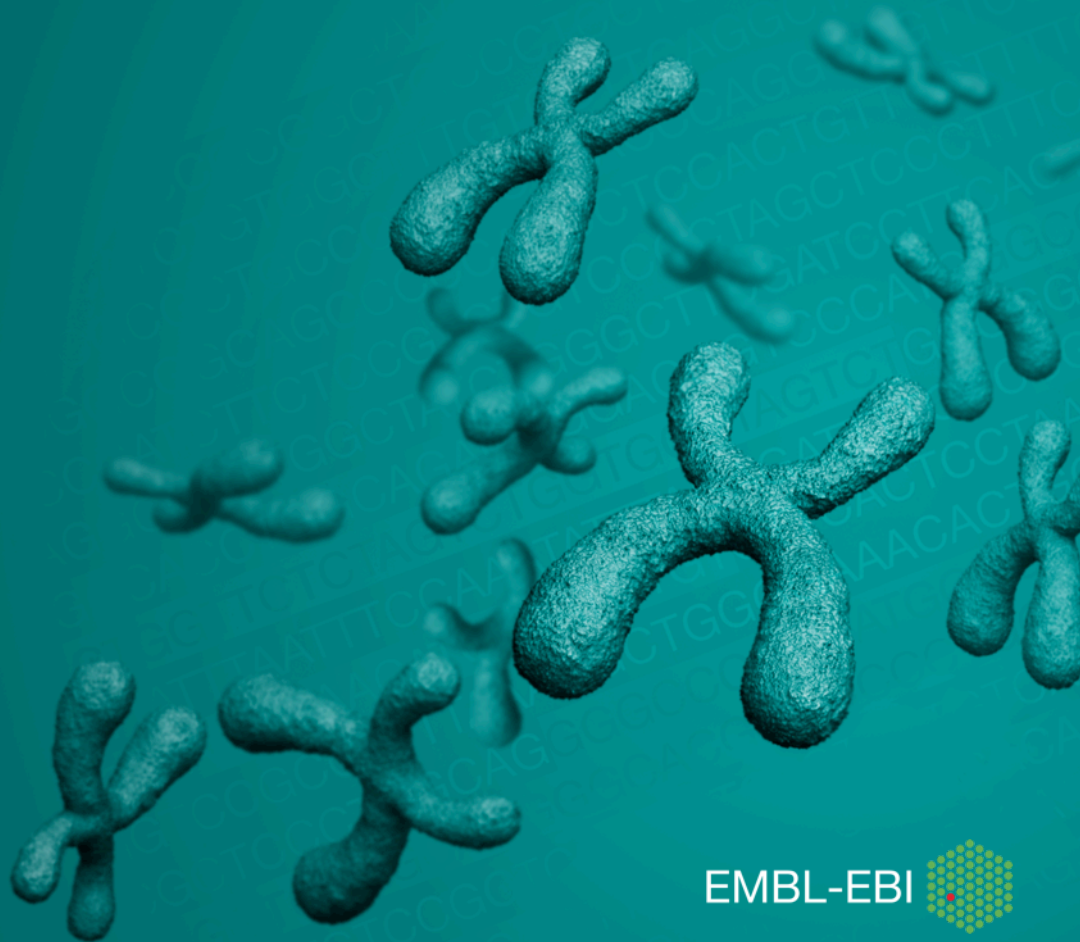


A High Quality Resequencing Analysis Pipeline for the NextGen Project Livestock Data

