Supplementary Material

Supplementary Tables

Supplementary Table S1. Recall and precision values for each broad phenotype category by submissions

		Recall	Precision
SID#1	Eye	0.33	0.50
	Neuro	0.43	0.43
	Connective	0.27	0.50
SID#2	Eye	0.17	0.17
	Neuro	0.00	0.00
	Connective	0.18	0.50
SID#3	Eye	0.17	0.14
	Neuro	0.29	0.50
	Connective	0.55	0.46
SID#4	Eye	0.50	0.50
	Neuro	0.43	0.43
	Connective	0.55	0.55
SID#5	Eye	0.50	0.38
	Neuro	0.43	0.33
	Connective	0.45	0.50
SID#6.1	Eye	0.00	0.00
	Neuro	0.14	0.14
	Connective	0.55	0.46
SID#6.2	Eye	0.00	0.00
	Neuro	0.14	0.50
	Connective	0.91	0.43
SID#7	Eye	0.33	0.33
	Neuro	0.43	0.33
	Connective	0.27	0.60
SID#8	Eye	0.50	0.43
	Neuro	0.29	0.33
	Connective	0.55	0.55

Supplementary Table S2. Number of gender matches for each submission

Submission ID	Gender match	Number of genomes not matched to any patient
SID#1	23	0
SID#2	**12**	12
SID#3	**24**	0
SID#4	**24**	0
SID#5	**12**	12
SID#6.1	23	0
SID#6.2	17	0
SID#7	18	3
SID#8	**24**	0

Supplementary Table S3. Benign nominated diagnostic variants predicted with the highest probability for correct genome-patient matches

Genome (patient)	Phenotype category	SID #	Genomic position (hg19)	Gene	Plausible gene for category?	Gene Phenotype Correlation	Transcript	Nucleotide change	Protein change	Classification	Gene disease associations
7(X)	Ophthalmologic	7	9:80647059:GC:TT	*GNAQ*	Yes	Poor	NM_002072.4	c.-908_909GC>TT	5'UTR	Benign	Somatic mutations cause Sturge Weber syndrome
		8	2:216251538:G:A	*FN1*	No	Poor	NM_212482.2	c.4486C>T	p.Arg1496Trp	Likely Benign	Glomerulopathy; Fibronectin deficiency; Spondylometaphyseal dysplasia
17(H)	Ophthalmologic	4	1:94568686:C:T	*ABCA4*	Yes	Good	NM_000350.2	c.455G>A	p.Arg152Gln	Likely Benign	Retinitis pigmentosa, red-cone dystrophy and other eye disorder (AR inheritance)
			1:94470320:C:T				NM_000350.2	c.6147+677G>A	Intronic	Likely Benign	
21(G)	Neurologic	6.1	12:7343108:GGCCTCTGAGGCAGTGAGTGTTCTTGAGGTGGAAAGCCCAGGTGCA:G	*PEX5*	Yes	Poor Partial	NM_001131023.1	c.192+32_192+76del	None (3 repeats to 2)	Benign	Peroxisome biogenesis disorders, Rhizomelic chondrodysplasia punctata
42(O)	Ophthalmologic	3	16:1621518:GA:TT	*IFT140*	Yes	Good	NM_014714.4	c.1541_1542delTCinsAA	p.Leu514Gln	Benign	Retinitis pigmentosa 80, Short-rib thoracic dysplasia 9 with or without polydactyly
56(N)	Connective	4	15:48700642:T:C	*FBN1*	Yes	Some	NM_000138.4	c.*2545A>G	3'UTR	Likely Benign	Marfan syndrome
			15:48849792:C:A				NM_000138.4	c.539-19787G>T	Deep intronic	Likely Benign	
		8	6:32012817:C:T	*TNXB*	Yes	Some	NM_019105.7	c.10887G>A	p.Lys3629=	Benign	Ehlers-Danlos like syndrome
68(J)	Neurological	8	11:108196896:C:T	*ATM*	Yes	Poor	NM_000051.3	c.6919C>T	p.Leu2307Phe	Likely Benign	Ataxia telangiectasia
71(L)	Connective	1	*2:75811731:G:A	*EVA1A*	No	None	NM_032181.2	None	Intergenic	Likely Benign	None
78(V)	Connective	6.2	15:59443263:CGTGCACTT:C	*MYO1E*	No	None	NM_004998.3	c.3080+2518_3080+2525delAAGTGCAC	Deep intronic	Likely Benign	Glomerulosclerosis, focal segmental 6
			1:40768483:T:TGGAG	*COL9A2*	Yes	Poor	NM_001852.4	c.1604-6_1604-3dupCTCC	Intronic near splice	Benign	Multiple epiphyseal dysplasia 2

			6:70972993:T:C	*COL9A1*	Yes	Poor	NM_001851.4	c.1349A>G	p.Glu450Gly	Likely Benign	Stickler syndrome type IV, Epiphyseal dysplasia, multiple, 6
93(F)	Connective	8	12:2791205:A:G	*CACNA1C*	No	Poor	NM_001129830.2	c.5534A>G	p.Lys1845Arg	Benign	Brugada syndrome
95(C)	Ophthalmologic	4	1:215953583:A:G	*USH2A*	Yes	Good	NM_206933.2	c.10741-200T>C	Deep intronic	Likely Benign	Usher syndrome
			1:215964830:T:G				NM_206933.2	c.9959-1206A>C	Deep intronic	Likely Benign	
99(B)	Neurological	8	11:793588:G:A	*SLC25A22*	Yes	Mild	NM_001191061.1	c.234C>T	p.Pro78=	Benign	Epileptic encephalopathy, early infantile, 3
102(A)	Connective	8	2:189974958:G:T	*COL5A2*	Yes	Partial	NM_000393.4	c.315C>A	p.Thr105=	Benign	Ehlers-Danlos syndrome-Classic

*Several systematic errors or differences in how variants were referenced for submission files for groups 1 and 7. All variants identified in genes on the minus strand were represented as the base of the minus strand while convention generally refers to only plus strand bases for genomic coordinates. Assumptions were made based on notes provided with submissions, gene orientation, and know variants at those loci. Assumed validated variants are listed here and do not match perfectly with original submission files.

Supplementary Table S4. Nominated secondary variants from groups 1, 4, and 7.

Genome	Genomic position (hg19)	Transcript	Nucleotide change	Protein change	Gene	Variant type	Classification	Disease inheritance pattern	Associated disease	Reporting suggestions
7	19:34262922:C:T	NM_001127896.1	c.229C>T	p.Arg77Trp	*CHST8*	missense	Pathogenic	AR	Peeling skin syndrome 3	do not report as secondary, only carrier status
7	1:76226846:A:G	NM_001127328.2	c.997A>G	p.Lys333Glu	*ACADM*	missense	Pathogenic	AR	Acyl-CoA dehydrogenase, medium chain, deficiency of	do not report as secondary, only carrier status
7	12:88454737:T:C	NM_025114.3	c.6392A>G	p.Glu2131Gly	*CEP290*	missense	VUS	AR	Senior-Loken syndrome 6; Joubert syndrome 5; Leber congenital amaurosis 10; Meckel syndrome 4	do not report as secondary
17	11:6411935:T:CGCTGGCGCTGGC	NM_000543.4	repeat variability	repeat variability	*SMPD1*	repeat variability	Benign	AR	Niemann-Pick disease, types A/B	do not report as secondary
17	14:88452941:A:G	NM_000153.4	c.334A>G	p.Thr112Ala	*GALC*	missense	VUS	AR	Krabbe disease	do not report as secondary
18	18:21134743:G:A	NM_000271.4	c.1532C>T	p.Thr511Met	*NPC1*	missense	VUS	AR	Niemann-Pick disease, types D/C1	do not report as secondary
18	18:21140367:G:A	NM_000271.4	c.709C>T	p.Pro237Ser	*NPC1*	missense	Benign	AR	Niemann-Pick disease, types D/C1	do not report as secondary
30	7:150644901:G:A	NM_000238.3	c.2758C>T	p.Arg920Trp	*KCNH2*	missense	VUS	AD	Long QT syndrome; Short QT syndrome	do not report as secondary
30	2:47707963:TATG:T	NM_000251.2	c.2590_25992delATG	p.Asp864del	*MSH2*	in frame deletion	VUS	AD/AR	Mismatch repair cancer syndrome; Colorectal cancer, hereditary nonpolyposis, type 1; Muir-Torre syndrome	do not report as secondary
30	13:32911607:CCTA:C	NM_000059.3	c.3119_3121delCTA	p.Thr1040del	*BRCA2*	in frame deletion	VUS	AD	Fanconi anemia, complementation group D1; Wilms tumor; other cancers	do not report as secondary
39	5:125887751:C:G	NM_001201377.1	c.1279G>C	p.Glu427Gln	*ALDH7A1*	missense	VLP	AR	Epilepsy, pyridoxine-dependent	do not report as secondary, only carrier status
42	2:46386838:G:C	NM_005400.2	c.2014G>C	p.Asp672His	*PRKCE*	missense	VUS	unknown	possible association with SHORT syndrome	do not report as secondary
42	*17:79767715:G:A	NM_000160.4	c.118G>A	p.Gly40Ser	*GCGR*	missense	VUS	AD	Diabetes mellitus, noninsulin-dependent	do not report as secondary
42	2:46325089:G:C	NM_005400.2	c.1592+11588G>C	non coding	*PRKCE*	non coding	VUS	unknown	possible association with SHORT syndrome	do not report as secondary
56	16:17232391:G:A	NM_022166.3	c.1588-3C>T	splice site	*XYLT1*	splice	VLP	AR	Desbuquois dysplasia 2	do not report as secondary, only carrier status
56	4:15504459:T:G	NM_001080522.2	C.351T>G	p.Ser117Arg	*CC2D2A*	missense	VLB	AR	Joubert syndrome 9; COACH syndrome; Meckel syndrome 6	do not report as secondary

56	1:171083232:G:T	NM_006894.5	c.913G>T	p.Glu305Ter	*FMO3*	nonsense	Pathogenic	AR	Trimethylaminuria (fish odor smell)	do not report as secondary, only carrier status
68	21:35742947:T:C	NM_172201.1	c.170T>C	p.Ile57Thr	*KCNE2*	missense	VUS	AD	Long QT syndrome 6; Atrial fibrillation, familial, 4	do not report as secondary
68	7:143018525:C:G	NM_000083.2	c.501C>G	p.Phe167Leu	*CLCN1*	missense	VUS	AD/AR	Myotonia congenita; Myotonia levior	do not report as secondary
71	17:72916740:C:T	NM_173477.4	c.191G>A	p.Trp64Ter	*USH1G*	nonsense	Pathogenic	AR	Usher syndrome, type 1G	do not report as secondary, only carrier status
71	*7:117188841:TTG:	NM_000492.3	c.1360_1362delTTG	p.Leu454del	*CFTR*	in frame deletion	VLB	AR	Cystic fibrosis; Congenital bilateral absence of vas deferens; Sweat chloride elevation without CF	do not report as secondary
71	*15:45400303:C:T	NM_014080.4	c.1516G>A	p.Asp506Asn	*DUOX2*	missense	VUS	AR	Thyroid dyshormonogenesis 6	do not report as secondary
76	7:117227832:G:T	NM_000492.3	c.1624G>T	p.Gly542Ter(*)	*CFTR*	nonsense	Pathogenic	AR	Cystic fibrosis; Congenital bilateral absence of vas deferens; Sweat chloride elevation without CF	do not report as secondary, only carrier status
76	1:156108510:C:T	NM_170707.3	c.1930C>T	p.Arg644Cys	*LMNA*	missense	VUS	AD	LMNA-related diseases	do not report as secondary
78	20:52789466:CCTT:C	NM_000782.4	c.428_430delAAG	p.Glu143del	*CYP24A1*	in frame deletion	Pathogenic	AR	Hypercalcemia, infantile, 1	do not report as secondary, only carrier status
79	1:237824218:C:T	NM_001035.2	c.8407C>T	p.Arg2803Trp	*RYR2*	missense	VUS	AD	Arrhythmogenic right ventricular dysplasia 2; Ventricular tachycardia, catecholaminergic polymorphic, 1	do not report as secondary
79	11:76925023:T:C	NM_000260.3	c.6557T>C	p.Leu2186Pro	*MYO7A*	missense	VLP	AR	Usher syndrome, type 1B; Deafness, AD 11 / AR 2	do not report as secondary
81	*X:153762634:G:A	NM_000402.4	c.653C>T	p.Ser218Phe	*G6PD*	missense	Pathogenic	XLD	Hemolytic anemia, G6PD deficient (favism)	do not report as secondary, unless thought to be cause of disease, most people unaffected, common cause of favism
92	6:26093141:G:A	NM_000410.3	c.845G>A	p.Cys282Tyr	*HFE*	missense	Pathogenic	AR	Hemochromatosis	do not report as secondary, only carrier status
97	21:47542052:C:T	NM_001849.3	c.1552C>T	p.Pro518Ser	*COL6A2*	missense	VLB	AD/AR	Bethlem myopathy 1; Ullrich congenital muscular dystrophy 1	do not report as secondary
99	2:50851527:C:T	NM_004801.5	c.831-772G>A	intronic	*NRXN1*	non coding	VUS	AR	Pitt-Hopkins-like syndrome 2	do not report as secondary
99	2:51093481:T:G	NM_004801.5	c.823+56312A>C	non coding	*NRXN1*	non-coding	VUS	AR	Pitt-Hopkins-like syndrome 2	do not report as secondary
99	2:50570600:T:TA	NM_004801.5	c.3365-106493dupT	non coding	*NRXN1*	non-coding	VLB	AR	Pitt-Hopkins-like syndrome 2	do not report as secondary

21, 71	2:48027755:T:C	NM_000179.2	c.2633T>C	p.Val878Ala	*MSH6*	missense	Benign	AD	Mismatch repair cancer syndrome; Colorectal cancer, hereditary nonpolyposis, type 5	do not report as secondary
81, 91	2:47641558:GT:G	NM_000251.2	c.942+2delT	splice site	*MSH2*	splice	VLP	AD	Mismatch repair cancer syndrome; Colorectal cancer, hereditary nonpolyposis, type 1; Muir-Torre syndrome	report as secondary finding

*Several systematic errors or differences in how variants were referenced for submission files for groups 1 and 7. All variants identified in genes on the minus strand were represented as the base of the minus strand while convention generally refers to only plus strand bases for genomic coordinates. In addition, insertions and deletions were generally off by 1 or 2 base pairs relative to convention. Assumptions were made based on notes provided with submissions, gene orientation, and know variants at those loci. Assumed validated variants are listed here and do not match perfectly with original submission files.

Patient Phenotype descriptions

Bioinformatic groups were given the following patient phenotype information text. Some patients had pedigrees and some growth charts that are not included here-in.

Patient 7(X)
Sex: Female
Indication for referral: High myopia and bilateral retinal hamartomas

Family history and pedigree
Paternal ethnicity:
1. Oman
2. Middle East
Maternal ethnicity:
1. East Indian
NO Other affected relatives
NO Consanguinity
NO Parents with at least 3 miscarriages

Prenatal and perinatal history
NO Multiple gestation
Gestation at delivery (weeks) 41
Maternal age at EDD (years) 27
DELIVERY
NO Premature birth
Primary Caesarian section
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
Notes:
Unremarkable pregnancy and emergency C-section at 41 weeks gestation. BW 2.8 kg. Well in immediate neonatal period.

