Stickler syndrome: present, and future

2013 17th National Stickler Syndrome Conference
Las Vegas, NV
July 13, 2013

Joseph V. Thakuria, MD, MMSc
Medical Genetics and Metabolism Attending Physician
Director, MGH Stickler Syndrome Program
Medical Director, Personal Genome Project
Harvard Catalyst CRC Protocol Review Committee
MGH Center for Human Genomics Research
Massachusetts General Hospital
Harvard Medical School

HARVARD
MEDICAL SCHOOL
TEACHING AFFILIATE

Stickler syndrome

- Estimated to be 1 in 7500-9000 *newborns*
- Complete penetrance, variable expressivity (even intrafamilial)
- No strict clinical diagnostic criteria. Consider when 2 or more of the following are present:
- Ophthalmologic: congenital or early cataracts, vitreous anomaly, retina detachment, myopia (typically -3 diopters; or any newborn with myopia)

Stickler syndrome (cont.)

- Craniofacial: midface hypoplasia, depressed nasal bridge, anteverted nares (all more pronounced in childhood), bifid uvula, cleft of the hard palate, micrognathia
- Audiologic: hypermobile inner ear systems, sensorineural and conductive hearing loss
- Orthopedic: early OA, hypermobility, mild spondyloepiphyseal dysplasia

Stickler syndrome known genes

- 5 genes currently known; respectively Stickler syndrome types I-V:
- *COL2A1:* 12q13.11; AD (50% risk to offspring), 80-90% of cases
- *COL11A1:* 1p21.1; AD, 10-20% of cases
- *COL11A2:* 6p21.32; AR (25% risk to offspring), rare
- *COL9A1:* 6q13; AR, rare
- *COL9A2:* 1p34.2; AR, rare

Stickler syndrome medical management

- Craniofacial: infants with more severe disease and Robin sequence may require tracheostomy and/or mandibular advancement procedures
- Eyes: correction of refractive errors with glasses; standard tx for retinal detachment; avoid activities like contact sports that may lead to retinal detachment
- Audiologic: standard tx for sensorineural and conductive hearing loss
- CV: consider abx prophy in some pts w/ MVP
- Ortho: symptomatic consider NSAIDs before and after activity in some patients

Stickler syndrome medical surveillance

- Annual exam by retinal specialist
- Audiologic exam every 6mos until age 5yo; then annually
- Screen for MVP (usu reserve echo for pts w/sxs)
- When a molecular diagnosis is made it's important to test at risk family members to determine who needs surveillance

Genetically related allelic disorders

- *COL2A1* (AD):
- achondrogenesis type II
- hypochondrogenesis
- spondyloepiphyseal dysplasia
- Kneist dysplasia
- spondyloperipheral dysplasia
- spondyometaphyseal dysplasia
- platyspondylic lethal skeletal dysplasia, Torrance type (PLSDT)
- osteoarthritis with mild chondrodysplasia
- avascular necrosis of femoral head, primary (ANFH)

Genetically related allelic disorders (cont.)

- *COL11A1* (AD): Marshall syndrome (ocular hypertelorism, flat facial profile)
- *COL11A2:*
- Autosomal recessive otospondylometaepiphyseal dysplasia (OSMED)
- Weissenbach-Zweymuller syndrome (WZS); AD
- Nonsyndromic sensorineural hearing loss (DFNA13); AD
- *COL9A1* and *COL9A2:* Multiple epiphyseal dysplasia 6 (EDM6) and multiple epiphyseal dysplasia 2 (EDM2)

The future

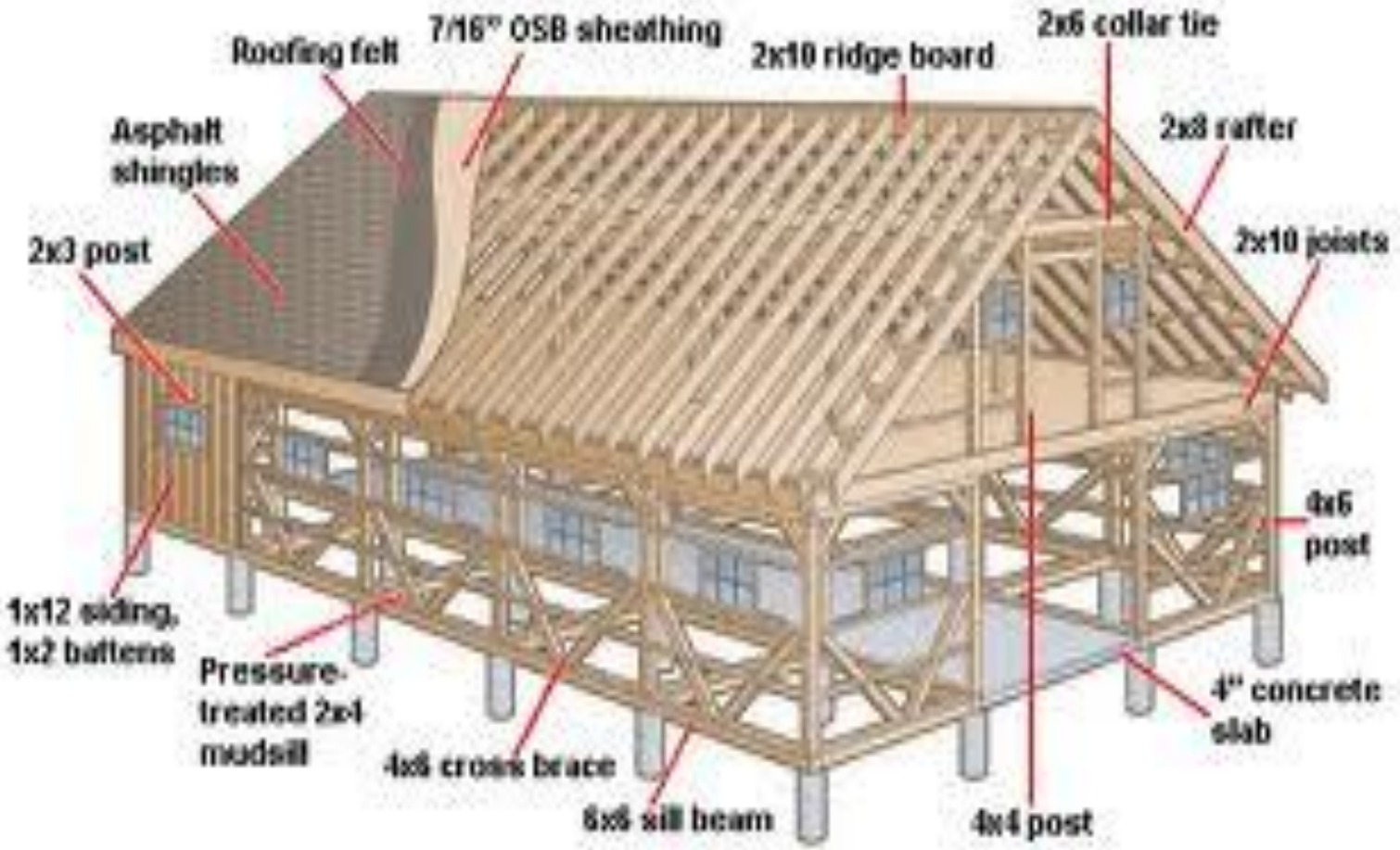
- To what extent do mutations in these same genes account for milder manifestations that look like common adult onset disease (OA, eye, HL)?
- What other genes are involved for those patients without a molecular diagnosis after testing of all 5 genes? (WGS/exome sequencing)
- In patients with a molecular diagnosis, what other genetic modifiers can be discovered that accounts for variable expression especially within families? (WGS/exome)

The future (cont.)

- Would sequence data on all 5 genes +/- other collagen genes help in predicting clinical severity? (WGS/exome)
- To what extent will other NGS modalities help (transcriptomes, immunomes, microbiomes, etc) understand the syndrome better? To what extent will these help in producing viable drug targets? (CFTR and SCD BUT drug screening also better)

Fundamentals:

1. Become familiar with basic human genomic stats/architecture
2. Apply variant filtering algorithms in genomic data analyses
3. Know the minimum sequencing coverage depths needed for accurate base calling and zygosity


Roofing felt
7/16" OSB sheathing
2x10 ridge board
2x6 collar tie
Asphalt shingles
2x8 rafter
2x3 post
2x10 joists
4x6 post
1x12 siding, 1x2 battens
Pressure-treated 2x4 mudsill
4x6 cross brace
4" concrete slab
6x6 sill beam
4x4 post

Fundamentals:

1. Become familiar with basic human genomic stats/architecture
2. Apply variant filtering algorithms in genomic data analyses
3. Know the minimum sequencing coverage depths needed for accurate base calling and zygosity


EXIT 12
NORTH
95
Portsmouth
NH
NORTH
93
Boston
2 MILES

Fundamentals:

1. Become familiar with basic human genomic stats/architecture
2. Apply variant filtering algorithms in genomic data analyses
3. Know the minimum sequencing coverage depths needed for accurate base calling and zygosity


SPEED
LIMIT
85

Variant filtering

- In whole genome/exome analysis:
- 1) Different filters and stringencies can be applied depending on preferences (discovery vs diagnostic) and specific case (eg, healthy study participant vs. patient with personal or family history of genetic disease)
- 2) Avg “hits” applying different filters (eg, not >1mil – though this is always where you start)
- 3) Some pitfalls (current state and surmountable)

Interspecies Variation

99.9% similarity in randomly chosen people (some data suggests 99.5%)

As low as 99.1% if you factor in CNVs

- Human = 1 in 1,000bp
- Chimp = 1 in 300
- Fruit flies = 1 in 80

This comparatively low level of DNA variation in humans (from being evolutionarily recent) works to our advantage in clinical analyses.

RNA analyses: greater diversity in humans

Human Variation

Individually: nonreference SNVs/genome= >3-4mil

Collectively: dbSNP (homo sapiens; build 137):
>53mil, w/ close to 36mil with freq data

- expected to significantly expand with new entries from emerging large scale next gen efforts (including 1KG, ClinSeq, PGP, etc)
- Genetics/env&accidents/ age (esp. extreme ages)

Relative Risk Ratios for Siblings of Probands

Disease	λ^r
Autism	150
Type I diabetes mellitus	35
Crohn's disease	25
Multiple sclerosis	24
Schizophrenia	12

Concordance Rates in Monozygotic and Dizygotic Twins

Concordance

Disorder	MZ (%)	DZ (%)	% difference
Psoriasis	72	15	57
Epilepsy (nontraumatic)	70	6	64
Autism	64	9	55
Bipolar disorder	62	8	54
Schizophrenia	46	15	31
Type I diabetes	40	4.8	35.2
Osteoarthritis	32	16	16
Cleft lip (with or without cleft palate)	30	2	28
Multiple sclerosis	17.8	2	15.8
Rheumatoid arthritis	12.3	3.5	8.8

Mutation rates (u) for different classes of polymorphic markers

	Mutation rate (u)
Minisattelites	1 X 10^-1 – 1 X10 ^-2
Microsatellites	1 X 10^-2 – 1 X10 ^-4
Structural polymorphisms	1 X 10^-3 – 1 X10 ^-5
Single base substitutions	1 X 10^-5 – 1 X10 ^-9
Retroelement insertions	1 X 10^-10 – 1 X10 ^-12

Clinical Genetics

- Clinical genetics to clinical genomics
- Historical technical limitations (starting with low resolution karyotypes) on our ability to recognize genetic disease has resulted in bias towards:
- severe disease
- monogenic: strongly acting single genes; predictive, highly penetrant
- early onset: often pediatric populations

Clinical Genetics (cont.)

