

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.



A handwritten signature in black ink, appearing to read 'Stephen Meyn'.

CAGI: The SickKids Genome Clinic Challenge

Genomes + phenotypic descriptions of 25 undiagnosed paediatric genetics patients

Stephen Meyn MD PhD FRCPC
Centre for Genetic Medicine
Program in Genomics and Genome Biology
Division of Clinical and Metabolic Genetics
The Hospital for Sick Children

Department of Molecular Genetics
Department of Paediatrics
University of Toronto



"SO - WHICH GENES HAVE BEEN BOTHERING YOU?"

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

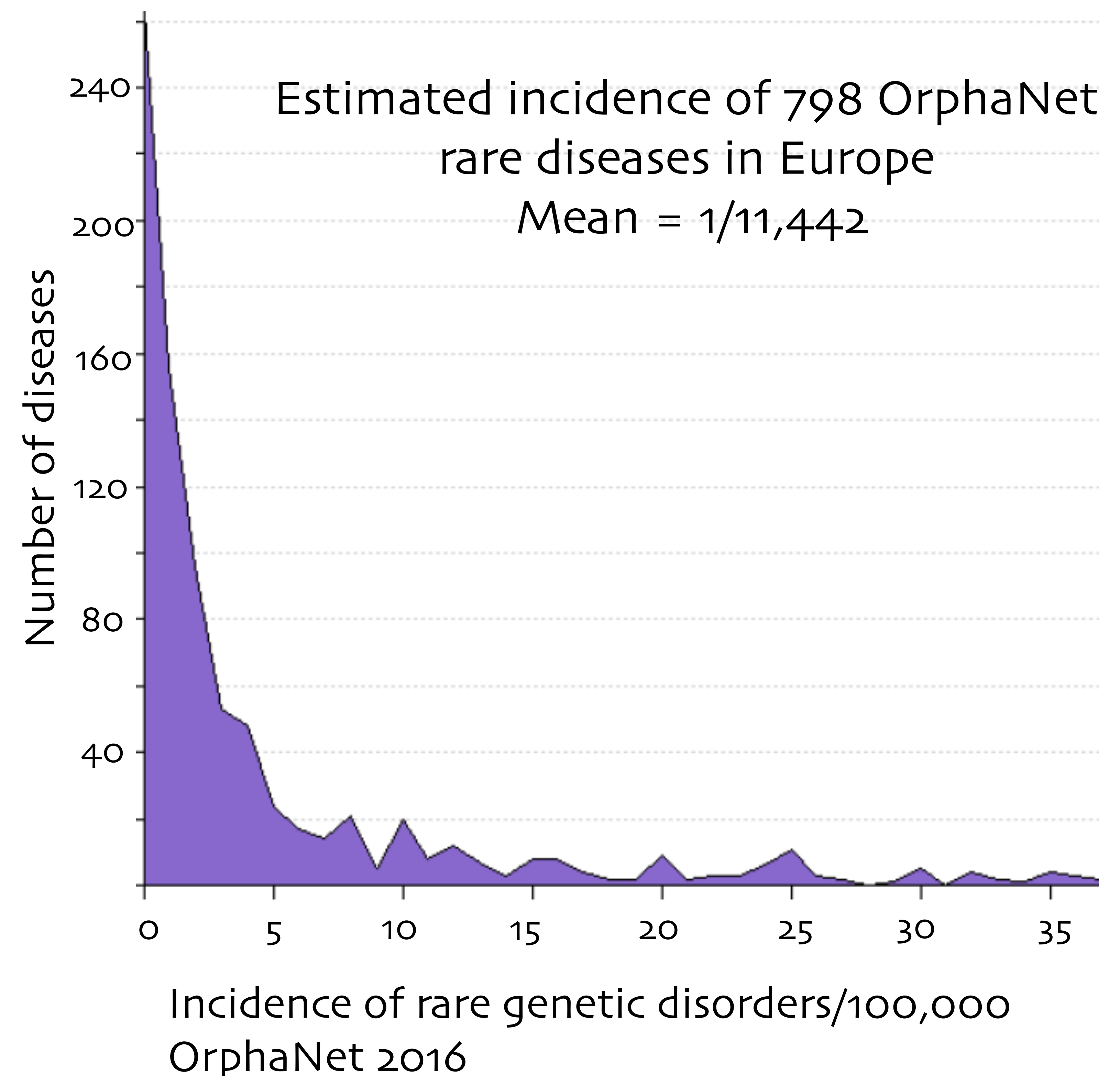
- few cases, impurity of control datasets

Segregation analyses not always possible

Disorders not always well characterized

Limited phenotypic description in databases

Functional validation rarely available



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

Fanconi anemia - 19 genes/pleiotropic phenotype



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

Fanconi anemia - 19 genes/pleiotropic phenotype



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

Fanconi anemia - 19 genes/pleiotropic phenotype

OMIM Clinical Synopsis

Radial aplasia	Congenital heart defect
Kidney malformation	Microcephaly
Cryptorchidism	Strabismus, Microphthalmia
Low birth weight	Ear anomaly, Deafness
Mental retardation	Pancytopenia/Bone Marrow Failure
Pigmentary changes	Leukemia

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

Fanconi anemia - 19 genes/pleiotropic phenotype

OMIM Clinical Synopsis

33% - Radial aplasia	10% - Congenital heart defect
20% - Kidney malformation	25% - Microcephaly
30% - Cryptorchidism	20% - Strabismus, Microphthalmia
10% - Low birth weight	10% - Ear anomaly, Deafness
10% - Mental retardation	90% - Risk of Pancytopenia/bone marrow failure
50% - Pigmentary changes	30% - Leukemia
	30% - Risk of head/neck squamous cell carcinoma

25% of FA patients have no phenotype besides bone marrow failure

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

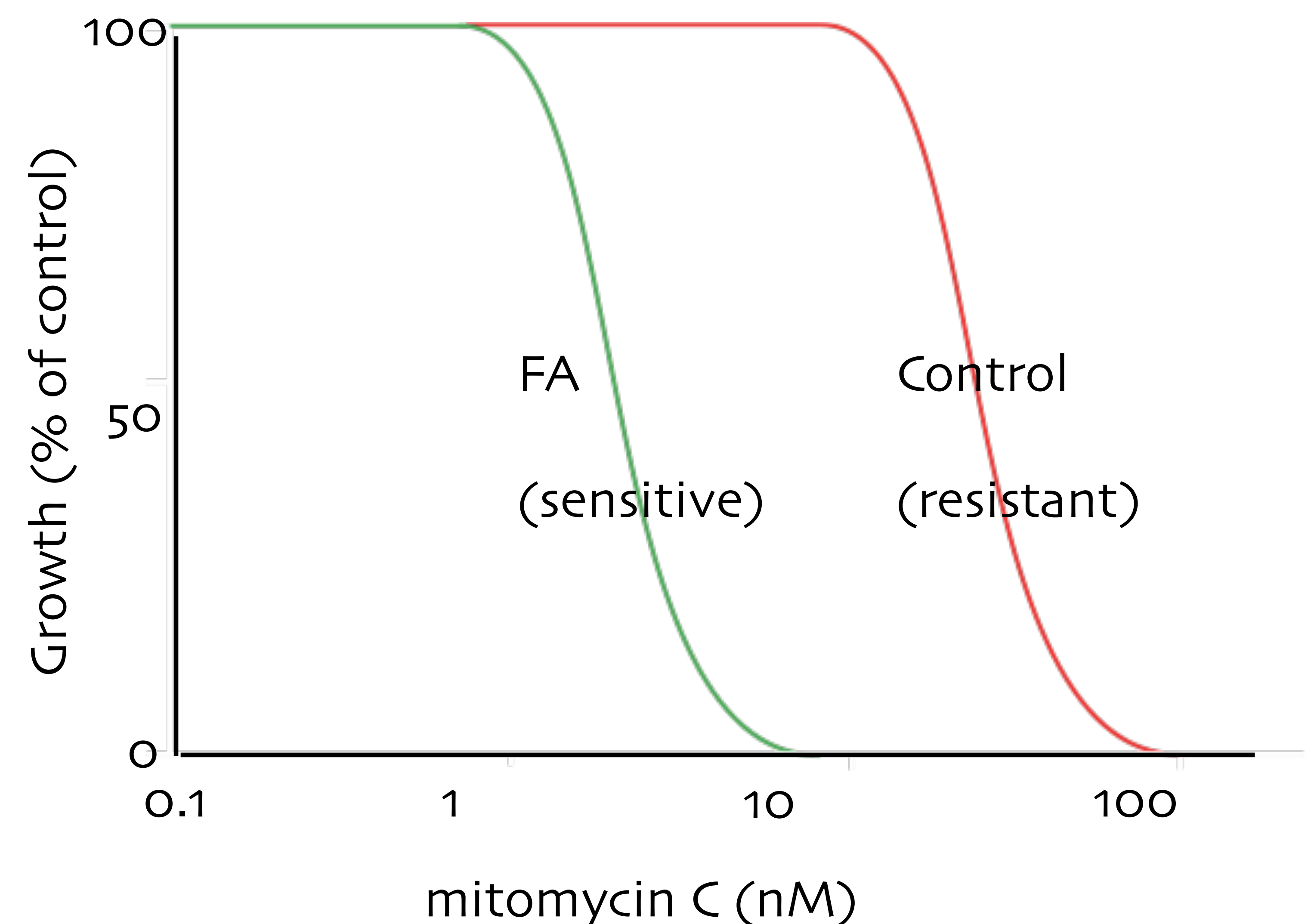
Segregation analyses not always possible

Disorders not always well characterized

Limited phenotypic description in databases

Functional validation rarely available

Fanconi anemia - Functional assessment



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

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

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

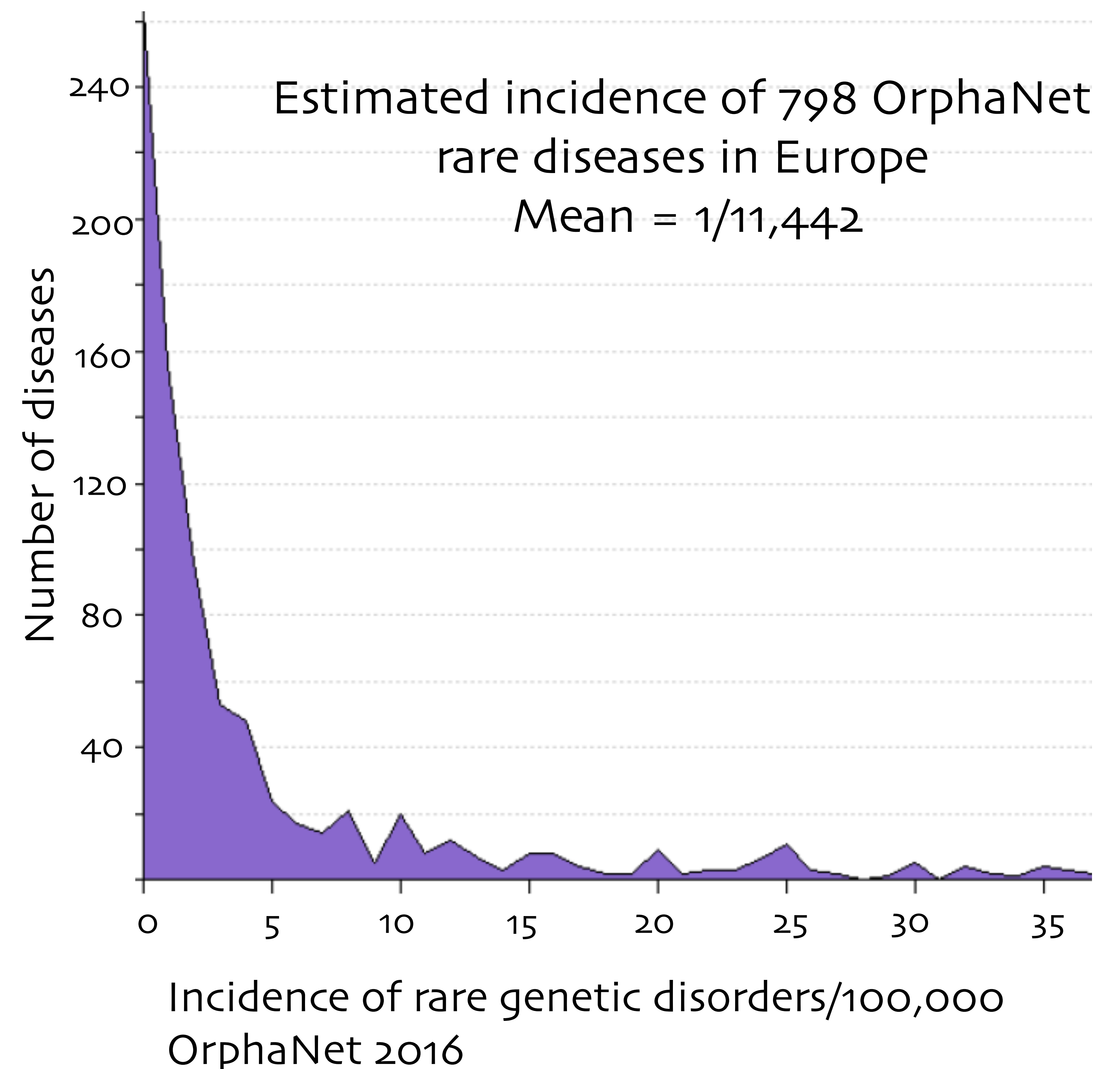
Total incidence of Mendelian disorders: 10+ %

