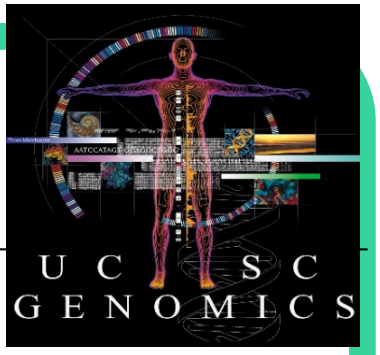
UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics Institute



Viewing the data:
UCSC Genome Browser and its possibilities

.

Robert Kuhn
UC Santa Cruz

.

Variant Prediction Training Course

Johor, Malaysia

.

August 27-30, 2018

@GenomeBrowser

Disclosures

Royalties from Browser licenses

Bioinformatics contract, Regeneron, Inc

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

UC SC GENOMICS

Acknowledgements

funding:

National Human Genome Research Institute (NHGRI)

California Institute for Regenerative Medicine (CIRM)

QB3 (UCBerkeley, UCSF, UCSC)

Chan Zuckerberg Initiative

Howard Hughes Medical Institute

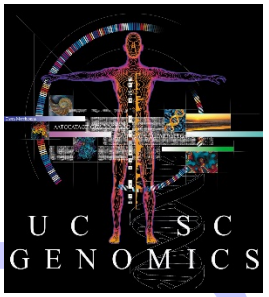
CIRM CALIFORNIA'S STEM CELL AGENCY

qb3 ucb·ucsc·ucsf

CHAN ZUCKERBERG INITIATIVE

HHMI HOWARD HUGHES MEDICAL INSTITUTE

UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics Institute



Acknowledgements

UCSC Browser team

- **David Haussler** – co-PI
- **Jim Kent** – Browser Concept, BLAT, Team Leader, PI

Engineering
Angie Hinrichs
Kate Rosenbloom
Hiram Clawson
Galt Barber
Brian Raney
Max Haeussler
Jonathan Casper
Christopher Lee

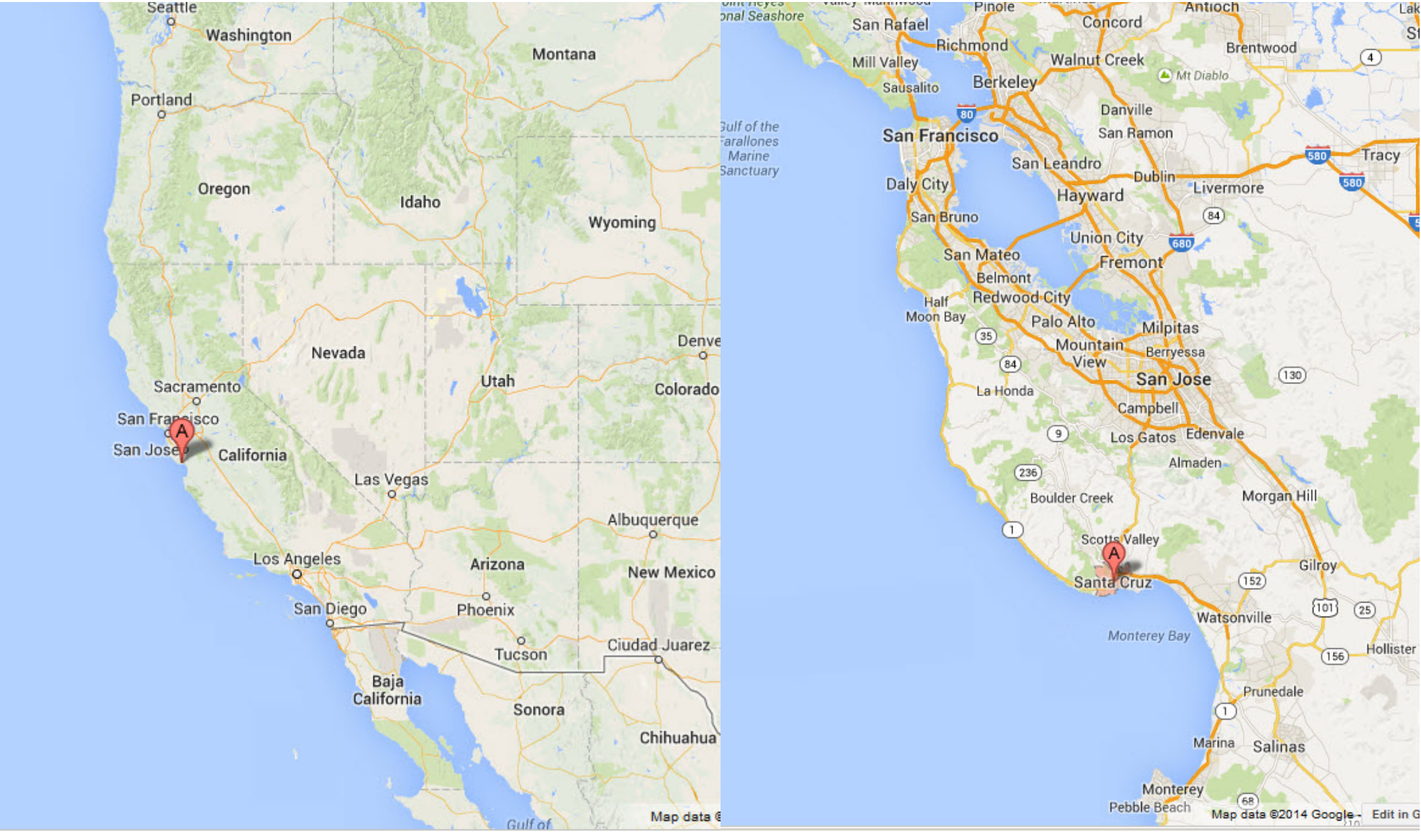
QA, Docs, Support
Brian Lee
Matt Speir
Jairo Navarro
Chris Villarreal
Lou Nassar
Daniel Smelter
Conner Powell