- We've come a long way:
- NBS: CF, SCD, enz def (eg, pku); expanding menu/expanding phenotypes; clinical outcomes data for newer offerings (UCD, FAOD) are actively being gathered
- Positive family history; more to offer (100 new genes/year); ACTA2 in FTAAD
- Chromosomal disorders; improving resolution; CGH microarray (diagnostic rate in DD: 5-20%)
- Other "Mendelian" disorders: high penetrance; most genes/pathways involved in rare monogenic disease also have some role in more common pediatric and adult diseases

Genetics Clinic:
July 2013

- Genetic testing: karyotype (~$1500), CGH microarray (~$3500->$900-1500), methylation studies, TIEF, variety of biochemical tests, mutation analysis of coding regions in 1-5 genes (depending on clinical history/presentation)
- Diagnostic sequencing costs/gene = $800-3500 (size dependent), avg $1000-2500/gene
- Other medical: eg, 1 day in a hospital; surgeries
- Benefits: Ultimate goal is targeted therapies; at the very least, individual risk assessment and family planning. In some cases (eg, cancer or ao diss risk), life-saving surveillance/preventative measures.
- Avoid “dx odyssey” and more invasive testing: W/o dx: More invasive testing: skin bx, muscle bx, MRI w/ sedation, spinal tap, even liver or heart bx when appropriate

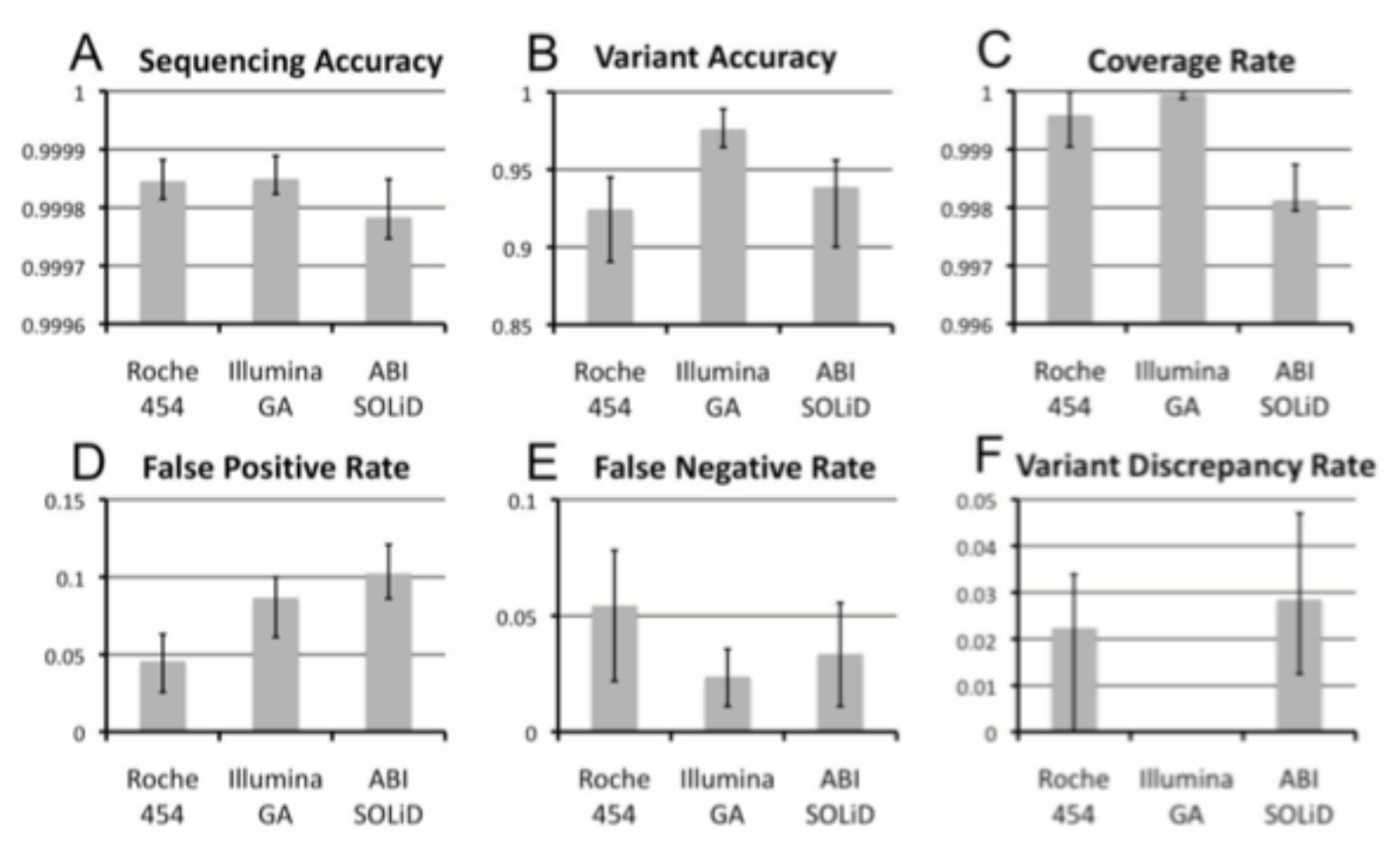
Examples of heterozygosity for a Mendelian d/o being a risk factor for complex dz

Gene mutation	Increased risk
Methylenetetrahydrofolate reductase (MTHFR)	Atherothrombotic disease
Factor V and prothrombin	Stroke, recurrent miscarriages
Alpha-1-antitrypsin	COPD
CFTR	Obstructive azoospermia, Chronic pancreatitis
Glycerol kinase	DM
Glucocerebrosidase	Parkinson disease
	Also XL d/o, GWAS hits in mendelian genes

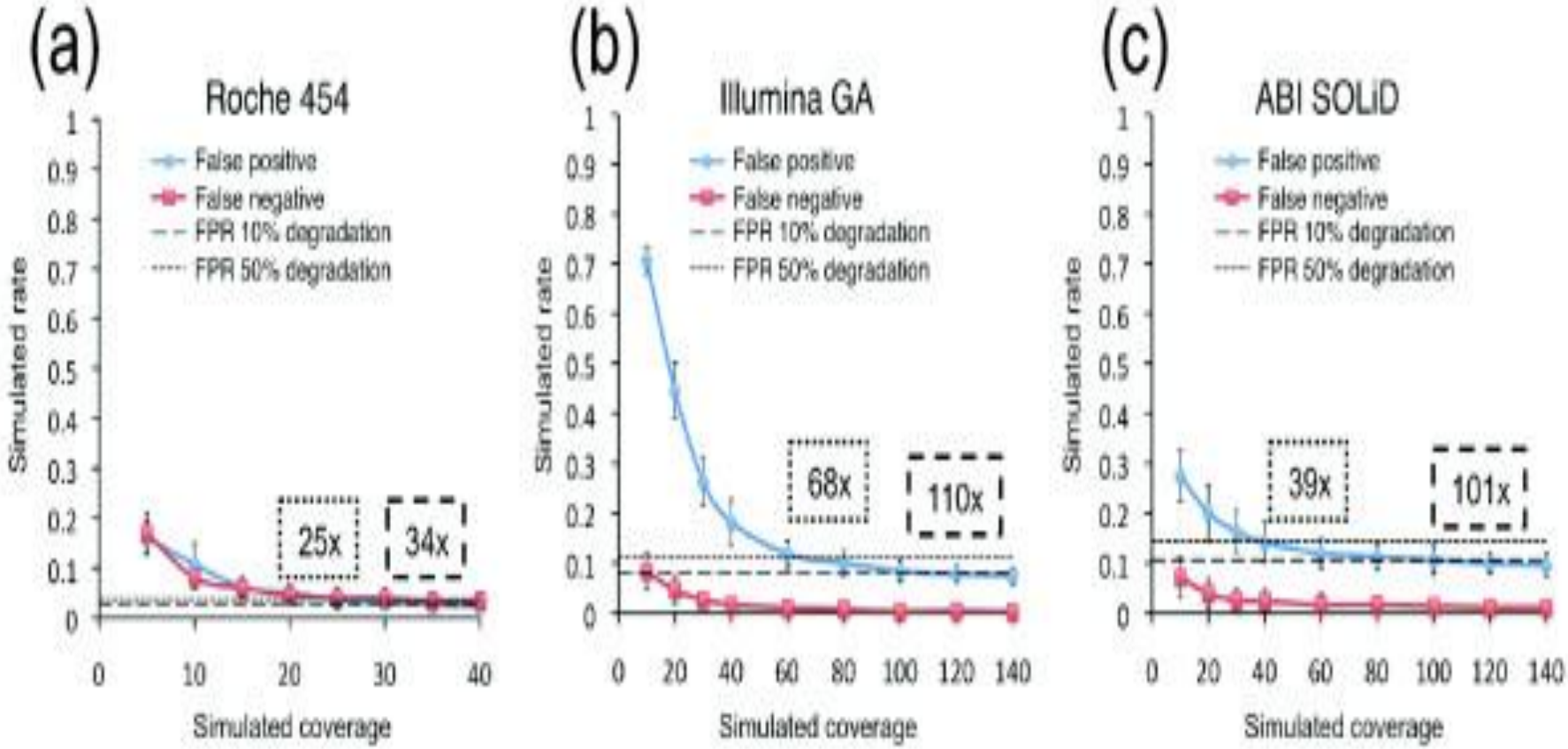
Table 1. Examples of Mendelian traits that have provided new understanding of the background of common diseases

