

Illumina Seminar Series

Munich 6. July 2009

Tim M Strom
Helmholtz Zentrum München &
Technische Universität München

Outline

- Introduction
 - Setup, Projects
 - Error rates
 - Data management

Tim M Strom

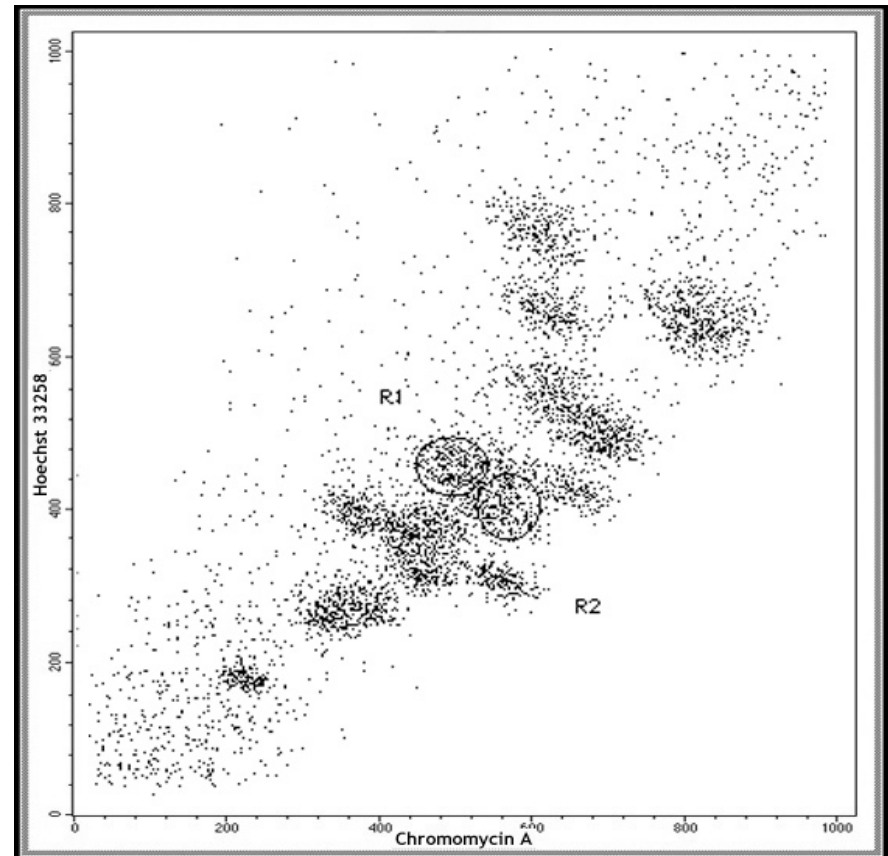
- Somatic mutations in Acute Myeloid Leukemia
- Sebastian Eck

Setup

- 2 GA II (30 runs, 100 GB mappable raw data)
(Storage 100 TB, 2 Servers 40 cores)
- Service at the Helmholtz Zentrum Munich and for external customers
- Synbreed: Whole-genome re-sequencing of cattle, chicken and maize -> SNP detection -> SNP arrays
- Mutation detection in ENU mutagenized mice
- Somatic mutations in AML and CLL

ENU mutagenized mice

- Linkage in 50 mice
- Chromosome sorting



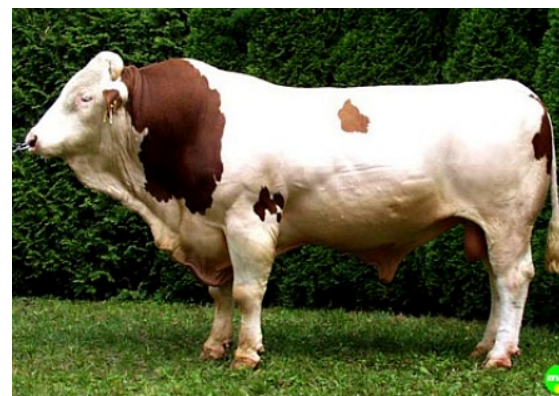
- Chromosome-specific library (~100ng)
- Sequencing of mutant and background strain

Chromosome sorting, Johannes Wienberg, Chrombios

SNP detection in cattle (genomic selection)

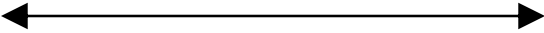
- Whole-genome resequencing of a single breeding bull

(50.000 inseminations in 2008)



- 24 GB, 7.4-fold read depth
- Comparison with the reference (bosTau4)
- 2.4 million SNPs, 2 million novel
- 9.318 coding SNPs
- False-positive detection rate 1%,
determined by genotyping 190 randomly selected SNPs
- Detection rate (homozygous) 74%, (heterozygous) 30%
determined by comparison with 50k SNP array

Error rate

Reads  Reference sequence

Sequencing errors (0.5-1.5%)
(depending on cluster density)

+

Genomic variability (0.1% in humans)

+

Alignment errors

!=

False-positive SNP detection rate (1-4%)