KiloKluster, Sys-admin
Jorge Garcia
Erich Weiler
Haifang Talc

Management
Ann Zweig

UCSC = UC Santa Cruz

!= USC, USCS, UCSF, UCSD….



Browser Team

Next-Gen Sequence Data

How do you visualize your Next-Gen Sequencing data?

Data flood

DECIPHER

71,000,000 mRNAs

54,000,000 SNPs

CNVs

30,000 genes

OMIM

GENOME 10K

3,000,000,000

nucleotides

Microarray expression

Links

http:// **genome-asia.ucsc.edu**

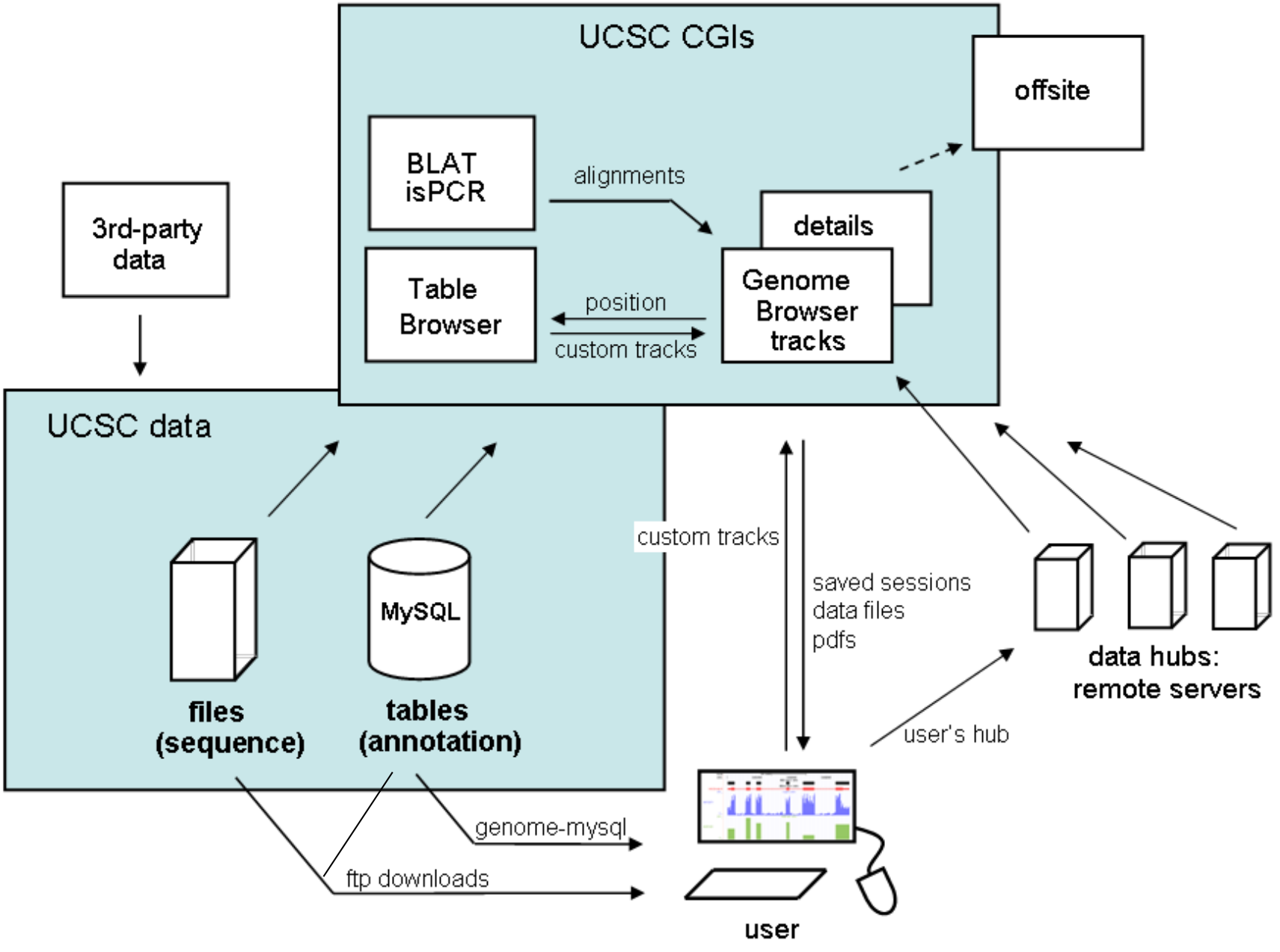
http:// **bit.ly/ucscMalaysia2018**

UCSC Genome Browser

Display engine for genomic annotations.
Consistent interface across genomes.

A tool for inquiry-driven discovery.

architecture



Training

YouTube training channel:  **bit.ly/ucscVideos**

On-site workshops (training link on main page):