But - individual Mendelian disorders are rare

- prior risk of a specific disorder is low
- 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+%

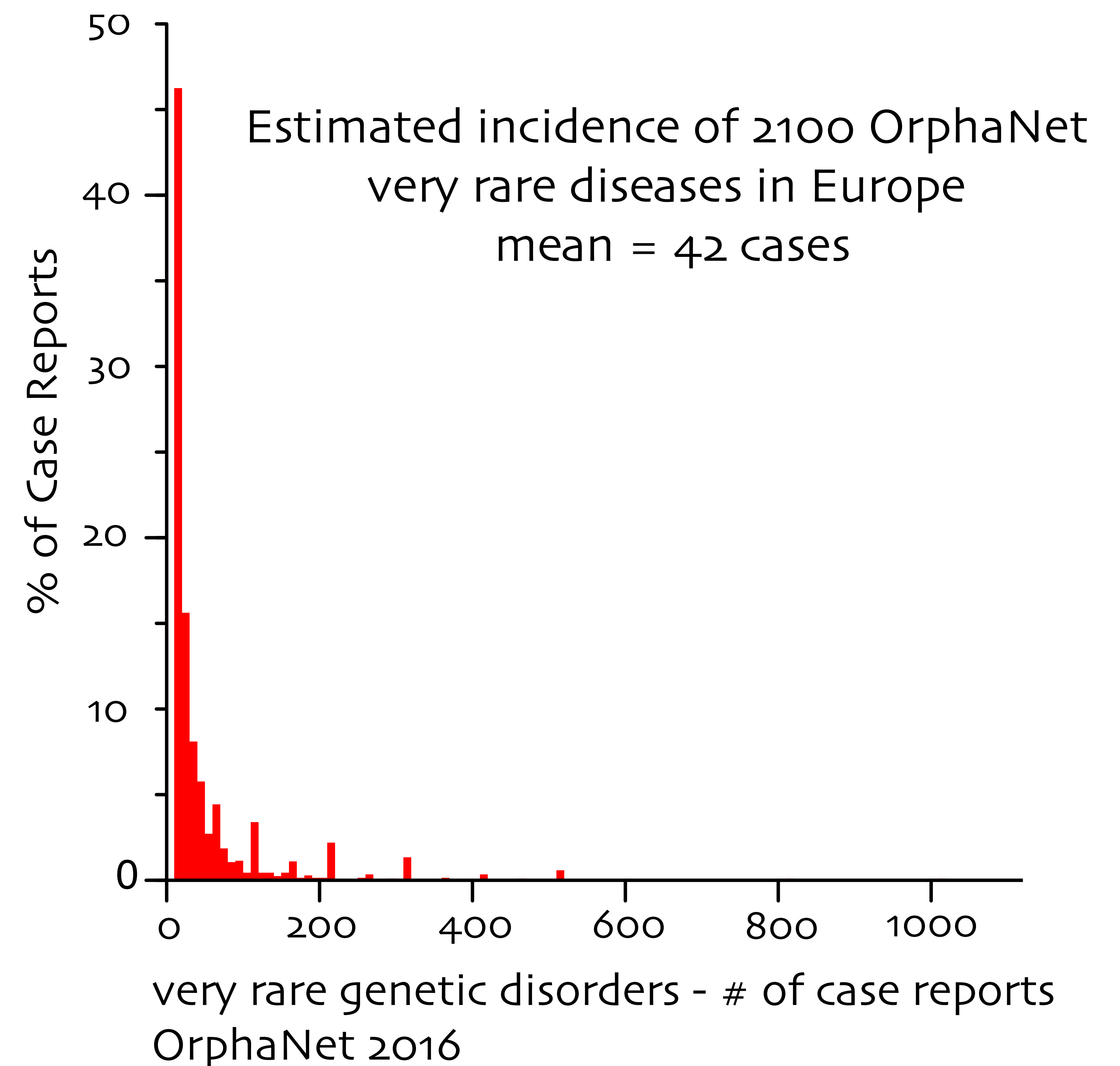
But - individual Mendelian disorders are rare

- 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+%

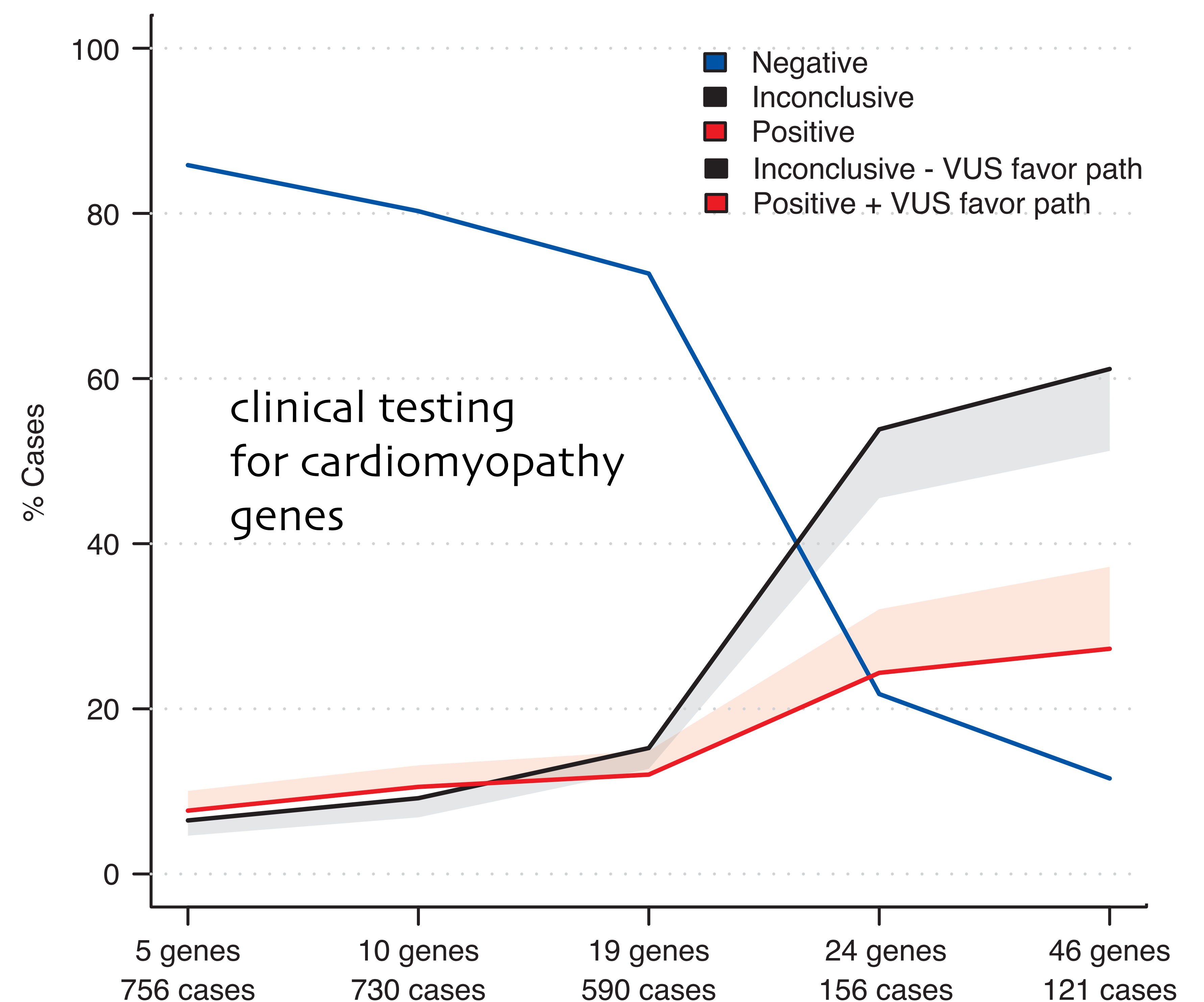
But - individual Mendelian disorders are rare

- 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



Pugh et al. Genet in Med 2014

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+ %

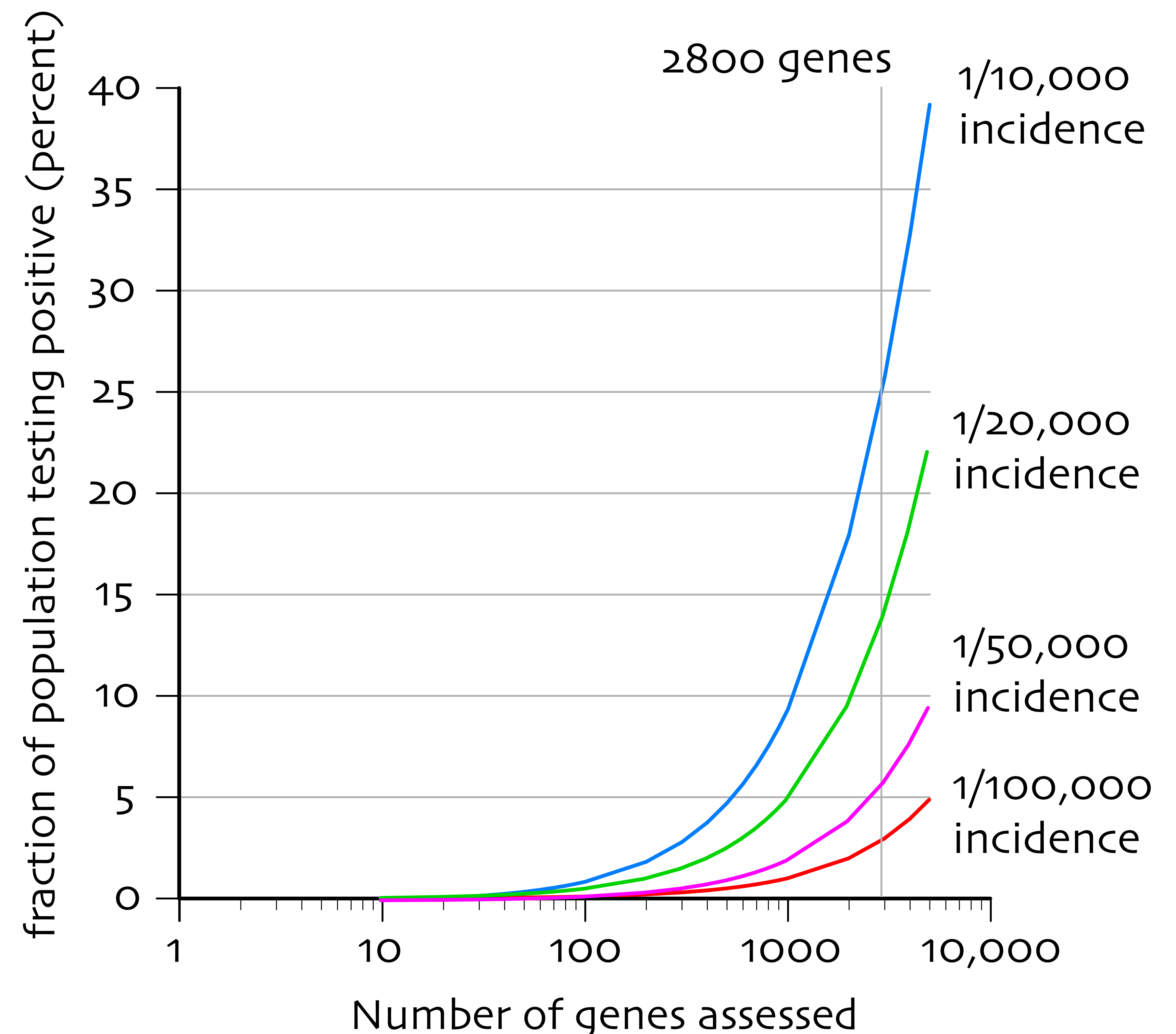
But - individual Mendelian disorders are rare