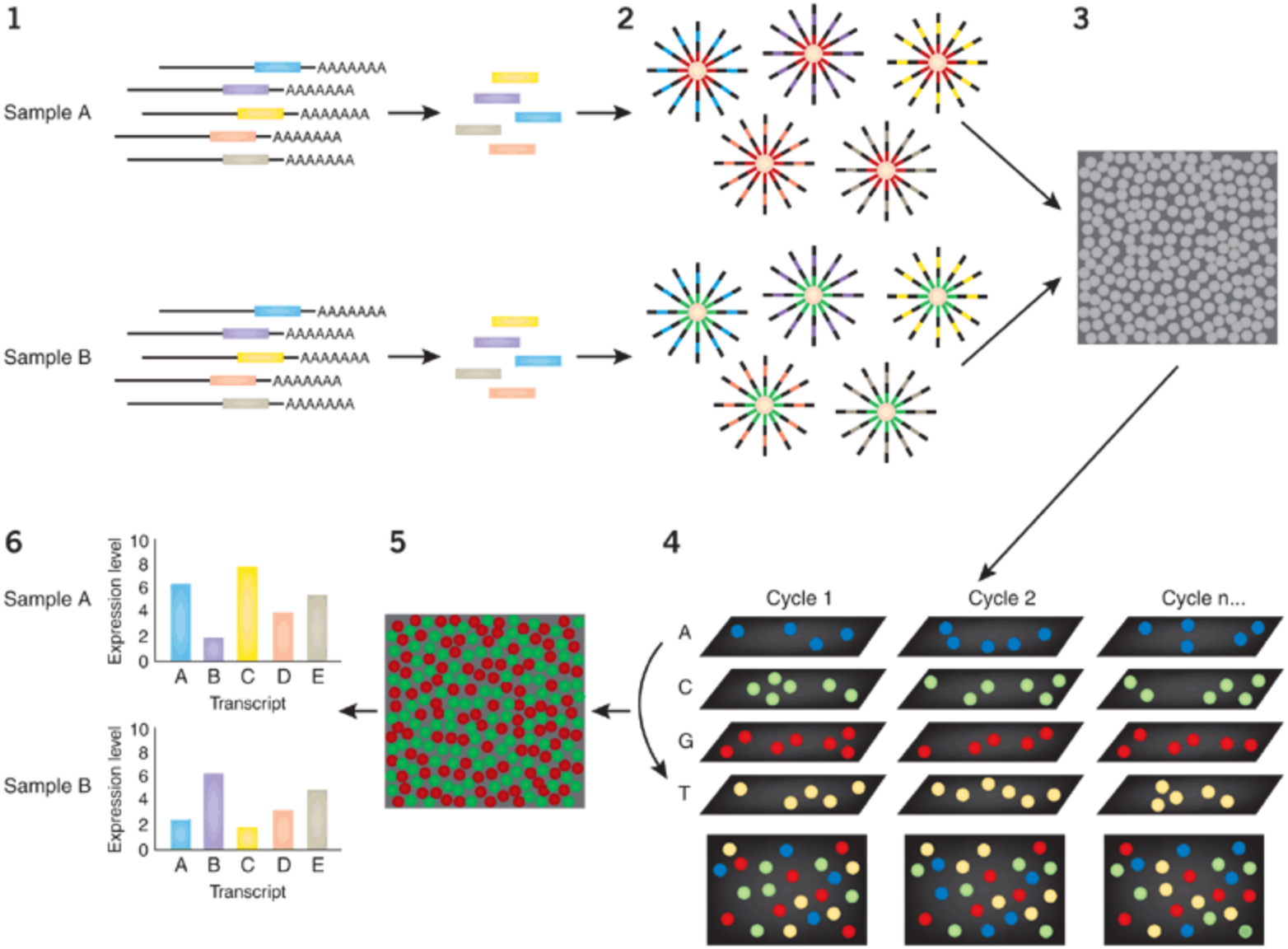
Common trait	Monogenic, Mendelian form	Gene/gene group mutated	Impact for common trait
Psychiatric disorders	Familial autism, rare syndromes translocations	Robo1, FOXP2, DISC1, neuroligins	Suggestive role established in various study samples Functional data in cellular systems
Alzheimer's disease	Familial early onset forms	Presenilins, APP	Molecular mechanism behind amyloid plaques
Epilepsy	Familial epilepsies	Ion channels, e.g. SCN1A, SCN1B, KCNQ2, KCNQ3	Strong hypothesis for role of ion channels and related pathways
Headache disorders	Familial hemiplegic migraine	Ion channels, CACNA1A, SCN1A, ATP1A2	Strong hypothesis for role of ion channels and related pathways
Cardiac arrhythmias	Long QT syndrome	Ion channels, e.g. KCNQ1, HERG, SCN5A, KCNE1, KCNE2	Suggestive background for unexplained sudden death
Dyslipidemias	Tangier's disease, low HDL, familial hyperlipidemias	ABCA1 transporter USF1 LDL receptor APOB, ApoA1, LCAT, ABCG5, ABCG8, ARH	Functional data on the effect of cellular cholesterol transport and metabolism
Hypertension	Rare syndromes causing hypo- or hypertension	Genes associated with renal salt balance e.g. ENaC, mineralocorticoid receptor gene	Several genes also associated with essential hypertension. Pinpointing basic mechanisms of blood pressure regulation
Obesity	Rare syndromes rodent models	Brain- and adipose tissue-derived hormones, leptin receptor, MC4R, Lipin	Importance of brain- and adipose tissue-derived hormones and the hormonal cross-talk between fat tissue and the hypothalamus-hypophysis axis
Diabetes	Familial forms, MODY	Glucokinase transcription factors	Lessons of mechanisms resulting in beta cell dysfunction
Autoimmune diseases	APECED, ALPS, IPEX	Transcription regulators, Fas-receptors and their ligands, caspases	The critical role of central tolerance in autoimmunity

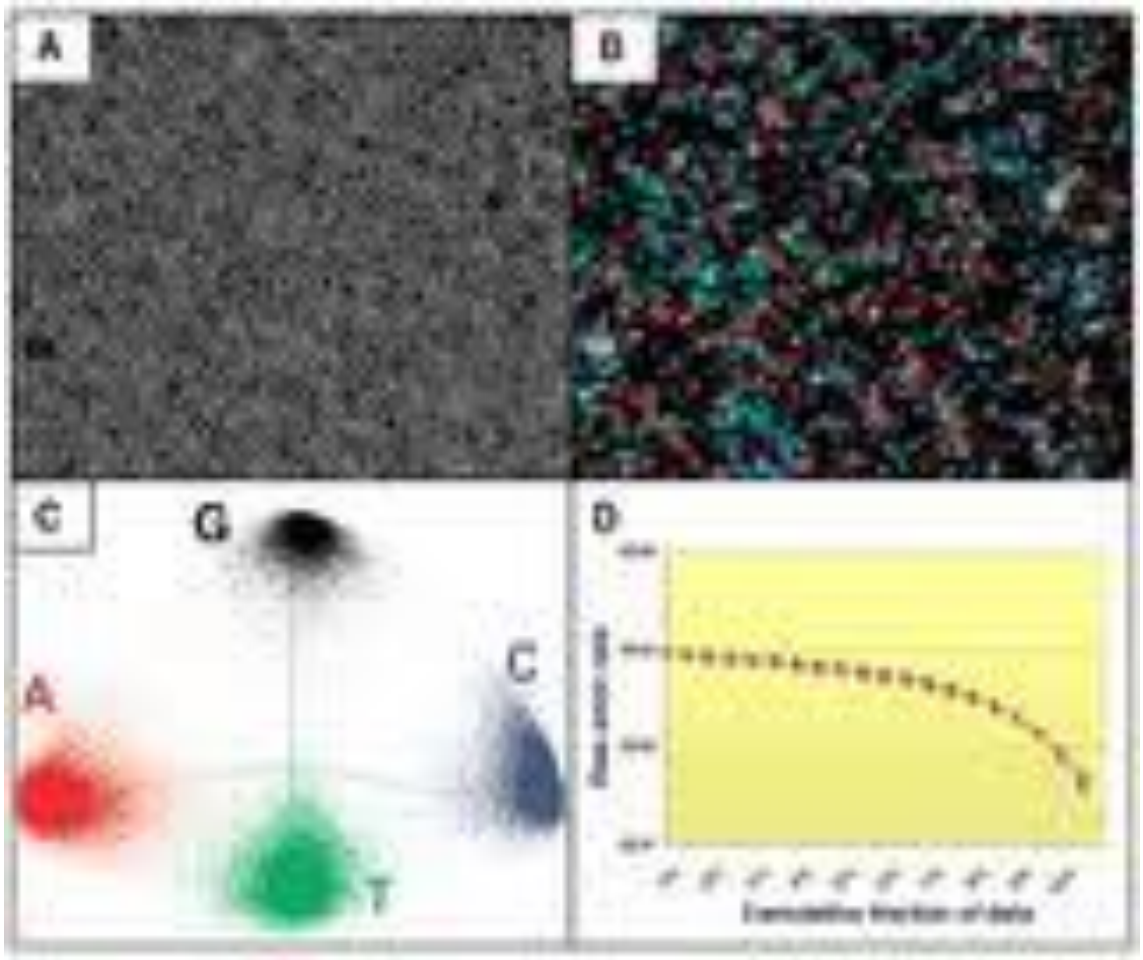
Novel genes and pathways identified through the study of exceptional families segregating a disease in Mendelian fashion have played a decisive role in the characterization of more common forms of diseases with milder phenotypes. This table attempts to summarize some of the best examples. A complete list of references is available at http://www.genome.helsinki.fi/publications/peltonen/table2.



Harismendy et al., Genome Biol. 2009, pmid: 19327155


(a)
Roche 454
False positive
False negative
FPR 10% degradation
FPR 50% degradation
25x
34x
Simulated rate
Simulated coverage
(b)
Illumina GA
False positive
False negative
FPR 10% degradation
FPR 50% degradation
68x
110x
Simulated rate
Simulated coverage
(c)
ABI SOLiD
False positive
False negative
FPR 10% degradation
FPR 50% degradation
39x
101x
Simulated rate
Simulated coverage


1
2
3
Sample A
AAAAAAA
Sample B
4
5
6
Cycle 1
Cycle 2
Cycle n...
A
C
G
T
Expression level
10
8
6
4
2
0
A B C D E
Transcript
Sample A
Sample B


A
B
C
G
A
C
T
D


p24.2
p23
p22.3
p22.1
p21.2
p21.1
p13.3
p13.1
p11.2
q11
q12
q13
q21.13
q21.31
q21.33
q22.2
q22.32
q31.1
q31.2
q32
q33.1
q33.2
q33.3
q34.11
q34.2
151 bp
,155,600 bp
22,155,620 bp
22,155,640 bp
22,155,660 bp
22,155,680 bp
22,155,700 bp
22,155,720 bp
22,155,740 bp

Rapid dx in ¾ neonates (sz, heterotaxy, derm)

DIAGNOSTICS

Rapid Whole-Genome Sequencing for Genetic Disease Diagnosis in Neonatal Intensive Care Units

Carol Jean Saunders,[1,2,3,4,5]* Neil Andrew Miller,[1,2,4]* Sarah Elizabeth Soden,[1,2,4]* Darrell Lee Dinwiddie,[1,2,3,4,5]* Aaron Noll,[1] Noor Abu Alnadi,[4] Nevene Andraws,[3] Melanie LeAnn Patterson,[1,3] Lisa Ann Krivohlavek,[1,3] Joel Fellis,[6] Sean Humphray,[6] Peter Saffrey,[6]
Zoya Kingsbury,[6] Jacqueline Claire Weir,[6] Jason Betley,[6] Russell James Grocock,[6] Elliott Harrison Margulies,[6] Emily Gwendolyn Farrow,[1] Michael Artman,[2,4] Nicole Pauline Safina,[1,4]
Joshua Erin Petrikin,[2,3] Kevin Peter Hall,[6] Stephen Francis Kingsmore[1,2,3,4,5]†

