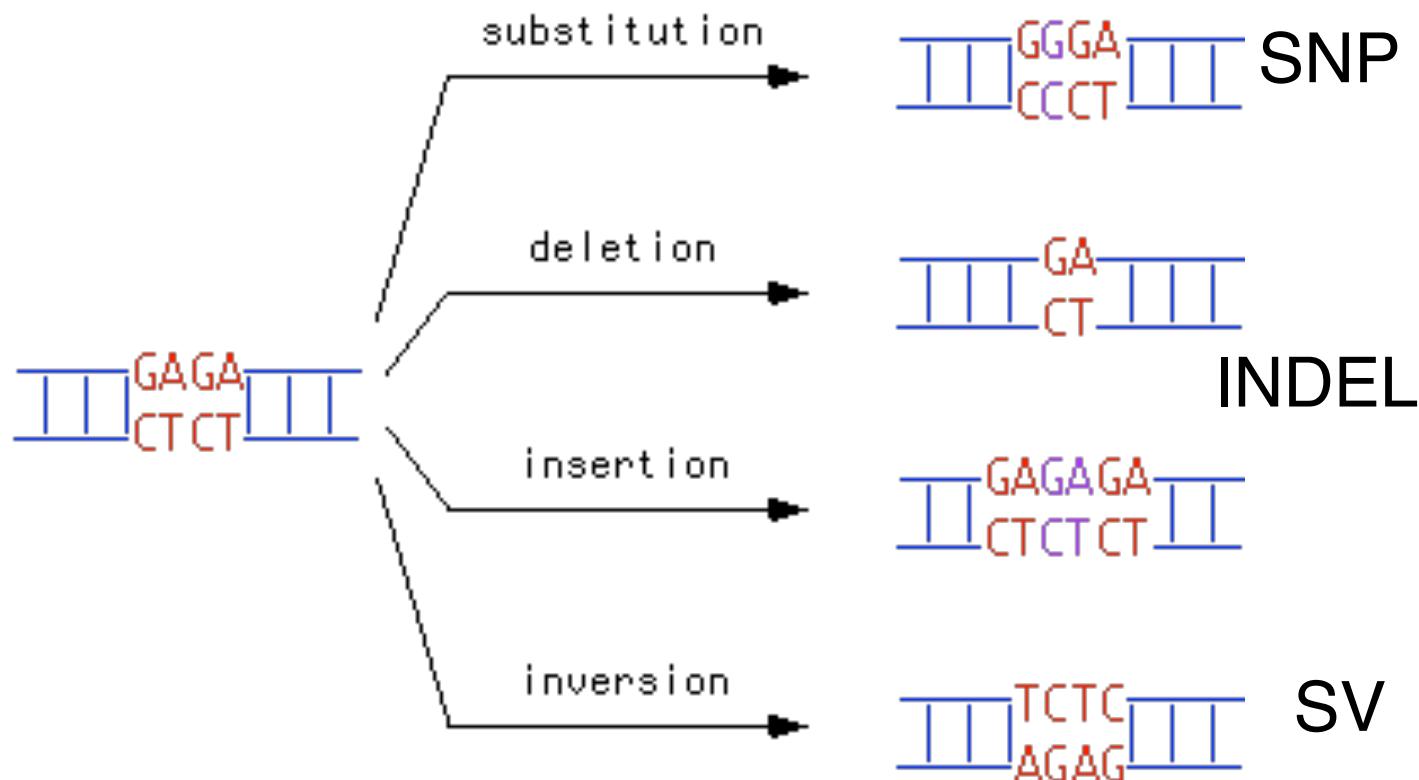


Identification of Disease Causing Candidate Mutations

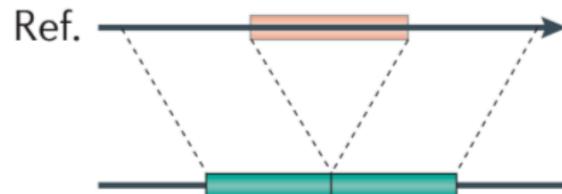
Cancer and Mendelian
Disease

Types of Variation

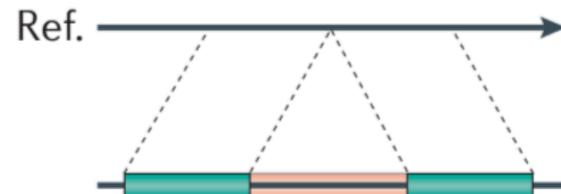


Types of Structural Variation

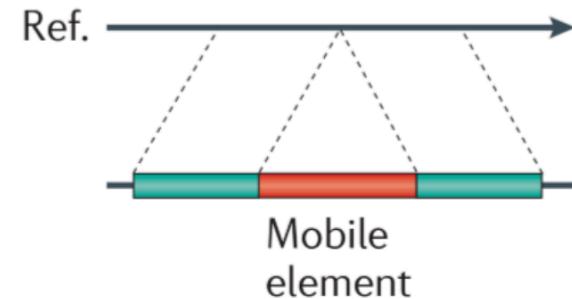
Deletion



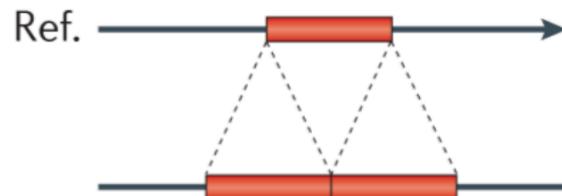
Novel sequence insertion



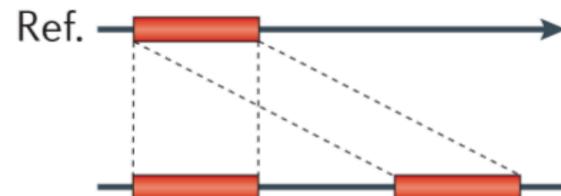
Mobile-element insertion



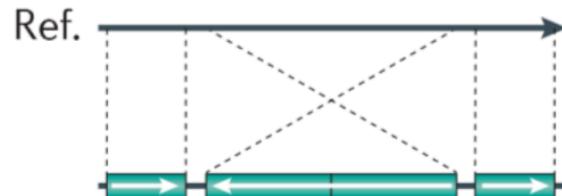
Tandem duplication



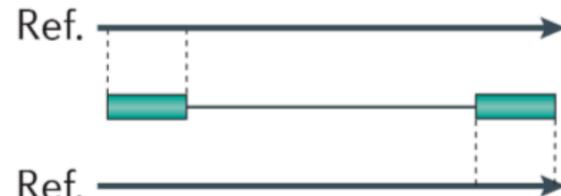
Interspersed duplication



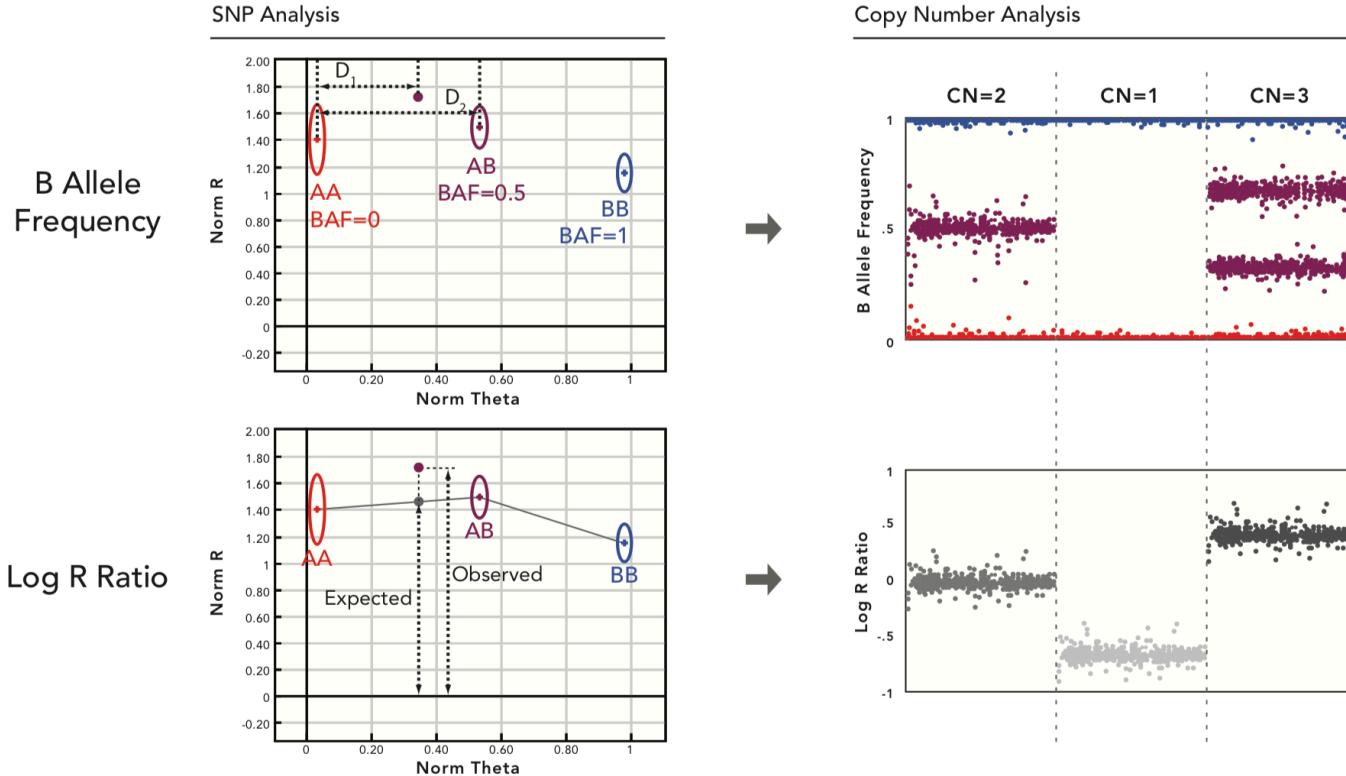
Inversion



Translocation



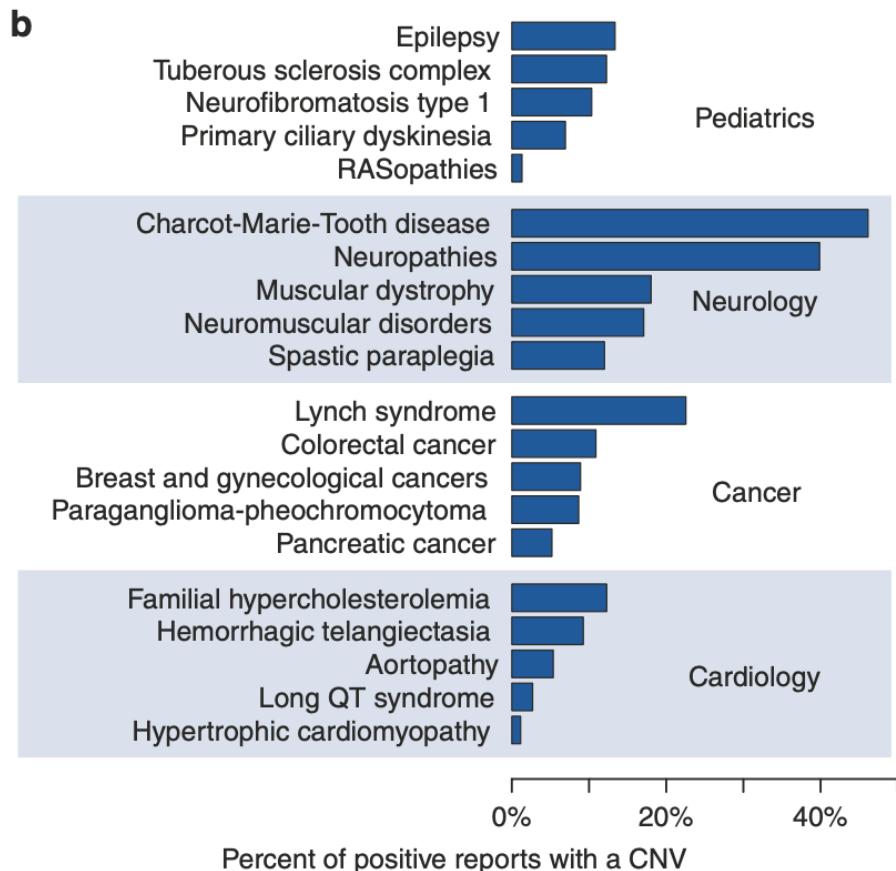
CNV Detection



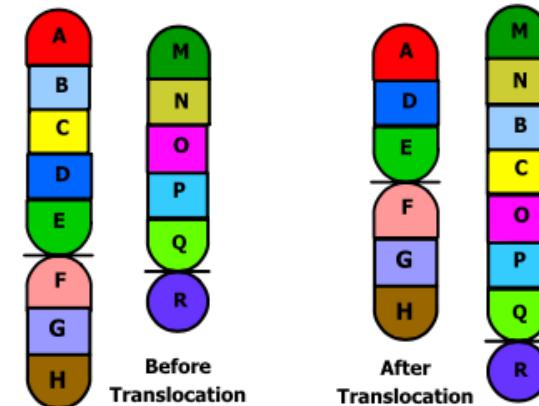
- Microarray
- Next Generation Sequencing