Wanted: tuning of the trade-off between error rate
and sequence output

Error rate

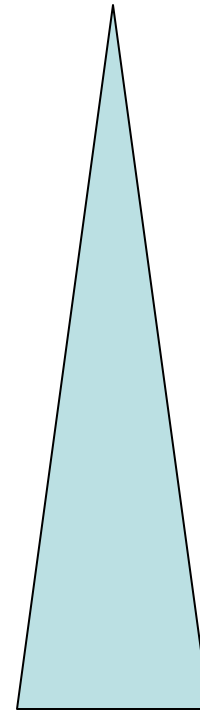
ChIP-Seq

Expression analysis

SNP detection

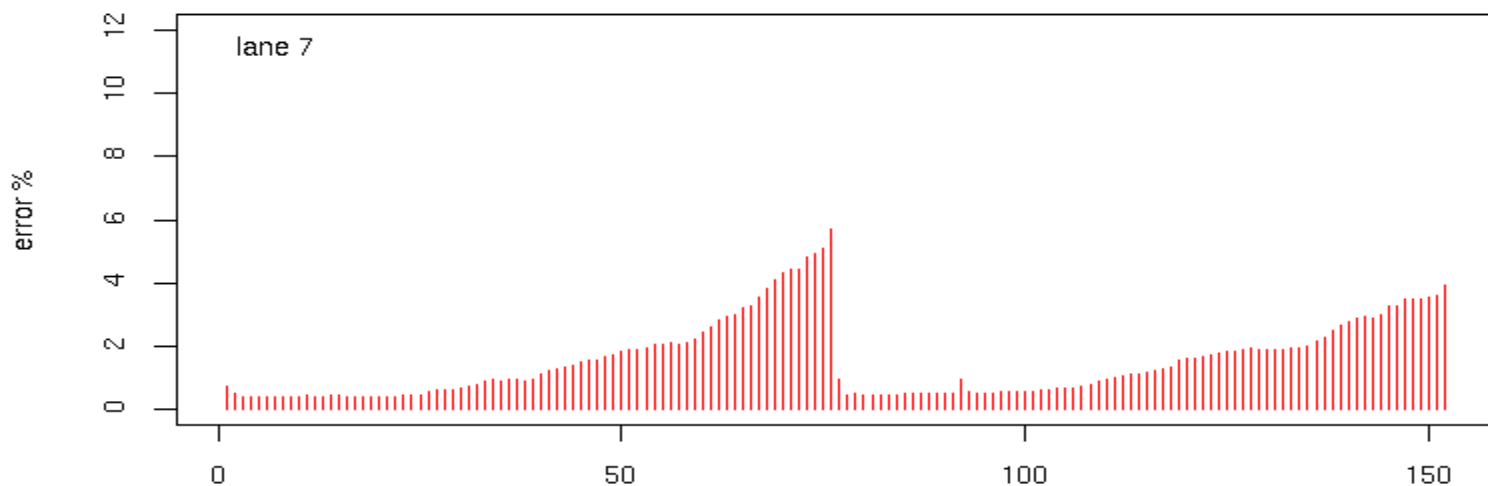
Mutation detection

De novo assembly