Sci Transl Med 3 October 2012

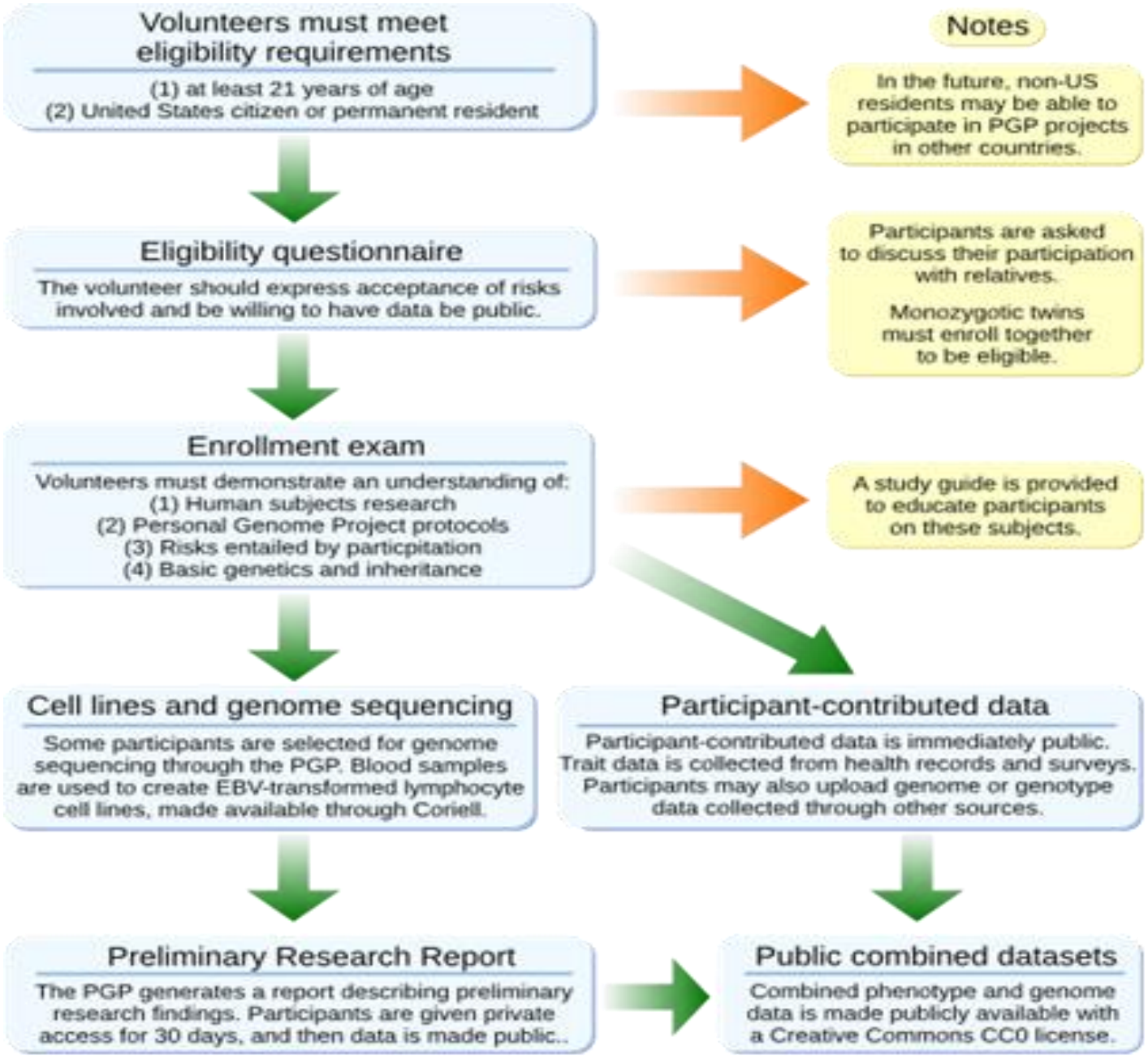
Genomics Research

Eg, Personal Genome Project:
http://www.personalgenomes.org/

- HMS IRB approved to enroll 100,000 participants
- Comprehensive phenotyping, genome sequencing, tissue banking and distribution, iPS and other cell line generation
- Costly: $10K/participant?; now $2500-3000 (WGS, blood draw, cell line, bioinfo)
- Time consuming: recruitment/enrollment/consenting
- Analysis=many variants with 1- 10 literature hits

PGP stats

- ~2000 fully enrolled; self-selected group
- >1000 have uploaded EHR's
- 200 samples currently in the sequencing pipeline
- Over 70 genomes with detailed phenotyping completed


Volunteers must meet eligibility requirements
(1) at least 21 years of age
(2) United States citizen or permanent resident
Notes
In the future, non-US residents may be able to participate in PGP projects in other countries.
Eligibility questionnaire
The volunteer should express acceptance of risks involved and be willing to have data be public.
Participants are asked to discuss their participation with relatives.
Monozygotic twins must enroll together to be eligible.
Enrollment exam
Volunteers must demonstrate an understanding of:
(1) Human subjects research
(2) Personal Genome Project protocols
(3) Risks entailed by particpitation
(4) Basic genetics and inheritance
A study guide is provided to educate participants on these subjects.
Cell lines and genome sequencing
Some participants are selected for genome sequencing through the PGP. Blood samples are used to create EBV-transformed lymphocyte cell lines, made available through Coriell.
Participant-contributed data
Participant-contributed data is immediately public.
Trait data is collected from health records and surveys.
Participants may also upload genome or genotype data collected through other sources.
Preliminary Research Report
The PGP generates a report describing preliminary research findings. Participants are given private access for 30 days, and then data is made public..
Public combined datasets
Combined phenotype and genome data is made publicly available with a Creative Commons CC0 license.

Variant Prioritization

- Given recent technological advances in sequencing, variant prioritization is critical
- NCBI ref/ PGP / Watson / Venter
- High priority: known variants in known disease genes especially those carrying significant medical risk with available therapeutic or preventative measures
- Medium priority: includes novel or functional variants or variants where disease segregation is incomplete or conflicting in known pathogenic genes; especially important to put fhx and phenotype in context
- Low priority: variants in genes with unknown function or clinical relevance

GET-Evidence: 6 lines of evidence

- 5 star rankings for each category:
- 4 variant-specific:
- Computational
- Functional
- Population data (case/controls;ORs)
- Familial disease segregation
- 2 disease-specific: severity and treatability

GET-Evidence (in silico)

- 5 star ranking system on:
- **Computational**: One star for each consistent prediction and one start subtracted for conflicting results from:
- BLOSUM100 ≤ -4
- Other variants in the same gene are associated with the disease (presence in GeneTests is automatically used as a proxy for this)
- Evolutionary conservation (-1 if not conserved)
- Presence in active domain
- Polyphen/SIFT prediction of pathogenic effect (-1 if predicted to be benign)
- Nonsense or frame shift mutation
- rank predictability of different programs

GET-Evidence (in vitro and non-human)

- **Functional**: One star for each experiment supporting the result, and penalize one star for conflicting results from:
- enzyme extracts, cell lines, animal models, etc...

GET-Evidence (human)

- **Case/Control: Odds Ratio**
- 0 stars for OR<1
- 1 star for 1<OR<1.5
- 2 stars for 1.5<OR<2
- 3 stars for 2<OR<3
- 4 stars for 3<OR<5
- 5 stars for OR>5

GET-Evidence (human cont.)

- **Familial Disease Segregation**
- (-1) if there is conflicting evidence and best LOD < 0.5
- (0) no familial evidence, or fails to meet the other thresholds
- (1) best LOD ≥ 1
- (2) best LOD ≥ 1.3
- (3) best LOD ≥ 1.5, more than one family
- (4) best LOD ≥ 3, more than one family
- (5) best LOD ≥ 5, more than one family

Clinical Importance: Disease severity

(0) = Benign or protective

(1) = Low expectation of having symptoms for this genotype / low penetrance (eg. increased risk of Crohn's disease from high frequency variant causing only a ~.2% elevated risk of disease even with OR=6)

(2) = Mild effect on quality of life or unlikely to be symptomatic (eg. Cystinuria)

(3) = Moderate effect on quality of life (eg. Familial Mediterranean fever)

(4) = Expected to cause disability and/or reduced life expectancy (eg. sickle cell disease)

(5) = Severe effect causing early lethality or severe disability (eg. Familial adenomatous polypopsis, adrenoleukodystrophy)

Clinical Importance: Treatability

(0) = No clinical evidence supporting intervention (ie, will not trigger any medical advice or discussion of reproductive options, surveillance, non-invasive or invasive action)

(1) = Uncurable;, intervention only to alleviate symptoms

(2) = Potentially treatable; intervention is not standardized or is in development or is controversial

(3) = Medical advice, surveillance and/or behaviorial risk reduction recommended; treatment is available if needed but not routinely recommended.

(4) = Treatable; treatment recommended and reduces the amount of mortality and/or morbidity but does not eliminate it (eg. sickle cell disease)

(5) = Treatable; limited or ongoing treatment routinely eliminates or markedly decreases the effect of the disease (eg. PKU)

MYBPC3 E334K

(MYBPC3 Glu334Lys)

Short summary

This variant is associated with hypertrophic cardiomyopathy in four Japanese and one Italian patient.

Variant quality

Computational	3	(This column soon to collect/display justification for ratings)
Molecular and cellular	1	
Clinical population	5	
Clinical family	-	
Clinical outcomes	3	

Impact

putative pathogenic

Inheritance pattern

dominant

Summary of published research, and additional commentary

Total cases/controls	case+	case–	control+	control–	odds ratio
Familial Hypertrophic Cardiomyopathy	5	269	0	300	**5.595**

Search

"GENE" or "GENE A123C":

search

Log in

Google login

Yahoo login

OpenID URL:

https://www.google.com/

Log in

Source Data

Genome	Publication	SNV's	Obtained
JCV	Levy, et al	3213401	
JW	Wheeler, et al	3322093	2060544
NA07022		3077756	
PGP1		2966187	
YH	Wang, et al	3074097	
AK	Kim, et al	3453653	
SJK	Ahn, et al	3439107	
NA18507(Bentley)	Bentley et al	3612498	
NA18507 (McKernan)	McKernan, et al	3866085	
P0	Pushkarev, et al	2805471	
Ng et al (exomesX8)	Ng, et al	Avg 48500 each	

HLG JR (Illumina)

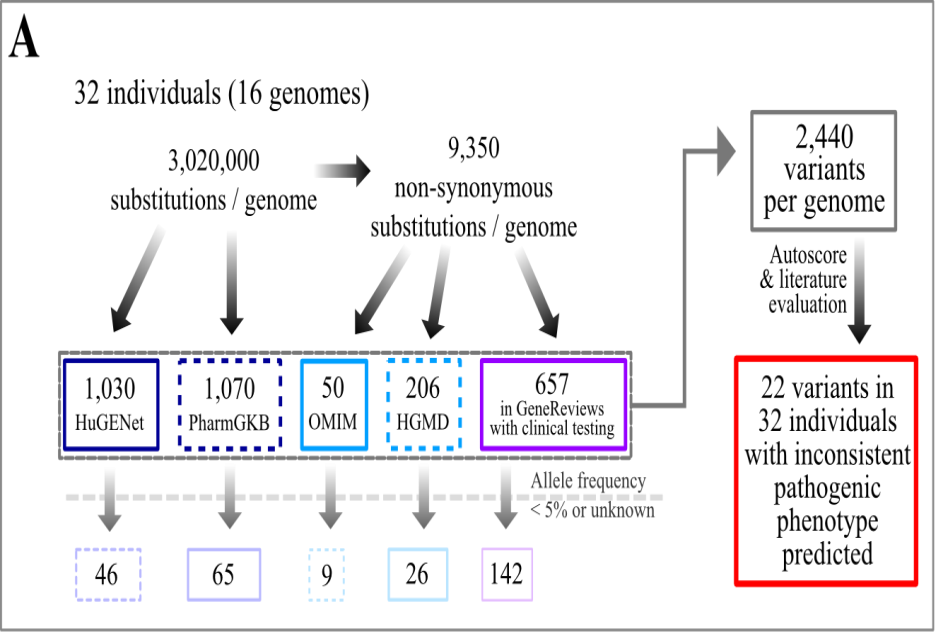
Variation type	Count	Novel proportion	FP rate/Mbp
All SNV's	3,076,869	10%	2.3-6.1%
Homozygous	1,097,899	2%	
Heterozygous	1,800,286	15%	
Transitions	2,858,818	------	
Transversions	1,316,837	------	
Coding	18,723	9%	
Nonsynonymous	9,286	11%	
Short insertions	168,909	37%	2.3-3.9
Short deletions	168,726	37%	1.8-3.0
Coding indels	528	56%	----------
Frameshifting short indels	310	62%	----------