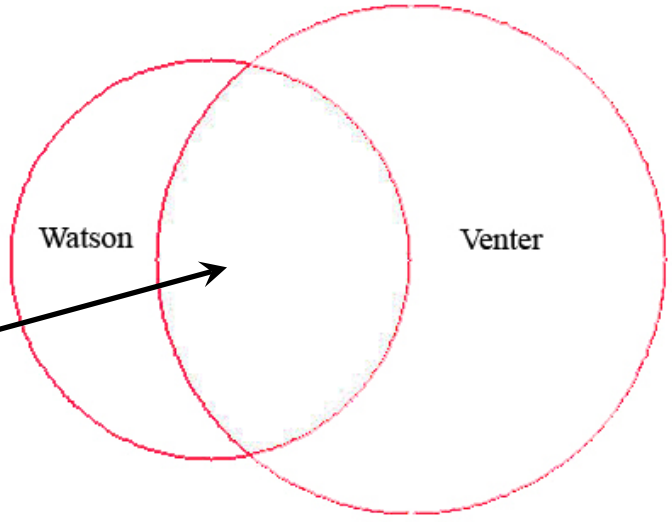
http:// genome.ucsc.edu/training/

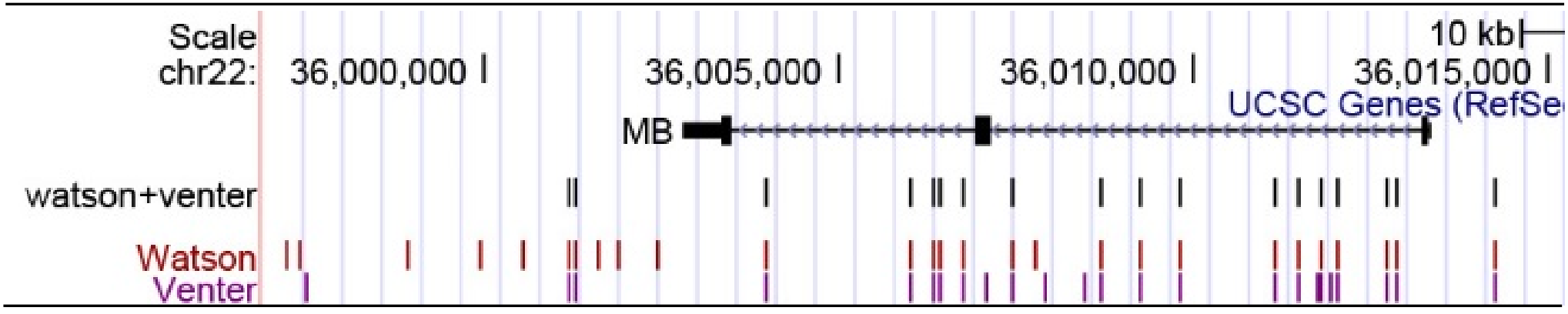
Variation

Watson: 2.06 million SNPs

Venter: 3.32 million SNPs

1.19 million in common





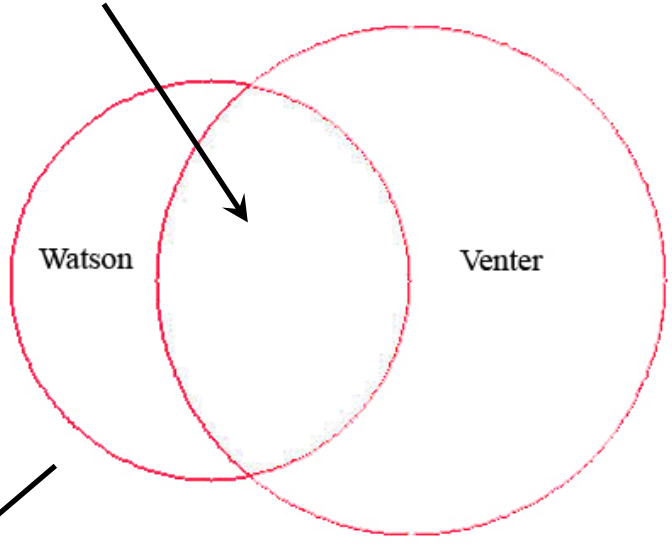
In 20 kb myoglobin region on chr22, Watson and Venter share 20 SNPs.
Watson has 9 unique SNPs, Venter 6.