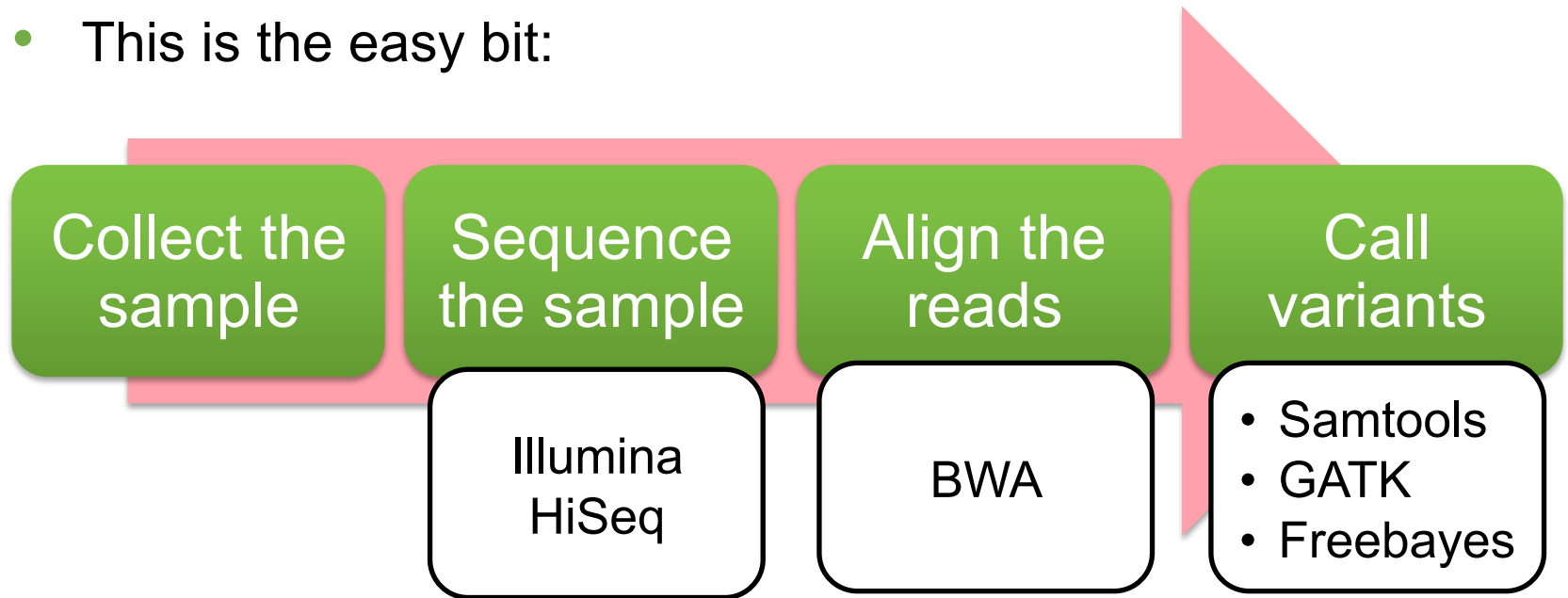
Ian Streeter

streeter@ebi.ac.uk



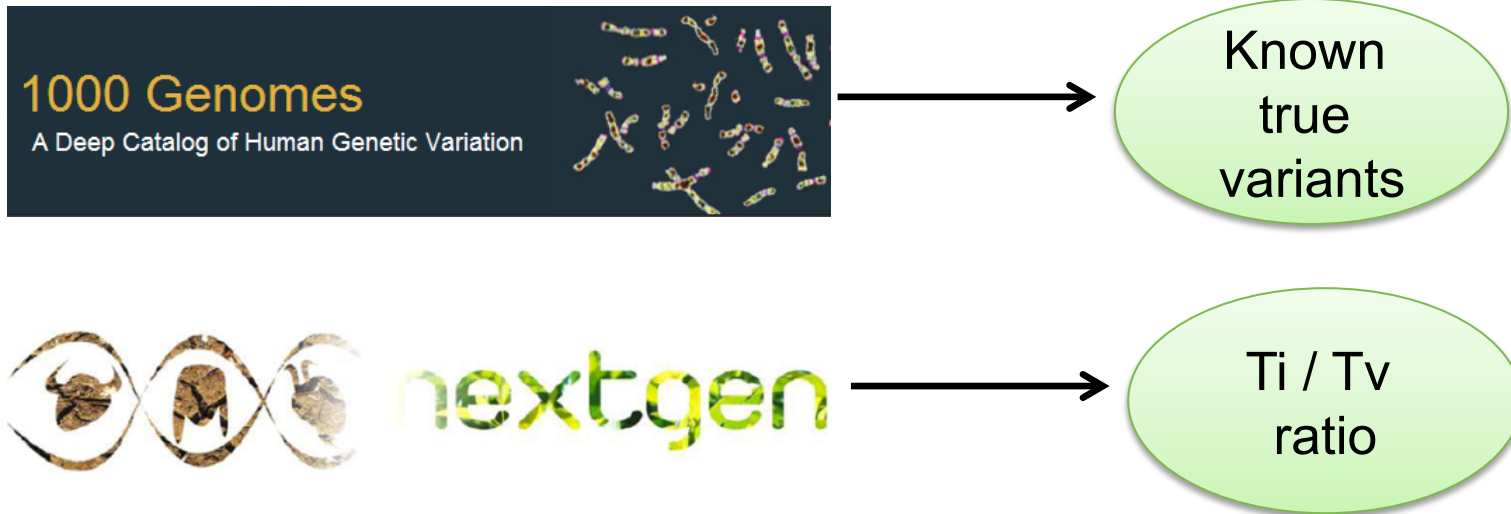
Genome Sequencing Pipeline

- This is the easy bit:



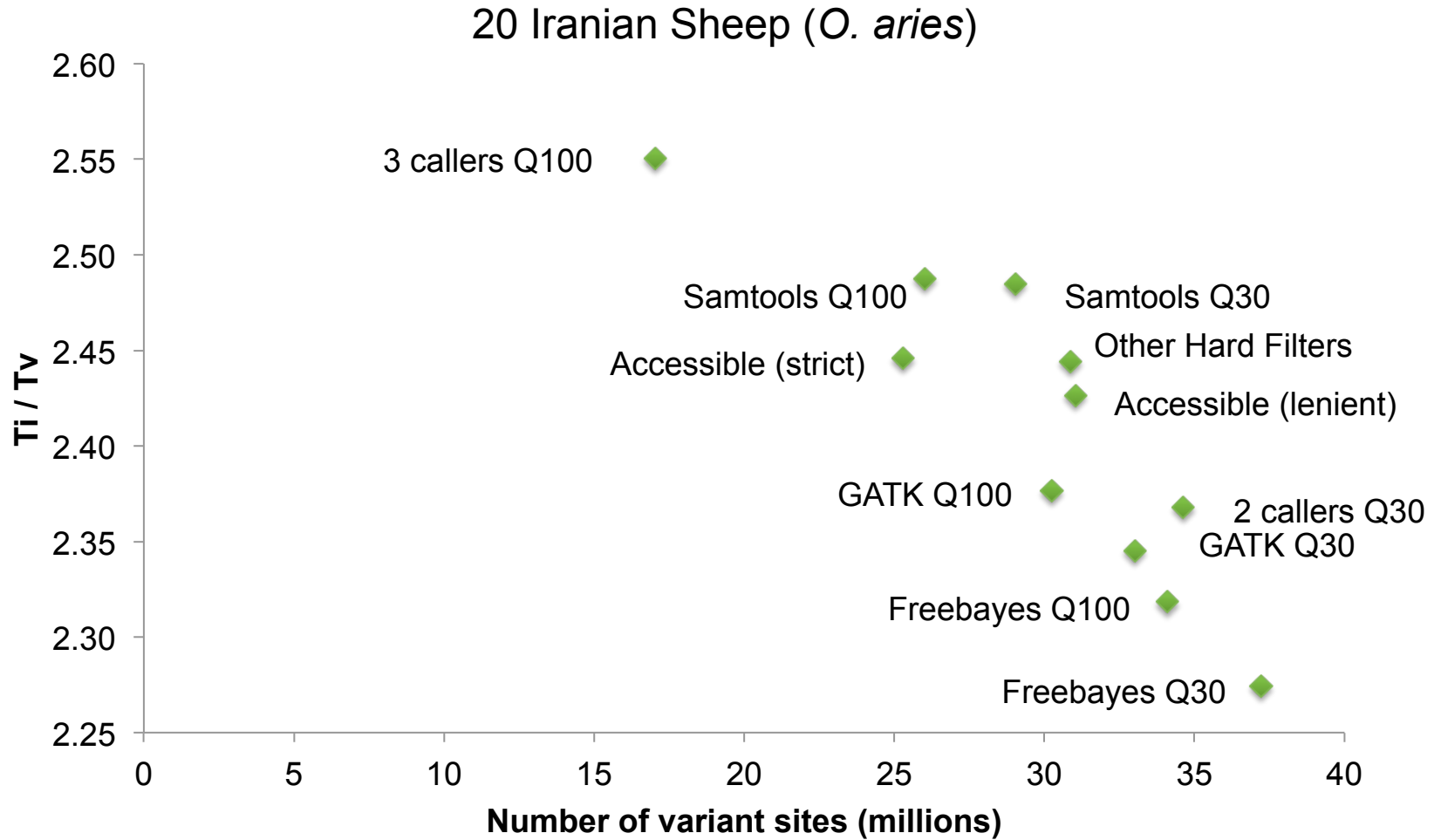
- How can you **assess / improve** your variant call set quality?

