

CAGI slide pack guidelines

I agree to my slides being used according to the following terms: *Stephen Meyn*

- Authors: please include THIS slide at the beginning of your presentation to indicate that you agree with reuse of your slides in CAGI presentations, according to the terms below, and digitally sign by writing your name above.
- To disseminate the CAGI results, I encourage all CAGI participants to remix these CAGI slides for inclusion in their oral and poster presentations
- CAGI participants may use or publish any content of these slides only in compliance with the CAGI Data Use Agreement
- CAGI participants may not cite or publish any content of these slides except with the written permission of the originating author or in the primary CAGI publications
- CAGI participants must acknowledge the original authors of these slides and CAGI 2016. Include the acknowledgement banner across the bottom
- CAGI participants must include the CAGI credits slide in their presentations
- Slides with a red slash (no remix) may not be included in a presentation unless all attendees are registered CAGI participants who signed the data use agreement
- Slides with a red X over the entire slide may not be used in any circumstances
- Slides with a blue slash (embargoed) may not be used unless the embargo is explicitly lifted
- Slides without slash or X may be tweeted unless flagged with the “No tweet” icon.
- CAGI participants using these slides must explain the “No tweet” icon, which means that the slide content should not be disseminated outside the presenting conference hall. If the presentation venue permits pervasive tweeting, you may not include these slides in your talk 
- Note for slide authors: we expect all CAGI participants to abide by these restrictions, and will make best effort to ensure they are followed. However, but bear in mind that we have limited means of enforcing them, and therefore the restrictions cannot be guaranteed.

CAGI 5: The SickKids Genome Clinic Challenge

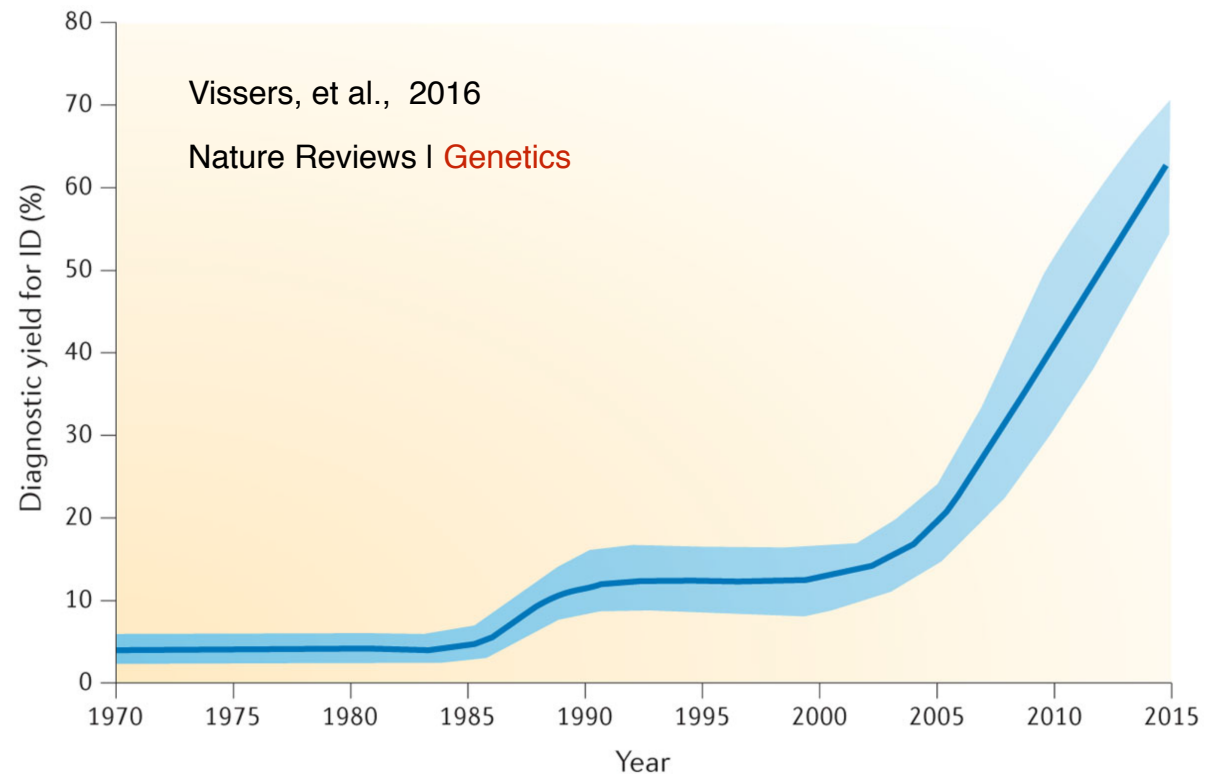
Genomes + phenotypic descriptions of 24 undiagnosed paediatric genetics patients



"So - WHICH GENES HAVE BEEN BOTHERING YOU?"

Identification of pathogenic variants by genome-wide analyses

Advances in molecular diagnosis have revolutionized clinical genetics



Analysis Method:

G-banded Karyotype

Microarray

Whole Exome Sequencing

Whole Genome Sequencing

First Clinical Use:

1960's

2005

2013

2014

Genomic features Detected:

~600-700

1,000,000+

50,000,000+

3,000,000,000+

Resolution:

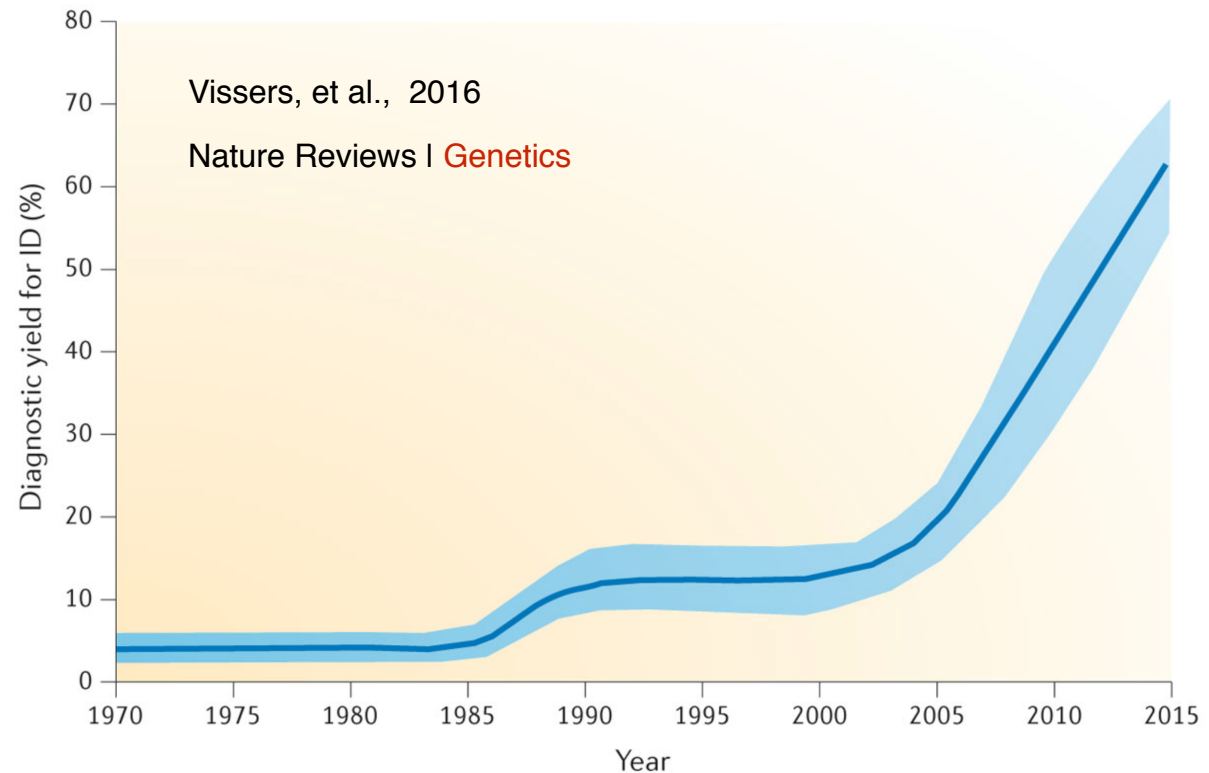
~5,000 kb

10-50 kb

single base

single base

Yet - the Diagnostic Revolution leaves most patients behind

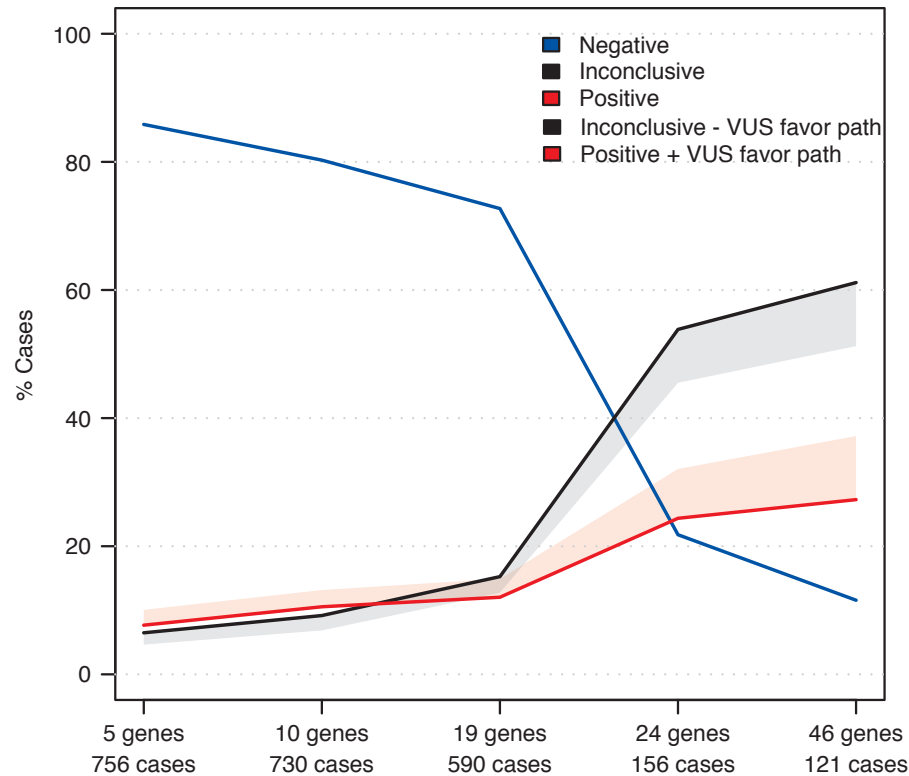


Why aren't the diagnostic rates for rare genetic disorders higher?

- Technical Deficiencies: Inability to sequence and align "difficult" genomic regions
- Analytical Challenges: Limited ability to correctly classify both coding and non-coding variants
- Most genome-wide analyses based on singleton exome sequencing
- Undiscovered disease genes
- Not all rare disorders are due to germ line genomic variants

Yet - the Diagnostic Revolution leaves most patients behind

clinical testing
for cardiomyopathy
genes



(Pugh, et al. 2014, Genet in Med)

Why aren't the diagnostic rates for rare genetic disorders higher?