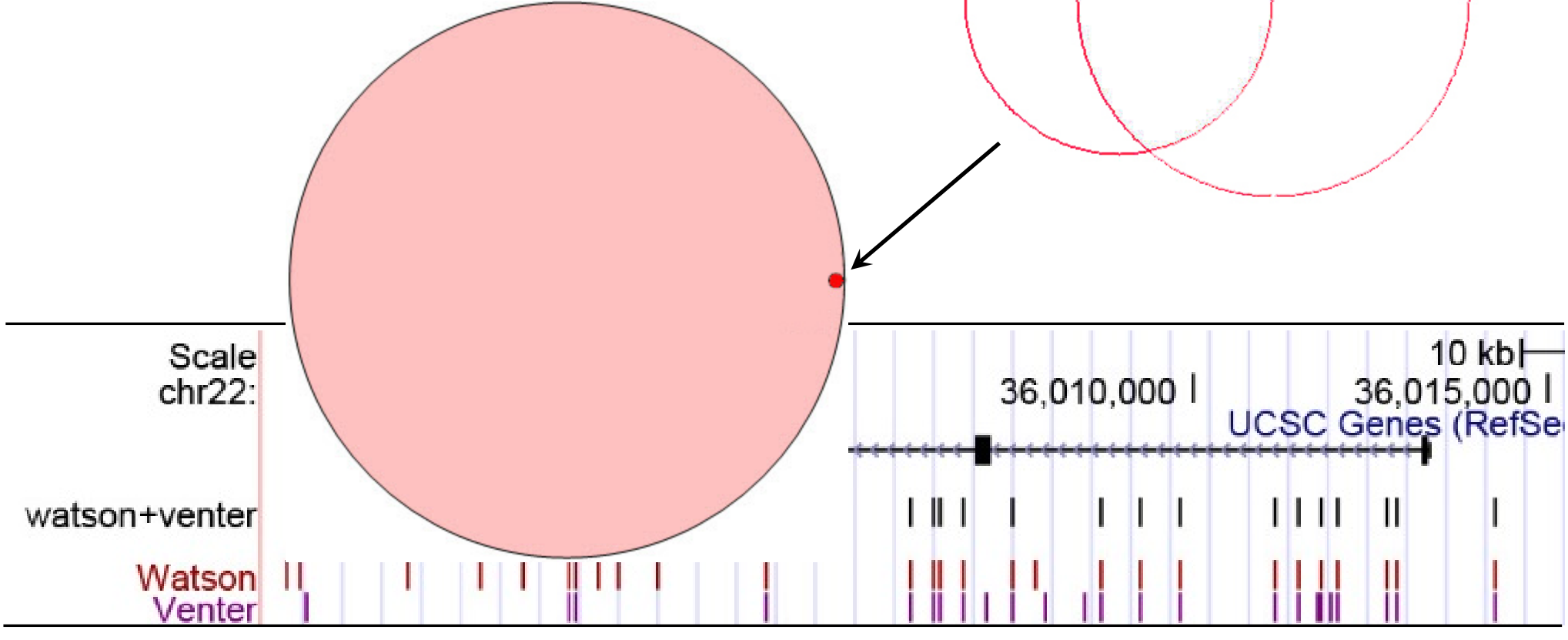
Variation

1.19 million
in common

Watson: 2.06 million SNPs

Venter: 3.32 million SNPs

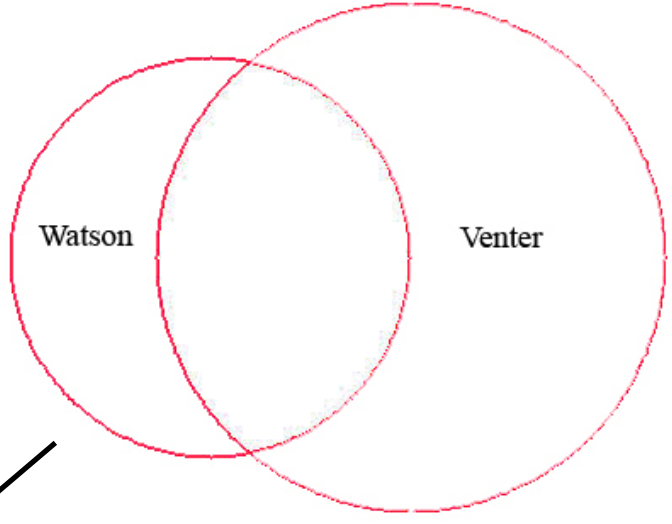




Variation

Watson: 2.06 million SNPs

Venter: 3.32 million SNPs



HGVS nomenclature

support for (some) HGVS:

NM_198056.2:c.1654G>T

NP_002993.1:p.Asp92Glu

NP_002993.1:p.D92E

BRCA1 Ala744Cys / BRCA1 p744

http://genome.soe.ucsc.edu/goldenPath/help/query.html#HGVS

= bit.ly/ucscHGVS

HGVS nomenclature

And now: HGVS output from Variation Annotation Integrator:

HGVS variant nomenclature

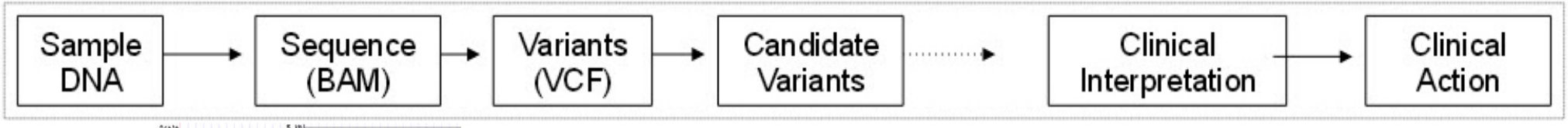
The Human Genome Variation Society (HGVS) has established a sequence variant nomenclature, an international standard used to report variation in genomic, transcript and protein sequences.

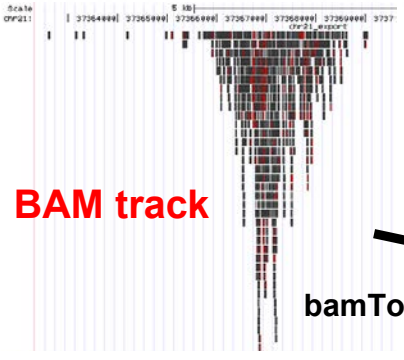
- ☐ Include HGVS genomic (g.) terms in output
- ☐ Include HGVS coding (c.) terms if applicable, otherwise noncoding (n.) terms, in output
- ☐ Include HGVS protein (p.) terms (if applicable) in output
- ☐ When including HGVS protein (p.) terms, add parentheses around changes to emphasize that they are predictions
- ☐ For variants that involve both a deletion and insertion, including multi-nucleotide variants, include the deleted sequence (e.g. show "delAGinsTT" instead of only "delinsTT")