CNV

- There are a number of inherited conditions that are caused by copy number changes (Down Syndrome 3 copies of Ch21)
- In Cancer, Loss of Tumor Suppressors or Amplification of Oncogenic genes can lead to tumor formation (e.g. Her2+)



Gene Fusions Translocations



- It's quite rare that inherited diseases are caused by translocation events
 - Robertsonian — type of translocation that is thought to cause some form of down syndrome
 - Balanced
- The discovery of BRC-ABL gene fusions as a driver mutation of Chronic Myeloid Leukemia (CML) led to the development of imatinib (selective inhibitor), which improves the overall survival rates of CML patients to 90% over 5 years and 88% over 8 years
- NTRK Fusions lead to transcription of chimeric proteins with constitutively activated or overexpressed kinase function conferring oncogenic potential. Vitrakvi, an NTRK inhibitor, has been approved for treatment in various tumor types, including soft tissue sarcoma, salivary gland, infantile fibrosarcoma, thyroid, lung, melanoma, colon, GIST, cholangiocarcinoma, appendix, breast and pancreas

Key Analysis and Files

- SNPs/Indels for each sample
 - VCF file — SampleID.annot.vcf.gz
- Copy Number Variation
 - SampleID.cnv.calls.cns
- Gene Fusions
 - SampleID.genefusion.txt

VCF

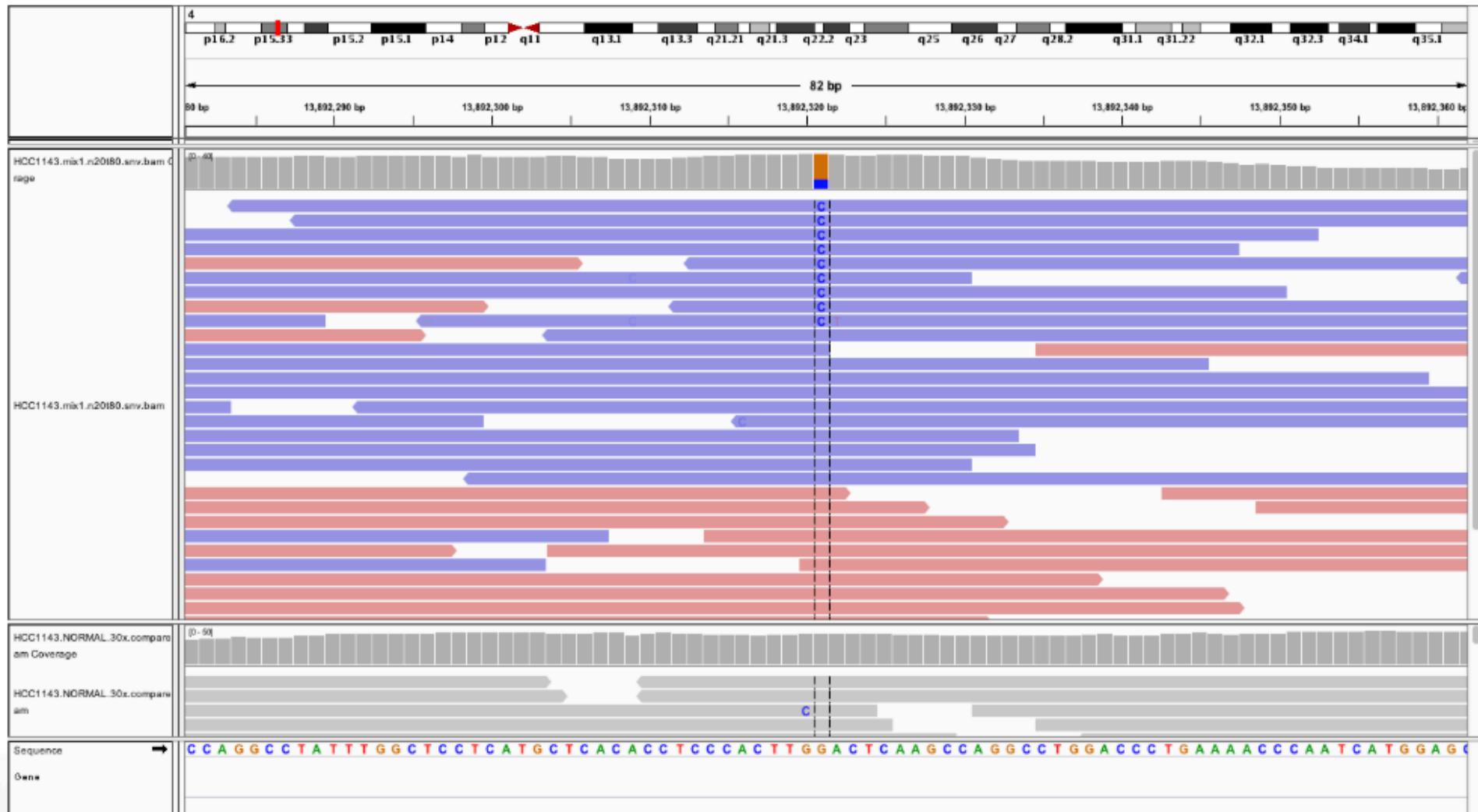
Every VCF file has three parts in the following order

- Meta-information lines (lines beginning with "##").
- One header line (line beginning with "#CHROM").
- Data lines contain marker and genotype data (one variant per line). A data line is called a VCF record.

Example VCF file

```
##fileformat=VCFv4.2
##FORMAT=<ID=GT,Number=1,Type=Integer,Description="Genotype">
##FORMAT=<ID=GP,Number=G,Type=Float,Description="Genotype Probabilities">
##FORMAT=<ID=PL,Number=G,Type=Float,Description="Phred-scaled Genotype Likelihoods">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT SAMP001 SAMP002
20 1291018 rs11449 G A . PASS . GT 0/0 0/1
20 2300608 rs84825 C T . PASS : GT:GP 0/1:.. 0/1:0.03,0.97,0
20 2301308 rs84823 T G . PASS : GT:PL ./... 1/1:10,5,0
```

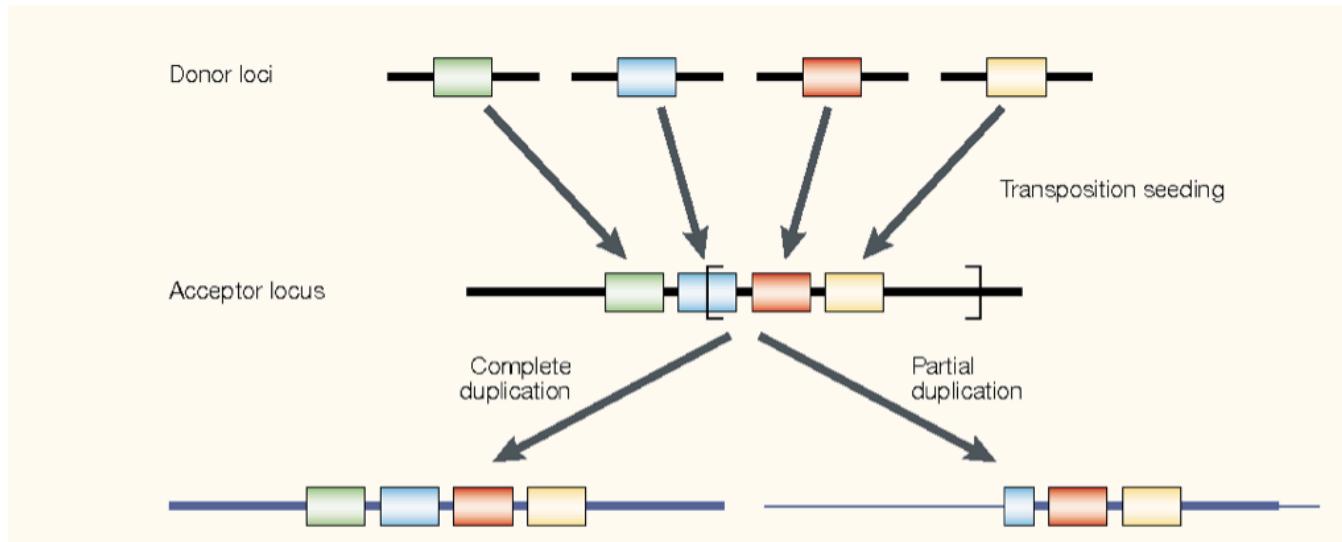
QC: StrandBias



QC:Simple Repeats



Segmental Duplication can be the cause of Alignment Errors in Alignments

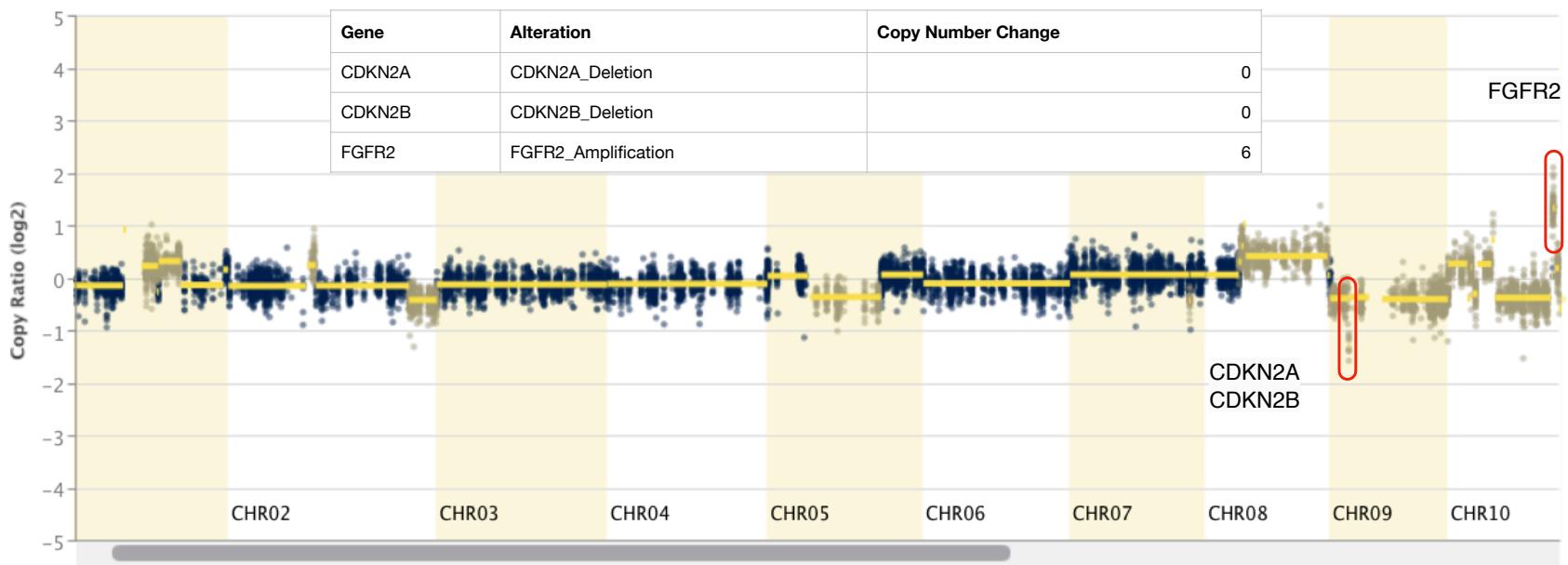
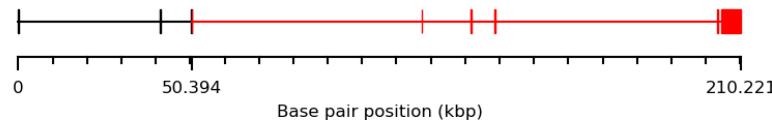