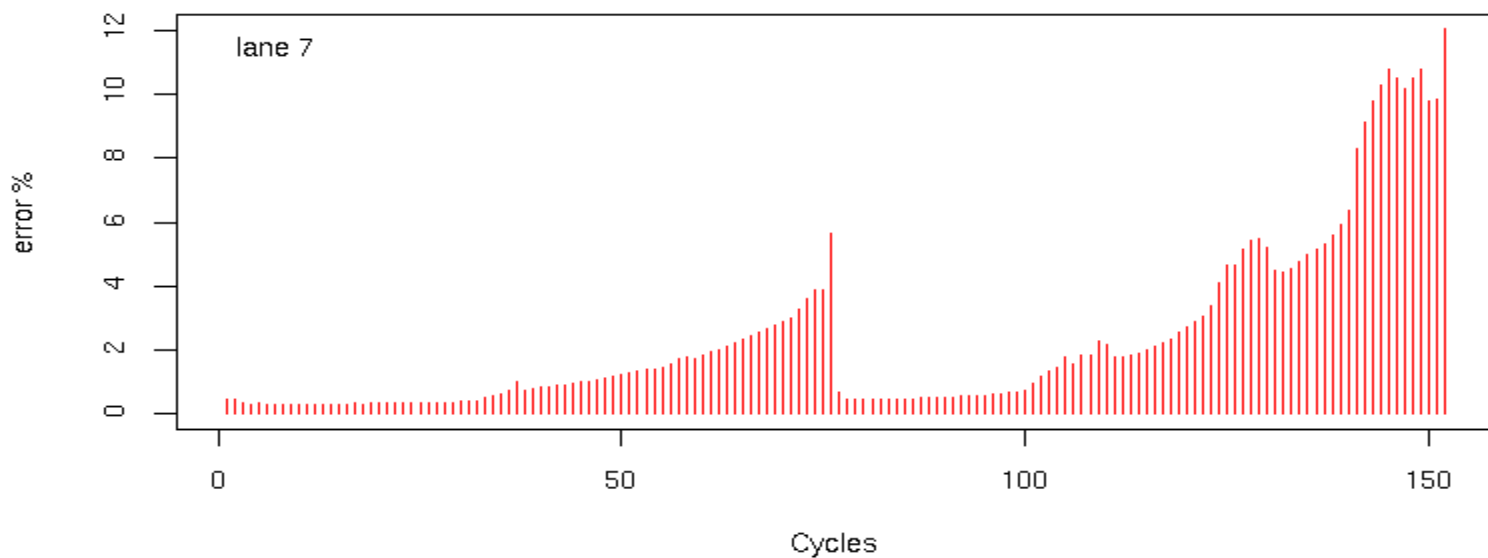


Error rate: Sequencing error

At least 10-fold higher than in capillary sequencing



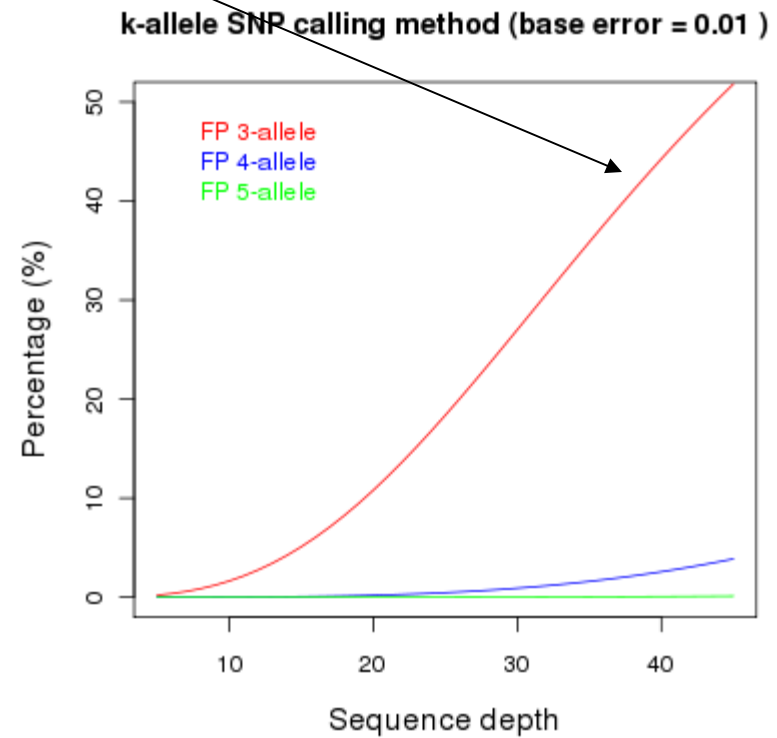
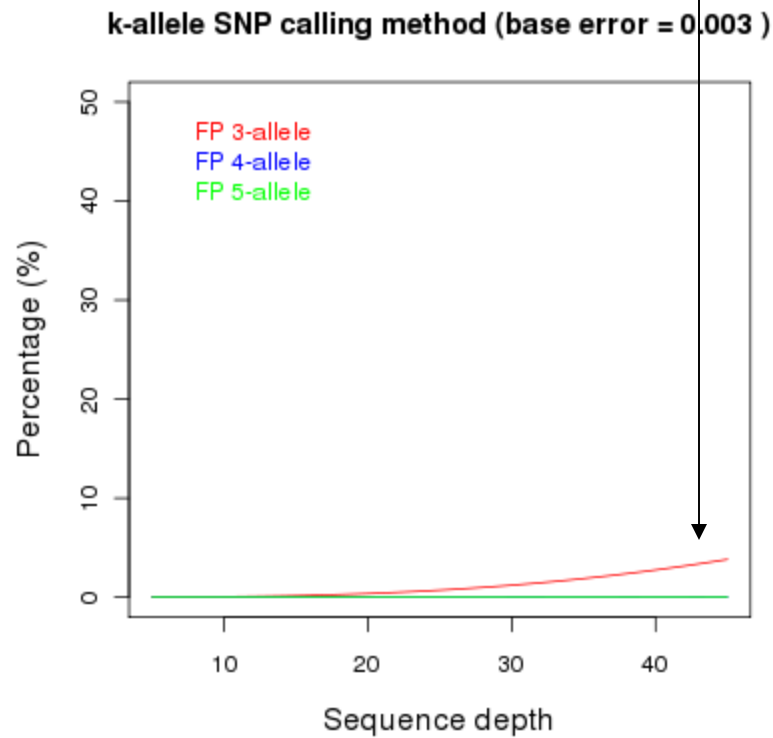
Read 1
1.21%
Read 2
1.12%



