Clinical Genomics

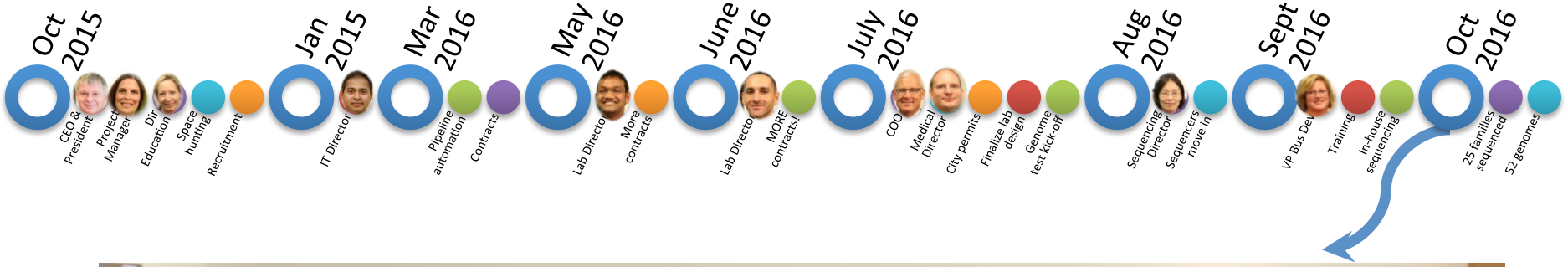
Narayanan "Ray" Veeraraghavan, PhD

Director of Informatics

Rady Children's Institute for Genomic Medicine, San Diego, CA USA

Rady Children's Institute

Genomic Medicine



Rady Children's Institute
Genomic Medicine

Outline

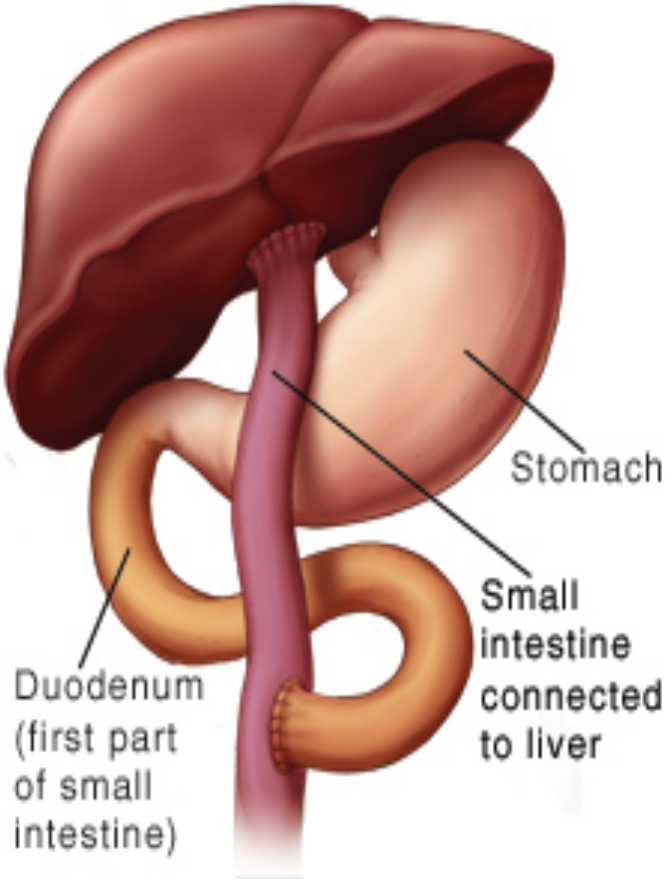
- WHY genomics in clinic?
- WHAT is Clinical Genomics
  - Steps involved in Clinical Genomics
  - Ever increasing bioinformatics tool set
- HOW to be a good clinical bioinformaticist
  - Skills to develop
  - Compliance
  - Resources

Let us begin with a story ……


Baby admitted
to regional
Level 4 NICU

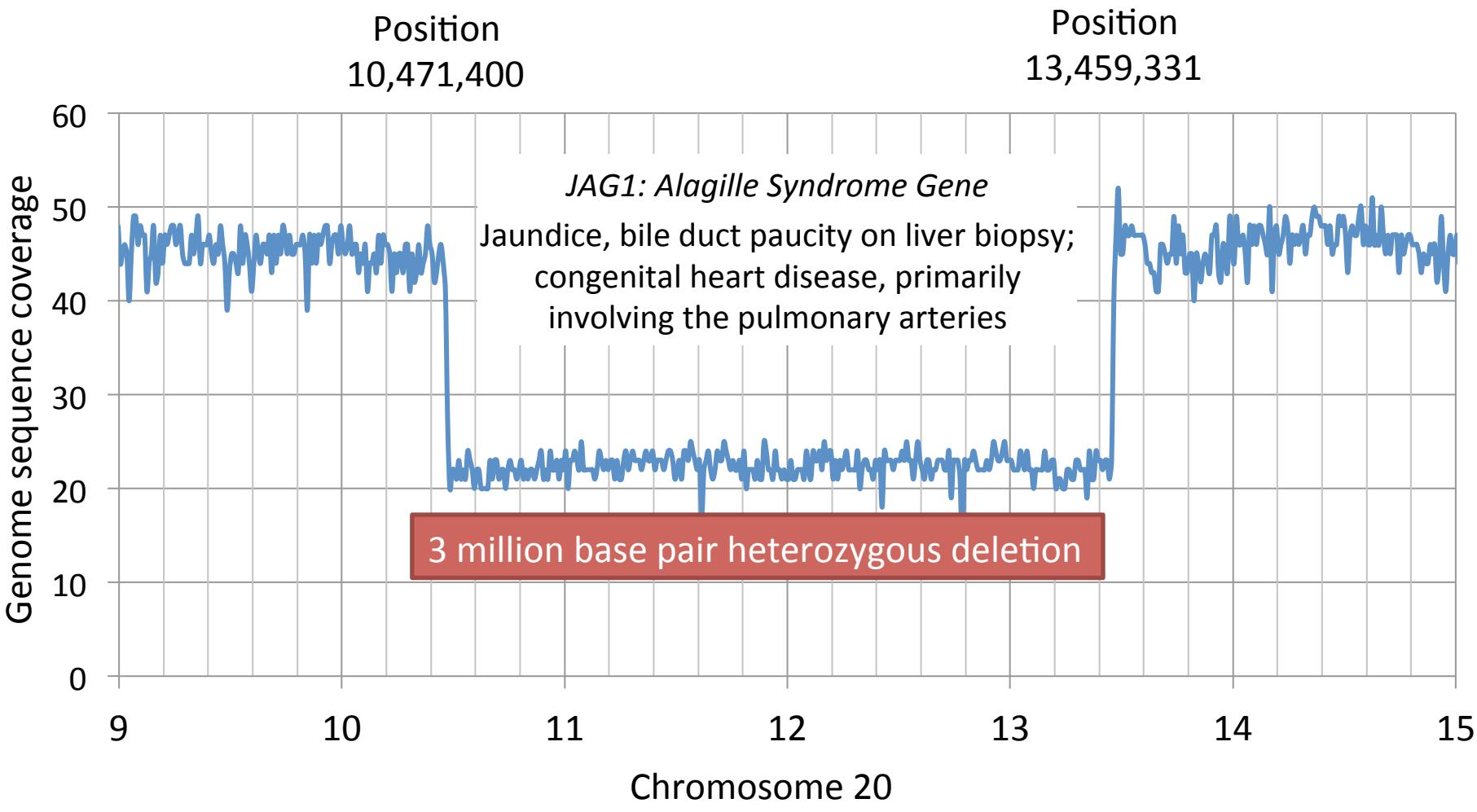
Rady Children's Hospital Baby 6026

- 2 month old admitted to PICU with severe jaundice & poor weight gain for 1 month
- Weight 0.05%^{ile}
- Acidotic, tachypneic
- Echo: Congenital heart disease, underdeveloped pulmonary arteries
- Clinical diagnosis: biliary atresia
  - Incidence 1 in 10,000
- Empiric treatment: Kasai procedure
  - Prognosis worsens with time to surgery
- Liver biopsy not diagnostic: giant cell hepatitis, sample too small to count bile ducts



Rady Children's Institute
Genomic Medicine

43 Hours Later: Provisional Diagnosis



Rady Children's Institute
Genomic Medicine

Clinical Impact & Outcome

- Kasai procedure and intraoperative cholangiogram scheduled at 11:00 am
- Genetic diagnosis communicated to clinical team just before operation – procedure cancelled
- Studies have shown kids with Alagille Syndrome w/ o Kasai have 50% decrease in mortality

*"Infants with Alagille syndrome **are occasionally misdiagnosed as biliary atresia** and subsequently undergo Kasai operation during infancy.*

*Among 15 children with Alagille syndrome, mortality was 60% among the Kasai group, and **10% among the non-Kasai group** . Liver transplantation was performed in 100% of the Kasai group, and **20% of the non-Kasai group**."*

Rady Children's Institute
Genomic Medicine

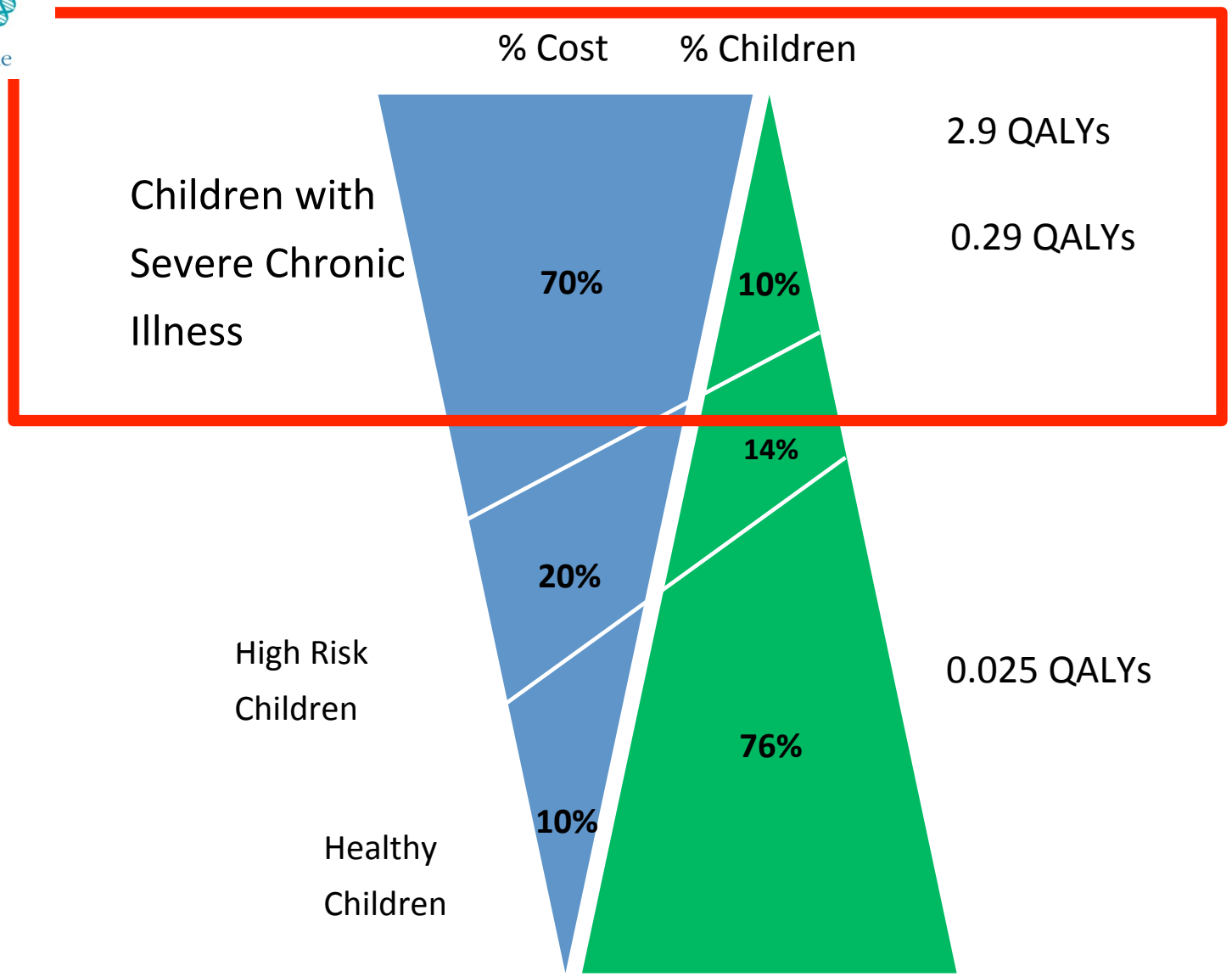
On an Average Day in the US...

Births

Babies born with 1 of 8,000 genetic diseases

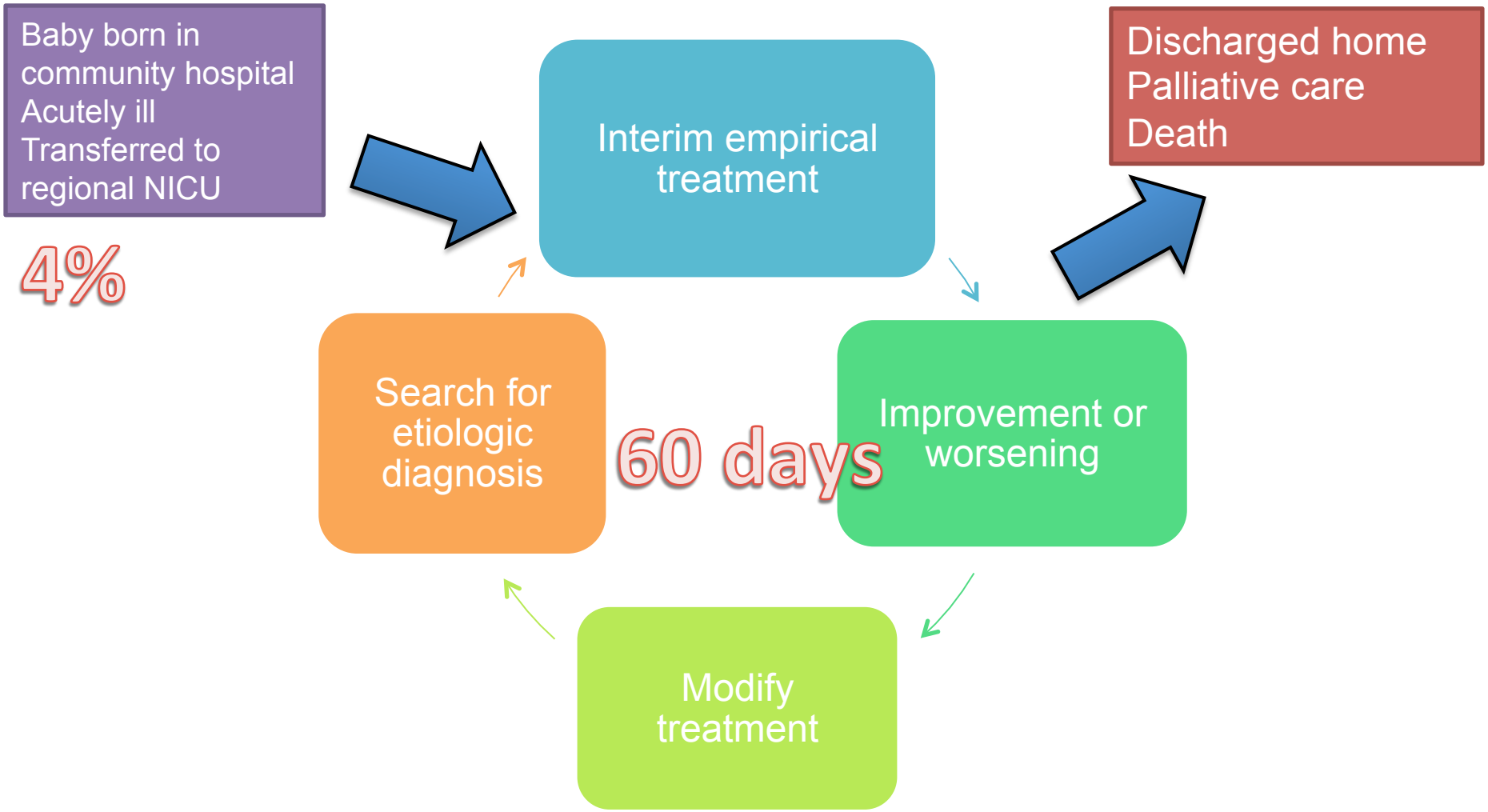
Babies die before 1st birthday

Rady Children's Institute
Genomic Medicine



Rady Children's Institute
Genomic Medicine

On an Average Day...



Rady Children's Institute
Genomic Medicine

...Rady Children's Hospital – San Diego

Baby born in community hospital
Acutely ill
Transferred to regional NICU

4%

Rapid Whole Genome Sequencing

Rapid Precision Medicine

2 days

RECORD GUINNESS WORLD RECORDS HOLDER

24% improved outcome
75% shorter hospital stay
85% decrease in cost

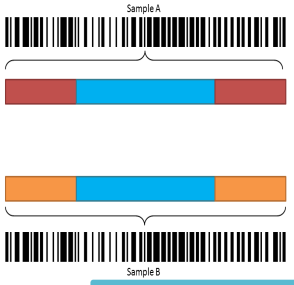
Based off first 42 babies seen

40% rapid diagnoses → 76% precision medicine → 77% improved outcome → 75% reduction in healthcare utilization
85% reduction in cost

Rady Children's Institute
Genomic Medicine

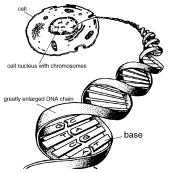
"Core" genomics workflow

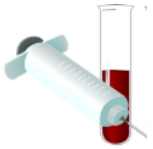
Illumina HiSeq



DNA library prep

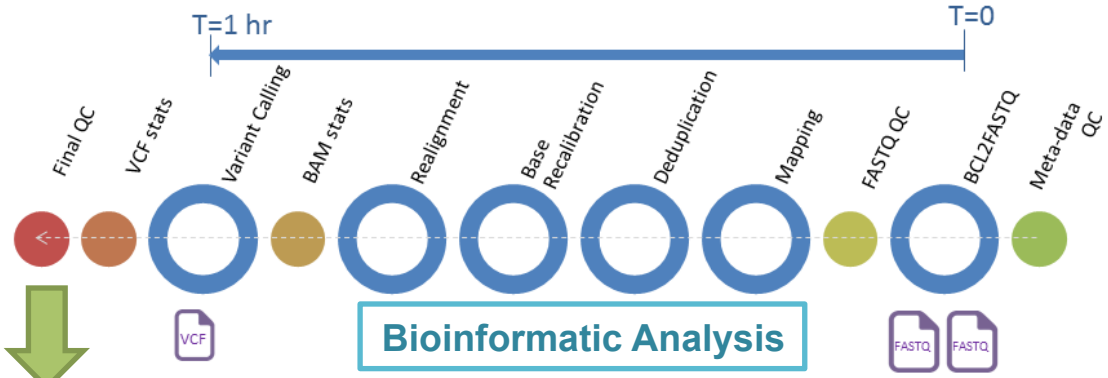
DNA extraction





Sample collection

Deep Phenotype Data

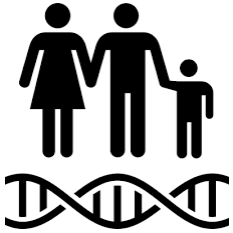


Bioinformatic Analysis

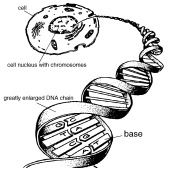
Variant interpretation & Clinical reporting

Biological

Patient enrollment



Sample collection



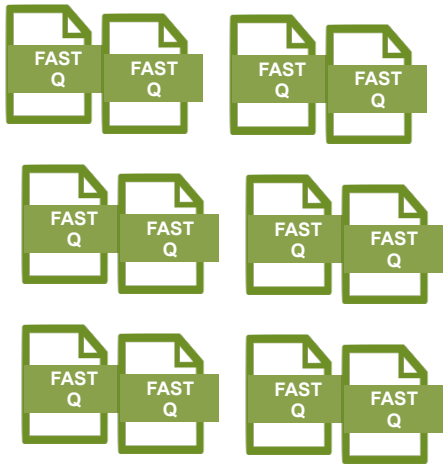
DNA extraction

Library prep

Genome sequencing

Digital

~1 million BCL files/ run (flowcell)



ASCII genome sequence

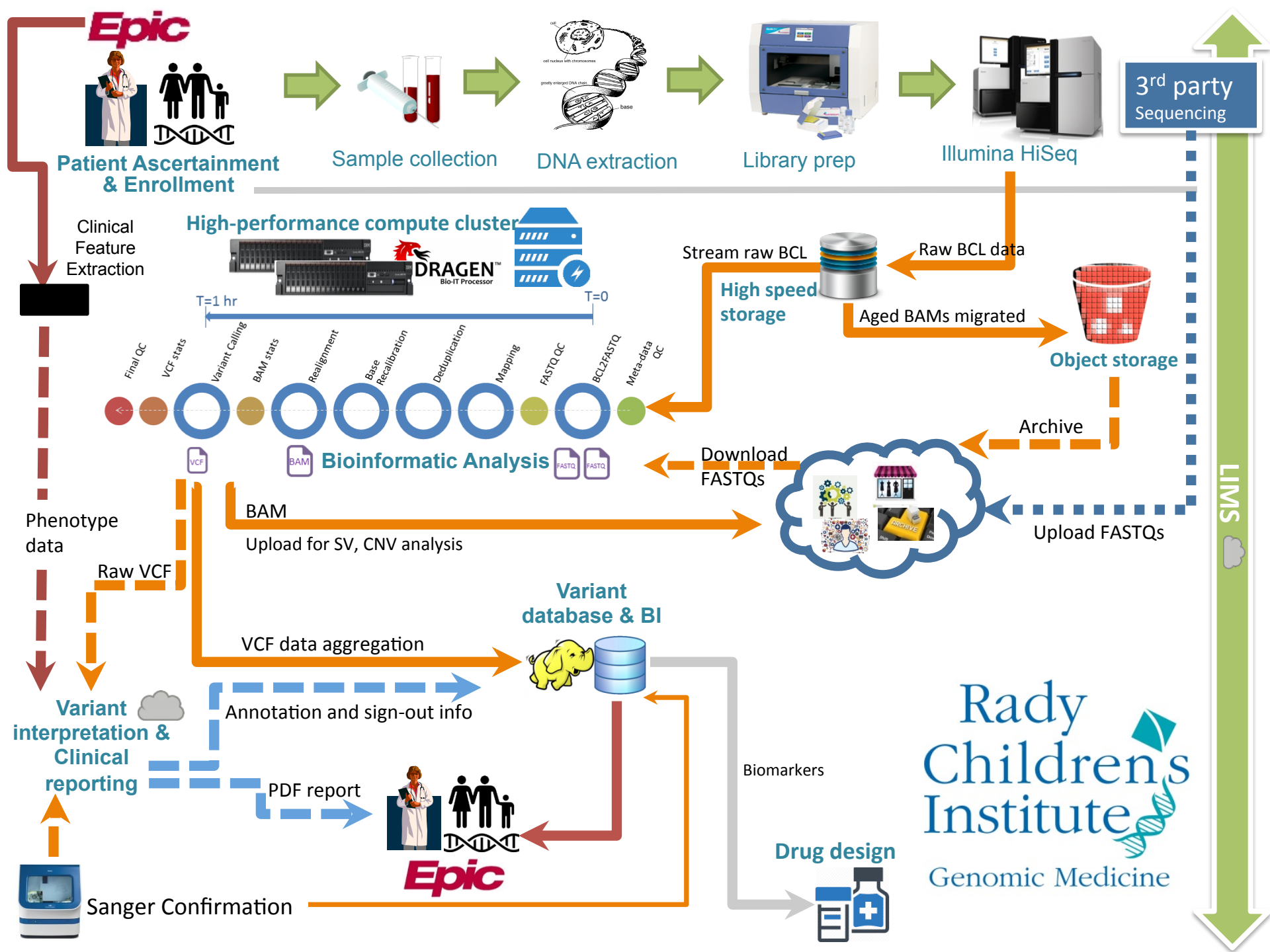
Binary genome alignments

ASCII Variant Call File

Clinical

Annotation, Prioritization and Interpretion

Genomics workflow


Epic
Patient Ascertainment & Enrollment
Sample collection
DNA extraction
Library prep
Illumina HiSeq
3rd party Sequencing
LIMS
Clinical Feature Extraction
High-performance compute cluster
DRAGEN™ Bio-iT Processor
T=1 hr
T=0
Final QC
VCF stats
Variant Calling
BAM stats
Realignment
Base Recalibration
Deduplication
Mapping
FASTQ QC
BCL2FASTQ
Meta-data QC
Bioinformatic Analysis
VCF
BAM
FASTQ
Stream raw BCL
Raw BCL data
High speed storage
Aged BAMs migrated
Object storage
Archive
Download FASTQs
Upload FASTQs
BAM
Upload for SV, CNV analysis
Phenotype data
Raw VCF
Variant database & BI
VCF data aggregation
Annotation and sign-out info
Variant interpretation & Clinical reporting
PDF report
Epic
Biomarkers
Drug design
Sanger Confirmation
Rady Children's Institute
Genomic Medicine

YOUR HEALTH IS PERSONAL SUMMER & FALL 2014

Genome

THE FAMILY CONDITION

HOW THE GENES YOU INHERIT AFFECT YOUR RISK FOR DISEASE

A GENOMIC REVOLUTION:

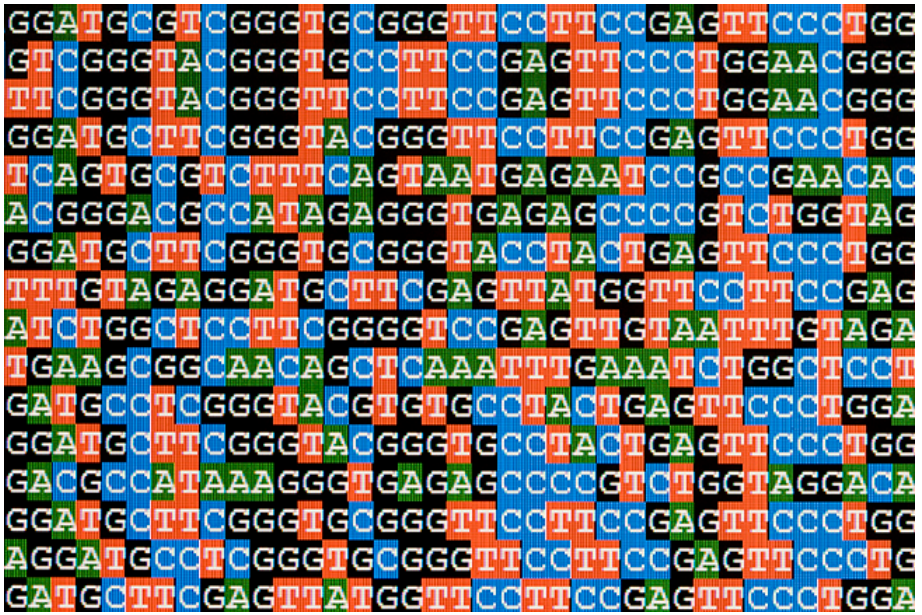
THE MAPPING OF THE HUMAN GENOME

THE SEQUENCING OF THE HUMAN GENOME WAS

THE LARGEST SINGLE UNDERTAKING IN THE HISTORY OF BIOLOGICAL SCIENCE.