Medical history
Medical and developmental history:
Nystagmus at 7 weeks, bilateral retinal hamartomas discovered at 2 months, and diagnosed with high myopia
at 6 months. Initially was assessed in London, England. There was a question of partial ocular albinism, and so
panel testing was done that was negative. After coming to Toronto, she was noted to not have blond fundi and the Ophthalmologists are not suspecting ocular albinism.
Global age of onset: Infantile onset

Measurements
Date: Age: 5m
Weight: 2.8 kg 0th pctl (-7.51SD)
Date: Age: 7y 5m
Weight: 26.4 kg 76th pctl (+0.7SD)
Height: 133.1 cm 96th pctl (+1.72SD)
BMI: 14.9 36th pctl (-0.37SD)
Head circumference: 50.0 cm 9th pctl (-1.32SD)

Clinical symptoms and physical findings
EYE DEFECTS
Nystagmus
Onset at 7 weeks of age
Retinal hamartoma Bilateral
Identified at 2 months of age
Severe Myopia -10, diagnosed at 6 months of age
NO Hypopigmentation of the fundus
EAR DEFECTS
Recurrent otitis media
CUTANEOUS
NO Hypopigmentation of the skin
Multiple cafe-au-lait spots
4-5
NO Fair hair
MUSCULOSKELETAL
Hyperextensibility of the finger joints
Hyperextensibility at elbow
GENITOURINARY
NO Abnormality of the kidney
Normal abdominal ultrasound
BEHAVIOR, COGNITION AND DEVELOPMENT
NO Global developmental delay
High functioning
Attention deficit hyperactivity disorder
Borderline
NEUROLOGICAL
NO Morphological abnormality of the central nervous system
Essentially normal brain MRI. Nonspecific foci of signal abnormality are seen in the subcortical white matter of
both frontal lobes.

Diagnosis
Additional comments:
?OCCULAR ALBINISM ?DEVELOPMENTAL EYE DISORDER

Patient 9(W)
Sex: Female

Family history and pedigree
Paternal ethnicity:
1. French Canadian
Maternal ethnicity:
1. British

Medical history
Global age of onset: Childhood onset

Clinical symptoms and physical findings
EYE DEFECTS
Nystagmus
CUTANEOUS
Fair hair
BLOOD AND BLOOD-FORMING TISSUES
Bruising susceptibility

Patient 17(H)
Sex: Male
Indication for referral: Retinitis pigmentosa

Family history and pedigree
Paternal ethnicity:
1. British
Maternal ethnicity:
1. Canadian

List health conditions found in family (describe the relationship with proband)
No history of eye disease
Maternal family history of sensorineural hearing loss
NO Other affected relatives
NO Consanguinity
NO Parents with at least 3 miscarriages

Global mode of inheritance: Autosomal recessive inheritance

Medical history
Medical and developmental history: Nyctalopia before the age of 4 years.
Global age of onset: Childhood onset

Clinical symptoms and physical findings
EYE DEFECTS

NO Nystagmus
Rod-cone dystrophy
ERG
Photophobia
Mild
Nyctalopia Onset <4 years, relatively stable
Hypoplasia of the fovea
Borderline
Color vision test abnormality
Mild
EAR DEFECTS
NO Hearing impairment

Patient 18(U)
Sex: Male
Indication for referral: Global developmental delay and epileptic encephalopathy.

Family history and pedigree
Paternal ethnicity:
1. not known
Maternal ethnicity:
1. Portuguese
List health conditions found in family (describe the relationship with proband)
Father with bipolar disorder. Mother with chronic anemia unknown cause. Maternal side females have history of
hypothyroidism.
NO Consanguinity
NO Parents with at least 3 miscarriages
Global mode of inheritance: Autosomal recessive inheritance

Prenatal and perinatal history
Gestation at delivery (weeks) 31

Assisted reproduction:
NO Conception after fertility medication
NO In vitro fertilization
NO Gestational surrogacy
APGAR score (1 minute) 5
APGAR score (5 minutes) 7
PRENATAL DEVELOPMENT
NO Oligohydramnios
NO Polyhydramnios
DELIVERY
Premature birth
NEONATAL GROWTH PARAMETERS

Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
Abnormal birth length
NO Small birth length (<-2SD)
NO Large birth length (>+2SD)
Abnormal head circumference at birth
NO Congenital microcephaly (<-3SD)
NO Congenital macrocephaly (>+2SD)

Medical history
Medical and developmental history:
Preterm, non-reassuring fetal heart rate, premature delivery at 31 weeks. Neonatal jaundice, sepsis and GI bleeding. Global developmental delay with history of regression after 4 years of age: never walked. First word at age 2 yrs, about fifty words at age 4 yrs, regression in language after 4 years of age. Seizure onset at age 2 years, intially febrile. Severe GDD wheelchair bound, no words. MRI showed thin corpus callsoum and periventricular leukomalasia.
Global age of onset: Infantile onset

Clinical symptoms and physical findings
GROWTH PARAMETERS
Weight for age
NO Decreased body weight (<-2SD)
Stature for age
NO Short stature (<-2SD)
NO Tall stature (>+2SD)
Head circumference for age
NO Microcephaly (<-3SD)
NO Macrocephaly (>+2SD)
NO Hemihypertrophy
NO Obesity
CRANIOFACIAL
NO Craniosynostosis
NO Cleft upper lip
NO Cleft palate
NO Abnormal facial shape
EYE DEFECTS
NO Visual impairment
NO Abnormality of the cornea
NO Coloboma
NO Abnormality of the anterior chamber
NO Cataract
NO Abnormality of the retina
NO Abnormality of the optic nerve
NO Microphthalmia

NO Nystagmus
NO Strabismus
NO Hypotelorism
NO Hypertelorism
EAR DEFECTS
Deafness
NO Sensorineural
NO Conductive
NO Preauricular pit
NO Preauricular skin tag
NO Abnormality of the outer ear
NO Abnormality of the inner ear
CUTANEOUS
NO Hyperpigmentation of the skin
NO Hypopigmentation of the skin
NO Capillary hemangiomas
NO Vascular skin abnormality
CARDIOVASCULAR
NO Atrial septal defect
NO Ventricular septal defect
NO Complete atrioventricular canal defect
NO Coarctation of aorta
NO Tetralogy of Fallot
NO Cardiomyopathy
NO Arrhythmia
RESPIRATORY
NO Congenital diaphragmatic hernia
NO Abnormality of lung morphology
MUSCULOSKELETAL
NO Skeletal dysplasia
NO Increased susceptibility to fractures
NO Lower limb undergrowth
NO Upper limb undergrowth
Camptodactyly
NO Finger
NO Toe
Syndactyly
NO Finger
NO Toe
Polydactyly
NO Preaxial
NO Postaxial
Oligodactyly
NO Hands
NO Feet
NO Scoliosis

NO Abnormality of the vertebral column
NO Flexion contracture
NO Talipes equinovarus
GASTROINTESTINAL
NO Esophageal atresia
NO Tracheoesophageal fistula
NO Gastroschisis
NO Omphalocele
NO Aganglionic megacolon
NO Cholestasis
NO Elevated hepatic transaminases
NO Exocrine pancreatic insufficiency
NO Diabetes mellitus
GENITOURINARY
NO Renal cyst
NO Horseshoe kidney
NO Abnormality of the ureter
NO Abnormality of the urethra
NO Ambiguous genitalia
NO Hypospadias
NO Cryptorchidism
BEHAVIOR, COGNITION AND DEVELOPMENT
Global developmental delay
Intellectual disability
Severe
NO Attention deficit hyperactivity disorder
NO Autism
NO Behavioral abnormality
NEUROLOGICAL
NO Generalized hypotonia
Seizures
NO Ataxia
NO Dystonia
NO Chorea
Spasticity
NO Spinal dysraphism
NO Morphological abnormality of the central nervous system

Patient 21(G)
Age: 5 years
Sex: Female
Indication for referral: Infantile onset epileptic encephalopathy and global developmental delay.

Family history and pedigree
Paternal ethnicity:
1. Romanian
Maternal ethnicity:
1. Romanian
List health conditions found in family (describe the relationship with proband)
None.
NO Consanguinity
NO Parents with at least 3 miscarriages
Global mode of inheritance: Autosomal recessive inheritance

Prenatal and perinatal history
Gestation at delivery (weeks) 34
Assisted reproduction:
NO Conception after fertility medication
In vitro fertilization
NO Gestational surrogacy
PRENATAL DEVELOPMENT
NO Oligohydramnios
NO Polyhydramnios
DELIVERY
NO Premature birth
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
Abnormal birth length
NO Small birth length (<-2SD)
NO Large birth length (>+2SD)
Abnormal head circumference at birth
NO Congenital microcephaly (<-3SD)
NO Congenital macrocephaly (>+2SD)

Medical history
Medical and developmental history: First seizure at age 4.5 months. Hypotonic since birth. Global developmental delay(chronological age 24 months) walking at age 21 months, wide-based, no pincer grasp and no words yet. Developmental age around 9-12 months. Still ongoing seizures. MRI shows delayed myelination as well as thin corpus callosum and small hipocampi.

Global age of onset: Infantile onset

Clinical symptoms and physical findings
GROWTH PARAMETERS
Weight for age
NO Decreased body weight (<-2SD)
Stature for age
NO Short stature (<-2SD)
NO Tall stature (>+2SD)
Head circumference for age
NO Microcephaly (<-3SD)
NO Macrocephaly (>+2SD)
NO Hemihypertrophy
NO Obesity
CRANIOFACIAL
NO Craniosynostosis
NO Cleft upper lip
NO Cleft palate
NO Abnormal facial shape
EYE DEFECTS
NO Visual impairment
NO Abnormality of the cornea
NO Coloboma
NO Abnormality of the anterior chamber
NO Cataract
NO Abnormality of the retina
NO Abnormality of the optic nerve
NO Microphthalmia
NO Nystagmus
NO Strabismus
NO Hypotelorism
NO Hypertelorism
EAR DEFECTS
Deafness
NO Sensorineural
NO Conductive
NO Preauricular pit
NO Preauricular skin tag
NO Abnormality of the outer ear
NO Abnormality of the inner ear
CARDIOVASCULAR
NO Atrial septal defect
NO Ventricular septal defect
NO Complete atrioventricular canal defect
NO Coarctation of aorta
NO Tetralogy of Fallot

NO Cardiomyopathy
NO Arrhythmia
RESPIRATORY
NO Congenital diaphragmatic hernia
NO Abnormality of lung morphology
MUSCULOSKELETAL
NO Skeletal dysplasia
NO Increased susceptibility to fractures
NO Lower limb undergrowth
NO Upper limb undergrowth
Camptodactyly
NO Finger
NO Toe
Syndactyly
NO Finger
NO Toe
Polydactyly
NO Preaxial
NO Postaxial
Oligodactyly
NO Hands
NO Feet
NO Scoliosis
NO Abnormality of the vertebral column
NO Flexion contracture
NO Talipes equinovarus
GASTROINTESTINAL
NO Esophageal atresia
NO Tracheoesophageal fistula
NO Gastroschisis
NO Omphalocele
NO Aganglionic megacolon
NO Cholestasis
NO Elevated hepatic transaminases
NO Exocrine pancreatic insufficiency
NO Diabetes mellitus
GENITOURINARY
NO Renal cyst
NO Horseshoe kidney
NO Abnormality of the ureter
NO Abnormality of the urethra
NO Ambiguous genitalia
BEHAVIOR, COGNITION AND DEVELOPMENT
Global developmental delay
Delayed fine motor development
Delayed gross motor development

Delayed speech and language development
NO Attention deficit hyperactivity disorder
NO Autism
NO Behavioral abnormality
NEUROLOGICAL
Generalized hypotonia
Seizures
Ataxia
NO Dystonia
NO Chorea
NO Spasticity
NO Spinal dysraphism
NO Morphological abnormality of the central nervous system

Patient 30(R)
Age: 12 years
Sex: Male
Indication for referral: Global developmental delay, epileptic encephalopathy
Family history and pedigree
Paternal ethnicity:
1. Phillipine
Maternal ethnicity:
1. Phillipine
List health conditions found in family (describe the relationship with proband)
None
NO Consanguinity
NO Parents with at least 3 miscarriages
Global mode of inheritance: Autosomal recessive inheritance

Prenatal and perinatal history
Assisted reproduction:
NO Conception after fertility medication
NO In vitro fertilization
NO Gestational surrogacy
PRENATAL DEVELOPMENT
NO Oligohydramnios
NO Polyhydramnios
DELIVERY
NO Premature birth
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
Abnormal birth length
NO Small birth length (<-2SD)

NO Large birth length (>+2SD)
Abnormal head circumference at birth
NO Congenital microcephaly (<-3SD)
NO Congenital macrocephaly (>+2SD)
Notes:
Pregnancy complicated with maternal gestational diabetes mellitus.

Medical history
Medical and developmental history: Infantile spasms at age 4 months. Intractable epilepsy followed. Various anti-epileptic medications, ketogenic diet non-responsive. MRI no focal lesion. Epilepsy surgery due to focal EEG features. Epilepsy surgery. Pathology reported focal cortical dysplasia type 1b in brain specimen.

Global age of onset: Infantile onset

Clinical symptoms and physical findings
GROWTH PARAMETERS
Weight for age
NO Decreased body weight (<-2SD)
Stature for age
NO Short stature (<-2SD)
NO Tall stature (>+2SD)
Head circumference for age
Microcephaly (<-3SD)
NO Macrocephaly (>+2SD)
NO Obesity
CRANIOFACIAL
NO Craniosynostosis
NO Cleft upper lip
NO Cleft palate
NO Abnormal facial shape
EYE DEFECTS
NO Visual impairment
NO Abnormality of the cornea
NO Coloboma
NO Abnormality of the anterior chamber
NO Cataract
NO Abnormality of the retina
NO Abnormality of the optic nerve
NO Microphthalmia
NO Nystagmus
NO Strabismus
NO Hypotelorism
NO Hypertelorism
EAR DEFECTS
Deafness

NO Sensorineural
NO Conductive
NO Preauricular pit
NO Preauricular skin tag
NO Abnormality of the outer ear
NO Abnormality of the inner ear
CUTANEOUS
NO Hyperpigmentation of the skin
NO Hypopigmentation of the skin
NO Capillary hemangiomas
NO Vascular skin abnormality
CARDIOVASCULAR
NO Atrial septal defect
NO Ventricular septal defect
NO Complete atrioventricular canal defect
NO Coarctation of aorta
NO Tetralogy of Fallot
NO Cardiomyopathy
NO Arrhythmia
RESPIRATORY
NO Congenital diaphragmatic hernia
NO Abnormality of lung morphology
MUSCULOSKELETAL
NO Skeletal dysplasia
NO Increased susceptibility to fractures
NO Lower limb undergrowth
NO Upper limb undergrowth
Camptodactyly
NO Finger
NO Toe
Syndactyly
NO Finger
NO Toe
Polydactyly
NO Preaxial
NO Postaxial
Oligodactyly
NO Hands
NO Feet
NO Scoliosis
NO Abnormality of the vertebral column
NO Flexion contracture
NO Talipes equinovarus
GASTROINTESTINAL
NO Esophageal atresia
NO Tracheoesophageal fistula

NO Gastroschisis
NO Omphalocele
NO Aganglionic megacolon
NO Cholestasis
NO Elevated hepatic transaminases
NO Exocrine pancreatic insufficiency
NO Diabetes mellitus
GENITOURINARY
NO Renal cyst
NO Horseshoe kidney
NO Abnormality of the ureter
NO Abnormality of the urethra
NO Ambiguous genitalia
NO Hypospadias
NO Cryptorchidism
BEHAVIOR, COGNITION AND DEVELOPMENT
Global developmental delay
NO Delayed fine motor development
NO Delayed gross motor development
NO Delayed speech and language development
NO Specific learning disability
Intellectual disability
NO Severe
NO Attention deficit hyperactivity disorder
NO Autism
NO Behavioral abnormality
NEUROLOGICAL
NO Generalized hypotonia
Ataxia
NO Dystonia
NO Chorea
NO Spasticity
NO Spinal dysraphism
NO Morphological abnormality of the central nervous system
Generalized tonic-clonic seizures
Absence seizures
Cortical dysplasia
Atonic seizures