Next-Gen Sequence Data

Data pipeline

Process Overview



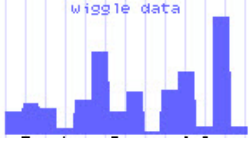


BAM track

bamToBedGraph

bedGraph track

bedGraphToBigWig



Wiggle track

bamToBigWig

slide modified from:

Tim Hubbard, King's College, London

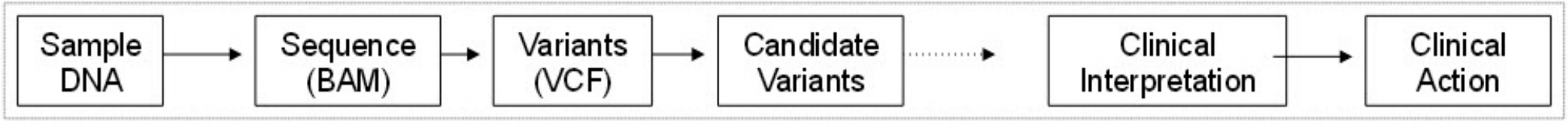
100K genomes (rare disease or cancer)

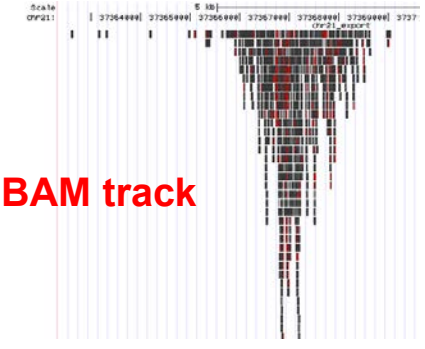
Genomics england

Next-Gen Sequence Data

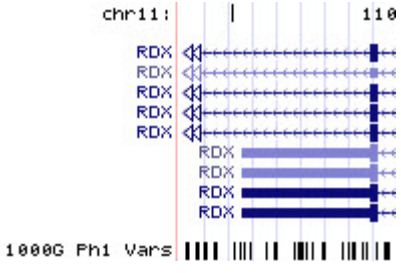
Data pipeline

Process Overview





BAM track



VCF track

Variant Annotation Integrator

slide modified from:

Tim Hubbard, King's College, London

100K genomes (rare disease or cancer)

