



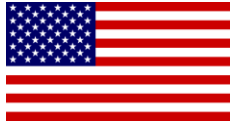
**Interbeef**  
**Dublin - 25 Nov 2014**

**John J. Crowley**



UNIVERSITY OF  
**ALBERTA**

# Genome Canada Project





# Genome Canada Project

- Involves 7 main Canadian breeds plus a hybrid plus 2 research populations
- Sequence 300+ influential animals
- Genotype ~18,000
- Feed efficiency (FE) and meat quality





# So far....

| Data Type                | n    |
|--------------------------|------|
| Sequence                 | 378  |
| HD                       | 4825 |
| 50k                      | 5007 |
| 50k with feed efficiency | 8027 |



# Sequencing

- ~8X coverage (1 or 2/breed @ 20+)
- Illumina HiSeq platform (and some SOLiD)
- Computing resources from WestGrid to simplify and speed up data analysis
  - 50 TB spinning disk storage (backed up to tape) on Silo storage system (3150 TB).
  - 30 core years and 6 TB on Jasper cluster (2880 cores, 355 TB).

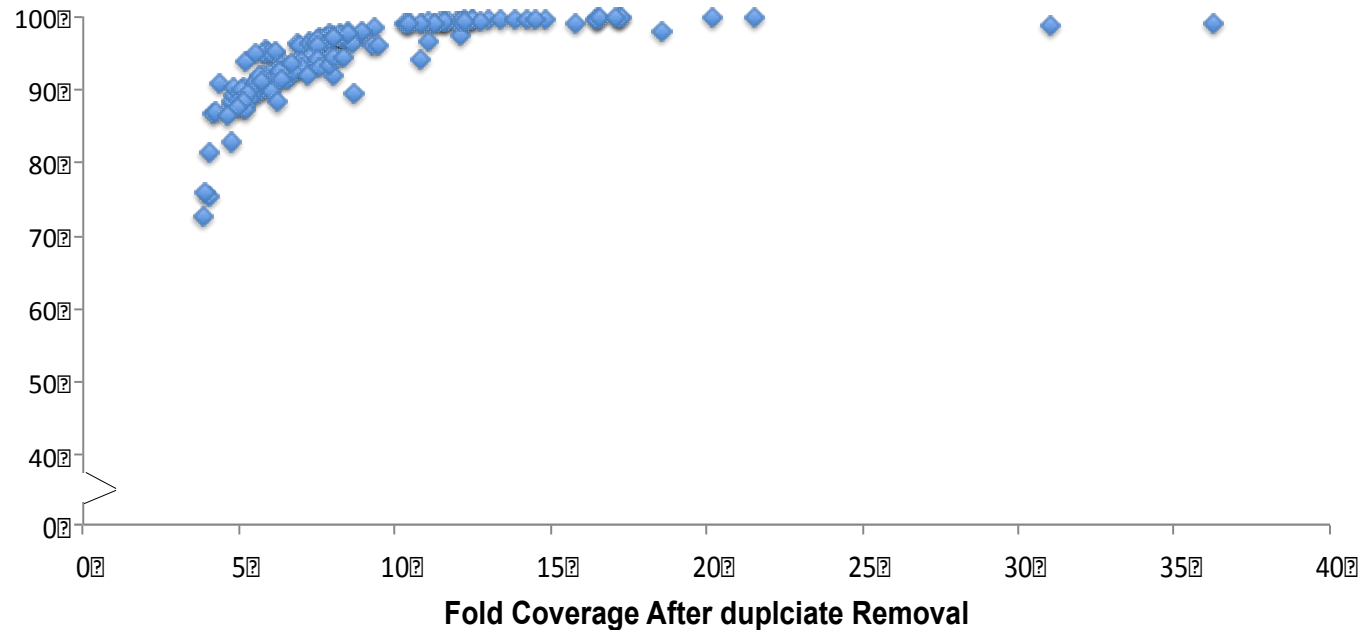




# Sequence processing tools

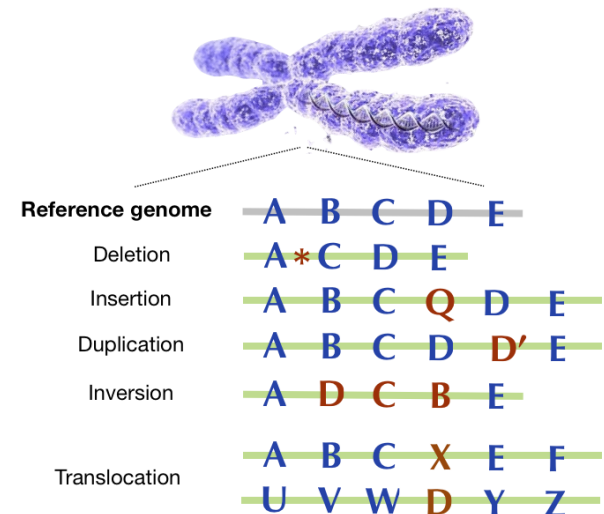
- **FastQC** – sequence data quality control checks
- **BWA** – mapping reads to reference.
- **Picard** – manipulating and summarizing BAM files, duplicate read detection.
- **SAMtools** – variant discovery and genotyping.
- **GATK** – variant discovery and genotyping.
- **Perl** and **Bash** scripts for submitting jobs, comparing sequence genotypes and reports