Patient 39(P)
Age: 12 years
Sex: Male
Indication for referral: Epileptic encephalopathy
Family history and pedigree
Paternal ethnicity:
1. Caucasian
Maternal ethnicity:
1. Caucasian
List health conditions found in family (describe the relationship with proband)
Mother with ADD and celiac disease.
paternal uncle's son diagnosis of Asperger's syndrome and recent diagnosis of epilepsy
NO Consanguinity
NO Parents with at least 3 miscarriages
Global mode of inheritance: Autosomal recessive inheritance

Prenatal and perinatal history
Assisted reproduction:
NO Conception after fertility medication
NO In vitro fertilization
NO Gestational surrogacy
PRENATAL DEVELOPMENT
NO Oligohydramnios
NO Polyhydramnios
DELIVERY
NO Premature birth
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
Abnormal birth length
NO Small birth length (<-2SD)
NO Large birth length (>+2SD)
Abnormal head circumference at birth
NO Congenital microcephaly (<-3SD)
NO Congenital macrocephaly (>+2SD)

Medical history
Medical and developmental history:
First seizure at age 2.5 years. Normal development until grade 1. Due to intractable seizures, in grade 1 requiring an
Individualized Educational Plan.
Ketogenic diet.
Global age of onset:
Infantile onset
Clinical symptoms and physical findings

GROWTH PARAMETERS
Weight for age
NO Decreased body weight (<-2SD)
Stature for age
NO Short stature (<-2SD)
NO Tall stature (>+2SD)
Head circumference for age
NO Microcephaly (<-3SD)
NO Macrocephaly (>+2SD)
NO Obesity
CRANIOFACIAL
NO Craniosynostosis
NO Cleft upper lip
NO Cleft palate
NO Abnormal facial shape
EYE DEFECTS
NO Visual impairment
NO Abnormality of the cornea
NO Coloboma
NO Abnormality of the anterior chamber
NO Cataract
NO Abnormality of the retina
NO Abnormality of the optic nerve
NO Microphthalmia
NO Nystagmus
NO Strabismus
NO Hypotelorism
NO Hypertelorism
EAR DEFECTS
Deafness
NO Sensorineural
NO Conductive
NO Preauricular pit
NO Preauricular skin tag
NO Abnormality of the outer ear
NO Abnormality of the inner ear
CUTANEOUS
NO Hyperpigmentation of the skin
NO Hypopigmentation of the skin
NO Capillary hemangiomas
NO Vascular skin abnormality
CARDIOVASCULAR
NO Atrial septal defect
NO Ventricular septal defect
NO Complete atrioventricular canal defect
NO Coarctation of aorta

NO Tetralogy of Fallot
NO Cardiomyopathy
NO Arrhythmia
RESPIRATORY
NO Congenital diaphragmatic hernia
NO Abnormality of lung morphology
MUSCULOSKELETAL
NO Skeletal dysplasia
NO Increased susceptibility to fractures
NO Lower limb undergrowth
NO Upper limb undergrowth
Camptodactyly
NO Finger
NO Toe
Syndactyly
NO Finger
NO Toe
Polydactyly
NO Preaxial
NO Postaxial
Oligodactyly
NO Hands
NO Feet
NO Scoliosis
NO Abnormality of the vertebral column
NO Flexion contracture
NO Talipes equinovarus
GASTROINTESTINAL
NO Esophageal atresia
NO Tracheoesophageal fistula
NO Gastroschisis
NO Omphalocele
NO Aganglionic megacolon
NO Cholestasis
NO Elevated hepatic transaminases
NO Exocrine pancreatic insufficiency
NO Diabetes mellitus
GENITOURINARY
NO Renal cyst
NO Horseshoe kidney
NO Abnormality of the ureter
NO Abnormality of the urethra
NO Ambiguous genitalia
NO Hypospadias
NO Cryptorchidism
Nephrolithiasis

BEHAVIOR, COGNITION AND DEVELOPMENT
NO Global developmental delay
NO Delayed fine motor development
NO Delayed gross motor development
NO Delayed speech and language development
NO Specific learning disability
Intellectual disability
NO Mild
NO Moderate
NO Severe
NO Attention deficit hyperactivity disorder
NO Autism
NO Behavioral abnormality
Dysarthria
NEUROLOGICAL
NO Generalized hypotonia
Seizures
NO Ataxia
NO Dystonia
NO Chorea
NO Spasticity
NO Spinal dysraphism
NO Morphological abnormality of the central nervous system
Generalized tonic-clonic seizures
Absence seizures
Generalized myoclonic seizures
Focal tonic seizures
Diagnosis
Additional comments: Epileptic encephalopathy. Normal cranial MRI.

Patient 42(O)
Age: 12 years
Sex: Female
Indication for referral: Early onset retinal dystrophy

Family history and pedigree
Paternal ethnicity:
1. Egyptian
Maternal ethnicity:
1. Egyptian
List health conditions found in family (describe the relationship with proband)
Mother's parents are distantly related.
Distant maternal relative with unspecified eye problem (see pedigree).
NO Other affected relatives
NO Consanguinity

NO Parents with at least 3 miscarriages
Global mode of inheritance: Autosomal recessive inheritance

Medical history
Medical and developmental history: Born with very small congenital cataracts. Later re-presented to Ophthamology clinic with nyctalopia. Allergies - environmental, food, medication: NKDA

Clinical symptoms and physical findings
EYE DEFECTS
Abnormal electroretinogram
Severe rod and cone dysfunction
Congenital cataract
Bilateral
Very small anterior pyramidal cataracts
Retinal dystrophy
Nyctalopia Age 5
EAR DEFECTS
NO Hearing impairment
RESPIRATORY
Asthma
Mild
METABOLISM/HOMEOSTASIS
Cystoid macular edema
Persistent on serial OCT

Patient 56(N)
Age: 11 years
Sex: Female
Indication for referral: Cerebral arteriovenous malformation
Family history and pedigree
Paternal ethnicity:
1. French Canadian
Maternal ethnicity:
1. Chinese
List health conditions found in family (describe the relationship with proband)
Paternal family history of cerebral aneurysm
NO Consanguinity
NO Parents with at least 3 miscarriages

Prenatal and perinatal history
NO Multiple gestation
Gestation at delivery (weeks) 37
Maternal age at EDD (years) 35
Paternal age at EDD (years) 36

DELIVERY
NO Abnormal delivery (Non-NSVD)
NO Premature birth
Notes: Reportedly unremarkable pregnancy and delivery.

Medical history
Allergies - environmental, food, medication:
NKDA
Measurements
Age: 17y 9m
Weight: 54.3 kg 43rd pctl (-0.18SD)
Height: 164.5 cm 59th pctl (+0.22SD)
BMI: 20.07 35th pctl (-0.4SD)
Head circumference: 56.5 cm 91st pctl (+1.36SD)
Age: 18y 2m
Weight: 55.8 kg 49th pctl (-0.03SD)
Height: 164.0 cm 55th pctl (+0.14SD)
BMI: 20.75 43rd pctl (-0.18SD)

Clinical symptoms and physical findings
CARDIOVASCULAR
Aortic dilatation
Borderline
Cardiac MRI suggestive of mild dilated aortic root, ascending aorta, aortic arch, and thoracic aorta. However,
this was discordant with results from echo showing aortic root and ascending aorta z-scores within normal limits. Ultimately not felt to have significant issue. Cerebral arteriovenous malformation Discovered on brain MRI after presenting with headaches, resected from right temporal region? small associated aneurysm.
NO Arterial tortuosity
MUSCULOSKELETAL
Joint hypermobility
Borderline
Beighton score 5
Fractures of the long bones
Left femur and right humerus fracture @ age 3 secondary to trauma.
Patellar subluxation
Bilateral knee/patellar instability for which she wears brace.
BEHAVIOR, COGNITION AND DEVELOPMENT
NO Cognitive impairment
NEUROLOGICAL
Headache

Patient 57(T)
Age: 9 years
Sex: Female
Indication for referral: Ehlers-Danlos syndrome, hypermobility type
Family history and pedigree
Paternal ethnicity:
1. Ashkenazi Jewish
2. Polish
Maternal ethnicity:
1. Ashkenazi Jewish
2. Polish
List health conditions found in family (describe the relationship with proband)
Hypermobility, joint dislocation, and IBS (see pedigree)
Other affected relatives
NO Consanguinity
NO Parents with at least 3 miscarriages

Prenatal and perinatal history
NO Multiple gestation
Gestation at delivery (weeks) Term birth
Maternal age at EDD (years) 27
Paternal age at EDD (years) 30
PREGNANCY HISTORY
Hyperemesis gravidarum (excessive vomiting)
DELIVERY
NO Abnormal delivery (Non-NSVD)
NO Premature birth
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
Notes:
Birth weight 7 lbs 6 oz.

Medical history
Allergies - environmental, food, medication:
1. Adhesives / tapes (rash)
Measurements
Age: 16y 2m
Weight: 47.4 kg 16th pctl (-1.0SD)
Height: 156.5 cm 18th pctl (-0.91SD)
BMI: 19.35 30th pctl (-0.52SD)
Head circumference: 52.0 cm 5th pctl (-1.64SD)

Clinical symptoms and physical findings
EAR DEFECTS

NO Hearing impairment
Audiology testing within normal limits.
CARDIOVASCULAR
Patent foramen ovale
On echo at age 15 1/2 years.
Postural hypotension with compensatory tachycardia
Tilt table testing pending for query POTS.
RESPIRATORY
Asthma
Shoulder dislocation
Multidirectional instability with frequent subluxation and occasional dislocation. Went to OR for left shoulder.
Stridor Likely secondary to conversion disorder. ? paradoxical vocal cord movements.
Sleep apnea
CPAP at night.
MUSCULOSKELETAL
Joint dislocation
Knee, finger, bilateral shoulder.
Joint hypermobility
Beighton score 7.
NO EMG abnormality
GASTROINTESTINAL
Abnormality of the gastrointestinal tract
Irritable bowel syndrome. Upper and lower scopes showed only non-specific changes.
BEHAVIOR, COGNITION AND DEVELOPMENT
Specific learning disability
Reportedly with above average IQ but specific learning disability in processing.
CONSTITUTIONAL SYMPTOM
Chronic pain
CONNECTIVE TISSUE
Atypical scarring of skin
NO Mastocytosis
Clinically suspected of having mast cell activation syndrome with history of hives/rashes NYD and sensitive skin.
No increase in mucosal mast cells on GI biopsies.
VOICE
Weak voice
Likely secondary to conversion disorder.
IMMUNE SYSTEM
NO Celiac disease

Patient 67(M)
Age: 13 years
Sex: Male
Indication for referral: Retinitis pigmentosa
Family history and pedigree
List health conditions found in family (describe the relationship with proband)
Mother has eye phenotype, so there was a question about X-linked RP.
NO Consanguinity
NO Parents with at least 3 miscarriages

Prenatal and perinatal history
Gestation at delivery (weeks) Term birth
DELIVERY
NO Premature birth
Notes:
Reportedly unremarkable pregnancy.

Medical history
Medical and developmental history: Glasses since age 2. Presented at ~ age 4 after noted to have chin down position and to be walking into things. Exam and investigations consistent with retinitis pigmentosa. Otherwise healthy.

Clinical symptoms and physical findings
EYE DEFECTS
Visual impairment
NO Nystagmus
Rod-cone dystrophy
Abnormality of color vision
Nyctalopia
Depigmented fundus
BEHAVIOR, COGNITION AND DEVELOPMENT
NO Cognitive impairment

Patient 68(J)
Age: 17 years
Sex: Female
Indication for referral: Mitochondrial disorder (query Pearson syndrome)

Family history and pedigree
Paternal ethnicity:
1. East Indian
Maternal ethnicity:
1. East Indian
List health conditions found in family (describe the relationship with proband)
Identical twin has similar multi-system health concerns.

Strong maternal family history of diabetes.
Other affected relatives
NO Consanguinity
NO Parents with at least 3 miscarriages

Prenatal and perinatal history
Multiple gestation
Gestation at delivery (weeks) 33
Maternal age at EDD (years) 30
Paternal age at EDD (years) 35
PREGNANCY HISTORY
Maternal diabetes
NO Maternal teratogenic exposure
DELIVERY
Abnormal delivery (Non-NSVD)
Caesarian section
Premature birth
Nuchal cord
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
Notes:
BW 1.59 kg

Measurements
Age: 7y 4m
Weight: 15.5 kg 0th pctl (-2.92SD)
Height: 107.2 cm 0th pctl (-2.85SD)
BMI: 13.49 8th pctl (-1.38SD)
Age: 11y 5m
Weight: 26.4 kg 2nd pctl (-1.99SD)
Height: 119.0 cm 0th pctl (-4.27SD)
BMI: 18.64 67th pctl (+0.45SD)
Age: 14y 9m
Weight: 31.0 kg 0th pctl (-3.62SD)
Height: 122.0 cm 0th pctl (-5.71SD)
BMI: 20.83 60th pctl (+0.26SD)
Head circumference: 54.0 cm 49th pctl (-0.03SD)
Age: 16y 7m
Weight: 34.3 kg 0th pctl (-3.7SD)
Height: 125.9 cm 0th pctl (-5.47SD)
BMI: 21.64 59th pctl (+0.24SD)

Clinical symptoms and physical findings
GROWTH PARAMETERS

Weight for age
Decreased body weight (<-2SD)
Stature for age
Short stature (<-2SD)
Delayed puberty
EYE DEFECTS
Ptosis
Retinal dystrophy
Corneal dystrophy
Corneal disease with possible endothelial dystrophy, status post right corneal transplant
CARDIOVASCULAR
Sinus tachycardia
MUSCULOSKELETAL
Scoliosis
Mild
Osteopenia
Pes planus
Bilateral
Hallux valgus
Bilateral
Delayed skeletal maturation
Vertebral compression fractures
Myopathy
Mild
Proximal
Ragged-red muscle fibers
Abnormal mitochondria in muscle tissue
Also in renal tissue
GENITOURINARY
Stage 5 chronic kidney disease
Renal transplant
BEHAVIOR, COGNITION AND DEVELOPMENT
NO Global developmental delay
Specific learning disability
Some learning difficulties, with IEP in school. E.g., in gr. 8, at gr. 5-6 level.
NEUROLOGICAL
Ataxia
Morphological abnormality of the central nervous system
Brain MRI showed mild prominence of lateral and third ventricles, as well as mild cerebellar volume loss. No
abnormal lactate peak on MRS.
Abnormality of the peripheral nervous system
Guillian-Barre syndrome
BLOOD AND BLOOD-FORMING TISSUES
Aplastic anemia
Severe

17 pRBC transfusions
METABOLISM/HOMEOSTASIS
Renal tubular acidosis
Fanconi renal syndrome in setting of sepsis
Increased serum lactate
Decreased activity of mitochondrial complex II
DIGESTIVE SYSTEM
Abnormality of pancreas morphology
Echogenic, with fatty infiltration changes, on ultrasound
Abnormality of exocrine pancreas physiology
Fecal fat excretion and fecal elastase abnormal
ENDOCRINE SYSTEM
Type I diabetes mellitus
IMMUNE SYSTEM
Recurrent lower respiratory tract infections

Patient 71(L)
Age: 9 years
Sex: Male
Indication for referral: Hypermobility, recurrent dislocation, query connective tissue disorder
Family history and pedigree
Paternal ethnicity:
1. Argentinian
Maternal ethnicity:
1. Italian
List health conditions found in family (describe the relationship with proband)
Mother and maternal relatives with joint hypermobility, but not necessarily to same degree as the proband.
NO Other affected relatives
NO Consanguinity
NO Parents with at least 3 miscarriages

Prenatal and perinatal history
NO Multiple gestation
Gestation at delivery (weeks) 38
Assisted reproduction:
NO Conception after fertility medication
Maternal age at EDD (years) 31
Paternal age at EDD (years) 32
PRENATAL DEVELOPMENT
Oligohydramnios
NO Polyhydramnios
NO Decreased fetal movement
DELIVERY

NO Premature birth
Primary Caesarian section
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
Small for gestational age (<-2SD)
Notes: Unremarkable pregnancy until oligohydramnios noted in third trimester. Born via emergency C-section at 38 weeks for fetal distress. Weight 2.38 kg, labeled as SGA and kept in NICU for 10 days. No resuscitation required at birth.