Read 1
0.89%
Read 2
2.61%

Error rate: false positive detection rate

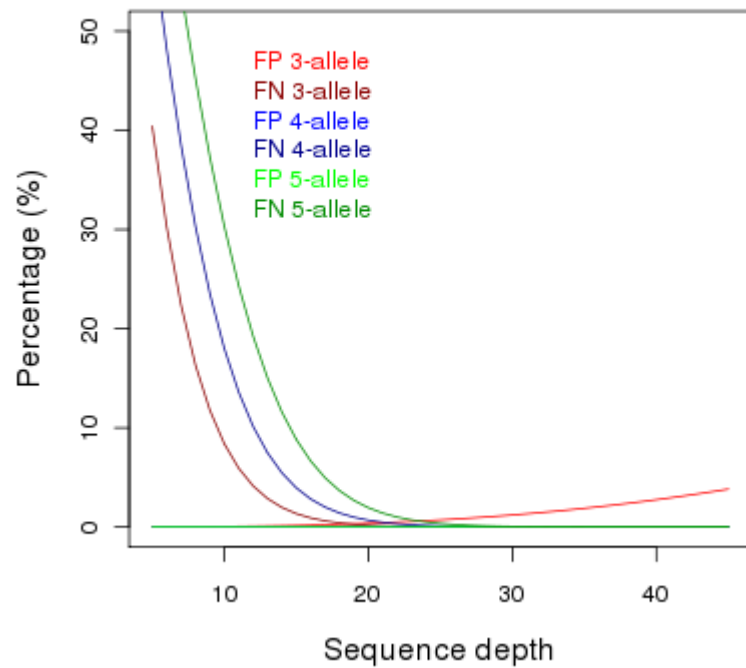
SNPs covered by at least 3 alleles



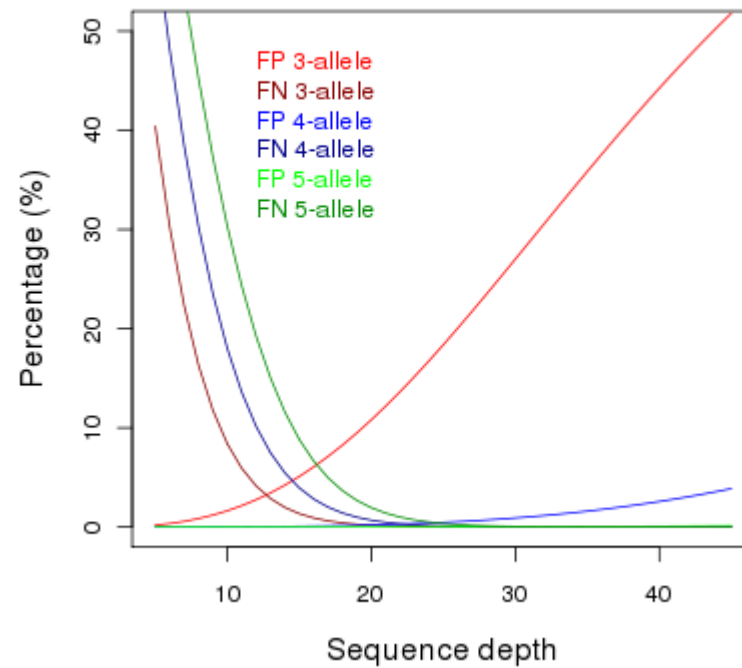
According to H Li, J Ruan and R Durbin, Genome Research, 2008

Error rate: false negative detection rate

k-allele SNP calling method (base error = 0.003)



k-allele SNP calling method (base error = 0.01)



-> Sequence depth ≥ 30

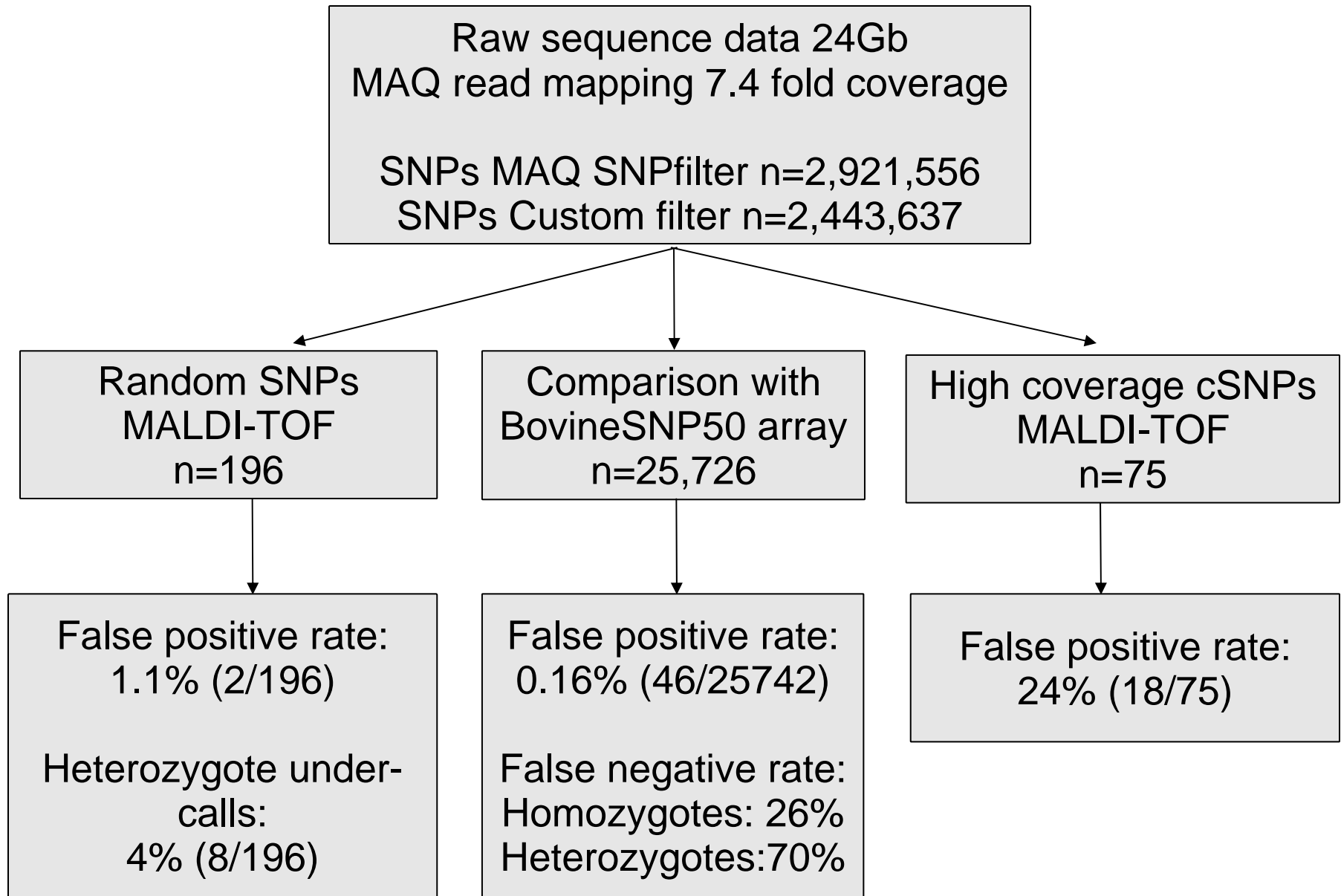
-> Error rate $< 1\%$ or Pred-like quality score ≥ 25