- 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

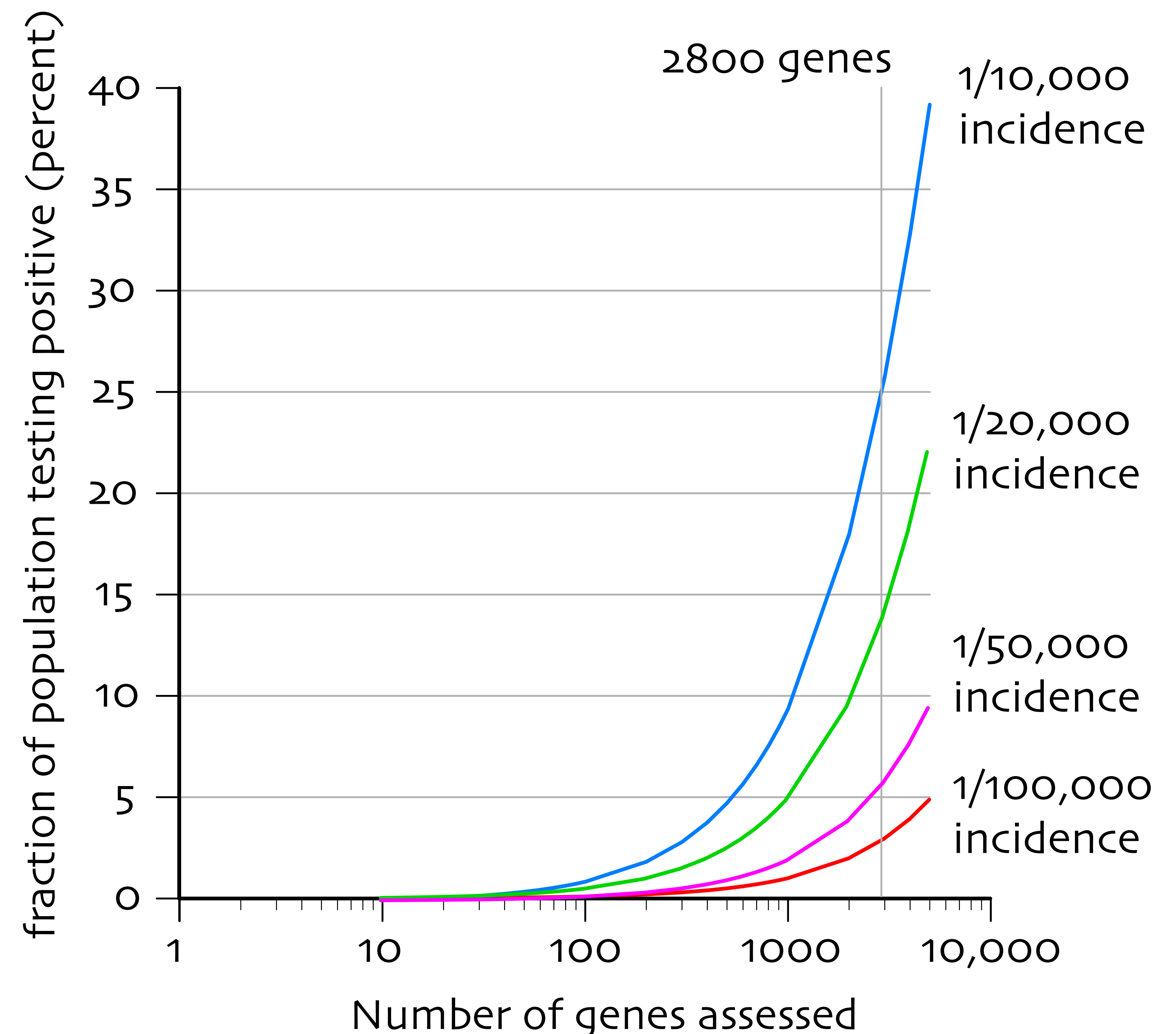
- prior risk of a specific disorder is low

- 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)



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
-

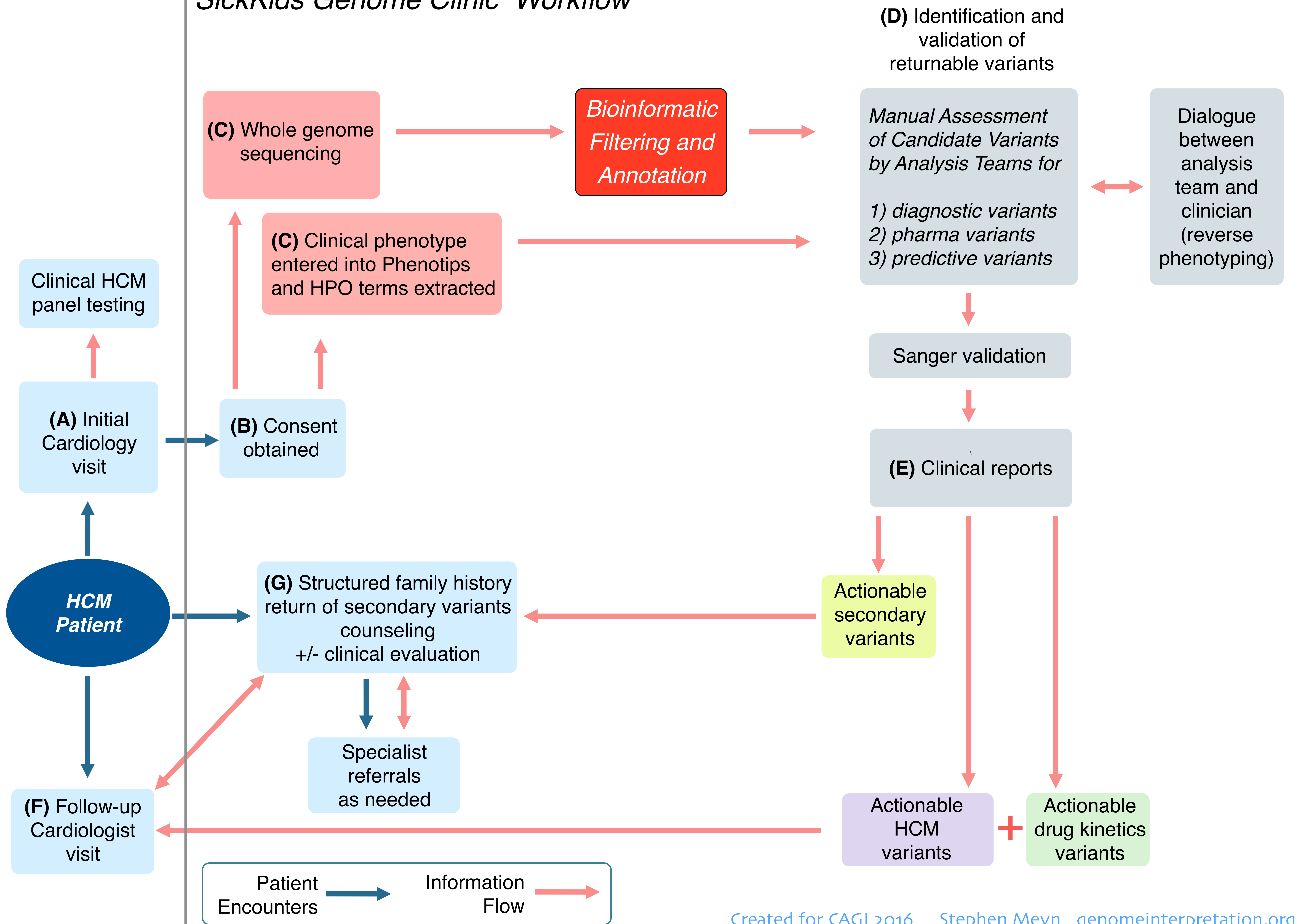
Inclusion Criteria:

- <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 2016
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

The CAGI cohort: Clinical microarray vs. WGS

Demographics:

100 consecutive cases
undergoing CMA (180K clinical chip)

- all seen by geneticists
- 5.5 years mean age
- 43% female/57% male
- 8% consanguineous parents
- 59% had at least 1 additional genetic test
- 57% Developmental delay

Genotyping:

WGS by Complete Genomics

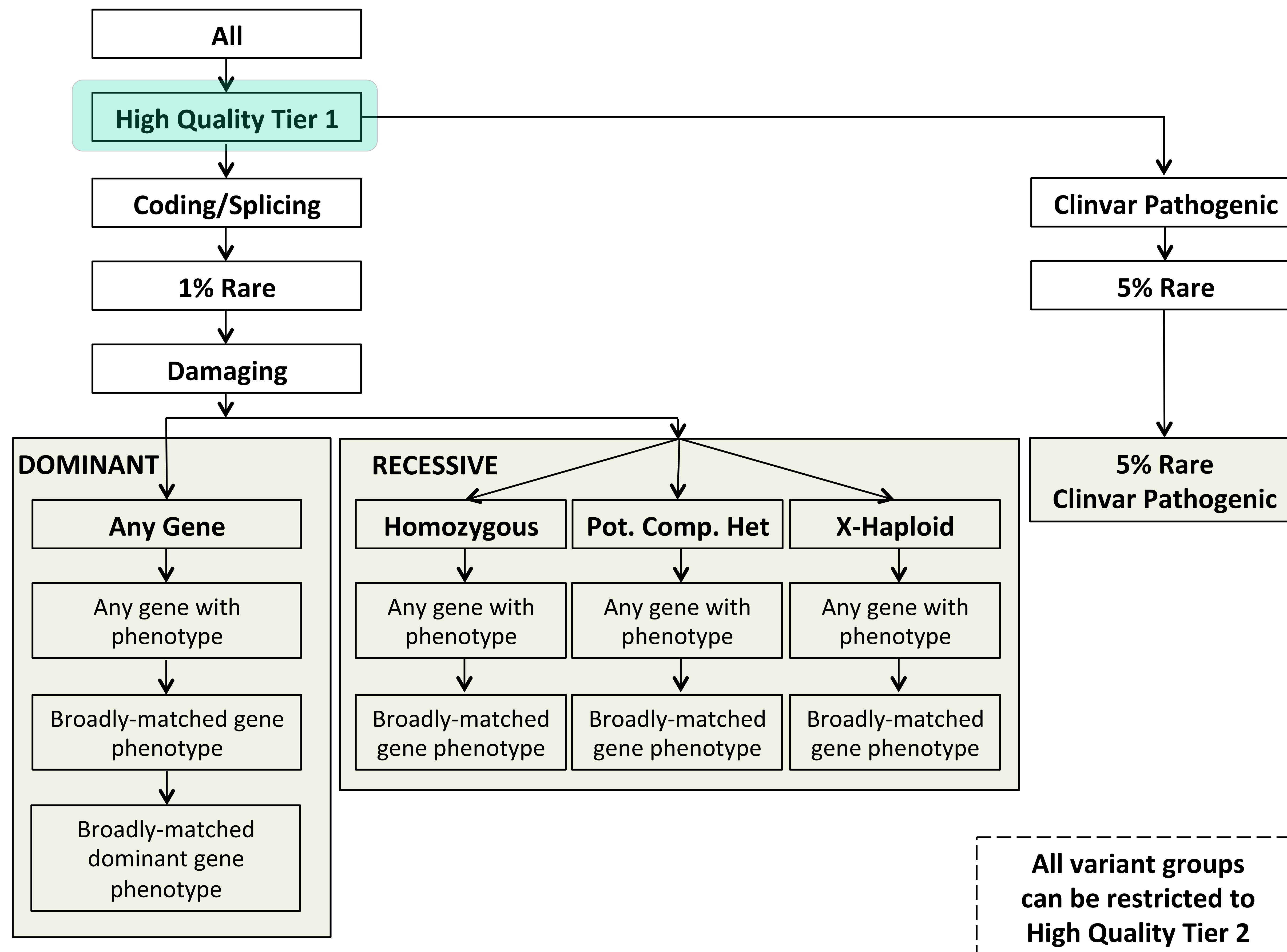
- 52x average coverage
- 98.6% >10x coverage
- 99.4% > 5x coverage
- ~3,700,000 variants/genome