Assessing Variant Quality

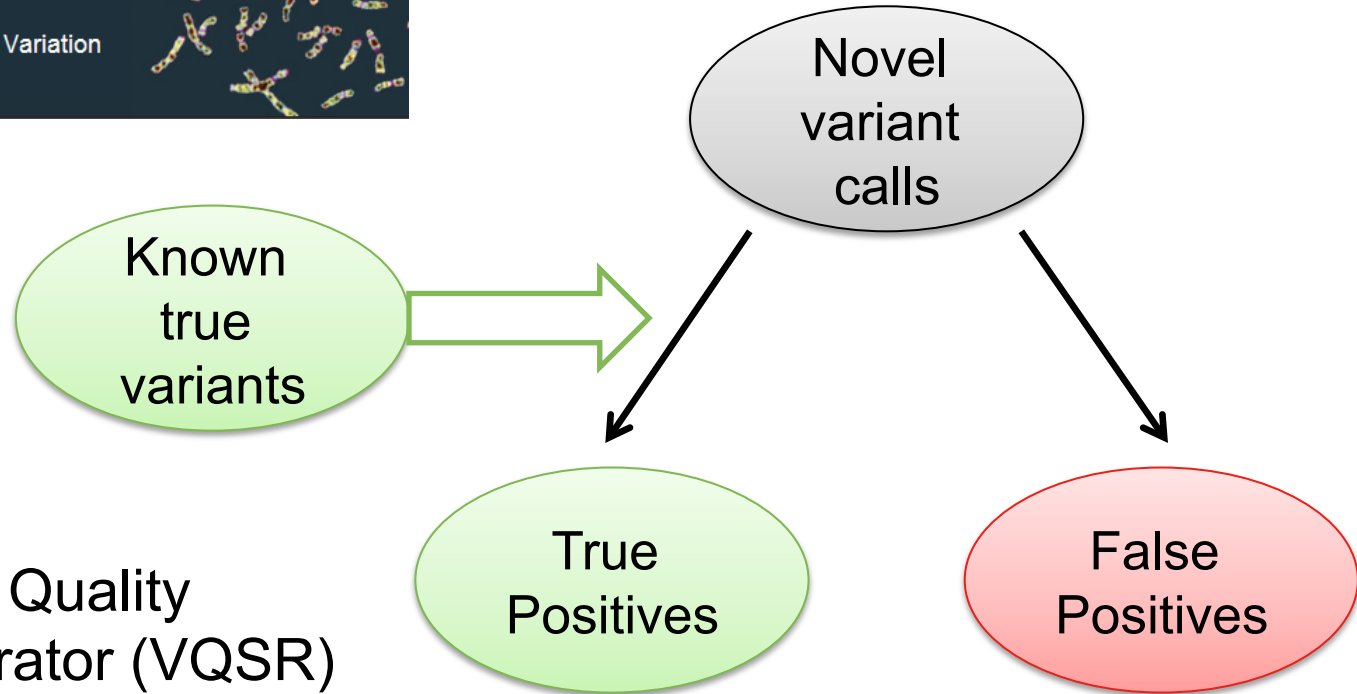
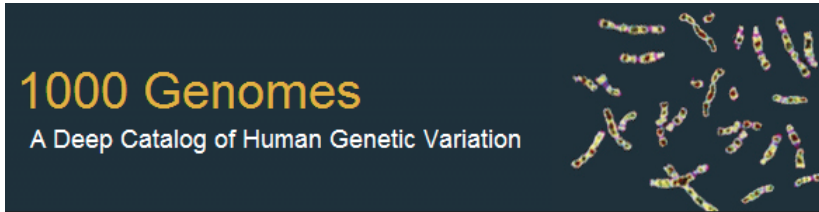


- Statistically more likely: A/C A/T C/G T/G
- Biologically more like: A/G C/T
- Good SNP set: High Ti / Tv
- Bad SNP set: Low Ti / Tv