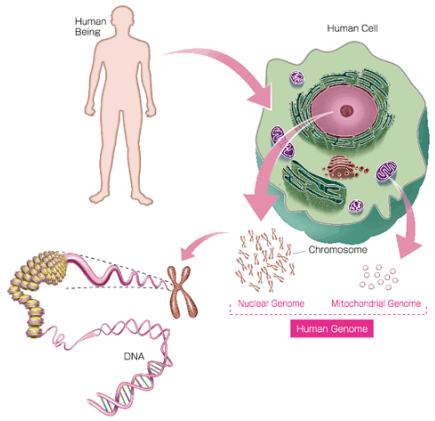
1990 ···· IT TOOK 13 YEARS ···· 2003

TO SEQUENCE 3 BILLION DNA BASE PAIRS AND IDENTIFY EACH HUMAN GENE.

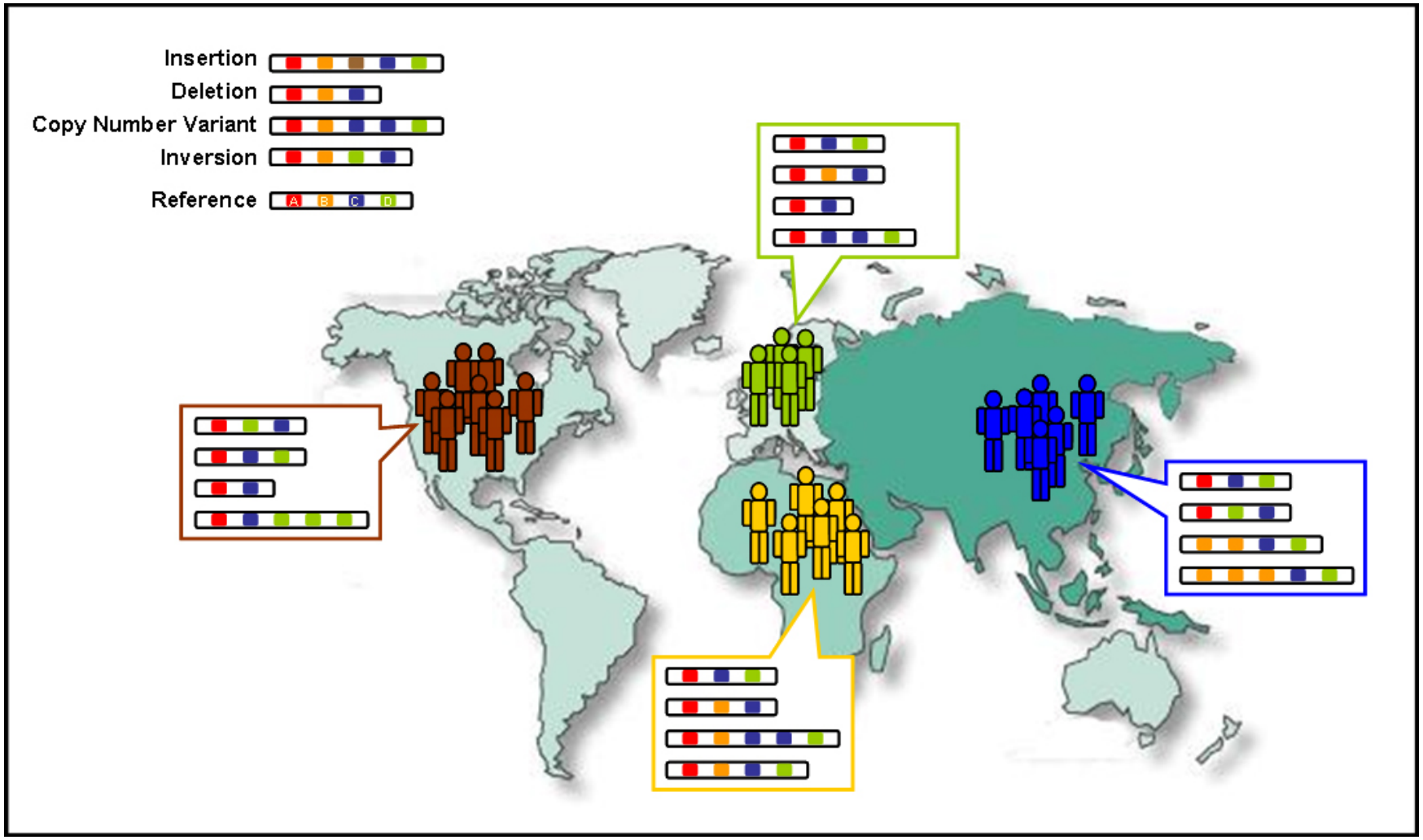


Only 10 years after the human genome was first mapped,

THE IMPACT IS ALREADY SIGNIFICANT AND WIDESPREAD. ···· 2013

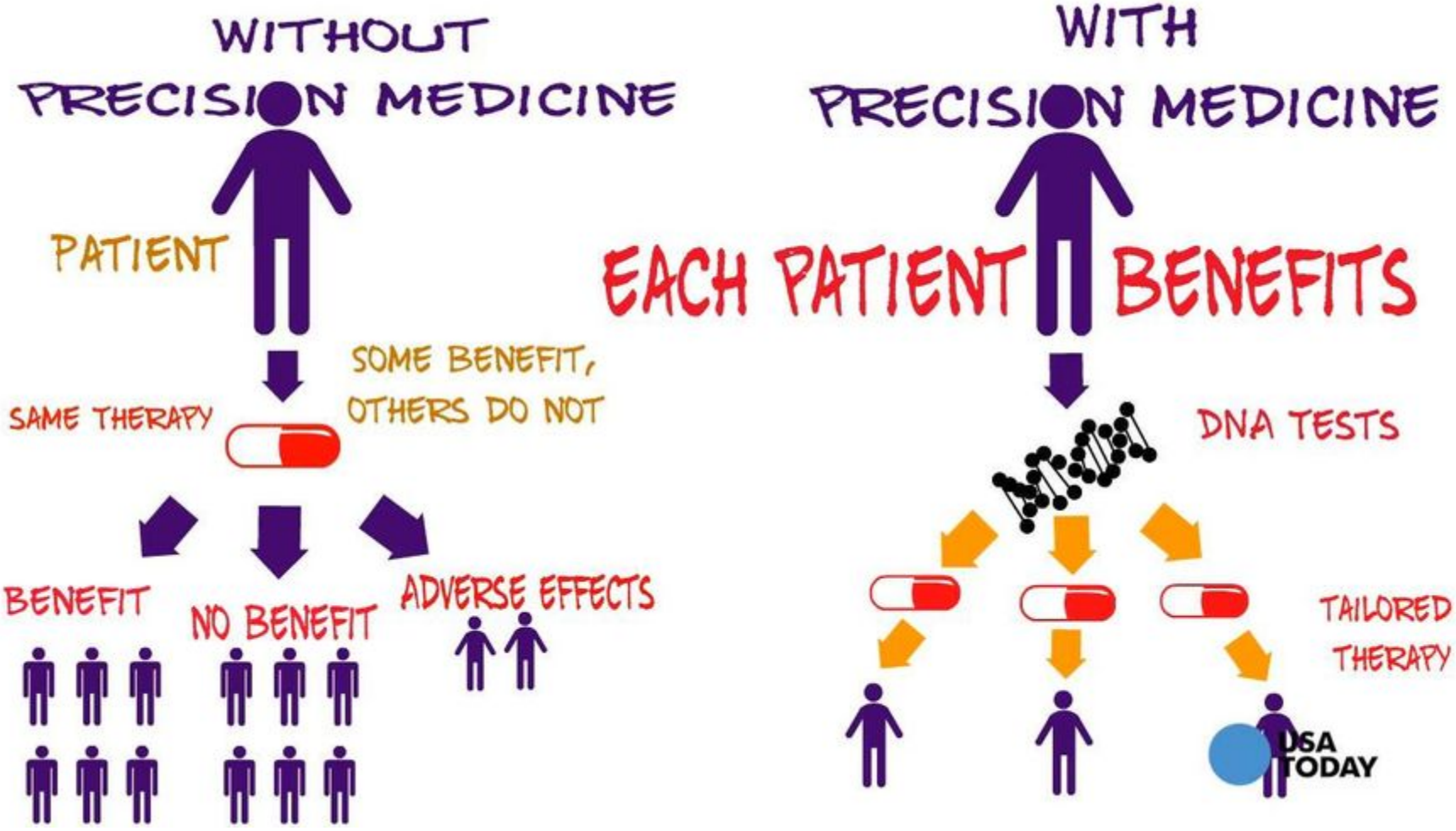


ZOOM


Insertion
Deletion
Copy Number Variant
Inversion
Reference
A
B
C
D

Kinds of data analytics

- Single patient level:
  - what genetic variation could be causative of XYZ clinical presentations
- Population level:
  - How frequent is variation XYZ in Hispanic population
  - What kinds of genetic variation do we typically observe for patients who have XYZ types of clinical presentations
  - Give me all patients with ABC phenotype and PQR genotype (genetic variation)
  - What genetic variation co-occur in LMN disease


WITHOUT
PRECISION MEDICINE
PATIENT
SAME THERAPY
SOME BENEFIT,
OTHERS DO NOT
BENEFIT
NO BENEFIT
ADVERSE EFFECTS
WITH
PRECISION MEDICINE
EACH PATIENT
BENEFITS
DNA TESTS
TAILORED
THERAPY
USA
TODAY

Alexis and Noah, born 1996

Colic ???

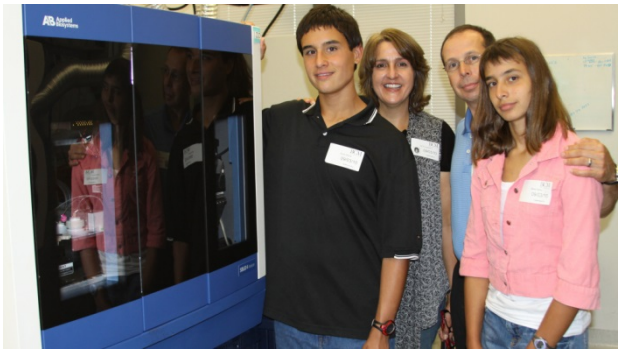
Twins wrongly diagnosed with cerebral palsy at age 2

2 yrs later, after multiple interventions

Low dopamine levels dopa-responsive dystonia

Google

Twins sleep through the night for the first time in their lives!



Beery twins, Age 14, 2010 @ Baylor College of Medicine, Houston, TX

https://www.youtube.com/watch?v=yUQFHecs8EQ

Rady Children's Hospital San Diego

Dr. Jennifer Friedman

9 yrs

2009, Cough + breathing problems

2011, Sequencing

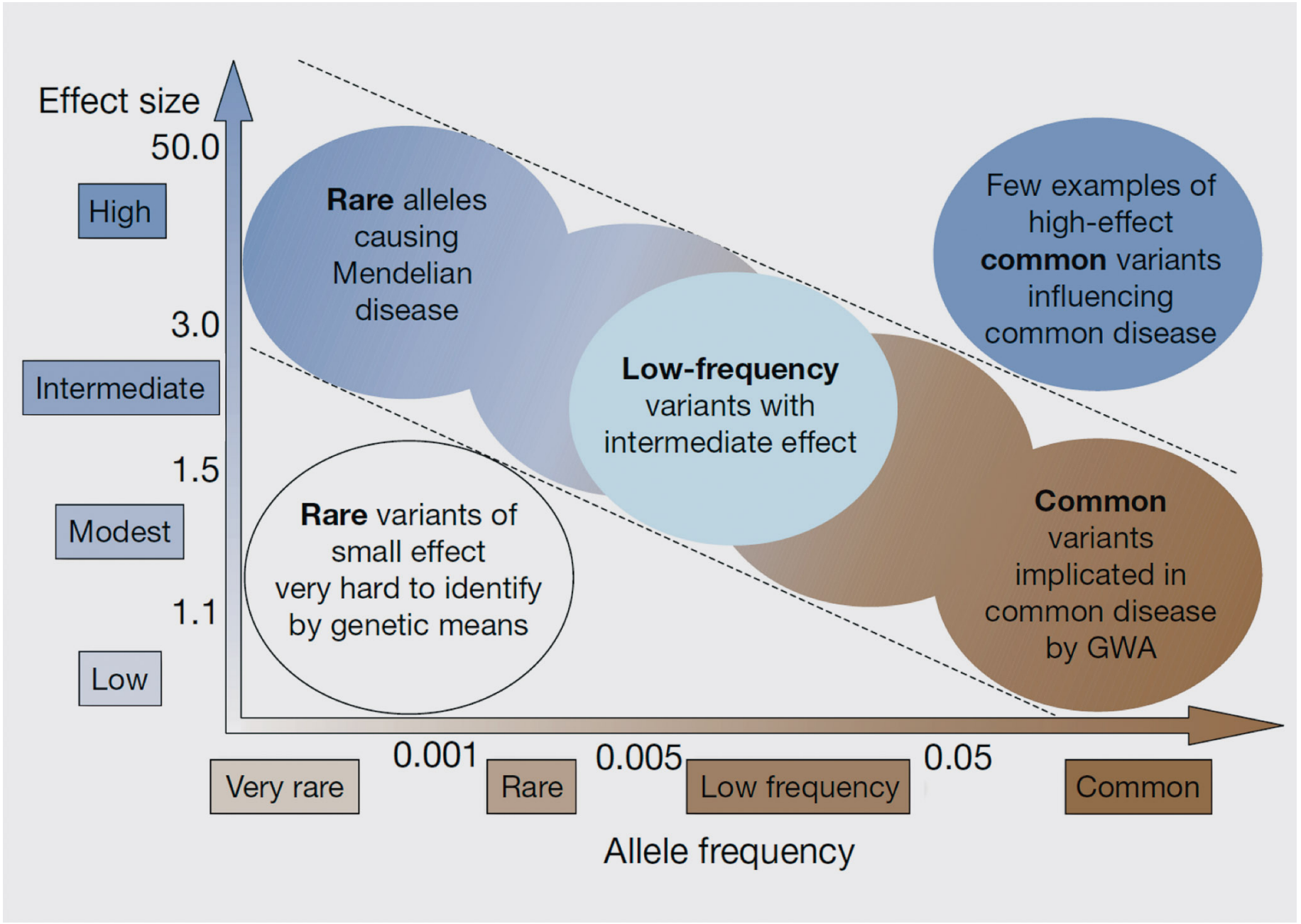
life technologies™

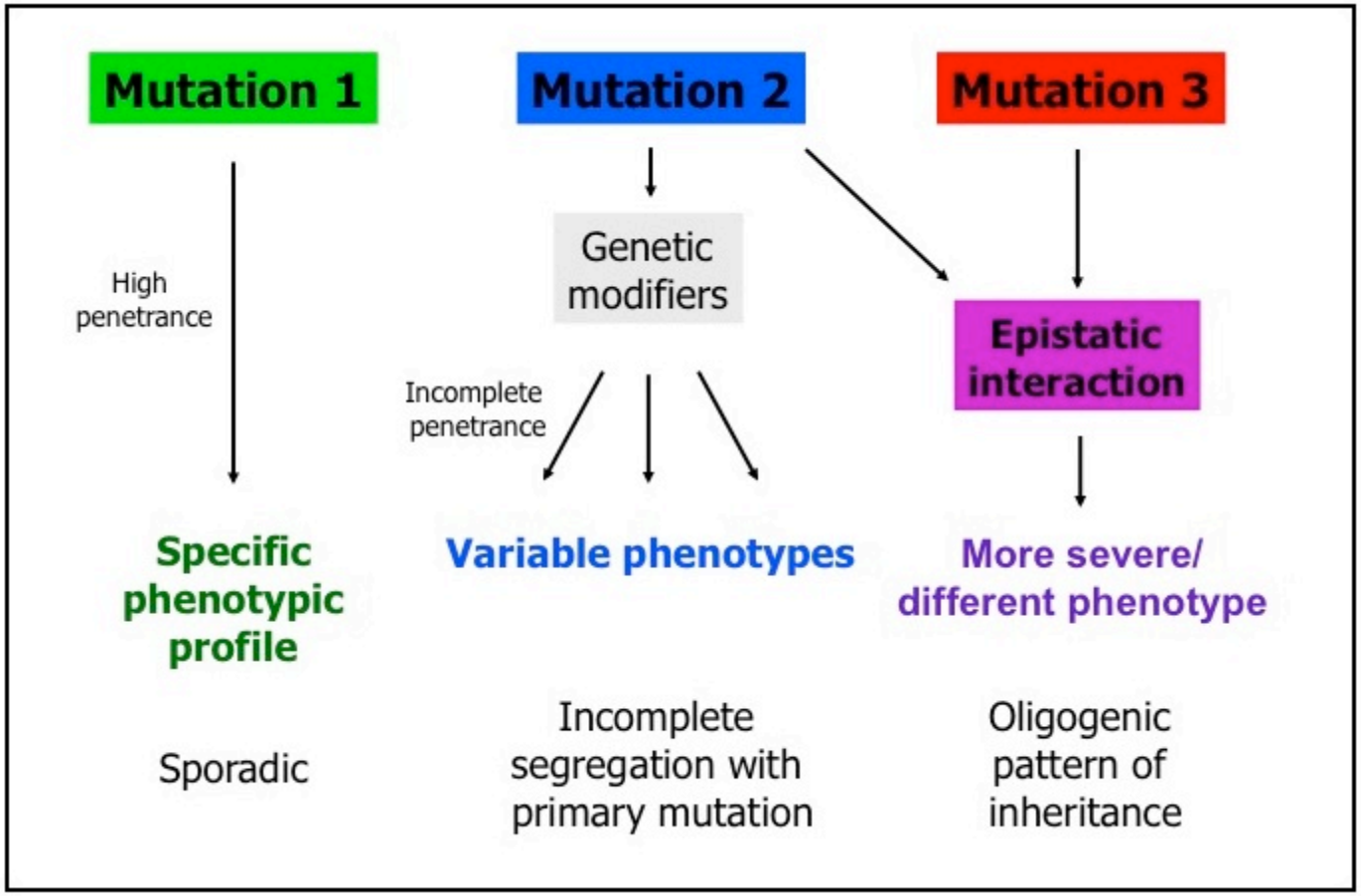
Mutations in a gene called *SPR* that encodes the enzyme sepiapterin reductase, which was previously linked to some cases of dopa-responsive dystonia

affect not only dopamine but also *serotonin* production

Usual suspects

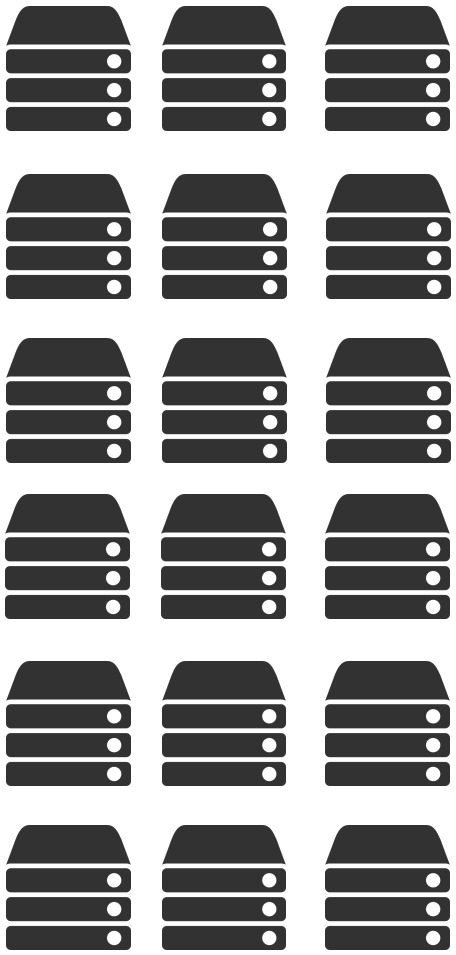
Rare Variants


Effect size
50.0
High
3.0
Intermediate
1.5
Modest
1.1
Low
Rare alleles causing Mendelian disease
Few examples of high-effect common variants influencing common disease
Low-frequency variants with intermediate effect
Rare variants of small effect very hard to identify by genetic means
Common variants implicated in common disease by GWA
0.001
0.005
0.05
Very rare
Rare
Low frequency
Common
Allele frequency

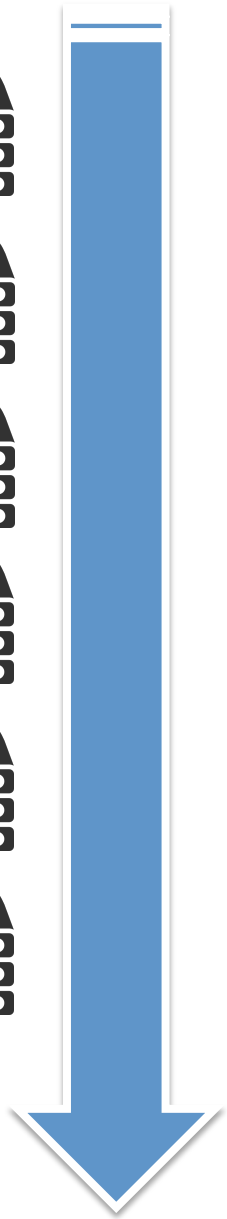

Mutation 1
Mutation 2
Mutation 3
High penetrance
Genetic modifiers
Incomplete penetrance
Epistatic interaction
Specific phenotypic profile
Variable phenotypes
More severe/ different phenotype
Sporadic
Incomplete segregation with primary mutation
Oligogenic pattern of inheritance

Common Bioinformatics Steps

Bioinformatics Pipeline



Compute cluster



Mapping
- **BWA**
- Alignment to GRCh37

BAM finishing
- **Sambamba**
- PCR de-duplication

Variant Calling
- **Platypus**
- SNV and INDEL

Annotation
- **VEP**
- dbSNP, HGMD, OMIM, etc

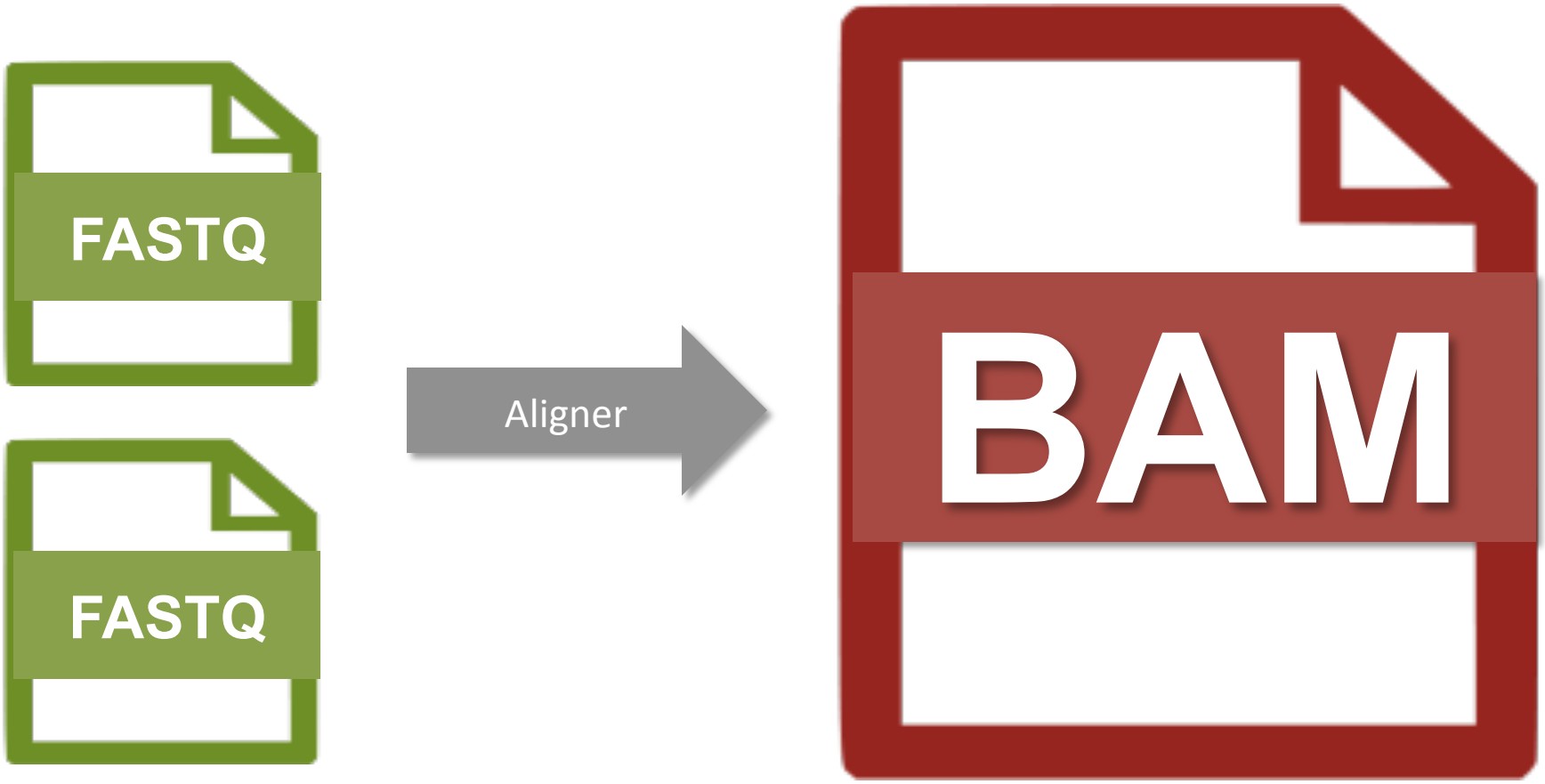
Variant Prioritization
- **Mendelian filtering**
- **Phenome + Variome**

