SEVENTH FRAMEWORK PROGRAMME

Quantomics

High resolution copy number variable regions in Brown Swiss dairy cattle and their value as markers

MA Dolezal[1,2,3], K Schlangen[1,2], F Panitz[4], L Pellegrino[1], M Soller[5], E Santus[6], M Jaritz[3], A Bagnato[1,7]

1 University of Milan

2 Vetmeduni Vienna

3 FH Campus Vienna

4 Aarhus University

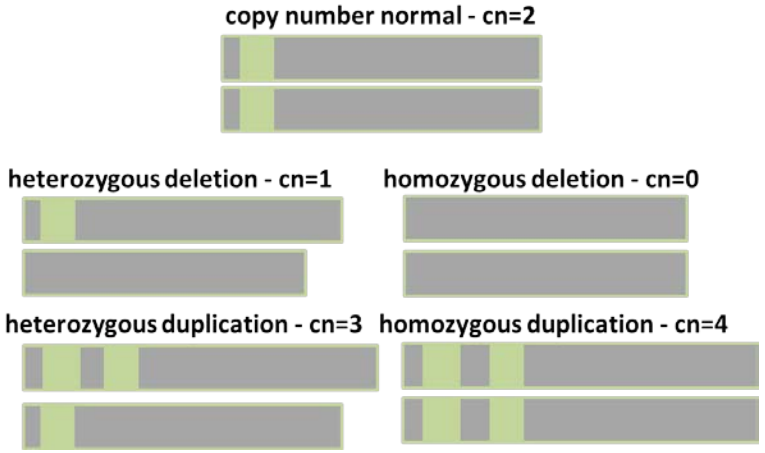
5 Hebrew University Jerusalem

6 ANARB

7 G&B platform Filarete Foundation - UNIMI

Background

- CNVs are duplications, insertions and deletions of chromosomal segments in comparison to a reference genome



- CNVs recognized as substantial source of genetic variation
- CNVs are summarized to copy number variable regions (CNVRs) at the population level

SEVENTH FRAMEWORK PROGRAMME

Quantomics

Objective

- evaluate potential contribution of CNVs as genetic markers for GWAS & GS in cattle

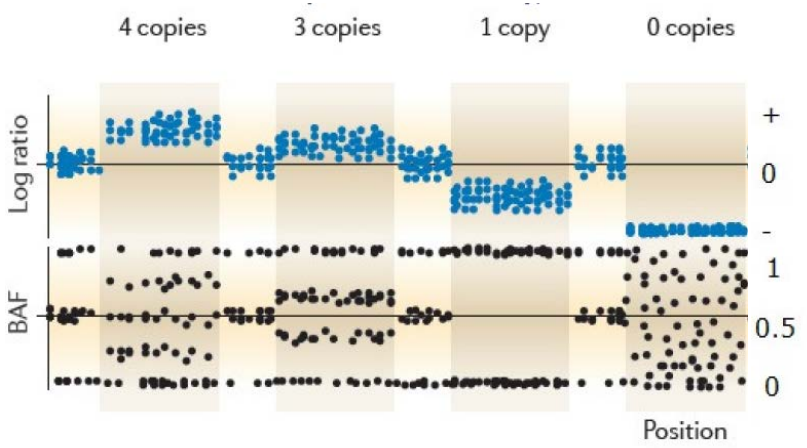
  - Polymorphic information content (PIC) of CNV loci
  - correlations between SNPs residing in CNVRs and their underlying CNVs
  - LD of SNPs residing in CNVRs with SNPs surrounding the CNVRs (adjacent SNPs)

SEVENTH FRAMEWORK PROGRAMME

Quantomics

Data for SNP array based CNV detection

- 192 BS bulls genotyped with Illumina HD chip
- Log R Ratios (LRR) - total signal intensities
- B allele frequencies (BAF) - allelic intensity ratio values



- LRR & BAF values for 735,239 SNPs on UMD3.1 autosome

Material & Methods – 1
CNV calling

- **CNV calling in 164 stringently quality filtered bulls**
  - **PennCNV[1] & genoCN[2]**
  - **reliable CNV calls ≥3 consecutive SNPs of the same type**

1 Wang et al. (2007) doi: 10.1101/gr.6861907
2 Sun et al. (2009) doi: 10.1093/nar/gkp493

Quantomics

Material & Methods – 2 definition of CNVRs

- within each algorithm summarisation of CNVs to CNVRs
  - union set of CNVs[1]

- high confidence set of CNVRs for population genetic analysis
  - intersection of overlapping CNVRs of same type[2] across algorithms