Sheep SNP Variant Calls

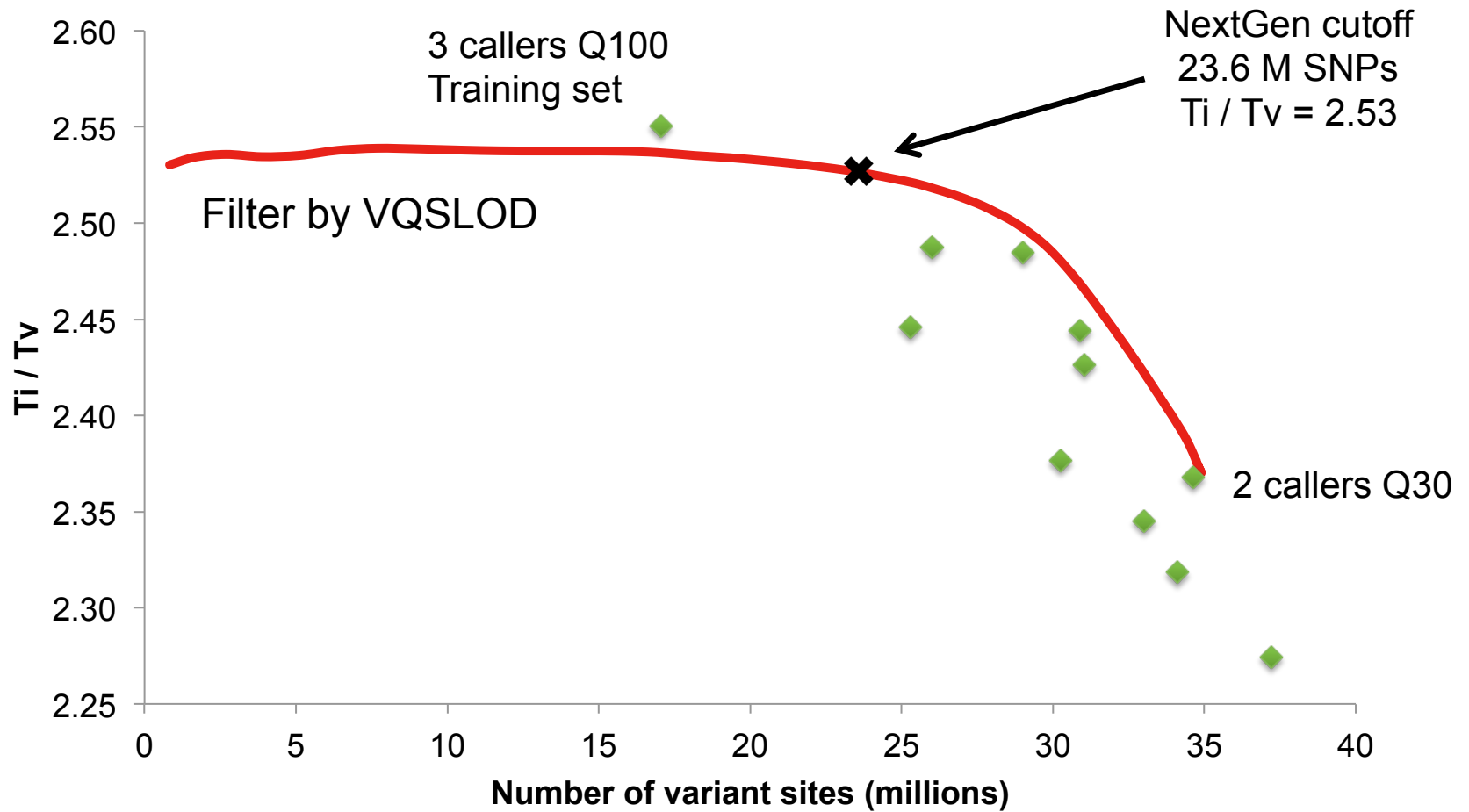


Variant Filtering by Truth Training Sets



- GATK Variant Quality Score Recalibrator (VQSR)
- U. Mich Support Vector Machine (SVM)