SV analysis

Common Bioinformatics Steps

- Note
  - We will keep things simple
  - Many nuances and side paths, but lets not get distracted
  - We will not get into algorithms or underlying math
  - We will not be afraid of the command line

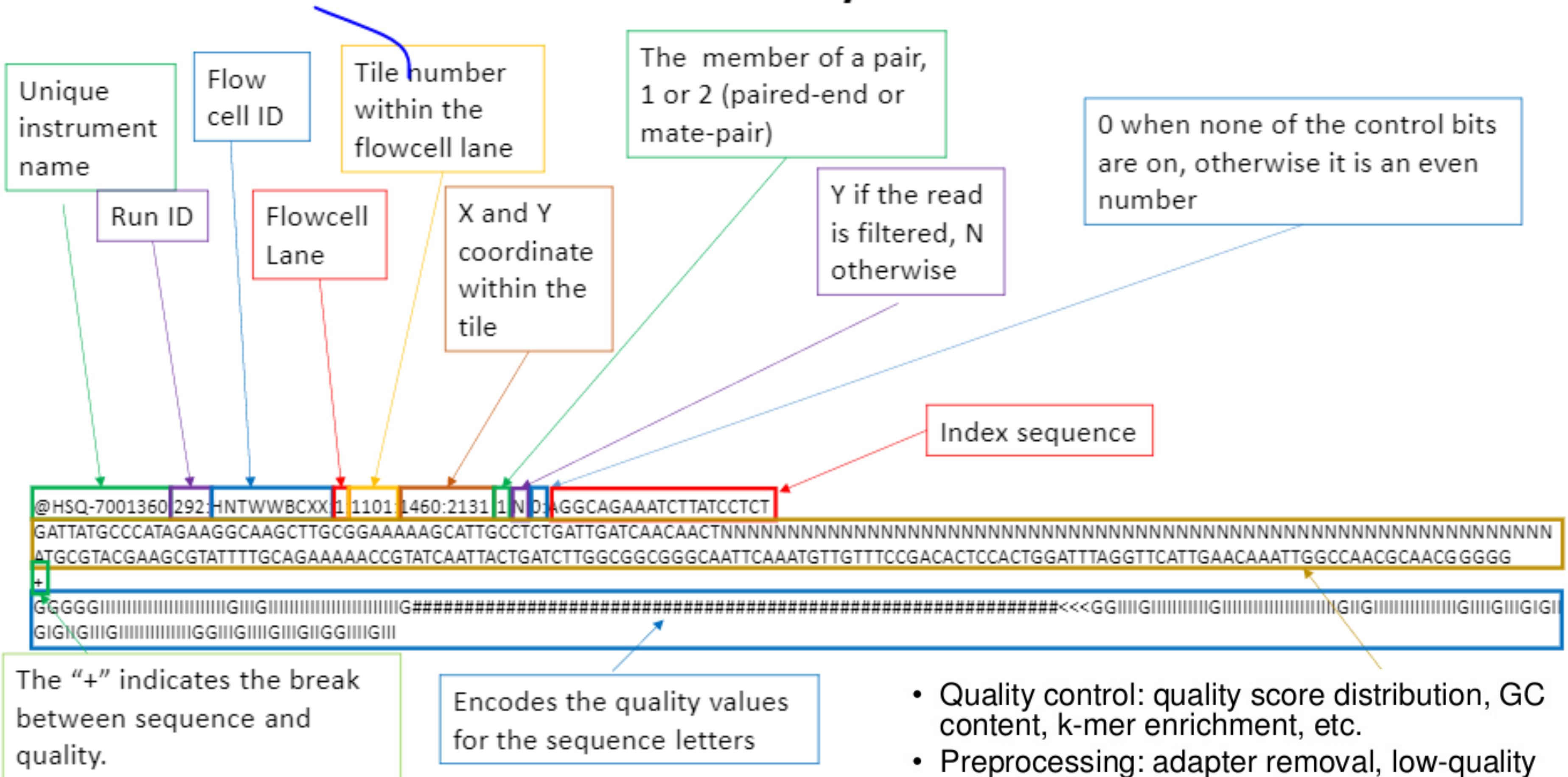
Mapping


FASTQ
FASTQ
Aligner
BAM

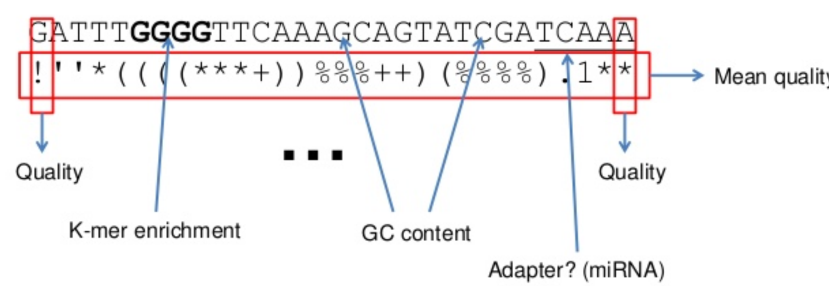
Many mapping tools out there....

- Bfast
- BioScope
- Bowtie
- BWA
- CLC bio
- CloudBurst
- Eland/Eland2
- GenomeMapper
- GnuMap
- Karma
- MAQ
- MOM
- Mosaik
- MrFAST/MrsFAST
- NovoAlign
- PASS
- PerM
- Phaster
- RazerS
- RMAP
- SSAHA2
- Segemehl
- SeqMap
- SHRiMP
- Slider/SliderII
- SOAP/SOAP2
- Srprism
- Stampy
- vmatch
- ZOOM
- .....

FASTQ File Format Analysis



- Quality control: quality score distribution, GC content, k-mer enrichment, etc.
- Preprocessing: adapter removal, low-quality reads filtering, etc.

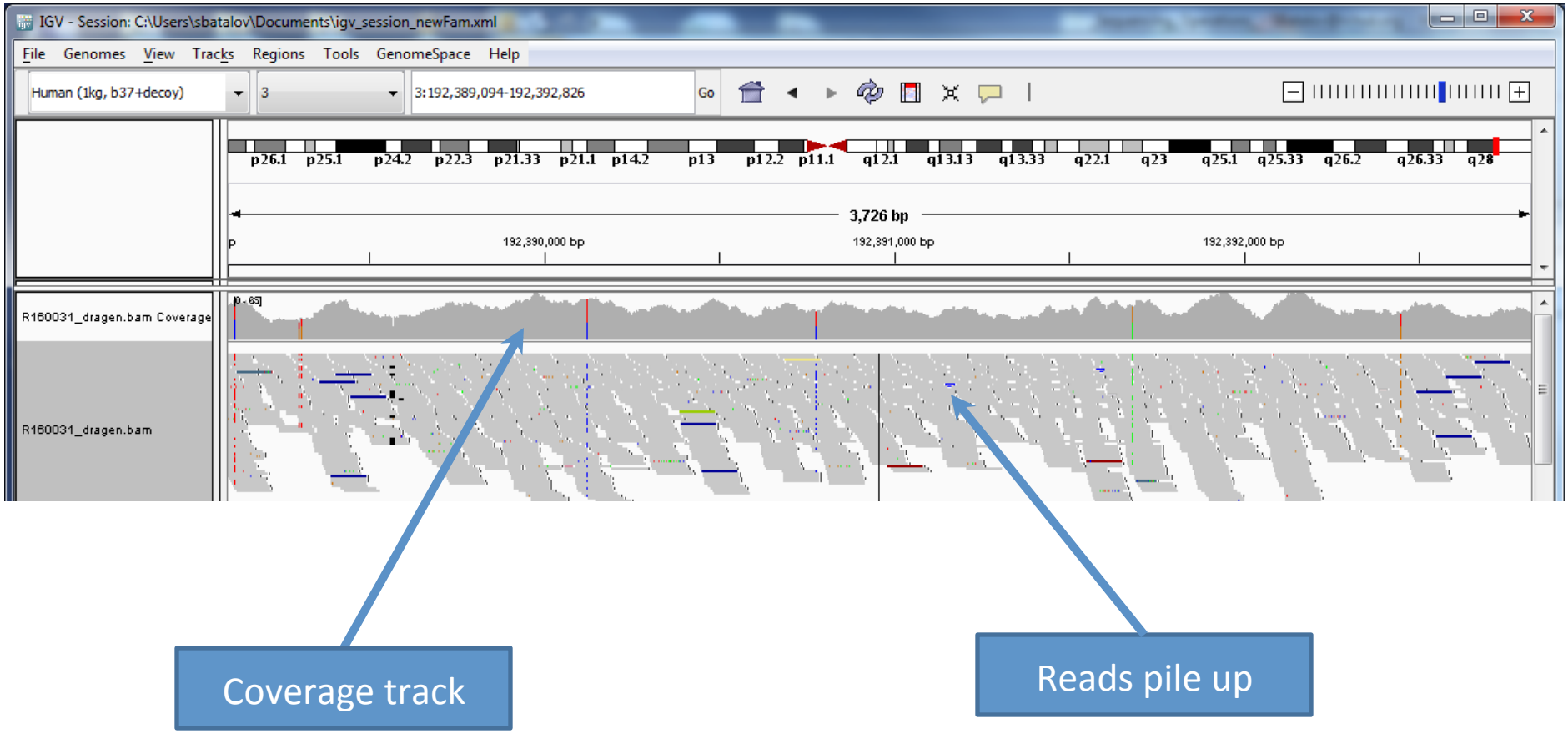


Here’s what a BAM looks like

Here's what a BAM really looks like

```
@HD     VN:1.0  SO:coordinate
@SQ     SN:chr20        LN:64444167
@PG     ID:TopHat       VN:2.0.14       CL:/srv/dna_tools/tophat/tophat -N 3 --read-edit-dist 5 --read-rea
lign-edit-dist 2 -i 50 -I 5000 --max-coverage-intron 5000 -M -o out /data/user446/mapping_tophat/index/chr
20 /data/user446/mapping_tophat/L6_18_GTGAAA_L007_R1_001.fastq
HWI-ST1145:74:C101DACXX:7:1102:4284:73714        16      chr20   190930  3       100M    *       0       0
       CCGTGTTTAAAGGTGGATGCGGTCACCTTCCCAGCTAGGCTTAGGGATTCTTAGTTGGCCTAGGAAATCCAGCTAGTCCTGTCTCTCAGTCCCCCCTCT
C    BBDCCDDCCDDDDCDDDDDDCDCCCDBC?DDDDDDDDDDDDDDDCCDCDDDDDDDDDDCCCCEDDDC?DDDDDDDDDDDDDDDDDDDDDBDHFFFFDC@@
    AS:i:-15        XM:i:3  XO:i:0  XG:i:0  MD:Z:55C20C13A9 NM:i:3  NH:i:2  CC:Z:=  CP:i:55352714   HI:i:0
HWI-ST1145:74:C101DACXX:7:1114:2759:41961        16      chr20   193953  50      100M    *       0       0
       TGCTGGATCATCTGGTTAGTGGCTTCTGACTCAGAGGACCTTCGTCCCCTGGGGCAGTGGACCTTCCAGTGATTCCCCTGACATAAGGGGCATGGACGA
G    DCDDDDEDDDDDDDCDDDDDDDCCCDDDCDDDDDEEC>DFFFEJJJJJIGJJJJIHGBHHGJIJJJJJJGJJJIJJJJJIHJJJJJJHHHHHFFFFFCCC
    AS:i:-16        XM:i:3  XO:i:0  XG:i:0  MD:Z:60G16T18T3 NM:i:3  NH:i:1
HWI-ST1145:74:C101DACXX:7:1204:14760:4030        16      chr20   270877  50      100M    *       0       0
       GGCTTTATTGGTAAAAAAGGAATAGCAGATTTAATCAGAAATTCCCACCTGGCCCAGCAGCACCAACCAGAAAGAAGGGAAGAAGACAGGAAAAAACCA
C    DDDDDDDDDCCDDDDDDDDDEEEEEEEFFFEFFEGHHHHFGDJJIHJJIJIJJJIIIIGGFJJIHIIIIJJJJJJIGHHFAHGFHJHFGGHFFFDD@BB
    AS:i:-11        XM:i:2  XO:i:0  XG:i:0  MD:Z:0A85G13    NM:i:2  NH:i:1
HWI-ST1145:74:C101DACXX:7:1210:11167:8699        0       chr20   271218  50      50M4700N50M     *       0
       0       GTGGCTCTTCCACAGGAATGTTGAGGATGACATCCATGTCTGGGGTGCACTTGGGTCTCCGAAGCAGAACATCCTCAAATATGACCTCTCG
accepted_hits.sam
```

Here's how a BAM can be better visualized



http://software.broadinstitute.org/software/igv/book/export/html/6

BRIEFINGS IN BIOINFORMATICS. VOL 11. NO 5. 473–483
Advance Access published on 11 May 2010
doi:10.1093/bib/bbq015

A survey of sequence alignment algorithms for next-generation sequencing

Heng Li and Nils Homer

Submitted: 3rd March 2010; Received (in revised form): 14th April 2010

Abstract

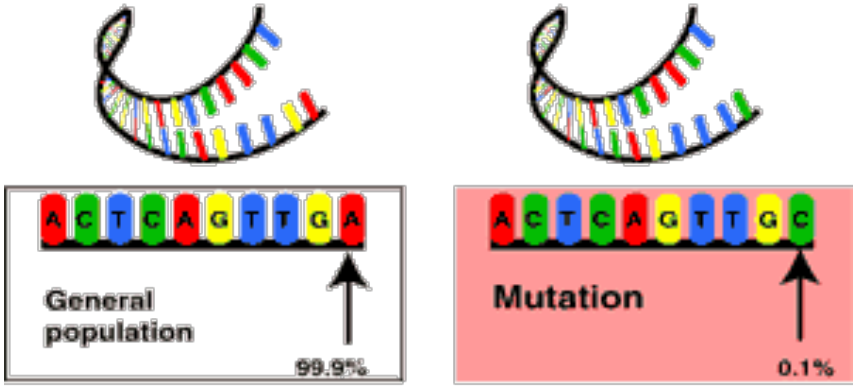
Rapidly evolving sequencing technologies produce data on an unparalleled scale. A central challenge to the analysis of this data is sequence alignment, whereby sequence reads must be compared to a reference. A wide variety of alignment algorithms and software have been subsequently developed over the past two years. In this article, we will systematically review the current development of these algorithms and introduce their practical applications on different types of experimental data. We come to the conclusion that short-read alignment is no longer the bottleneck of data analyses. We also consider future development of alignment algorithms with respect to emerging long sequence reads and the prospect of cloud computing.

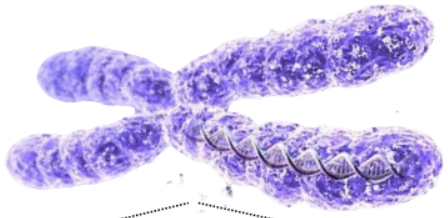
Variant Calling

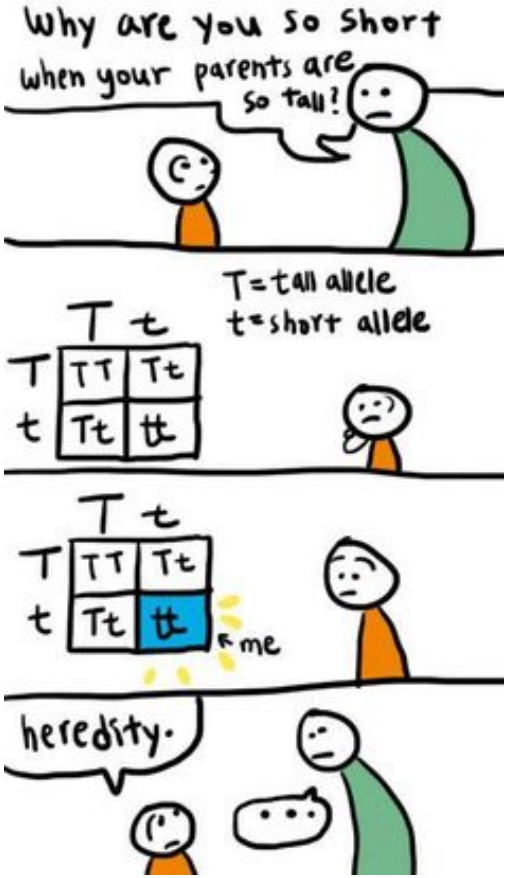

BAM
Variant Caller
VCF

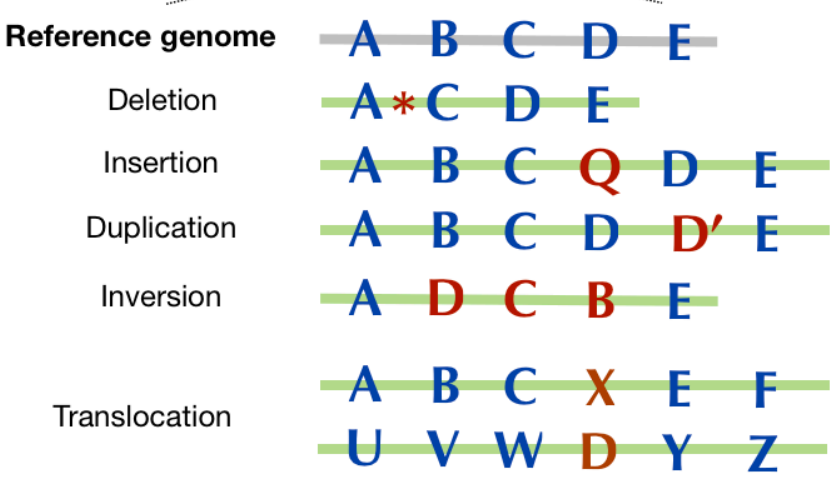
Many Variant Calling tools out there….

- GATK
- Platypus
- Varscan2
- Atlas2
- Senteion
- FreeBayes
- SNVer
- SAMtools
- VarDict
- LoFreq
- …….


A C T C A G T T G A
General population
99.9%
A C T C A G T T G C
Mutation
0.1%




Why are you so short
when your parents are so tall?
T=tall allele
t=short allele
T t
T TT Tt
t Tt tt
T t
T TT Tt
t Tt tt
me
heredity.


Reference genome
A B C D E
Deletion
A * C D E
Insertion
A B C Q D E
Duplication
A B C D D′ E
Inversion
A D C B E
Translocation
A B C X E F
U V W D Y Z

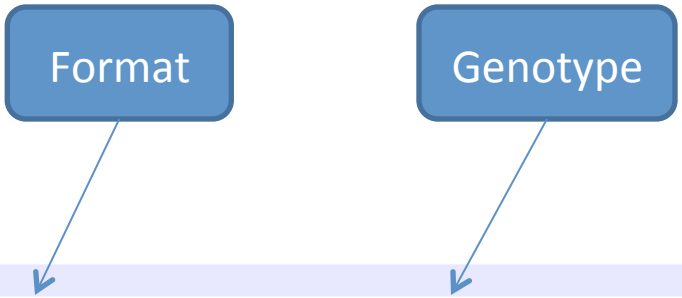
Some of the genomic artifacts that influence variant calling

- Coverage at that particular base
- Coverage around that particular base
- Base quality at that particular base
- Base quality around that particular base
- Strandedness – is there enough support in the "forward" and "reverse" strands
- Variant ratio = # of variant reads/# of reference reads

Here's what a VCF file looks like

```
##fileformat=VCFv4.0
##fileDate=2017-09-03
##source=Platypus_Version_0.8.1
##platypusOptions={'assemblyRegionSize': 1500, 'trimReadFlank': 0, 'assembleBadReads': 1, 'bamFiles': ['NA24385_bw
##INFO=<ID=FR,Number=.,Type=Float,Description="Estimated population frequency of variant">
##INFO=<ID=MMLQ,Number=1,Type=Float,Description="Median minimum base quality for bases around variant">
##INFO=<ID=TCR,Number=1,Type=Integer,Description="Total reverse strand coverage at this locus">
##INFO=<ID=HP,Number=1,Type=Integer,Description="Homopolymer run length around variant locus">
##INFO=<ID=WE,Number=1,Type=Integer,Description="End position of calling window">
##INFO=<ID=Source,Number=.,Type=String,Description="Was this variant suggested by Playtypus, Assembler, or from a
##INFO=<ID=FS,Number=.,Type=Float,Description="Fisher's exact test for strand bias (Phred scale)">
##INFO=<ID=WS,Number=1,Type=Integer,Description="Starting position of calling window">
##INFO=<ID=PP,Number=.,Type=Float,Description="Posterior probability (phred scaled) that this variant segregates">
##INFO=<ID=TR,Number=.,Type=Integer,Description="Total number of reads containing this variant">
##INFO=<ID=NF,Number=.,Type=Integer,Description="Total number of forward reads containing this variant">
##INFO=<ID=TCF,Number=1,Type=Integer,Description="Total forward strand coverage at this locus">
##INFO=<ID=NR,Number=.,Type=Integer,Description="Total number of reverse reads containing this variant">
##INFO=<ID=TC,Number=1,Type=Integer,Description="Total coverage at this locus">
##INFO=<ID=END,Number=.,Type=Integer,Description="End position of reference call block">
##INFO=<ID=MGOF,Number=.,Type=Integer,Description="Worst goodness-of-fit value reported across all samples">
##INFO=<ID=SbPval,Number=.,Type=Float,Description="Binomial P-value for strand bias test">
##INFO=<ID=START,Number=.,Type=Integer,Description="Start position of reference call block">
##INFO=<ID=ReadPosRankSum,Number=.,Type=Float,Description="Mann-Whitney Rank sum test for difference between in po
##INFO=<ID=MQ,Number=.,Type=Float,Description="Root mean square of mapping qualities of reads at the variant posit
##INFO=<ID=QD,Number=1,Type=Float,Description="Variant-quality/read-depth for this variant">
##INFO=<ID=SC,Number=1,Type=String,Description="Genomic sequence 10 bases either side of variant position">
##INFO=<ID=BRF,Number=1,Type=Float,Description="Fraction of reads around this variant that failed filters">
##INFO=<ID=HapScore,Number=.,Type=Integer,Description="Haplotype score measuring the number of haplotypes the vari
##INFO=<ID=Size,Number=.,Type=Integer,Description="Size of reference call block">
##FILTER=<ID=GOF,Description="Variant fails goodness-of-fit test.">
##FILTER=<ID=badReads,Description="Variant supported only by reads with low quality bases close to variant positio
##FILTER=<ID=alleleBias,Description="Variant frequency is lower than expected for het">
##FILTER=<ID=hp10,Description="Flanking sequence contains homopolymer of length 10 or greater">
##FILTER=<ID=Q20,Description="Variant quality is below 20.">
##FILTER=<ID=HapScore,Description="Too many haplotypes are supported by the data in this region.">
##FILTER=<ID=MQ,Description="Root-mean-square mapping quality across calling region is low.">
##FILTER=<ID=strandBias,Description="Variant fails strand-bias filter">
```

```
##FILTER=<ID=MQ,Description="Root-mean-square mapping quality across calling region is low.">
##FILTER=<ID=strandBias,Description="Variant fails strand-bias filter">
##FILTER=<ID=SC,Description="Variants fail sequence-context filter. Surrounding sequence is low-complexity">
##FILTER=<ID=QualDepth,Description="Variant quality/Read depth ratio is low.">
##FILTER=<ID=REFCALL,Description="This line represents a homozygous reference call">
##FILTER=<ID=QD,Description="Variants fail quality/depth filter.">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Unphased genotypes">
##FORMAT=<ID=GQ,Number=.,Type=Integer,Description="Genotype quality as phred score">
##FORMAT=<ID=GOF,Number=.,Type=Float,Description="Goodness of fit value">
##FORMAT=<ID=NR,Number=.,Type=Integer,Description="Number of reads covering variant location in this sample">
##FORMAT=<ID=GL,Number=.,Type=Float,Description="Genotype log10-likelihoods for AA,AB and BB genotypes, where A =
##FORMAT=<ID=NV,Number=.,Type=Integer,Description="Number of reads containing variant in this sample">
#CHROM  POS ID  REF ALT QUAL    FILTER  INFO    FORMAT  NA24385
1   10439   .   AC  A   68  PASS    BRF=0.0;FR=0.9998;HP=4;HapScore=2;MGOF=39;MMLQ=38;MQ=46.0;NF=1;NR=1;PP=68;QD=3
1   846338  .   A   G   44  PASS    BRF=0.0;FR=0.9998;HP=2;HapScore=2;MGOF=0;MMLQ=39;MQ=60.0;NF=1;NR=1;PP=44;QD=39
1   866319  .   G   A   45  PASS    BRF=0.0;FR=0.9998;HP=3;HapScore=2;MGOF=0;MMLQ=39;MQ=60.0;NF=2;NR=0;PP=45;QD=39
1   871334  .   G   T   314 PASS    BRF=0.0;FR=1.0000;HP=1;HapScore=2;MGOF=0;MMLQ=37;MQ=60.0;NF=2;NR=7;PP=314;QD=3
1   876499  .   A   G   168 PASS    BRF=0.0;FR=1.0000;HP=6;HapScore=2;MGOF=0;MMLQ=37;MQ=60.0;NF=5;NR=1;PP=168;QD=3
1   877831  .   T   C   346 PASS    BRF=0.17;FR=1.0000;HP=2;HapScore=2;MGOF=3;MMLQ=35;MQ=60.0;NF=9;NR=1;PP=346;QD=
1   878314  .   G   C   342 PASS    BRF=0.0;FR=0.5000;HP=2;HapScore=2;MGOF=2;MMLQ=37;MQ=60.0;NF=7;NR=6;PP=342;QD=2
1   879676  .   G   A   517 PASS    BRF=0.15;FR=1.0000;HP=3;HapScore=1;MGOF=9;MMLQ=38;MQ=60.0;NF=7;NR=7;PP=517;QD=
1   879687  .   T   C   540 PASS    BRF=0.15;FR=1.0000;HP=3;HapScore=1;MGOF=9;MMLQ=38;MQ=60.0;NF=7;NR=8;PP=540;QD=
1   880238  .   A   G   351 PASS    BRF=0.0;FR=1.0000;HP=2;HapScore=2;MGOF=0;MMLQ=39;MQ=60.0;NF=3;NR=7;PP=351;QD=3
1   881627  .   G   A   463 PASS    BRF=0.13;FR=1.0000;HP=3;HapScore=2;MGOF=3;MMLQ=38;MQ=60.0;NF=6;NR=7;PP=463;QD=
1   883625  .   A   G   45  PASS    BRF=0.5;FR=0.9998;HP=3;HapScore=2;MGOF=0;MMLQ=39;MQ=60.0;NF=1;NR=1;PP=45;QD=39
1   887560  .   A   C   379 PASS    BRF=0.0;FR=1.0000;HP=1;HapScore=2;MGOF=1;MMLQ=39;MQ=60.0;NF=2;NR=9;PP=379;QD=3
1   887801  .   A   G   697 PASS    BRF=0.05;FR=1.0000;HP=2;HapScore=1;MGOF=0;MMLQ=39;MQ=58.63;NF=13;NR=5;PP=697;Q
1   888639  .   T   C   273 PASS    BRF=0.31;FR=1.0000;HP=1;HapScore=2;MGOF=17;MMLQ=38;MQ=50.09;NF=3;NR=5;PP=273;Q
1   888659  .   T   C   120 PASS    BRF=0.2;FR=1.0000;HP=1;HapScore=2;MGOF=23;MMLQ=34;MQ=47.33;NF=1;NR=3;PP=120;QD
1   889158  .   GA  CC  571 PASS    BRF=0.25;FR=1.0000;HP=2;HapScore=1;MGOF=8;MMLQ=39;MQ=60.0;NF=5;NR=3;PP=571;QD=
1   892745  .   G   A   159 PASS    BRF=0.0;FR=1.0000;HP=2;HapScore=2;MGOF=0;MMLQ=38;MQ=60.0;NF=0;NR=5;PP=159;QD=3
1   894573  .   G   A   433 PASS    BRF=0.0;FR=1.0000;HP=1;HapScore=2;MGOF=0;MMLQ=36;MQ=60.0;NF=8;NR=5;PP=433;QD=3
1   896271  .   C   T   337 PASS    BRF=0.13;FR=1.0000;HP=2;HapScore=2;MGOF=0;MMLQ=36;MQ=60.0;NF=3;NR=7;PP=337;QD=
1   897325  .   G   C   691 PASS    BRF=0.09;FR=1.0000;HP=2;HapScore=1;MGOF=1;MMLQ=38;MQ=60.0;NF=9;NR=10;PP=691;QD
1   897564  .   T   C   122 PASS    BRF=0.0;FR=1.0000;HP=2;HapScore=2;MGOF=0;MMLQ=39;MQ=60.0;NF=4;NR=0;PP=122;QD=3
1   898323  .   T   C   506 PASS    BRF=0.24;FR=1.0000;HP=1;HapScore=1;MGOF=0;MMLQ=36;MQ=60.0;NF=3;NR=11;PP=506;QD
1   899928  .   GGGGAGGGGGGCGCG CGGGAGGGGTCCGCA 164 PASS    BRF=0.0;FR=0.9998;HP=3;HapScore=3;MGOF=0;MMLQ=39;MQ=60
```