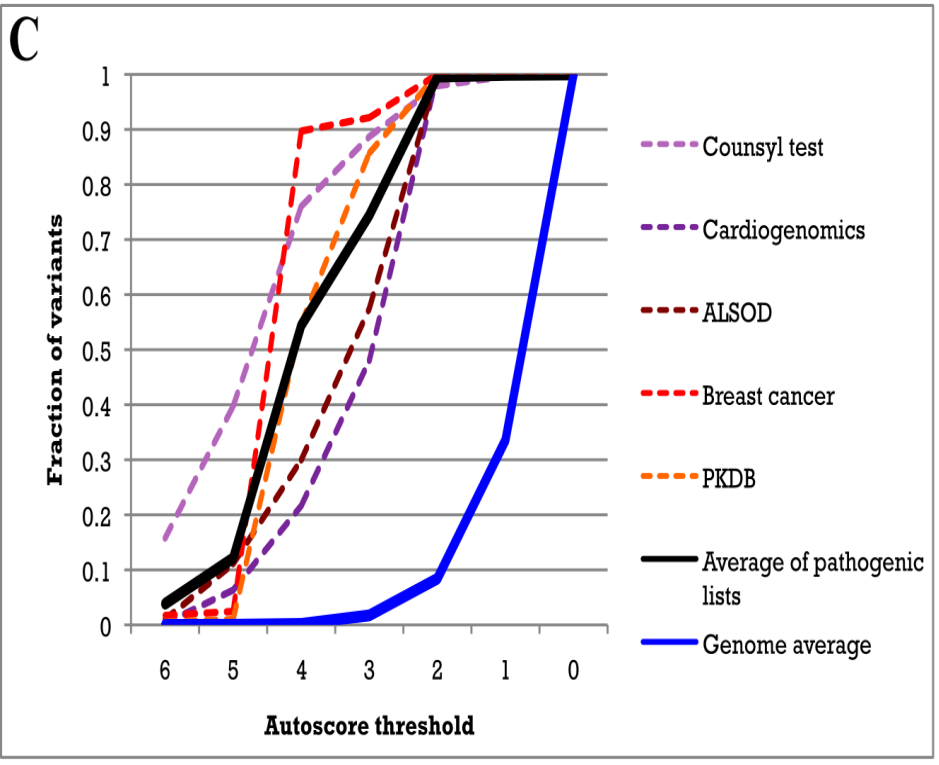
GET Output

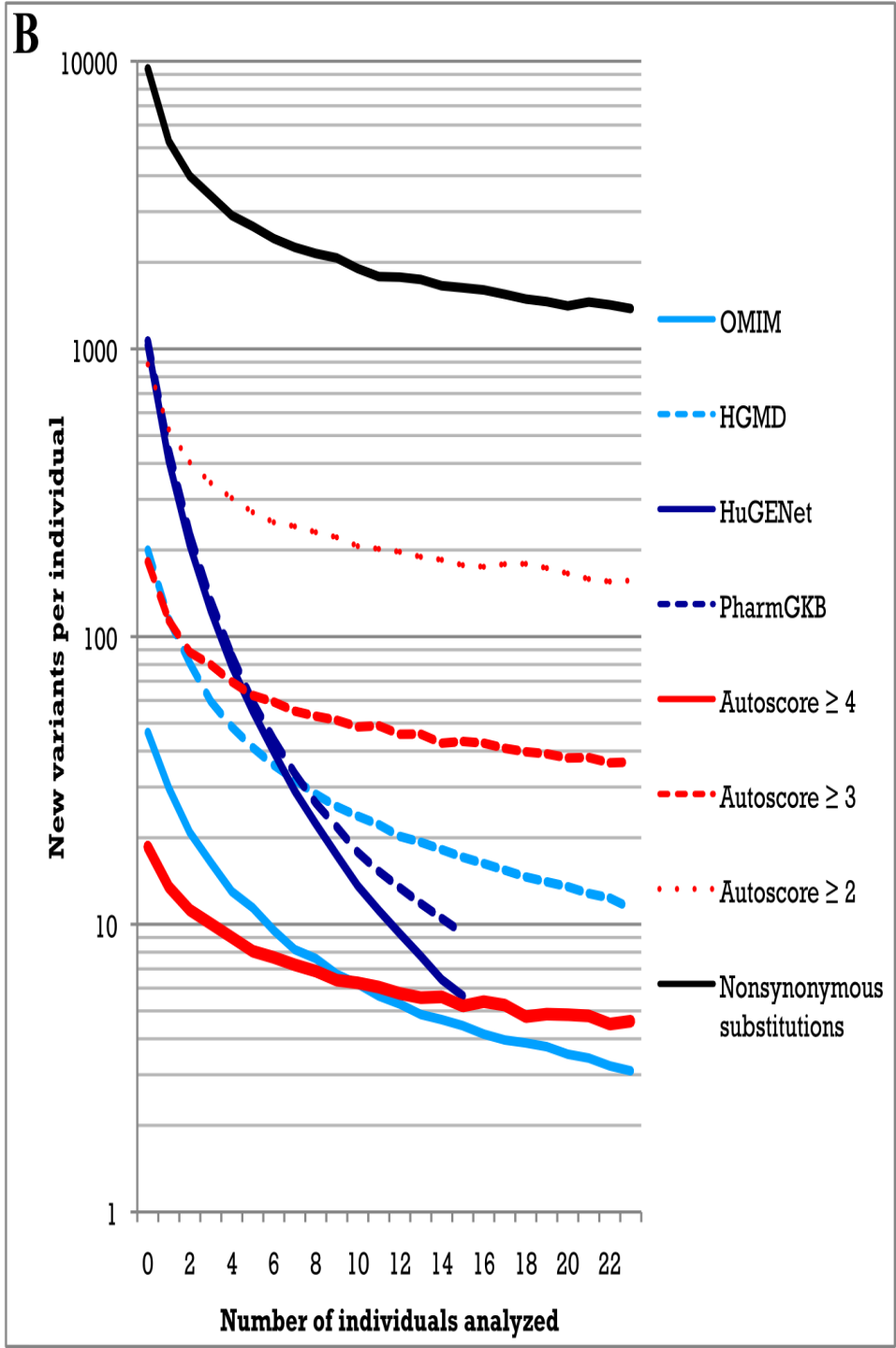
Genome	ns	HGMD	OMIM
JCV	8895	265	60
JW	6831	243	63
NA07022	9310	314	75
PGP1	8217	260	64
YH	9060	267	55
AK	10119	257	62
SJK	9714	268	68
NA18507(Bentley)	9995	259	62
NA18507 (McKernan)	10360	245	61
P0	9200	280	69
Ng et al (exomesX8)	Avg 8500	Avg 250 each	Avg 60

Freq filter

- TAF of 0.05 or less or unknown freq
- Avg HGMD hits down from over 250 to 25-50
- Avg OMIM hits down from over 60 to ~10
- Filter out variants: low risk conferring, in vitro study, few case reports, conflicting/nonreplicable lit hits
- Manual curation: 0-2 medically significant hits per person
- MYL2 A13T CMP variant: after bioinformatics, “manual” assessment is no different from clinic. Community efforts help.
- Ti/Tv, hom/het, total nt’s/#’s in mutation classes, etc.

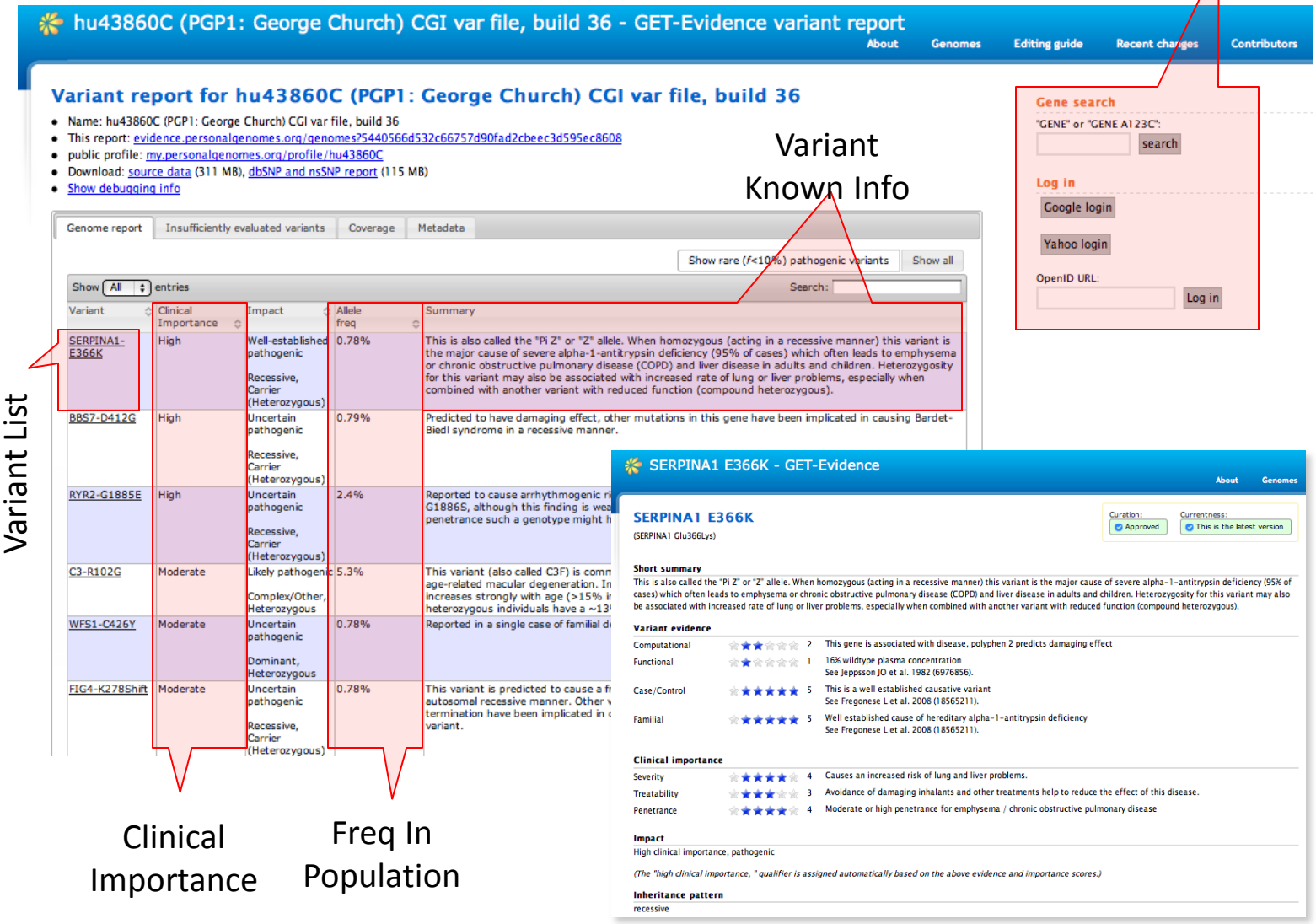

A
32 individuals (16 genomes)
3,020,000 substitutions / genome
9,350 non-synonymous substitutions / genome
2,440 variants per genome
Autoscore & literature evaluation
22 variants in 32 individuals with inconsistent pathogenic phenotype predicted
1,030 HuGENet
1,070 PharmGKB
50 OMIM
206 HGMD
657 in GeneReviews with clinical testing
Allele frequency < 5% or unknown
46
65
9
26
142