Typically these can be detected with low alignment quality scores
Mapping scores can be seen in IGV

CNVs and Gene Fusions

FusionName	Junction Read Count	Spanning Frag Count	Large Anchor Support	FFPM	Left Break Dinuc	Right Break Dinuc	CDS LEFT RANGE	CDS RIGHT RANGE	PROT FUSION TYPE
STRN--NTRK2	81	34	YES_LDAS	15.8797	GT	AG	1-412	1445-2517	INFRAME
STRN--NTRK2	17	28	YES_LDAS	6.2138	GT	AG	1-412	1397-2517	INFRAME

STRN-NTRK2
ENST00000263918-ENST00000277120



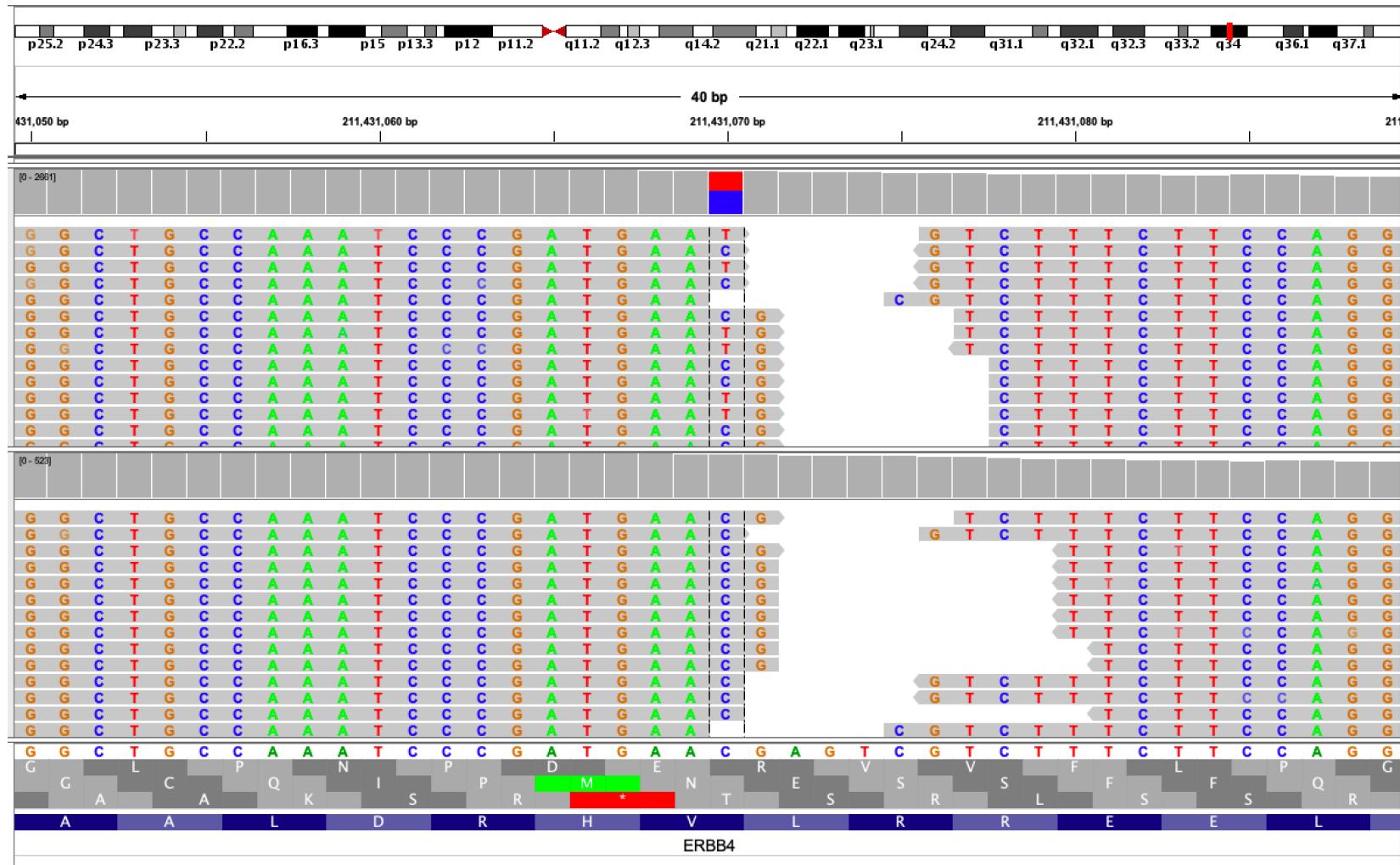
Variant Annotation

- What annotation does
 - Is the variant in a coding exon, and if so does it change an amino acid of the translated protein? If the variant changes the protein, is that likely to be deleterious? Is it in or near a splice site? Does it disturb gene regulation?
 - Has the reference allele been conserved in evolution?
 - Has the variant been seen before? What is its frequency in a population of interest?
 - Does the variant appear in a disease database?
- What annotation does not do
 - The various annotations enable you to prioritize variants for further investigation, but do not by themselves identify causality.

Gene Annotations and Effects

- snpEff
 - Changes affecting genes
 - Changes affecting regulatory regions
 - ENCODE
 - Epigenome Roadmap
 - NextProt
 - proteomic annotations
 - Motifs
- VEP
 - Changes affecting genes
 - Changes affecting regulatory regions
 - Integrated with downstream tools like cBioporal and GenVisR

Mutational Effects



AAC -> AAT

Val -> Ile

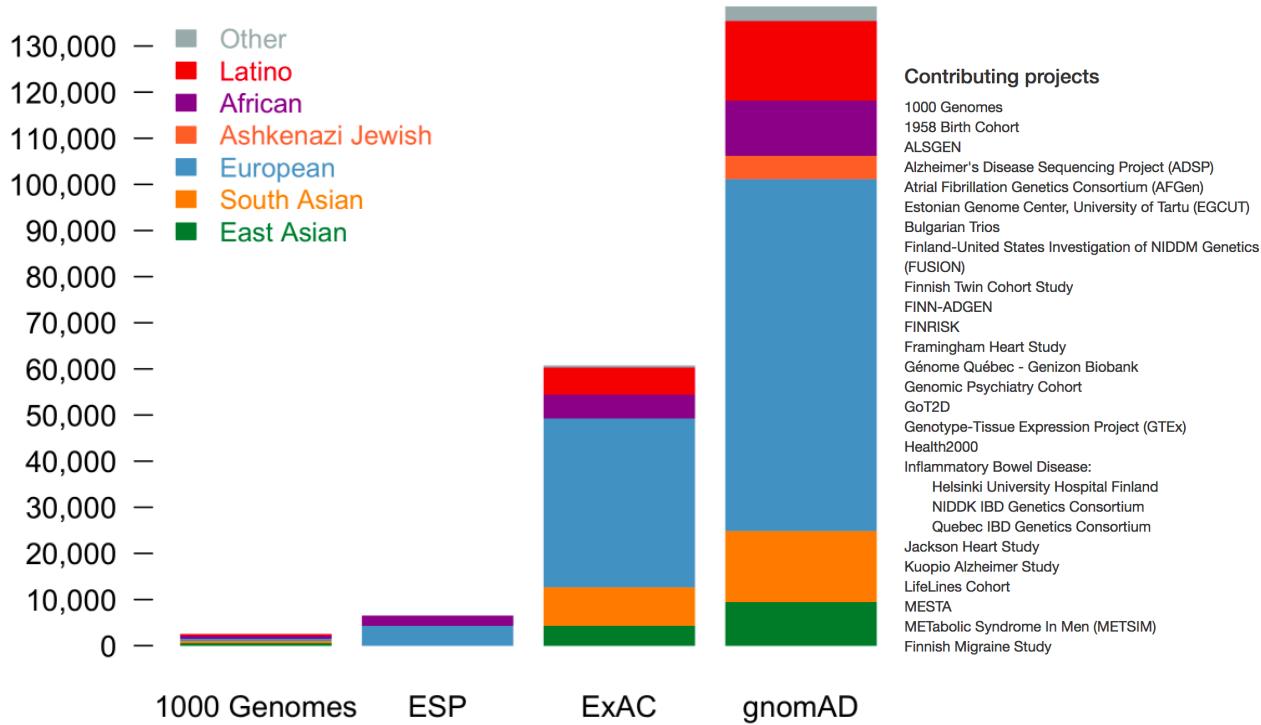
Effect Can be Unique to a Transcript

Allele	Effects	Impact	Gene	↑ Ensemble Gene Id	Feature Type	Feature Id	Transcript Biotype	Exon #	Coding Notation	Protein Notation	Protein Position	CD Pos
A	["missense_variant"]	MODERATE	PDE4DIP	ENSG00000178104	transcript	ENST00000369356.8	protein_coding	37/44	c.5965C>A	p.Gln1989Lys	1989/2362	6256
Other VCF Annotations (11 rows)												
Change	Allele	Effects	Impact	Gene	↑ Ensemble Gene Id	Feature Type	Feature Id	Transcript Biotype	Exon #	Coding Notation	Protein Notation	Protein Position
☒	A	["missense_variant"]	MODERATE	PDE4DIP	ENSG00000178104	transcript	ENST00000618462.4	protein_coding	39/46	c.5647C>A	p.Gln1883Lys	1883/2240
☒	A	["missense_variant"]	MODERATE	PDE4DIP	ENSG00000178104	transcript	ENST00000369354.7	protein_coding	37/44	c.5965C>A	p.Gln1989Lys	1989/2346
☒	A	["missense_variant"]	MODERATE	PDE4DIP	ENSG00000178104	transcript	ENST00000585156.5	protein_coding	40/47	c.6373C>A	p.Gln2125Lys	2125/2482
☒	A	["missense_variant"]	MODERATE	PDE4DIP	ENSG00000178104	transcript	ENST00000524974.5	protein_coding	39/46	c.6220C>A	p.Gln2074Lys	2074/2431
☒	A	["missense_variant"]	MODERATE	PDE4DIP	ENSG00000178104	transcript	ENST00000530130.2	protein_coding	4/7	c.433C>A	p.Gln145Lys	145/307
☒	A	["intron_variant"]	MODIFIER	PDE4DIP	ENSG00000178104	transcript	ENST00000530062.5	nonsense-mediated_decay	1/5	c.90-5682C>A		
☒	A	["intron_variant"]	MODIFIER	PDE4DIP	ENSG00000178104	transcript	ENST00000526664.5	processed_transcript	4/4	n.502+697C>A		
☒	A	["non_coding_transcript_exon_variant"]	MODIFIER	PDE4DIP	ENSG00000178104	transcript	ENST00000527901.1	retained_intron	3/3	n.523C>A		
☒	A	["non_coding_transcript_exon_variant"]	MODIFIER	PDE4DIP	ENSG00000178104	transcript	ENST00000479369.6	retained_intron	7/7	n.1815C>A		
☒	A	["non_coding_transcript_exon_variant"]	MODIFIER	PDE4DIP	ENSG00000178104	transcript	ENST00000481227.6	retained_intron	4/5	n.574C>A		
☒	A	["intron_variant"]	MODIFIER	RP4-791M13.4	ENSG00000255148	transcript	ENST00000532137.1	antisense	2/2	n.493-2301G>T		