- Technical Deficiencies: Inability to sequence and align "difficult" genomic regions
- Analytical Challenges: Limited ability to correctly classify both coding and non-coding variants
- Most genome-wide analyses based on singleton exome sequencing
- Undiscovered disease genes
- Not all rare disorders are due to germ line genomic variants

Diagnostic Variants - Challenges to validation

The generic child in the genetics clinic:

Phenotypic description may be information poor (developmental delay + dysmorphic features)

Vast number of potential genetic diagnoses, each of which is rare

Comparing variant frequencies in cases vs. controls not straight-forward

- few cases, impurity of control datasets

Segregation analyses not always possible

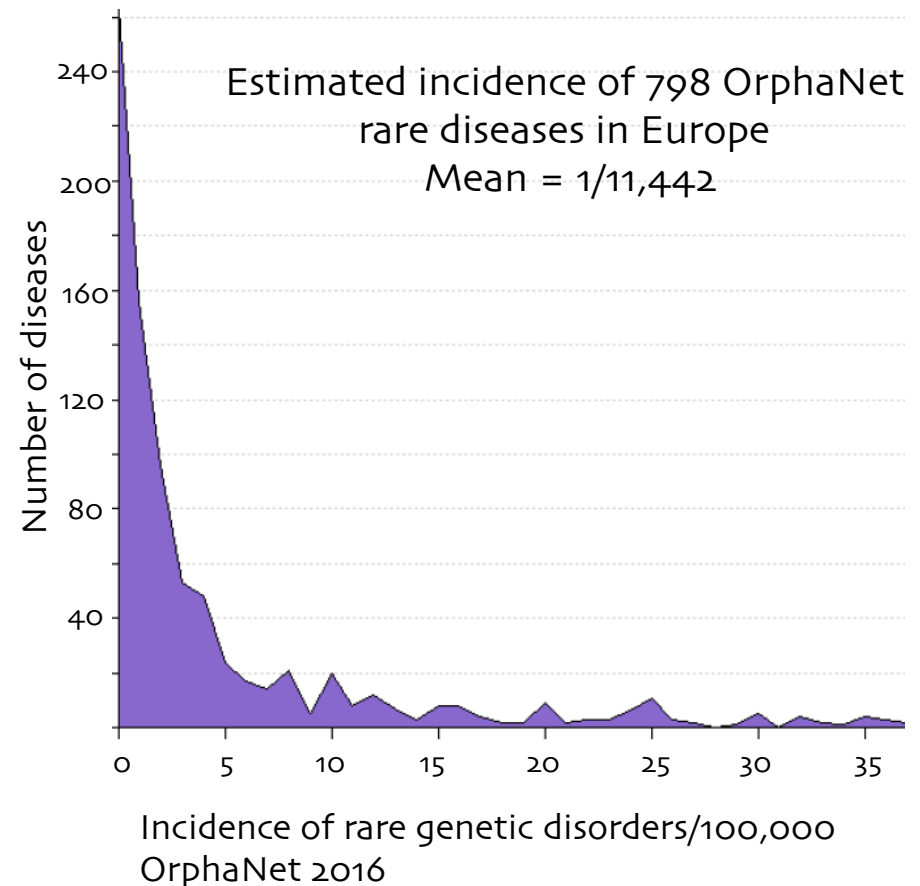
Disorders not always well characterized

Limited phenotypic description in databases

Functional validation rarely available

High reliance on plausible phenotypic fit as determined by clinical expert

Multiple genetic etiologies possible



The SickKids Genome Clinic Project

Designed to be a test bed for clinical research in Genomic Medicine:

Working definition: A new clinical paradigm in which knowledge gained from analyzing the informational content of an individual's genome is used to guide health care decisions throughout life, in order to anticipate, diagnose and manage disease.

Project Scope: Pilot the practice of paediatric genomic medicine

- Assess WGS as both a diagnostic and predictive test

- Develop and validate new bioinformatics tools for clinical genomics

- Discover new disease genes

- Determine clinical utility of pharmacogenomic and secondary MAVs in children

- Conduct health policy/health economics inquiries

- Investigate ethics of consent and psychosocial impact of genomic knowledge

- Develop new models of clinical care

Inclusion Criteria:

- <18 years of age, both parents available, adequate English

- Child is suspected of having a primary genetic disorder

- Current standard of care includes diagnostic genetic testing

- Working diagnosis is polygenic with incomplete clinical test sensitivity

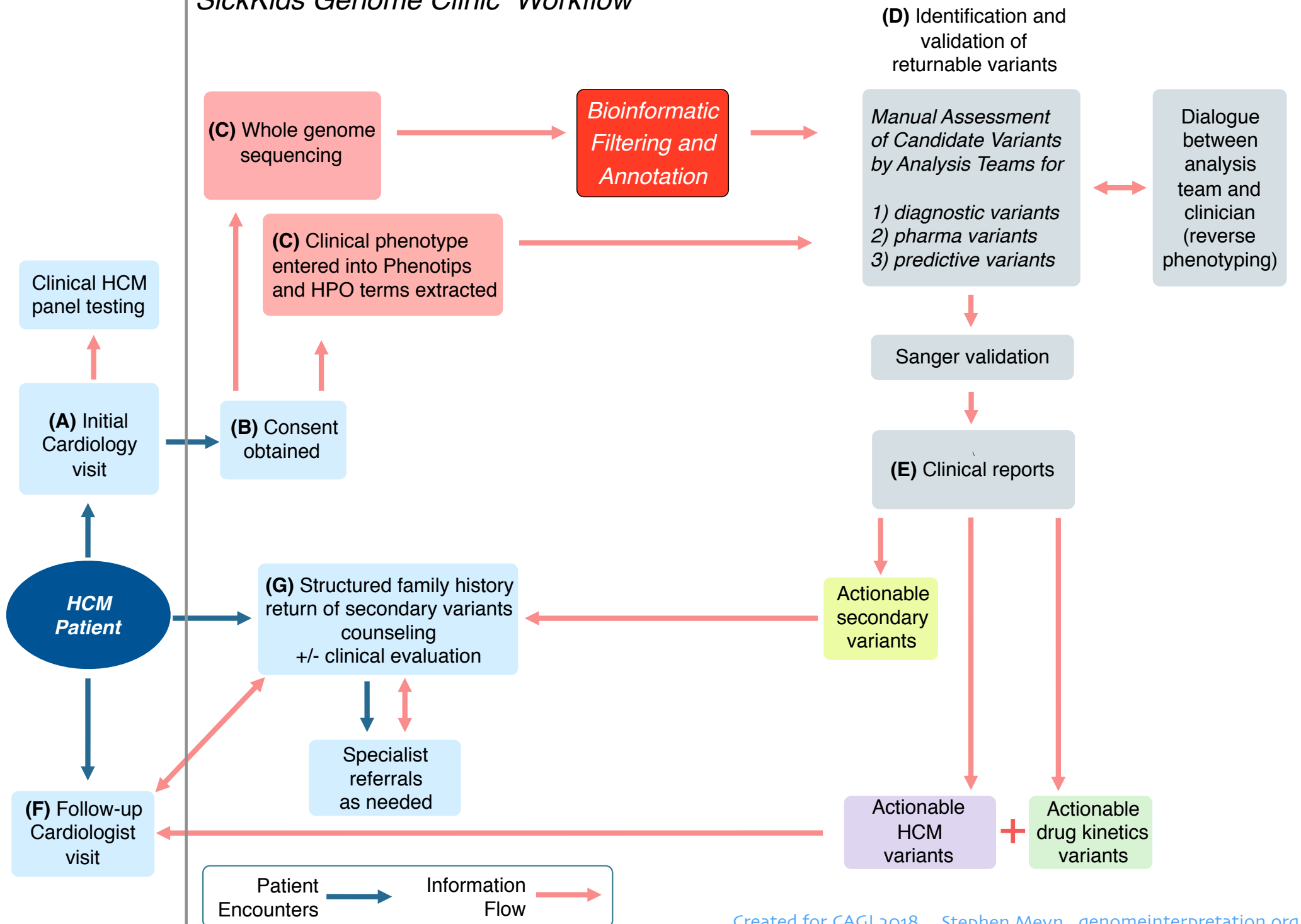
Created for CAGI 2018

Stephen Meyn

genomeinterpretation.org

For more information see Bowdin, et al., Hum Mutation (2014) and Bowdin, et al., Clin Genetics (2015)

SickKids Genome Clinic Workflow



SickKids Genome Clinic: Acquiring phenotypic data

Goal: Efficient/accurate collection of phenotypic data for database use

Challenges:

Most genotype/phenotype databases store little phenotypic information

Evocative phenotypic descriptions/free text difficult for machines to understand
e.g., - “still walking 25 years after onset”

The same phenotype can be described multiple ways:

e.g., - “generalized amyotrophy” vs. “generalized muscle atrophy”

Standard phenotype ontologies exist but are hard to use

- SNOMED, London Dysmorphology Database, Human Phenotype Ontology
- Human Phenotype Ontology: >10,000 terms
- data entry post clinic visit tedious and prone to errors

e.g., - clinicians’ descriptions of congenital malformation/anomaly:

- | | |
|---------------------------|-------------------------------------|
| • congenital anomaly | • cong. Mal |
| • congenital malformation | • cong. malfor |
| • congenital anamoly | • congenital malform |
| • congenital anomly | • congenital m. |
| • congential anomaly | • multiple congenital anomalies |
| • congential anomaly | • multiple congenital abnormalities |
| • Cong. m. | • multiple congenital abnormalities |

Acquiring phenotypic data via Phenotips

INFORMATICS

Human Mutation



PhenoTips: Patient Phenotyping Software for Clinical and Research Use

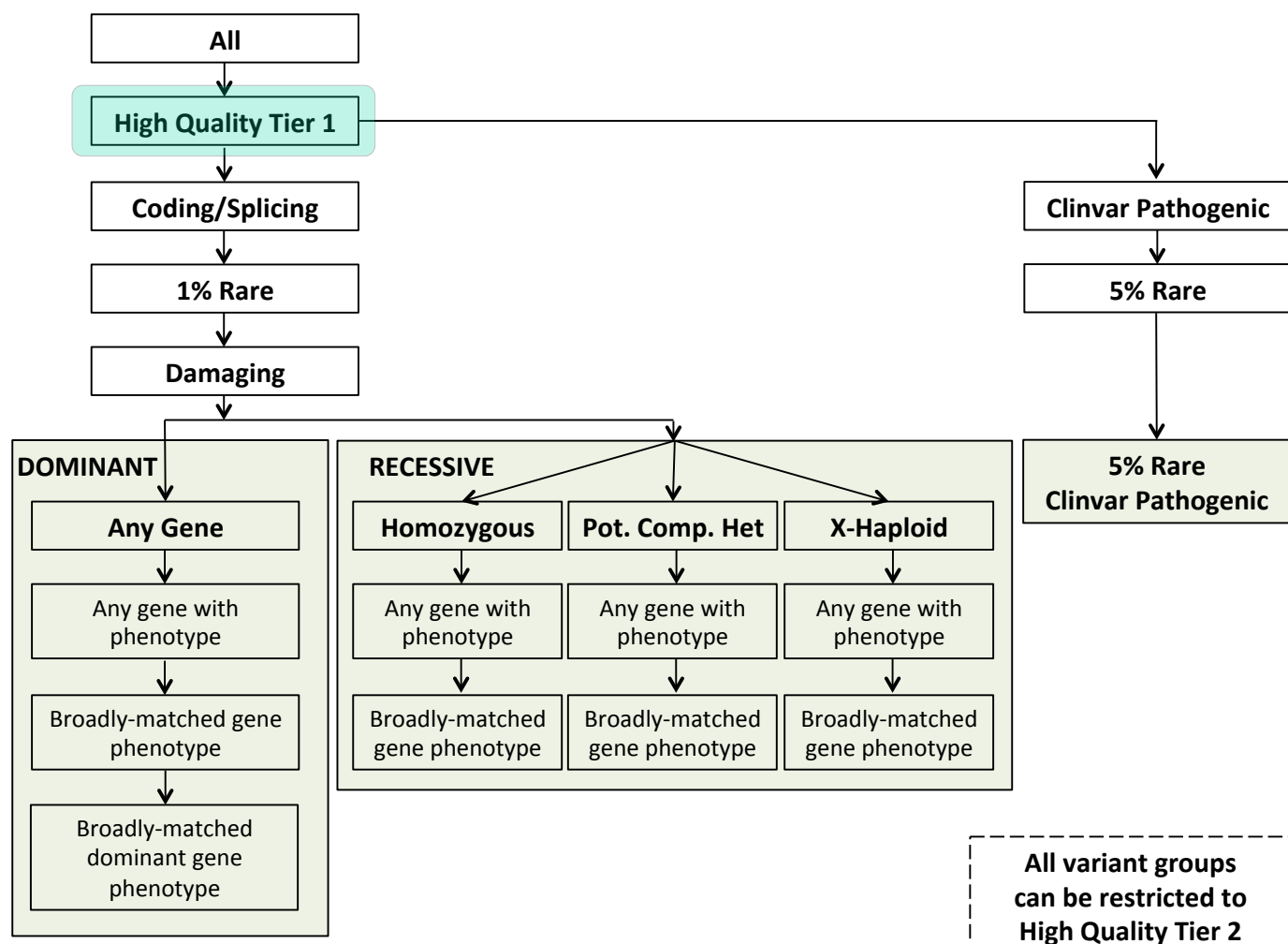
Marta Girdea,^{1,2} Sergiu Dumitriu,¹ Marc Fiume,¹ Sarah Bowdin,^{3,4} Kym M. Boycott,⁵ Sébastien Chénier,^{6†} David Chitayat,^{4,7} Hanna Faghfoury,^{8,9} M. Stephen Meyn,^{4,10} Peter N. Ray,⁶ Joyce So,^{8,11} Dimitri J. Stavropoulos,⁶ and Michael Brudno^{1,2,10*}
Human Mutation: 34: 1057-1064 (2013)