C
Fraction of variants
Autoscore threshold
Counsyl test
Cardiogenomics
ALSOD
Breast cancer
PKDB
Average of pathogenic lists
Genome average

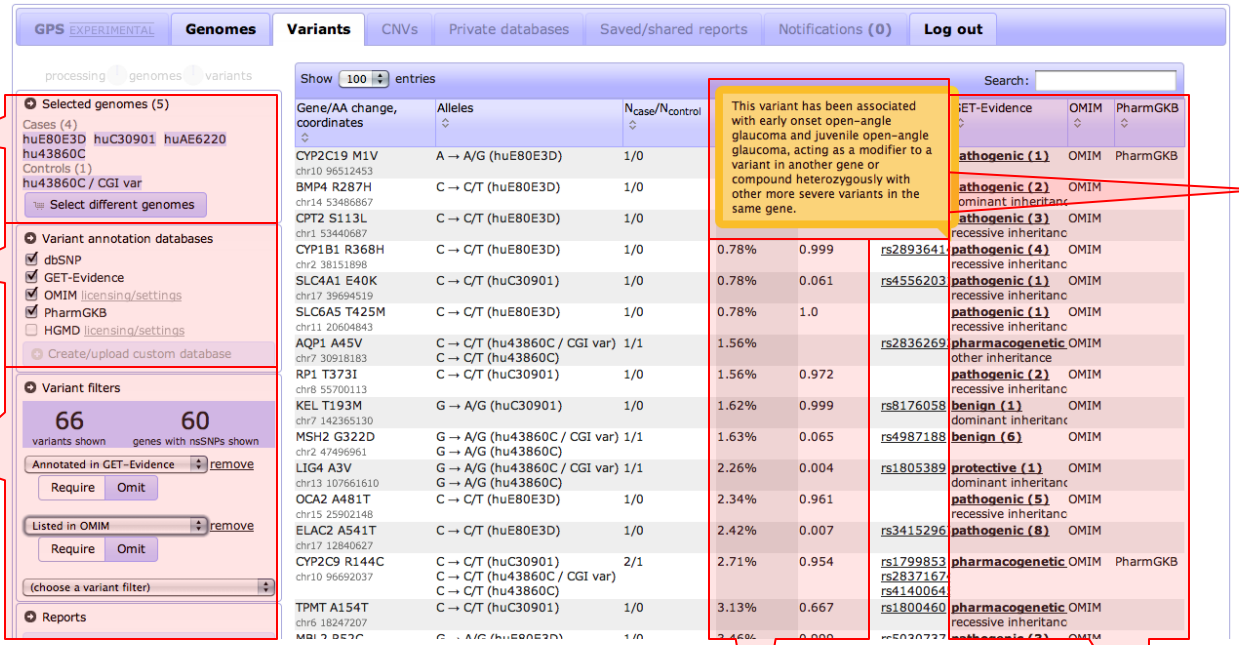

B
New variants per individual
Number of individuals analyzed
OMIM
HGMD
HuGENet
PharmGKB
Autoscore ≥ 4
Autoscore ≥ 3
Autoscore ≥ 2
Nonsynonymous substitutions

GET Evidence

Community Curation



Quickly Filter Variants Across Multiple Genomes & DBs



Multiple Genomes

Multiple Databases

Filters Down To Manageable Number of Variants

Pop-Up Summary Data

Sort By Frequency (Rare/Common)

Link to Variant DBs

Low frequency variants causing AA change in genomes from the general population

Genomes	Unfiltered	Testable	Testable, 5% MAF threshold	Testable, 1% MAF threshold	Testable, unknown frequency	Testable, 5% +unk	Testable, 1% + unk
62	703046	57063	9124	5624	2878	6246	2756
1	11339.45	920.37	98.77	90.71	46.42	100.74	44.29

Clinical Utility of Whole Genome Sequencing in a Healthy Adult

- CC/HPI: 53yo generally healthy, asymptomatic male with *JAK2* V617F mutation discovered on wgs through PGP genomic research study
- PMHx:
- Ocular migraines: ~1x/mo
- Pneumonia X1 at age 38yo
- Seasonal allergies
- History X 1 of sudden flank pain in 2010 while playing tennis. CT abd showed ?adrenal hematoma
- Medications: none
- Allergies: NKDA
- SHx: plays tennis regularly, +Tob Hx X 12pk yrs, quit age 30
- FHx: non relevant

Clinical Utility of Whole Genome Sequencing in a Healthy Adult (cont.)

- Physical exam: normal; pertinent negatives = no splenomegaly on exam (details in manuscript)
- Laboratory data:
- Historically, **plts** in the 501K/uL-779K/uL range on 4 different occasions spanning from 2/10/06-present
- I ordered the most recent CBC on 2/7/12 when I saw him in Medical Genetics Outpatient Clinic at MGH. **PLTs = 723K/uL**
- At the time I also ordered **d-dimer** (helping rule out active clot/inflammatory state), **PT/PTT** (ruling out coagulopathy) and **wbc w/ diff** (helping rule out in infection) = **all were wnl** {Values in manuscript}

Clinical Utility of Whole Genome Sequencing in a Healthy Adult (cont.)

- Careful history taking reveals no obvious cause for his thrombocytosis during the time of these 4 episodes.
- He meets WHO and PVSG criteria for essential thrombocytosis
- Essential thrombocytosis or thrombocythemia or ET:
- Definition: plt count >450K-500K; see diagnostic criteria

JAK2 V617F

- Clear assoc w/PV (90%/hom 30%), PMF (50%/ hom 60%), and ET (50%/ hom rare); <0.001% - 0.02% freq in general population; Danish data n= >10K; **why? HFE assoc HC**
- **tyrosine kinase** essential for the **function of the erythropoietin and thrombopoietin receptors** (but not the granulocyte colony-stimulating factor receptor)
- PV, PMF, and ET survival is usually measured in decades, and transformation to acute leukemia is uncommon in the absence of exposure to mutagenic agents

JAK2 V617F

- - Most patients with **PV** can live long lives without functional impairment when their red cell mass is effectively managed with phlebotomy alone
- **PMF:**characteristic features of extramedullary hematopoiesis: teardrop-shaped red cells, nucleated red cells, myelocytes, and promyelocytes
- 10% of patients spontaneously transform to an aggressive form of acute leukemia for which therapy is usually ineffective
- **ET:** can evolve into PV or PMF after a period of many years
- Survival of patients with ET is not different than for the general population/ no tx w/o sxs
- Plan CLIA seq conf (BWH); CBC/O2 sat/splenomeg/smear/LAP/B12 level +/- heme f/u

ET

- In this instance and, typically, more data available on diagnosed disease than sequence data (currently anyway)

ET

n	thrombosis	AML	myelofibrosis	
435 X 15yrs	17%	2%	4%	
891 X 15yrs	22%	2%	9%	

ET

- High risk:
- >60yo
- Prev hx of thrombosis

- Low risk:
- <60yo
- No hx of thrombosis
- Plt ct <1mil/uL

ET Diagnostic criteria

- Need to rule out reactive thrombocytosis:
- Fe def
- Acute blood loss
- Vit B12
- EtOH rebound
- Ca
- IBD
- Celiac
- Rheum
- MI
- Infxn
- Pancreatitis
- RF
- meds

WHO ET Diagnostic criteria

- Sustained PLT ct >450K X10^9/L; “during work up period”
- BM bx showing only megakaryocyte lineage
- JAK2 V617F
- Not meeting WHO criteria for PV, PMF, CML, MDS
- Tx: baby asa daily; thrombocytosis>bleeding; heme f/u

Clinical Utility of Whole Genome Sequencing in a Healthy Adult (cont.)

freq	nsSNPs (unknown MAF included)	nsSNPs (unknown MAF excluded)		
10%	1,839	1254		
1%	876	291		
0.1%	654	69		
0.01%	605 hits NOT containing JAK2 mut	20		

Clinical Utility of Whole Genome Sequencing in a Healthy Adult (cont.)

Freq = 0.1%	Unknown MAF included	Unknown MAF excluded		
dbSNP	393	17		
Get-ev	547 (includes affected genome)	67 (listed); 1 (annotated)		
Testable genes	38	9		
OMIM	1	1		

Clinical Utility of Whole Genome Sequencing in a Healthy Adult (cont.)

OMIM/Freq = 0.1%	OMIM/Freq= 1% (N.B.: negligibly small number of variants with unknown MAF currently in OMIM)	OMIM/Freq= 10%		
1	2	13		

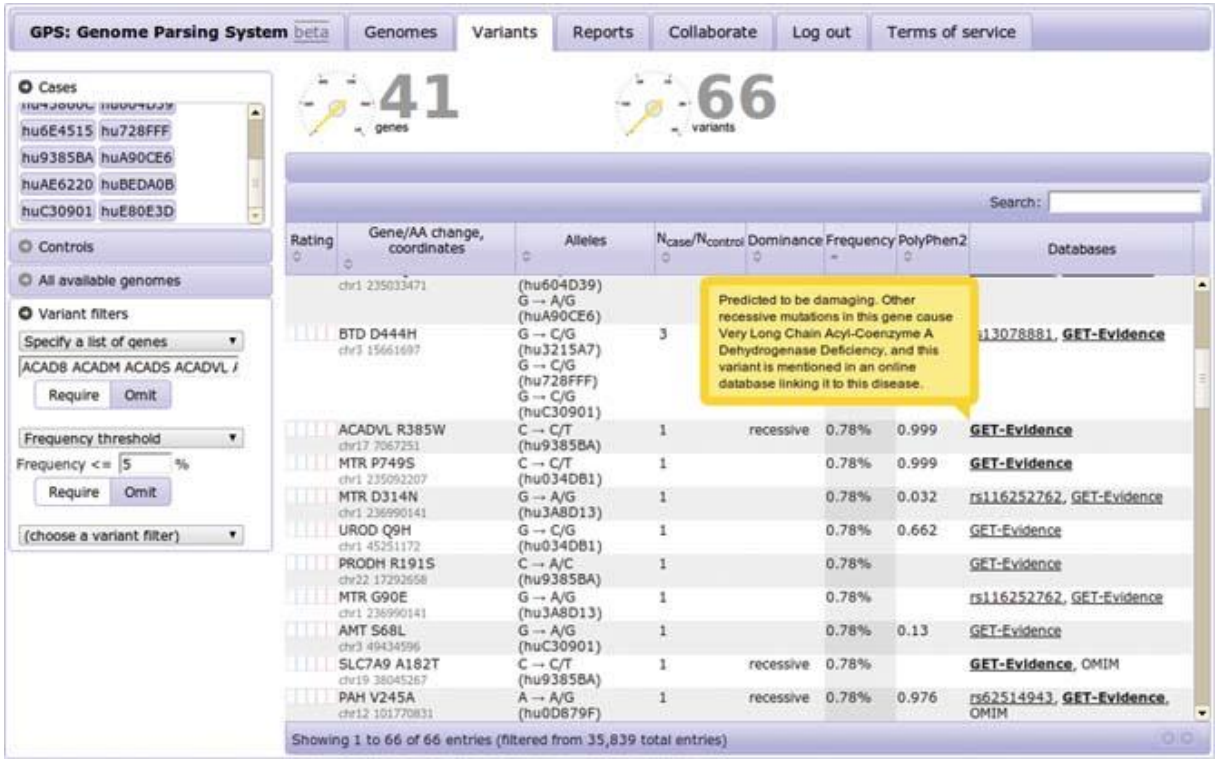
Clinical Utility of Whole Genome Sequencing in a Healthy Adult (cont.)