According to H Li, J Ruan and R Durbin, Genome Research, 2008

Error rate: Systematic error

```
M.A.Q Viewer
GGCTTCCTCCCTCTTTGGGTAGCCCTCAAGCCACTGGACTTTGAAGTCAGCAGGTAATTCTCCAAATGGAAGAACTTTTTTTTTTTTTTTTAAAGGCAGA
GGCTTCCTCCCTCTTTGGGTAGCCCTMAAGCCACTGGACTTTGAAGTCAGCAGGTAATTCTCCAAATGGAAGAACTTTTTTTTTTTTTTTTAAAGGCAGA
GgCtt.      tgggggaaacccaaAagcccccggactTTAAGTCA  cAgGtAAtTCTCCaAATGGAAGAACTTTTTTTTTTTT
TCTTCa     tggggatagacgtAaAGccActGgAcTTTgAAGtCA  gaggtaAtTctccaAAagGgAgaActTTTTTTTTTT
gGCTTc     ttGggtAgCccaaAAGCCACTGGACTTTGAAGTCA  caGgtAATTCTcCAAAATGGAAGAACTTTTTTTTTTT
GGCTTC     ttGggtagccctAaAGccActGGACTTTGAAGTCA  cAgGtAAtTCTcCAAAATGGAAGAACTTTTTTTTTTT
GGCtTC     tttgggtagCcoctAAGCCCaCTGGACTTTGAAGTCA  aAgGTAATTcTcCAAAATGGAAGAACTTTTTTTTTTT
GGCTTC     atggggaaacccaaAAGccActGGAcTTTGAAGTCA  caggtaAATTCTcCAAAATGGAAGAACTTTTTTTTTTT
GGCTTC     tttGggtAgcCctAaAGCCACTGGACTTTGAAGTCA  aaggtaAAatccccAAAaGGAAaAACTTTTTTTTTTT
GGCTTC     ttggggtacccctAaAGCCACTGGACTTTGAAGTCA  caggtaAaTtcccaAAagGaaAaActTTTTTTTTTT
GGCTTC     tttGggtAagccctAaAGCCACTGGACTTTGAAGTCA  caggtaAtTcttcaAAtgGaaAgaActTTTTTTTTTT
GGCTTC     ttggggtAgccctAaAGCCACTGGACTTTGAAGTCA  aaggtaATtCTcCaaAtGgAAGAAcTTTTTTTTTTt
GGCTTC     TTTGGGgAgcCCTCAaaccCaCtGgAcTtTgAaGTca  aAgGTAAtTctcCAAAATGGAAGAAcTTTTTTTTTTT
GGCTTC     tggggagccctAaAGCCaCtGGACTTTGAAGTCAG  aagtaATTtTcCAAAATGGAAGAACTTTTTTTTTTT
GGCTTC     tgggaaacccaaAAGCCaACTGGAcTTTGAaGTCAG  aggtaatTctccaAATGGAAGAACTTTTTTTTTTTtTTT
GGCTTC     ttGggtagcccaAaAGCCACTGGACTTTGAAGTCAG  AggTaATTcTccAAATGGAAGAACTTTTTTTTTTTT
GGCTTC     ctgggtaccccaAaAGCCACTGGACTTTGAAGTCAG  cggtaAtTatccaaaTgGaaGaaactTTttTTtTttT
GGCTTC     tgggaaaccccaAaAGCCACTGGACTTTGAAGTCAG  agGtaAtTctcCAAAATgGaaAgaActTTTTTTTTTTT
GGCTTC     ttgggtAgCccaaAaAGCCACTGGACTTTGAAGTCAG  aGgaAAttctcCAAAAGgGaaAgActTTTTTTTTTTT
GGCTTC     tgGgaaagCccaaAaAGCCCaCTGGACTTTGAAGTCaG  agGTAATTcTcCAAAATGGAAGAACTTTTTTTTTTTT