VQSR Filters



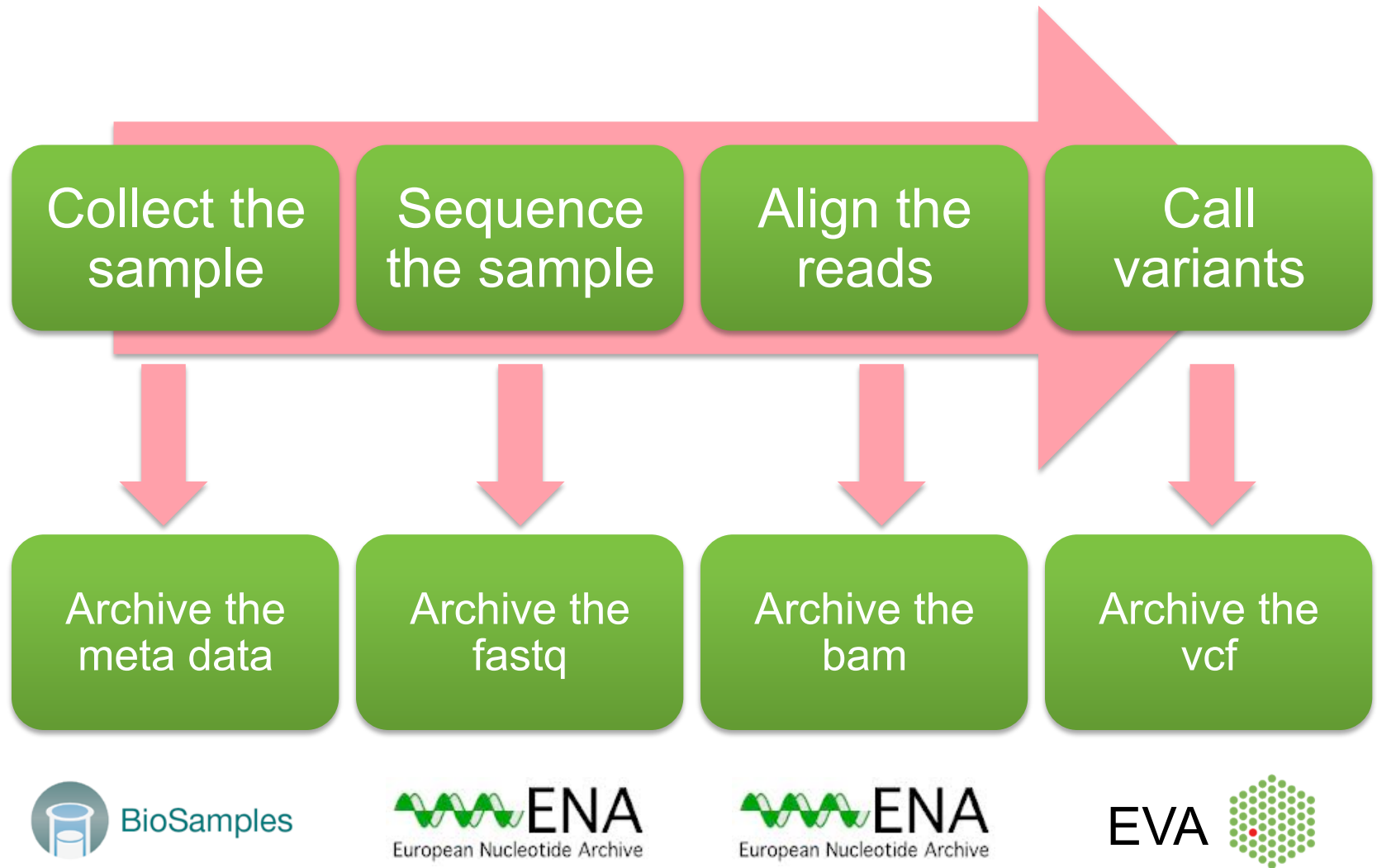
VQSR for SNPs and small indels

| Population | Samples | SNPs | Indels |
|-----------------|---------|--------|--------|
| Iran goat | 20 | 21.1 M | 1.80 M |
| Iran wild goat | 18 | 16.1 M | 1.36 M |
| Morocco goat | 161 | 29.6 M | 2.14 M |
| Iran sheep | 20 | 23.6 M | 2.21 M |
| Iran wild sheep | 14 | 26.6 M | 2.71 M |
| Morocco sheep | 161 | 40.7 M | 2.81 M |
| Uganda cattle | 25 | 27.2 M | 2.08 M |
| Iran cattle | 8 | 18.6 M | 1.59 M |

Interlude.



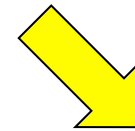
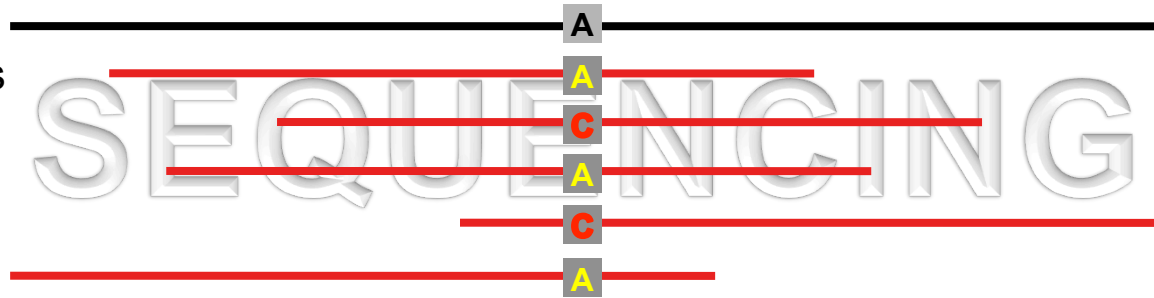
Genome Sequencing Pipeline