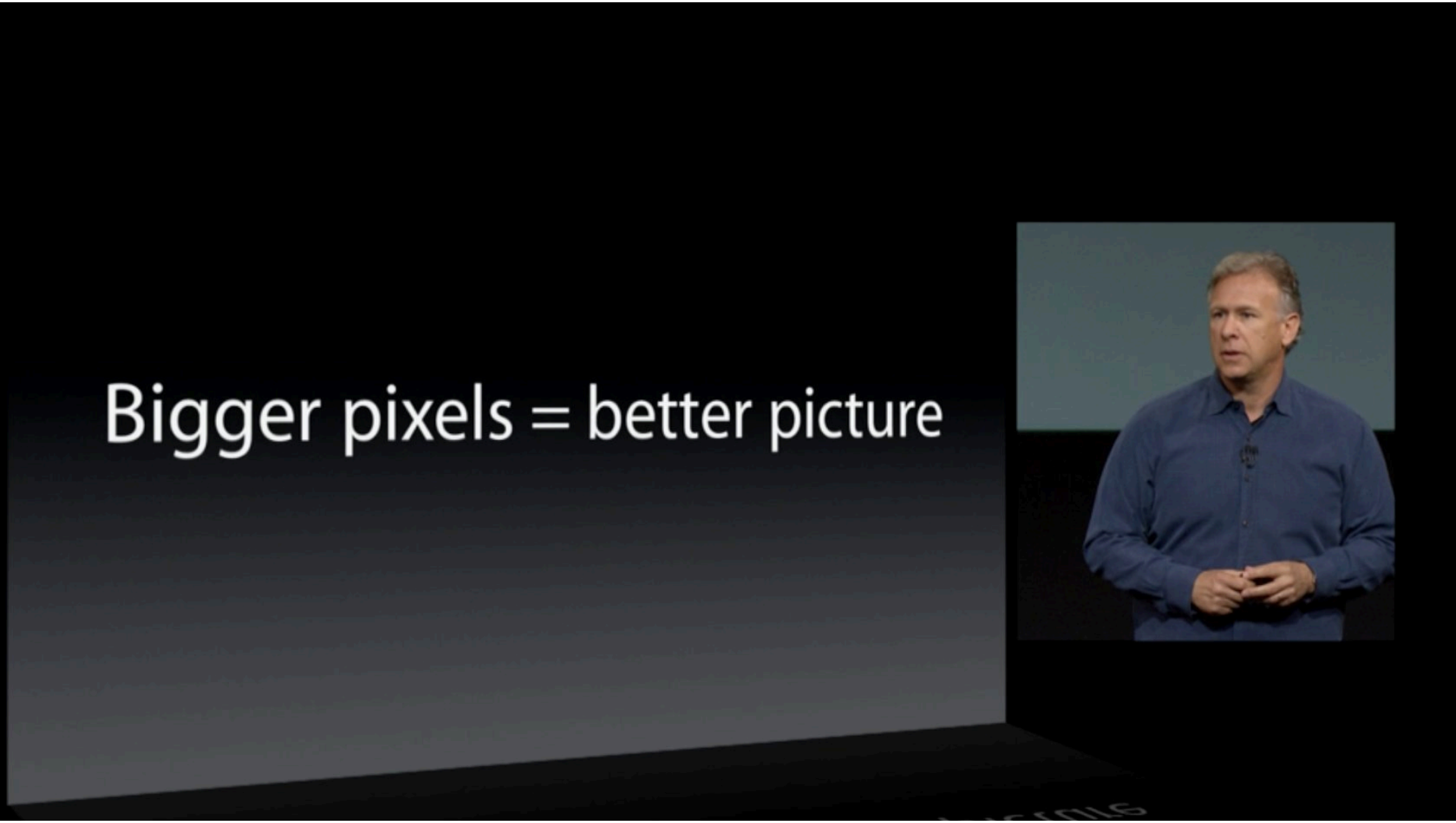

```
TCF=1;TCR=1;TR=2;WE=10447;WS=10429  GT:GL:GOF:GQ:NR:NV  1/1:-7.8,-0.3,0.0:39:5:2:2
s;TC=2;TCF=1;TCR=1;TR=2;WE=846346;WS=846328    GT:GL:GOF:GQ:NR:NV  1/1:-7.9,-0.3,0.0:0:5:2:2
;TC=2;TCF=2;TCR=0;TR=2;WE=866327;WS=866309 GT:GL:GOF:GQ:NR:NV  1/1:-8.0,-0.3,0.0:0:5:2:2
us;TC=9;TCF=2;TCR=7;TR=9;WE=871342;WS=871324    GT:GL:GOF:GQ:NR:NV  1/1:-34.9,-2.41,0.0:0:24:9:9
us;TC=6;TCF=5;TCR=1;TR=6;WE=876507;WS=876489    GT:GL:GOF:GQ:NR:NV  1/1:-20.3,-1.51,0.0:0:15:6:6
ypus;TC=10;TCF=9;TCR=1;TR=10;WE=877839;WS=877821    GT:GL:GOF:GQ:NR:NV  1/1:-38.1,-2.71,0.0:3:27:10:10
us;TC=38;TCF=19;TCR=19;TR=13;WE=878322;WS=878304    GT:GL:GOF:GQ:NR:NV  1/0:-38.26,0.0,-81.46:2:99:38:13
ypus;TC=14;TCF=7;TCR=7;TR=14;WE=879695;WS=879666    GT:GL:GOF:GQ:NR:NV  1/1:-55.0,-3.74,0.0:9:37:14:14
ypus;TC=15;TCF=7;TCR=8;TR=15;WE=879695;WS=879666    GT:GL:GOF:GQ:NR:NV  1/1:-56.9,-3.91,0.0:9:39:15:15
us;TC=10;TCF=3;TCR=7;TR=10;WE=880246;WS=880228 GT:GL:GOF:GQ:NR:NV  1/1:-38.6,-2.71,0.0:0:27:10:10
us;TC=13;TCF=6;TCR=7;TR=13;WE=881635;WS=881617 GT:GL:GOF:GQ:NR:NV  1/1:-49.8,-3.61,0.0:3:36:13:13
s;TC=2;TCF=1;TCR=1;TR=2;WE=883633;WS=883615    GT:GL:GOF:GQ:NR:NV  1/1:-8.0,-0.3,0.0:0:5:2:2
s;TC=11;TCF=2;TCR=9;TR=11;WE=887568;WS=887550  GT:GL:GOF:GQ:NR:NV  1/1:-41.4,-3.01,0.0:1:30:11:11
typus;TC=18;TCF=13;TCR=5;TR=18;WE=887809;WS=887791 GT:GL:GOF:GQ:NR:NV  1/1:-73.2,-5.42,0.0:0:54:18:18
typus;TC=8;TCF=3;TCR=5;TR=8;WE=888647;WS=888629    GT:GL:GOF:GQ:NR:NV  1/1:-30.8,-2.11,0.0:17:21:8:8
ypus;TC=4;TCF=1;TCR=3;TR=4;WE=888667;WS=888649 GT:GL:GOF:GQ:NR:NV  1/1:-15.5,-0.9,0.0:23:10:4:4
ypus;TC=9;TCF=6;TCR=3;TR=8;WE=889167;WS=889148  GT:GL:GOF:GQ:NR:NV  1/1:-62.4,-2.11,0.0:8:21:9:8
s;TC=5;TCF=0;TCR=5;TR=5;WE=892753;WS=892735    GT:GL:GOF:GQ:NR:NV  1/1:-19.4,-1.2,0.0:0:12:5:5
us;TC=13;TCF=8;TCR=5;TR=13;WE=894581;WS=894563 GT:GL:GOF:GQ:NR:NV  1/1:-46.8,-3.61,0.0:0:36:13:13
ypus;TC=10;TCF=3;TCR=7;TR=10;WE=896279;WS=896261    GT:GL:GOF:GQ:NR:NV  1/1:-37.2,-2.71,0.0:0:27:10:10
ypus;TC=19;TCF=9;TCR=10;TR=19;WE=897333;WS=897315  GT:GL:GOF:GQ:NR:NV  1/1:-72.6,-5.42,0.0:1:54:19:19
s;TC=4;TCF=4;TCR=0;TR=4;WE=897572;WS=897554    GT:GL:GOF:GQ:NR:NV  1/1:-15.7,-0.9,0.0:0:10:4:4
ypus;TC=14;TCF=3;TCR=11;TR=14;WE=898331;WS=898313  GT:GL:GOF:GQ:NR:NV  1/1:-54.1,-3.91,0.0:0:39:14:14
SbPval=1.0;Source=Platypus;TC=5;TCF=0;TCR=5;TR=2;WE=899950;WS=899918    GT:GL:GOF:GQ:NR:NV  1/1:-42.5,-1.2,0.0:0:12:5:2
```

Now we have the capability to evaluate the quality of sequencing and bioinformatic analyses

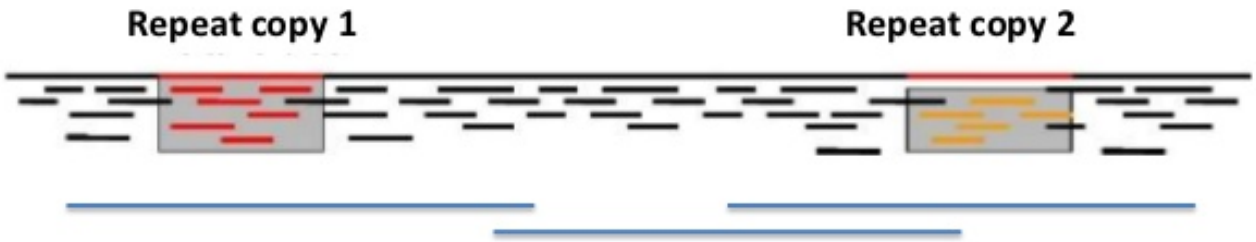
Utah trio (NA12878) Bioinformatics

CORIELL TRIO 1 (UTAH TRIO)	NA12878-proband	NA12891-dad	NA12892-mom	UNITS	REFERENCE RANGE
Sex	**F verified**	**M verified**	**F verified**		**trio order verified**
Yield: raw/bulk	**201.5**	**141.1**	**161.4**	Gbp	> 180
% mapped	**99.24%**	**99.11%**	**99.15%**	pct	98-100
% duplicates	**8.82%**	**7.83%**	**8.17%**	pct	< 15%
Yield	**181.9**	**128.8**	**146.8**	Gbp	> 130
Insert size: Mean +/- std.dev	**388.2 +/- 84.1**	**400.0 +/- 87.8**	**393.3 +/- 85.3**	bp	300-480
Average and median coverage across genome	**55.0**	**39.0**	**43.0**	x	> 40
Average coverage over OMIM genes	**59.0**	**42.0**	**47.0**	x	> 40
# of OMIM genes with coverage at <10X (and list)	**249**	359	**260**	ENST	< 2% (282)
# of OMIM genes with 100% coverage at >=10X	**98.2%**	**97.5%**	**98.2%**	pct	> 98%
# of OMIM genes with 100% coverage at >=20X	**97.7%**	**91.9%**	**96.9%**	pct	> 94%
# of OMIM genes with 100% coverage at >=30X	**95.8%**	**56.1%**	**83.1%**	pct	> 80%
# of genes with 100% coverage at >=40X	**81.5%**	**7.3%**	**23.8%**	pct	
Variation (VCF) metrics					
# of calls Total	**4850124**	**4783261**	**4841927**		2.5-6.0M
# of PASS calls	**4782957**	**4692663**	**4758545**		2.5-6.0M
# of calls Total coding	**27117**	**26583**	**26730**		25000-30000
Total # of SNVs	**3989749 [82.26%]**	**3953443 [82.65%]**	**3996596 [82.54%]**		
Total # of Indels	**860375 [17.74%]**	**829818 [17.35%]**	**845331 [17.46%]**		
Hom/Het ratio (in coding regions)	**0.58 (0.60)**	**0.58 (0.59)**	**0.58 (0.59)**	ratio	0.5-0.61
Ti/Tv ratio (in coding regions)	**2.00 (2.86)**	**2.00 (2.89)**	**2.01 (2.88)**	ratio	2-2.2 (2.8-3)
# of het calls (# of hom call)	**3087082 (1763042)**	**2980726 (1802535)**	**3085568 (1756359)**	units	
In-silico sample swap check	**PASS**	**PASS**	**PASS**		Mendelian test
Automated upload of VCF to Omicia	**PASS**	**PASS**	**PASS**		
Precision metrics compared to GIAB: total (on SNPs)	**0.9895 (0.9915)**	**n/a**	**n/a**		GIAB only available
Sensitivity metrics : total (on SNPs)	**0.9947 (0.9975)**	**n/a**	**n/a**		for NA12878
Update LIMS	**TBD**	**TBD**	**TBD**		
Download annotated VCF to RCI	**N/A**	**N/A**	**N/A**		

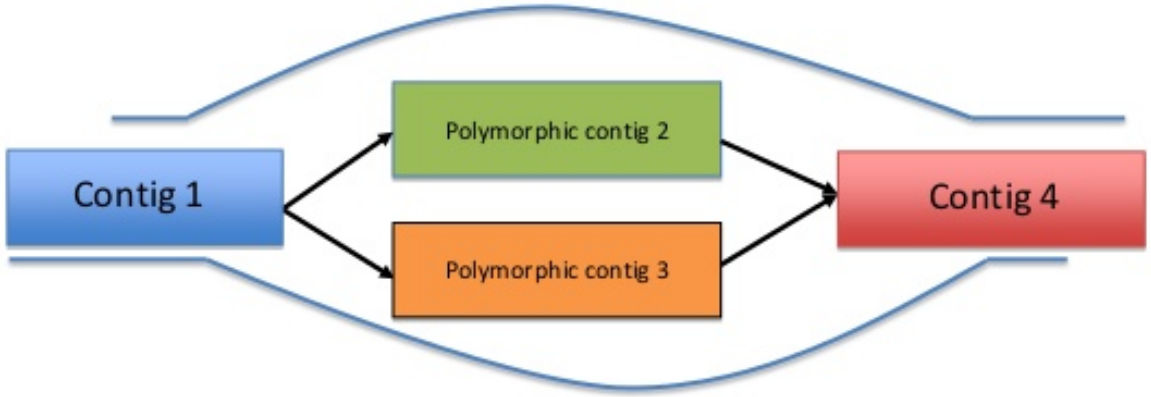
Just a slight detour …..



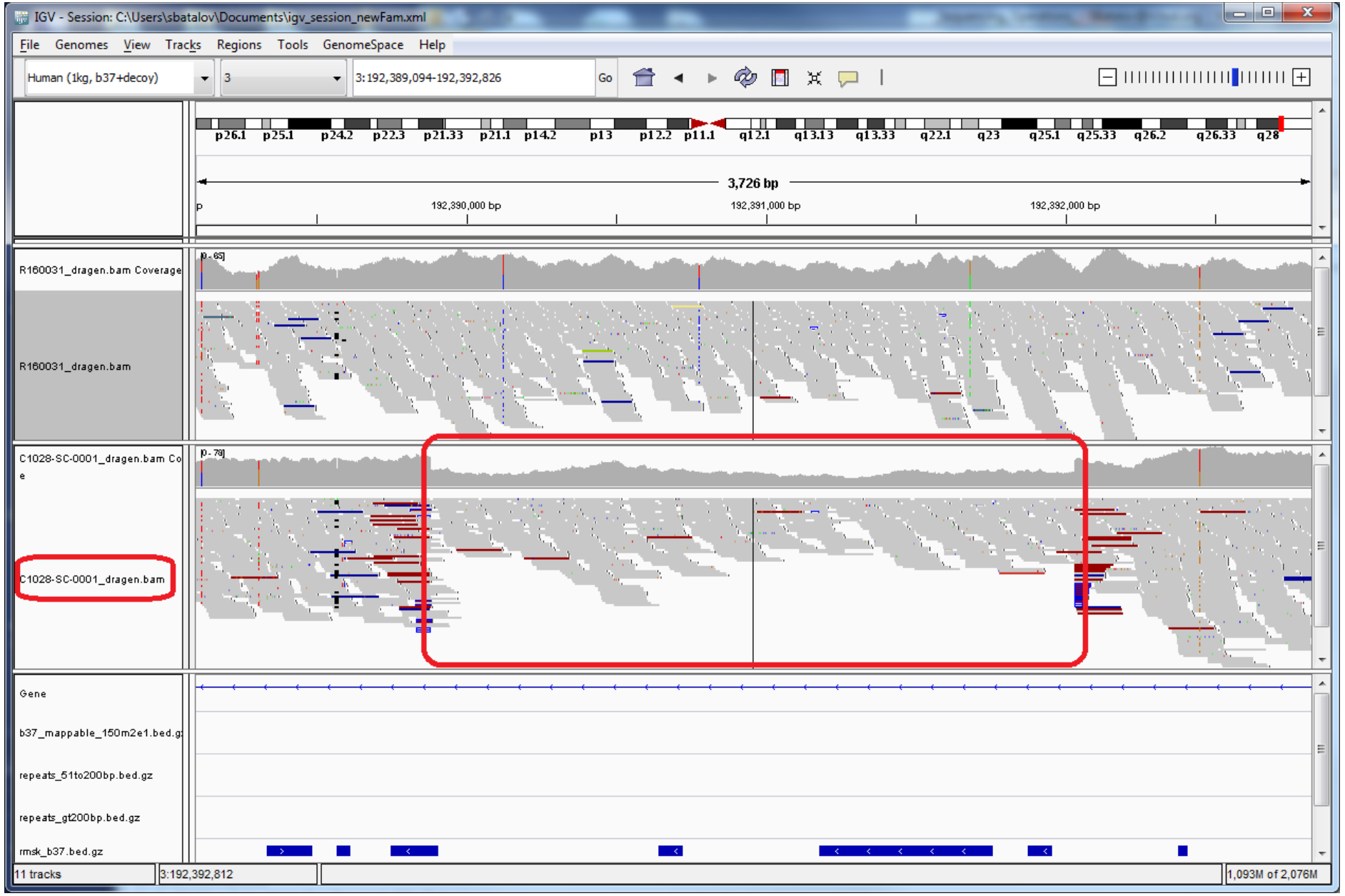
Longer reads?