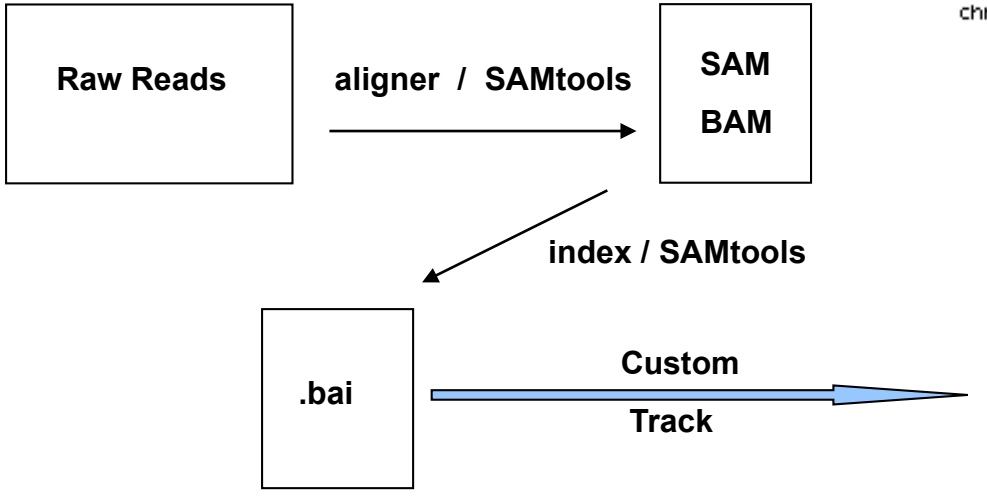
Genomics england

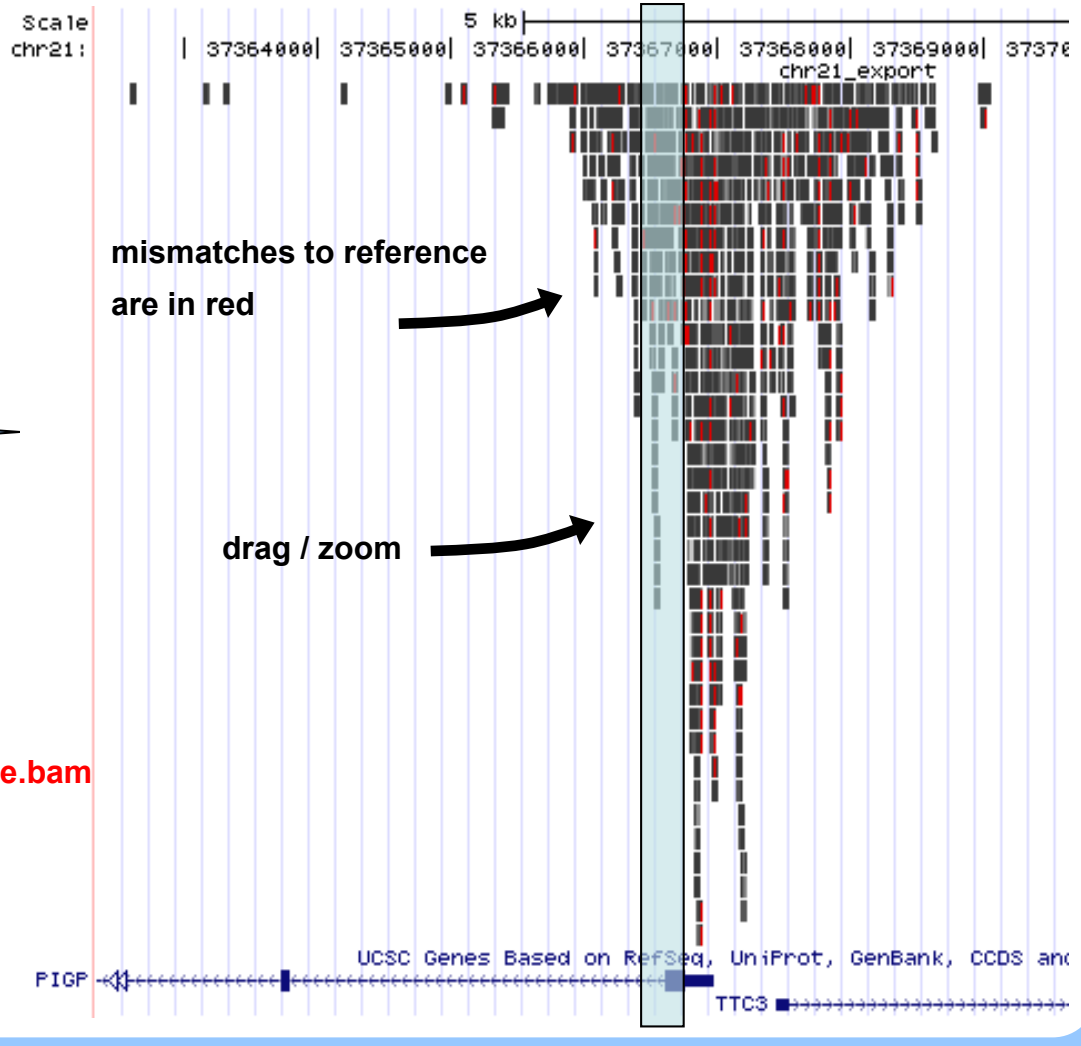
Next-Gen Sequence Data

How do you visualize your Next-Gen Sequencing data? -- BAM file

millions of short reads

files too large to upload (timeout)





Custom Track:

Make BAM, .bai files available to the web (http:, https: or ftp:)

Upload only the *location* of the data.