OMIM/MAF = 0.1%	OMIM/MAF= < 1%	OMIM/MAF = < 10%		
N=117	N=117	N=117		
49	150	255		
pharmGKB/0.1 % = 7 (no genes)				
HugeNET= n/a				

Future work

- Individualize genome analyses based on reported phenotype
- Try different algorithms (eg, inc freq/stricter gene thresholds)
- Manual curation of the literature
- Contact clinical labs

Back to the future: From genome to metabolome



Human Mutation
Volume 33, Issue 5, pages 809-812, 13 APR 2012 DOI: 10.1002/humu.22073
http://onlinelibrary.wiley.com/doi/10.1002/humu.22073/full#fig1

Avg # of potentially actionable metabolome hits per individual in the general population

- 88 metabolic genes 16 pgp people
- Clia biochem testing
- Avg 4-9 variants per person at less than 5% freq
- For 117 people:

<5%

with unknowns: 303 or avg 2.59/person
without unknowns: 196 or avg 1.68/person

<10%
with unknowns: 315 or 2.69/person
without unknowns: 208 or 1.78/person

Table 2. Planned Biochemical Phenotyping for 200+ PGP Participants with Whole-Genome Data
Plasma amino acids
Urine organic acids
Plasma acylcarnitines
Urine acylglycines
Sodium
Potassium
Chloride
Bicarbonate
Blood urea nitrogen
Creatinine
Glucose
NH_4 level
Carnitine profile (free and total)
Folate level
Zinc level
B12 level
Urine-reducing substances
Lipid profile
Hemoglobin electrophoresis
Pyridoxine level
Biotin level
Urine galactitol
Galactose-1-phosphate
Copper level
Ceruloplasmin
Magnesium level
Carbohydrate-deficient transferrin
Urine and plasma porphobilinogen
Urine and plasma delta-aminolevulinic acid
RBC plasmalogens
Pipecolic acid
Plasma very-long-chain fatty acids

Genomic Architecture

- Sequence architecture considerations:
- Sequence **gaps**:
- There are still gaps in “full” genomes largely due to repetitive sequences
- Fraction of chromosome sequenced ranges from 44.4% (Y-chr; an outlier) to 98.8% (chr17)
- Avg is 92.8%

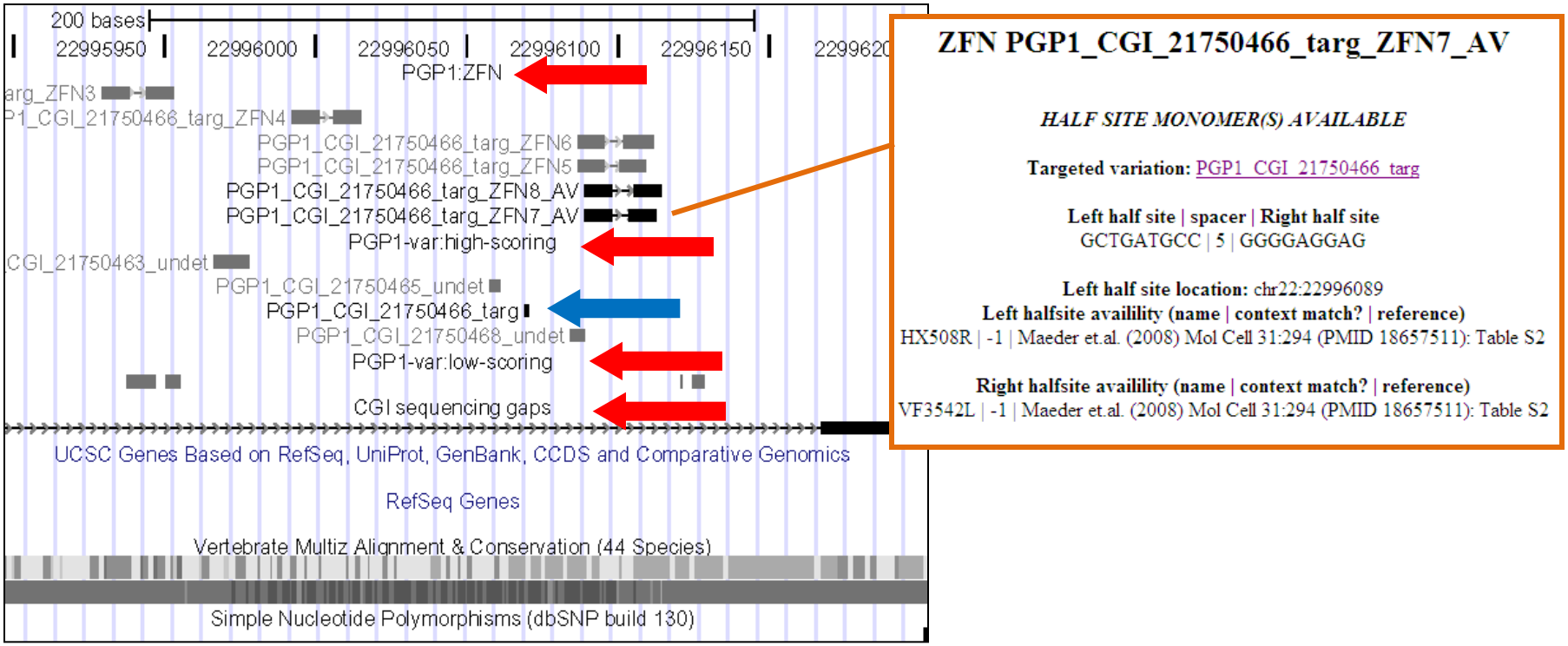
General Genomic Stats (cont.)

- **Gene density** is not uniform across chromosomes (it's positively associated with GC content)
- Highest gene density=chr19 (size=55.8Mb; 1457 genes; gene density=26.12 genes/sequenced Mb)
- Lowest gene density= Y chr (25.7Mb; 94 genes; 3.66 genes/sequenced Mb)

General Genomic Stats (cont.)

- **Pseudogenes**: >5000 annotated pseudogenes in NCBI release 36.2 of the human genome sequence (median size=1200nucleotides; ((per pseudogene??))
- **Exon number**: seems to correlate with protein size rather than gene size

UCSC Genome Browser custom annotation tracks for evaluation of PGP1 variations.



Screen shot of UCSC Genome Browser illustrating custom annotation tracks developed for evaluation of PGP1 variations. Shown are tracks for ZFN sites, high-scoring variations, low-scoring variations, and sequencing gaps (red arrows, see text). The high-scoring variation at chr22:22996070 described in the text is shown (blue arrow) along with its targeting ZFNs. Clicking on a feature calls up auxiliary information. Shown is the description called up for the ZFN targeted to chr22:22996070 described in the text.

Clinical Utility of Whole Genome Sequencing in a Healthy Adult

MP Ball[1], AW Zaranek[1,2], T Clegg[2], W Vandewege[2], PM Campeau[5], D Kuter[3], GM Church[1], JV Thakuria[1,2,4]
[1]Department of Genetics, Harvard Medical School, Boston, MA 02115
[2]Clinical Future, Inc., Cambridge, MA 02142
[3]Division of Hematology and Oncology, Massachusetts General Hospital, Boston, MA 02114
[4]Division of Medical Genetics and Metabolism, Massachusetts General Hospital, Boston, MA 02114
[5]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030

INTRODUCTION

Over the past two years, a growing body of literature has demonstrated clinical utility of exome and whole genome sequencing (WGS) in primarily pediatric patients carefully selected for high likelihood of strongly penetrant, rare, Mendelian disorder. As a result, exome sequencing has recently become clinically available in the United States. While clinical utility is currently a) expected in highly selected cases and b) likely to inform reproductive risk in most patients of childbearing age and their family members given the expected burden of carrier status for recessive disorders in all individuals, the degree of clinical utility of whole genome sequencing in healthy adults has not yet been firmly established.

CASE SUMMARY

Here, we describe clinical utility of whole genome sequencing in an asymptomatic, healthy, 53 year-old male leading to diagnosis of essential thrombocythemia (ET), as well as appropriate treatment, and prevention of ET complications. WGS performed through participation in The Personal Genome Project (PGP) identified a known pathogenic *JAK2* V617F mutation using the Genome-Environment-Trait-Evidence (GET-Evidence) automated system. Medical follow up revealed an elevated platelet count of 723,000/uL (reference range:150,000-400,000/uL) which remained elevated on repeat studies. Secondary causes of thrombocytosis were ruled out and the diagnosis of ET was subsequently made. Low dose ASA 81mg po daily was started to prevent thromboembolic complications and symptomatically treat intermittent episodes of scotomata that were described by the patient upon eliciting further medical history after WGS revealed heterozygosity for *JAK2* V617F.

PRESENTATION

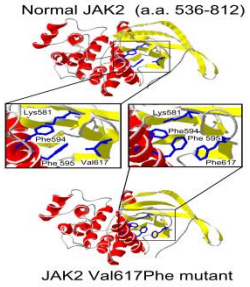
53yo M w/ PMHx of ocular migraines (~1x/mo), pneumonia (X1 at age 38yo), and seasonal allergies, presents for follow up of *JAK2* V617F mutation discovered through WGS as part of his participation in the Personal Genome Project (PGP) genomic research study. The patient is generally healthy and plays tennis regularly. He has a 12pack year history but quit smoking at age 30. He describes one incident of sudden flank pain in 2010 while playing tennis. Evaluation in the ED at that time included CT abd w/ contrast showing a questionable adrenal hematoma. ROS was otherwise negative. He was on no medications, has no known drug allergies, FHx was noncontributory and negative for coagulopathy, stroke, myelofibrosis, polycythemia vera, or thrombocytosis. Physical examination was normal and pertinent negatives included no findings of splenomegaly, petechiae, or bruising. History of intermittent scotomata bilaterally was revealed on detailed history taking.

LABORATORY DATA

WGS was performed on DNA from whole blood and library preparation and sequencing done by Complete Genomics, Inc. (CGI) through the PGP. Data was annotated using their 2.X pipeline (matching against the NCBI build 37 reference genome). Genome-Environment-Traits Evidence system (GET-Evidence) identified the pathogenic *JAK2* V617F mutation. Confirmatory Sanger confirmation was performed at Brigham and Womens Hospital. The remainder of his clinial laboratory testing was done at Massachusetts general Hospital. Review of his outpatient platelet values showed a range of 501K/uL-779K/uL on 4 different occasions spanning from 2/10/06-present. A repeat CBC revealed an elevated PLT ct of 723,000/uL. Studies to rule out secondary causes of thrombocytosis included d-dimer, PT/PTT, CBC w/ diff, all returned normal and he was found to meet WHO and PVSG criteria for a diagnosis of essential thrombocythemia (ET) with the exception of bone marrow biopsy which we deferred.

TREATMENT

The patient was started on low dose ASA 81mg po daily to decrease risk of serious thromboembolic events and he will continue regular follow up with Hematology and Medical Genetics for treatment and surveillance of his ET.