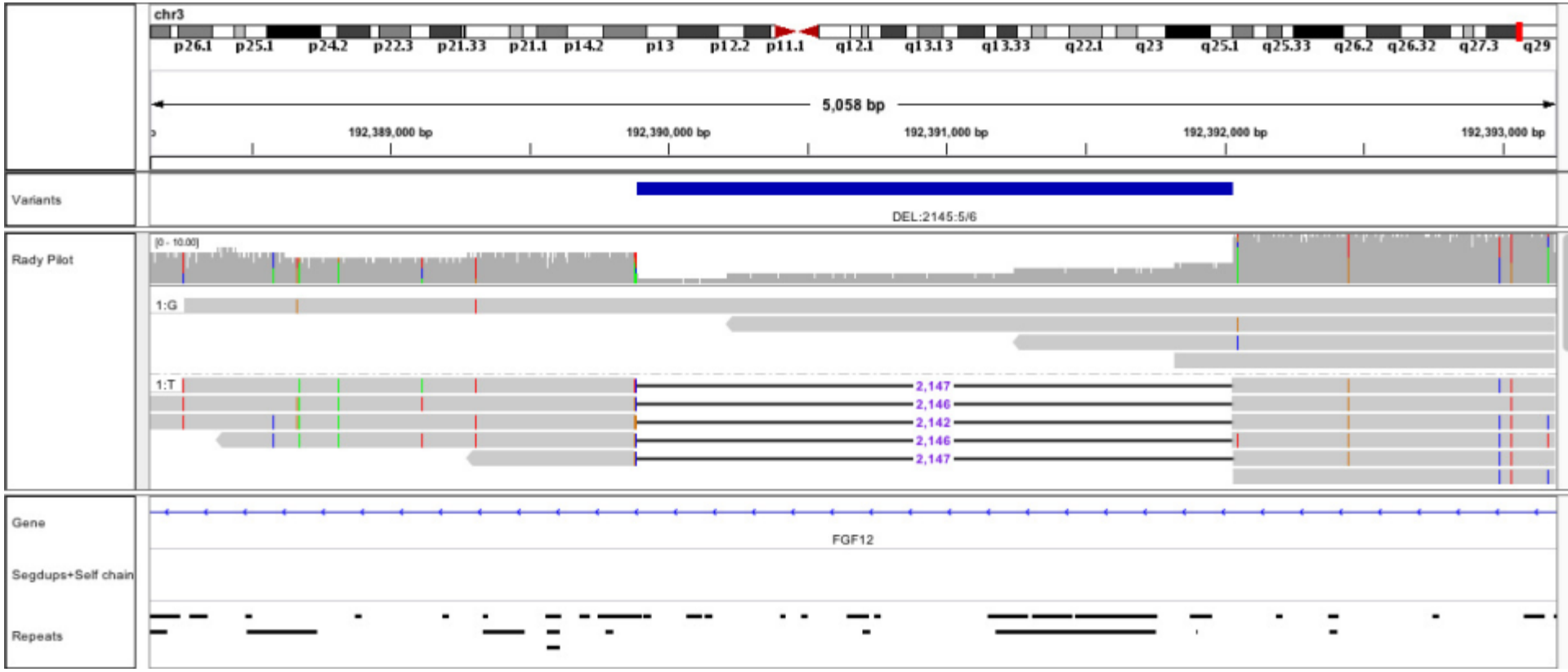


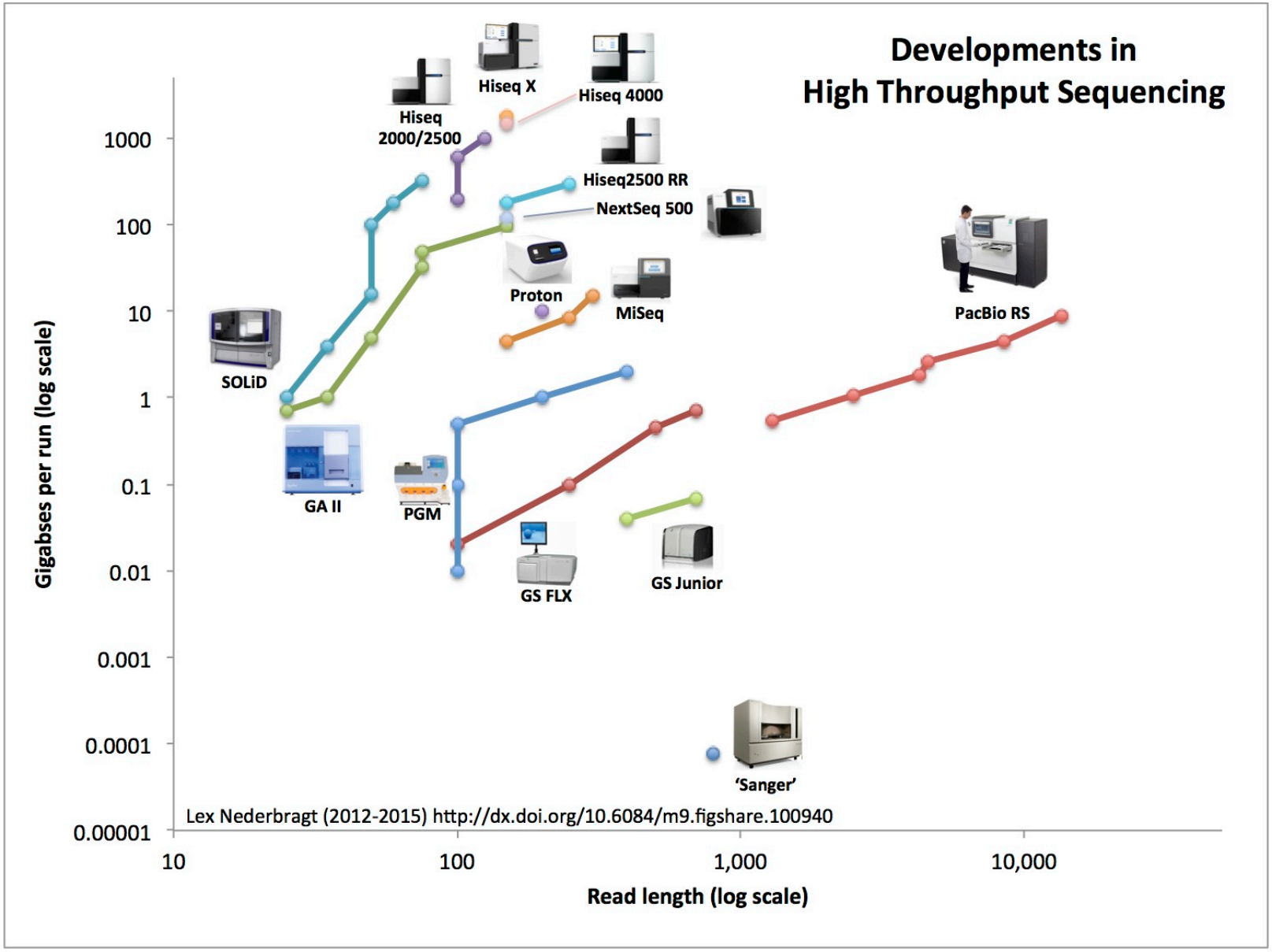
Long reads can span repeats and heterozygous regions



NORWEGIAN SEQUENCING CENTRE


IGV - Session: C:\Users\sbatalov\Documents\igv_session_newFam.xml
File Genomes View Tracks Regions Tools GenomeSpace Help
Human (1kg, b37+decoy)
3
3:192,389,094-192,392,826
Go
p26.1 p25.1 p24.2 p22.3 p21.33 p21.1 p14.2 p13 p12.2 p11.1 q12.1 q13.13 q13.33 q22.1 q23 q25.1 q25.33 q26.2 q26.33 q28
3,726 bp
192,390,000 bp
192,391,000 bp
192,392,000 bp
R160031_dragen.bam Coverage
R160031_dragen.bam
C1028-SC-0001_dragen.bam Co
e
C1028-SC-0001_dragen.bam
Gene
b37_mappable_150m2e1.bed.g
repeats_51to200bp.bed.gz
repeats_gt200bp.bed.gz
rmsk_b37.bed.gz
11 tracks
3:192,392,812
1,093M of 2,076M

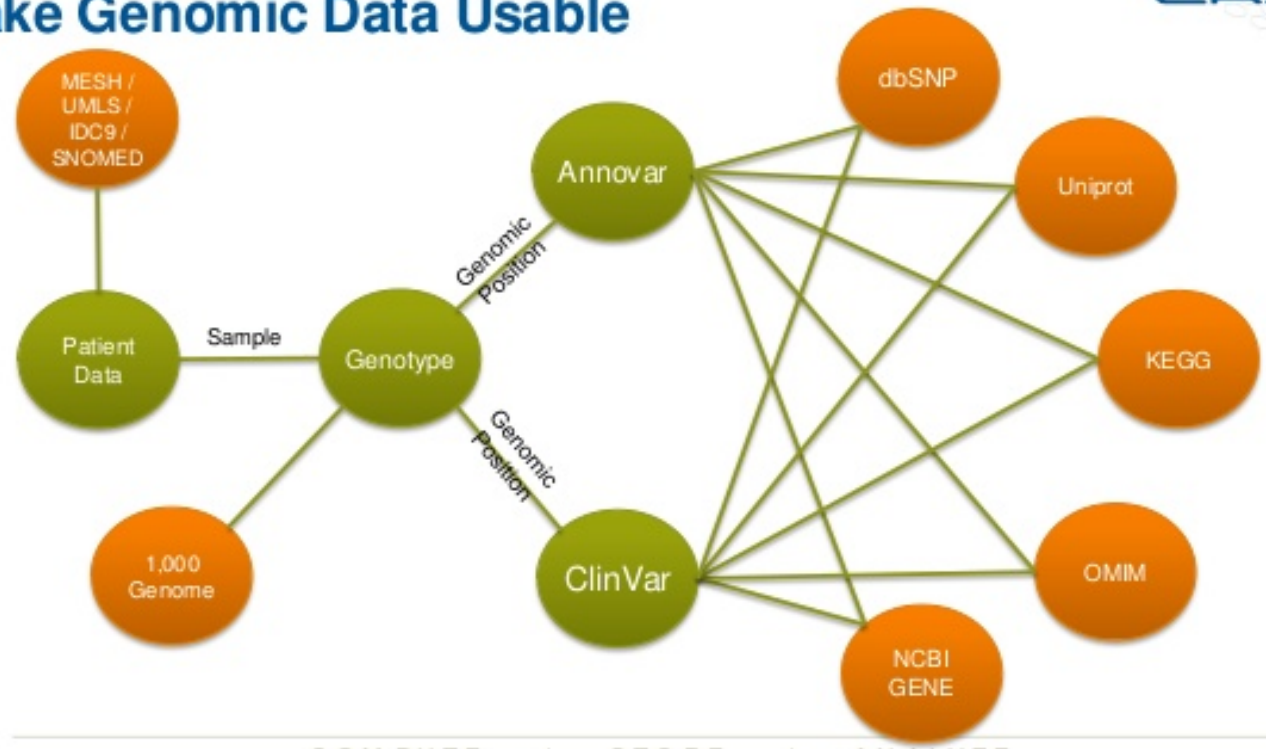

chr3
p26.1 p25.1 p24.2 p22.3 p21.33 p21.1 p14.2 p13 p12.2 p11.1 q12.1 q13.13 q13.33 q22.1 q23 q25.1 q25.33 q26.2 q26.32 q27.3 q29
5,058 bp
192,389,000 bp
192,390,000 bp
192,391,000 bp
192,392,000 bp
192,393,000 bp
Variants
DEL:2145:5/6
Rady Pilot
[0 - 10.00]
1:G
1:T
2,147
2,146
2,142
2,146
2,147
Gene
FGF12
Segdups+Self chain
Repeats


Developments in
High Throughput Sequencing
Hiseq X
Hiseq 4000
Hiseq
2000/2500
Hiseq2500 RR
NextSeq 500
Proton
MiSeq
PacBio RS
SOLiD
GA II
PGM
GS FLX
GS Junior
'Sanger'
Gigabses per run (log scale)
1000
100
10
1
0.1
0.01
0.001
0.0001
0.00001
10
100
1,000
10,000
Read length (log scale)
Lex Nederbragt (2012-2015) http://dx.doi.org/10.6084/m9.figshare.100940

Back to our bioinformatics pipeline….

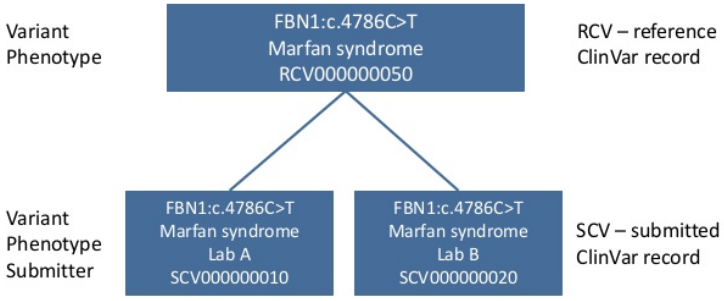
Annotation

Make Genomic Data Usable



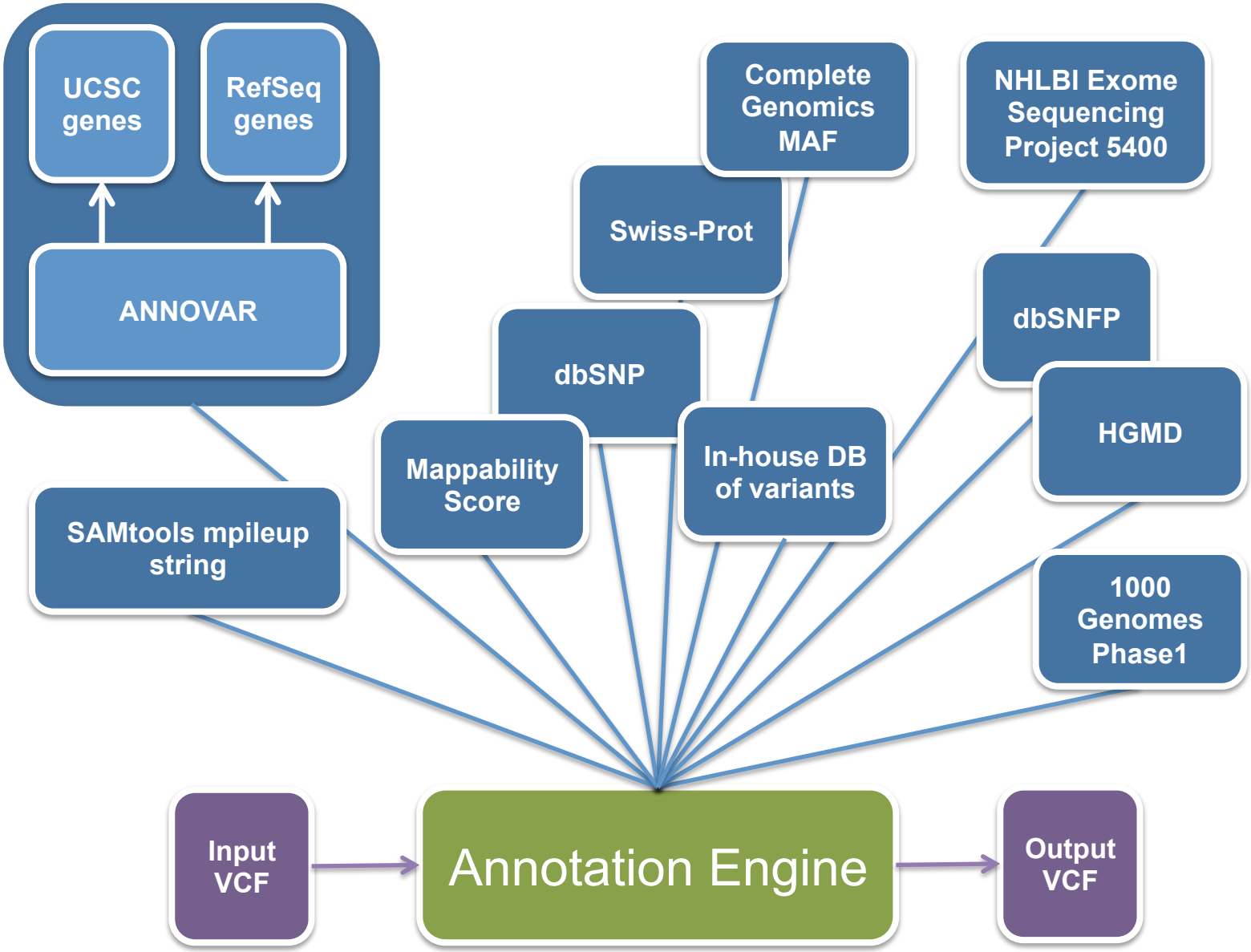
COMPUTE | STORE | ANALYZE

ClinVar aggregates by variant and phenotype



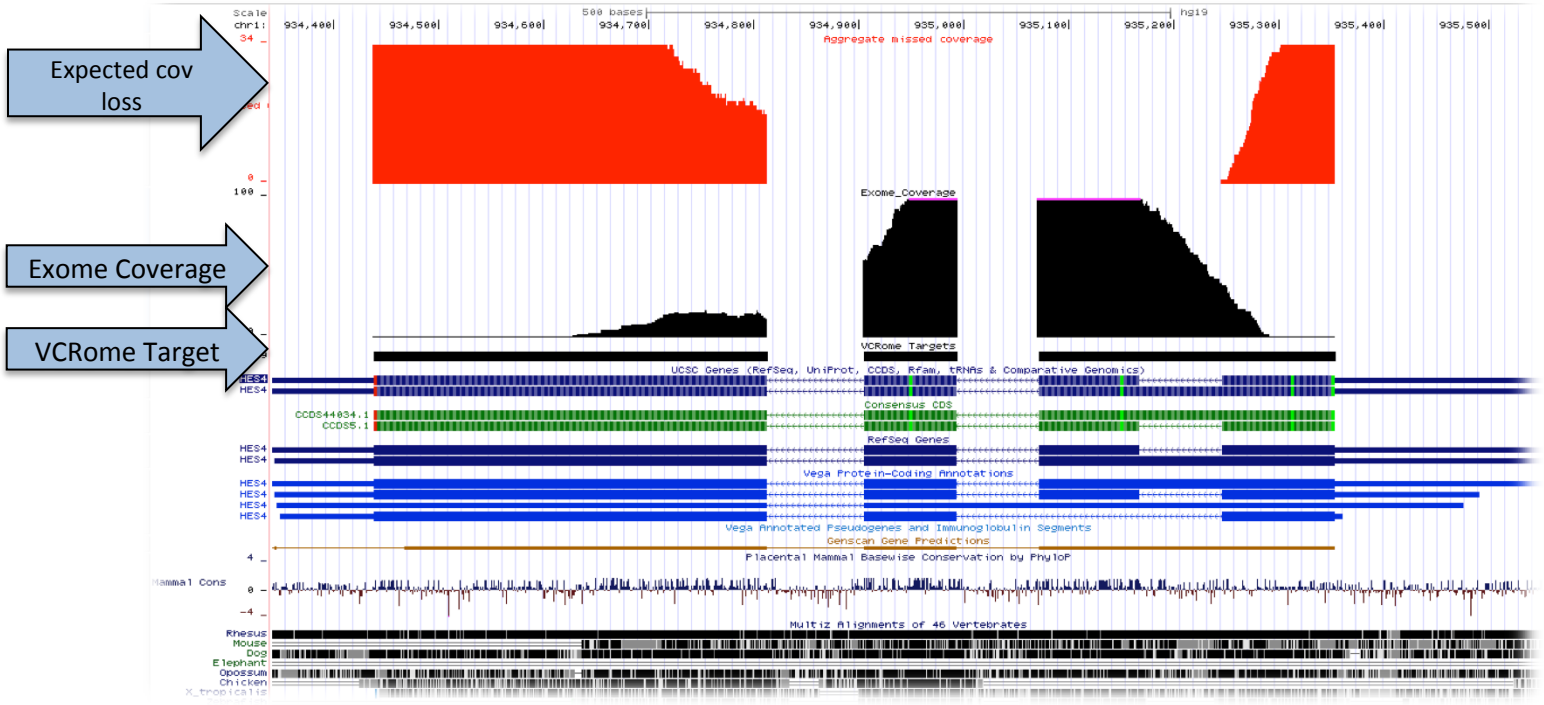
Integrate info from multiple public and private data bases

Variant Annotation Sources



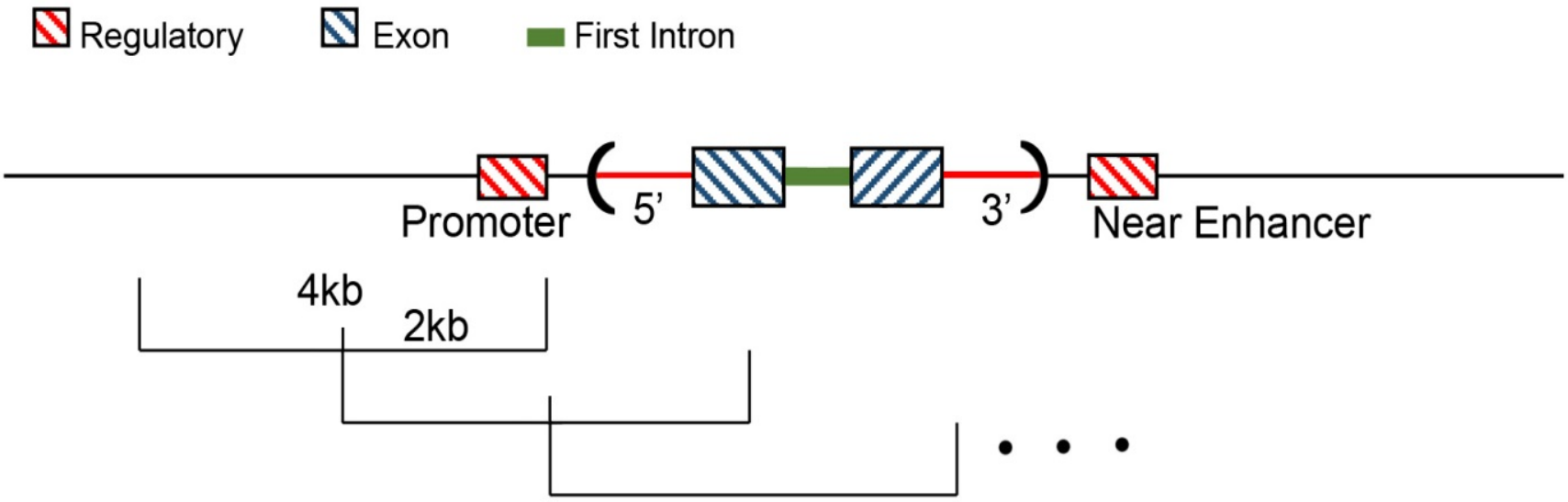
Variant Annotator

Exome Coverage and Identification (ExCID) Report



ExCID has a visualization component which allows the user to see coverage across a region of interest and correlate that with portions of the exon target information and regions of expected coverage loss.

Annotated functional motifs



Regulatory Domain

- Promoter
- UTR3
- UTR5
- Enhancer

Regulatory Domain Annotation

All SNVs – annotated using the WGSA pipeline, version 05 (Liu et al. 2016).

- Promoter – based on the overlap between the permissive set of Cap Analysis of Gene Expression (CAGE) peaks in FANTOM5 data (FANTOM Consortium and the RIKEN PMI and CLST (DGT) 2014) and the 5 kb upstream region (ANNOVAR(Wang et al. 2010) annotation, refSeq gene model - upstream).
- UTR3' – if variant overlaps a 3' untranslated region (ANNOVAR annotation, RefSeq gene model – UTR3).
- UTR5' – if variant overlaps a 5' untranslated region (ANNOVAR annotation, RefSeq gene model – UTR5).

Enhancer

The enhancers and the target genes of the enhancers were defined based on the permissive set of enhancers and enhancer-promoter pairs reported by the FANTOM5 project (Andersson et al. 2014).

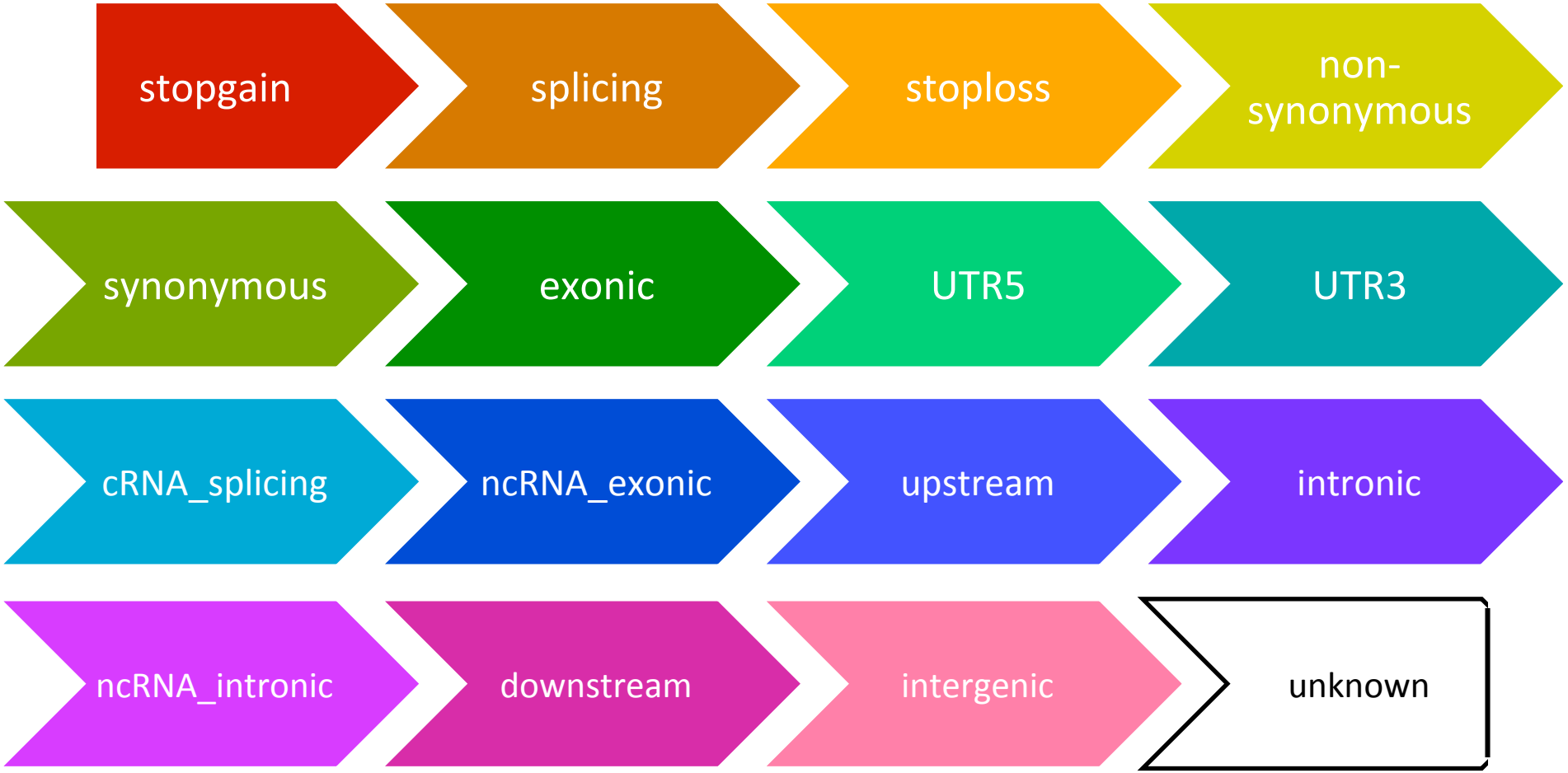
Multiple assignment of SNVs

Among 770,137 SNVs (MAF≤0.05) annotated to **regulatory domain** (MAC≥3) in AA (ARIC),

11.39% SNVs (87,718 variants) were assigned to more than 1 gene.