Phenotyping:

HPO.ID	HPO.name	Case Count
HP:0000707	Abnormality of the nervous system	77
HP:0000152	Abnormality of head and neck	70
HP:0000924	Abnormality of the skeletal system	68
HP:0001507	Growth abnormality	44
HP:0001574	Abnormality of the integument	38
HP:0000478	Abnormality of the eye	34
HP:0001626	Abnormality of the cardiovascular system	32
HP:0003011	Abnormality of the musculature	27
HP:0000598	Abnormality of the ear	26
HP:0001438	Abnormality of the abdomen	25
HP:0000119	Abnormality of the genitourinary system	18
HP:0002086	Abnormality of the respiratory system	16
HP:0003549	Abnormality of connective tissue	13
HP:0002715	Abnormality of the immune system	9
HP:0001939	Abnormality of metabolism/homeostasis	7
HP:0001871	Abnormality of blood/blood-forming tissues	6
HP:0002664	Neoplasm	5
HP:0000818	Abnormality of the endocrine system	5
HP:0000769	Abnormality of the breast	2
HP:0001197	Abnormality of prenatal development or birth	1
HP:0001608	Abnormality of the voice	1

Bioinformatics Pipeline for Primary Diagnosis

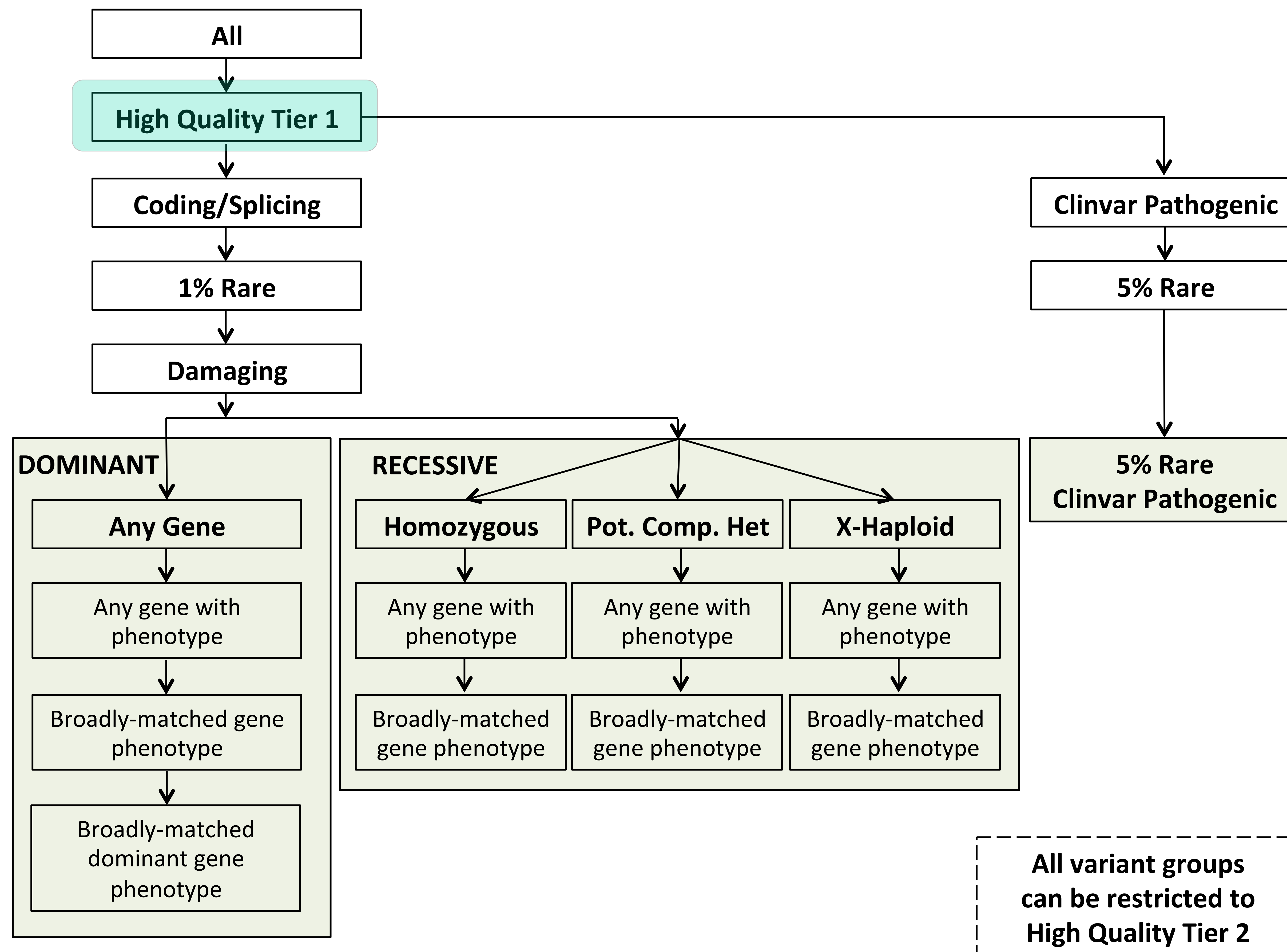
Primary Disorder Pipeline: a “Genetics-driven” approach



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

Bioinformatics Pipeline for Primary Diagnosis

Primary Disorder Pipeline

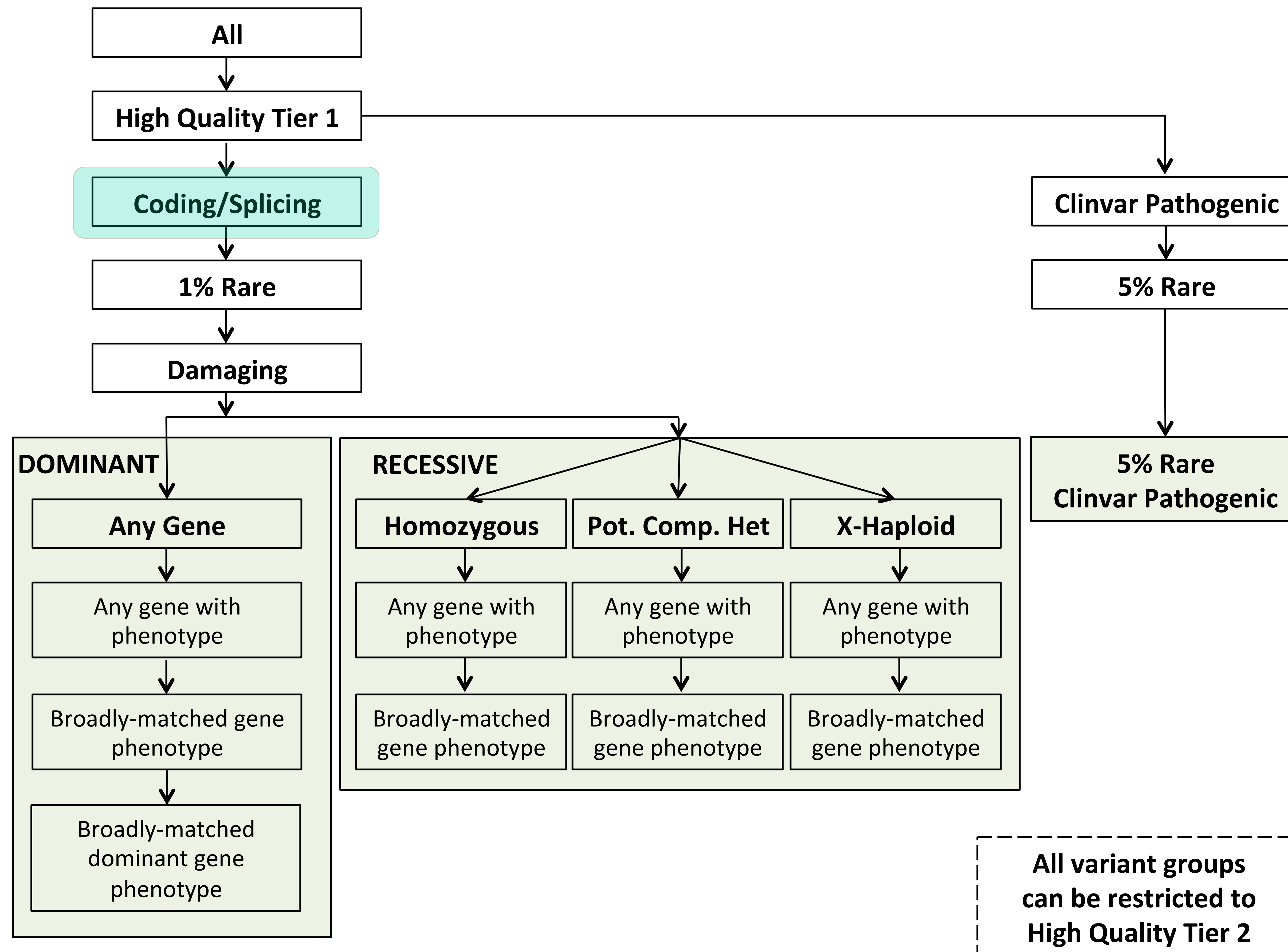


High Quality Tiers:

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

Bioinformatics Pipeline for Primary Diagnosis

Primary Disorder Pipeline



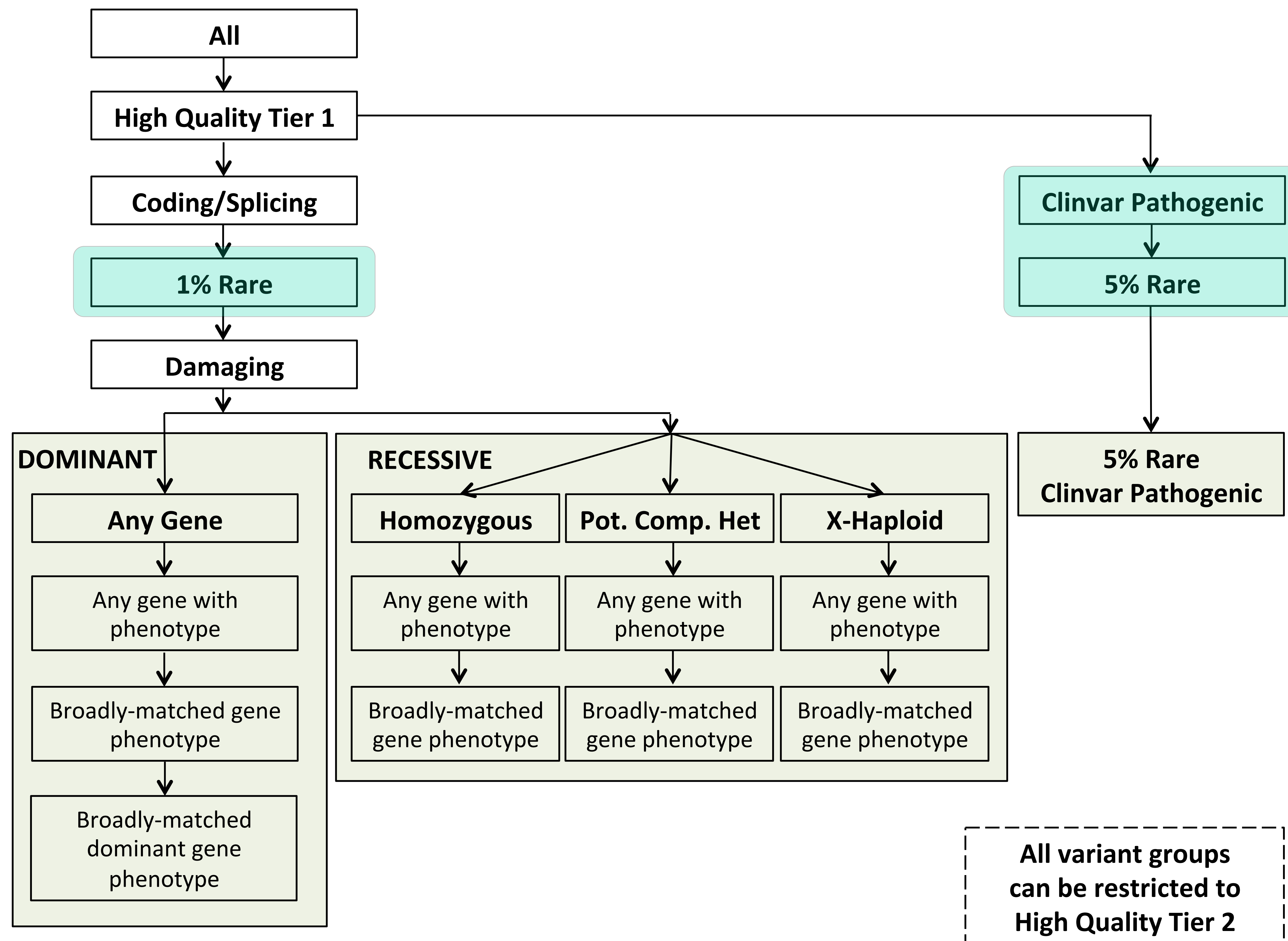
Regions examined:

Initial pass:
- coding/splicing

Subsequent analyses:
- expanded to other
gene features
(e.g., promoter sequences)

Bioinformatics Pipeline for Primary Diagnosis

Primary Disorder Pipeline



Filtering for rare variants

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

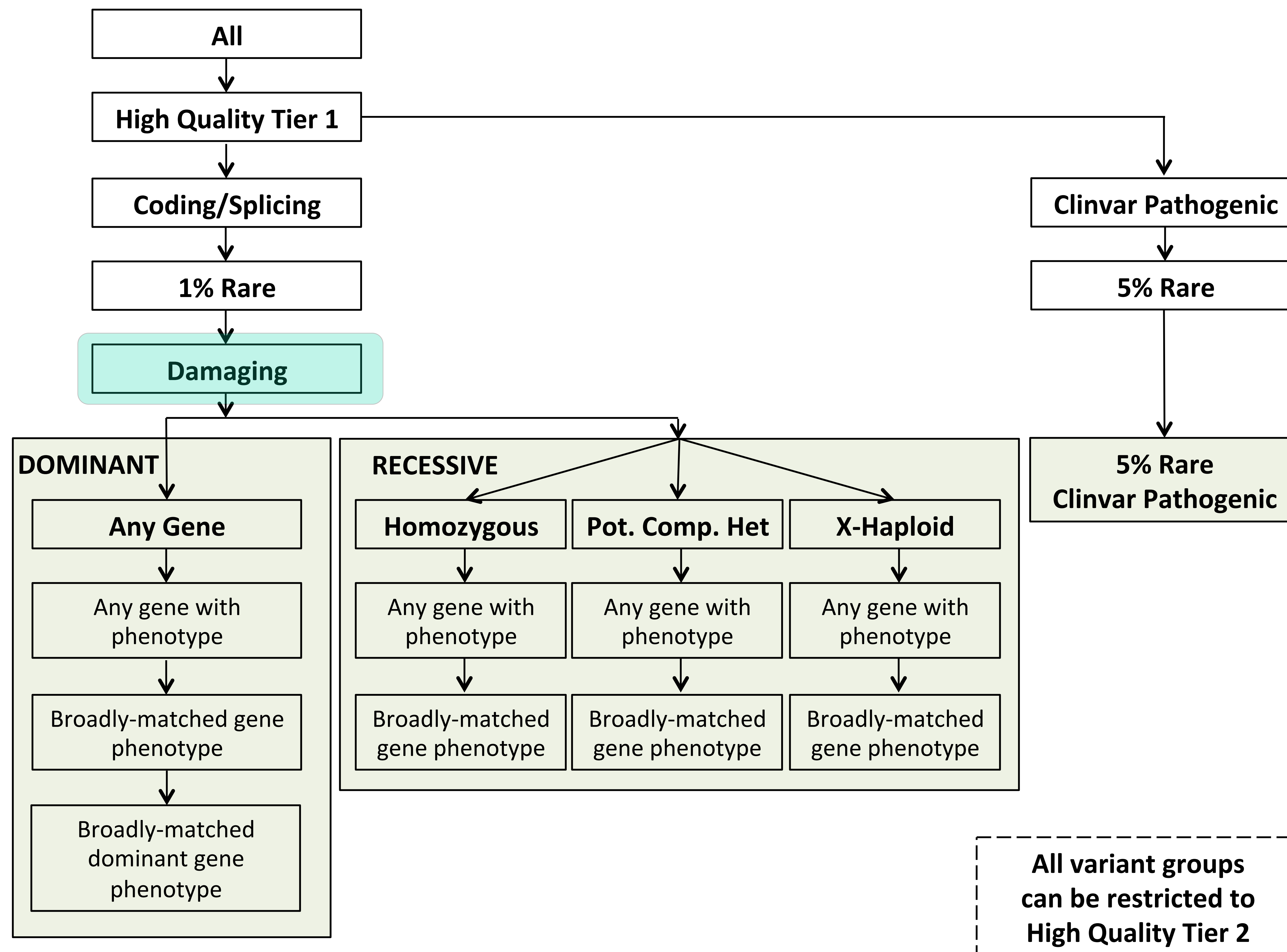
or

<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

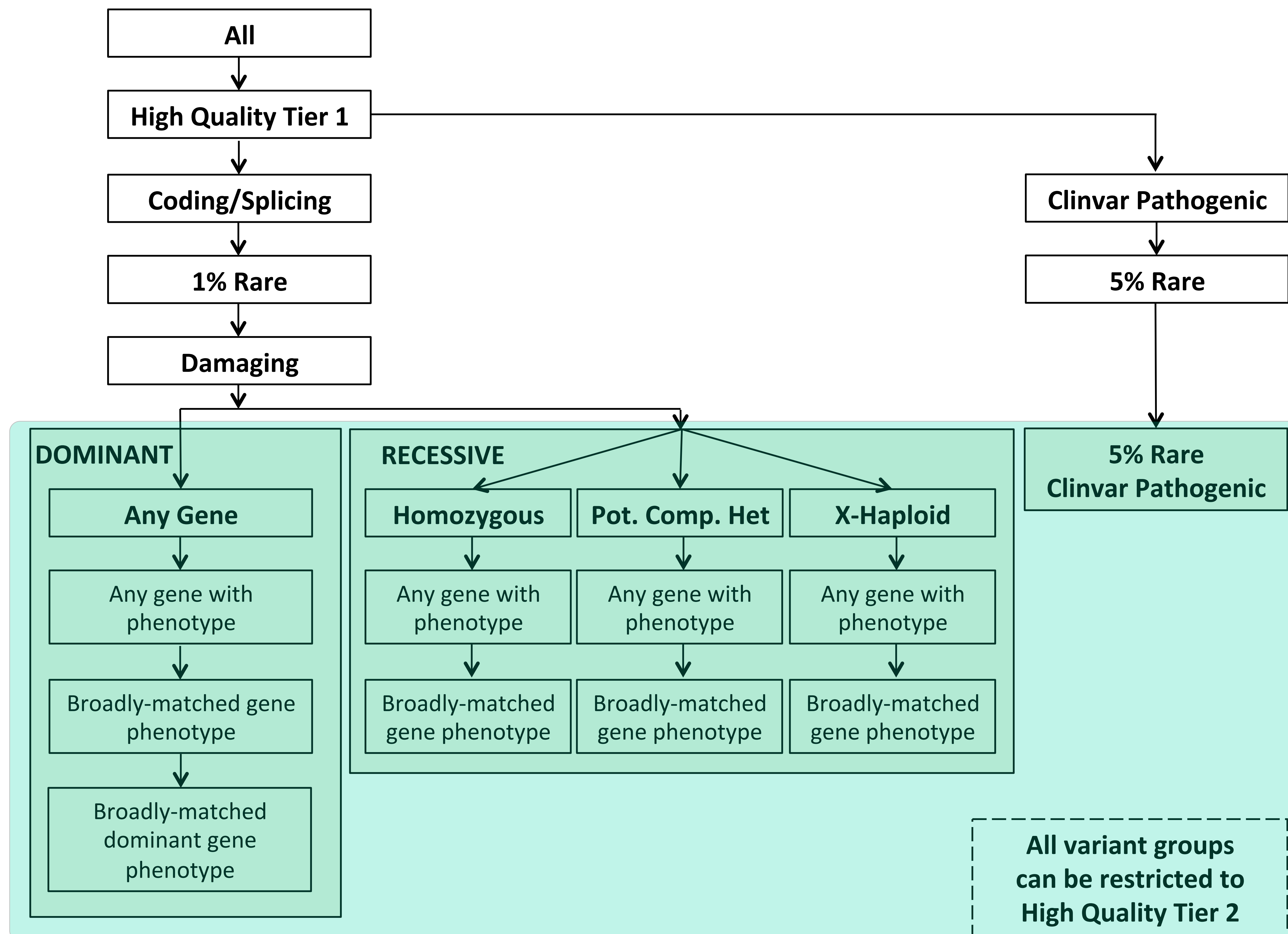
- minimal conservation plus at least one of:
 - High conservation
 - SIFT
 - Polyphen
 - Mutation Assessor

Non-frameshift in/del/sub

- mod/high conservation
- most not overlap dbSNP

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

Diagnostic yield: 30 diagnostic SNVs detected by WGS

Gene	Disease	Inheritance
EP300	Rubinstein-Taybi	AD
MC4R	obesity	AD
SMARCB1	Coffin-Siris	AD
Larp7	Alazami	AR
KAT6B	Chr8:97,145 - 98,301	AD
GDF5	Chr16:15,507 - 16,400	AD
PANK2	neurodegeneration/iron	AR
NGLY1	deglycosylation disorder	AR
PIK3R2	MPPH	AD
SPTAN1	early infantile epilepsy	AD
EXT2	multiple exostoses type 2	AD
TSEN54	Pontocerebellar hypoplasia 2a	AR
NSD1	Sotos	AD
CBL	Noonan-like with JMML	AD
PACS1	AD mental retardation 17	AD

Gene	Disease	Disease
SETD5	AD mental retardation 23	AD
PIK3R1	SHORT	AD
GJB2	AR hearing loss	AR
CCM2	cerebral cavernous malformation	AD
VWF	von Willibrand	AD
TYR	oculocutaneous albinism 1	AR
MC4R	obesity	AD
COL4A1	Angiopathy	AD
PLVAP	protein losing enteropathy	AR
NGLY1	congenital glycosylation Iv	AR
COG5	congenital glycosylation Ili	AR
ATP1A3	CAPOS	AD
VPS53	Pontocerebellar hypoplasia 2E	AR
SMARCA2	Nicolaides-Baraitser	AD
ALDH18A1	Cutis laxa IIIa	AR