This 3D model of *JAK2* (amino acids 536-812) demonstrates that a change from a Valine to a Phenylalanine at position 617 induces positional changes in residues 595 and 594, leading ultimately in a displacement of the Lysine at position 581, which is critical for the kinase activity of the protein. These observations by Drs. Bandaranayake et al. help explain why a mutation far from the active site ultimately result in an activation of the kinase activity. The model was generated with Swiss-PDB viewer using PDB files 4FVP and 4FVR.

DISCUSSION

The *JAK2* V617F mutation results in decreased tyrosine kinase activity essential for the function of the erythropoietin and thrombopoietin receptors and confers a significantly increased risk for polycythemia vera (PV), primary myelofibrosis (PMF) and essential thrombocythemia (ET) against background. Population based data from 10,507 participants in the Copenhagen City Heart Study showed the prevalence of the mutation to be 0.2% and was seen in 18 participants. According to this data, risk of hematologic malignancy is 27% (seen in 4/15 participants) and risk of myeloproliferative disease is 13% (or 2/15) vs. the background risk of 1% and 0.2% respectively. (Nielsen, et al, 2010)

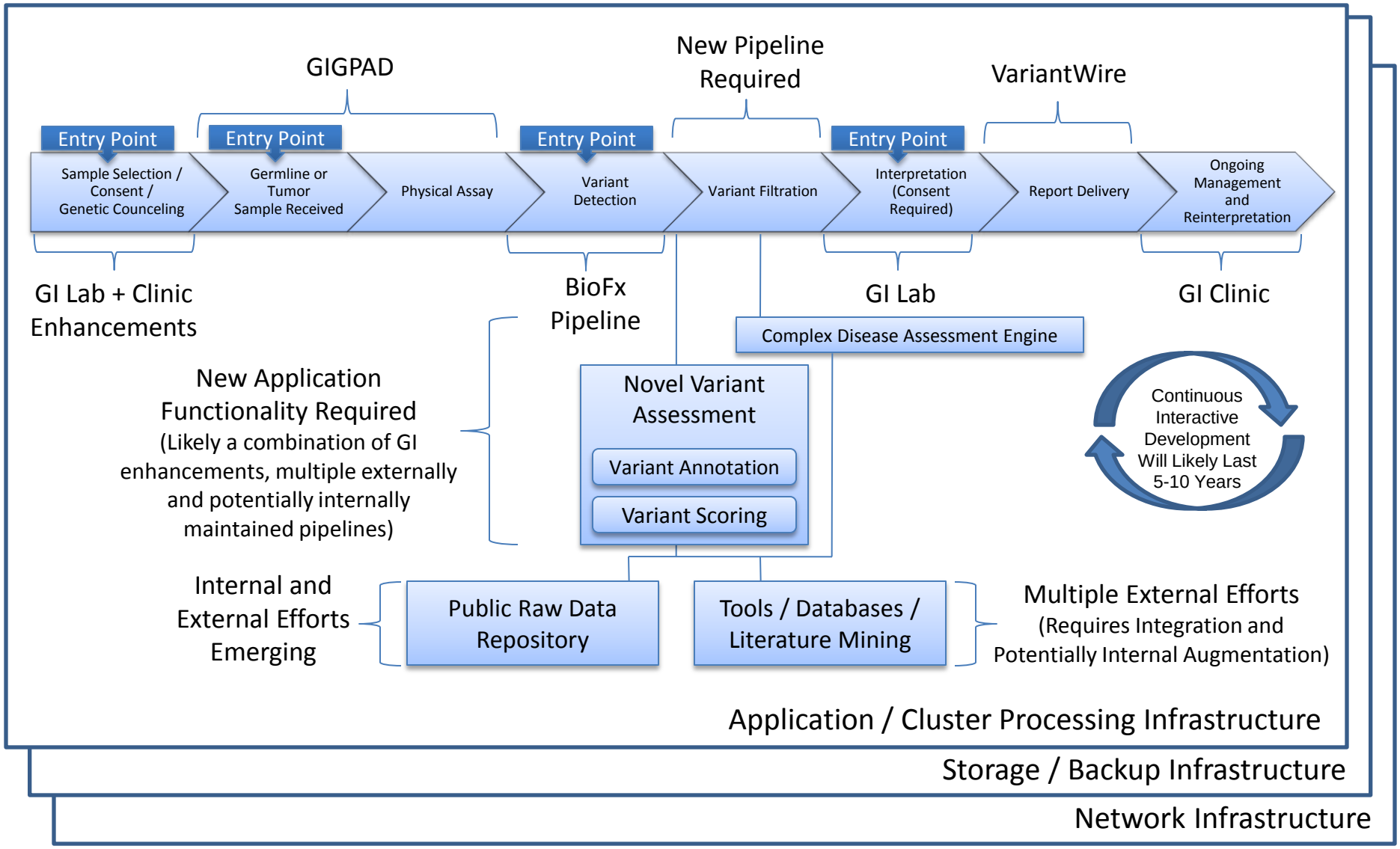
Based on data from Passamnoti, et al (2004) and Barbui, et al (2011), this patient's risk is significantly modified downward given his subsequent diagnosis of ET in medical follow up. (see below)

n	thrombosis	AML	myelofibrosis	
435 X 15yrs	17%	2%	4%	
891 X 15yrs	22%	2%	9%	

Assuming the *JAK2* V617F has characteristics shared by other medically actionable mutations incidentally found, using the Genome Parsing System (GPS), a secure, private web-service for genomic and phenotypic data management and variant filtration, we generated a variety of variant filtering algorithms that would correctly identify *JAK2* V617F and tested these against 100+ PGP genomes and a variety of disease-specific mutation databases to evaluate sensitivity and specificity. When excluding carrier status for serious recessive disorders, our initial screen checking for OMIM cited variants with a MAF <0.1% as many as 1 out of 15 cases in the general population have variants which should prompt careful literature review. Subsequent manual review has revealed many of these to be likely benign in our analysis of >100 PGP whole genomes. We are currently deploying other variant filtering algorithms to determine the extent of clinical utility in the general population when filtering out other rare pathogenic variants. Even with clinical utility seen in as low as 1 in 100 patients, such screening may become standard optional testing that is comparable in benefit with currently recommended medical screening in the general adult population, such as mammogram, colonoscopy, and cervical Pap smear – that can be invasive and need to be performed on a regular basis. The risk of cervical cancer for example being <1% in U.S. women.

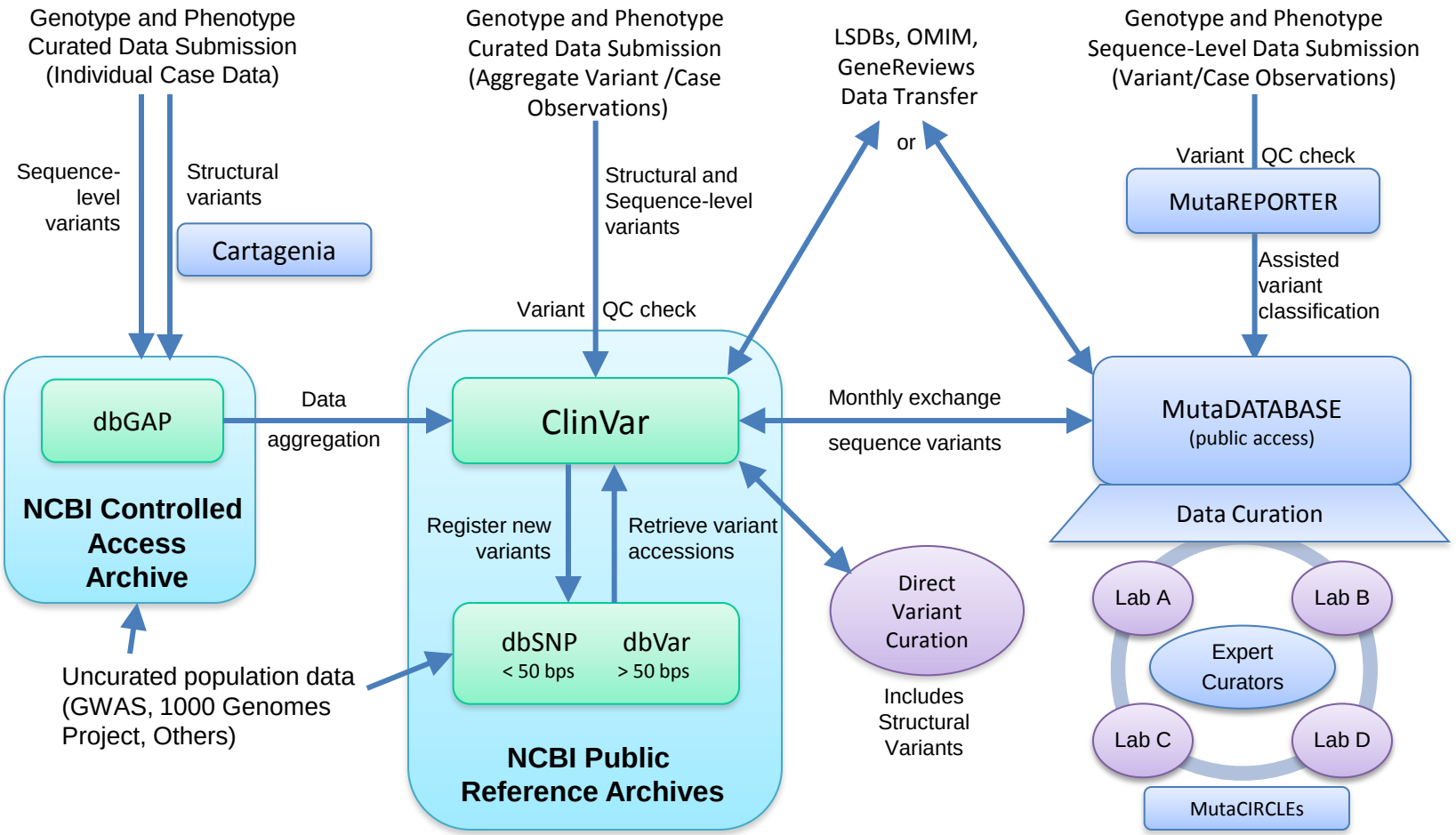
References
1. Bandaranayake RM, et al, Crystal structures of the JAK2 pseudokinase domain and the pathogenic mutant V617F. Nat Struct Mol Biol. 2012 Aug;19(8):754-9.
2. Nielsen C, et al, The JAK2 V617F somatic mutation, mortality and, cancer risk in the general population, *Hematologica* 2011;96(03):450-453.
3. Passamonti F, et al, Life expectancy and prognostic factors for survival in patients with polycythemia vera and essential thrombocythemia. *Am J Med*, 2004;117:755.
4. Barbui T, et al, Survival and disease progression in essential thrombocythemia are significantly influenced by accurate morphologic diagnosis: an international study. *J Clin Oncol* 2011; 29:3179.

Whole Genome Sequencing IT Requirements



Copyright Partners HealthCare 2011

Overview of Data Flows and Systems



Arvados:

- an open source platform for managing genomic data. (includes all data generated by the PGP)
- Arvados can sit on Amazon Web Services or OpenStack.

- Learn more:
  http://arvados.org
  - download the source code, and get involved with the community

Mass. Gen. Hosp. M3D

- Animal studies: zebrafish, yeast, plant; murine models through collaboration with JAX lab and local investigators
- Transcriptomes
- Fxnl fibro cx's
- Immunomes
- Disease modifiers/ new tx's
- UPDATES in 2014!


www.earthinpictures.com

Thank you

- Dr. Ruth Liberfarb
- SIP
- Stickler syndrome community

Thank you