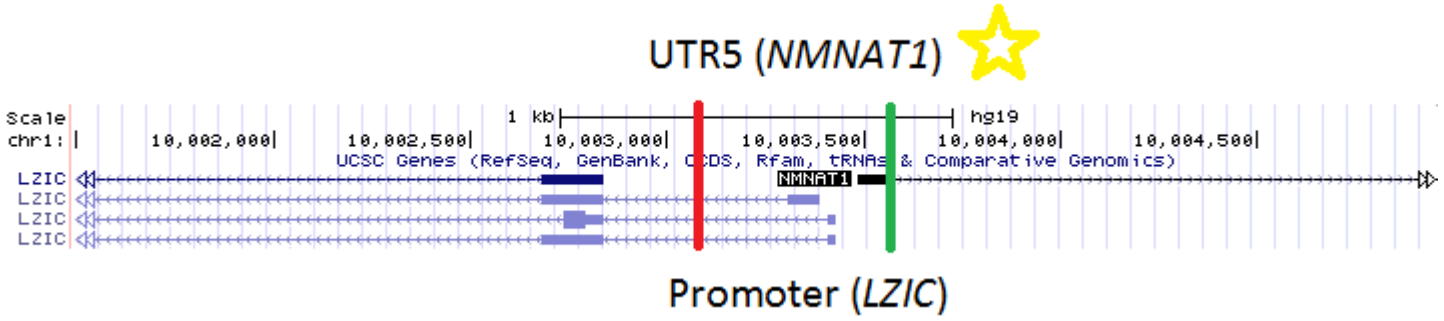
$$Unique\ SNVs\ assigned\ to\ different\ genes\ /\ Unique\ SNVs = 87{,}718/770{,}137 = 0.1139$$

If SNV has multiple gene-based annotations due to overlapping genes, the most “deleterious” consequence (reported by ANNOVAR using the RefSeq gene model) is used, based on the following order:

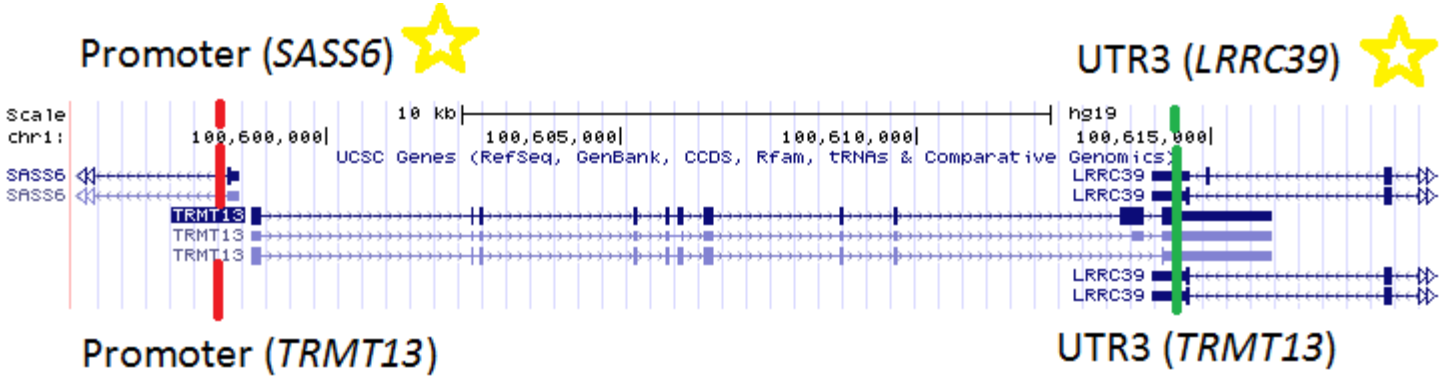


Multiple SNV assignment: examples

1:10003077 and 1:10003539



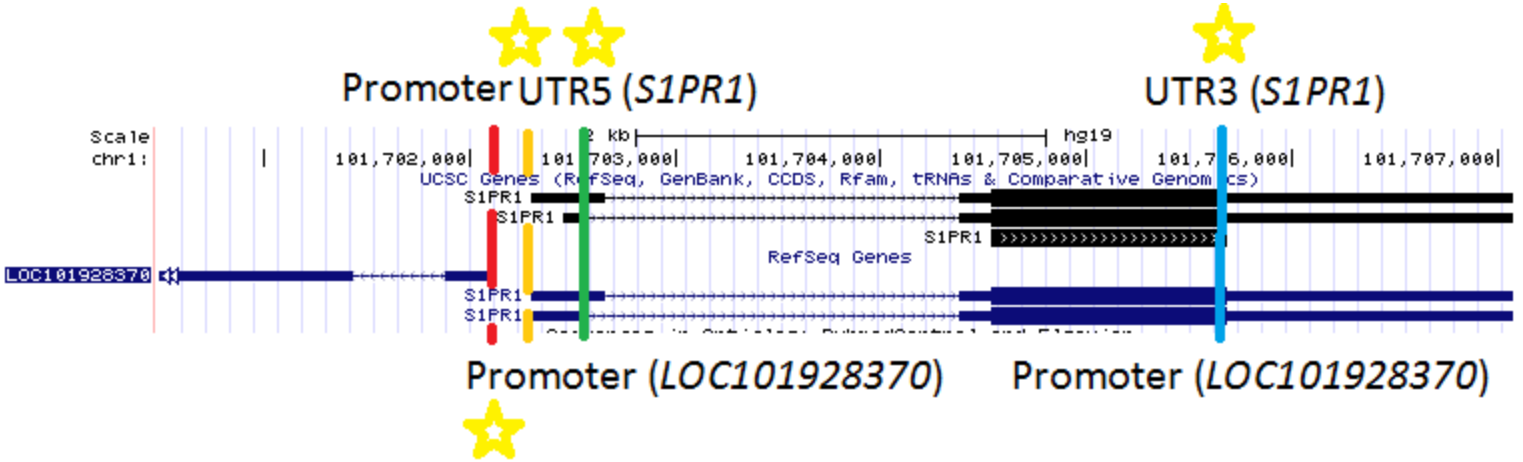
1:100598518 and 1:100614563



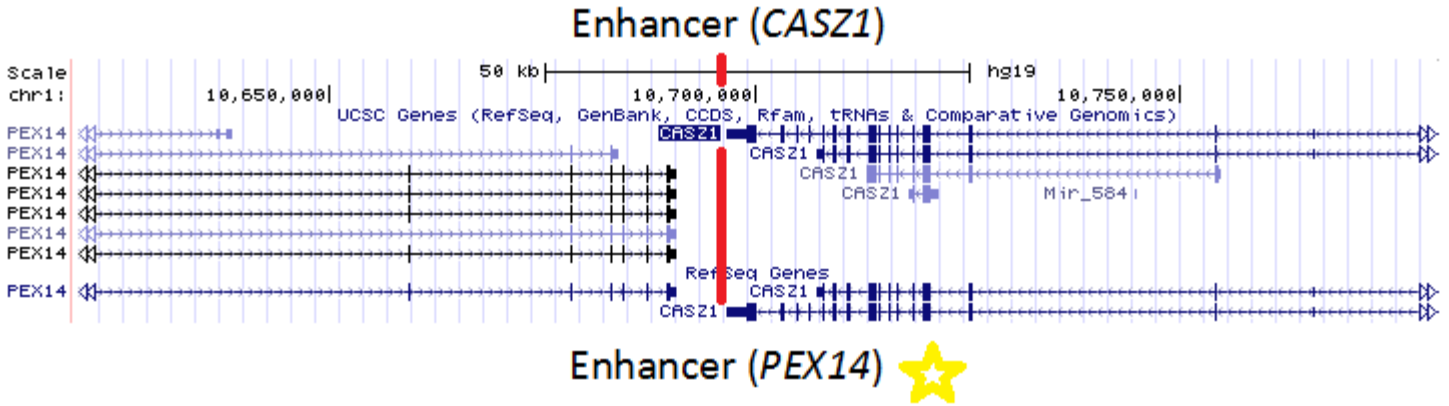
- Signifies which variant would have been chosen, if each SNV was assigned to be "unique" to one regulatory unit.