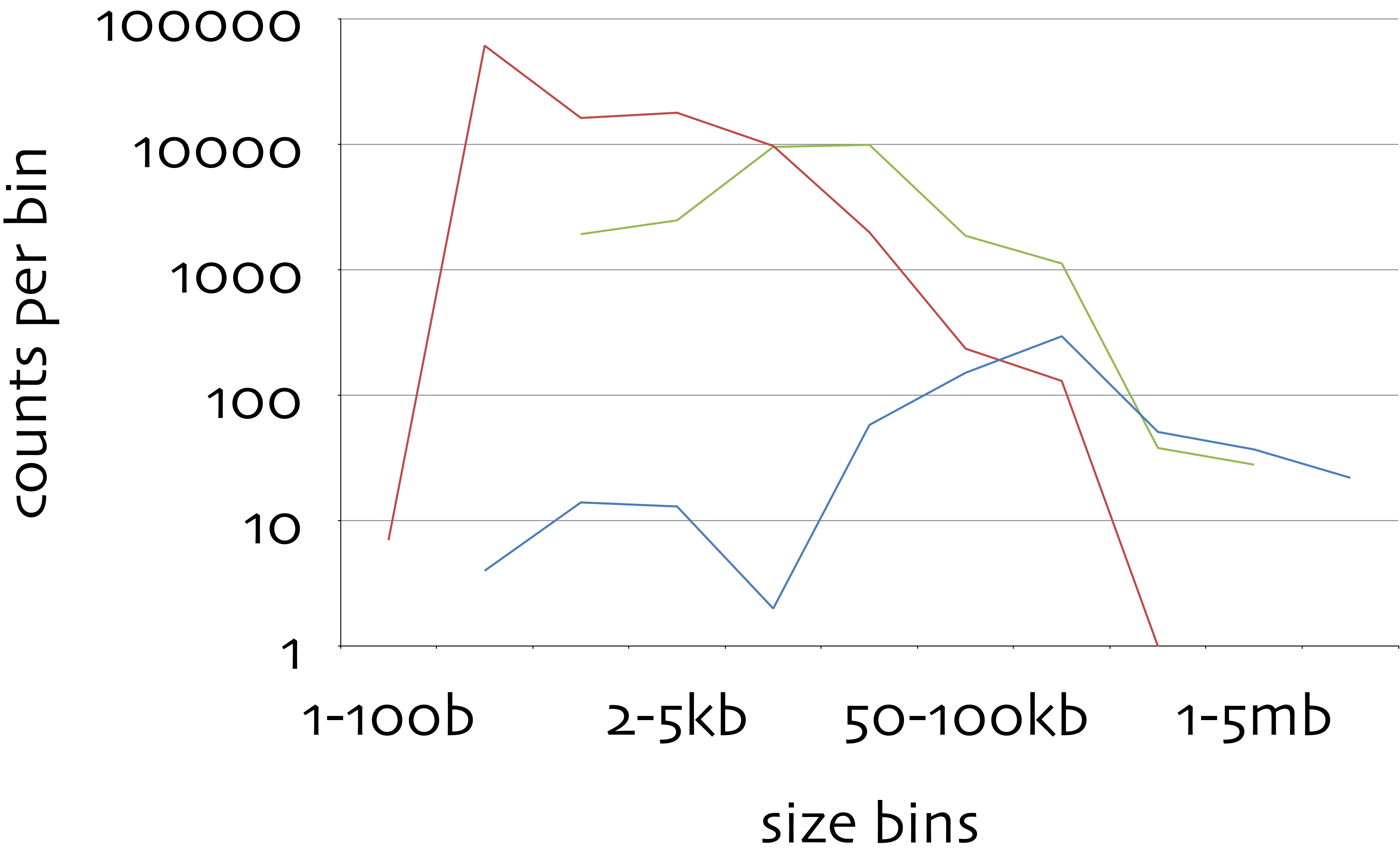
4/100 patients had two pathogenic diagnostic variants

First variant	Disease	Second variant	Disease
NGLY1	congenital glycosylation Iv	COG5	congenital glycosylation Ili
TYR	oculocutaneous albinism 1	MC4R	obesity
ATP1A3	CAPOS syndrome	2p16.3 deletion	contiguous gene syndrome
CCM2	cerebral cavernous malformations	8q22.1 deletion	contiguous gene syndrome

Diagnostic yield: Structural variants detected by WGS

All structural variants

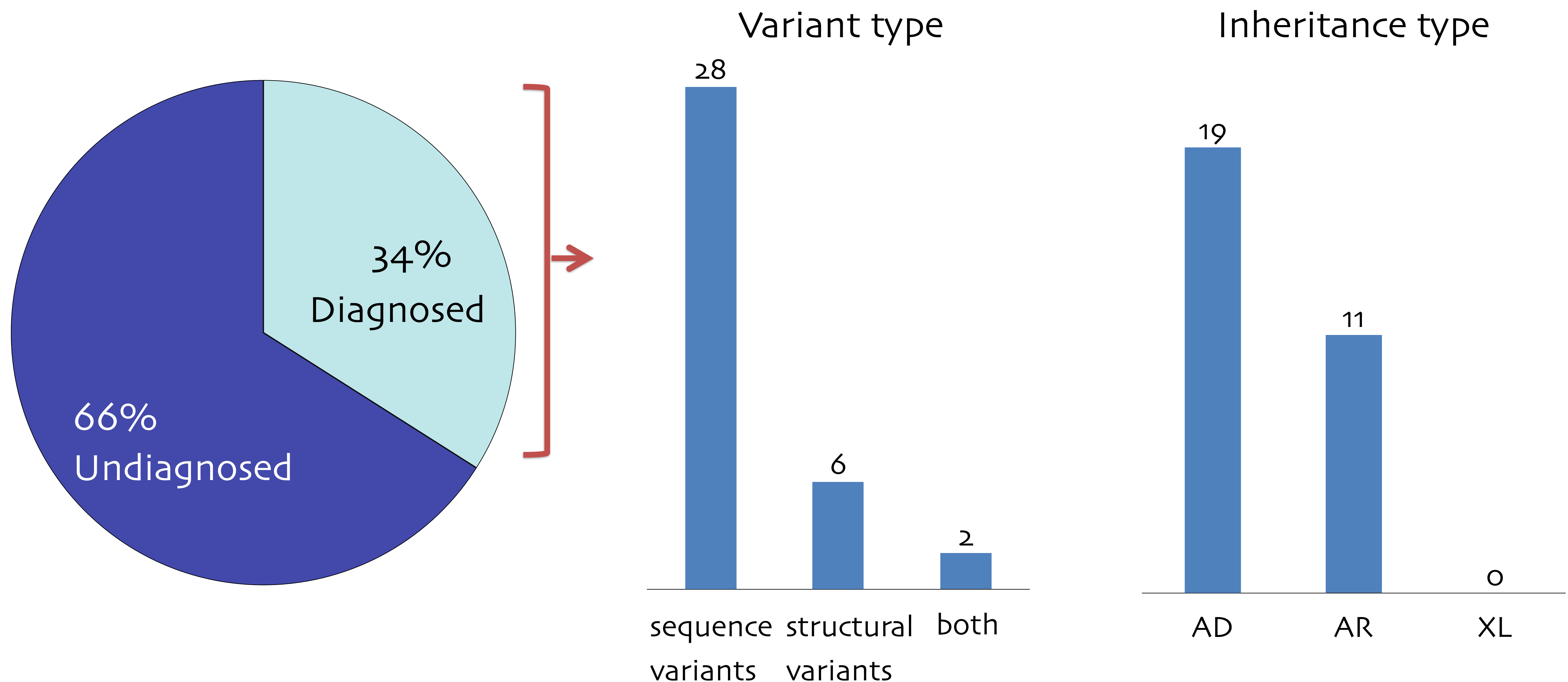
	micro array	read depth	paired end
#/sample	~7	252	1598
Mean size	366kb	30kb	14kb
Median size	116kb	10kb	496bp



Clinically significant variants

Gain/Loss	Clinical Microarray (kb)	Size (Kb)	WGS CNV (kb)	Origin
Loss	ChrX:60 - 91,873	91,813	ChrX:176 - 91,866	De Novo
Gain	ChrX: 91,877 - 155,174	63,297	ChrX:92,796 - 155,260	
Loss	Chr4:33 - 7,608	7,575	Chr4:68 - 7,616	De Novo
Loss	Chr22:18,713 - 21,440	2,727	Chr22:18,888 - 21,466	De Novo
Loss	Chr10:30,822 - 32,872	2,050	Chr10:30,814 - 32,892	De Novo
Loss	Chr22: 35,931 - 37,272	1,342	Chr22:35,890 - 37,302	N/A
Loss	Chr8:97,145 - 98,301	1,156	Chr8:97,134 - 98,308	De Novo
Loss	Chr16:15,507 - 16,400	894	Chr16:15,480 - 16,296	De Novo
Loss	Chr2:51,021 - 51,358	337	Chr2:51,020 - 51,374	De Novo

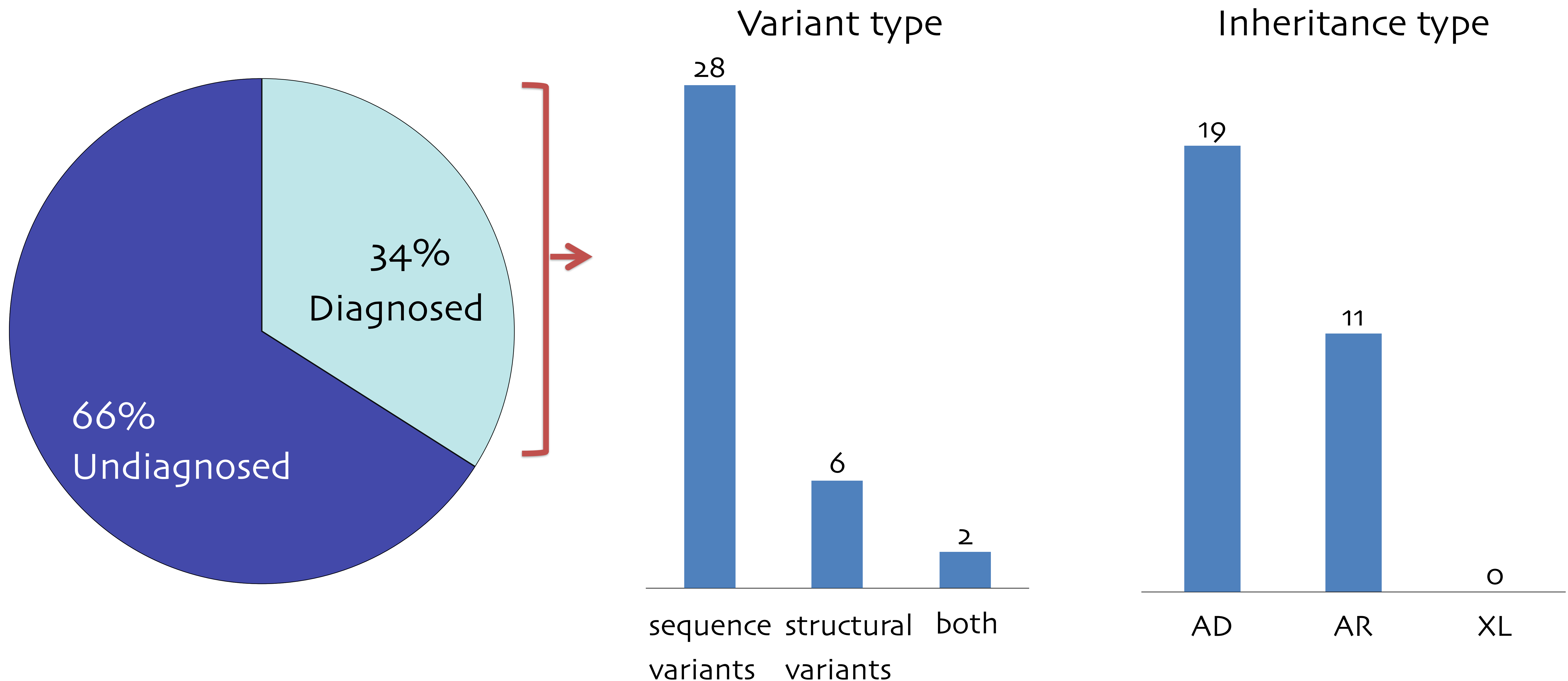
Summary: Diagnostic variants detected by WGS



34% Overall diagnostic rate

- CNV diagnostic rate: 8%, similar to rate over same time period
- Sequence level diagnostic rate: 28%, consistent with WES yield
- Most sequence level diagnostic variants are AD
- 44% of all diagnostic variants are de novo

