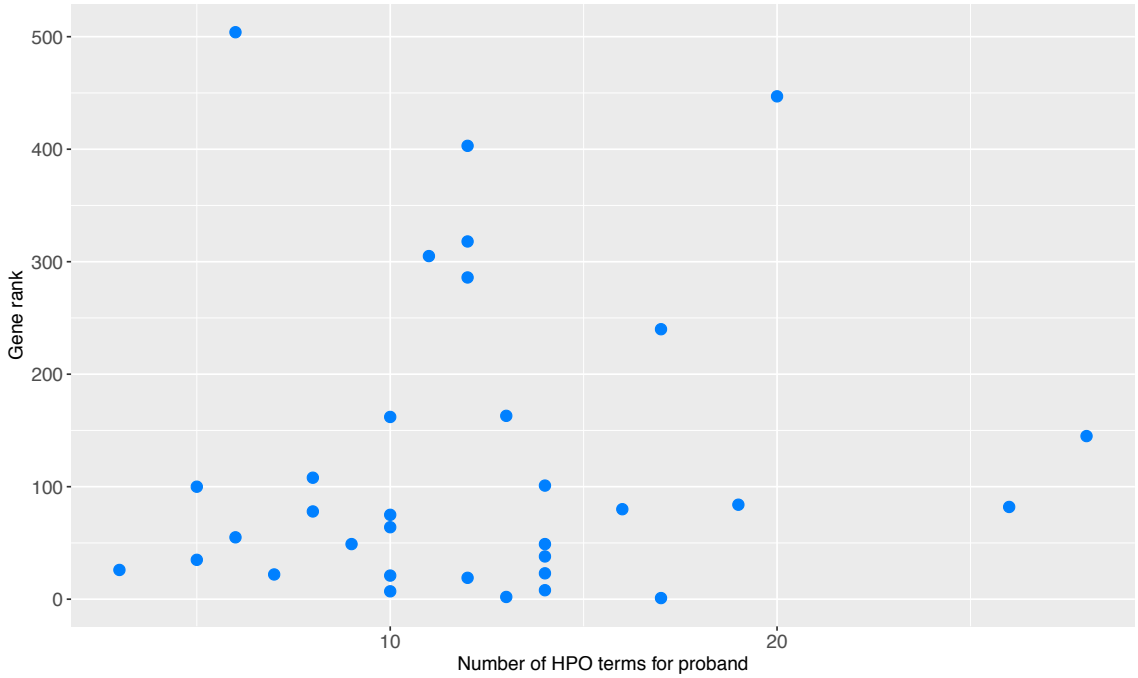
Supplementary Figure 7. Probands from the PICU grouped by phenotype similarity



Supp. Fig. 7 Probands from the paediatric intensive care unit (PICU) were hierarchically clustered by the semantic similarity score of their HPO term profiles and divided into 11 groups. Group numbers match those in Table 2 and Supplementary Table 7. Leaves are colored by the proband's age at recruitment (red: less than 2 years old; blue: over 2 years old)