More examples

1:101702176, 1:101702359, 1:101702566, 1:101705827

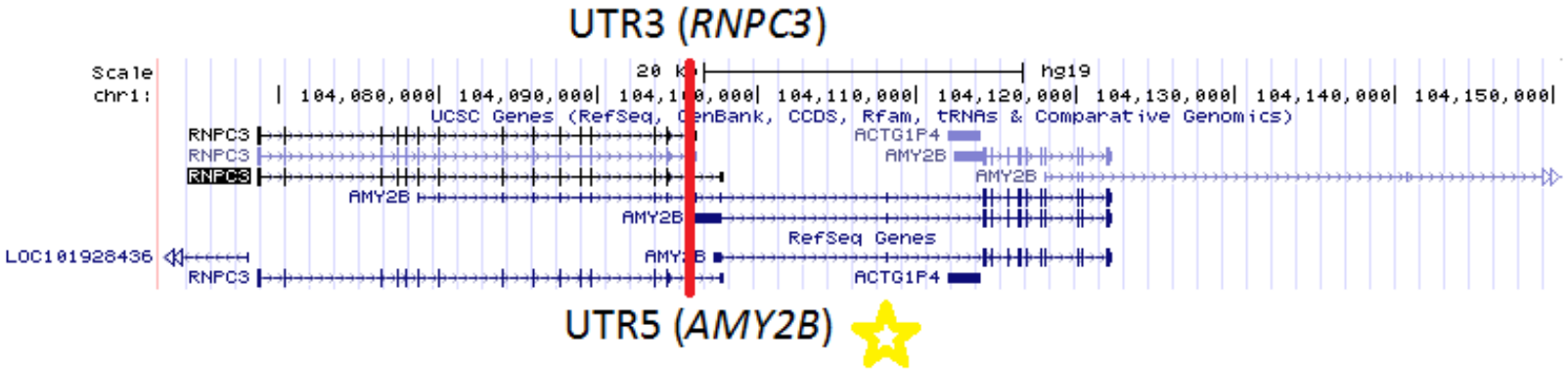


1:10695323



… and one more

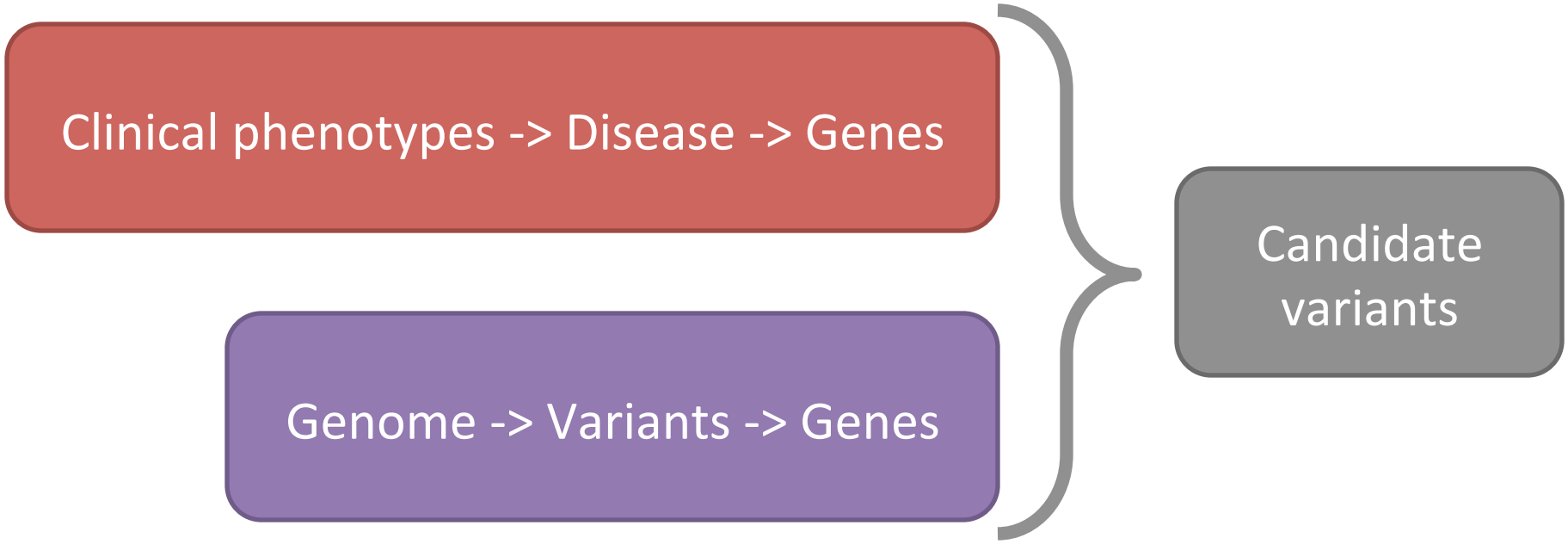
1:104097688



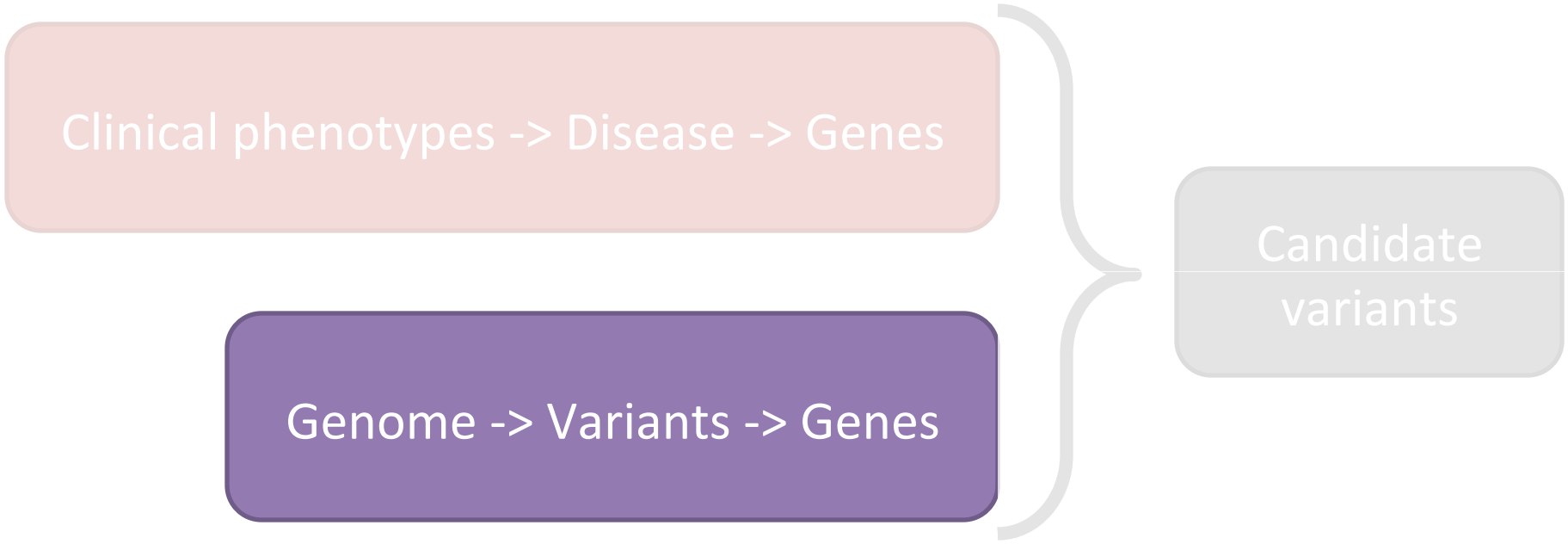
Break

Variant Filtering and Prioritization

Prioritization

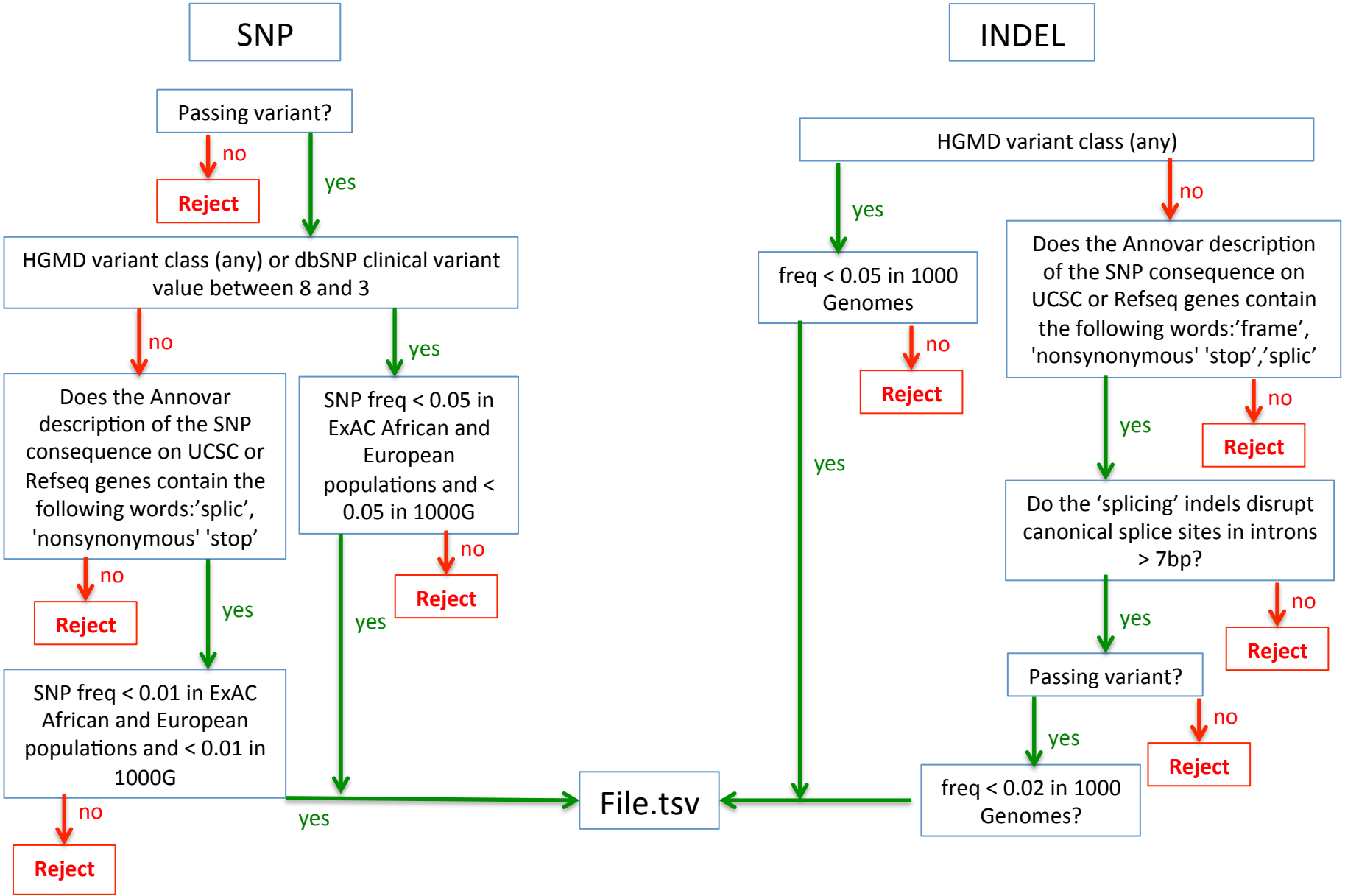


Prioritization – Genome Based



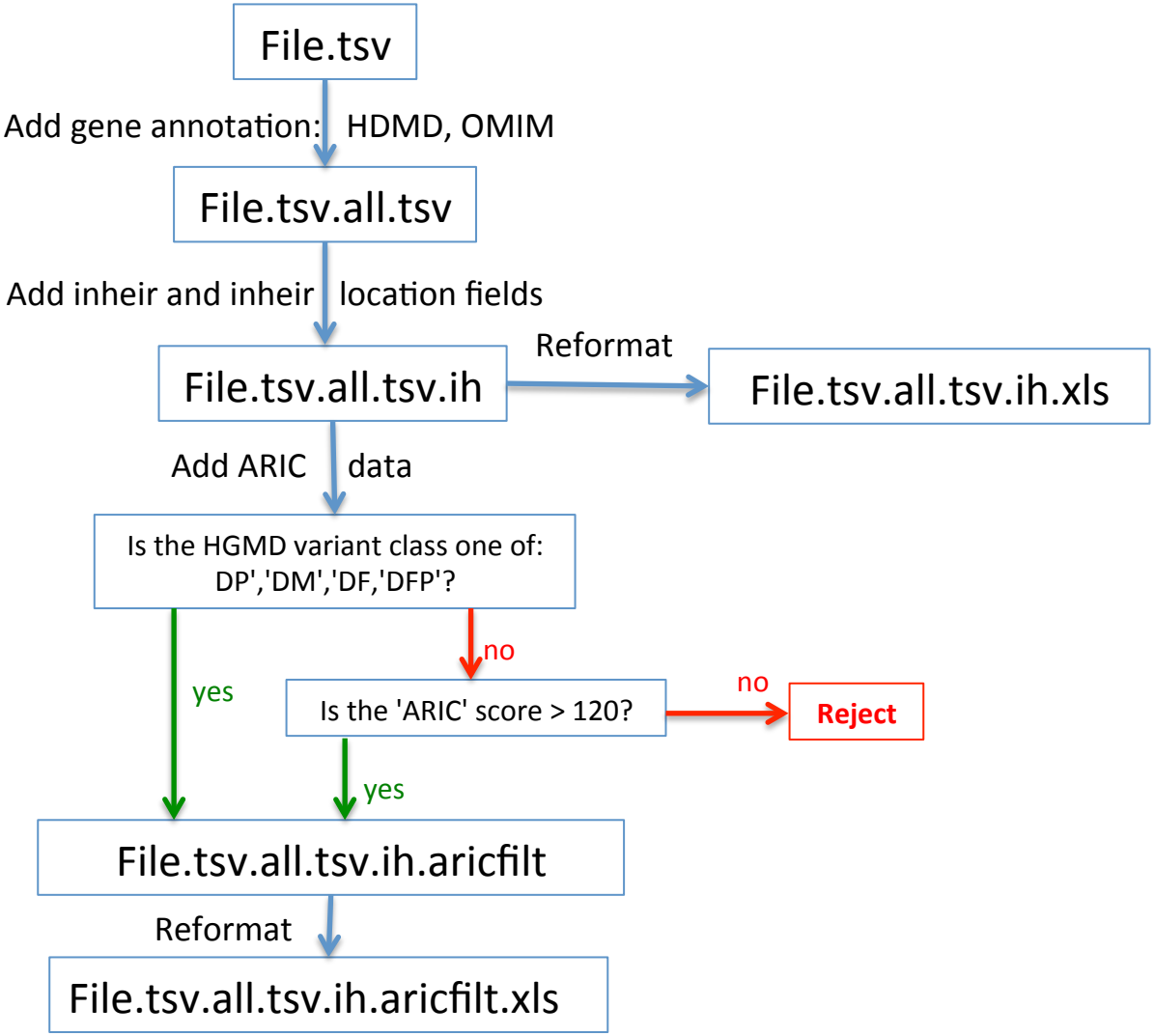
Mendelian Filter

Example filtering workflow



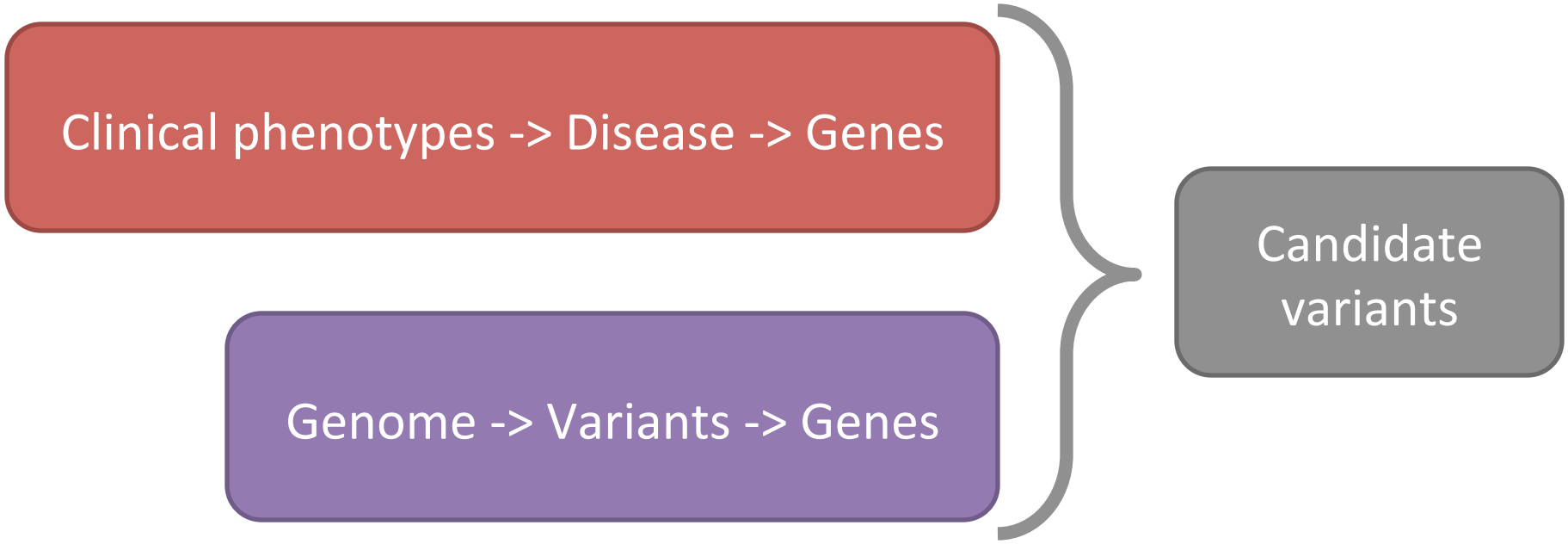
Mendelian Filtering

Secondary filtering

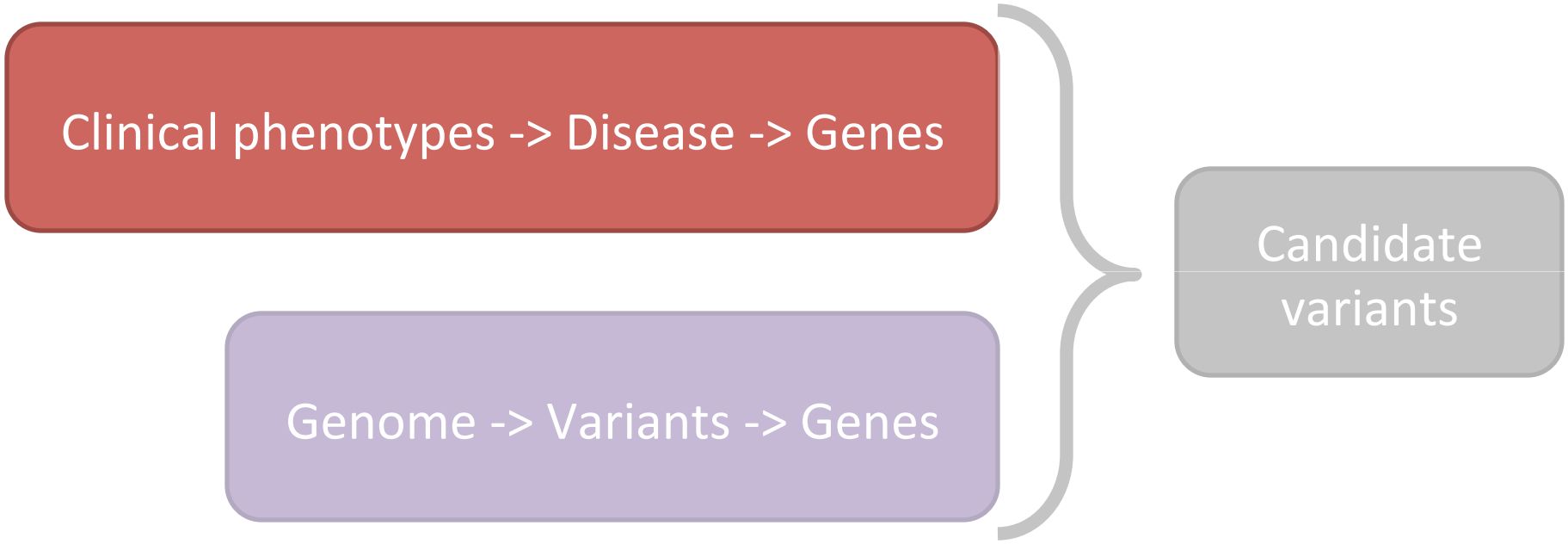


Baylor College of Medicine, Houston
Human Genome Sequencing Center

Prioritization



Phenotype Based Prioritization

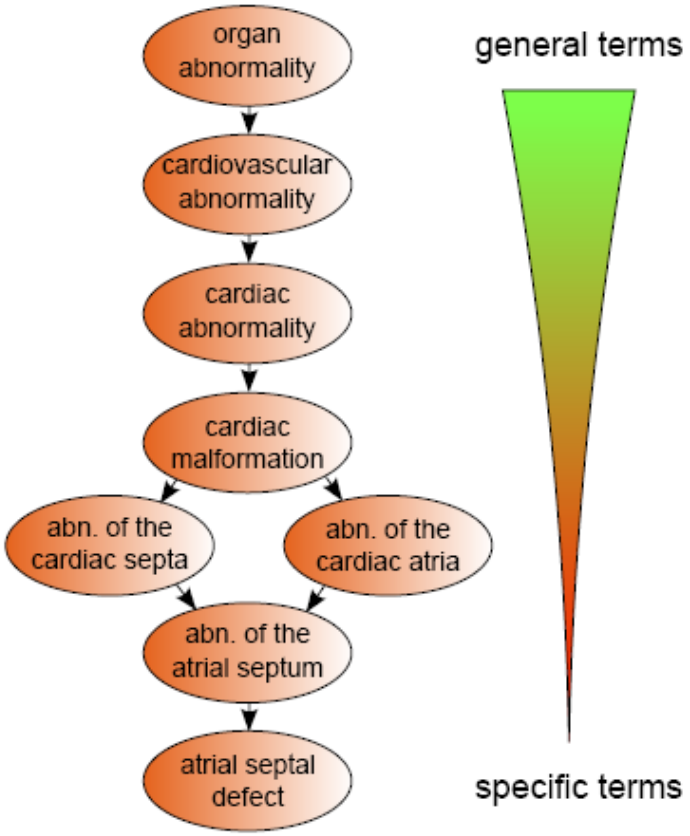


The Human Phenotype Ontology (HPO*)

www.human-phenotype-ontology.org

- The HPO provides a standardized vocabulary of phenotypic abnormalities encountered in human genetic syndromes

- Phenotypic features are formally represented as terms of a directed acyclic graph:
  - **Terms are related to parent terms by „is a“ relationships, representing subclasses of more general parent terms**
  - **Multiple parentage allows the representation of different aspects of phenotypic abnormalities**



*** Robinson P, Köhler S, Bauer S, Seelow D, Horn D, Mundlos:** The Human Phenotype Ontology: A Tool for annotating and analyzing human hereditary disease, *Am J Hum Genet. 2008 Nov*

Using the HPO to calculate Phenotypic Similarities

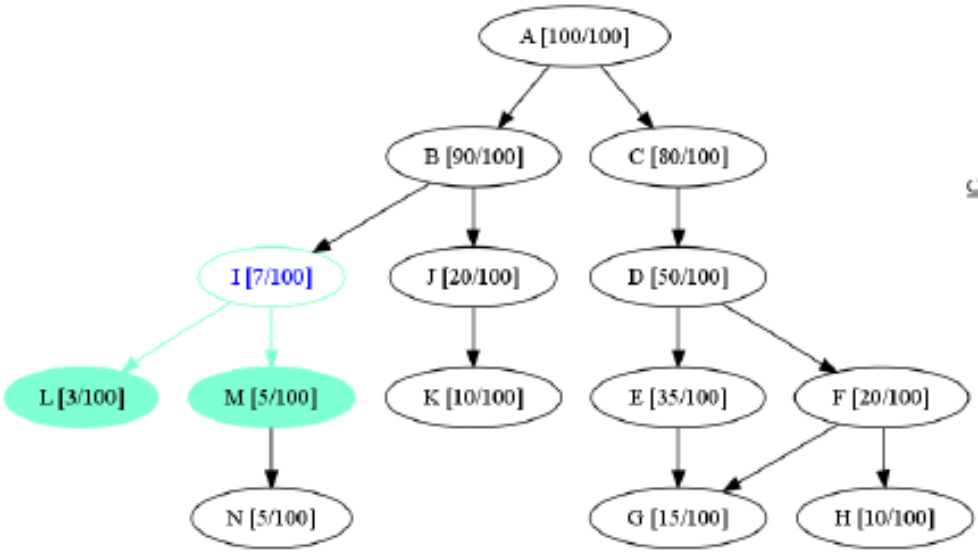
- The importance of a clinical phenotypic finding of the differential diagnosis depends on its specificity

- The specificity of a phenotypic feature, *t,* is represented by its information cotent (IC), defined as negative natural logarithm of its frequency of occurrence:

$$IC(t) = -\ln p(t)$$

- The similarity between two terms t_1 and t_2 is defined as the IC of their most specific common ancestor:

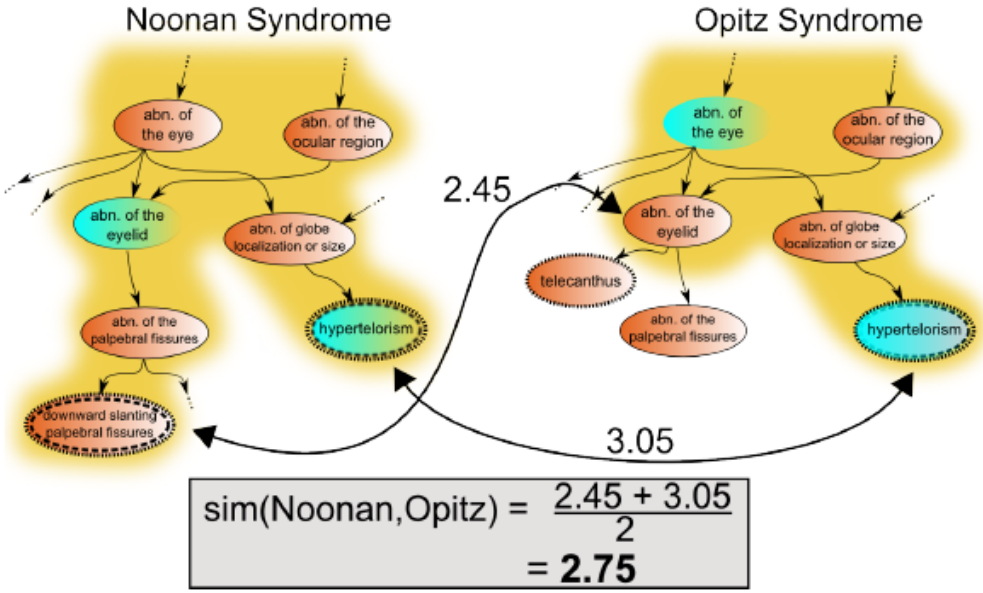
$$sim(t_1, t_2) = \max_{a \in A(t_1, t_2)} \left(-\ln p(a)\right)$$

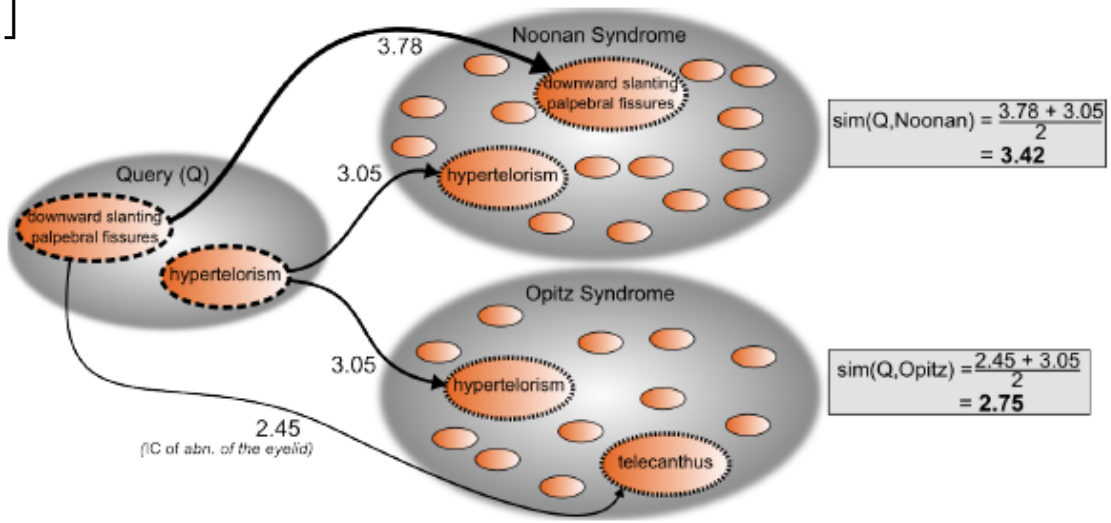


Ontological Similarity Search

- Term similarity measures may be used to compute similarity scores between genetic syndromes or phenotypic feature query sets:

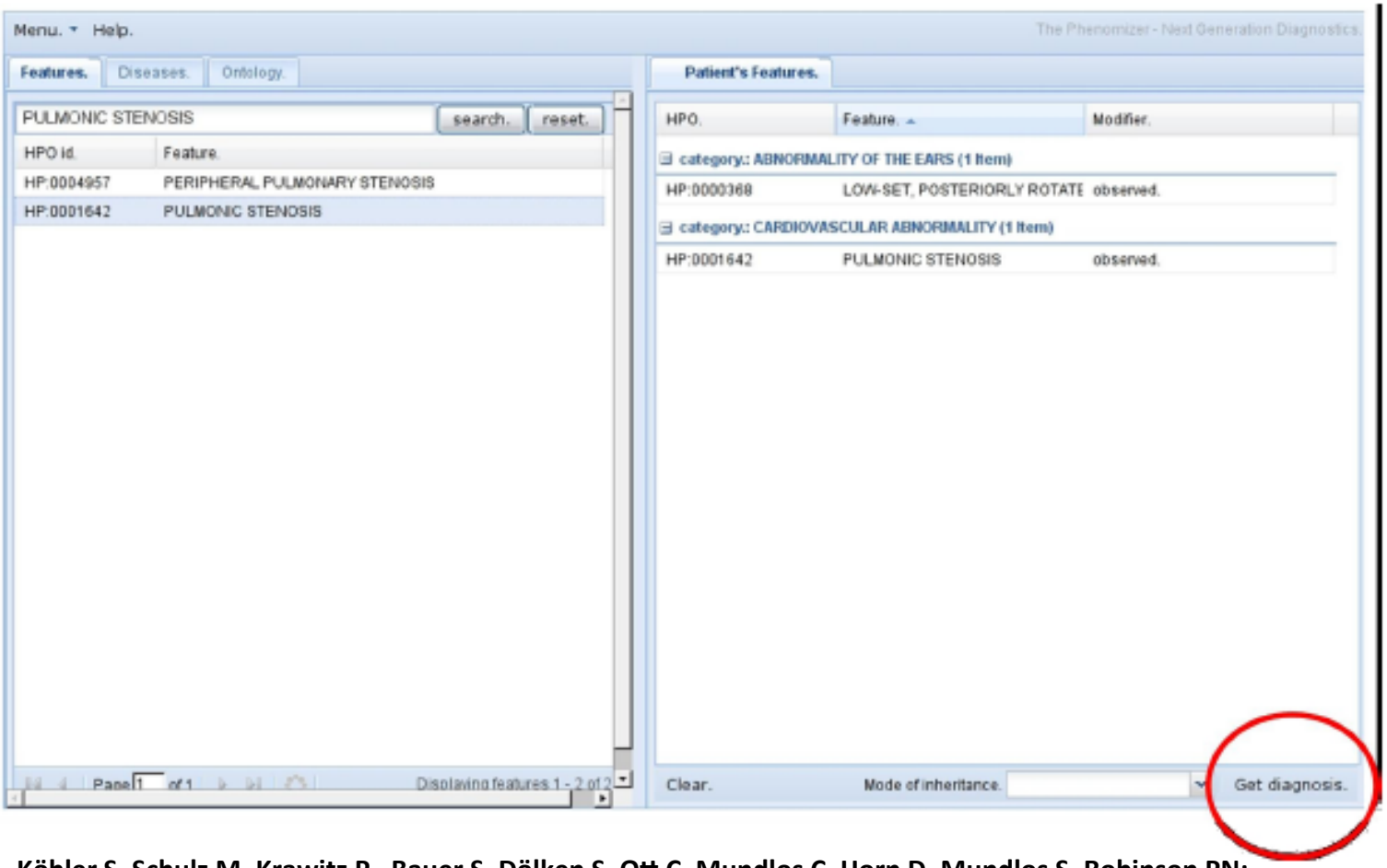
$$sim(D_1 \rightarrow D_2) = avg\left[\sum_{t_i \in D_1} \max_{t_j \in D_2}(sim(t_i, t_j))\right]$$





The Phenomizer*: A Tool for Diagnostics

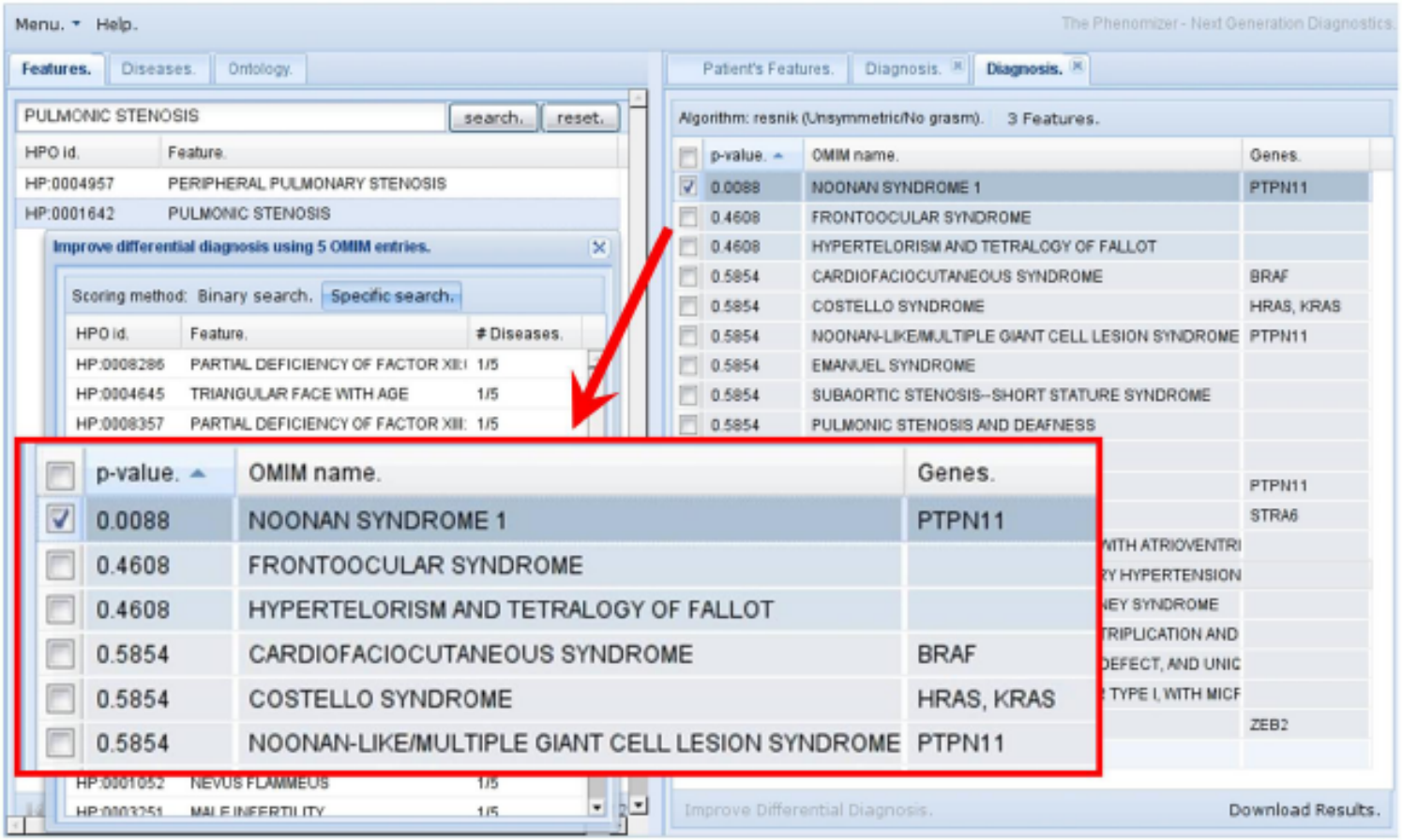
http://compbio.charite.de/Phenomizer/Phenomizer.html



Köhler S, Schulz M, Krawitz P, Bauer S, Dölken S, Ott C, Mundlos C, Horn D, Mundlos S, Robinson PN:
Clinical Diagnostics in Human Genetics with Semantic Similarity Searches in Ontologies, *Am J Hum Genet. 2009 Sep.*

The Phenomizer*: A Tool for Diagnostics

http://compbio.charite.de/Phenomizer/Phenomizer.html



Köhler S, Schulz M, Krawitz P, Bauer S, Dölken S, Ott C, Mundlos C, Horn D, Mundlos S, Robinson PN:
Clinical Diagnostics in Human Genetics with Semantic Similarity Searches in Ontologies, *Am J Hum Genet. 2009 Sep.*

Deep Phenotype Data

The Phenomizer

Patient's Features.

HPO.	Feature.	Modifier.	Num diseas...
category.: Abnormality of metabolism/homeostasis (2 Items)			
HP:0002908	Conjugated hyperbilirubinemia	observed.	13 of 7994
HP:0001942	Metabolic acidosis	observed.	63 of 7994
category.: Abnormality of the abdomen (1 Item)			
HP:0001396	Cholestasis	observed.	125 of 7994
category.: Abnormality of the cardiovascular system (2 Items)			
HP:0001643	Patent ductus arteriosus	observed.	199 of 7994
HP:0004971	Pulmonary artery hypoplasia	observed.	2 of 7994
category.: Abnormality of the respiratory system (1 Item)			
HP:0004971	Pulmonary artery hypoplasia	observed.	2 of 7994
category.: Growth abnormality (1 Item)			
HP:0001508	Failure to thrive	observed.	370 of 7994

Clear. Mode of inheritance. Get diagnosis.

Rady Children's Institute
Genomic Medicine

Phenomizer or Phenolyzer

The Phenomizer

Patient's Features. | **Diagnosis.**

Algorithm: resnik (Unsymmetric). | 6 Features.

	p-val...	...	Disease name.	Genes.
	0.0080	...	#613404 ARTHROGRYPOSIS, RENAL DYSFUNCTION, A...	VPS33B, VIP...
	0.0080	...	%612726 HARDIKAR SYNDROME	
	0.0107	...	#208085 ARTHROGRYPOSIS, RENAL DYSFUNCTION, A...	VPS33B, VIP...
	0.0160	...	CHOLESTASIS, PROGRESSIVE FAMILIAL INTRAHEPATI...	ATP8B1
	0.0160	...	#614582 COMBINED OXIDATIVE PHOSPHORYLATION D...	MRPL3
	0.0160	...	#615595 COMBINED OXIDATIVE PHOSPHORYLATION D...	LYRM4, NFS1
	0.0160	...	HYPERBILIRUBINEMIA, CONJUGATED, TYPE III	
	0.0160	...	BILIRUBIN, SERUM LEVEL OF, QUANTITATIVE TRAIT LO...	
	0.0196	...	#610205 ALAGILLE SYNDROME 2; ALGS2	NOTCH2
	0.0240	...	#229600 FRUCTOSE INTOLERANCE, HEREDITARY;;FRU...	ALDOB
	0.0437	...	CHOLESTASIS, PROGRESSIVE FAMILIAL INTRAHEPATI...	TJP2, ABCB11
	0.0454	...	#605711 MULTIPLE MITOCHONDRIAL DYSFUNCTIONS ...	NFU1
	0.0456	...	#251880 MITOCHONDRIAL DNA DEPLETION SYNDROM...	DGUOK
	0.0489	...	BILIARY MALFORMATION WITH RENAL TUBULAR INSU...	
	0.0489	...	#614300 HYPERMETHIONINEMIA DUE TO ADENOSINE ...	ADK
	0.0489	...	#614096 COMBINED OXIDATIVE PHOSPHORYLATION D...	AARS2

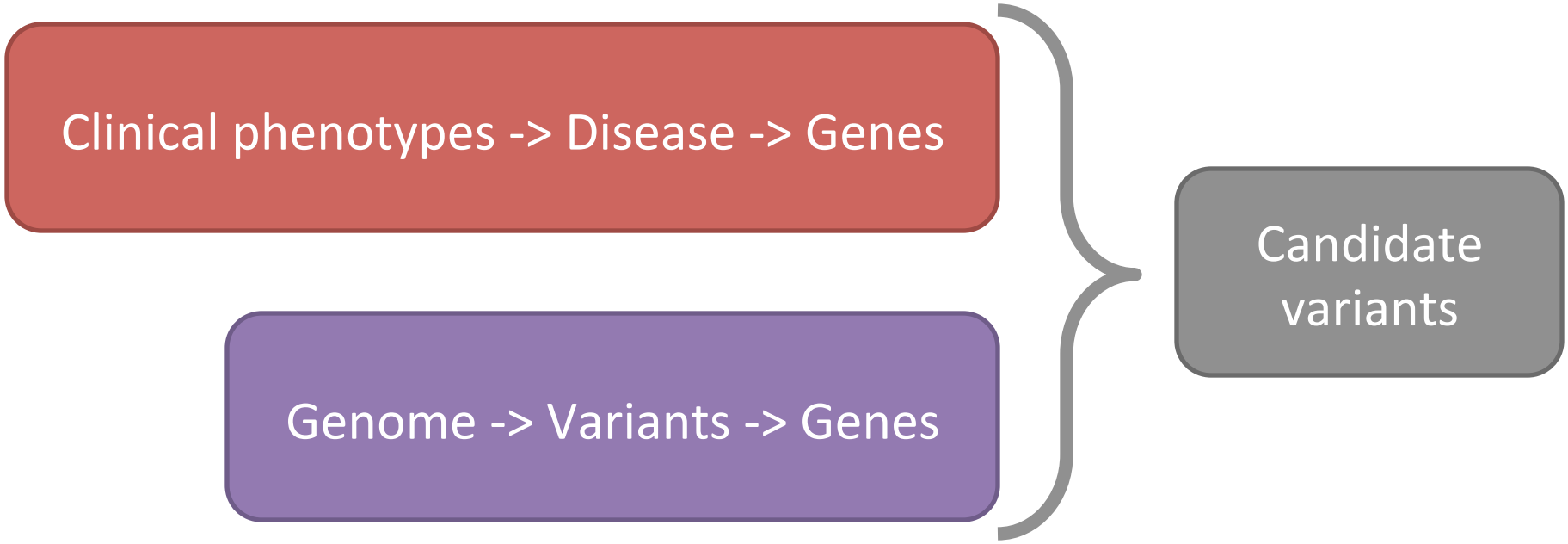
Page 1 of 268 | Improve Differential Diagnosis. | Download Results.