Population Allele Frequency

- 1000 Genomes
 - A project to sequence ~3K people from a diverse set of genetically isolated populations and known admixture populations
- ExAC (Exome Aggregation Consortium)
 - An aggregation of several large exome sequencing projects (~60K WES)
 - Obsolete ie not maintained integrated in GnomAD
- GnomAD (Genome Aggregation Database)
 - An aggregation of several large exome sequencing projects (~125K WES) and whole genome sequencing projects (15K WGS)

GnomAD



Contributing projects

1000 Genomes
1958 Birth Cohort
ALSGEN
Alzheimer's Disease Sequencing Project (ADSP)
Atrial Fibrillation Genetics Consortium (AFGen)
Estonian Genome Center, University of Tartu (EGCUT)
Bulgarian Trios
Finland-United States Investigation of NIDDM Genetics (FUSION)
Finnish Twin Cohort Study
FINN-ADGEN
FINRISK
Framingham Heart Study
Gérome Québec - Genizon Biobank
Genomic Psychiatry Cohort
GoT2D
Genotype-Tissue Expression Project (GTEx)
Health2000
Inflammatory Bowel Disease:
Helsinki University Hospital Finland
NIDDK IBD Genetics Consortium
Quebec IBD Genetics Consortium
Jackson Heart Study
Kuopio Alzheimer Study
LifeLines Cohort
MESTA
Metabolic Syndrome In Men (METSIM)
Finnish Migraine Study

Myocardial Infarction Genetics Consortium (MiGen):
Leicester Exome Seq
North German MI Study
Ottawa Genomics Heart Study
Pakistan Risk of Myocardial Infarction Study (PROMIS)
Precocious Coronary Artery Disease Study (PROCARDIS)
Registre Gironi del COR (REGICOR)
South German MI Study
Variation in Recovery: Role of Gender on Outcomes of Young AMI Patients (VIRGO)
National Institute of Mental Health (NIMH) Controls
NHBLI-GO Exome Sequencing Project (ESP)
NHBLI TOPMed
Schizophrenia Trios from Taiwan
Sequencing Initiative Suomi (SISU)
SIGMA-T2D
Swedish Schizophrenia & Bipolar Studies
T2D-GENES
GoDARTS
T2D-SEARCH
The Cancer Genome Atlas (TCGA)

Variant ID	Source	Consequence	Annotation	Flags	Allele Count	Allele Number	Allele Frequency	Number of Homozygotes
12-25362738-C-G	E	p.Met188Ile †	missense		1	248966	4.02e-6	0
12-25362738-TAC-T	E	p.Val186AsnfsTer25 †	frameshift	LC LoF	1	249450	4.01e-6	0
12-25362740-C-T	E	p.Val186Ile †	missense		3	249276	1.2e-5	0
12-25362742-C-T	E	p.Cys185Tyr †	missense		1	249326	4.01e-6	0
12-25362743-ACTTTGTA	E	p.Thr183_Lys184del †	inframe deletion		5	280748	1.78e-5	0
12-25362748-G-T	E	p.Thr183Lys †	missense		1	249262	4.01e-6	0
12-25362755-ACCT-A	E	p.Lys180del †	inframe deletion		5	249584	2e-5	0

Prioritization of Variants

- Variants knowns to cause the disease/condition
- Not common in the general population
- Loss/Gain of function mutations
 - Splice
 - Stop gain/loss
 - Start gain/loss
 - Coding frame shifts
- Non-synonymous Mutations
 - Amino Acid Changes
- Variants Likely to Change Expression
 - Transcription Factor Binding Sites
 - miRNA Targets

Variant Prioritization

- In Cancer (tumor/normal pairs)
 - somatic mutation identification
- In small family studies (trios)
 - rare germline variation
- In large populations (GWAS)
 - Covered this afternoon

Germline SNV/Indel Calling

(GATK, Samtools, etc)

- Looks for mismatches and gaps in the alignment to **the reference**
- Assumes **2 copies** of each allele
- In some cases uses a database of **known common SNPs** to support variant calls

Somatic SNV/Indel Calling

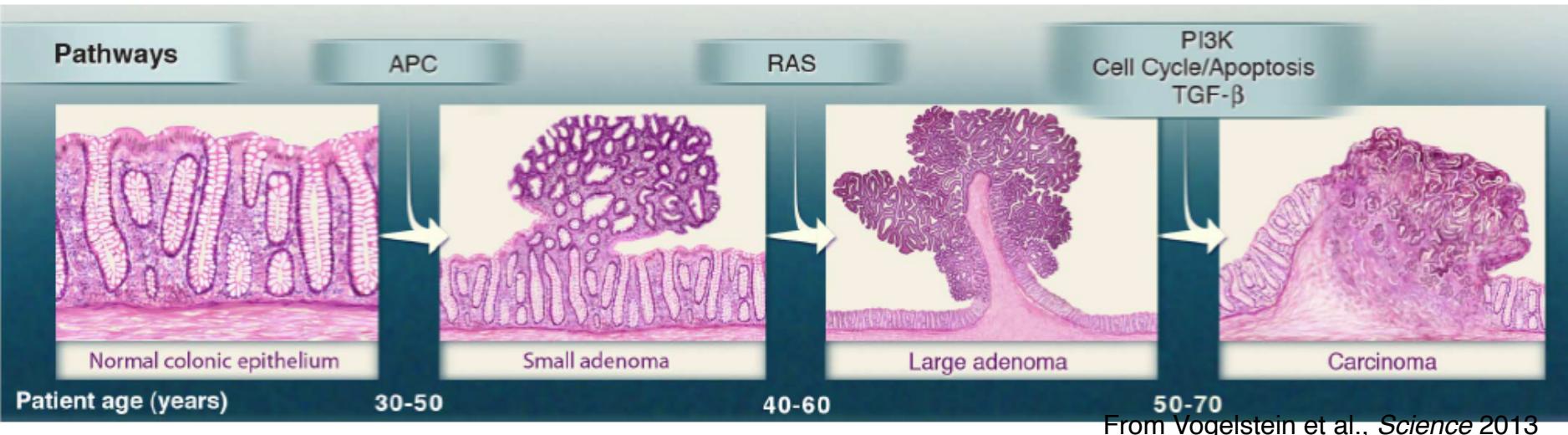
(Mutect2, VarScan, etc)

- **Compares tumor/normal** mismatches and gaps in the alignment to the reference
- Sensitive to **low frequency mutation**
- In some cases uses a database of **known cancer** mutations to support variant calls

Variant Browsing Tools and Packages

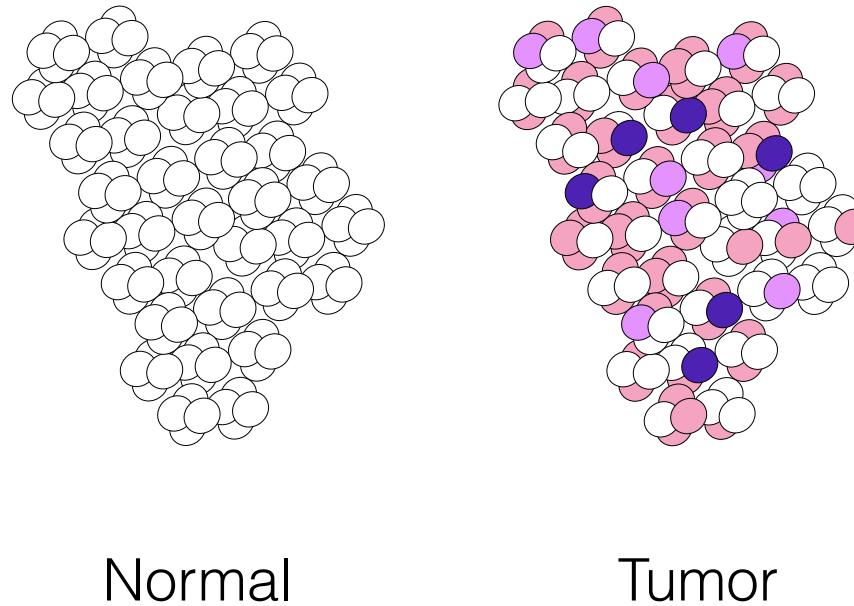
- Commercial
 - Ingenuity Variant Analysis
 - Golden Helix
 - Variant Studio
- Free Command Line
 - GEMINI
- Free GUI
 - KGGSeq — 30GB Download
 - FMFilter
- Free Webtools
 - gene.jobio.io

Finding Drivers of Cancer

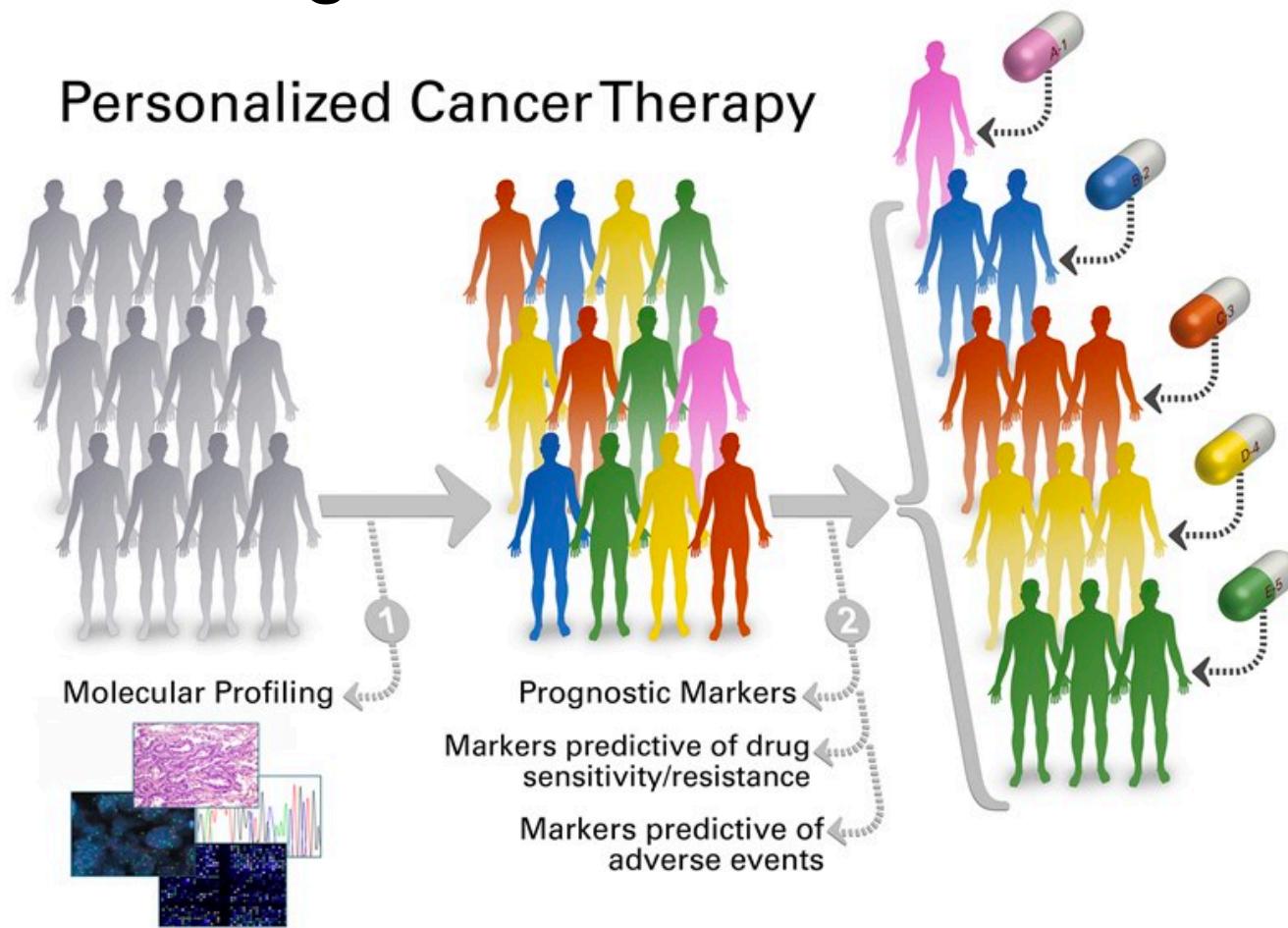


Progressive accumulation of pathogenic mutations in oncogenes culminates in the development of cancer