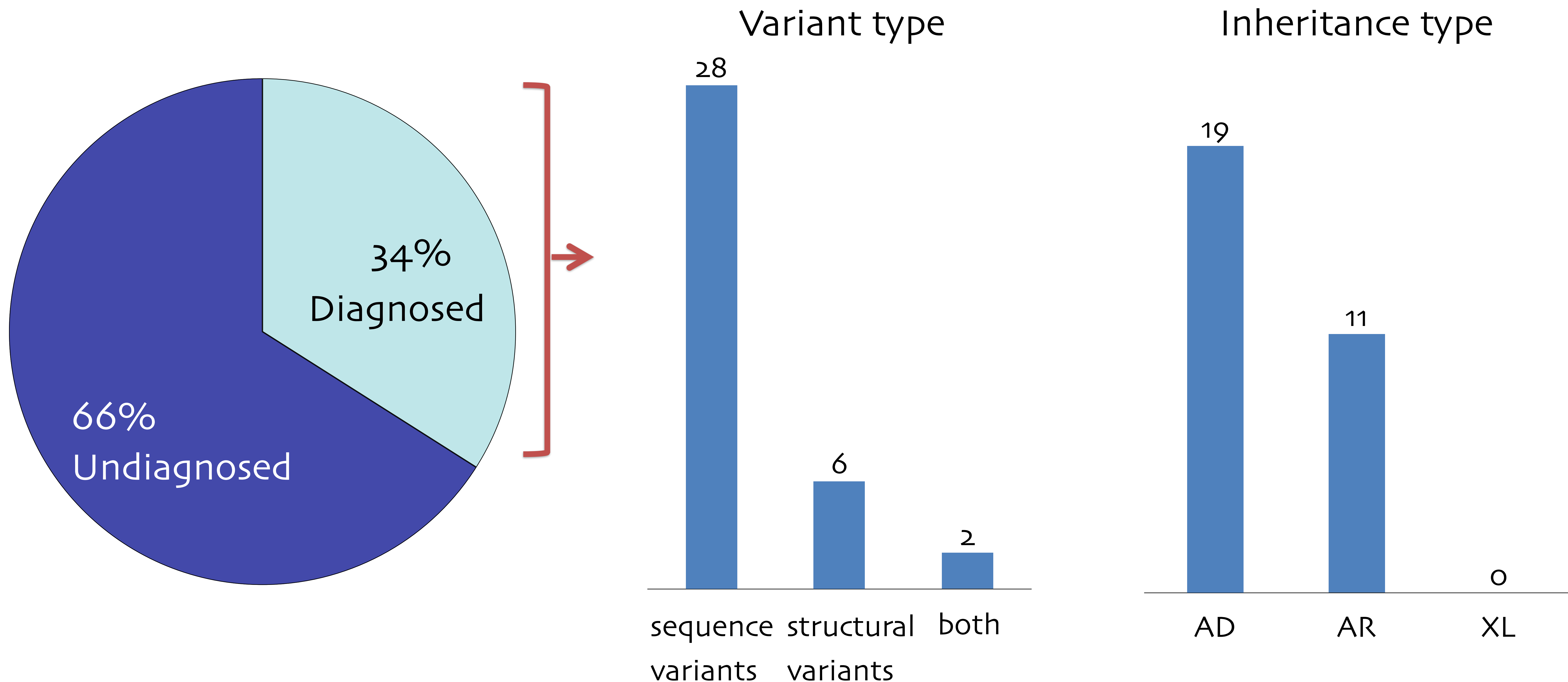
Summary: Diagnostic variants detected by WGS



Limitations

- Only “virtual exome” and “virtual microarray” has been examined
- Large In/Dels and small CNVs are hard to interpret due to lack of population data
- Trios not analysed
- Russell-Silver syndrome and uniparental disomy chromosome 14 were missed

Summary: Diagnostic variants detected by WGS



The benefits of genetics-driven hypothesis-free testing

- 59 microarray cohort patients underwent hypothesis-driven molecular testing
- average of 2 hypothesis driven tests/patient (range 1-13)
- hypothesis-driven testing diagnosed 6 of these 59 patients
- hypothesis-free WGS diagnosed 23 of these 59 patients

CAGI: The SickKids Genome Clinic Challenge

Methods

Team 1

Focus on coding variants - excluded non-coding variants and synonymous variants

Exomiser used to identify candidate variants based on HPO terms extracted from phenotype

Eliminate variants with MAF >1% in ExAc, 1000 genomes and ESP databases

Assessed pathogenicity with Polyphen2, MutationTaster and SIFT

Prioritized variants using Exomiser's hiPHIVE algorithm (examines gene and local interactome)

Top 20 variants assessed manually and ranked based on phenotype matching, likely pathogenicity, and rarity

Team 2:

dcGO (database of domain-centric ontologies) used to match SNPs to phenotypes

fathmm (functional analysis through hidden Markov models) used to predict variant pathogenicity

1000 genomes to assess MAF

Team 3:

Combination of Exomiser, Phenomizer, Phenolyzer, GeneMania, and Phen-Gen

Team 4:

Hierarchical filtering:

Eliminate variants with MAF <5% in 1000 Genomes and ExAc and weight MAF in controls

Classification in databases - E.g., variants reported as disease causing in HGMD or ClinVar

Variant type - E.g., LOF variants prioritized over non-synonymous variants

In silico prediction of non-synonymous variant pathogenicity

Degree of phenotype overlap between OMIM phenotype and patient's phenotype

CAGI: The SickKids Genome Clinic Challenge

Initial assessment of results: Diagnostic variants

4 CAGI teams - 187 variants proposed for 25 patients

- 60 variants were not in genes with OMIM disease phenotypes (gene discovery)
- Almost all variants for AD disorders with 11 X-Linked, 4 AR, no structural variants proposed

MAF were problematic for some variants from teams 2 and 4

- Almost all of Team 1's 26 variants were novel
- 23/55 Team 2 variants had MAF >1%
- All 52 Team 3 variants had MAF <0.01%
- 12/56 Team 4 variants had MAF >1%

Vast majority of variants appeared to be chosen for ability to partially explain phenotype

But - for vast majority, variant's phenotype had little if any overlap with patient's phenotype

Most variants only proposed by 1 CAGI team

- 10 variants proposed by 2 CAGI teams (9/10 have OMIM disease phenotypes)
- 5 variants proposed by 1 CAGI team and SickKids (5/5 have OMIM disease phenotypes)
- Team 2's 55 variants were not proposed by any other team

CAGI: The SickKids Genome Clinic Challenge

Initial assessment of results: Examples from 15 diagnostic variants nominated by 2 teams

- 1011 - COL1A1:c.2782_2789del:p.P928fs (Ehlos Danlos syndrome) proposed by Team 4 and SickKids - partial explanation of phenotype - reverse phenotyping pending
- 1024 - CREBBP:exon31:c.6697_6699TTG p.Gln2233Leu (Rubinstein Taybi syndrome) proposed by Team 4 and SickKids - reverse phenotyping - did not fit clinical phenotype
- 1035 - NIPBL:c.1801dupA:p.P600fs (Cornelia de Lange syndrome) proposed by Team 4 and SickKids - reverse phenotyping - good fit but didn't validate
- 1060 - EHMT1:c.C1253T:p.S418L (Kleefstra syndrome) proposed by 2 teams + SickKids Partial overlap with clinical phenotype - reverse phenotyping and validation pending
- 1106 - KCNB1:c.T623C:p.L208P (Epileptic encephalopathy) proposed by Teams 1 and 3 Reverse phenotyping - excellent phenotype fit with patient - Diagnostic Hit
- 1105 - SMC3:c.A2996T:p.K999M (Cornelia de Lange) proposed by Teams 3 and 4 Reverse phenotyping - very good phenotype fit with patient - Diagnostic Hit

CAGI: The SickKids Genome Clinic Challenge

Initial assessment of results: Predictive variants

3 CAGI teams made predictions for secondary variants

- Teams 3 and 4 only examined the ACMG 2013 gene lists

100 predictive variants proposed by CAGI teams:

- Team 2: 61 variants; Team 3: 5 ACMG variants; Team 4: 34 ACMG variants;
SickKids: 13 variants (after validation)

No variants were proposed by more than 1 CAGI team

- Team 2 proposed many variants in genes without a known clinical phenotype
- Large majority of disease-associated variants were benign or VUS (ACMG 2015 guidelines)
- 9 CAGI variants appeared to be pathogenic/likely pathogenic
- 5 variants (all from Team 4) were also picked by SickKids
- 4 variants were not in the SickKids secondary pipeline

Secondary Variants: the 56 ACMG genes

Published in final edited form as:

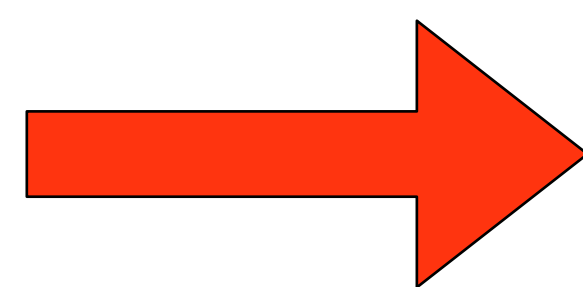
Genet Med. 2013 July ; 15(7): 565–574. doi:10.1038/gim.2013.73.

ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing

Robert C. Green, MD, MPH^{1,2}, Jonathan S. Berg, MD, PhD³, Wayne W. Grody, MD, PhD^{4,5,6}, Sarah S. Kalia, ScM, CGC¹, Bruce R. Korf, MD, PhD⁷, Christa L. Martin, PhD, FACMG⁸, Amy McGuire, JD, PhD⁹, Robert L. Nussbaum, MD¹⁰, Julianne M. O'Daniel, MS, CGC¹¹, Kelly E. Ormond, MS, CGC¹², Heidi L. Rehm, PhD, FACMG^{2,13}, Michael S. Watson, MS, PhD, FACMG¹⁴, Marc S. Williams, MD, FACMG¹⁵, and Leslie G. Biesecker, MD¹⁶

ACMG gene assessment in 25 CAGI challenge genomes:

Initial Sickkids
assessment



TP53 (3' UTR variant) -	Li Fraumeni cancer predisposition syndrome
RYR2:c.C3320T:p.T1107M -	catecholaminergic tachycardia
BRCA1:c.1016dupA:p.K339fs-	Breast/ovarian cancer syndrome
BRCA1:c524dupA:p.K175fs-	Breast/ovarian cancer syndrome
MLH1:c.1376_1377insAGAGAAGAGA:p.S459fs -	Lynch syndrome

Secondary Variants: the 56 ACMG genes

Published in final edited form as:

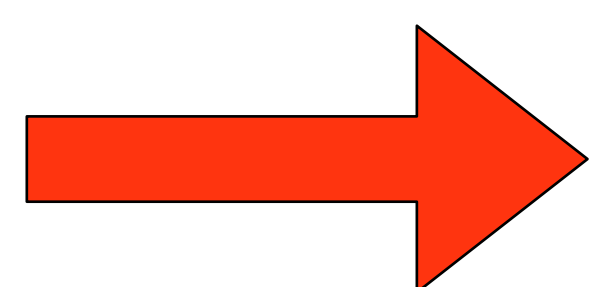
Genet Med. 2013 July ; 15(7): 565–574. doi:10.1038/gim.2013.73.

ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing

Robert C. Green, MD, MPH^{1,2}, Jonathan S. Berg, MD, PhD³, Wayne W. Grody, MD, PhD^{4,5,6}, Sarah S. Kalia, ScM, CGC¹, Bruce R. Korf, MD, PhD⁷, Christa L. Martin, PhD, FACMG⁸, Amy McGuire, JD, PhD⁹, Robert L. Nussbaum, MD¹⁰, Julianne M. O'Daniel, MS, CGC¹¹, Kelly E. Ormond, MS, CGC¹², Heidi L. Rehm, PhD, FACMG^{2,13}, Michael S. Watson, MS, PhD, FACMG¹⁴, Marc S. Williams, MD, FACMG¹⁵, and Leslie G. Biesecker, MD¹⁶

ACMG gene assessment in 25 CAGI challenge genomes:

Initial Sickkids
assessment



Validated

TP53 (3' UTR variant) -
RYR2:c.C3320T:p.T1107M -

Li-Fraumeni cancer predisposition syndrome
catecholaminergic tachycardia

Didn't Sanger validate

BRCA1:c.1016dupA:p.K339fs- Breast/ovarian cancer syndrome

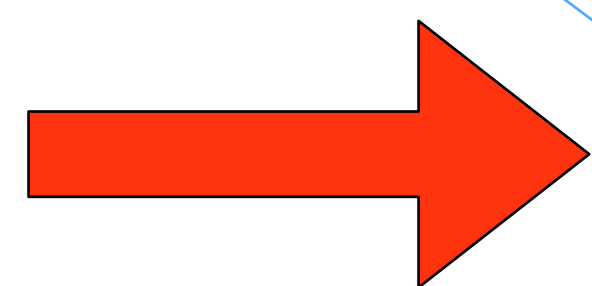
BRCA1:c.524dupA:p.K175fs- Breast/ovarian cancer syndrome