Medical history
Medical and developmental history: Noted at ~5 months to be hyperextensible. Came to Canada from Argentina in grade 1. Hypermobility has improved with age.
Measurements
Age: 15y 11m
Weight: 60.0 kg 44th pctl (-0.15SD)
Height: 164.0 cm 13th pctl (-1.11SD)
BMI: 22.31 74th pctl (+0.66SD)
Head circumference: 58.0 cm 93rd pctl (+1.47SD)
Age: 16y 6m
Weight: 62.5 kg 45th pctl (-0.11SD)
Height: 175.0 cm 54th pctl (+0.1SD)
BMI: 20.41 44th pctl (-0.16SD)
Head circumference: 57.5 cm 84th pctl (+1.01SD)
Arm span: 178.5 cm = Height + 3.5 cm
Age: 18y 4m
Weight: 66.9 kg 44th pctl (-0.16SD)
Height: 176.5 cm 51st pctl (+0.02SD)
BMI: 21.48 44th pctl (-0.15SD)

Clinical symptoms and physical findings
CRANIOFACIAL
High palate
Malar flattening
Mandibular prognathia
Triangular face
Midface retrusion
EYE DEFECTS
Astigmatism
Left > right
Hypermetropia
NO Abnormal eye morphology
Reportedly normal dilated exam
CARDIOVASCULAR
NO Arrhythmia
Normal Holter
NO Orthostatic hypotension

No concern regarding POTS
NO Abnormal heart morphology
Reportedly normal echo
MUSCULOSKELETAL
Joint hypermobility
Beighton score initially 9/9, but as he aged he became less hypermobile.
Pes cavus
Right
Pes planus
Left
Hammertoe
Right
Bilateral talipes equinovarus
Congenital onset
Clinodactyly of the 5th finger
Recurrent patellar dislocation
Right
Soccer player
Ankle contracture
Mild
Bilateral
Scheuermann-like vertebral changes
On spine MRI and x-ray. Mild thoracolumbar left scoliosis.
Infantile muscular hypotonia
Few details. Apparently resolved, as exam records from Canada do not report hypotonia.
Long fingers
NO EMG abnormality
Initial EMG showed non-specific pattern consistent with increased small polyphasic low voltage motor units not
really in keeping with a muscle process. Repeat EMG was normal.
GENITOURINARY
Bilateral cryptorchidism
Total of 3 operations in Argentina
NEUROLOGICAL
NO Morphological abnormality of the central nervous system
Brain MRI showing only non-specific tiny white matter signal intensity in right frontal lobe.
NO Motor delay
BLOOD AND BLOOD-FORMING TISSUES
Bruising susceptibility
CONNECTIVE TISSUE
Inguinal hernia
Contracture of the proximal interphalangeal joint of the 5th finger

Patient 76(Q)
Age: 11 years
Sex: Female
Indication for referral: Ehlers-Danlos syndrome (EDS) type III, hypermobility type

Family history and pedigree
Paternal ethnicity:
1. Anglo-Celtic Australian
Maternal ethnicity:
1. Anglo-Celtic Australian
Other affected relatives
NO Consanguinity
NO Parents with at least 3 miscarriages

Prenatal and perinatal history
NO Multiple gestation
Gestation at delivery (weeks) 38
PREGNANCY HISTORY
Maternal diabetes
DELIVERY
NO Abnormal delivery (Non-NSVD)
NO Premature birth
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
Notes: Uncomplicated pregnancy and term vaginal delivery in New Zealand, BW 7 lbs 2 oz.

Medical history
Allergies - environmental, food, medication:
1. Mango (anaphylaxis)
Measurements
Age: 18y 1m
Weight: 61.0 kg 71st pctl (+0.56SD)
Height: 158.5 cm 24th pctl (-0.69SD)
BMI: 24.28 81st pctl (+0.87SD)
Head circumference: 54.5 cm 48th pctl (-0.04SD)

Clinical symptoms and physical findings
CARDIOVASCULAR
Abnormal heart morphology
Chiari network in right atrium on echocardiogram.
Vasovagal syncope
Positive tilt table test response for vasovagal presyncope.
RESPIRATORY

Shoulder subluxation
MUSCULOSKELETAL
Joint hypermobility
Beighton score >=7.
Pes planus
Fractured hand bones
Wrist fracture.
Multiple joint dislocation
NO Osteopenia
Normal BMD study.
NEUROLOGICAL
Migraine
Ocular migraine.
Hydromyelia
NO Motor delay
BLOOD AND BLOOD-FORMING TISSUES
NO Abnormal bleeding
DIGESTIVE SYSTEM
Constipation
Also with early satiety. Seen by GI in the past.
CONSTITUTIONAL SYMPTOM
Chronic pain
Especially in lower legs, since age 10.
CONNECTIVE TISSUE
Mastocytosis
Also with intermittent lips and tongue paresthesias and skin flushing.
NO Atypical scarring of skin
ENDOCRINE SYSTEM
Abnormality of the thyroid gland
Right posterior mid-lobe hypoechoic nodules on ultrasound, with normal thyroid function.
IMMUNE SYSTEM
Allergy
NO Celiac disease

Patient 78(V)
Age: 6 years
Sex: Female
Indication for referral: Ehlers-Danlos syndrome, hypermobility type
Family history and pedigree
Paternal ethnicity:
1. English
Maternal ethnicity:
1. English
2. French

List health conditions found in family (describe the relationship with proband)
Mother, maternal aunt, and maternal grandfather with hypermobility.
Brother and father with EDS type 3 diagnosis.
Paternal uncle with unspecified psychiatric issues, and paternal female first-cousin with seizures and mild autism
spectrum disorder.
Other affected relatives
NO Consanguinity
NO Parents with at least 3 miscarriages

Prenatal and perinatal history
NO Multiple gestation
Gestation at delivery (weeks) 37
Maternal age at EDD (years) 35
Paternal age at EDD (years) 34
APGAR score (1 minute) 9
APGAR score (5 minutes) 9
PREGNANCY HISTORY
NO Maternal hypertension
NO Maternal diabetes
NO Maternal teratogenic exposure
PRENATAL DEVELOPMENT
NO Intrauterine growth retardation
NO Oligohydramnios
NO Polyhydramnios
DELIVERY
NO Abnormal delivery (Non-NSVD)
NO Premature birth
Precipitous labour
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
Notes: Reportedly uncomplicated pregnancy with normal ultrasounds. BW 6 lbs 11 oz.

Medical history
Medical and developmental history: Followed in Hamilton in early years. Initially seen by Genetics there in because of GDD and FTT. Allergies - environmental, food, medication: NKDA

Measurements
Age: 3y 11m
Weight: 12.5 kg 3rd pctl (-1.85SD)
Height: 90.0 cm 0th pctl (-2.9SD)
BMI: 15.43 55th pctl (+0.12SD)
Head circumference: 47.0 cm 1st pctl (-2.4SD)

Outer canthal distance: 7.5 cm 38th pctl (-0.29SD)
Inner canthal distance: 3.0 cm 82nd pctl (+0.91SD)
Interpupilary distance: 5.0 cm 60th pctl (+0.25SD)
Arm span: 90.0 cm = Height

Clinical symptoms and physical findings
GROWTH PARAMETERS
Stature for age
Short stature (<-2SD)
Failure to thrive in infancy
EYE DEFECTS
Hypermetropia
Borderline
Glasses. Also with "delayed visual myelination" - resolved
EAR DEFECTS
NO Hearing impairment
CUTANEOUS
Hyperextensible skin
Abnormality of the toenails
Short, somewhat hypoplastic, and easy breakability
NO Eczema
CARDIOVASCULAR
Urticaria
Intermittent, not always clear precipitant
NO Abnormality of the cardiovascular system
Normal echocardiogram, with normal aortic measurements and no mitral valve abnormalities
RESPIRATORY
Cough
Chronic
MUSCULOSKELETAL
Talipes equinovarus
Treated with therapy and SMO. Other reports say bilateral calcaneovalgus deformity
Joint dislocation
Joint hypermobility
Beighton score 6/9
Infantile muscular hypotonia
BEHAVIOR, COGNITION AND DEVELOPMENT
Global developmental delay
Early on, especially fine and gross motor. At time of last follow-up, gross motor normal.
NEUROLOGICAL
Seizures
Clinically suspected per parents on the basis of occasional stiffening spells beginning at 4 months of age, but
two normal EEGs, never treated with anti-epileptic medication, and resolved in childhood.

NO Morphological abnormality of the central nervous system
Essentially normal brain MRI in Hamilton, with exception of possibly delayed myelination.
NO EEG abnormality
BLOOD AND BLOOD-FORMING TISSUES
Bruising susceptibility
DIGESTIVE SYSTEM
Constipation
CONSTITUTIONAL SYMPTOM
Chronic pain
Multisystem. Seen by Psychiatry because of disruptive behaviours that were attributed to pain
IMMUNE SYSTEM
Allergy
Multiple foods

Patient 79(K)
Age: 11 years
Sex: Male
Indication for referral: query Ehlers-Danlos syndrome, hypermobility type, possible ADHD

Family history and pedigree
List health conditions found in family (describe the relationship with proband)
mother - hypermobility
NO Consanguinity

Measurements
Age: 8y 6m
Weight: 36.2 kg 97th pctl (+1.84SD)
Height: 147.7 cm 100th pctl (+2.99SD)
BMI: 16.59 67th pctl (+0.43SD)

Clinical symptoms and physical findings
GROWTH PARAMETERS
Stature for age
Tall stature (>+2SD)
CUTANEOUS
NO Poor wound healing
CARDIOVASCULAR
Tachycardia
occasional when running
NO Abnormal echocardiogram
normal aortic root and arch, no MVP or MR, no ASD/VSD/PDA/LVOTO/RVOTO good biventricular

function, normal chamber size
RESPIRATORY
Pectus excavatum "slight"
Apnea
NO Cyanosis
MUSCULOSKELETAL
NO Increased susceptibility to fractures
NO Scoliosis
Joint hyperflexibility
particularly wrists and elbows
NO Joint dislocation
GENITOURINARY
NO Enuresis
BEHAVIOR, COGNITION AND DEVELOPMENT
NO Loss of consciousness
NEUROLOGICAL
Poor fine motor coordination
difficulty with buttons and zippers
BLOOD AND BLOOD-FORMING TISSUES
NO Abnormal bleeding
DIGESTIVE SYSTEM
Encopresis
NO Hepatomegaly
CONSTITUTIONAL SYMPTOM
Pain
daily neck, chest, knees, feet pain worse in morning, characterized as stiffness
IMMUNE SYSTEM
Allergy
Abnormality of the tonsils
hypertrophy
Abnormality of nasopharyngeal adenoids
Hypertrophy

Patient 81(I)
Age: 11 years
Sex: Female
Indication for referral: chronic abdominal pain, dysmotility, hyperextesibility, neurogenic bladder dysmotility, superior mesenteric artery syndrome

Family history and pedigree
Paternal ethnicity:
1. Scottish
Maternal ethnicity:
1. British
List health conditions found in family (describe the relationship with proband)

mother myopia and talipes

Clinical symptoms and physical findings
GROWTH PARAMETERS
NO Disproportionate tall stature
CRANIOFACIAL
NO Abnormal facial shape
Malar flattening
Mild
NO Abnormality of the palate
EYE DEFECTS
NO Myopia
CUTANEOUS
NO Hyperextensible skin
NO Molluscoid pseudotumors
NO Poor wound healing
NO Abnormal elasticity of skin
CARDIOVASCULAR
NO Tachycardia
RESPIRATORY
Pectus excavatum
Mild
MUSCULOSKELETAL
NO Scoliosis
Arachnodactyly
Joint hypermobility
Beighton score 2/9 (elbow)
Joint laxity
NO Syndactyly
GASTROINTESTINAL
Gastrointestinal dysmotility
GENITOURINARY
Neurogenic bladder
Urinary retention
Ovarian cyst
BLOOD AND BLOOD-FORMING TISSUES
NO Bruising susceptibility
DIGESTIVE SYSTEM
Pancreatitis
Nausea and vomiting
Abdominal distention
CONSTITUTIONAL SYMPTOM
Abdominal pain
Chronic
Arthralgia
shoulders, hips, knees

CONNECTIVE TISSUE
NO Scarring

Patient 91(E)
Age: 11yrs
Sex: Female
Indication for referral: Global developmental delay, autism spectrum disorder, epilepsy, atrial septal defect

Family history and pedigree
Paternal ethnicity:
1. Nova scotia
Maternal ethnicity:
1. Welsh
2. English
3. Irish
List health conditions found in family (describe the relationship with proband)
(1) Autism spectrum disorder: brother
(2) Depression/anxiety: mother (both), brother (anxiety)
(3) Learning difficulties: maternal female first cousins x2
(4) Query ADHD: father
(5) Query bipolar disorder: maternal female first cousin
(6) Atrial septal defect: maternal first cousin once removed
(7) Aortic aneurysm: maternal great uncle
(8) Recurrent miscarriages: maternal grandmother (up to 5, all male, including one set of twins)
NO Consanguinity
NO Parents with at least 3 miscarriages

Prenatal and perinatal history
NO Multiple gestation
Gestation at delivery (weeks) 36
Maternal age at EDD (years) 28
Paternal age at EDD (years) 34
DELIVERY
Abnormal delivery (Non-NSVD)
Premature birth
Secondary Caesarian section
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
EAR
Vertigo
Notes: BW 5 lbs 12.5 oz (2.62 kg)

Medical history
Allergies - environmental, food, medication:
1. amoxicillin (rash)

Measurements
Age: 5y 5m
Weight: 19.1 kg 51st pctl (+0.04SD)
Height: 101.8 cm 2nd pctl (-2.04SD)
BMI: 18.43 96th pctl (+1.76SD)
Head circumference: 50.4 cm 36th pctl (-0.35SD)
Outer canthal distance: 8.0 cm 61st pctl (+0.27SD)
Inner canthal distance: 3.3 cm 97th pctl (+1.95SD)
Interpupilary distance: 5.4 cm 86th pctl (+1.07SD)
Left ear length: 5.5 cm 64th pctl (+0.35SD)
Right ear length: 5.2 cm 29th pctl (-0.55SD)
Age: 6y 11m
Weight: 22.4 kg 52nd pctl (+0.05SD)
Height: 111.2 cm 4th pctl (-1.7SD)
BMI: 18.12 92nd pctl (+1.38SD)
Head circumference: 51.5 cm 48th pctl (-0.04SD)
Age: 10y 3m
Weight: 54.4 kg 98th pctl (+2.14SD)
Height: 139.2 cm 42nd pctl (-0.2SD)
BMI: 28.08 100th pctl (+2.9SD)

Clinical symptoms and physical findings
GROWTH PARAMETERS
Weight for age
Increased body weight (>+2SD)
Obesity
CRANIOFACIAL
Abnormal facial shape
looks different from parents, coarse features, hooded eyelids, short upslanting palpebral fissures
High palate
One note reported bifid uvula, but other notes reported normal uvula.
Nasolacrimal duct obstruction
Left
CUTANEOUS
Single transverse palmar crease
Left
Abnormality of the hair
Streak of blond hair in right anterior parietal area
CARDIOVASCULAR
Secundum atrial septal defect
Noted at birth, large, surgically repaired. Also with spontaneously closed VSD.