Tumors acquire mutations and can be heterogeneous



Personalize Cancer Therapy targets the “oncogenic drivers” of a tumor



Precision medicine is an approach to patient care that allows doctors to select treatments that are most likely to help patients based on the molecular mechanism of their disease

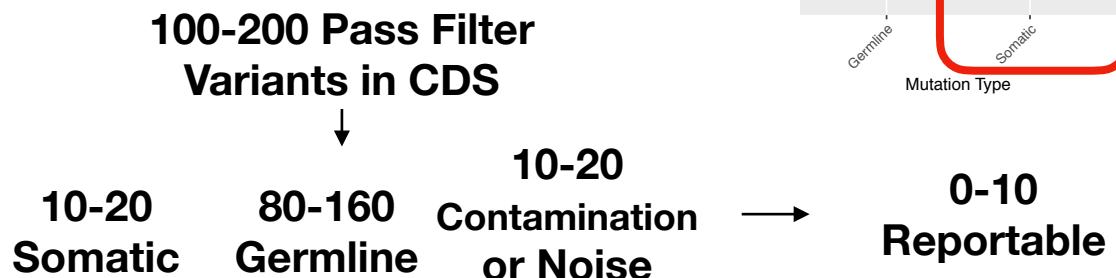
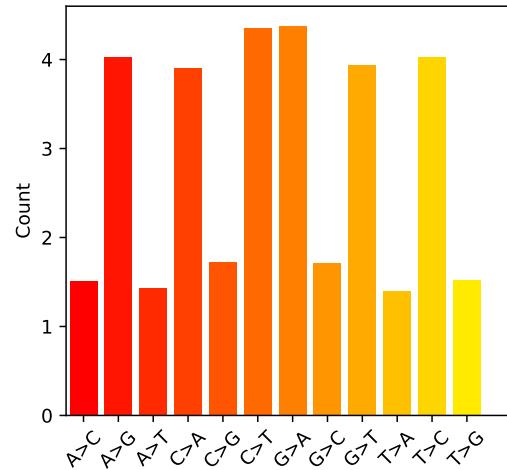
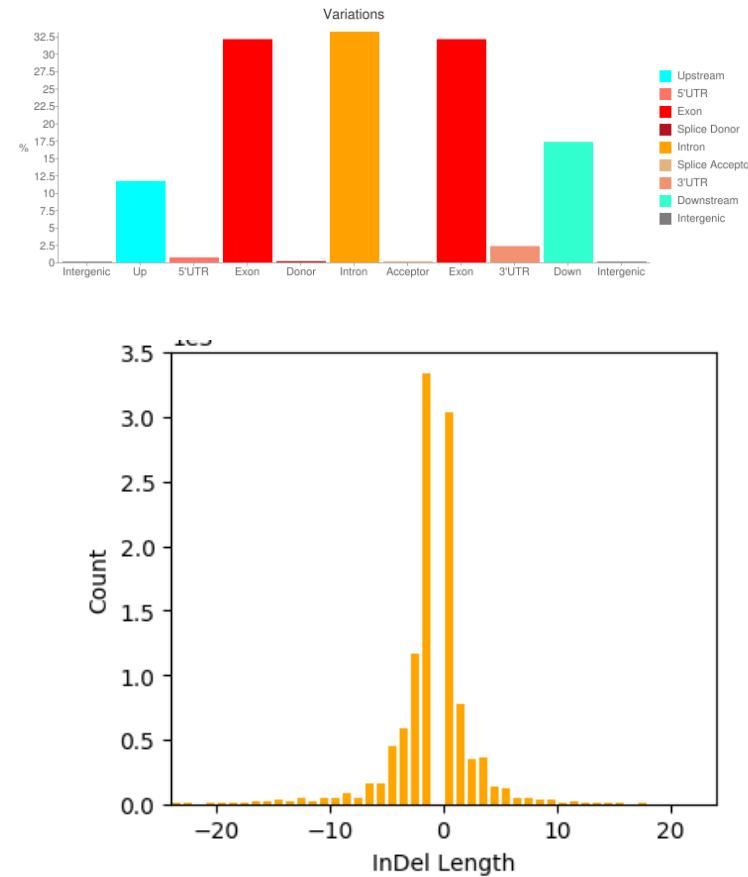
Genetic Changes in Tumors as a Proxy for Changes in Cellular Function

- An oncogene is a gene that has the potential to cause cancer because it promotes cell growth or represses apoptosis
- In tumor cells, oncogenes are often mutated or expressed at high levels.
- A tumor suppressor gene can have a damping or repressive effect on the regulation of the cell cycle and/or promote apoptosis
- In tumor cells, alteration in TSGs cause a loss or reduction in its function, thus promoting cell growth or the repression of apoptosis

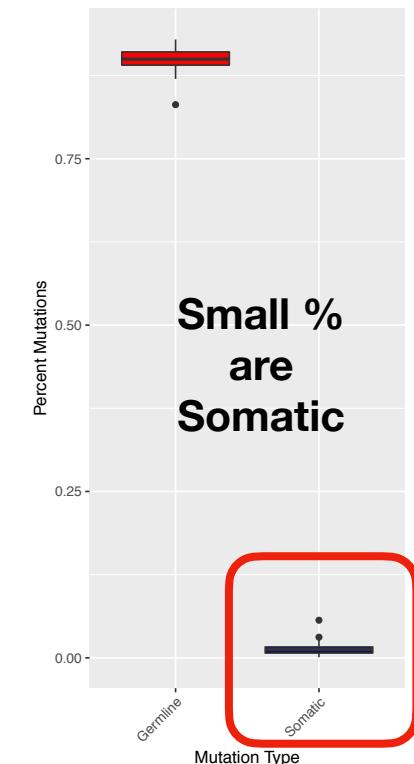
How do we find the oncogenic driver?

1. QC Checks
2. Find the genes effected by these mutations
3. Determine the likely functional consequence of these mutations
4. Differentiate rare human variation with somatic mutations
5. Determine which mutations are likely to be “clinically actionable.”
 1. FDA approved drugs
 2. Clinical Trials
6. Determine if the profile of these mutation are clinical actionable (immunotherapies)

SNV and Indels

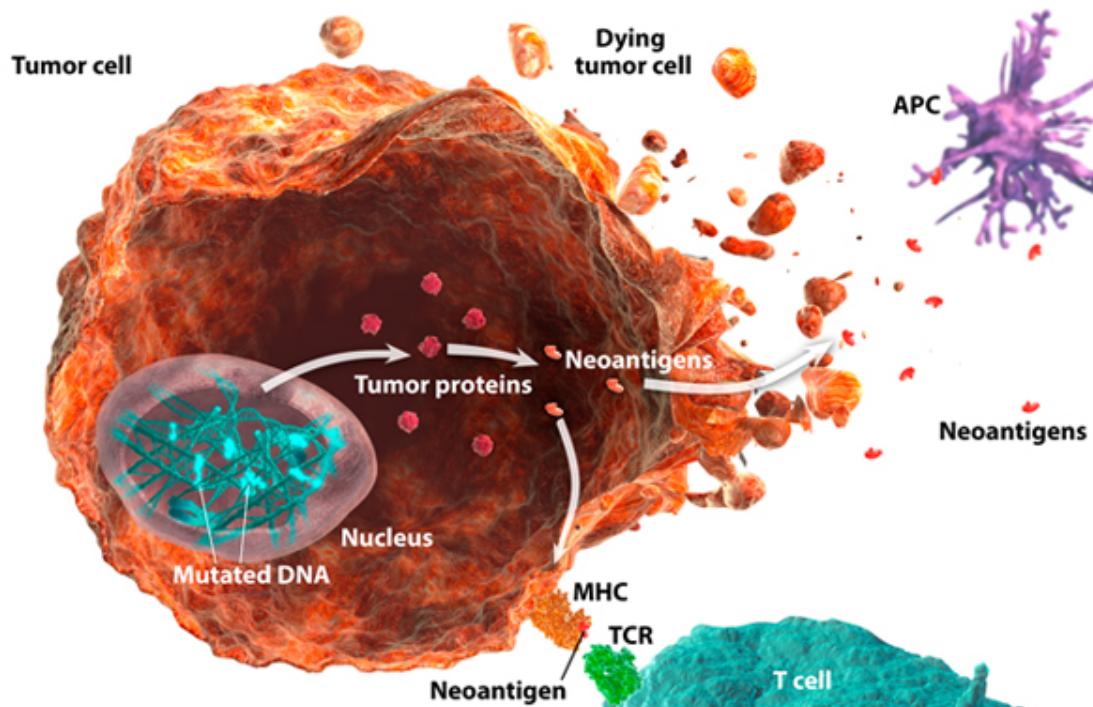


MSH6	p.Lys632fs	Likely pathogenic
ATM	p.Ser274fs	Likely pathogenic
FANCM	p.Ser1746*	Likely pathogenic
PMS1	p.Glu543*	Likely pathogenic



Tumor Mutational Burden

- Tumor mutational burden is a predictor for the response to immunotherapies
- Tumor cells with high TMB may have more neoantigens, with an associated increase in cancer-fighting T cells in the tumor microenvironment and periphery. These neoantigens can be recognized by T cells, inciting an anti-tumor response.



QC Checks

- Depth < 20
- Strand Bias, Repeats and low complexity sequence
- MAF (Normal) * 10.< MAF (Tumor)
- In COSMIC > 5 Subject
 - Tumor: Alt Read Ct < 3
 - Tumor: MAF < 0.05
- Others
 - Tumor: Alt Read CT < 8
 - Tumor: MAF < 0.05
 - Called by multiple variant calling algorithms

Databases for Cancer Studies and “Knowledge”

- Clinical Interpretation of Variants in Cancer (CIVIC)
 - Manually Curated Knowledgebase
- Catalog of Somatic Mutation in Cancer (COSMIC)
 - Gene Fusions
 - Gene Census
 - Curated Genes
 - Drug Resistance (so far 9 genes)
 - Genome Wide Screens

Variant Functional Classification in Cancer

- Tier 1
 - An FDA approved drug to treat cancers with this mutation
- Tier 2
 - A clinical trial is available for patients with this mutation
- Tier 3
 - Variant of Unknown Significant (VUS)
- Tier 4
 - Benign

Reporting

INDICATED THERAPIES

DRUGS	VARIANT	LEVEL	INDICATION
	IDH2 p.Arg140Gln	FDA-Approved	The IDH2-targeted inhibitor enasidenib is FDA-approved for the treatment of patients with IDH2-mutant acute myeloid leukemia.