Goals: Make deep phenotyping at the bedside “faster than paper.”

- Predictive Search (with synonyms): Allows rapid data entry with HPO terms
Highly tolerate of misspellings
Selection of negatives
- Extensive quantitative measurements (e.g. height, weight, eye measurements)
 - automatic plotting of percentiles and selection of relevant phenotypic terms
- Built-in Exam Assistance: suggest related but unselected phenotypes, so none are missed
- Built-in Diagnosis Assistance: suggests potential differential Dx from OMIM

Bioinformatics Pipeline for Primary Diagnosis

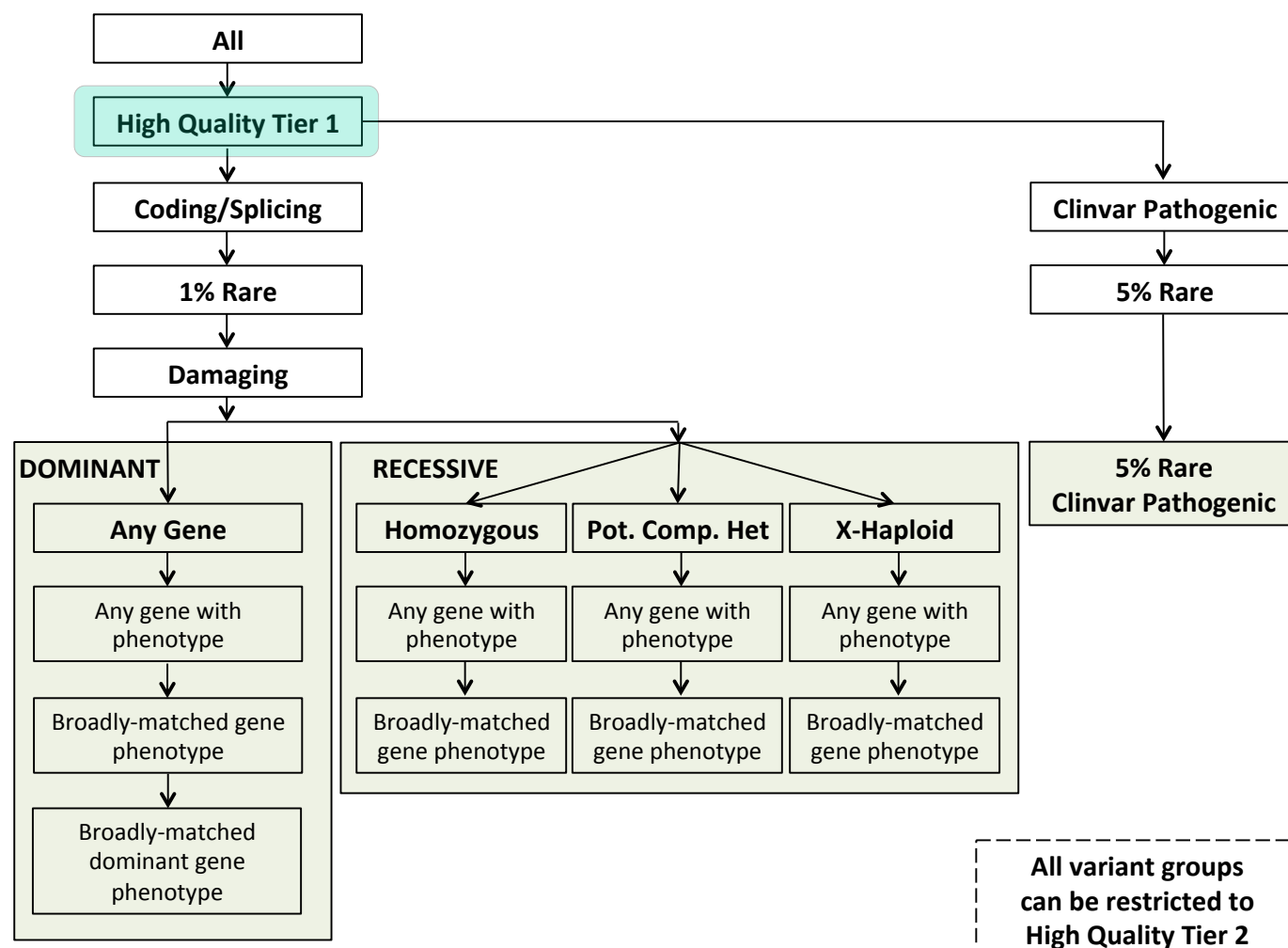
Primary Disorder Pipeline



Pipelines developed by:
Christian Marshall
Daniele Merico
Thomas Nalpathamkalam
Giovanna Pellecchia
Bhooma Thiruvahindrapuram

Bioinformatics Pipeline for Primary Diagnosis

Primary Disorder Pipeline



High Quality Tiers:

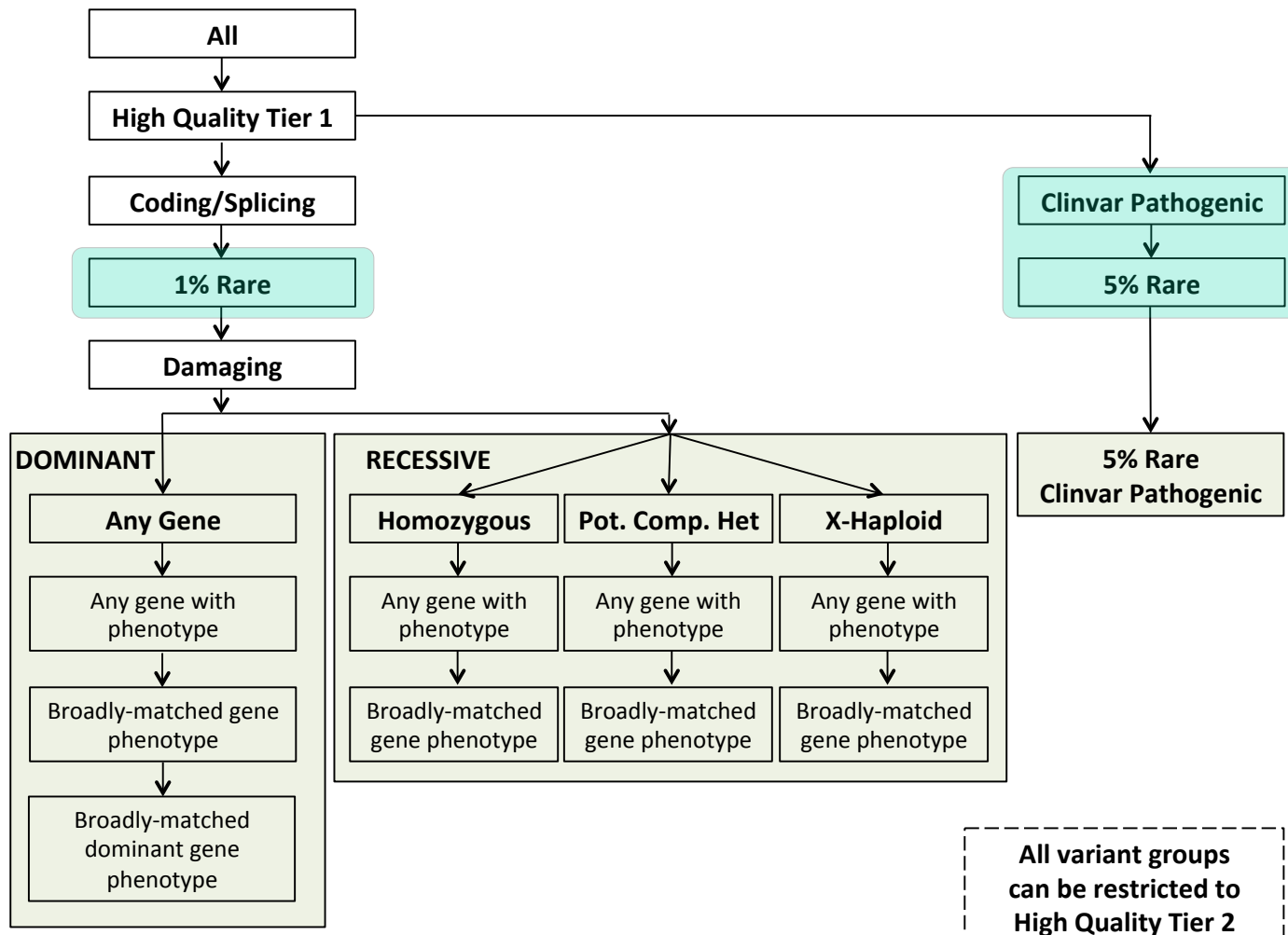
1 - default CG filters

2 - must also pass:
autosomal ploidy=2
diploid QC score
minimum depth >5
variant/ref read ratio

All variant groups
can be restricted to
High Quality Tier 2

Bioinformatics Pipeline for Primary Diagnosis

Primary Disorder Pipeline



Filtering for rare variants

must be <1% MAF in all of:
ExAc/Gnomad
1000 genomes
NHLBI-ESP 6500 exomes
CG diversity panel
TCAG reference genomes