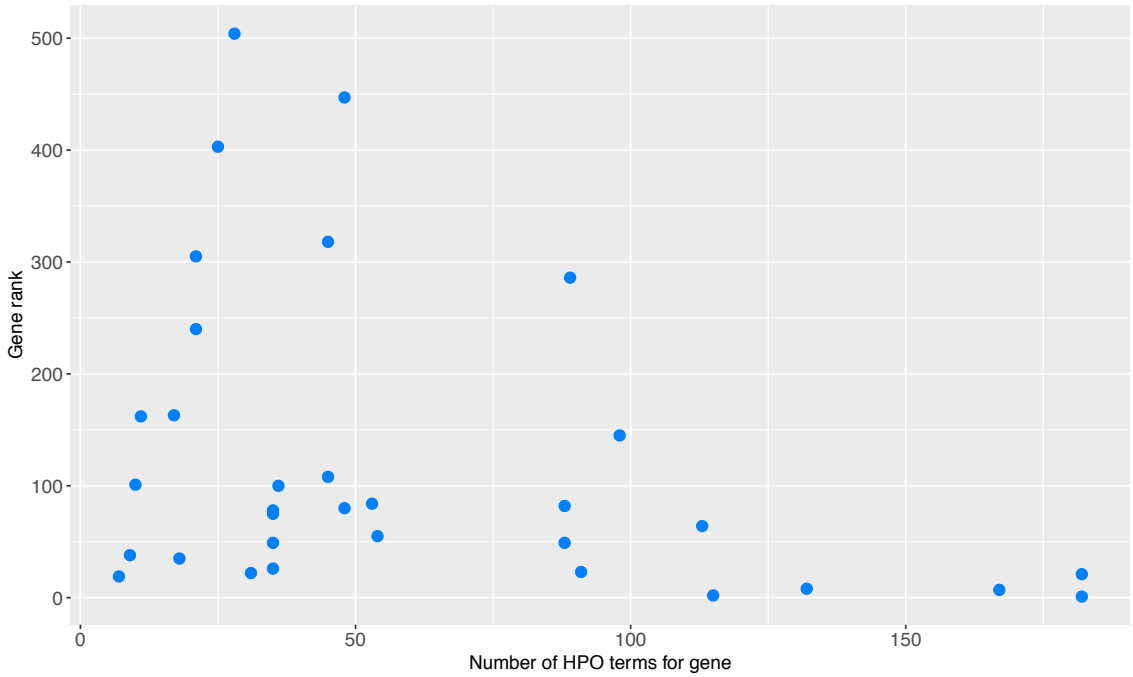
Supplementary Figure 8. Correlation between diagnosed gene similarity rank and number of HPO terms

a. Number of proband's HPO terms vs diagnosed gene rank in similarity to proband



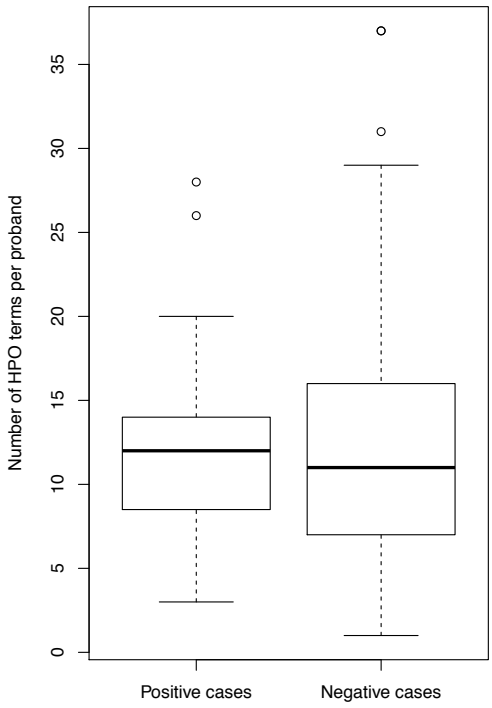
Number of non-redundant HPO terms assigned to each proband plotted against where the diagnosed gene ranked in phenotype similarity score to the proband compared to all HPO-typed genes in OMIM.

b. Number of gene's HPO terms vs diagnosed gene rank in similarity to proband



Number of non-redundant HPO terms associated with the diagnosed gene in OMIM plotted against where the diagnosed gene ranked in phenotype similarity score to the proband compared to all HPO-typed genes in OMIM.

Supplementary Figure 9. Number of phenotypes for diagnosed and non-diagnosed cases



Supp. Fig. 9 Boxplots of the number of HPO terms per proband for the 40 diagnoses cases (Positive cases) and the 155 cases for which we have not reported a genetic diagnosis (Negative cases). The positive case mean is 12.2 and the negative case mean is 11.7 (no significant increase for positive cases, Student's t-test p = 0.3422).
Bar: median (positive: 12, negative 11); Box: IQR (interquartile range, 25%-75%); Whiskers: +/-1.58 IQR/sqrt(n)