1385-Gene Pan-Cancer Mutation Test

MRN 93947728 BROWN,ALLISANDRA

page 1/5



TITLE	PHASE	TARGETS	LOCATIONS	NCT ID
CLINICAL TRIALS				
TITLE	PHASE	TARGETS	LOCATIONS	NCT ID
A Master Protocol for Biomarker-Based Treatment of AML (The Beat AML Trial)	Phase 1/Phase 2	RUNX1 RUNX1T1	Spencer Kalk 919-227-5843 spencer.kalk@incresearch.com Dallas, TX	NCT03013998

Indicated Therapies

← Clinical Trials

Clinical Annotations

TIER 1 - VARIANTS OF STRONG CLINICAL SIGNIFICANCE

VARIANT	COMMENT
IDH2 p.Arg140Gln Pos: chr15:90088702 ENST: ENST00000330062.7 VAF: 37.61% Depth: 3023	<p>Epidemiology: IDH1 and IDH2 mutations are most common in patients with cytogenetically normal AML, and the prevalence of IDH1 and IDH2 missense mutations among patients with AML is between 5% and 20% (PMID: 23999441)</p> <p>Gene Function: IDH2, isocitrate dehydrogenase 2, is an enzyme that catalyzes the conversion of isocitrate to alpha-ketoglutarate (alpha-KG), and is involved in the tricarboxylic acid cycle (PMID: 28711227, 28980701). Mutations in IDH2 are associated with aberrant conversion of alpha-KG to 2-HG, which is an oncogenic metabolite, and leads to altered epigenetics, RNA methylation, cellular signaling, hypoxic response, and DNA repair.</p> <p>Therapy: The IDH2-targeted inhibitor enasidenib is FDA-approved for the treatment of patients with IDH2-mutant acute myeloid leukemia.</p> <p>Variant Function: R140Q confers a gain of function to IDH2, enabling conversion of alpha-ketoglutarate to the onco-metabolite 2HG (R(-)-2-hydroxyglutarate), results in increased 2HG levels in patient samples, and is transforming in cell culture (PMID: 20171147, PMID: 23558173).</p>

UT Southwestern

Lyda Hill Department of Bioinformatics

Variant Prioritization

- In Cancer (tumor/normal pairs)
 - somatic mutation identification
- In small family studies (trios)
 - rare germline variation
- In large populations (GWAS)
 - Covered this afternoon

Intersample QC

There are many reasons why the genotyped samples may not represent the intended relationships with each other

- Sample swaps
- Duplicate samples
- Sample contamination
- Cryptic relatedness
- Pedigree errors/False paternity

PEDDY

Peddy seeks discrepancies between relationships and sexes as indicated in a .PED file with those inferred from genotypes.

- Works directly on the VCF and PED
- It samples ~23,000 well-behaved exome sites in the genome
- Generates interactive plots for QC
- Runs on a trio in a few seconds.
- Runs on 2K samples in ~10 minutes.

Common FP

- Large genes
 - TTN
 - USH2A
 - MUC16
 - FLG
- Lots of paralogs/part of gene family

Don't rule out if phenotype makes sense!

- *TTN* mutations cause dilated cardiomyopathy and muscular dystrophy
- *MUC1* mutations cause medullary cystic kidney disease
- *KRT** gene mutations cause ichthyosis, keratoderma, keratosis

Count	String	Description
91	LOC	LOC genes
22	ENS	Ensembl genes
21	FAM	FAM proteins
15	GOL	Golgi-like GOLGA8E
13	PRA	PRAMEF genes
9	NBP	Nuclear breakpoint family
7	POT	POTE ankyrin domain family
6	DEF	defensins
5	OR2	Olfactory receptor
5	MUC	Mucins
5	KRT	Keratins
4	WAS	WAS protein family homolog
4	ANK	ankyrins
3	TRI	tri-partite motif containing
3	OR1	Olfactory receptor
3	FRG	FSHD region gene

Pseudogene database: <http://pseudofam.pseudogene.org>
<http://massgenomics.org/2013/06/ngs-false-positives.html>

Variant Browsing Tools and Packages

- Commercial
 - Ingenuity Variant Analysis
 - Golden Helix
- Free Command Line
 - VAAST — Variant, Annotation, Analysis Search Tool
 - GEMINI
- Free GUI
 - KGGSeq — 30GB Download
 - FMFilter
- Free Webtools
 - gene.iobio.io
 - <https://bystro.io>
 - GenEsysV
 - Phevor — Phenotype Driven Variant Ontological ReRanking Tool
 - <http://weatherby.genetics.utah.edu/phevor2/index.html>
 - Phenovar
 - <https://phenovar.med.usherbrooke.ca/>

Table 1 Comparison of existing open-source software tools with similar functions

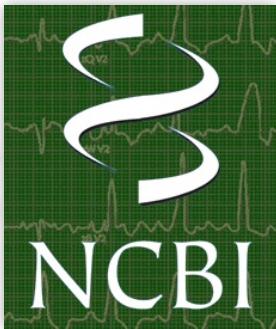
From: [GenESysV: a fast, intuitive and scalable genome exploration open source tool for variants generated from high-throughput sequencing projects](#)

Features	GenESysV	GEMINI	BrowseVCF	VCF-Miner	Mendel,MD	BiERapp
Graphical User ^a Interface	Yes	No	Yes	Yes	Yes	Yes
Study type	Single cohort complex disease, Case/Control, and Mendelian inheritance	Single cohort complex disease and Mendelian inheritance	Single cohort complex and Mendelian inheritance	Single cohort complex disease and Mendelian inheritance	Mendelian only	Single cohort complex disease, Case/Control, and Mendelian inheritance
Whole genome, exome or target study	All	All	All	All	WES or targeted study	WES or target study
Can handle studies with large numbers of samples	Yes	Yes	No	No	No	No
Database Type	Elasticsearch	Sqlite3	Wormtable & BerkeleyDB	MongoDB	PostgreSQL	SQLite & MongoDB
Flag variants for further filtering	Yes	No	No	No	No	No

^aFeatures listed here are not exhaustive

NCBI Resources

MedGen



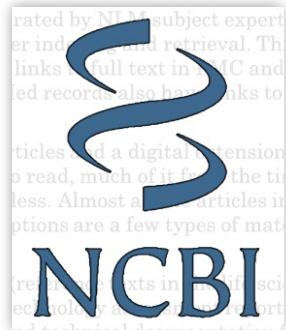
Genetic Testing Registry
(GTR)

ClinicalTrials.gov

PubMed Health

Books

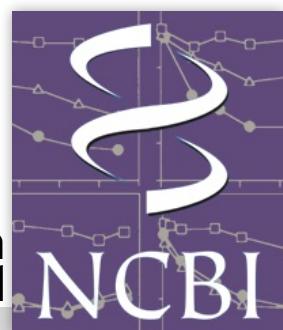
NLM Catalog
(Journals)



PubMed

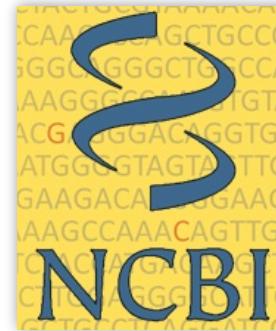
PubMed Central
(PMC)

PubChem
Compound



dbGaP

ClinVar



Genome Assembly
dbSNP

Epigenomics

dbVar

SRA

Nucleotide

GEO

BioSystems



Gene

UniGene

HomoloGene



Protein

Conserved Domains
Database (CDD)

Structure

UT Southwestern

MedGene

NCBI Resources How To

Sign in to NCBI

MedGen

MedGen

Dystonia AND Spasticity AND Sleep disturbance AND Hyperactivity

Search

Create alert Limits Advanced

Help

Full Report ▾

Send to: ▾

Table of contents

Disease characteristics

Additional descriptions

Clinical features

The following clinical feature is unrelated to
Sepiapterin reductase deficiency

Term Hierarchy

Recent clinical studies

Genetic Testing Registry

Analyte (1)

Deletion/duplication analysis (9)

Linkage analysis (1)

Sequence analysis of the entire coding
region (31)

Targeted variant analysis (1)

See all (32)

Clinical resources

OMIM

Orphanet

ClinicalTrials.gov

Molecular resources

OMIM

View SPR variations in ClinVar

RefSeqGene

Coriell Institute for Medical Research

Sepiapterin reductase deficiency

MedGen UID: 120642 • Concept ID: C0268468 • Disease or Syndrome

Synonyms: Dopa-Responsive Dystonia Due to Sepiapterin Reductase Deficiency; SPR deficiency

Modes of inheritance: *Autosomal recessive inheritance* (HPO, OMIM, Orphanet)
Autosomal dominant inheritance (HPO)

SNOMED CT: Sepiapterin reductase deficiency (45116002); 7,8-Dihydrobiopterin synthetase deficiency (45116002); Biopterin deficiency (45116002)

Gene (location): SPR (2p13.2)

OMIM®: 612716

Orphanet: ORPHA70594

Disease characteristics

Go to: ▾

Excerpted from the GeneReview: Sepiapterin Reductase Deficiency

The phenotypic spectrum of sepiapterin reductase deficiency (SRD), which ranges from significant motor and cognitive deficits to only minimal findings, has not been completely elucidated. Clinical features in the majority of affected individuals include motor and speech delay, axial hypotonia, dystonia, weakness, and oculogyric crises; symptoms show diurnal fluctuation and sleep benefit. Other common features include parkinsonian signs (tremor, bradykinesia, masked facies, rigidity), limb hypertonia, hyperreflexia, intellectual disability, psychiatric and/or behavioral abnormalities, autonomic dysfunction, and sleep disturbances (hypersomnolence, difficulty initiating or maintaining sleep, and drowsiness). Most affected individuals have nonspecific features in infancy including developmental delays and axial hypotonia; other features develop over time. [from GeneReviews]

Full text of GeneReview (by section):

[Summary](#) | [Diagnosis](#) | [Clinical Characteristics](#) | [Genetically Related \(Allelic\) Disorders](#) | [Differential Diagnosis](#) | [Management](#) | [Genetic Counseling](#) | [Resources](#) | [Molecular Genetics](#) | [References](#) | [Chapter Notes](#)

Authors:

Jennifer Friedman [view full author information](#)

Go to: ▾

Additional descriptions

From NCBI curation

Sepiapterin reductase deficiency is a condition characterized by movement problems, most often a pattern of involuntary, sustained muscle contractions known as dystonia. Other movement problems can include muscle stiffness (spasticity), tremors, problems with coordination and balance (ataxia), and involuntary jerking movements (chorea). People with sepiapterin reductase deficiency can experience episodes called oculogyric crises. These episodes involve abnormal rotation of the eyeballs; extreme irritability and agitation; and pain, muscle spasms, and uncontrolled movements, especially of the head and neck. Movement abnormalities are often worse late in the day. Most affected individuals have delayed development of motor skills such as sitting and crawling, and they typically are not able to walk unassisted. The problems with movement tend to worsen over time. People with sepiapterin reductase deficiency may have additional signs and symptoms including an unusually small head size (microcephaly), intellectual disability, seizures, excessive sleeping, and mood swings.

From OMIM

SPR deficiency results in neurologic deterioration due to severe dopamine and serotonin deficiencies in the central nervous system caused by a

Gene

NCBI Resources How To

Sign in to NCBI

Gene

Gene

Advanced

Search

Help

Full Report

Send to:

Hide sidebar >>

SPR sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase) [*Homo sapiens* (human)]

Gene ID: 6697, updated on 4-Sep-2016

Summary

Official Symbol SPR provided by HGNC

Official Full Name sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase) provided by HGNC

Primary source HGNC:HGNC:1125

See related Ensembl:ENSG00000116096 HPRD:01632; MIM:182125; Vega:OTTHUMG00000129777

Gene type protein coding

RefSeq status REVIEWED

Organism *Homo sapiens*

Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as SDR38C1

Summary This gene encodes an aldo-keto reductase that catalyzes the NADPH-dependent reduction of pteridine derivatives and is important in the biosynthesis of tetrahydrobiopterin (BH4). Mutations in this gene result in DOPA-responsive dystonia due to sepiapterin reductase deficiency. A pseudogene has been identified on chromosome 1. [provided by RefSeq, Jul 2008]

Orthologs [mouse](#) [all](#)

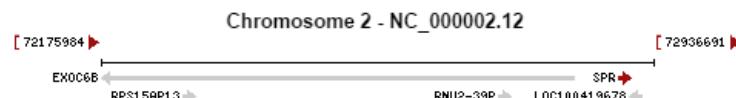
Genomic context

Location: 2p14-p12

See SPR in [Genome Data Viewer Map Viewer](#)

Exon count: 3

Annotation release	Status	Assembly	Chr	Location
108	current	GRCh38.p7 (GCF_000001405.33)	2	NC_000002.12 (72887383..72892160)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	2	NC_000002.11 (73114512..73119289)



- Table of contents
- Summary
 - Genomic context
 - Genomic regions, transcripts, and products
 - Expression
 - Bibliography
 - Phenotypes
 - Variation
 - Pathways from BioSystems
 - Interactions
 - General gene information
 - Markers, Related pseudogene(s), Homology, Gene Ontology
 - General protein information
 - NCBI Reference Sequences (RefSeq)
 - Related sequences
 - Additional links
 - Locus-specific Databases