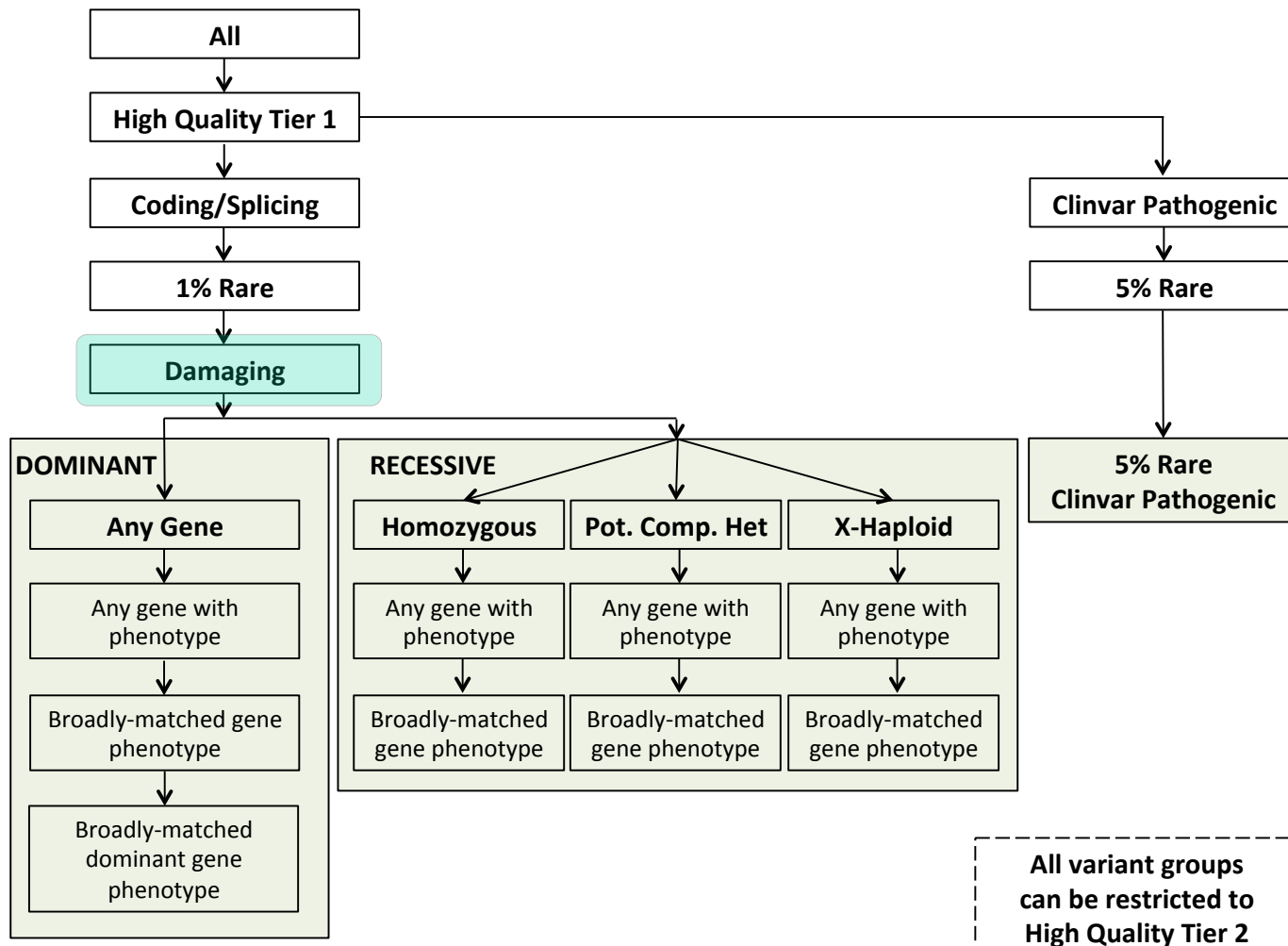
or

<5% MAF + pathogenic
in Clinvar or HGMD

All variant groups
can be restricted to
High Quality Tier 2

Bioinformatics Pipeline for Primary Diagnosis

Primary Disorder Pipeline



Putative damaging:

Loss of function

- frameshift
- splicing
- stop/gain

Missense

- minimal conservation plus at least one of:
 - High conservation
 - SIFT
 - Polyphen
 - Mutation Assessor

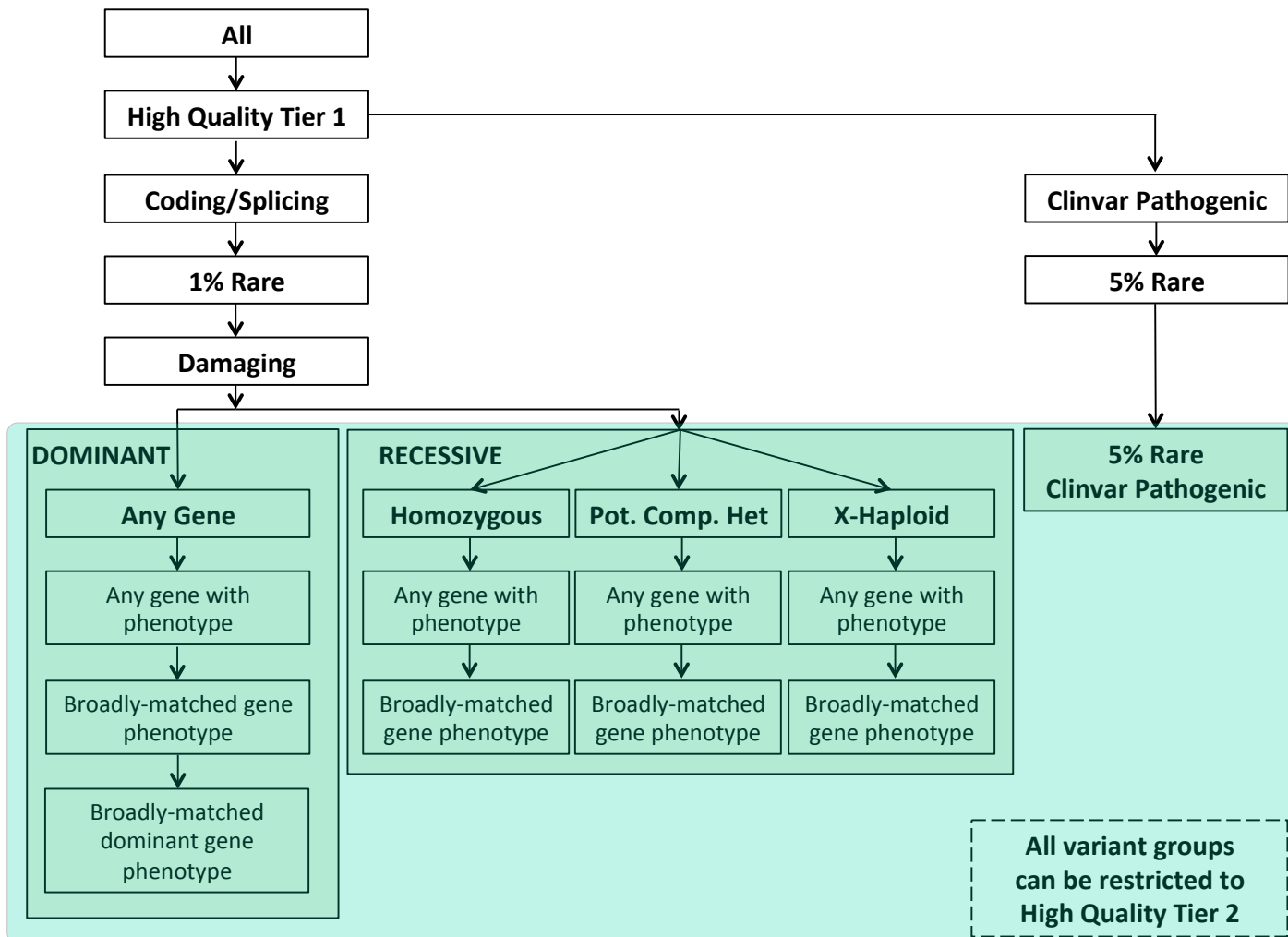
Non-frameshift in/del/sub

- mod/high conservation
- most not overlap dbSNP

All variant groups
can be restricted to
High Quality Tier 2

Bioinformatics Pipeline for Primary Diagnosis

Primary Disorder Pipeline



Variant Candidate Tiers:

0 - No associated phenotype

1 - phenotype exists
no match to patient

2 - matched phenotype
(HPO, MGI/MPO)

3 - matched phenotype
and known dominant

The CAGI 5 cohort: Panel Sequencing vs. singleton WGS

Demographics:

103 consecutive cases tested by clinical multi-gene panels

- patients from 10 specialty clinics
- 8 years mean age
- 61% of European ancestry
- 49.5% female/50.5% male
- 8.7% consanguineous parents
- 20% had positive family histories
- 57% Developmental delay

Genotyping:

Singleton WGS (Illumina)

- 37x average coverage
- 98% >10x coverage
- 93% > 20x coverage
- ~3,900,000 variants/genome

Referral clinic:		Total cohort (n=103)		Diagnosed cohort (n=42)	
		n	(%)	n	(%)
	Metabolics	45	(43.7)	20	(47.6)
	Ophthalmology	24	(23.3)	13	(31.0)
	Joint laxity/ Hypermobility	15	(14.6)	0	(0.0)
	Rheumatology	5	(4.9)	0	(0.0)
	Respirology	4	(3.9)	1	(2.4)
	Neurology	3	(2.9)	3	(7.1)
	Cardiology	2	(1.9)	1	(2.4)
	Development	2	(1.9)	2	(4.8)
	Nephrology	2	(1.9)	2	(4.8)
	Endocrinology	1	(1.0)	0	(0.0)

Diagnostic Rates:

- 24% for conventional panel tests
- 41% for WGS
- 6 cases identifiable by WGS but not by WES
 - e.g., deep intronic variant, snRNA, CNV

The CAGI 5 cohort: Panel Sequencing vs. singleton WGS

Demographics:

103 consecutive cases tested by clinical multi-gene panels

- patients from 10 specialty clinics
- 8 years mean age
- 61% of European ancestry
- 49.5% female/50.5% male
- 8.7% consanguineous parents
- 20% had positive family histories
- 57% Developmental delay

Conventional Testing:

- median of 3 genetic tests
- 44% had clinical CMA testing
- 81.5% had WES testing

Referral clinic:		Total cohort (n=103)		Diagnosed cohort (n=42)	
		n	(%)	n	(%)
	Metabolics	45	(43.7)	20	(47.6)
	Ophthalmology	24	(23.3)	13	(31.0)
	Joint laxity/ Hypermobility	15	(14.6)	0	(0.0)
	Rheumatology	5	(4.9)	0	(0.0)
	Respirology	4	(3.9)	1	(2.4)
	Neurology	3	(2.9)	3	(7.1)
	Cardiology	2	(1.9)	1	(2.4)
	Development	2	(1.9)	2	(4.8)
	Nephrology	2	(1.9)	2	(4.8)
	Endocrinology	1	(1.0)	0	(0.0)

Diagnostic Rates:

- 24% for conventional panel tests
- 41% for WGS
- 6 cases identifiable by WGS but not by WES
 - e.g., deep intronic variant, snRNA, CNV

The CAGI 5 cohort: Panel Sequencing vs. singleton WGS

Demographics:

103 consecutive cases tested by clinical multi-gene panels

- patients from 10 specialty clinics
- 8 years mean age
- 61% of European ancestry
- 49.5% female/50.5% male
- 8.7% consanguineous parents
- 20% had positive family histories
- 57% Developmental delay

Conventional Testing:

- median of 3 genetic tests
- 44% had clinical CMA testing
- 81.5% had WES testing

18 diagnoses made only through WGS:

Diagnostic categories (n=18*)	# of cases
AD	5
AR	8
X-linked	5
mtDNA variant	3
Partial dx	5
Research candidate genes	2
Types of diagnostic variants	n=18
Exonic SNVs	12
Deep intronic variants	2
Splice variants	2
snRNA	1
Small CNV	1

Limitations

- Only “virtual exome” and “virtual microarray”
- Large In/Dels and small CNVs are hard to interpret due to lack of population data
- Singleton analyses