GgCTTC     ttgggaaacCcaaaAAGCCaACTGGaCtTTTGAAGTCAG  agGTaATTctcCAAAATGGAAGAACTTTTTTTTTTTT
GGCTTC     gtgggtAaacccaaaAgCcActtGacTTTGAAGTCAG  aggtAAtTCTcCAAAATgGAAgAACTTTTTTTTTTTT
GGCTTC     ttgggaaacccctAaAGCCACTGGACTTTGAAGTCAG  AgGTAATtCTcCAAAATGGAAGAAcTTTTTTTTTTT
GGCTTC     TTGGGTAGCCCTCAAGCCACTGGAcTTTgaagtcaG  AGGTaATTCTcCAAAATGGAAGAACTTTTTTTTTTTT
GGCTTC     TTGGGTAGCCCTCAAGCCACTGGACTTTGAaGTcAg  agGtaAtTcTcCAAAATGGAAGAACTTTTTTTTTTTT
GGCTTC     TTGGGTAGCCCTCAAGCCACTGGACTTTgAAGTCaG  aggtaAttctccAAAgTGAAGaaTTTTTTTTTTTTT
GGCTTC     TTGGGTAGcCCTCAAgCcaCTGgACTTtgaagtcaG  aggtaAtTctcAAAATGGAAGAACTTTTTTTTTTTT
GGCTTC     TTGGGTAGCCCTCAAGCCACTGGACTTTGaaGTcAg  aggtaatTCTcCAAAATGgAAGAACTTTTTTTTTTTT
GGCTTC     TTGGGTAGCCCTCAAGCCACTGGACTTTGAAGTcAg  agGTAATTCTcCAAAATGGAAGAACTTTTTTTTTTTT
GGCTTC     tgggtacCctAaAGCCCaCTGGaCtTTTGAAGTCAGC  GGTaATTCTCCAAATGGAAGAACTTTTTTTTTTTT
GGCTTC     agggaaacccataAAGCCACTGGACTTTGAAGTCAGC  GGAATCTCTTTTTCTTTTTCTTTTTCTTCAAAAa
GGCTTC     gGggtacCccaaAaAGCCACTGGACTTTGAAGTCAGC  gggAaAttctccaAAagGaaAaActTTTTTTTTTTT
GGCTTC     gGggaAgcccaaaAaAGCCACTGGACTTTGAAGTCAGC  agtaAatctccAAATgGaaAaActTTTTtTTTTtTt
GGCTTC     tGgtAcCccctAaAGCCACTGGACTTTGAAGTCAGC  aggaatttctcaAAagggaAACTTTTTTTTTtTt
GGCTTC     ggggcaacCcaaaAaAGCCACTGGACTTTGAAGTCAGC  gGtaAtTcTCCAAATGGAAGAACTTTTTTTTTTTT
GGCTTC     ggggtagccctAaAGCCACTGGACTTTGAAGTCAGC  ggtAAtTctacAAATGGAAGAAcTTTTTTTTTTTTT
GGCTTC     ggggaacCcaaaAaAGCCACTGGACTTTGAAGTCAGC  ggtAAtTatCcaAAATGGAAGAACTTTTTTTTTTTT
GGCTTC     gGggtagcCctAaAGCCACTGGACTTTGAAGTCAGC  ggtAaTTctcCAAAATGGAAGAACTTTTTTTTTTTT
GGCTTC     ggggtagCcoctAaAGCCACTGGACTTTGAAGTCAGC  ggtAaTctcCaAAAgGaaAgAcTTTTTTTTTTTTT
GGCTTC     tgggtAgccCtAaAGtCACTGGACTTTGAAGTCAGC  gtaAaTctccCAAAAGaaAaTTTTTTTTTTTTTT
gGCTTCct.  ggggaaagccctAaAgCCACTGGACTTTGAAGTCAGC  ggtAAtTctcCAAAATGGAAGAAcTTTTTTTTTTTTT
Read: 173930241-l_Index_Read1:5:29:1059:496#ATCACG/1[30]IG 29
```