Genome Browsers

Genome Data Viewer

Map Viewer

Variation Viewer (GRCh37.p13)

Variation Viewer (GRCh38)

1000 Genomes Browser (GRCh37.p13)

Ensembl

UCSC

NCBI CS

Gene

Gene

Advanced

Search

Help

Full Report

SPR sepiapterin reductase

Gene ID: 6697, updated on 4-3-

Summary

Official Symbol

Official Full Name

Primary source

See related

Gene type

RefSeq status

Organism

Lineage

Also known as

Summary

Orthologs

Genomic context

Location: 2p14-p12

Exon count: 3

Annotation release

108

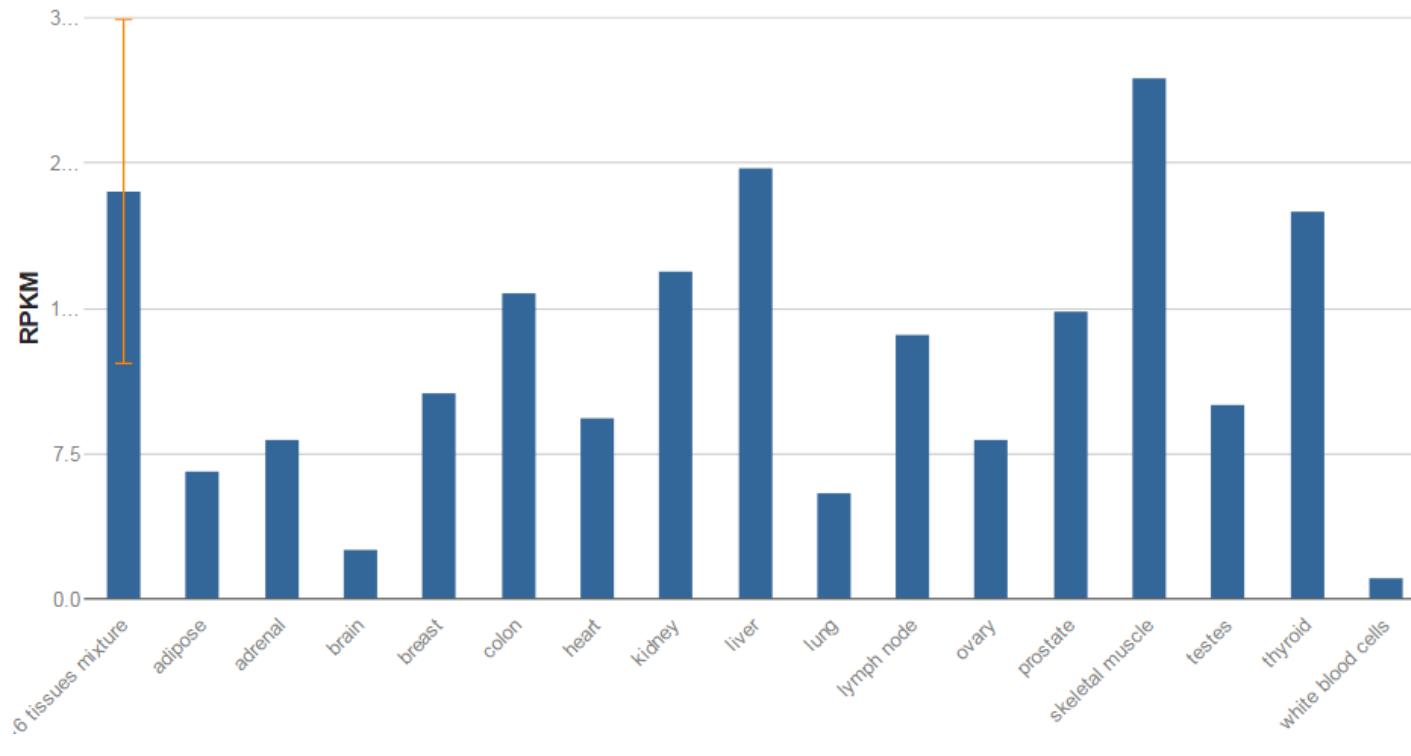
105

Expression

Illumina bodyMap2 transcriptome

[See details](#)

- Project title: Illumina bodyMap2 transcriptome
- Description: Transcription profiling by high throughput sequencing of individual and mixture of 16 human tissues RNA.
- BioProject: [PRJEB2445](#)
- Analysis date: n/a



RPS15AP13

RNU2-39P

LOC100419678

EMX1

Databases for Known Disease Associated Variation

- ClinVar
 - ClinVar is a freely accessible, public archive of reports of the relationships among human variations and phenotypes, with supporting evidence
- GWAS Catalog
 - The Catalog is a quality controlled, manually curated, literature-derived collection of all published genome-wide association studies assaying at least 100,000 SNPs and all SNP-trait associations with p-values < 1.0×10^{-5}
- Decipher
 - The DECIPHER database contains data from 20305 patients who have given consent for broad data-sharing; DECIPHER also supports more limited sharing via consortia.

Variation ID: ?

89533

Review status: ?

★★★☆ reviewed by expert panel

Interpretation ?

Go to: ☰ ▲

Clinical significance: [Pathogenic](#)

Last evaluated: Sep 5, 2013

Number of submission(s): 2

Condition(s): Lynch syndrome [[MedGen](#) - [Orphanet](#) - [OMIM](#)][See supporting ClinVar records](#) ↗**Allele(s)** ?

Go to: ☰ ▲

NM_000179.2(MSH6):c.458-?_627+?del

Allele ID: 95007

Variant type: Deletion

- HGVS:
- NM_000179.2:c.458-?_627+?del
 - LRG_219t1:c.458-?_627+?del
 - LRG_219:g.(?-17748)_(17917_-?)del

Assertion and evidence details

Go to: ☰ ▲

[Clinical assertions](#)[Summary evidence](#)[Supporting observations](#)

ClinVar

GermlineFilter:

Clinical significance (Last evaluated)	Review status (Assertion method)	Collection method	Condition(s) (Mode of inheritance)	Origin	Citations	Submitter - Study name	Submission accession
Pathogenic (Sep 5, 2013)	reviewed by expert panel <ul style="list-style-type: none"> • Guidelines v1.9 	research	Lynch syndrome [MedGen Orphanet OMIM]	germline	<ul style="list-style-type: none"> • Other citation ↗ 	International Society for Gastrointestinal Hereditary Tumours (InSiGHT)	SCV000108216.2
Pathogenic (Oct 13, 2016)	criteria provided, single submitter <ul style="list-style-type: none"> • ACMG Guidelines, 2015 • ACMG Guidelines, 2015 	clinical testing	Lynch syndrome [MedGen Orphanet OMIM]	germline	<ul style="list-style-type: none"> • PubMed (2) [See all records that cite these PMIDs] 	Department of Pathology and Laboratory Medicine, Sinai Health System - The Canadian Open Genetics Repository (COGR)	SCV000591006.1

Dicpher

Phenotypic abnormality in open-access patients in DECIPHER

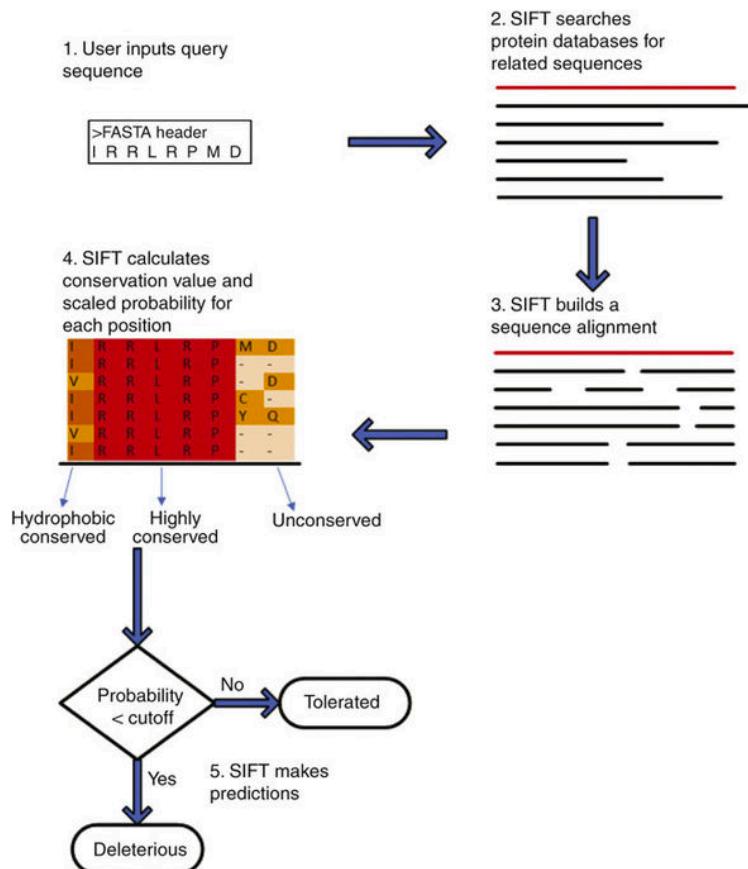
[Hide redundant paths](#)[Simple view](#)[About](#)

- Abnormality of the cardiovascular system
- Abnormality of blood and blood-forming tissues
- Abnormality of the endocrine system
- Abnormality of the musculature
- Abnormality of the digestive system
- Abnormality of the skeletal system
- Abnormality of limbs
- Abnormality of prenatal development or birth
- Growth abnormality
- Abnormality of the genitourinary system
- Abnormality of connective tissue
- Abnormality of the ear
- Abnormality of metabolism/homeostasis
- Abnormal cellular phenotype
- Abnormality of the eye
- Abnormality of the nervous system
- Neoplasm
- Constitutional symptom
- Abnormality of the immune system
- Abnormality of the voice
- Abnormality of the respiratory system
- Abnormality of head or neck
- Abnormality of the breast
- Abnormality of the integument

If No Known Variant

- Likely to effect protein function
 - SIFT
 - PolyPhen
- Likely to be evolutionarily conserved
 - GERP
 - PhastCons
 - PolyP

SIFT



SIFT: sorting intolerant from tolerant

- For single amino acid substitution
- Based on protein sequence
- Estimated by conservation
- Structure information is not used
- Score from 0 to 1, at or below 0.05 is damaging

Nature Protocols volume 4
pages 1073–1081 (2009)

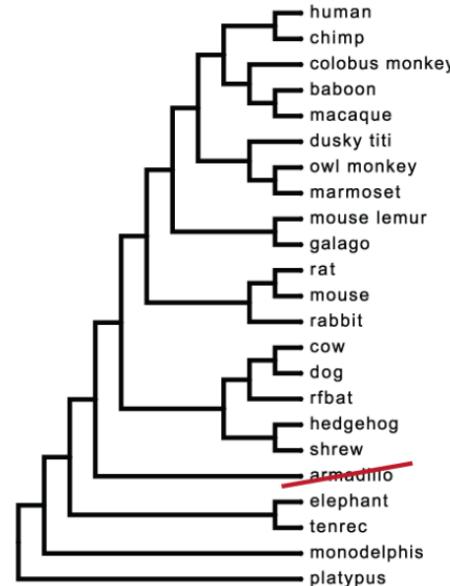
GERP

GERP: Genomic Evolutionary Rate Profiling

- Genome-wide, single-base resolution
- Range of -12.3 to 6.17, with 6.17 being the most conserved