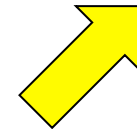
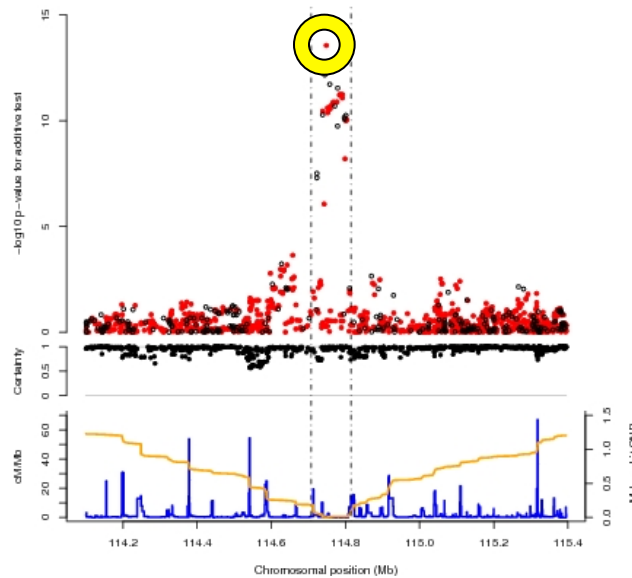
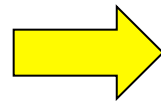
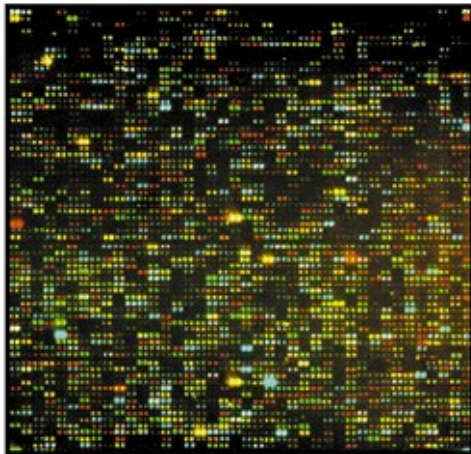
Variant Effect Predictor

Ve!P

Ref
Reads



SNPs
Indels
CNVs



Variant Effect Predictor



- www.ensembl.org
- Web interface or perl API



Search: for

e.g. **BRCA2** or **rat X:100000..200000** or **coronary heart disease**

Browse a Genome

The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Favourite genomes



Human
GRCh37



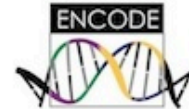
Mouse
GRCm38



Zebrafish
Zv9

★ [Change favourites](#)