MUSCULOSKELETAL
Clinodactyly of the 5th finger
Bilateral
BEHAVIOR, COGNITION AND DEVELOPMENT
Global developmental delay
MRI brain: Few tiny nonspecific periventricular white matter signals are seen, otherwise the brain parenchyma
appears unremarkable.
Attention deficit hyperactivity disorder
Developed tics on Biphentin
NEUROLOGICAL
Seizures
First diagnosed. Multiple types: GTC, absence, partial complex
Autistic behavior
Neuropsychology assessment

Patient 92(S)
Age: 9 years
Sex: Female
Indication for referral: Evaluation for Ehlers-Danlos, overweight, sleep apnea, hyperlipidemia, mild limb length asymmetry, hyperextensible joints, pain, easy bruising, delayed wound healing with wide hypertrophic distended cigarette paper thin scars, sensitive to all allergens, develops hives, very sensitive to insect bites,

Family history and pedigree
Paternal ethnicity:
1. English
Maternal ethnicity:
1. Italians
List health conditions found in family (describe the relationship with proband)
Mat fam hx of obesity, hypertension, hyperlipidemia, cardiac events
Other affected relatives
NO Consanguinity
NO Parents with at least 3 miscarriages
Medical history
Medical and developmental history: she has hyperextensible joints, easy bruising, delayed wound healing, frequent joint subluxations in her hips, knees and elbows
Allergies - environmental, food, medication:
1. substances as normal saline and scratching herself
Global age of onset: Juvenile onset

Measurements
Age: 18y 1m
Weight: 74.8 kg 96th pctl (+1.75SD)
Height: 161.0 cm 38th pctl (-0.32SD)

BMI: 28.86 97th pctl (+1.88SD)

Clinical symptoms and physical findings
CUTANEOUS
Soft skin
MUSCULOSKELETAL
Joint hypermobility
Beighton score 2/9
Limb pain
CONNECTIVE TISSUE
Cigarette-paper scars

Patient 93(F)
Age: 10 years
Sex: Male
Indication for referral: query Ehlers-Danlos hypermobility type,

Family history and pedigree
List health conditions found in family (describe the relationship with proband)
Mom dx asthma as child, feeding issues, hypotonia, dyslexia, psychiatric issues
NO Consanguinity

Prenatal and perinatal history
DELIVERY
NO Premature birth
Premature rupture of membranes
NEONATAL GROWTH PARAMETERS
Increased body weight
CARDIOVASCULAR SYSTEM
Syncope
DIGESTIVE SYSTEM
Nausea
NERVOUS SYSTEM
Seizures
Notes:
born at 10lbs, unusual head size

Measurements Date:
Age: 18y 0m
Weight: 46.9 kg 0th pctl (-2.73SD)
Height: 162.0 cm 3rd pctl (-1.9SD)
BMI: 17.87 5th pctl (-1.68SD)
Head circumference: 53.0 cm 1st pctl (-2.21SD)
Outer canthal distance: 10.0 cm 95th pctl (+1.67SD)
Inner canthal distance: 3.0 cm 28th pctl (-0.59SD)

Interpupilary distance: 6.5 cm 86th pctl (+1.08SD)
Arm span: 161.0 cm = Height - 1.0 cm

Clinical symptoms and physical findings
GROWTH PARAMETERS
Weight for age
Decreased body weight (<-2SD)
Failure to thrive
was born 10lbs, lost weight in first few weeks and never gained it back.
CRANIOFACIAL
NO Bifid uvula
EAR DEFECTS
Hearing impairment
CUTANEOUS
Fragile skin
Mild
Poor wound healing
CARDIOVASCULAR
NO Syncope
NO Palpitations
RESPIRATORY
Asthma
MUSCULOSKELETAL
Muscular hypotonia
Joint dislocation
Infantile onset
Joint hypermobility
Knee, not elbow Beighton score 4/9
Clinodactyly of the 5th finger
Knee dislocation
Juvenile onset
multiple knee dislocations
Prominent proximal interphalangeal joints
Hypermobility of distal interphalangeal joints
GASTROINTESTINAL
Crohn's disease
NEUROLOGICAL
Motor delay
sat up at 1yr did not crawl but had bum shuffle first steps with walker at 24 months
needed stroller for long
distances at 5yrs
NO Autonomic dysregulation
BLOOD AND BLOOD-FORMING TISSUES
Bruising susceptibility
METABOLISM/HOMEOSTASIS
Food intolerance

CONNECTIVE TISSUE
Scarring
"distended thin scar" right elbow

Patient 95(C)
Age 13 years
Sex: Male
Indication for referral: vision problems at night, hearing issues resolved after tube insertion, delayed speech (started at age 4), mild redgreen color deficiency on testing, right iris heterochromia, normal discs, retinal pigment epithelial stippling, early pigmentary changes in mid and far periphery, photoreceptor retinal layers disrupted outside fovea, consistent with AR early onset retinal dystrophy

Family history and pedigree
Paternal ethnicity:
1. Scottish
Maternal ethnicity:
1. Polish
2. Irish
List health conditions found in family (describe the relationship with proband)
Pat aunt "racing heart"
Pat grnd mother d. 49 "sudden cardiac death"
NO Consanguinity
NO Parents with at least 3 miscarriages

Clinical symptoms and physical findings
EYE DEFECTS
Abnormality of color vision
Mild
mild red-green color deficiency on testing
Retinal dystrophy
Abnormality of retinal pigmentation
retinal pigment epithelial stippling
Central heterochromia
NO Optic disc hypoplasia

Genotype information
LIST OF GENES
Gene Status Strategy Comments
1 array CGH retinal dystrophy genes Negative OGT v2
2 ADH7 Negative
VUS: EYS c.3568+5T>C
VUS: GPR125 c.2107C>T (p.ARg703Trp)
VUS: MYO7A c.4159G>A (p.Asp1387Asn)
VUS: OTX2 c.707C>A (p.Thr236Asn)

Patient 97(D)
Age 19 yrs
Sex: Female
Indication for referral: connective tissue disorder (EDS hypermobility), myopia, cardiovascular dysautonomia, generalized musculoskeletal pain, heaches with mydriasis and floaters, IBS, hx syncope with postural hypertension, myopia, sx excision of thyroglossal duct cyst, ?Chiari malformation, poor heat tolerance, multiple allergies,easy bruising, wide scars, chokes on food, things that aren't smooth,

Clinical symptoms and physical findings
CRANIOFACIAL
NO Bifid uvula
EYE DEFECTS
Myopia
Mydriasis
Vitreous floaters
NO Visual loss
CARDIOVASCULAR
Hypertension
postural
Syncope
Abnormality of the cardiovascular system
MUSCULOSKELETAL
Joint hypermobility
Beighton 4/9
GASTROINTESTINAL
Inflammation of the large intestine
NEUROLOGICAL
Dysphagia
Headache
BLOOD AND BLOOD-FORMING TISSUES
Bruising susceptibility
METABOLISM/HOMEOSTASIS
Heat intolerance
CONSTITUTIONAL SYMPTOM
Pain
ENDOCRINE SYSTEM
Thyroglossal cyst

Patient 99(B)
Age: 6 years
Sex: Male
Indication for referral: Seizure disorder including infantile spasms

Family history and pedigree
List health conditions found in family (describe the relationship with proband)
(1) Mother with anxiety
(2) Maternal great aunt with breast cancer in her 40s
(3) Maternal great grandmother with ovarian cancer in her 50s
NO Other affected relatives
NO Consanguinity
NO Parents with at least 3 miscarriages

Prenatal and perinatal history
NO Multiple gestation
Gestation at delivery (weeks) Term birth
Maternal age at EDD (years) 23
Paternal age at EDD (years) 23
PREGNANCY HISTORY
Maternal hypertension
DELIVERY
NO Abnormal delivery (Non-NSVD)
NO Premature birth
NEONATAL GROWTH PARAMETERS
Abnormal birth weight
NO Small for gestational age (<-2SD)
NO Large for gestational age (>+2SD)
METABOLISM/HOMEOSTASIS
NO Neonatal hypoglycemia
DIGESTIVE SYSTEM
NO Prolonged neonatal jaundice
NO Poor suck
RESPIRATORY SYSTEM
NO Neonatal respiratory distress
IMMUNE SYSTEM
NO Neonatal sepsis
Notes:
Pregnancy-induced hypertension did not require treatment. Maternal fever during labour, treated with antibiotics. BW ~3 kg.

Medical history
Allergies - environmental, food, medication:
1. Ibuprofen (eye swelling)
2. Nuts (eye swelling)

Measurements
Age: 8m
Weight: 9.3 kg 67th pctl (+0.43SD)
Height: 70.0 cm 21st pctl (-0.82SD)
BMI: 18.98 87th pctl (+1.15SD)
Head circumference: 41.9 cm 0th pctl (-2.81SD)
Age: 1y 3m
Weight: 12.7 kg 96th pctl (+1.72SD)
Height: 79.0 cm 32nd pctl (-0.46SD)
BMI: 20.35 99th pctl (+2.55SD)
Head circumference: 45.0 cm 3rd pctl (-1.92SD)
Age: 4y 0m
Weight: 17.13 kg 62nd pctl (+0.32SD)
Height: 104.3 cm 56th pctl (+0.15SD)
BMI: 15.75 63rd pctl (+0.32SD)
Head circumference: 49.3 cm 12th pctl (-1.16SD)

Clinical symptoms and physical findings
MUSCULOSKELETAL
Osteopenia
During hospitalization at 10 months of age, had multiple issues attributed to ACTH side effects: low Vitamin D,
hypocalcemia, osteopenia, nephrocalcinosis, and kidney injury.
BEHAVIOR, COGNITION AND DEVELOPMENT
Global developmental delay
Borderline
His development was disrupted by onset of seizures, but after appropriate treatment he caught up. At time of
last assessment, there were no concerns with his development.
NEUROLOGICAL
Seizures
In addition to infantile spasms, has had tonic, absence, and complex partial seizures
Infantile spasms
Onset at 5 months of age
ABNORMAL TEST RESULT
Prolonged QT interval
Borderline
Seen by Cardiology, not felt to have prolonged QT syndrome

Patient 102(A)
Age: 10 years
Sex: Male
Indication for referral: poor balance, hyperflexible joints Beighton score 6/9, speech and language delay, ADD, asthma, pectus carinatum, mild pes planus, high arched palate, dental crowding, triangular chin, mild micrognathia, large eyes

Family history and pedigree
Paternal ethnicity:
1. Caucasian
Maternal ethnicity:
1. Caucasian
NO Consanguinity

Prenatal and perinatal history
Notes: amniocenthesis normal
walked ~18 months
single words by 2 years, speech delay

Clinical symptoms and physical findings
CRANIOFACIAL
High palate
Pointed chin
"triangular"
Micrognathia
Mild
Dental crowding
EYE DEFECTS
Visual impairment
Large eyes
NO Ectopia lentis
EAR DEFECTS
NO Hearing impairment
CARDIOVASCULAR
NO Syncope
NO Abnormal echocardiogram
RESPIRATORY
Pectus carinatum
Asthma
Childhood onset
MUSCULOSKELETAL
Osteopenia
Joint hypermobility
able to put both legs around his head and spontaneously sublux his shoulder, hips and fingers Beighton score
6/9

Pes planus
Mild
NO Pathologic fracture
BEHAVIOR, COGNITION AND DEVELOPMENT
Delayed speech and language development
18-20 months behind peers expressive language more affected than receptive speech and language therapy
and occupational therapy
Attention deficit hyperactivity disorder
NEUROLOGICAL
Anxiety
BLOOD AND BLOOD-FORMING TISSUES
NO Bruising susceptibility
NO Persistent bleeding after trauma
CONSTITUTIONAL SYMPTOM
Pain
nonspecific

Prediction method descriptions

CAGI5

Group 1

We first used our text-mining framework TPX [Joseph et al., 2012] for semi-automated HPO coding for each phenotype case. This resulted in a set of HPO codes for each case. We reviewed this output and came up with a final set of HPO codes by manually adding/removing/modifying as required.

Using the set of HPO terms for each case, we queried the HANRD (*Heterogeneous Association Network for Rare Diseases)* network for a ranked list of genes. HANRD is a heterogeneous network consisting of entities such as genes, phenotypes, diseases and pathways as nodes while associations between these entities are represented as weighted edges. The weight of an edge represents the score of the association between the entity pairs. Existing association networks usually view ontological associations as distinct from the network of other heterogeneous associations [Ullah et al., 2013]. We instead combined pairwise ontological and curated associations into a single heterogeneous association network. Motivated by the recent progress in spectral graph convolutions [Hammond et al., 2011; Kipf and Welling, 2017], we developed an information propagation algorithm GCAS (Graph Convolution-based Association Scoring) that performs information propagation on the initial ontological and curated association network and infers novel binary associations between the entities of the network (Paper submitted). These inferred associations are added to the aforementioned initial network, and the resulting network of ontological, curated and inferred associations is called HANRD (paper accepted). HANRD gives us a set of ranked genes based on the input phenotype.

We also used each set of phenotypes to query TPXRD. TPXRD is a heterogeneous dataset of rare disease associations from MEDLINE abstracts generated using modules of the TPX framework. It contains association pairs such as disease-gene, phenotype-phenotype and phenotype-disease extracted from MEDLINE using TPX. We have previously described TPX, a web-based text-mining tool that

supports real-time entity assisted search and navigation of the MEDLINE repository whilst continuing to use PubMed as the underlying search engine. Although the TPX tool is primarily meant to search PubMed, specific modules of the TPX framework such as the dictionary-based named entity recognition (NER), acronym handler and association extraction (AE) were re-purposed for extracting rare disease entity association pairs from literature. Rare disease-specific dictionaries for rare diseases, phenotypes and genes from multiple sources were created, and resolution of conflicts and overlaps amongst these dictionaries was done. MEDLINE abstracts related to rare diseases were identified using terms from the disease dictionary. We then apply the information propagation algorithm GCAS (Graph Convolution-based Association Scoring) that performs information propagation on the initial ontological and curated association network and infers novel binary associations between the entities of the network. These inferred associations are added to the aforementioned initial network, and the resulting network of ontological, curated and inferred associations is called TPXRD (paper in progress). Thus, TPXRD also gives us a set of ranked genes based on the input phenotype.

Using the VCF files, we performed variant prioritization using VPR (variant prioritization), our in-house variant prioritization pipeline. Here, variants are prioritized independent of gene information using an in-house scoring scheme. The score, ranging from 0 to 1, is a weighted combination of global minor allele frequency, conservation information, and functional information. Allele frequency scores are derived from public data sources such as 1000 genomes and conservation scores from GERP, PhyloP, and PhastCons. Functional scoring is done depending on region and mutation type. Variant effect predictions are combined from sources like CADD, REVEL, LINSIGHT etc. and prior knowledge of variant from data sources like Clinvar is also considered. VPR gives us a set of ranked genes based on the input genotype.

We then manually look at the different ranked gene lists and try and match a phenotypic case to possible genotypic cases. This involved comparing the top genes from HANRD/TPXRD with those from VPR. This was done via the intermediary disease link. Figure 1 shows the overall approach of our method.