# Sequencing – concordance with HD



# Variant Discovery

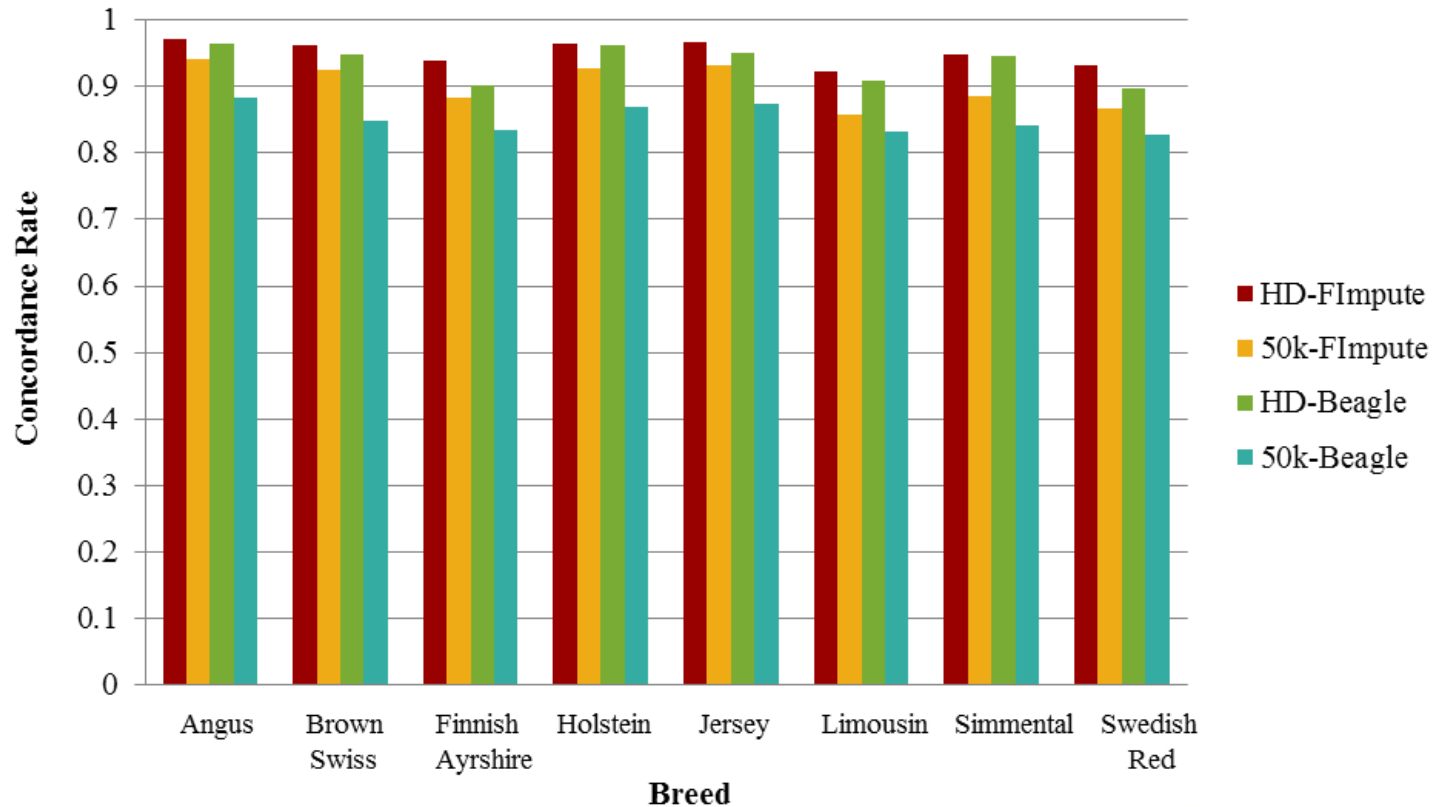
- Output of run 4.0 of 1000 bulls
  - 35,201,962 SNP
  - 1,714,895 INDELs
  - Other variants to be quantified
    - CNVs, inversions, translocations
    - Causative mutation should be in the data
    - Fine mapping of QTL







# Imputation to Sequence



Average accuracy (Concordance rate) of single breed whole-genome sequence imputation of 28.3 million SNP genotypes using FImpute and Beagle software



# Prediction Accuracy

| Dataset | n    | Breeds                       |
|---------|------|------------------------------|
| ANC     | 1946 | AN, CH                       |
| KIN     | 923  | Beef synt. 1 & 2, Dairy synt |
| PG1     | 4359 | HE, AN, RA, CH, M4, TX       |
| UoG     | 1972 | AN, CH, SM, LM, PD, GV       |

|     | GBLUP1      | GBLUP2      | BayesC      |
|-----|-------------|-------------|-------------|
| ADG | 0.35 (0.56) | 0.29 (0.57) | 0.33 (0.52) |
| DMI | 0.46 (0.70) | 0.37 (0.60) | 0.54 (0.71) |
| RFI | 0.35 (0.46) | 0.34 (0.51) | 0.24 (0.63) |

\*Highest accuracy in parenthesis



# Funders and Sponsors





# Questions?