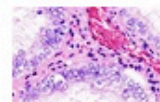
ENCODE data in Ensembl



Variant Effect Predictor



Gene expression in different tissues



Find SNPs and other variants for my gene

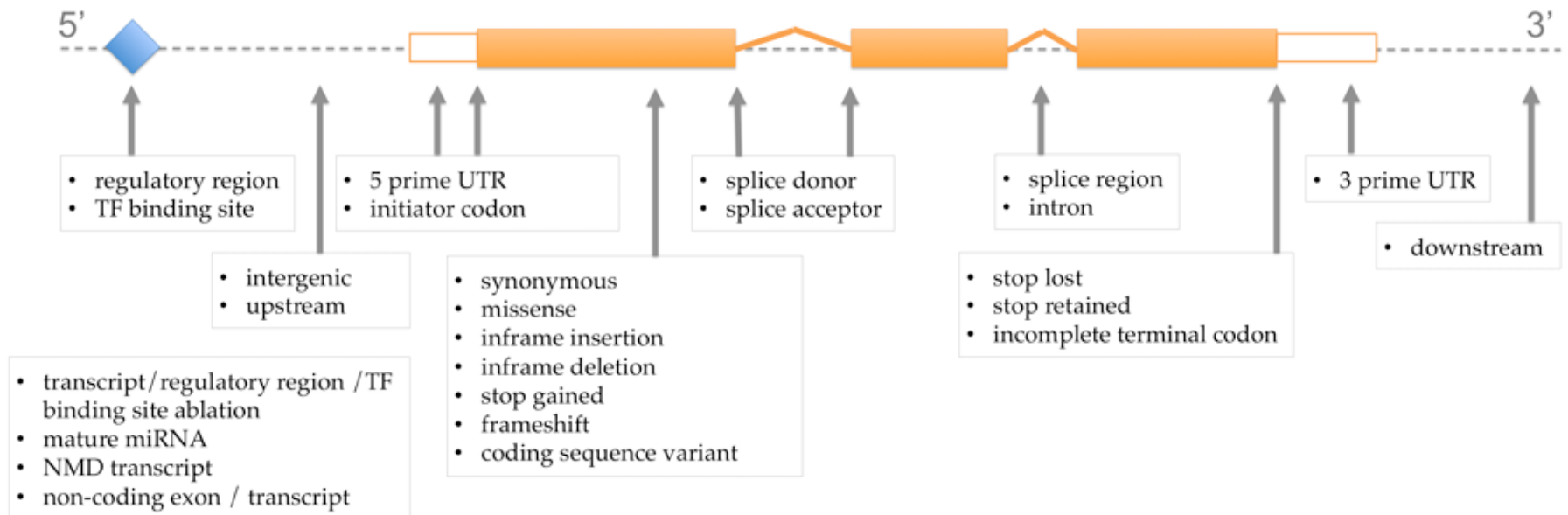
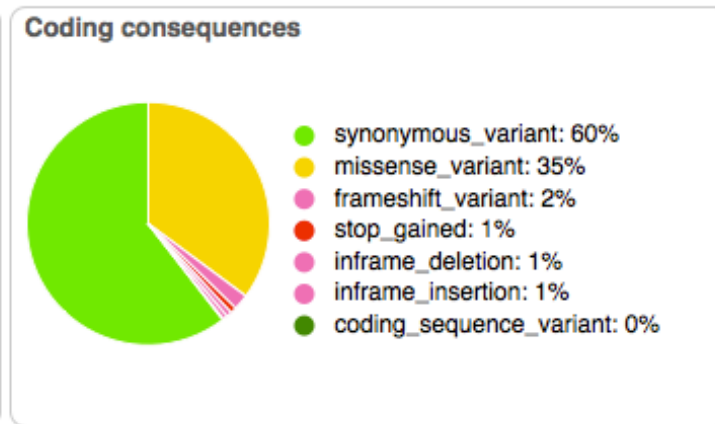
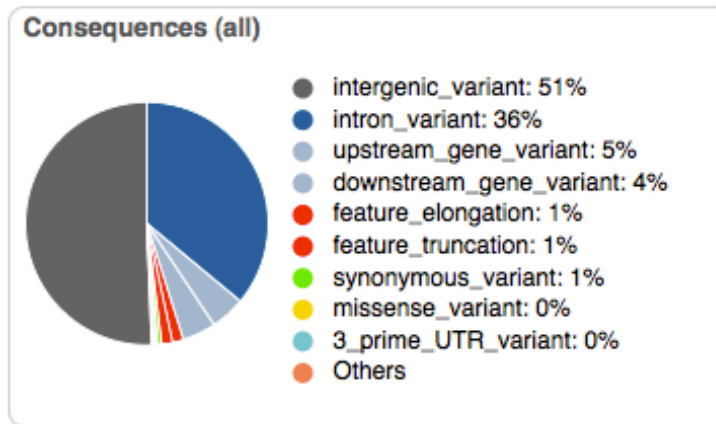
```
GTATATACATTC  
CCTRAAAGTCTT  
CTTCTAAATTCT  
GRAACATTTTCC
```

Retrieve gene sequence

```
GCCTGACTTCGCGTGG
```

Compare genes across species

Variant Effect Predictor



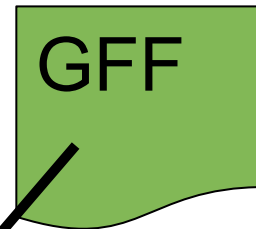
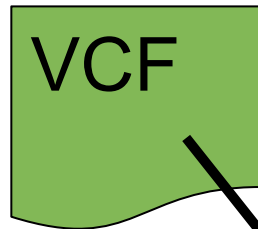
Variant Effect Predictor

- No Ensembl annotation? No problem.

nextgen



38.5 M variants
in goats



Eukaryotic Genome
Annotation Pipeline

Ve!P

Functional
consequences

The Future for Livestock Genomics?



- Sequence itself is not sufficient to understand which parts of animal genomes are functionally important

- Human genomics: more than just whole genome sequencing



International
Cancer Genome
Consortium



RNA-seq
profiling

Chromatin
accessibility
(DNase I)

Transcription
factor binding

Histone
modifications
(ChIP seq)

Methylome
(bisulfite seq)

Genome
conformation
(Hi-C)

Thanks!

- EMBL-EBI
 - Laura Clarke
 - Paul Flicek
 - Vertebrate Genomics
- NextGen partners:
 - CNRS, Grenoble
 - Cardiff University
 - Università Cattolica del Sacro Cuore, Piacenza
 - Parco Tecnologico Padano
 - EMBL-EBI



- EPFL, Lausanne
- Università Teramo
- Makerere University, Uganda
- INRA-Morocco
- Gorgan University, Iran
- CSIRO, Brisbane
- Genoscope, France