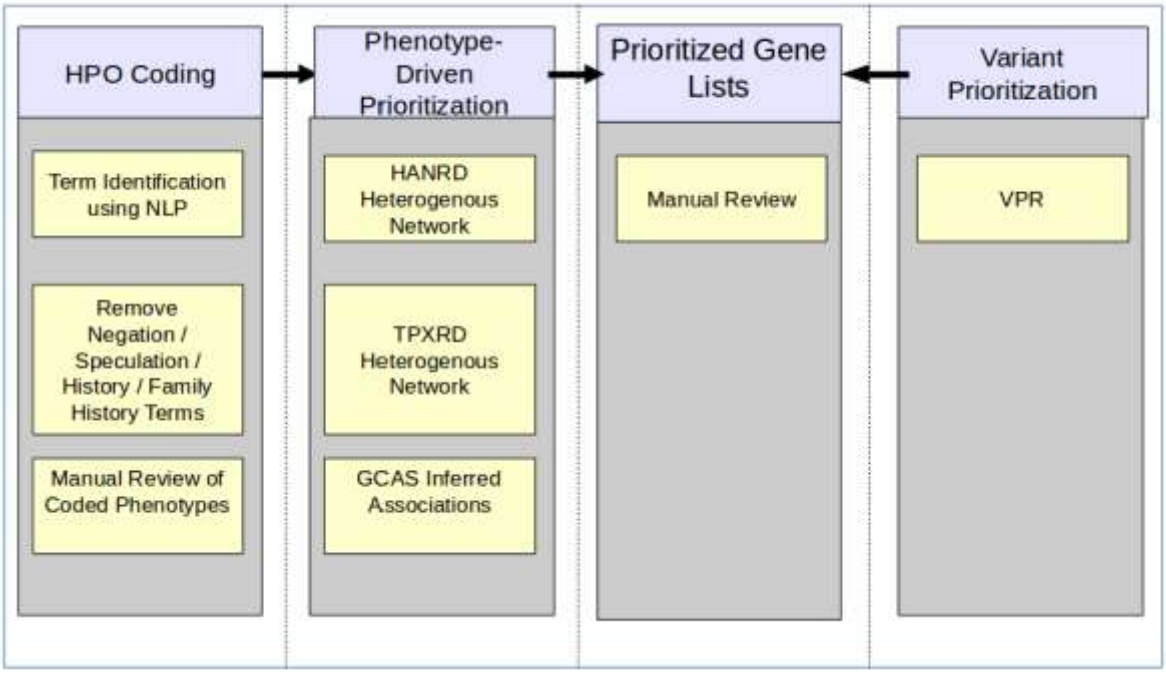

Figure 1. Overall Approach followed using phenotype-driven prioritization networks HANRD and TPXRD along with variant prioritization method VPR.

References

Hammond DK, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory. Applied and Computational Harmonic Analysis 2011, 30(2):129{150.

Kipf TN, Welling M: Semi-supervised classification with graph convolutional networks. ICLR 2017.

Joseph, T., Saipradeep, V. G., Raghavan, G. S. V., Srinivasan, R., Rao, A., Kotte, S., and Sivadasan, N. (2012). Tpx: Biomedical literature search made easy. Bioinformation, 8(12):578.

Smedley D and Robinson PN. Phenotype-driven strategies for exome prioritization of human Mendelian disease genes. Genome Med. 2015 30;7:81.

Ullah M, Aono M, Seddiqui M. Estimating a ranked list of human hereditary diseases for clinical phenotypes by using weighted bipartite network. In Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, Volume 2013 2013:3475{3478.

Group 2

To correctly associate clinical panels with genomes, we started from the symptoms and signs described in the clinical reports. We mapped them into the Human Phenotype Ontology (HPO) and the OMIM classifications. We used in-house resources, such as eDGAR (Babbi et al, 2017) and PhenPath (Babbi et al., 2019), to link phenotypes to panel-specific sets of candidate genes. We then searched in genomes for variations in these genes. To this aim, we retained from vcf files only the variations marked as "PASSED" and annotated their outcome using VEP (McLaren et al, 2016).

We scored the pathogenicity of each variant firstly by looking whether it is associated to some disease in the literature (mainly considering the UniProtKB annotation). Other missense variations in candidate genes were scored with SNPs&GO (Calabrese et al, 2009); a method based on Support Vector Machines for the prediction of deleterious single amino acid polymorphisms using protein functional annotation.

The output of SNPs&GO returns the effect (*Disease* associated variant or *Neutral* variant) associated with a Reliability Index (RI) that is a number scoring from 0 (unreliable) and 10 (reliable). We checked our predictions also considering the sex of the individual, to confirm that the genome-clinical panel association is plausible.

When no clear association emerged (e.g. when different variants in different genes had similar pathogenicity scores) we did not indicate any variant in the final submission file (since the submission of the causative variant was optional). However, due to the format of the submission file, it was not possible to leave the variant field completely empty, and thus we used the code 1:0:-:- to specify that we had no specific causative variant.

References

Babbi G, Martelli PL, Profiti G, Bovo S, Savojardo C, Casadio R. (2017) *eDGAR: a database of Disease-Gene Associations with annotated Relationships among genes*. BMC Genomics. 5:554

Babbi G, Martelli PL, Casadio R. (2019) *PhenPath: a tool for characterizing biological functions underlying different phenotypes.* BMC Genomics. (accepted, in press)

Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. (2009) *Functional annotations improve the predictive score of human disease-related mutations in proteins*. Hum Mutat. 30:1237-1244

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. (2016). *The Ensembl Variant Effect Predictor.* Genome Biology. 17(1):122.

Group 3

VCF files were analyzed using standard parameters, including variant quality, allele frequency, functional damage prediction and gene-phenotype associations, using a variety of tools and databases. Gender was considered in phenotype-genotype matches, ethnic origin was not taken into account.

Group 4

Collection of SickKids challenge phenotypes and the corresponding gene list

A total of 213 clinical phenotype descriptions were extracted from the CAGI phenotype data provided for 24 children. These phenotypes were used to extract a total 6239 potentially relevant genes from the Human Phenotype Ontology-based database (HPO) (Build #139) [1] and the dbNSFP database (version 3.5a) [2]. We also used the list of 319 genes from RetNet database [3] for searching for eye disorder related variants. The gene list for secondary variants was taken from the Table in the 2017 ACMG guidelines [4].

Annotation of VCF files and QC filters

The VCF files (including SNV and Indels) provided for this challenge from Illumina Hiseq were annotated using the Varant [5] tools, including region of occurrence (intron, exon, splice site or intergenic), observed minor allele frequencies (MAF), mutation type, predicted impact on protein function, and previously established associated phenotypes reported in ClinVar [6]. The RefGene [7] gene definition file was used for gene and transcript annotations in Varant. In addition, in-house scripts were written to further annotate the VCF files with HGMD [8] disease related variants and with dbscSNV [9] variants that potentially alter splicing. We also used the Annovar annotation tool [10] to extract Genome Aggregation Database (GnomAD) frequency data [11], Eigen scores [12] and GERP++ scores [13]. Chromosome M was annotated and searched for pathogenic variants using MSeqDR mv tool [14]. We used only high quality variants for further analysis - for SNVs, and Indels, we used the ‘PASS’ filter.

Diagnostic variant identification

A hierarchical scheme was used for identification of diagnostic variants, based on the strength of the evidence for disease relevance. All accepted high quality variants in the selected gene list with the highest population frequency <1% in the GnomAD exomes [11], or GnomAD genomes [11], or 1000 genome data [15] or in the ExAC database [11]) were first categorized into ordered tiers as follows:

Category 1: Variants with HGMD annotation of either DM (disease-causing mutation) or DP (disease-associated polymorphism), and/or reported in ClinVar with pathogenic or likely pathogenic clinical significance status.

Category 2: Nonsense mutation, direct splicing mutation disrupting either splice donor or acceptor site, frameshift or non-frameshift mutation, splice altering variant predicted by the dbscSNV database [9], and non-synonymous mutations predicted as damaging by SNPs3D profile and stability methods [16, 17], SIFT [18], PolyPhen-2 [19], Vest [20], REVEL [21] and CADD [22]. For inclusion of a non-synonymous variant in Category 2, at least 60% of these methods were required to return a prediction of deleterious. This threshold is based on a calibration against HGMD. (Note that for any given variant not all methods may return a result, hence the non-obvious cutoff).

Category 3: Non-synonymous mutations predicted as damaging by one or more of the above non-synonymous impact prediction methods, with the deleterious prediction agreement fraction < 0.6.

Category 4: Benign non-synonymous mutations (zero reporting non-synonymous methods predicting deleterious).

Category 5: Variants annotated as close to a splice acceptor or splice donor site.

Category 6: Variants annotated as UTR and intronic. Pathogenicity of these noncoding variants are based on CADD [22], Eigen [12] and GERP++ [13] scores.

Variants from all categories were further categorized into ordered tiers according to their rarity in the population data.

Frequency bin 1: Novel mutations (not seen in any of 1000 genomes, ExAC, gnomAD exomes and genomes databases).

Frequency bin 2: Variants with population frequency > 0 and <= 0.001.

Frequency bin 3: Variants with population frequency > 0.001 and <= 0.005.

Frequency bin 4: Variants with population frequency > 0.005 and < 0.01.

Binned variants were further filtered for an appropriate inheritance model using the OMIM inheritance pattern. Variants were assigned to autosomal dominant, autosomal recessive, compound heterozygous, pseudo autosomal recessive, or X-linked recessive models.

Profile matching with genomes

For each phenotypic profile, each phenotips term was assigned a subjective value from 0 to 1, according to its importance. For example, if a connective tissue disorder is the most serious and definitive term in the profile, it was scored the highest. If seizure is also part of that profile with borderline occurrence, then that was assigned a lower value than would be the case if the term occurred in a profile where seizure is the most serious phenotype. We then calculated a weighted matching score between the phenotypic profile and each variant-carrying gene in a genome. We selected top five scoring genomes for each clinical profile for further analysis. For each genome, we examined the evidence supporting the top five scoring variant-carrying genes, considering gender match, inheritance pattern and correspondence with the OMIM disease description.

Searching for Predictive Secondary variants

Here we followed the rules in ACMG (2017) [4] to extract predictive secondary variants from 59 genes. We searched for clinical variants and loss of function variants in those genes according to the Table 1 in [4].

References

1. Sebastian Köhler, Sandra C Doelken, Christopher J. Mungall, Sebastian Bauer, Helen V. Firth, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucl. Acids Res. (1 January 2014) 42 (D1): D966-D974 doi:10.1093/nar/gkt1026

2. Liu X, Wu C, Li C and Boerwinkle E. 2015. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Non-synonymous and Splice Site SNVs. Human Mutation. Published Online First 10 NOV 2015 | DOI: 10.1002/humu.22932.

3. SP Daiger, BJF Rossiter, J Greenberg, A Christoffels, W Hide. Data services and software for identifying genes and mutations causing retinal degeneration. Invest. Ophthalmol.Vis. Sci. 39:S295, 1998.
4. Kalia, S. S. et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet. Med. 19, 249–255 (2017).
5. VARANT: http://compbio.berkeley.edu/proj/varant/Home.html
6*. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM & Maglott DR: ClinVar: public archive of relationships among sequence variation and human phenotype. Nucl. Acids Res.2014; 42:D980-D985. doi:10.1093/nar/gkt1113.*
7. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, *et al*: RefSeq: an update on mammalian reference sequences. Nucl AcidsRes.2014; 42:756–63.
8. Stenson PD, et al: Human Gene Mutation Database (HGMD): 2003 update. Hum Mutat.2003; 21:577-581.
9. Jian X, Boerwinkle E and Liu X: In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res.2014; 42:13534–44.
10. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data Nucleic Acids Research, 38:e164, 2010.
11. Lek M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature, 536, 285-291 (2016).
12. Ionita-Laza I. et al. A spectral approach integrating functional genomic annotations for coding and noncoding variants. Nat. Genet. 48, 214-220. (2016).
13. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S (2010) Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. PLoS Comput Biol 6(12): e1001025.
14. Shen L. et al. MSeqDR mvTool: A mitochondrial DNA Web and API resource for comprehensive variant annotation, universal nomenclature collation, and reference genome conversion. Hum Mutat. 2018 Mar 14. doi: 10.1002/humu.23422.
15. The 1000 Genome Project Consortium: An integrated map of genetic variation from 1,092 human genomes. Nature.2012; 491:56–65. doi:10.1038/nature11632
16. Yue P, Melamud E, Moult J: SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics. 2006; 7:166.
17. Yue P, and Moult J. Loss of protein structure stability as a major causative factor in monogenic disease. J. Mol. Biol. 353, 459-473. (2005).
18. Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc.2009; 4:1073-1081. doi:10.1038/nprot.2009.86.
19. Adzhubei I, Jordan DM, Sunyaev SR: Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet 2013, Chapter 7:Unit7.20, doi:10.1002/0471142905.hg0720s76
20. Carter H. et al. Identifying mendelian disease genes with the variant effect scoring tool. BMC Genomics. 14(Suppl 3): S3. (2013).
21. Ioannidis et al. REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. Am. J. Hum Genet. 99, 877-885. (2016).

22. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J: A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet.2014; 46:310-315. doi: 10.1038/ng.2892.

Group 5

Evolutionary Action (EA) diffusion in a gene-disease network

In order to predict the disorder class (eye disorder, neurogenetic disease, or connective tissue disorder) for each individual's genome sequencing data set, the predictors calculated the effect of the genetic variants on the fitness of each gene (see A below). This fitness effect was used as the input of a diffusion process over a network of genes and diseases (see B below). The diffusion signal on each of the three disorder classes was used to calculate the probability of each genome to be linked to each disease.

In order to match each individual's genome sequencing data set to a clinical report file, the predictors used again the diffusion process (see B below), but the signal was measured at each specific symptom of the clinical description files. This process generated overwhelming information. To narrow down the potential matches, the predictors identified the gender and predicted the ethnic origin of each genome (see D below). Because the clinical reports varied a lot in the amount, type, and the detail level, the genomes were matched to clinical reports manually, by weighing the various accumulated links of genotype to phenotype. Only 12 of the 24 genomes were matched to clinical reports, while for the rest 12 genomes the submitted probabilities were arbitrarily set to 10% (value of 0.1) for all remaining potential matches after narrowing them down based on gender and ethnic group predictions.

No predictions of diagnostic variants nor of secondary variants were submitted.

Detailed Calculations

A. Gene fitness effect. The predictors computed the fitness of each gene in each individual's genome based on the predicted impact of all variants called in that gene. The variant impact was calculated according to the Evolutionary Action (EA) method for each missense variant (see C below), where a value of 0 corresponds to wild-type level gene activity and a value of 100 corresponds to complete loss of gene function. Nonsense and start loss variants were given impact of 100, while synonymous variants were given impact of 0. Insertions, deletions, and variants without a PASS filter were

given impact of 0 (although insertions and deletions may have strong impact on gene function, the ability to separate them into passenger and loss of function variants might be questionable). The fitness effect on a gene was defined to be equal to: 0 if there was no mutation, EA/100 if there was one mutation, or 1-∏(1-EAm/100) if there were multiple mutations in that gene (∏ indicates the product for all mutations m). To account for the different ability of genes to tolerate mutations, the predictors weighted the fitness effect with a gene importance score (multiplied them). The ability of each gene to tolerate mutations was calculated as the average EA score of all variants in that gene found by the gnomAD project (Lek et al. 2016). This average EA value was transformed into a fraction coverage (0 means the gene can tolerate mutations with the highest impact and 1 means the gene can only tolerate mutations with the lowest impact), which coverage was used as the gene importance weighting factor of the gene fitness effect.

B. Diffusion on the gene disease network. A gene and disease network was constructed based on multiple sources (Stark et al. 2006; Gutierrez-Sacristan et al. 2015; Szklarczyk et al. 2015; Davis et al. 2017). For each individual's genome, the predictors run a diffusion process of weighted gene fitness effects (see A above) over the network and measured the diffusion signal outcome at disease labels (Lin et al. 2018). The disease labels could be either the disorder classes or specific symptoms of the clinical reports. The diffusion signal on each disease label was normalized for the 24 individual's genomes and it was reported as a fraction coverage. For example, for each disease label, the individual with the lowest diffusion signal had coverage of 0 (0/24) and the individual with the highest diffusion signal had coverage of 0.96 (23/24). This coverage was used as the probability that the individual has the particular phenotype.