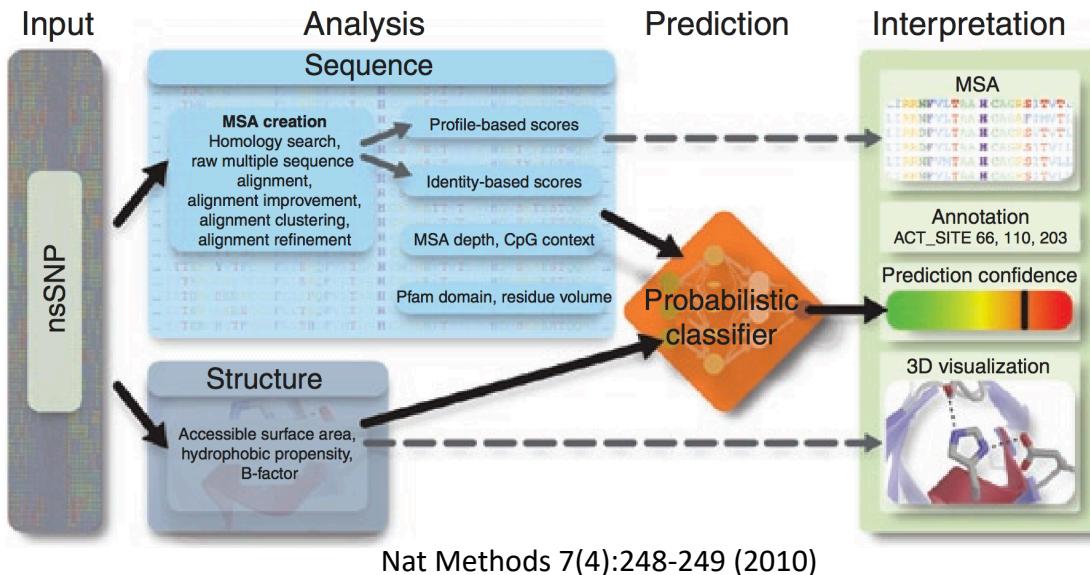
Multiple Sequence Alignment

human	AATACGG	A	ACTTCATTCA	T
chimp	AATATGG	A	ACTTCATTCA	T
colobus monkey	AGTATGG	A	ACTTCATTCA	T
baboon	AGTATGG	A	ACTTCATTCA	T
macaque	AGTATGG	A	ACTTCATTCA	T
dusky titi	AGTATGG	A	ACTTCATTCA	T
owl monkey	AGTATGG	A	ACTTCATTCA	T
marmoset	AGTATGG	A	ACTTCATTCA	T
mouse lemur	AGTACGG	A	ACTTCATTCA	T
galago	AGTACGG	A	ACTTCATTCA	T
rat	AGTATGG	A	ACATCGTC	CATT
mouse	AGTATGG	A	ACATCTTC	CATT
rabbit	AGTATGG	A	ACATCATT	CATT
cow	AGTATGG	A	ACATCATT	CATT
dog	AGTACGG	A	ACATCATT	CATT
rfbat	AGTATGG	A	ACATCGTC	CATT
hedgehog	AGTATGG	A	ACATCATT	CATT
shrew	AGTATGG	G	ACATCC	TTCA
armadillo	-----	-	-----	-
elephant	AGTATGG	A	ACATCGTC	CATT
tenrec	AGTATGG	A	ACATCGTC	CATT
monodelphis	AGTATGG	G	ACATCTTC	CATT
platypus	AGTATGG	A	ACGTCA	TTCATT



PLoS Comput Biol. 2010 Dec 2;6(12)

CADD: Consensus



PolyPhen: Polymorphism Phenotyping

- 8 protein sequence features
- 3 protein structure features
- Between 0 and 1, larger being more deleterious

CADD: Combined Annotation Dependent Depletion

- Generate 63 distinct annotations, including scores from PhastCons, GERP, PhyloP, SIFT and PolyPhen, etc., ENCODE data (summarized at various levels), gene body annotations...
- Build a support vector machine (SVM) that estimates, for a given variant, whether it is likely to be observed or simulated, based on its combined annotation profile
- The C score ranges from 1 to 99, with a higher score indicating greater deleteriousness. Values ≥ 10 are predicted to be the 10% most deleterious substitutions, ≥ 20 indicate the 1% most deleterious.
- Score all possible ~ 8.6 billion possible SNVs of hg19

Phenomizer

Menu. ▾ Support the Phenomizer. Help.

The Phenomizer

Features. **Diseases.** **Ontology.**

muscle hypotonia

HPO id.	Feature.
HP:0001252	Muscular hypotonia

Patient's Features. **Diagnosis.**

HPO.	Feature. ▲	Modifier.	Num diseases.
category.: Abnormality of head or neck (1 Item)			
HP:0002307	Drooling	observed.	29 of 7994
category.: Abnormality of the musculature (1 Item)			
HP:0001252	Muscular hypotonia	observed.	1099 of 7994
category.: Abnormality of the nervous system (5 Items)			
HP:0002307	Drooling	observed.	29 of 7994
HP:0001332	Dystonia	observed.	193 of 7994
HP:0001263	Global developmental delay	observed.	772 of 7994
HP:0002322	Resting tremor	observed.	13 of 7994
HP:0002360	Sleep disturbance	observed.	148 of 7994

The Phenomizer

Patient's Features. **Diagnosis.**

Algorithm: resnik (Unsymmetric). | 6 Features.

p-value. ▲	Disease Id.	Disease name.
0.0011	OMIM:606324	PARKINSON DISEASE 7, AUTOSOMAL RECESSIVE EARLY-ONSET
0.0011	OMIM:605909	PARKINSON DISEASE 6, AUTOSOMAL RECESSIVE EARLY-ONSET
0.0011	OMIM:615493	#615493 MENTAL RETARDATION, AUTOSOMAL RECESSIVE 37; MRT37
0.0011	OMIM:300911	#300911 PARKINSONISM WITH SPASTICITY, X-LINKED; XPDS
0.0011	OMIM:611092	#611092 MENTAL RETARDATION, AUTOSOMAL RECESSIVE 6; MRT6
0.0014	OMIM:261630	#261630 HYPERPHENYLALANINEMIA, BH4-DEFICIENT, C; HPABH4C;;HYPERPHENYLALANINE
0.0014	OMIM:261640	#261640 HYPERPHENYLALANINEMIA, BH4-DEFICIENT, A; HPABH4A;;HYPERPHENYLALANINE

Other Supporting Evidence

- Phenotypes observed in model organisms
(Monarch Initiative - <http://monarchinitiative.org>)

Gene-based predictions

- ClinGen haploinsufficiency score <https://www.ncbi.nlm.nih.gov/projects/dbvar/clingen/>
- ExAC Probability of loss of function intolerance (pLI)

ACMG Guidelines

	Evidence for pathogenicity					
	Benign			Pathogenic		
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in <i>trans</i> with a dominant variant BP2 Observed in <i>cis</i> with a pathogenic variant BP2		For recessive disorders, detected in <i>trans</i> with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Caveats

- Same apparent condition, different underlying genes
- Same condition, same gene, different inheritance patterns
- Different conditions, same gene, same inheritance pattern
- Different conditions, same gene, different inheritance patterns
- Compound inheritance:
 - SNV+CNV
 - SNV+indel
- Mosaicism:
 - Tissue from the somatic mosaic parent contains the mutation
 - Tissue from the somatic mosaic proband does not contain mutation
- Non-coding
 - Synonymous mutations that actually affect splicing
 - Intronic mutations that activates a pseudoexon
 - Mutations affecting lncRNA, enhancers, etc
- Complex Events, e.g. repeat expansions
- Polygenic conditions
- Uniparental disomy
- Imprinting
- Environmental rather than genetic

Variant Prioritization

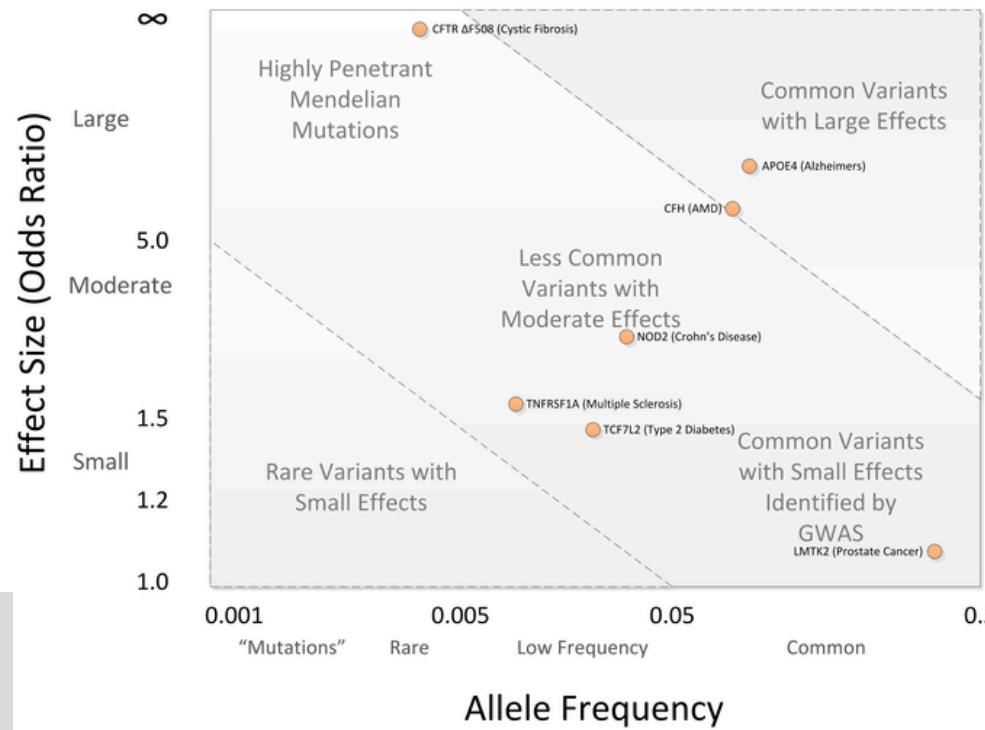
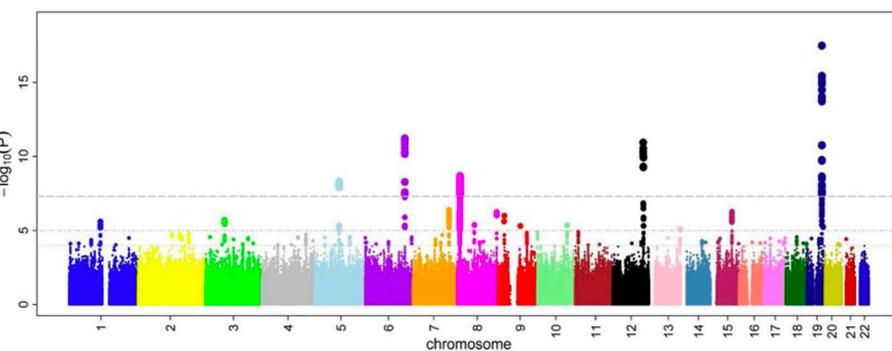
- In Cancer (tumor/normal pairs)
 - somatic mutation identification
- In small family studies (trios)
 - rare germline variation
- In large populations (GWAS)
 - common variants associated with traits

Complex Diseases

- Influenced by variation in many genes often acting together with the environment
- Cluster in families but do not follow Mendelian inheritance patterns
- Disease susceptibility influenced by many alleles having small effect
- Common disease such as cardiovascular disease and type 2 diabetes

GWAS

- Genome Wide Association studies examines associations between single-nucleotide polymorphisms (SNPs) and traits using statistical methods like Fisher Exact Test
- Often these associations have varying contributions to the trait (effect size).



	Cases	Controls
Ref Allele	X	Y
Alt Allele	A	B

PheWAS

- Phenome-wide association studies (PheWAS) examine the causal linkage between known sequence differences and any type of trait, including molecular, biochemical, cellular, and especially clinical diagnoses and outcomes..
- For example, given a single nucleotide polymorphism (SNP) identified by GWAS (SNP: rs17234657) and association with infection, one may conclude that the SNP increases susceptibility of the host.
- In contrast, with PheWAS new putative associations may be identified through interrogation of phenomic markers within the EHR. Hence, an alternative mechanism is identified, where rs17234657 is found to be associated with an increase in autoimmune disease and the treatment used (immunosuppressive medication) is the cause of the infection.