MLH1:c.1376_1377insAGAGAAGAGA:p.S459fs - Lynch syndrome

Secondary Variants: the 56 ACMG genes

ACMG gene assessment in 25 CAGI challenge genomes:

Initial Sickkids
assessment



TP53 (3' UTR variant)	Li Fraumeni cancer predisposition syndrome
RYR2:c.C3320T:p.T1107M	catecholaminergic tachycardia
BRCA1:c.1016dupA:p.K339fs	Breast/ovarian cancer syndrome
BRCA1:c.524dupA:p.K175fs	Breast/ovarian cancer syndrome
MLH1:c.1376_1377insAGAGAAGAGA:p.S459fs	Lynch syndrome

Team 4: 34 Candidate ACMG Variants:

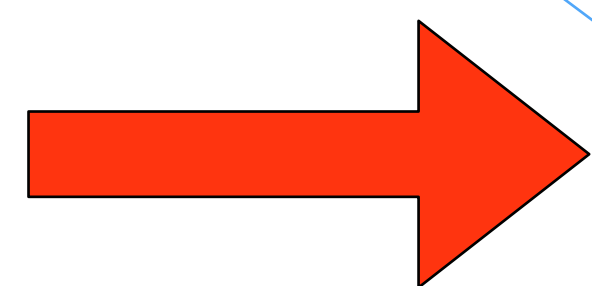
1011 RYR2:c.G5654A:p.G1885E
 1013 NTRK1:c.C1810T:p.H604Y
 1013 NTRK1:c.G1838T:p.G613V
 1013 SCN5A:c.G3578A:p.R1193Q
 1014 SCN5A:c.A1673G:p.H558R
 1014 KCNQ1:c.C1799T:p.T600M
 1019 KCNH2:c.T2848C:p.S950P
 1028 PKP2:c.G209T:p.S70I
 1038 MYH11:c.G4756A:p.D1586N
 1048 TSC1:c.A1760G:p.K587R
 1048 MYLK:c.C479G:p.P160R
 1048 MYLK:c.G1968T:p.W656C
 1052 APC:c.G3152A:p.R1051K
 1052 BRCA1:c.1016dupA:p.K339fs
 1052 MSH6:c.C3524G:p.T1175S

1064 MUTYH:c.C1544T:p.S515F
 1064 MUTYH:c.G1255A:p.A419T
 1064 DSC2:c.56_57CC
 1076 DSP:c.A913T:p.I305F
 1081 MYLK: splice variant
 1086 MYH7:c.G2400T:p.E800D
 1088 MYLK:c.C5441T:p.T1814I
 1088 FBN::c.C8363T:p.T2788M
 1096 TMEM43:c.C934T:p.R312W
 1099 APC:c.C1240T:p.R414C
 1099 MYBPC3:c.G472A:p.V158M
 1099 PKP2:c.G209T:p.S70I
 1105 PKP2:c.G209T:p.S70I
 1106 MYLK:c.C3748T:p.R1250C
 1106 DSP:c.A913T:p.I305F

Secondary Variants: the 56 ACMG genes

ACMG gene assessment in 25 CAGI challenge genomes:

Initial Sickkids
assessment



TP53 (3' UTR variant)	Li Fraumeni cancer predisposition syndrome
RYR2:c.C3320T:p.T1107M	catecholaminergic tachycardia
BRCA1:c.1016dupA:p.K339fs	Breast/ovarian cancer syndrome
BRCA1:c.524dupA:p.K175fs	Breast/ovarian cancer syndrome
MLH1:c.1376_1377insAGAGAAGAGA:p.S459fs	Lynch syndrome

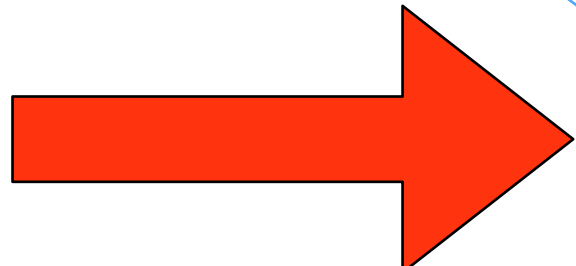
Team 4: 34 Candidate ACMG Variants: (30 Benign or VUS)

1011 RYR2:c.G5654A:p.G1885E
 1013 NTRK1:c.C1810T:p.H604Y
 1013 NTRK1:c.G1838T:p.G613V
 1013 SCN5A:c.G3578A:p.R1193Q
 1014 SCN5A:c.A1673G:p.H558R
 1014 KCNQ1:c.C1799T:p.T600M
 1019 KCNH2:c.T2848C:p.S950P
 1028 PKP2:c.G209T:p.S70I
 1038 MYH11:c.G4756A:p.D1586N
 1048 TSC1:c.A1760G:p.K587R
 1048 MYLK:c.C479G:p.P160R
 1048 MYLK:c.G1968T:p.W656C
 1052 APC:c.G3152A:p.R1051K
 1052 BRCA1:c.1016dupA:p.K339fs
 1052 MSH6:c.C3524G:p.T1175S

1064 MUTYH:c.C1544T:p.S515F
 1064 MUTYH:c.G1255A:p.A419T
 1064 DSC2:c.56_57CC
 1076 DSP:c.A913T:p.I305F
 1081 MYLK: splice variant
 1086 MYH7:c.G2400T:p.E800D
 1088 MYLK:c.C5441T:p.T1814I
 1088 FBN::c.C8363T:p.T2788M
 1096 TMEM43:c.C934T:p.R312W
 1099 APC:c.C1240T:p.R414C
 1099 MYBPC3:c.G472A:p.V158M
 1099 PKP2:c.G209T:p.S70I
 1105 PKP2:c.G209T:p.S70I
 1106 MYLK:c.C3748T:p.R1250C
 1106 DSP:c.A913T:p.I305F

Secondary Variants: the 56 ACMG genes

ACMG gene assessment in 25 CAGI challenge genomes:

Initial Sickkids assessment		TP53 (3' UTR variant)	Li Fraumeni cancer predisposition syndrome
		RYR2:c.C3320T:p.T1107M	catecholaminergic tachycardia
		BRCA1:c.1016dupA:p.K339fs	Breast/ovarian cancer syndrome
		BRCA1:c524dupA:p.K175fs	Breast/ovarian cancer syndrome
		MLH1:c.1376_1377insAGAGAAGAGA:p.S459fs	Lynch syndrome

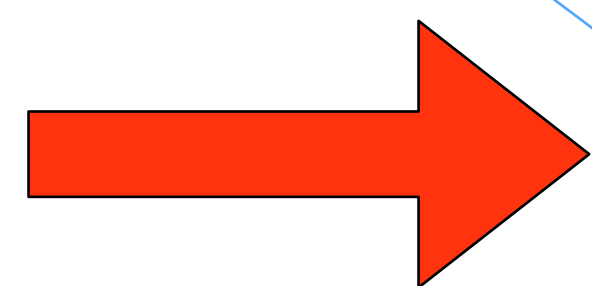
Team 4: 34 Candidate ACMG Variants: (30 Benign or VUS)

	ExAc MAF		ExAc MAF
1011 RYR2:c.G5654A:p.G1885E	2.30%	1064 MUTYH:c.C1544T:p.S515F	1%
1013 NTRK1:c.C1810T:p.H604Y	4.20%	1064 MUTYH:c.G1255A:p.A419T	
1013 NTRK1:c.G1838T:p.G613V	5.40%	1064 DSC2:c.56_57CC	
1013 SCN5A:c.G3578A:p.R1193Q		1076 DSP:c.A913T:p.I305F	3.90%
1014 SCN5A:c.A1673G:p.H558R		1081 MYLK: splice variant	
1014 KCNQ1:c.C1799T:p.T600M		1086 MYH7:c.G2400T:p.E800D	
1019 KCNH2:c.T2848C:p.S950P		1088 MYLK:c.C5441T:p.T1814I	
1028 PKP2:c.G209T:p.S70I	2.40%	1088 FBN::c.C8363T:p.T2788M	
1038 MYH11:c.G4756A:p.D1586N		1096 TMEM43:c.C934T:p.R312W	
1048 TSC1:c.A1760G:p.K587R	1.80%	1099 APC:c.C1240T:p.R414C	
1048 MYLK:c.C479G:p.P160R	1.60%	1099 MYBPC3:c.G472A:p.V158M	16%
1048 MYLK:c.G1968T:p.W656C		1099 PKP2:c.G209T:p.S70I	4.3%
1052 APC:c.G3152A:p.R1051K		1105 PKP2:c.G209T:p.S70I	4.3%
1052 BRCA1:c.1016dupA:p.K339fs		1106 MYLK:c.C3748T:p.R1250C	
1052 MSH6:c.C3524G:p.T1175S		1106 DSP:c.A913T:p.I305F	2.90%

Secondary Variants: the 56 ACMG genes

ACMG gene assessment in 25 CAGI challenge genomes:

Initial Sickkids
assessment



TP53 (3' UTR variant)	Li Fraumeni cancer predisposition syndrome
RYR2:c.C3320T:p.T1107M	catecholaminergic tachycardia
BRCA1:c.1016dupA:p.K339fs	Breast/ovarian cancer syndrome
BRCA1:c.524dupA:p.K175fs	Breast/ovarian cancer syndrome
MLH1:c.1376_1377insAGAGAAGAGA:p.S459fs	Lynch syndrome

Team 4: 34 Candidate ACMG Variants: (4 Likely pathogenic/Possibly pathogenic)

1052 APC:c.6573dupA:p.G2191fs	very rare fs (1x in ExAc) but 6:41 reads and occurred 2x in dataset
1064 BRCA1:c.524dupA:p.K175fs	novel fs near end of gene but 5:26 reads and didn't validate
1106 FBN1:c.G1027A:p.G343R	0.015% MAF, SIFT/PolyPhen path, 1 case report
1105 MLH1:c.1376_1377insAGAGAAGAGA:p.S459fs	novel fs, affects 100% transcripts, but 3:14 reads

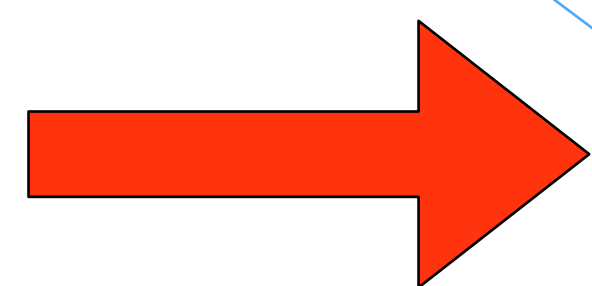
Team 3: 5 Candidate predictive variants:

1019 FRMD4A:c.T2188G:p.F730V	1038 CHEK2:c.A1298G:p.Y433C
1024 TNXB:c.T9023A:p.L3008Q	1088 CHEK2:c.A1298G:p.Y433C
1096 BRCA2:c.A7234T:p.T2412S	

Secondary Variants: the 56 ACMG genes

ACMG gene assessment in 25 CAGI challenge genomes:

Initial Sickkids
assessment



TP53 (3' UTR variant)	Li Fraumeni cancer predisposition syndrome
RYR2:c.C3320T:p.T1107M	catecholaminergic tachycardia
BRCA1:c.1016dupA:p.K339fs	Breast/ovarian cancer syndrome
BRCA1:c.524dupA:p.K175fs	Breast/ovarian cancer syndrome
MLH1:c.1376_1377insAGAGAAGAGA:p.S459fs	Lynch syndrome

Team 4: 34 Candidate ACMG Variants: (4 Likely pathogenic/Possibly pathogenic)

1052 APC:c.6573dupA:p.G2191fs	very rare fs (1x in ExAc) but 6:41 reads and occurred 2x in dataset
1064 BRCA1:c.524dupA:p.K175fs	novel fs near end of gene but 5:26 reads and didn't validate
1106 FBN1:c.G1027A:p.G343R	0.015% MAF, SIFT/PolyPhen path, 1 case report
1105 MLH1:c.1376_1377insAGAGAAGAGA:p.S459fs	novel fs, affects 100% transcripts, but 3:14 reads

Team 3: 5 Candidate predictive variants:

1019 FRMD4A:c.T2188G:p.F730V	VUS (insufficient evidence)
1024 TNXB:c.T9023A:p.L3008Q	VUS (insufficient evidence)
1096 BRCA2:c.A7234T:p.T2412S	VUS (insufficient evidence)

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