track name=trackName type=bam bigDataUrl=http://path/file.bam

The Browser fetches only tiny portion of the file

http://samtools.sourceforge.net/

sample ChIP-seq data courtesy Charles Nicolet, UC Davis

Next-Gen Sequence Data

zoom to base level

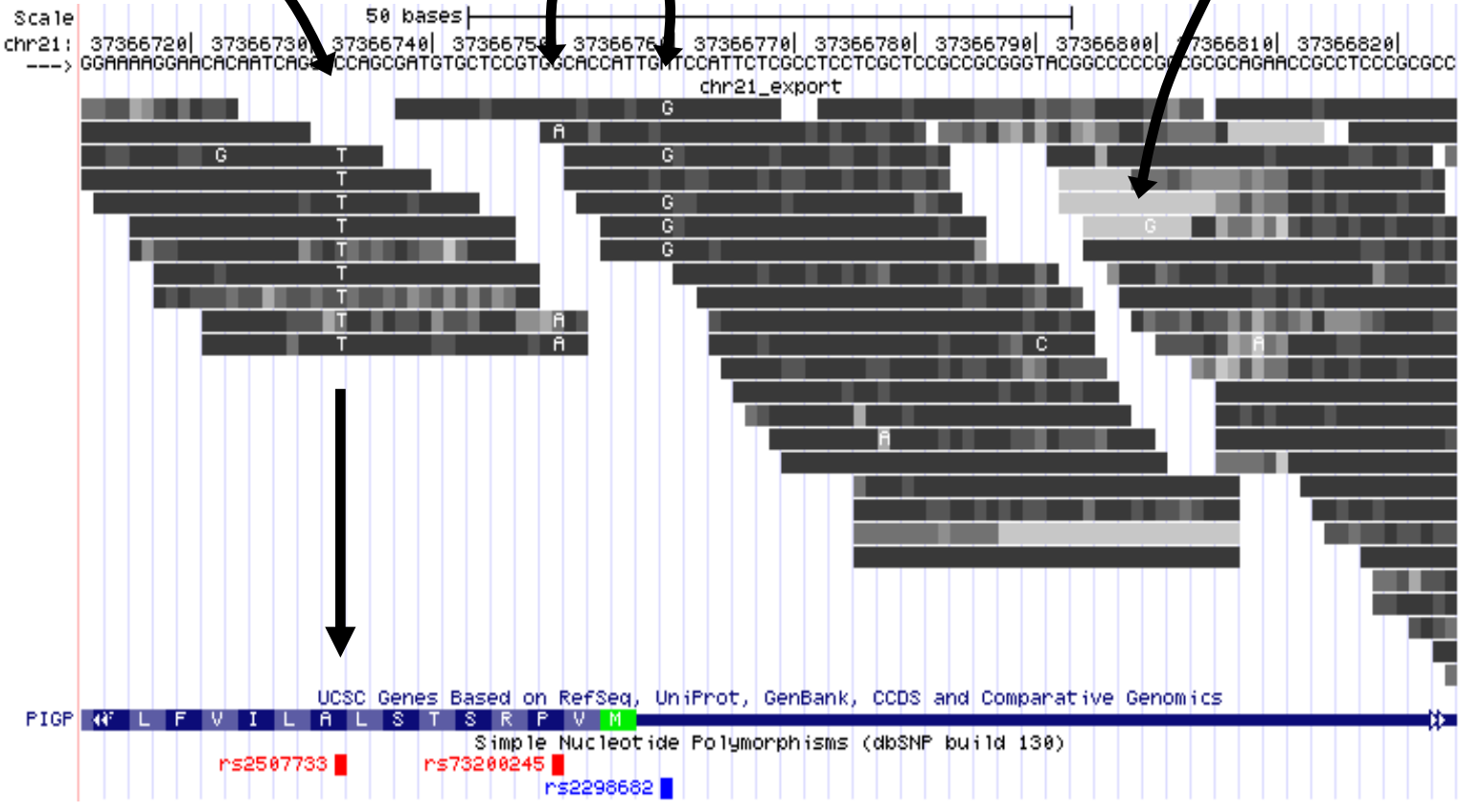
view alignment details

---

homozygous mismatch to reference is same as a known non-synonymous SNP

possible heterozygotes

lower quality scores shown in lighter color



UCSC genes

SNP track

Browser in exome sequencing

Display

- **Read alignments: BAM, CRAM**
- **Coverage: (BAM), wiggle**
- **Variants: pgSNP, VCF, (HGVS, rs#)**

Predictive

- **Variant Annotation Integrator**
  - **SIFT, PolyPhen, Mutation{Taster,Assessor} …**
- **Data tracks**
  - **Benign, Pathogenic**
  - **CNVs, SNPs**

live demo:

http:// genome-asia.ucsc.edu