Hypothesis-driven vs. Hypothesis-free genetic testing

“hypothesis-free” genome-wide testing increased diagnostic yield over targeted testing

59 microarray cohort patients also underwent hypothesis-driven molecular testing

- average of 2 hypothesis driven tests/patient (range 1-13)
- hypothesis-driven testing diagnosed 6 of these 59 patients (10%)
- hypothesis-free WGS diagnosed 23 of these 59 patients (39%)

103 panel cohort patients underwent hypothesis-driven molecular testing

- median of 19 genes examined by hypothesis-driven testing
- hypothesis-driven testing diagnosed 25 of these 103 patients (24.3%)
- hypothesis-free WGS diagnosed 42 of these 103 patients (40.8%)

CAGI 5: The SickKids Genome Clinic Challenge

The Cohort:

Unlinked genomes and phenotypic descriptions for 24 undiagnosed children from the 103 patient SickKids Genome Clinic Panel Sequencing Cohort

- 6 ophthalmologic disorders (conventional analysis dx rate: 54%)
- 7 neurogenetic/metabolic disorders (conventional analysis dx rate: 48%)
- 11 connective tissue disorders (conventional analysis dx rate: 0%)

The Challenge:

Predict which class of disease is associated with each genome - $p[0,1]$

Match genomes to specific clinical descriptions - $p[0,1]$

Identify diagnostic variants underlying the predictions

Identify predictive secondary variants conferring high risk of additional genetic disorder

CAGI 5: The SickKids Genome Clinic Challenge

The Cohort:

Unlinked genomes and phenotypic descriptions for 24 undiagnosed children from the 103 patient SickKids Genome Clinic Panel Sequencing Cohort

- 6 ophthalmologic disorders (conventional analysis dx rate: 54%)
- 7 neurogenetic/metabolic disorders (conventional analysis dx rate: 48%)
- 11 connective tissue disorders (conventional analysis dx rate: 0%)

The Teams:

tcsTeam - Sadhna Rana (Tata Consultancy Services, India)

BolognaBiocomputing - Rita Casadio (University of Bologna, Italy)

Lancet - Simon Fishilevich (Weizmann Institute of Science, Israel)

moult_lab - Lipika Pal (University of Maryland, United States)

Lichtarge Lab - Panagiotis Katsonis (Baylor College of Medicine, United States)

ANA - Aashish Adhikari (University of California, Berkeley, United States)

JP SickKids5 - Jennifer Poitras (QIAGEN, United States)

Mooney-Radivojac - Rafael Guerrero (Indiana University, United States)

CAGI 5: The SickKids Genome Clinic Challenge

Methods

Tata Consultancy Service (Sadhna Rana):

Phenotyping: Text mining for HPO terms (TPX software), followed by manual QC

Gene prioritization: Ranked using HANRD (Heterogeneous Association Network for Rare Diseases) network, TPXRD (PubMed text mining)

Variant assessment: Ranked by MAF, evolutionary conservation, in silicon prediction, ClinVar/HDMD

BolognaBiocomputing (Rita Casadio, University of Bologna):

Phenotyping: Reorganized phenotypes using HPO and OMIM classifications

Gene prioritization: Used eDGAR: a database of Disease-Gene Associations with annotated Relationships

Variant assessment: Ranked by known or predicted pathogenicity using SNPs&GO and UniProt

Lancet (Simon Fishilevich, Weizmann Institute):

Phenotyping: Manual curation and transformation into Boolean expressions

Gene prioritization: VarElect software + GeneCards/MalaCards databases

Variant assessment: VarElect software then manual QC

moult_lab (Lipika Pal, Baltimore):

Phenotyping: Phenotypic descriptions transformed into phenotypic scores

Gene Prioritization: Ranking driven by matches between ranked variants and phenotypic scores

Variant assessment: Filtered for quality, MAF, reported and predicted pathogenicity

Lichtarge Lab (Panagiotis Katsonis, Baylor):

Phenotyping: No special processing/filtering

Gene prioritization: Ranked primarily by evolutionary relationship to phenotype

Variant assessment: Ranked primarily by evolutionary burden

Additional Factors: Used sex and ethnic origin information from genomic data

CAGI 5: The SickKids Genome Clinic Challenge

Methods

ANA (Aashish Adhikari, Brenner Lab, Berkeley):

Phenotyping: No special processing/filtering

Gene prioritization: Genes with rare variants in associated with patients' phenotypes in Monarch database

Variant assessment: MAF filtering of protein coding variants

Additional Factors: Used sex and ethnic origin information from genomic data

ANA (Aashish Adhikari, Brenner Lab, Berkeley) Late:

Phenotyping: No special processing/filtering

Gene prioritization: Genes with rare variants in associated with patients' phenotypes encyclopedia of rare disease annotation (eRAM) and in Monarch database

Variant assessment: MAF filtering of protein coding variants

Additional Factors: Used sex and ethnic origin information from genomic data

JP SickKids5 (Jennifer Poitras, Qiagen):

Commercial Software (Inguinity Variant Analysis)

Phenotyping: No special processing/filtering

Gene Prioritization: Genes known to be associated with patients' phenotypes

Variant assessment: Filtered in quality, MAF, and known pathogenicity

Mooney/Radivojac (Rafael Guerrero, Indiana University):

Phenotyping: No special processing/filtering

Gene prioritization: Built lists from disease genes associated with specific phenotypes + their interactors

Variant assessment: Focused on ranking coding variants (Mutpred2 and Mutpred-LOF)

Additional Factors: Used sex and ethnic origin information from genomic data

CAGI 5: The SickKids Genome Clinic Challenge

The Challenge: Predict which disease category is associated with each genome

Overall Results

Clin Phenotype	Diagnostic Hit Rate	Correct Classifications of CAGI Unsolved Cases
Ophthalmologic	13/20 pts (65%)	6 pts - 14 matches
Neurologic	23/48 pts (48%)	7 pts - 14.5 matches
Connective Tissue	0/15 pts (0%)	11 pts - 47.5 matches

CAGI 5: The SickKids Genome Clinic Challenge

The Challenge: Predict which disease category is associated with each genome

Overall Results

Clin Phenotype	Diagnostic Hit Rate	Correct Classifications of CAGI Unsolved Cases
Ophthalmologic	13/20 pts (65%)	6 pts - 14 matches
Neurologic	23/48 pts (48%)	7 pts - 14.5 matches
Connective Tissue	0/15 pts (0%)	11 pts - 47.5 matches

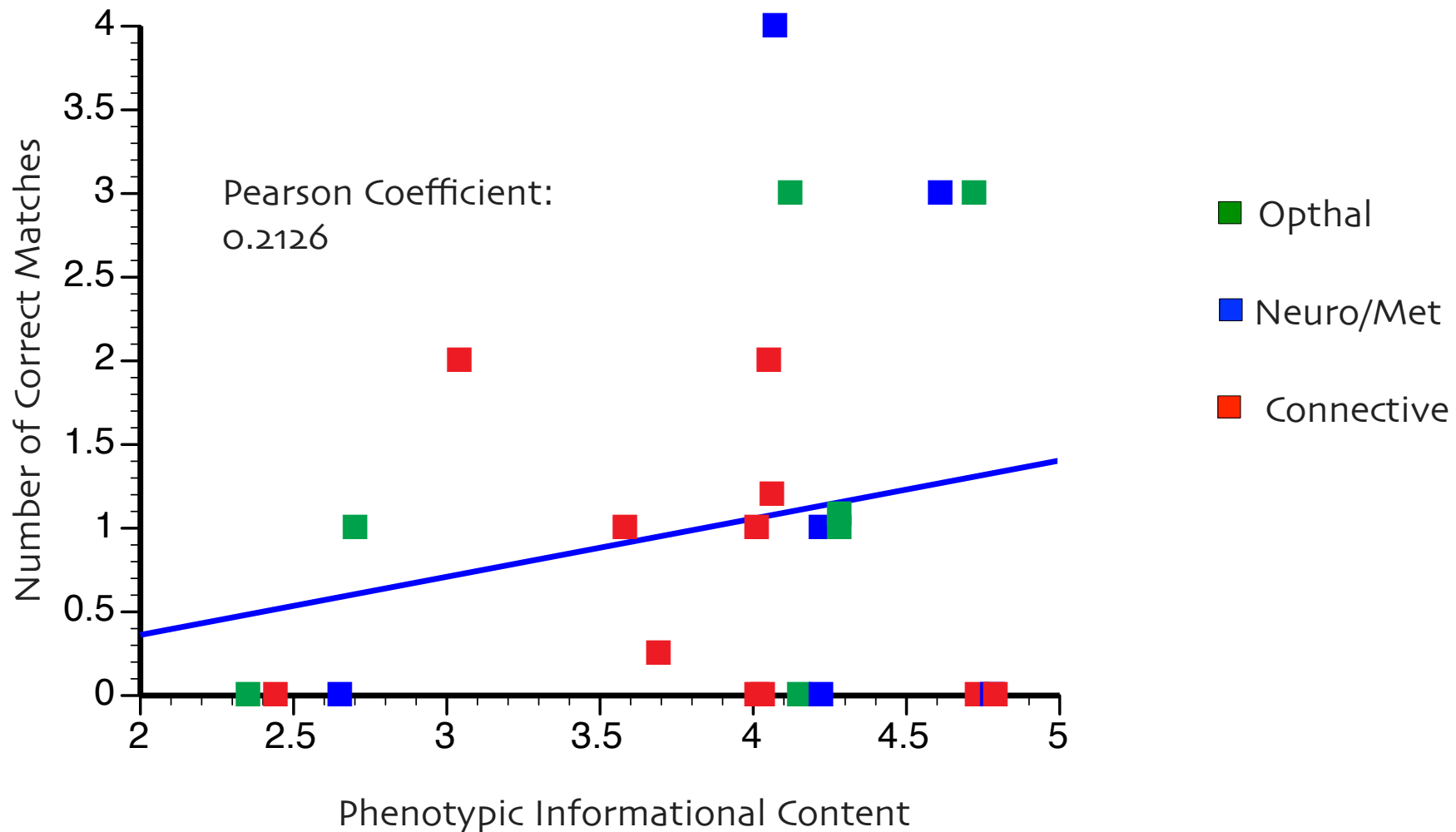
Team Results

Team	1st	1st or 2nd	2nd	2nd or 3rd	3rd
TCS Team	8	2		10	
Bologna Biocomputing	3			9	
Lancet	9		9		6
Moult Lab	11	1	2	10	
Lichtarge Lab	11	1	8		4
ANA (Berkeley)	8		10		6
ANA Late	9	2	6		7
JP SickKids5	6	2		10	
Mooney/Radivojac	11		6		7

CAGI 5: The SickKids Genome Clinic Challenge

The Challenge: Predict which patient is associated with each genome

Overall Results: Do information rich phenotypes improve matching of patients to genomes?