C. The Evolutionary Action (EA) method predicts the fitness effect of the genetic variants (Katsonis and Lichtarge 2014). EA does not involve any training, because it relies on a formal equation of the genotype-phenotype relationship. The terms of this equation were calculated using protein homology data. Briefly, the EA equation states that the fitness effect of a mutation equals the product of the sensitivity of the mutated position with the magnitude of the change. The sensitivity of the position is calculated by quantifying the correlation of the residue variations with phylogenetic branching within

an alignment of homologous sequences (Lichtarge et al. 1996; Mihalek et al. 2004; Lichtarge and Wilkins 2010). The magnitude of the change is calculated from substitution likelihood according to numerous sequence alignments for the given context (strata of sensitivity of the position, and optionally additional stratification based on structural features). The calculated product is then normalized to represent the percentile rank of each variant within the protein in the scale of 0 (benign) to 100 (pathogenic). The EA scores are available for all human variants at:
http://mammoth.bcm.tmc.edu/EvolutionaryAction

D. Using gender and ethnic information. To identify the gender, we used the concordance of reads in the X chromosome. For male genome sequencing data, a large fraction of X chromosome calls appears to be homozygous, while for female sequencing data that fraction is distinctly lower. To identify the ethnic background, we calculated the proximity of each genome to the genomes of ethnic groups available in the 1000 Genomes Project (The 1000 Genomes Project Consortium et al. 2015). The predictions were at the level of major ethnic groups (EUR, AFR, AMR, EAS, and SAS).

References

Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ. 2017. The Comparative Toxicogenomics Database: update 2017. Nucleic Acids Res 45: D972-D978.

Gutierrez-Sacristan A, Grosdidier S, Valverde O, Torrens M, Bravo A, Pinero J, Sanz F, Furlong LI. 2015. PsyGeNET: a knowledge platform on psychiatric disorders and their genes. Bioinformatics 31: 3075-3077.

Katsonis P, Lichtarge O. 2014. A formal perturbation equation between genotype and phenotype determines the Evolutionary Action of protein-coding variations on fitness. Genome research 24: 2050-2058.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536: 285-291.

Lichtarge O, Bourne H, Cohen F. 1996. An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol 257: 342 - 358.
Lichtarge O, Wilkins A. 2010. Evolution: a guide to perturb protein function and networks. Curr Opin Struct Biol 20: 351-359.
Lin CH, Konecki DM, Liu M, Wilson SJ, Nassar H, Wilkins AD, Gleich DF, Lichtarge O. 2018. Multimodal Network Diffusion Predicts Future Disease-Gene-Chemical Associations. Bioinformatics doi:10.1093/bioinformatics/bty858.
Mihalek I, Res I, Lichtarge O. 2004. A family of evolution-entropy hybrid methods for ranking protein residues by importance. J Mol Biol 336: 1265 - 1282.
Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. 2006. BioGRID: a general repository for interaction datasets. Nucleic Acids Res 34: D535-539.
Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP et al. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43: D447-452.
The 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA et al. 2015. A global reference for human genetic variation. Nature 526: 68-74.

Group 6

Our strategy involved predicting three different pieces of information from the genomes based on what is presented in the clinical descriptions.

A. Prediction of sex of the genomes to match to reported sex in the clinical notes

B. Prediction of ethnicity of the genome to match to the parental ethnicities reported in clinical notes

C. Prediction of phenotypes associated with genes carrying pathogenic mutations to match to the clinical symptoms reported in the clinical notes

For A., BCFtools +guess-ploidy function was used to guess the ploidy. Using cutoffs for the log P(Haploid)/nSites values of male if >-2 and female <-3, all but two genomes (WGS-NGS-017-03,WGS-NGS-018-03) were assigned a sex. Each genome/clinical

note pair was assigned a probability based on the predicted sex for that genome and the total occurrence of that sex in all 24 clinical notes.

For B., predictions were made using akt ancestry tools and the principal components were projected onto the components already available for 1000 genomes samples. The top 20 closest samples in the 1000 genomes were used to infer ethnicity.

For C., predictions were made using multiple criteria. First, for each genome, an in-house tool (varant) was used for annotation. All protein-altering variants with 1000G MAF < 2% were considered (PA-vars). The pdf clinical notes were converted to text using pdf2text tool and clinical terms describing patient features were extracted for each clinical note. These terms for each note was searched against a gene-phenotype database (Monarch initiative for submission 1 and eRAM for submission 2), and the genes were sorted by with highest number of matching terms. The genes matching phenotype terms from all clinical notes were then intersected with genes containing protein altering variants in all genomes. To score each pair of genome-clinical note, the following were considered.

a. Recessive genes (according to dbNSFP P(rec) > 0.6) in the top 25 percentile of term matches with >1 PA-vars were given a score +1

b. Dominant and also essential genes (according to dbNSFP P(rec) < 0.3) in the top 25 percentile of term matches with >0 PA-vars were given a score +1

c. Essential genes with low RVIS (according to dbNSFP) in the top 25 percentile of the term matches with >0 PA-vars were given a score +1

d. Genes with the highest number of term matches in Homo Sapiens in Monarch initiave genotype-phenotype database with >0 PA-vars were given a score +1

e. Genes with the highest number of term matches with direct mouse experimental evidence in Monarch initiative with >0 PA-vars were given a score +1

f. Genes in the top 25 percentile of term matches with StopGain, FrameShiftInsert,FrameshiftDelete were given a score of +1

For each genome/clinical note pair, the highest scoring gene was recorded in the respective comment field. The corresponding variants were recorded in the DV fields and the remaining variants from the rest of the lower scoring genes were recorded in the PSV column. For a genome, the highest score for each clinical note was gathered

and unit normalized across all clinical notes. These scores for each genome were then divided by 1/3 of the sum of scores and added to the probability scores from A. and B. to obtain the P-A, P-B, P-C..etc. reported probability values.

A second submission (group 6.2) was also made, where instead of the Monarch initiative, we used eRAM (http://www.unimd.org/eram/) - encyclopedia of rare disease annotation to obtain gene-phenotype relationships. Four random terms from the clincal notes were randomly selected 500 times and queried in the database together (AND of the terms) and the resulting genes were intersected with PA-vars like with the Monarch Initiative data. The second submission took a more restrictive approach, for a given genome of a predicted sex, only clinical notes with consistent matching sex were considered further. In this restricted set of clinical notes, the genome/clinical note pairs were scored by using a., b., c. and f. from above.

Group 7

VCFs for each individual were uploaded into Ingenuity Variant Analysis (QIAGEN). Ingenuity Variant Analysis utilizes curated content from the literature as well as external databases for us in basic variant filtering on the following parameters: Variants were filtered based on quality (phred score of 20), population frequency (<3% in gnomAD), ACMG classification (pathogenic and likely pathogenic). Additionally, since the QIAGEN knowledgebase contains a gene-disease framework built on curated literature and clinical databases, I looked for variants in genes with a known relationship to eye disorders, connective tissue disorders, or neurological dysfunction. After reviewing all variant specific citations and data provided within Ingenuity Variant Analysis, the most compelling pathogenic and likely pathogenic variants were reported, and where possible, the variants were qualitatively matched to the medical record by visual inspection.

Group 8

Summary

To predict correspondence between phenotypes and genomes, we calculated scores for all genome-phenotype pairs and assigned the most likely connections using a bipartite matching algorithm. We obtained these scores from three independent factors: 1) pathogenic mutations (predicted by the MutPred suite) in genes related to the reported phenotype, 2) genetic similarity to the reported ancestry, and 3) the presence of Y-chromosome variants. We searched for pathogenic variants in gene sets composed of known risk genes and their putative interactors, which we predicted using a propagation algorithm on protein-protein interaction networks. Variants associated with ancestry and sex hold the strongest signal for determining the identity of a genomic sample, and they were key in guiding our matching algorithm. Beyond these guiding variants, however, matching genomes to phenotypes proved to be a challenging task. By narrowing the search for pathogenic variants to those in known disease genes, we sought to reduce false positives and simplify variant interpretation. Our approach of combining the pathogenicity scores of all variants in a single probability score per genome-phenotype implied that we lost focus on better predicting single causal variants.

Methods

We used previously reported disease genes as seeds on the human protein-protein interaction network for running a network propagation algorithm [6]. The propagation algorithm was performed in a 5-fold cross-validation manner so as to get an initial score between 0 and 1 for all genes. We then used the AlphaMax algorithm [2, 3] to estimate the proportion of the risk genes in the human genome and calibrate those initial scores to be proper probability scores measuring the likelihood of a gene being associated with the disease [3]. We built a total of eight gene sets related to the diseases of interest: Ehlers-Danlos Syndrome (19 seed genes), autism (69), retinitis pigmentosa (52), nystagmus (9), ataxia (3), as well as general eye anomalies (DOID:102), connective tissue diseases (DOID:65), and congenital neurologic anomalies (DOID:2490). We classified each case as belonging in one of these disease categories, and considered only the corresponding gene lists when calculating their phenotype-genome scores. However, since more than one disease phenotype was reported in some cases, we

allowed those to have a secondary disease category (included in a weighed average with its primary counterpart).

We annotated protein coding variation using custom scripts to extract nucleotide sequence from the CCDS database and predict the mutant amino acid sequence. We assigned pathogenicity prediction scores to missense and stop gain variants with MutPred2 [8] and MutPred-LOF [7], respectively. For each gene in every individual genome, we included only the variant with the highest pathogenicity prediction score (which may include multiple CCDS isoforms) in further analyses. To assess significance of these pathogenic variants, we generated an empirical null distribution using the beta family based on the MutPred [5] scores of variants present in gnomAD [4], within the genes for each phenotype. The eight disease scores for each genome were computed as products of the p-values (from the beta distribution) for the individual's variants in the corresponding gene set.

We estimated genetic similarity to the reported ancestry by comparing the genomes to relevant samples from the 1000 Genomes dataset [1] with identity-by-state calculated with the SNPRelate package in R [9]. We chose 1000-G populations (or super-populations, when appropriate) that were close to the reported ethnicity of the parents and estimated relative similarity scores for each genome.

The last component of a phenotype-genotype score was the sex of the individual. We inferred sex from genomic data from SNPs mapped to the Y chromosome (a genome was assumed male if there were any Y-chromosome variants, and female otherwise).

The total score used for bipartite matching was the weighted average of the similarity and disease scores, multiplied by a sex factor (=1 only if the inferred and reported sex matched). We repeated the bipartite matching process 1000 times varying the weight assigned to the similarity score ($p \sim U[0.2; 0.8]$) relative to the disease score ($1 \sim p$), and the weight of the secondary disease ($q \sim U[0; 0.3]$) relative to the primary category ($1 \sim q$).

References

[1] 1000 Genomes Project Consortium, G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles, and G. A. McVean. A map of human genome variation from population-scale sequencing. Nature, 467(7319):1061{1073, 2010.

[2] S. Jain, M. White, and P. Radivojac. Estimating the class prior and posterior from noisy positives and unlabeled data. arXiv:1606.08561, 2016.

[3] S. Jain, M. White, and P. Radivojac. Estimating the class prior and posterior from noisy positives and unlabeled data. In Advances in Neural Information Processing Systems, NIPS 2016, pages 2693{2701, 2016.

[4] K. J. Karczewski et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv, page 531210, 2019.

[5] B. Li, V. G. Krishnan, M. E. Mort, F. Xin, K. K. Kamati, D. N. Cooper, S. D. Mooney, and P. Radivojac. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics, 25(21):2744{2750, 2009.

[6] E. Nabieva, K. Jim, A. Agarwal, B. Chazelle, and M. Singh. Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. Bioinformatics, 21 Suppl 1:i302{310, 2005.

[7] K. A. Pagel, V. Pejaver, G. N. Lin, H. J. Nam, M. Mort, D. N. Cooper, J. Sebat, L. M. Iakoucheva, S. D. Mooney, and P. Radivojac. When loss-of-function is loss of function: assessing mutational signatures and impact of loss-of-function genetic variants. Bioinformatics, 33:i389{i398, 2017.

[8] V. Pejaver, J. Urresti, J. Lugo-Martinez, K. A. Pagel, G. N. Lin, H. J. Nam, M. Mort, D. N. Cooper, J. Sebat, L. M. Iakoucheva, S. D. Mooney, and P. Radivojac. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. bioRxiv 134981, 2017.

[9] X. Zheng, D. Levine, J. Shen, S. Gogarten, C. Laurir, and B. Weir. A high-performance computing toolset for relatedness and Principal Component Analysis of SNP data. Bioinformatics, 28(24):3326{3328, 2012.

CAGI4

Group 9

Following download and unzipping of the genome data and phenotypes files the data were prepared for analysis with the Exomiser(1,2) using the following steps:

1. The ASM/vcfBeta-[ASM-ID].vcf.bz2 were uncompressed to vcf format.
2. The HPO terms in the phenotype PDF files were extracted manually into plain text format files.
3. The mode of inheritance, where provided, were also extracted manually and merged with the genome and phenotype identifiers from the clinical_genomes_answer_key.txt
4. A Python script was run which used the outputs from steps 1-3 to write out a set of 25 Exomiser 7 yaml format analysis scripts (sup_1) and an exomiser batch analysis script listing the paths to 25 yaml files. The exomiser requires a list of HPO ids for phenotypic analysis so this script converted the HPO terms into a list of HPO ids using the OLS beta webservice(3).

The yaml script configured the Exomiser to:

- Exclude non-exonic (intergenic, intronic, upstream, downstream or intronic) or synonymous variants
- Exclude variants having a maximum minor allele frequency (MAF) of 1.0% in all the Thousand genome, ESP and ExAC datasets.
- Include variant pathogenicity predictors from the Polyphen2, MutationTaster and SIFT resources as well as scores derived from the variant type (e.g. missense or frameshift insertions)
- Score genes according to phenotypic similarity with the observed patient phenotype using the hiPHIVE prioritiser. This uses know human gene disease phenotypes, and mouse and fish knockout phenotypes along with random walk analysis of physical interaction networks.

These steps were chosen according to the guidelines in (4), which explains the steps and data sources in much full detail.

The data were analysed using the exomiser version 7.2.1 downloaded from (5) on a laptop equipped with 16GB RAM and an Intel 4th generation Core i7 processor running Windows 7 64-bit, Java version 8. This exomiser version is capable of analysing whole genomes using only a moderate amount of RAM. The analysis was run using the command:

java -Xms8G -Xmx12G -jar exomiser-cli-7.2.1.jar --analysis-batch {path to analysis batch file}

The resulting HTML files containing the top 20 candidate genes were inspected by hand and the most likely (rarest, most pathogenic) variants from the best phenotypic matching gene were included in the submission file.