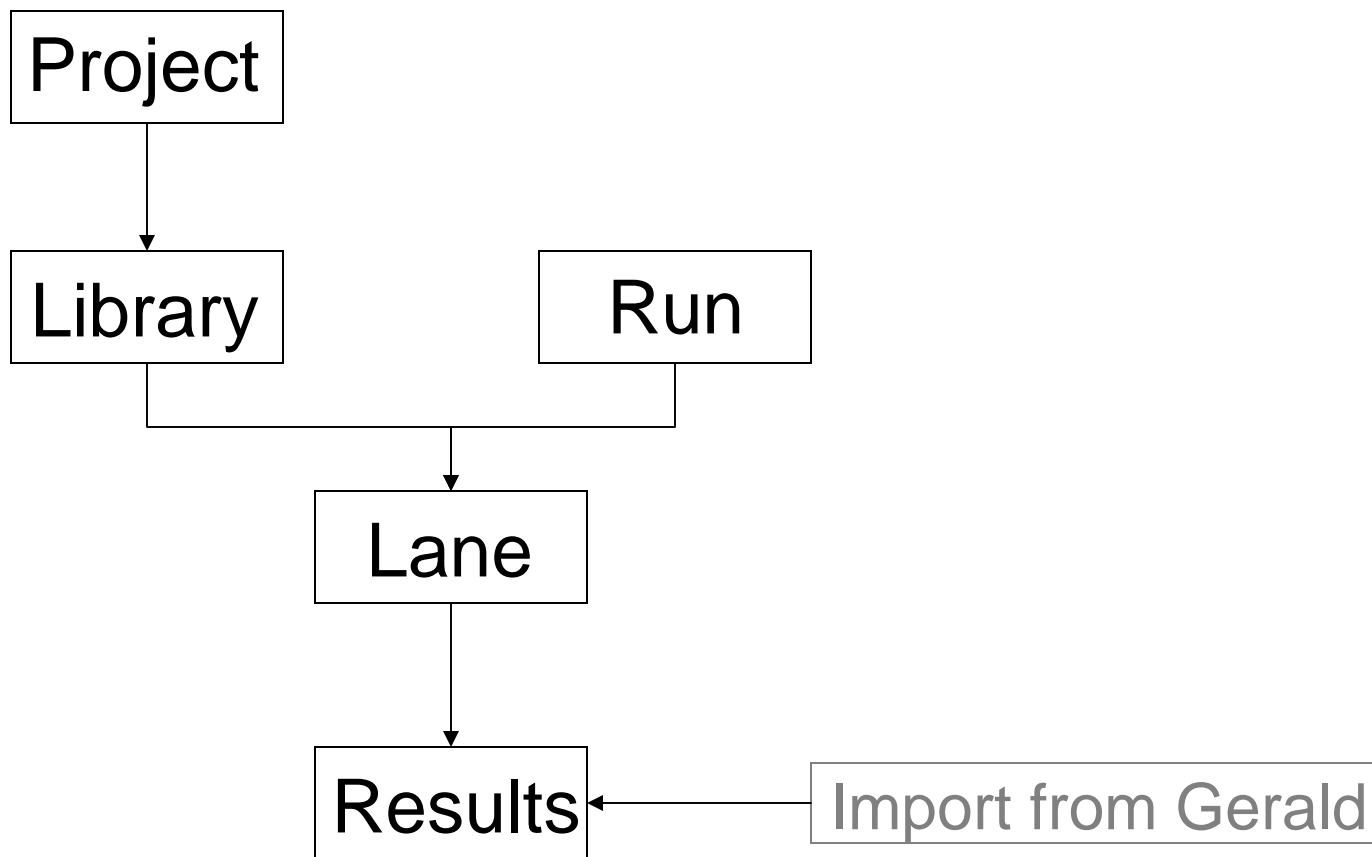
Error rate: Alignment error



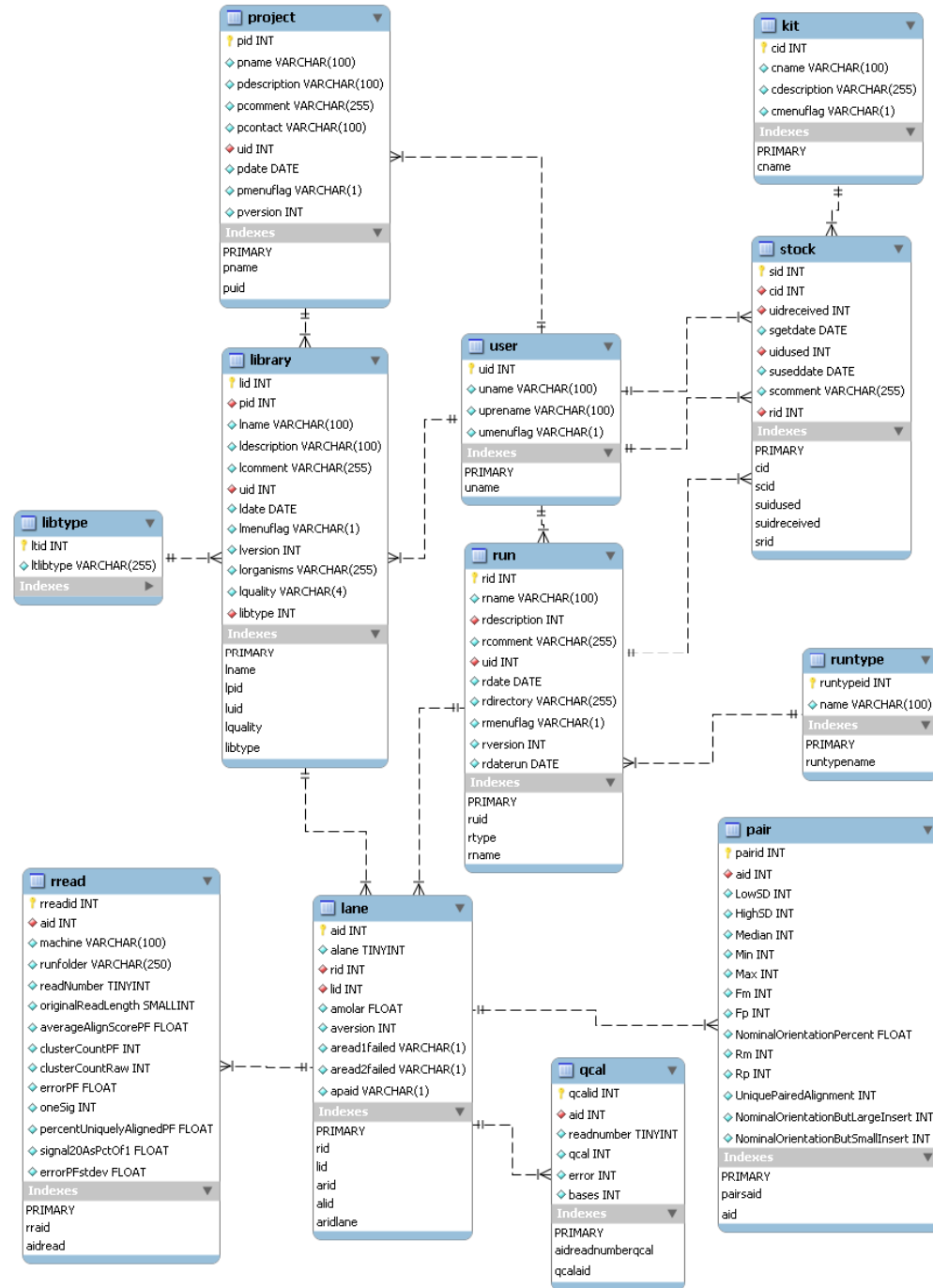
Exome non-synonymous SNPs

- Synonymous SNPs 10413 (novel 649)
- Non-synony. SNPs 10389 (novel 769)
 - 1600-2000 MAF < 0.05
- ~1500 (14% of 10389) may affect protein function (SIFT)
- 105 premature stop codon
 - 31 duplicated regions
 - 31 hypothetical genes
 - 22 members of gene families
 - 9 unique members
- 739 coding indels (52% tandem repeats)
- 632 nsSNPs in 443 OMIM disease genes

LIMS



LIMS



LIMS

[New Project](#)
[New Run](#)
[Search Projects](#)
[New Stock](#)
[Search Stocks](#)
[New Kit](#)
[List Kits](#)
[New User](#)
[List Users](#)
[Statistics](#)

All runs and lanes (including failed)

Runs 13
 Lanes 88
 Read1 failed 28
 Read2 failed 16

Summary

n Reads	Yield (Gb)	Sum Clusters (PF)	Mean Clusters (raw)	Mean Clusters (PF)	Mean 1st Cycle Int (PF)	Mean % Error Rate (PF)
9+ Reads, % Error Rate <= 2	23.815	6615771	119903	72700	224	1.14

n	Project	Library	Description	Machine	Run	Lane	Read	Length	pMol	Yield (Mbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% PF C
1	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	1	1	36	2	260	120580	72459	240	60
2	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	1	2	36	2	260	120580	72459	152	60
3	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	2	1	36	2	272	122683	75632	247	61