CAGI 5: The SickKids Genome Clinic Challenge

The Challenge: Predict which patient is associated with each genome

Team Results:

Team	1st	1st or 2nd	2nd	Sex Mismatch?
TCS Team	1	0	0	2
Bologna Biocomputing	1.2 (2)	0	0	0
Lancet	1	0	1	1
Moult Lab	5.08 (6)	0	1.5 (4)	0
Lichtarge Lab	4.125 (5)	0	2	0
ANA (Berkeley)	1.25 (3)	0	0	didn't exclude
ANA Late	1.25 (2)	0	0.5 (1)	didn't exclude
JP SickKids5	1	0	0	3
Mooney Radivojac	6	0	6	2

CAGI 5: The SickKids Genome Clinic Challenge

The Challenge: Predict which variant is associated with a specific patient

Summary Results

Genome	Patient	Category	Group	Gene	Diseases	Plausible phenotypic match?	Variants	Max MAF
WGS-NGS-007	X	eye	Group 7	PROM1	Somatic mutations cause Sturge Weber syndrome	unlikely	4:16024957, 9:80647059:GC:TT	both novel
	X	eye	Group 8	FN1	Glomerulopathy, fibronectin deficiency, Spondylometaphyseal dysplasia	unlikely	2:216251538:G:A	0.78%
WGS-NGS-017	H	eye	Group 4	ABCA4	retinitis pigmentosa, rod-cone dystrophy and other eye disorders(AR)	Possible	1:94568686:C:T, 1:94470320:C:T	0.3% Euro
WGS-NGS-030	R	neuro	Group 6	A2ML1	Noonan-like malformation syndrome	Very unlikely	12:9013783:C:T, 12:9020600:G:A,	4% & 1.8% in Africans
WGS-NGS-056	N	connective	Group 4	FBN1	Marfan syndrome	Possible (Marfanoid)	15:48700642:T:C, 15:48849792:C:A	1x Gnomad/Novel
			Group 8	TNXB	AR Ehlers-Danlos like syndrome	Unlikely	6:32012817:C:T	12% in ASH
WGS-NGS-067	M	eye	Group1	TYR	AR ocular-cutaneous albinism	Unlikely	11:89017973:C:T	1% in Finns
WGS-NGS-068	J	neuro	Group 3	LMNA	progeria, muscular dystrophy, lipodystrophy, and others	Possible	1:156105054:G:T	novel
			Group 6	WFS1	Wolfram syndrome	Partial Fit	4:6304086:C:T,	0.01%
			Group 8	ATM	Ataxia telangiectasia, Cancer	Very unlikely	11:108196896:C:T	3% in ASH
WGS-NGS-078	V	connective	Group 2	COL1A2	Ehlers-Danlos syndrome, Osteogenesis imperfecta	Possible (EDS III)	7:94049587:C:T	2x in Gnomad
			Group 6 late	MYO1E	Glomerulosclerosis, focal segmental, 6	Very unlikely	15:59443263:CGTGCACTT:C, 15:59443263:CGTGCACTT:C	1st - 6% MAF in Finns
WGS-NGS-093	F	connective	Group 4	EP300	Rubinstein Taybi syndrome	Unlikely	22:41574743:A:G	novel
			Group 8	CACNA1C	Brugada syndrome	Very unlikely	12:2791205:A:G	99% in Europeans
WGS-NGS-095	C	eye	Group 2	no gene listed			no variant listed	
			Group 4	USHA2	Usher syndrome	Possible, not ideal	1:215953583:A:G, 1:215964830:T:G	0.49% is East Asians, novel
WGS-NGS-097	D	connective	Group 6 late	TNXB	Ehlers-Danlos syndrome - Classic	Possible	6:32064371:A:G, 6:32064371:A:G	0.23% in East Asians
WGS-NGS-099	B	neuro	Group 4	PIGT	Multiple congenital anomalies-hypotonia-seizures syndrome 3 - AR	Unlikely	20:44044789:C:T	1 x in Gnomad
			Group 6	HSD17B4	Perrault syndrome (seizures +)	Unlikely	5:118832244:C:G,	7% in Africans
			Group 8	SLC25A22	Epileptic encephalopathy, early infantile, 3	Possible	11:793588:G:A	12% in Europeans
WGS-NGS-102	A	connective	Group 8	COL5A2	Ehlers-Danlos syndrome - Classic	Possible	2:189974958:G:T	97% in Europeans

CAGI 5: The SickKids Genome Clinic Challenge

The Challenge: Predict which variant is associated with a specific patient

Summary Results

Genome	Patient	Category	Group	Gene	Diseases	Plausible phenotypic match?	Variants	Max MAF
WGS-NGS-007	X	eye	Group 7	PROM1	Somatic mutations cause Sturge Weber syndrome	unlikely	4:16024957, 9:80647059:GC:TT	both novel
	X	eye	Group 8	FN1	Glomerulopathy, fibronectin deficiency, Spondylometaphyseal dysplasia	unlikely	2:216251538:G:A	0.78%
WGS-NGS-017	H	eye	Group 4	ABCA4	retinitis pigmentosa, rod-cone dystrophy and other eye disorders(AR)	Possible	1:94568686:C:T, 1:94470320:C:T	0.3% Euro
WGS-NGS-030	R	neuro	Group 6	A2ML1	Noonan-like malformation syndrome	Very unlikely	12:9013783:C:T, 12:9020600:G:A,	4% & 1.8% in Africans
WGS-NGS-056	N	connective	Group 4	FBN1	Marfan syndrome	Possible (Marfanoid)	15:48700642:T:C, 15:48849792:C:A	1x Gnomad/Novel
			Group 8	TNXB	AR Ehlers-Danlos like syndrome	Unlikely	6:32012817:C:T	12% in ASH
WGS-NGS-067	M	eye	Group1	TYR	AR ocular-cutaneous albinism	Unlikely	11:89017973:C:T	1% in Finns
WGS-NGS-068	J	neuro	Group 3	LMNA	progeria, muscular dystrophy, lipodystrophy, and others	Possible	1:156105054:G:T	novel
			Group 6	WFS1	Wolfram syndrome	Partial Fit	4:6304086:C:T,	0.01%
			Group 8	ATM	Ataxia telangiectasia, Cancer	Very unlikely	11:108196896:C:T	3% in ASH
WGS-NGS-078	V	connective	Group 2	COL1A2	Ehlers-Danlos syndrome, Osteogenesis imperfecta	Possible (EDS III)	7:94049587:C:T	2x in Gnomad
			Group 6 late	MYO1E	Glomerulosclerosis, focal segmental, 6	Very unlikely	15:59443263:CGTGCACTT:C, 15:59443263:CGTGCACTT:C	1st - 6% MAF in Finns
WGS-NGS-093	F	connective	Group 4	EP300	Rubinstein Taybi syndrome	Unlikely	22:41574743:A:G	novel
			Group 8	CACNA1C	Brugada syndrome	Very unlikely	12:2791205:A:G	99% in Europeans
WGS-NGS-095	C	eye	Group 2	no gene listed			no variant listed	
			Group 4	USHA2	Usher syndrome	Possible, not ideal	1:215953583:A:G, 1:215964830:T:G	0.49% is East Asians, novel
WGS-NGS-097	D	connective	Group 6 late	TNXB	Ehlers-Danlos syndrome - Classic	Possible	6:32064371:A:G, 6:32064371:A:G	0.23% in East Asians
WGS-NGS-099	B	neuro	Group 4	PIGT	Multiple congenital anomalies-hypotonia-seizures syndrome 3 - AR	Unlikely	20:44044789:C:T	1 x in Gnomad
			Group 6	HSD17B4	Perrault syndrome (seizures +)	Unlikely	5:118832244:C:G,	7% in Africans
			Group 8	SLC25A22	Epileptic encephalopathy, early infantile, 3	Possible	11:793588:G:A	12% in Europeans
WGS-NGS-102	A	connective	Group 8	COL5A2	Ehlers-Danlos syndrome - Classic	Possible	2:189974958:G:T	97% in Europeans

Pt J

GROWTH PARAMETERS

Weight for age
Decreased body weight (<-2SD)
Stature for age
Short stature (<-2SD)
Delayed puberty

EYE DEFECTS

Ptosis
Retinal dystrophy
Corneal dystrophy

CARDIOVASCULAR

Sinus tachycardia

MUSCULOSKELETAL

Scoliosis
Mild
Osteopenia
Pes planus
Hallux valgus
Delayed skeletal maturation
Vertebral compression fractures
Myopathy
Ragged-red muscle fibers
Abnormal mitochondria in muscle tissue

GENITOURINARY

Stage 5 chronic kidney disease
Renal transplant in 2012

BEHAVIOR, COGNITION AND DEVELOPMENT

NO Global developmental delay

Specific learning disability

Some learning difficulties,

NEUROLOGICAL

Ataxia
Morphological abnormality of the CNS
Brain MRI mild cerebellar volume loss
Guillain-Barre syndrome in 2008

BLOOD AND BLOOD-FORMING TISSUES

Aplastic anemia
METABOLISM/HOMEOSTASIS
Renal tubular acidosis

CAGI 5: The SickKids Genome Clinic Challenge

The Challenge: Predict which variant is associated with a specific patient

Summary Results

Genome	Patient	Category	Group	Gene	Diseases	Plausible phenotypic match?	Variants	Max MAF
WGS-NGS-007	X	eye	Group 7	PROM1	Somatic mutations cause Sturge Weber syndrome	unlikely	4:16024957, 9:80647059:GC:TT	both novel
	X	eye	Group 8	FN1	Glomerulopathy, fibronectin deficiency, Spondylometaphyseal dysplasia	unlikely	2:216251538:G:A	0.78%
WGS-NGS-017	H	eye	Group 4	ABCA4	retinitis pigmentosa, rod-cone dystrophy and other eye disorders(AR)	Possible	1:94568686:C:T, 1:94470320:C:T	0.3% Euro
WGS-NGS-030	R	neuro	Group 6	A2ML1	Noonan-like malformation syndrome	Very unlikely	12:9013783:C:T, 12:9020600:G:A,	4% & 1.8% in Africans
WGS-NGS-056	N	connective	Group 4	FBN1	Marfan syndrome	Possible (Marfanoid)	15:48700642:T:C, 15:48849792:C:A	1x Gnomad/Novel
			Group 8	TNXB	AR Ehlers-Danlos like syndrome	Unlikely	6:32012817:C:T	12% in ASH
WGS-NGS-067	M	eye	Group1	TYR	AR ocular-cutaneous albinism	Unlikely	11:89017973:C:T	1% in Finns
WGS-NGS-068	J	neuro	Group 3	LMNA	progeria, muscular dystrophy, lipodystrophy, and others	Possible	1:156105054:G:T	novel
			Group 6	WFS1	Wolfram syndrome	Partial Fit	4:6304086:C:T,	0.01%
			Group 8	ATM	Ataxia telangiectasia, Cancer	Very unlikely	11:108196896:C:T	3% in ASH
WGS-NGS-078	V	connective	Group 2	COL1A2	Ehlers-Danlos syndrome, Osteogenesis imperfecta	Possible (EDS III)	7:94049587:C:T	2x in Gnomad
			Group 6 late	MYO1E	Glomerulosclerosis, focal segmental, 6	Very unlikely	15:59443263:CGTGCACTT:C, 15:59443263:CGTGCACTT:C	1st - 6% MAF in Finns
WGS-NGS-093	F	connective	Group 4	EP300	Rubinstein Taybi syndrome	Unlikely	22:41574743:A:G	novel
			Group 8	CACNA1C	Brugada syndrome	Very unlikely	12:2791205:A:G	99% in Europeans
WGS-NGS-095	C	eye	Group 2	no gene listed			no variant listed	
			Group 4	USHA2	Usher syndrome	Possible, not ideal	1:215953583:A:G, 1:215964830:T:G	0.49% is East Asians, novel
WGS-NGS-097	D	connective	Group 6 late	TNXB	Ehlers-Danlos syndrome - Classic	Possible	6:32064371:A:G, 6:32064371:A:G	0.23% in East Asians
WGS-NGS-099	B	neuro	Group 4	PIGT	Multiple congenital anomalies-hypotonia-seizures syndrome 3 - AR	Unlikely	20:44044789:C:T	1 x in Gnomad
			Group 6	HSD17B4	Perrault syndrome (seizures +)	Unlikely	5:118832244:C:G,	7% in Africans
			Group 8	SLC25A22	Epileptic encephalopathy, early infantile, 3	Possible	11:793588:G:A	12% in Europeans
WGS-NGS-102	A	connective	Group 8	COL5A2	Ehlers-Danlos syndrome - Classic	Possible	2:189974958:G:T	97% in Europeans