References

(1) Improved exome prioritization of disease genes through cross-species phenotype comparison. Robinson PN, Köhler S, Oellrich A, Sanger Mouse Genetics Project, Wang K, Mungall CJ, Lewis SE, Washington N, Bauer S, Seelow D, Krawitz P, Gilissen C, Haendel M and Smedley D. Genome research 2014;24;2;340-8 PUBMED: 24162188; PMC: 3912424; DOI: 10.1101/gr.160325.113

(2) http://www.sanger.ac.uk/science/tools/exomiser

(3) http://www.ebi.ac.uk/ols/beta/

(4) Next-generation diagnostics and disease-gene discovery with the Exomiser. Damian Smedley, Julius O B Jacobsen, Marten Jäger, Sebastian Köhler, Manuel Holtgrewe, Max Schubach, Enrico Siragusa, Tomasz Zemojtel, Orion J Buske, Nicole L Washington, William P Bone, Melissa A Haendel & Peter N Robinson. Nature Protocols 10, 2004–2015 (2015) DOI: 10.1038/nprot.2015.124

(5) ftp://ftp.sanger.ac.uk/pub/resources/software/exomiser/downloads/exomiser/

Supplementary Information:

sup_1 - The Exomiser YAML script template. The fields $vcf_file, $inheritance_mode, $hpo_list and $out_file_prefix were replaced for each genome and phenotype set.

```
#Exomiser Analysis Template.
# These are all the possible options for running exomiser. Use this as a template for
# your own set-up.
---
analysis:
   vcf: $vcf_file
   ped:
   # AUTOSOMAL_DOMINANT, AUTOSOMAL_RECESSIVE, X_RECESSIVE or UNDEFINED
   modeOfInheritance: $inheritance_mode
   analysisMode: PASS_ONLY
   geneScoreMode: RAW_SCORE
   hpoIds: $hpo_list
   frequencySources: [
      THOUSAND_GENOMES,
      ESP_AFRICAN_AMERICAN, ESP_EUROPEAN_AMERICAN, ESP_ALL,
      EXAC_AFRICAN_INC_AFRICAN_AMERICAN, EXAC_AMERICAN,
      EXAC_SOUTH_ASIAN, EXAC_EAST_ASIAN,
      EXAC_FINNISH, EXAC_NON_FINNISH_EUROPEAN,
      EXAC_OTHER
   ]
   pathogenicitySources: [POLYPHEN, MUTATION_TASTER, SIFT]
   steps: [
      variantEffectFilter: {remove: [UPSTREAM_GENE_VARIANT,
         INTERGENIC_VARIANT,
         CODING_TRANSCRIPT_INTRON_VARIANT,
```

```
        NON_CODING_TRANSCRIPT_INTRON_VARIANT,
        SYNONYMOUS_VARIANT,
        DOWNSTREAM_GENE_VARIANT,
        SPLICE_REGION_VARIANT]},
    frequencyFilter: {maxFrequency: 1.0},
    pathogenicityFilter: {keepNonPathogenic: true},
    inheritanceFilter: {},
    omimPrioritiser: {},
    hiPhivePrioritiser: {}
  ]
outputOptions:
  outputPassVariantsOnly: true
  numGenes: 20
  outputPrefix: $out_file_prefix
  outputFormats: [HTML, TSV-GENE, TSV-VARIANT, VCF]
```

Group 10

Identifying variants responsible for the phenotypic abnormalities (CAGI4 SickKids challenge) involved running our in-house outlier-phenotype predictor algorithm. The tool relies on SUPERFAMILY (1) to annotate genetic variants falling within SCOP (2) domains with FATHMM (3) pathogenicity scores as well as data on allele frequency from the 1000 Genomes project (4). Each variant is linked with the phenotype terms it affects by transferring gene ontology annotations to the SCOP domain level using dcGO (5). The prediction algorithm runs an unsupervised learning pipeline that clusters genotypes into groups exhibiting levels of high similarity (indicating that multiple individuals have the same combination of variants associated with a phenotype) and identifying outliers from these clusters (those individuals that have a rare combination of variants, which were predicted to have strong functional effects) i.e. individuals likely to have an outlier phenotype. If a “sick kid” was an outlier for a phenotype term that

matched the clinical description (subjective judgement), we prioritised the causative variants by FATHMM score and allele rarity.

References

(1) Matt E. Oates, Jonathan Stahlhacke, Dimitrios V. Vavoulis, Ben Smithers, Owen J.L. Rackham, Adam J. Sardar, Jan Zaucha, Natalie Thurlby, Hai Fang, Julian Gough, The SUPERFAMILY 1.75 database in 2014: a doubling of data, *Nucleic Acids Research*, Volume 43, Issue D1, 28 January 2015, Pages D227–D233.

(2) Antonina Andreeva, Dave Howorth, John-Marc Chandonia, Steven E. Brenner, Tim J. P. Hubbard, Cyrus Chothia, Alexey G. Murzin, Data growth and its impact on the SCOP database: new developments, Nucleic Acids Research, Volume 36, Issue suppl_1, 1 January 2008, Pages D419–D42.

(3) Shihab, HA, Gough, J, Cooper, DN *et al*. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden markov models. *Hum Mutat* 2013; **34**: 57– 65.

(4) 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA,Abeca

(5) Hai Fang, Julian Gough, dcGO: database of domain-centric ontologies on functions, phenotypes, diseases and more, *Nucleic Acids Research*, Volume 41, Issue D1, 1 January 2013, Pages D536–D544.

Group 11

Phenotype to Disease/Phenotype to Genes

Since the original challenge was a matching challenge, my approach was to use the phenotypic information to find possible diseases and/or corresponding genes, if any. The phenotypic information already uses the HPO controlled-vocabulary since the referring clinical geneticist has entered this information using *Phenotips*, a Human Phenotype Ontology-based database [Girdea M et al., 2013]. However, there were certain verbatim term mentions from the phenotypic files that needed to be mapped to a HPO term.

Several tools that metrics that measure the semantic similarity between ontology annotations are present. Matching a term is not about exact match, it should also search across the HPO ontology. The Resnik measure seems the most common one when searching in the ontology [Resnik, 1995]. The HPOSim paper compares measures specific to the HPO context [Deng Y et al., 2015].

Phenomizer

I picked Phenomizer [Köhler et al., 2009 and Köhler et al.,2013] as my choice of tool to go from a set of HPO IDs to a scored and ranked set of OMIM/Orphanet IDs that could explain these IDs. Phenomizer was run online from the website http://compbio.charite.de/phenomizer/.

A set of phenotypic terms could be explained by 2 (or more?) disease terms, referred to in the paper as a complex phenotype with two genetic disorders [Stavropoulous et al, 2016]. Given that there could be complex phenotypes as well as the list of HPO IDs for some cases being >8, it could be important to identify critical IDs and give them higher weight. Phenomizer allows one to mark a HPO term as mandatory or observed, with the default being observed. By default, the tool uses the Resnik measure. It automatically converts the input query to a symmetric one, with the message “It may be appropriate to use the 'symmetric' mode for queries of size larger than 5. Should 'symmetric' be checked? (You can change that in the menu later on)”. Ideally this should be asymmetric, however in order to get possible disorders, I even ran this as symmetric.

I manually looked at the mapped disease terms to see coverage. For this, I picked OMIM over Orphanet as OMIM has a section called Clinical Synopsis in which one can choose to display all HPO IDs associated with the clinical phenotypes using the Display Options->Show Clinical IDs.

Phenolyzer

I picked Phenolyzer [Yang H et al, 2015] as my my choice of tool to go from a set of HPO IDs to a set of scored and ranked genes that could potentially explain the phenotype. Phenolyzer was run locally after installation.

I thus created the Phenomizer and Phenolyzer outputs.

Input File(s)

I only considered the vcf format as input. This file contains the small variant, copy number variation (CNV), structural variation (SV), and mobile element insertion (MEI) calls made by the Complete Genomics Assembly Pipeline for a single genome, and conforms to the VCF 4.1 specification. This could mean ignoring crucial input information present in the Master Variations file. Also, I am relying on the tool that converted the Master Variations file to vcf format. This tool produces a vcfBeta, so not sure if there is any loss-of-information when going from the Complete Genomics format to vcf.

My first tool for gene and variant prioritization was Exomiser [Smedley D et al., 2015]. It takes a vcf file as input and a set of phenotypes encoded using the Human Phenotype Ontology (HPO) it will annotate, filter and prioritize likely causative variants using:

• The functional annotation of variants is handled by Jannovar and uses UCSC KnownGene transcript definitions and hg19 genomic coordinates.

• Predicted pathogenicity data is extracted from the dbNSFP resource. Variant frequency data is taken from the 1000 Genomes, ESP and ExAC datasets

• Variants filtered for maximum allele frequency of 1%, or in some cases 2%

• Pathogenicity Filter: Retained all non-pathogenic missense variants

• The hiPHIVE algorithm combines data on the rarity of the variant and its predicted pathogenicity along with the similarity of the human, mouse, zebrafish phenotypes as well as a guilt-by-association approach using protein–protein associations for those genes that have no data in any of the species.

A typical Exomiser command that I used was

java -Xmx10g -jar exomiser-cli-7.2.0.jar -F 1 -v 1099.vcf --ped 1099.ped --hpo-ids HP:0009889, HP:0006466, HP:0007911, HP:0005709, HP:0001773, HP:0007598, HP:0001263, HP:0010864, HP:0000369, HP:0002194, HP:0200055, HP:0003701, HP:0000219, HP:0000664, HP:0001629 --prioritiser=hiphive

Genemania

Exomiser uses protein-protein interaction data to identify possible genes involved. The basis of the interaction wasn't always apparent to me. So I used Genemania [Warde-

Farley et al., 2010] to visualize the specific interactions, with emphasis on physical interactions between the entities.

Phen-Gen

Another tool that I used for gene and variant prioritization was Phen-Gen [Javed A et al, 2014]. Phen-Gen works as follows (below summary from Smedley and Robinson, 2015): It uses a Bayesian framework to compare predicted deleterious variants in the patient's exome and known patient symptoms to prior knowledge of human disease-gene associations and gene interactions. Coding variants are analyzed using a unifying framework to predict the damaging impact of non-synonymous, splice-site and indel variants.

• Any variant that has a MAF above 1 % is removed from further analysis.

• Genes are only retained for further analysis if the predicted damaging score for the variants exceeds that seen for 99 % of the 1000 Genomes dataset.

• These remaining genes are then analyzed using the Phenomizer algorithm to match semantically the proband's phenotypes encoded using HPO to known disease-gene associations.

A typical Phen-Gen command that I used was

perl phen-gen.pl input_phenotype=1099_CAGIV_HPO.txt input_vcf=1099.vcf input_ped=1099.ped

Final Variant Prioritization

Based on the inputs from Exomiser, Phen-Gen and Genemania, I picked variants based on quality, predicted pathogenicity, MAF frequencies and literature survey.

References

1. PhenoTips: patient phenotyping software for clinical and research use. Girdea M et al. Hum Mutat. 2013;34:1057-65.

2. HPOSim: An R Package for Phenotypic Similarity Measure and Enrichment Analysis Based on the Human Phenotype Ontology. Deng Y et al. PLoS One. 2015;10:e0115692

3. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. Köhler et al. Am J Hum Genet (2009) vol. 85 (4) pp. 457-64
4. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Köhler et al. Nucleic Acids Research (2013)
5. Using information content to evaluate semantic similarity in a taxonomy. Resnik P. Proceedings of the 14th international joint conference on Artificial intelligence—Volume 1. IJCAI'95 San Francisco, CA, USA: Morgan Kaufmann Publishers Inc; pp. 448–453
6. Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. Yang H et al. Nat Methods. 2015;12:841-3
7. Next-generation diagnostics and disease-gene discovery with the Exomiser. Smedley D et al. Nat Protoc. 2015;10:2004-15
8. Phen-Gen: combining phenotype and genotype to analyze rare disorders. Javed A et al. Nat Methods. 2014;11:935-7.
9. Phenotype-driven strategies for exome prioritization of human Mendelian disease genes. Smedley D and Robinson PN. Genome Med. 2015 30;7:81.
10. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Warde-Farley D et al. Nucleic Acids Res. 2010;38 Suppl:W214-20

Group 12

The genomes of 25 SickKids children were matched with their phenotypic description with the help of the published data in supplementary Table 4 of Stavropoulos, D.J., et al. (2016) [1]. Then we searched for diagnostic variants and predictive secondary variants for each child from the provided whole genome sequencing data, including SNVs, Indels, CNVs, SVs and MEI. At the end of this write-up we have included tables of the selected diagnostic variants and secondary variants, so as to provide more complete information than allowed in the submission file. In the comments fields of the submission file we have provided the gene name, mechanism, phenotype agreement, and confidence (see confidence definitions below).

Collection of SickKids challenge phenotypes and corresponding gene list

Clinical phenotype descriptions were collected from CAGI phenotype data for 25 children with suspected genetic disorders and also from the supplementary Table 4 of Stavropoulos, D.J., et al. (2016) [1], matched by genome id of these 25 children. We extracted 3306 genes corresponding to 243 phenotypes from the Human Phenotype Ontology-based database (HPO) (Build #102) [2] and the dbNSFP database (version 3.1a) [3]. The gene list for secondary variants was taken from the Table in the 2013 ACMG guidelines [4].

Annotation of VCF files and QC filters

The VCF files (including SNVs, Indels, CNVs, SVs and MEI) provided for this challenge from Complete Genomics assembly pipeline 2.5 were annotated using the Varant [5] tool, including region of occurrence (intron, exon, splice site or intergenic), observed minor allele frequencies (MAF), mutation type, predicted impact on protein function, and previously established associated phenotypes reported in ClinVar [6]. The RefGene [7] gene definition file was used for gene and transcript annotations in Varant. In addition, in-house scripts were written to further annotate the VCF files with HGMD [8] disease related variants and with dbscSNV [9] variants that potentially alter splicing.

We used high quality data for further analysis and different QC filters were used for different types of data as mentioned in Complete Genomics assembly pipeline 2.5. For SNVs, Indels and SVs, we used the 'PASS' filter; for CNVs, we used the ploidy score and CNV type score >30 for those segments where called ploidy is not equal to 2 and regions which are not hypervariable; for MEI data, we used 'sns95' as filter status and CGA_IS (measure of confidence that there is a mobile element insertion) > 15.

Diagnostic variants: identification and interpretation

A hierarchical scheme was used for identification of diagnostic variants, based on the strength of the evidence for disease relevance. All accepted high quality variants in the selected gene list with population frequency <5% in either 1000 genome data [10] or in the ExAC database [11]) were first categorized into ordered tiers as follows:

Category 1: Variants with HGMD annotation of either DM (disease-causing mutation) or DP (disease-associated polymorphism), and/or reported in ClinVar with pathogenic or likely pathogenic status.

Category 2: Nonsense mutation, direct splicing mutation disrupting either splice donor or acceptor site, frameshift or non-frameshift mutation, splice altering variant predicted by the dbscSNV database [9], and non-synonymous mutations predicted as damaging by one or more of SNPs3D [12], SIFT [13], PolyPhen-2[14] and CADD [15]. For inclusion of a non-synonymous variant in Category 2, at least 60% of these methods were required to return a prediction of deleterious. This threshold is based on a calibration against HGMD. (Note that for any given variant not all methods may return a result, hence the non-obvious cutoff).

Category 3: Non-synonymous mutations predicted as damaging by one or more of the above non-synonymous impact prediction methods, with the deleterious prediction agreement fraction < 0.6.

Category 4: Benign non-synonymous mutations (according to reporting of non-synonymous impact prediction methods).

Category 5: Variants annotated as close to a splice acceptor or splice donor site.

Category 6: Variants annotated as UTR and intronic.

Category 7: All CNV, SV and MEI variants overlapping with the selected gene list not included in an earlier category.

Except as noted below, only variants in Categories 1, 2 and 3 where accepted for final submission.

Variants from category 1, 2, and 3 were further categorized into ordered tiers according to their rarity in the population data.

Frequency bin 1: Novel mutations (not seen in 1000 genomes or ExAC).

Frequency bin 2: Variants with population frequency > 0 and <= 0.005.

Frequency bin 3: Variants with population frequency > 0.005 and <= 0.01.

Frequency bin 4: Variants with population frequency > 0.01 and < 0.05.

Generally, variants in lower frequency bins were preferred.

Binned variants were further filtered for an appropriate inheritance model using the OMIM inheritance pattern. Variants were assigned to autosomal dominant, autosomal

recessive, compound heterozygous, pseudo autosomal recessive, or X-linked recessive models. Some of variant genotypes had phase information and for these we checked for consistency with a compound heterozygous model, where appropriate. For compound heterozygous cases where one variant belonged to category 1, 2, and 3, we also considered variants in categories 4, 5, 6 and 7 to provide the second variant.

With these criteria, we found several Category 1, 2 and 3 variants for many individuals, but with the corresponding phenotype not an exact match for the phenotypes provided. In these cases we searched for partial overlap with the reported phenotypes and made a judgment call as to whether the variant could be relevant. We tagged these variants with three different confidence levels:

Probable Match: When the OMIM disease description for the gene matches the individual's phenotype.

Possible Match: When there is a partial overlap of the OMIM disease phenotypes with the the individual's phenotype.

Speculative Match: When the variant is unlikely to be causative, for example an SV in an intergenic region near an appropriate gene or completely within an intron of an appropriate gene, such that is no obvious mechanism of action.

Searching for Predictive Secondary variants

Here we followed the rules in Stavropoulos, D.J., et al. (2016) [1] to extract predictive secondary variants. We followed the same protocol as for diagnostic variants, described above, using inheritance models from the Table in the of 2013 ACMG guidelines [4].