4	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	2	2	36	2	272	122683	75632	145	61
5	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	3	1	36	2	271	123040	75482	245	61
6	Rind Vanstein	42563_IB2	paired end	HWI-CAS106	307FYAAXX	3	2	36	2	271	123040	75402	143	61
7	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	5	1	36	1	187	70149	52027	236	74
8	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	5	2	36	1	187	70149	52027	138	74
9	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	6	1	36	1	187	69845	52069	239	74
10	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	6	2	36	1	187	69845	52069	138	74
11	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	7	1	36	1	189	71117	52745	239	74
12	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	7	2	36	1	189	71117	52745	138	74
13	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	8	1	36	1	183	69332	50966	246	73
14	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307FYAAXX	8	2	36	1	183	69332	50966	149	73
15	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307J5AAXX	7	1	36	3	305	150488	84989	264	56
16	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307J5AAXX	7	2	36	3	305	150488	84989	245	56
17	Rind Vanstein	42563_IB2	paired end	HWI-EAS186	307J5AAXX	8	1	36	3	301	148202	83740	252	56
18	Rind Vanstein	42563_IB2	paired end	HWI-FAS186	307J5AAXX	8	2	36	3	301	148202	83740	237	56

Thank you!

Anna Benet-Pagès, Sebastian Eck and
Kerstin Schwerdtner