Pt J

GROWTH PARAMETERS

Weight for age
Decreased body weight (<-2SD)
Stature for age
Short stature (<-2SD)
Delayed puberty
EYE DEFECTS
Ptosis
Retinal dystrophy
Corneal dystrophy
CARDIOVASCULAR
Sinus tachycardia

MUSCULOSKELETAL

Scoliosis
Mild
Osteopenia
Pes planus
Hallux valgus
Delayed skeletal maturation
Vertebral compression fractures
Myopathy
Ragged-red muscle fibers
Abnormal mitochondria in muscle

GENITOURINARY

Stage 5 chronic kidney disease

Renal transplant in 2012

BEHAVIOR, COGNITION AND DEVELOPMENT

NO Global developmental delay

Specific learning disability

Some learning difficulties,

NEUROLOGICAL

Ataxia

Morphological abnormality of the CNS

Brain MRI mild cerebellar volume loss

Guillian-Barre syndrome in 2008

BLOOD AND BLOOD-FORMING TISSUES

Aplastic anemia

METABOLISM/HOMEOSTASIS

Renal tubular acidosis

LNMA - HG progeria, Emery Dreyfus MD, CMT 2B1,
Mandibuloacral dysplasia, lipodystrophy, cardiomyopathy:
Short stature, lipodystrophy, scleroderma,
decreased joint mobility, osteolysis, cardiomyopathy,
myopathy, osteolysis

CAGI 5: The SickKids Genome Clinic Challenge

The Challenge: Predict which variant is associated with a specific patient

Summary Results

Genome	Patient	Category	Group	Gene	Diseases	Plausible phenotypic match?	Variants	Max MAF
WGS-NGS-007	X	eye	Group 7	PROM1	Somatic mutations cause Sturge Weber syndrome	unlikely	4:16024957, 9:80647059:GC:TT	both novel
	X	eye	Group 8	FN1	Glomerulopathy, fibronectin deficiency, Spondylometaphyseal dysplasia	unlikely	2:216251538:G:A	0.78%
WGS-NGS-017	H	eye	Group 4	ABCA4	retinitis pigmentosa, rod-cone dystrophy and other eye disorders(AR)	Possible	1:94568686:C:T, 1:94470320:C:T	0.3% Euro
WGS-NGS-030	R	neuro	Group 6	A2ML1	Noonan-like malformation syndrome	Very unlikely	12:9013783:C:T, 12:9020600:G:A,	4% & 1.8% in Africans
WGS-NGS-056	N	connective	Group 4	FBN1	Marfan syndrome	Possible (Marfanoid)	15:48700642:T:C, 15:48849792:C:A	1x Gnomad/Novel
			Group 8	TNXB	AR Ehlers-Danlos like syndrome	Unlikely	6:32012817:C:T	12% in ASH
WGS-NGS-067	M	eye	Group1	TYR	AR ocular-cutaneous albinism	Unlikely	11:89017973:C:T	1% in Finns
WGS-NGS-068	J	neuro	Group 3	LMNA	progeria, muscular dystrophy, lipodystrophy, and others	Possible	1:156105054:G:T	novel
			Group 6	WFS1	Wolfram syndrome	Partial Fit	4:6304086:C:T,	0.01%
			Group 8	ATM	Ataxia telangiectasia, Cancer	Very unlikely	11:108196896:C:T	3% in ASH
WGS-NGS-078	V	connective	Group 2	COL1A2	Ehlers-Danlos syndrome, Osteogenesis imperfecta	Possible (EDS III)	7:94049587:C:T	2x in Gnomad
			Group 6 late	MYO1E	Glomerulosclerosis, focal segmental, 6	Very unlikely	15:59443263:CGTGCACTT:C, 15:59443263:CGTGCACTT:C	1st - 6% MAF in Finns
WGS-NGS-093	F	connective	Group 4	EP300	Rubinstein Taybi syndrome	Unlikely	22:41574743:A:G	novel
			Group 8	CACNA1C	Brugada syndrome	Very unlikely	12:2791205:A:G	99% in Europeans
WGS-NGS-095	C	eye	Group 2	no gene listed			no variant listed	
			Group 4	USHA2	Usher syndrome	Possible, not ideal	1:215953583:A:G, 1:215964830:T:G	0.49% is East Asians, novel
WGS-NGS-097	D	connective	Group 6 late	TNXB	Ehlers-Danlos syndrome - Classic	Possible	6:32064371:A:G, 6:32064371:A:G	0.23% in East Asians
WGS-NGS-099	B	neuro	Group 4	PIGT	Multiple congenital anomalies-hypotonia-seizures syndrome 3 - AR	Unlikely	20:44044789:C:T	1 x in Gnomad
			Group 6	HSD17B4	Perrault syndrome (seizures +)	Unlikely	5:118832244:C:G,	7% in Africans
			Group 8	SLC25A22	Epileptic encephalopathy, early infantile, 3	Possible	11:793588:G:A	12% in Europeans
WGS-NGS-102	A	connective	Group 8	COL5A2	Ehlers-Danlos syndrome - Classic	Possible	2:189974958:G:T	97% in Europeans

Pt J

GROWTH PARAMETERS

Weight for age
Decreased body weight (<-2SD)
Stature for age
Short stature (<-2SD)
Delayed puberty
EYE DEFECTS
Ptosis
Retinal dystrophy
Corneal dystrophy
CARDIOVASCULAR
Sinus tachycardia

MUSCULOSKELETAL

Scoliosis
Mild
Osteopenia
Pes planus
Hallux valgus
Delayed skeletal maturation
Vertebral compression fractures
Myopathy
Ragged-red muscle fibers
Abnormal mitochondria in muscle

GENITOURINARY

Stage 5 chronic kidney disease
Renal transplant in 2012

BEHAVIOR, COGNITION AND DEVELOPMENT

NO Global developmental delay

Specific learning disability
Some learning difficulties,
NEUROLOGICAL

Ataxia
Morphological abnormality of the CNS
Brain MRI mild cerebellar volume loss
Guillian-Barre syndrome in 2008

BLOOD AND BLOOD-FORMING TISSUES

Aplastic anemia
METABOLISM/HOMEOSTASIS
Renal tubular acidosis

WFS1 - Wolfram syndrome
diabetes mellitus, optic atrophy
diabetes insipidus, deafness
Occasional: renal abnormalities, ataxia,
mental retardation

CAGI 5: The SickKids Genome Clinic Challenge

The Challenge: Predict which variant is associated with a specific patient

Summary Results

Genome	Patient	Category	Group	Gene	Diseases	Plausible phenotypic match?	Variants	Max MAF
WGS-NGS-007	X	eye	Group 7	PROM1	Somatic mutations cause Sturge Weber syndrome	unlikely	4:16024957, 9:80647059:GC:TT	both novel
	X	eye	Group 8	FN1	Glomerulopathy, fibronectin deficiency, Spondylometaphyseal dysplasia	unlikely	2:216251538:G:A	0.78%
WGS-NGS-017	H	eye	Group 4	ABCA4	retinitis pigmentosa, rod-cone dystrophy and other eye disorders(AR)	Possible	1:94568686:C:T, 1:94470320:C:T	0.3% Euro
WGS-NGS-030	R	neuro	Group 6	A2ML1	Noonan-like malformation syndrome	Very unlikely	12:9013783:C:T, 12:9020600:G:A,	4% & 1.8% in Africans
WGS-NGS-056	N	connective	Group 4	FBN1	Marfan syndrome	Possible (Marfanoid)	15:48700642:T:C, 15:48849792:C:A	1x Gnomad/Novel
			Group 8	TNXB	AR Ehlers-Danlos like syndrome	Unlikely	6:32012817:C:T	12% in ASH
WGS-NGS-067	M	eye	Group1	TYR	AR ocular-cutaneous albinism	Unlikely	11:89017973:C:T	1% in Finns
WGS-NGS-068	J	neuro	Group 3	LMNA	progeria, muscular dystrophy, lipodystrophy, and others	Possible	1:156105054:G:T	novel
			Group 6	WFS1	Wolfram syndrome	Partial Fit	4:6304086:C:T,	0.01%
			Group 8	ATM	Ataxia telangiectasia, Cancer	Very unlikely	11:108196896:C:T	3% in ASH
WGS-NGS-078	V	connective	Group 2	COL1A2	Ehlers-Danlos syndrome, Osteogenesis imperfecta	Possible (EDS III)	7:94049587:C:T	2x in Gnomad
			Group 6 late	MYO1E	Glomerulosclerosis, focal segmental, 6	Very unlikely	15:59443263:CGTGCACTT:C, 15:59443263:CGTGCACTT:C	1st - 6% MAF in Finns
WGS-NGS-093	F	connective	Group 4	EP300	Rubinstein Taybi syndrome	Unlikely	22:41574743:A:G	novel
			Group 8	CACNA1C	Brugada syndrome	Very unlikely	12:2791205:A:G	99% in Europeans
WGS-NGS-095	C	eye	Group 2	no gene listed			no variant listed	
			Group 4	USHA2	Usher syndrome	Possible, not ideal	1:215953583:A:G, 1:215964830:T:G	0.49% is East Asians, novel
WGS-NGS-097	D	connective	Group 6 late	TNXB	Ehlers-Danlos syndrome - Classic	Possible	6:32064371:A:G, 6:32064371:A:G	0.23% in East Asians
WGS-NGS-099	B	neuro	Group 4	PIGT	Multiple congenital anomalies-hypotonia-seizures syndrome 3 - AR	Unlikely	20:44044789:C:T	1 x in Gnomad
			Group 6	HSD17B4	Perrault syndrome (seizures +)	Unlikely	5:118832244:C:G,	7% in Africans
			Group 8	SLC25A22	Epileptic encephalopathy, early infantile, 3	Possible	11:793588:G:A	12% in Europeans
WGS-NGS-102	A	connective	Group 8	COL5A2	Ehlers-Danlos syndrome - Classic	Possible	2:189974958:G:T	97% in Europeans

Pt J

GROWTH PARAMETERS

Weight for age
Decreased body weight (<-2SD)
Stature for age
Short stature (<-2SD)
Delayed puberty
EYE DEFECTS
Ptosis
Retinal dystrophy
Corneal dystrophy
CARDIOVASCULAR
Sinus tachycardia

MUSCULOSKELETAL

Scoliosis
Mild
Osteopenia
Pes planus
Hallux valgus
Delayed skeletal maturation
Vertebral compression fractures
Myopathy
Ragged-red muscle fibers
Abnormal mitochondria in muscle

GENITOURINARY

Stage 5 chronic kidney disease
Renal transplant in 2012

BEHAVIOR, COGNITION AND DEVELOPMENT

NO Global developmental delay

Specific learning disability
Some learning difficulties,
NEUROLOGICAL

Ataxia
Morphological abnormality of the CNS
Brain MRI mild cerebellar volume loss
Guillian-Barre syndrome in 2008

BLOOD AND BLOOD-FORMING TISSUES

Aplastic anemia
METABOLISM/HOMEOSTASIS
Renal tubular acidosis

ATM - ataxia telangiectasia
progressive ataxia, immunodeficiency,
cerebellar atrophy, cancer, diabetes

CAGI 5: The SickKids Genome Clinic Challenge

Initial assessment of results: Diagnostic variants

8 CAGI teams - 9 Assessments

Predict which class of disease is associated with each genome

Slightly better than chance, unexpectedly better for connective tissue disorders

Match genomes to specific clinical descriptions

Three teams substantially better than chance.