P-value Calculation

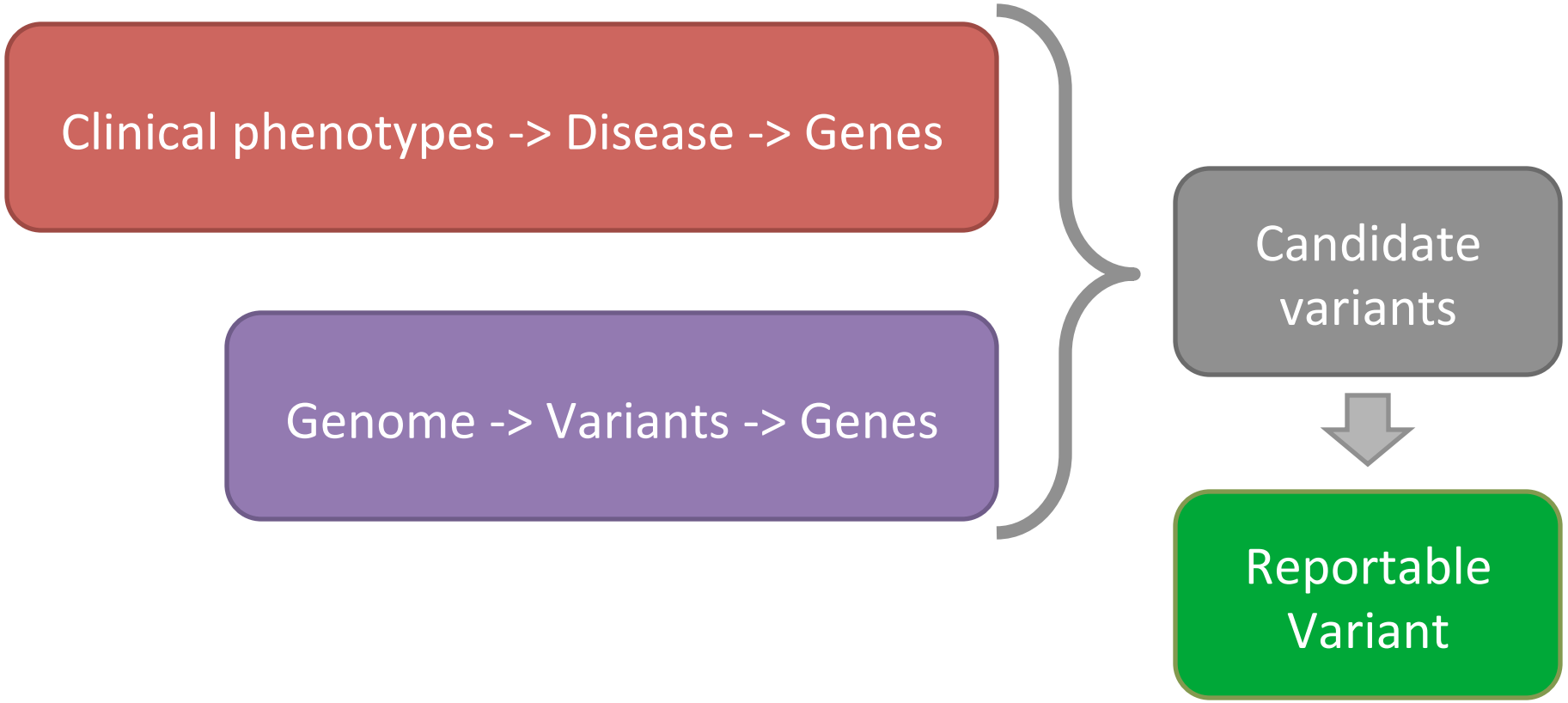
- The raw similarity score *S* depends on the number and specificty of terms both of the query *Q* and of the diseases *D* represented in the database
- The distribution of similarity scores can be used to obtain the significance of search results
- The P-value for the null hypothesis that a similiarty score of S or greater for a set of query terms Q and a disease D has been observed by chance is defined as:

$$P(s \geq S) = \frac{number\ of\ \ queries\ such\ that\ sim(Q,D) \geq S}{total\ number\ of\ \ possible\ queries}$$

Prioritization



Now we have a “reportable” variant



ACMG Guidelines

ACMG STANDARDS AND GUIDELINES | Genetics inMedicine

Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD[1], Nazneen Aziz, PhD[2,16], Sherri Bale, PhD[3], David Bick, MD[4], Soma Das, PhD[5], Julie Gastier-Foster, PhD[6,7,8], Wayne W. Grody, MD, PhD[9,10,11], Madhuri Hegde, PhD[12], Elaine Lyon, PhD[13], Elaine Spector, PhD[14], Karl Voelkerding, MD[13] and Heidi L. Rehm, PhD[15]; on behalf of the ACMG Laboratory Quality Assurance Committee

Disclaimer: These ACMG Standards and Guidelines were developed primarily as an educational resource for clinical laboratory geneticists to help them provide quality clinical laboratory services. Adherence to these standards and guidelines is voluntary and does not necessarily assure a successful medical outcome. These Standards and Guidelines should not be considered inclusive of all proper procedures and tests or exclusive of other procedures and tests that are reasonably directed to obtaining the same results. In determining the propriety of any specific procedure or test, the clinical laboratory geneticist should apply his or her own professional judgment to the specific circumstances presented by the individual patient or specimen. Clinical laboratory geneticists are encouraged to document in the patient's record the rationale for the use of a particular procedure or test, whether or not it is in conformance with these Standards and Guidelines. They also are advised to take notice of the date any particular guideline was adopted and to consider other relevant medical and scientific information that becomes available after that date. It also would be prudent to consider whether intellectual property interests may restrict the performance of certain tests and other procedures.

https://www.acmg.net/docs/Standards_Guidelines_for_the_Interpretation_of_Sequence_Variants.pdf

Pathogenic

(i) 1 Very strong (PVS1) AND
  (a) ≥1 Strong (PS1–PS4) OR
  (b) ≥2 Moderate (PM1–PM6) OR
  (c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) OR
  (d) ≥2 Supporting (PP1–PP5)
(ii) ≥2 Strong (PS1–PS4) OR
(iii) 1 Strong (PS1–PS4) AND
  (a)≥3 Moderate (PM1–PM6) OR
  (b)2 Moderate (PM1–PM6) AND ≥2 Supporting (PP1–PP5) OR
  (c)1 Moderate (PM1–PM6) AND ≥4 supporting (PP1–PP5)

Likely pathogenic

(i) 1 Very strong (PVS1) AND 1 moderate (PM1–PM6) OR
(ii) 1 Strong (PS1–PS4) AND 1–2 moderate (PM1–PM6) OR
(iii) 1 Strong (PS1–PS4) AND ≥2 supporting (PP1–PP5) OR
(iv) ≥3 Moderate (PM1–PM6) OR
(v) 2 Moderate (PM1–PM6) AND ≥2 supporting (PP1–PP5) OR
(vi) 1 Moderate (PM1–PM6) AND ≥4 supporting (PP1–PP5)

Benign

(i) 1 Stand-alone (BA1) OR
(ii) ≥2 Strong (BS1–BS4)

Likely benign

(i) 1 Strong (BS1–BS4) and 1 supporting (BP1– BP7) OR
(ii) ≥2 Supporting (BP1–BP7)

Uncertain significance

(i) Other criteria shown above are not met OR
(ii) the criteria for benign and pathogenic are contradictory

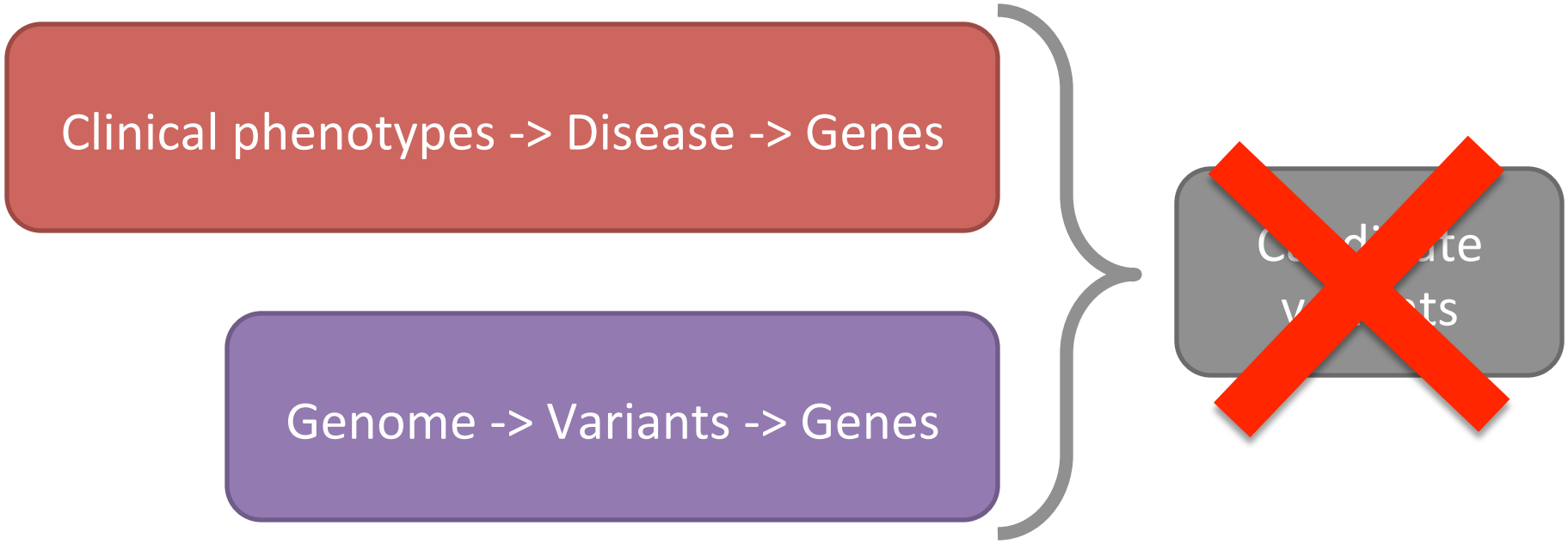
	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased 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	

	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 *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans 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			

Fortunately…

- http://www.medschool.umaryland.edu/Genetic_Variant_Interpretation_Tool1.html/

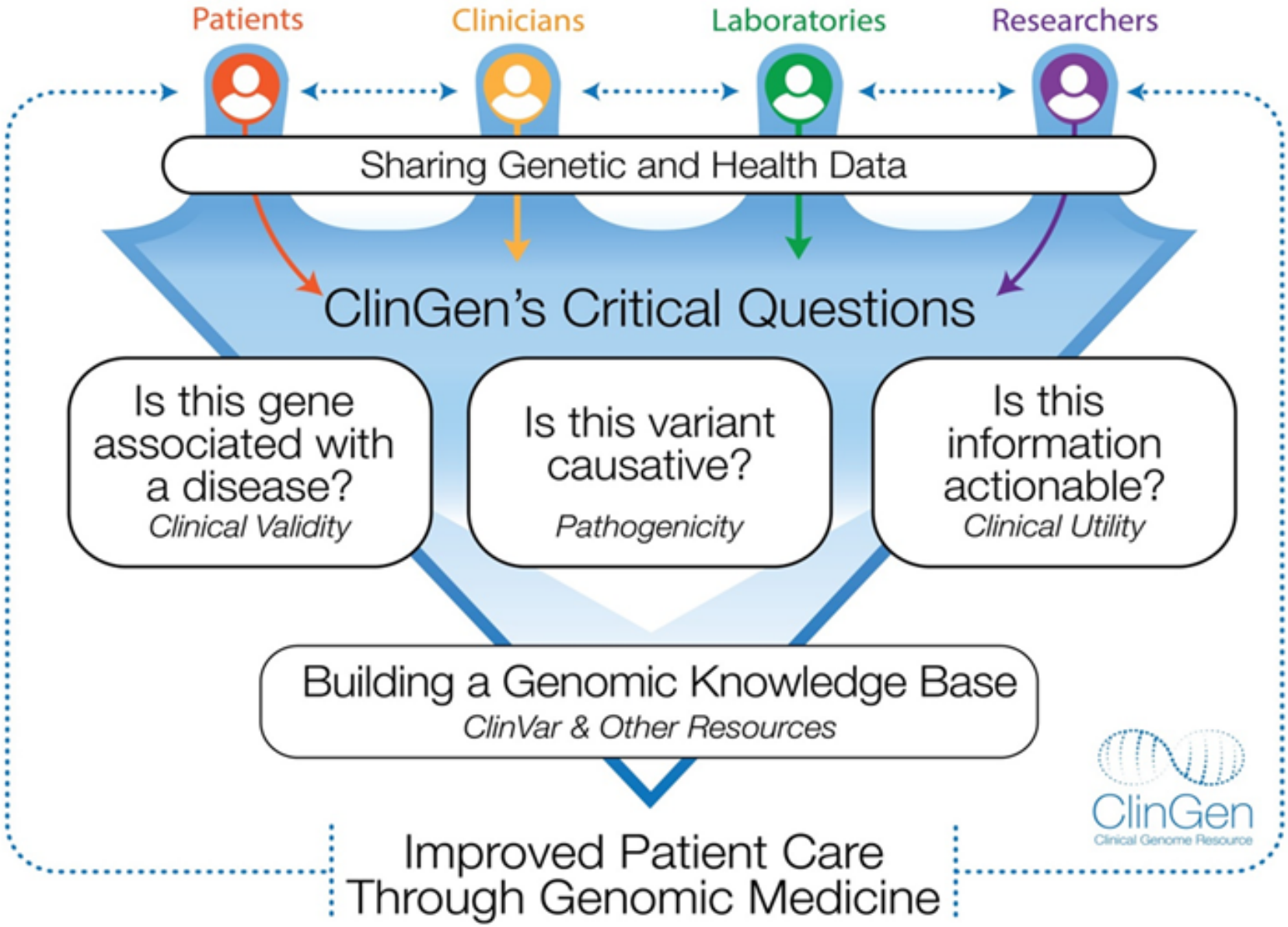
What if you are unlucky?



Matchmaker Exchange

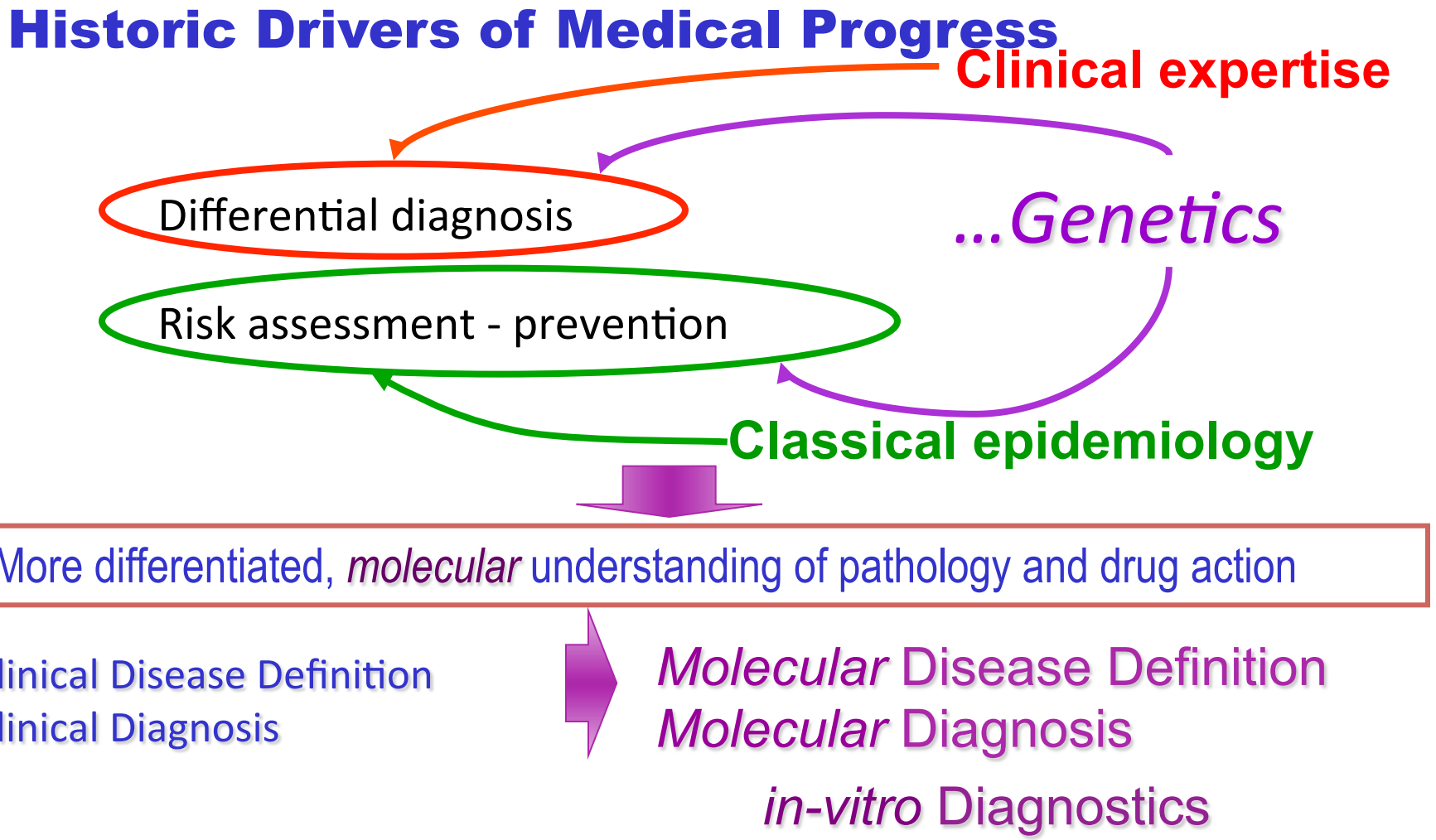
Genomic discovery through the exchange of phenotypic & genotypic profiles

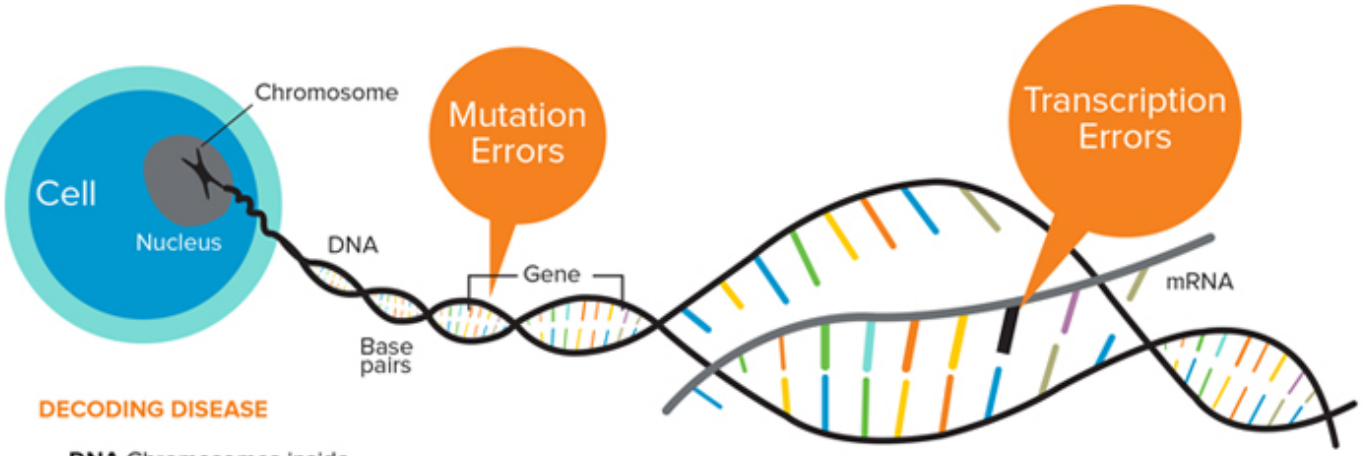

ClinGen
Clinical Genome Resource
AGTC
Centers for Mendelian Genomics
JOHNS HOPKINS
MEDICINE
BCM
Baylor College of Medicine
CARE forRARE
Phenome Central
ClinGen Genome Connect
Gene Matcher
Café Variome
Phenome Central
monarch INITIATIVE
Monarch
DECIPHER
wellcome trust sanger institute
DECIPHER GRCh37
LOVD
Leiden Open Variation Database
LOVD
Matchmaker Exchange
Undiagnosed Diseases Network
Undiagnosed Diseases Network
BROAD INSTITUTE
BROAD RDAP
GENESIS/ GEM.app
MIAMI
THE GENESIS PROJECT
PEER
RD Connect
PrivateAccess
Genetic Alliance
PEER
RD Connect


Patients
Clinicians
Laboratories
Researchers
Sharing Genetic and Health Data
ClinGen's Critical Questions
Is this gene associated with a disease?
Clinical Validity
Is this variant causative?
Pathogenicity
Is this information actionable?
Clinical Utility
Building a Genomic Knowledge Base
ClinVar & Other Resources
ClinGen
Clinical Genome Resource
Improved Patient Care Through Genomic Medicine

Just the start…….

Medical Progress: Evolution or Revolution?



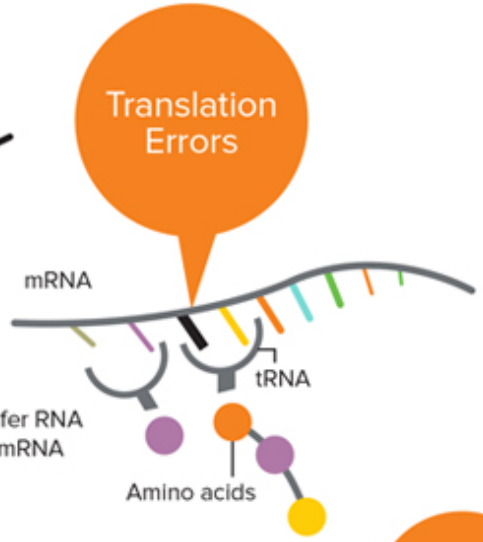


DECODING DISEASE

DNA Chromosomes inside every cell's nucleus contain the code of life—3 billion base pairs that tell the body how to build and repair itself. Errors can lead to disease.

GENES The main working parts of DNA, genes are sequences that instruct the body to perform functions. Humans have about 20,000.

TRANSCRIPTION To express the code, messenger RNA (mRNA) is transcribed from the gene sequence.



TRANSLATION Transfer RNA (tRNA) translates the mRNA into amino acids.

PROTEINS The amino acid fold into proteins, which perform functions throughout the body. Using advanced computing power, NANTHEATH is sequencing those proteins.

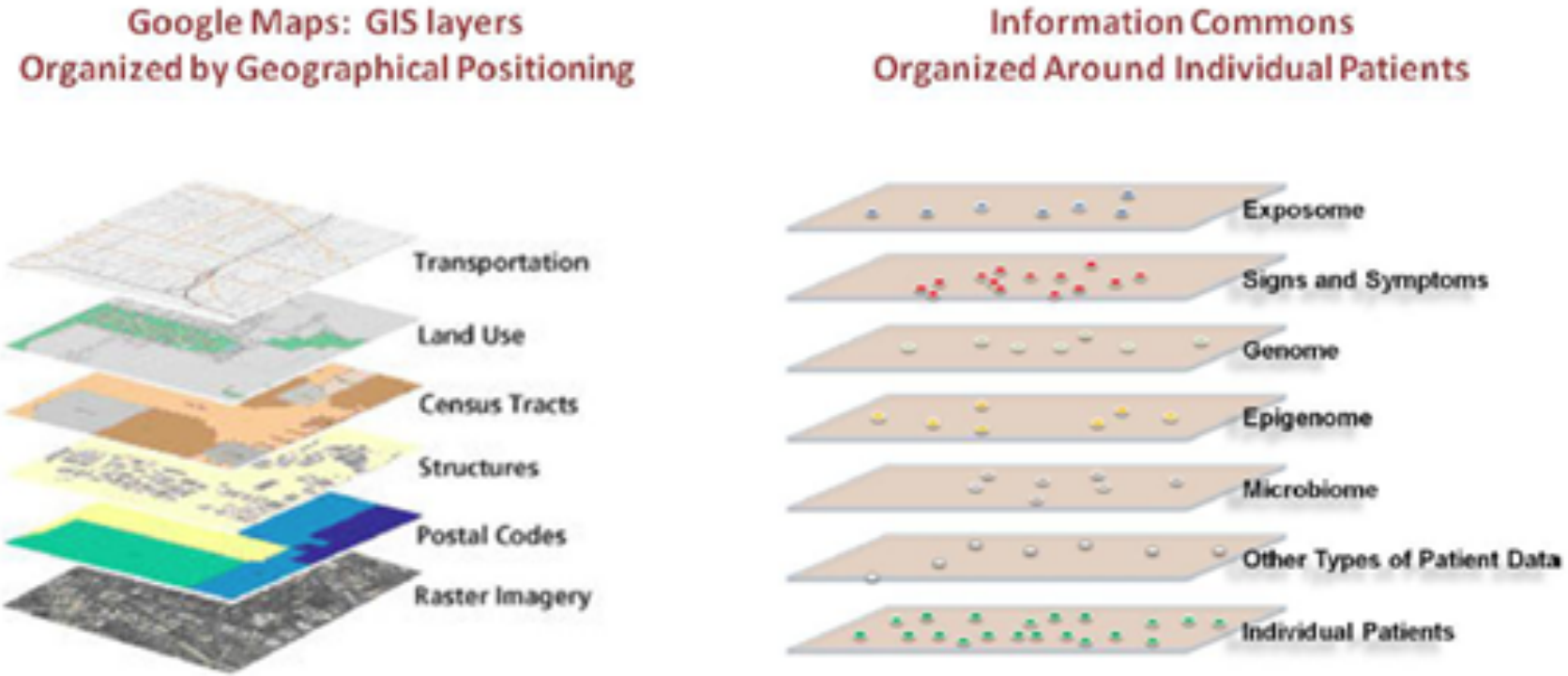


Figure 1-2: The proposed, individual-centric Information Commons (right panel) is somewhat analogous to a layered Geographical Information System (left panel). In both cases, the bottom layer defines the organization of all the overlays. However, in a GIS, any vertical line through the layers connects related snippets of information since all the layers are organized by geographical position. In contrast, data in each of the higher layers of the Information Commons will overlay on the patient layer in complex ways (e.g., patients with similar microbiomes and symptoms may have very different genome sequences). Source: FPA 2011 (left panel).

http://www.nimh.nih.gov/images/news-items/information_commons_142568_2.png