Identify diagnostic variants underlying the predictions

Not all teams included non-coding variants

Clinical phenotypes were not usually good matches for variants' predicted phenotype

MAF was a problem for some teams

But many variants were unlikely despite matching pt/genome. Were sex and ethnicity clues?

Next time: Trios and RNA-Seq?

Acknowledgements

Hospital for Sick Children/Centre for Genetic Medicine
Genome Clinic Working Group

- Stephen Meyn
- Sarah Bowdin
- Peter Ray
- Ronni Cohn
- Stephen Scherer
- Nasim Monfared
- Cheryl Shuman
- Christian Marshall
- James Stavropoulos
- Rebekah Jobling
- Michael Brudno
- Marta Girdea
- Shinya Ito
- Iris Cohn
- Participating Clinicians
- Patients and Parents
- Daniele Merico
- Thomas Nalpathamkalam
- Giovanna Pellecchia
- Bhooma Thiruvahindrapuram
- Randi Slotnik-Shaul
- Michael Szego
- Robin Hayeems

Secondary Predictive Variants

Predictive medically actionable variants predict hidden or future disease states:

High penetrant risk variants for Mendelian disorders

Attenuated risk variants for AD Mendelian disorders

Protective variants

Carrier variants for recessive disorders



"WHAT'S WRONG, SCOTTY? ACCORDING TO YOUR DNA, YOU SHOULD BE HITTING .324."

Secondary Predictive Variants

Predictive medically actionable variants predict hidden or future disease states:

High penetrant risk variants for Mendelian disorders

Attenuated risk variants for AD Mendelian disorders

Protective variants

Carrier variants for recessive disorders

Constraints on predictive variant searches

Total incidence of Mendelian disorders: 10+%

But - individual Mendelian disorders are rare

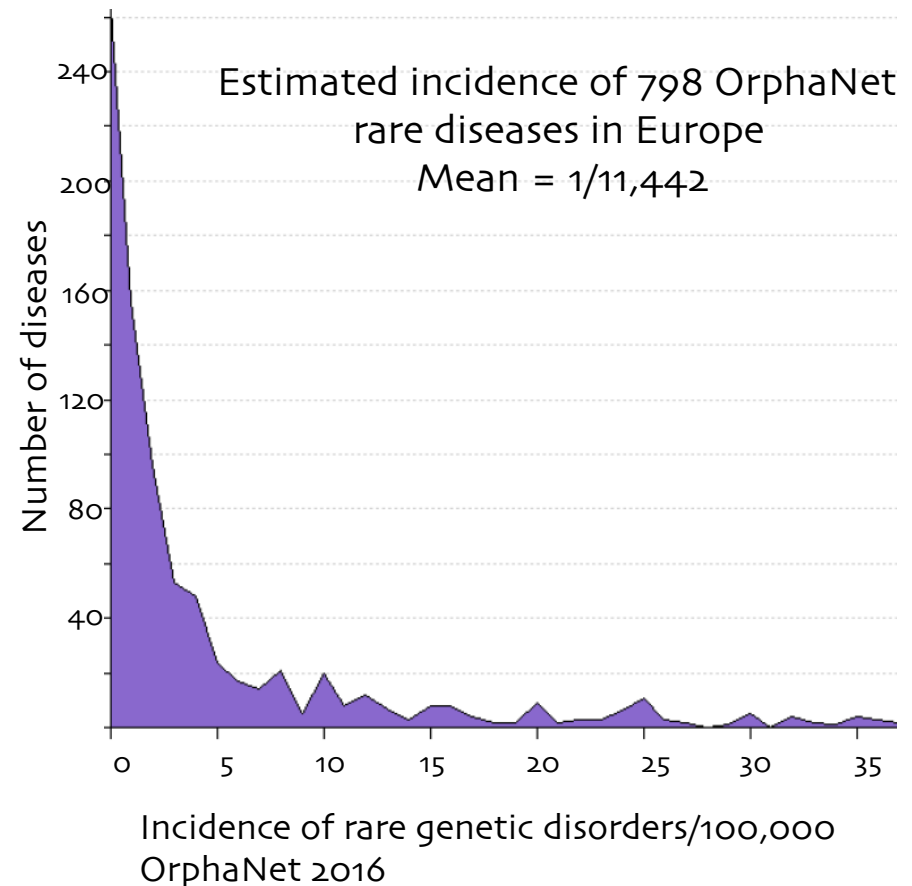
- prior risk of a specific disorder is low

Created for CAGI 2018

- VUS common

Therefore reporting criteria must be stringent and some risk variants left unreported

Incidence of genetic disease limits number of predictive variants/person



Secondary Predictive Variants

Predictive medically actionable variants predict hidden or future disease states:

High penetrant risk variants for Mendelian disorders

Attenuated risk variants for AD Mendelian disorders

Protective variants

Carrier variants for recessive disorders

Constraints on predictive variant searches

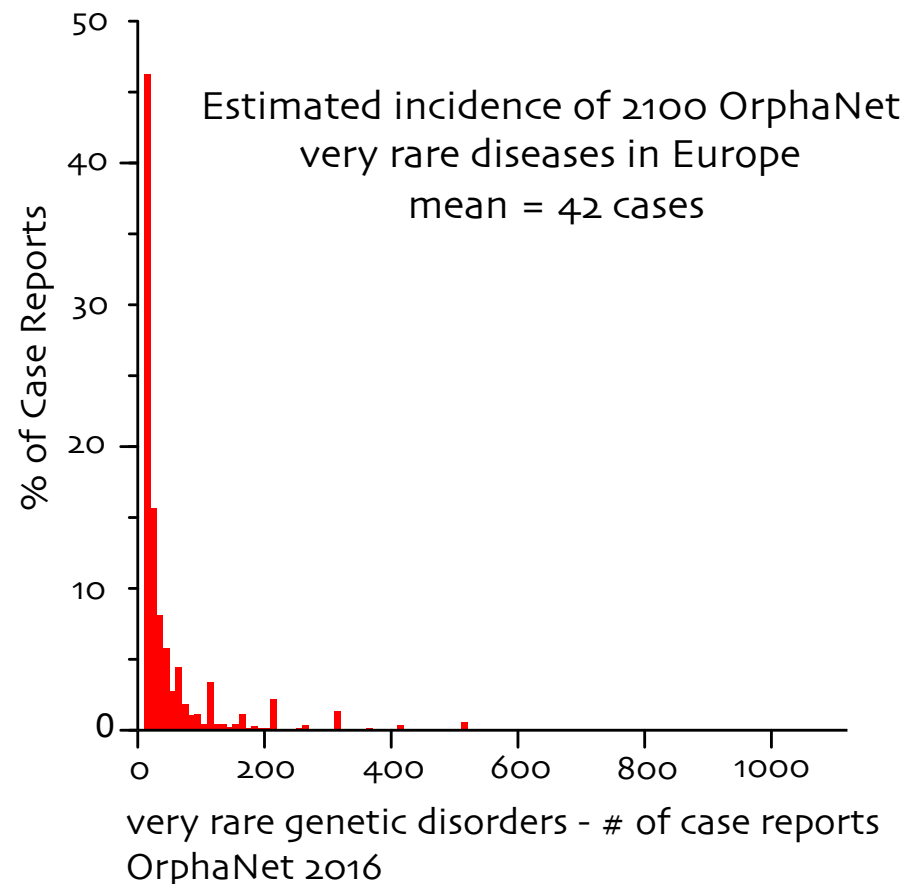
Total incidence of Mendelian disorders: 10+%

But - individual Mendelian disorders are rare

- prior risk of a specific disorder is low
- VUS common

Therefore reporting criteria must be stringent and some risk variants left unreported

Incidence of genetic disease limits number of predictive variants/person



Secondary Predictive Variants

Predictive medically actionable variants predict hidden or future disease states:

High penetrant risk variants for Mendelian disorders

Attenuated risk variants for AD Mendelian disorders

Protective variants

Carrier variants for recessive disorders

Constraints on predictive variant searches

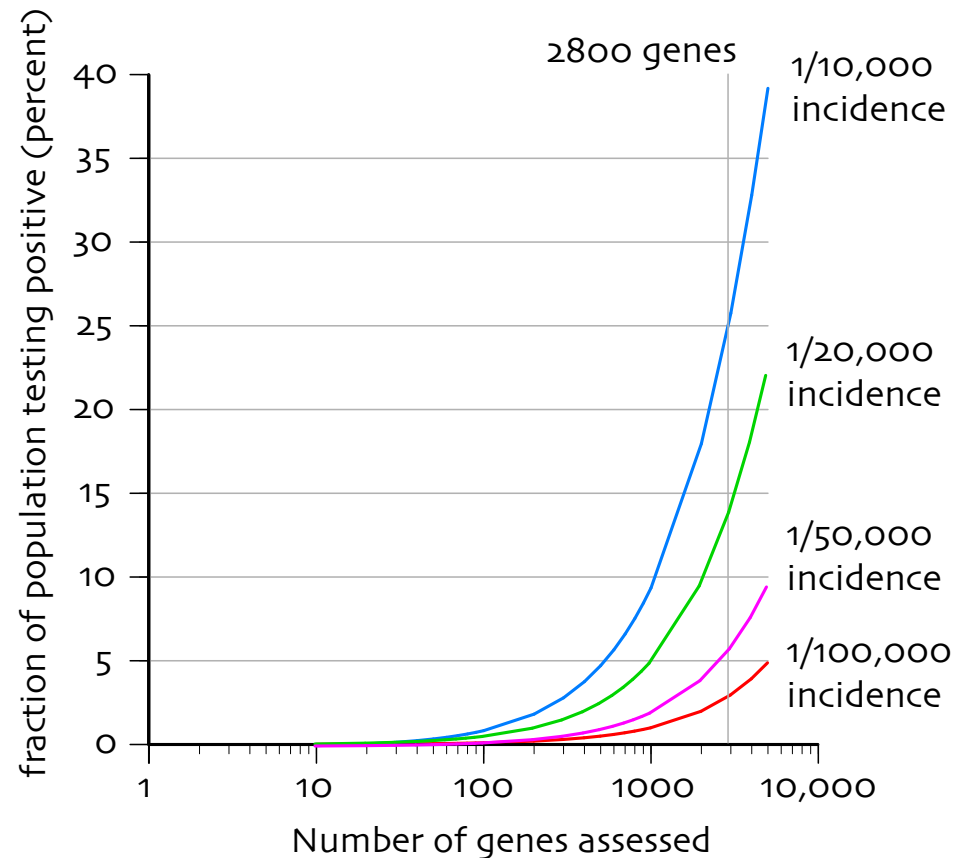
Total incidence of Mendelian disorders: 10+%

But - individual Mendelian disorders are rare

- prior risk of a specific disorder is low
- VUS common

Therefore reporting criteria must be stringent and some risk variants left unreported

Incidence of genetic disease limits number of predictive variants/person



Secondary Predictive Variants

Predictive medically actionable variants predict hidden or future disease states:

High penetrant risk variants for Mendelian disorders

Attenuated risk variants for AD Mendelian disorders

Protective variants

Carrier variants for recessive disorders

Constraints on predictive variant searches

Total incidence of Mendelian disorders: 10+%

But - individual Mendelian disorders are rare

- prior risk of a specific disorder is low
- VUS common

Therefore reporting criteria must be stringent and some risk variants left unreported

Incidence of genetic disease limits number of predictive variants/person

Exclude GUS (genes of unknown significance)