1 Redon et al. (2006) doi: 10.1038/nature05329
2 Wain et al. (2009) doi:10.1371/journal.pone.0008175

SEVENTH FRAMEWORK PROGRAMME

Quantomics

Material & Methods – 3 identification of „real“ alleles

- genoCN[1] employs a 3 copy number state model
  - 0-1-2 copies per haploid
  - possible alleles: 0, A, B, AA, BB and AB
  - total allelic content with highest posterior probability
    - eg. cn=3 AAB , possible alleles AA,B or AB,A
    - not equivalent to knowing the real alleles
- allele calling & phasing with polyHap[2] v2.0

1 Sun et al. (2009) doi: 10.1093/nar/gkp493
2 Su et al. (2010) doi: 10.1093/bioinformatics/btq157

Material & Methods – 4
population genetic characterization

- $PIC = 1 - \sum_{i=1}^{n} p_i^2$

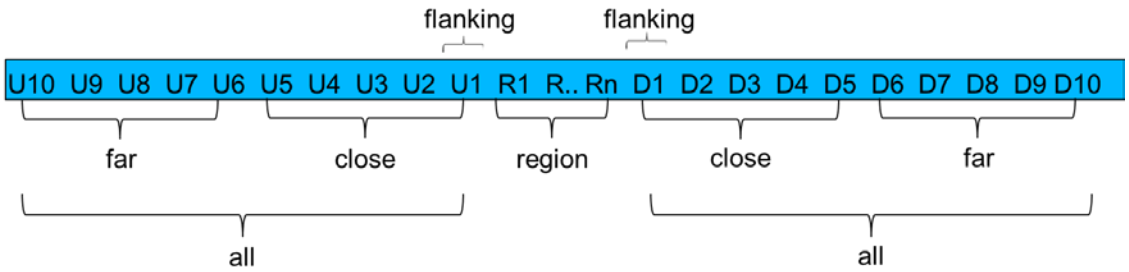
- **LD between SNPs residing within CNVRs and their underlying CNV**
  - **standard metrics incorrect[1]**
  - **r_C^2 correctly quantifies covariance[1]**

1 Wineinger (2011) doi: 10.3389/fgene.2011.00017

Material & Methods – 5
population genetic characterization

- **Global LD between SNPs in CNVRs & neighbouring SNPs: Wn (Cramer's V[1,2])**

$$W_n = \left[\frac{\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{D_{ij}^2}{p_i q_j}}{\min(I-1, J-1)} \right]^{1/2} = \left[\frac{\frac{X_{LD}^2}{2N}}{\min(I-1, J-1)} \right]^{\frac{1}{2}}$$

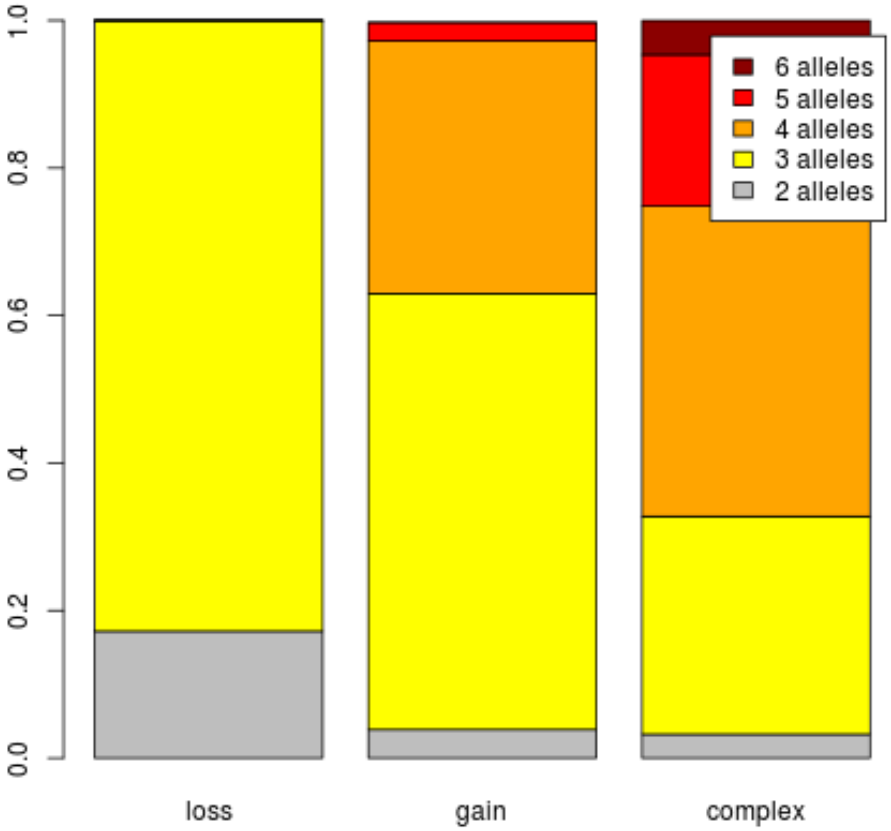


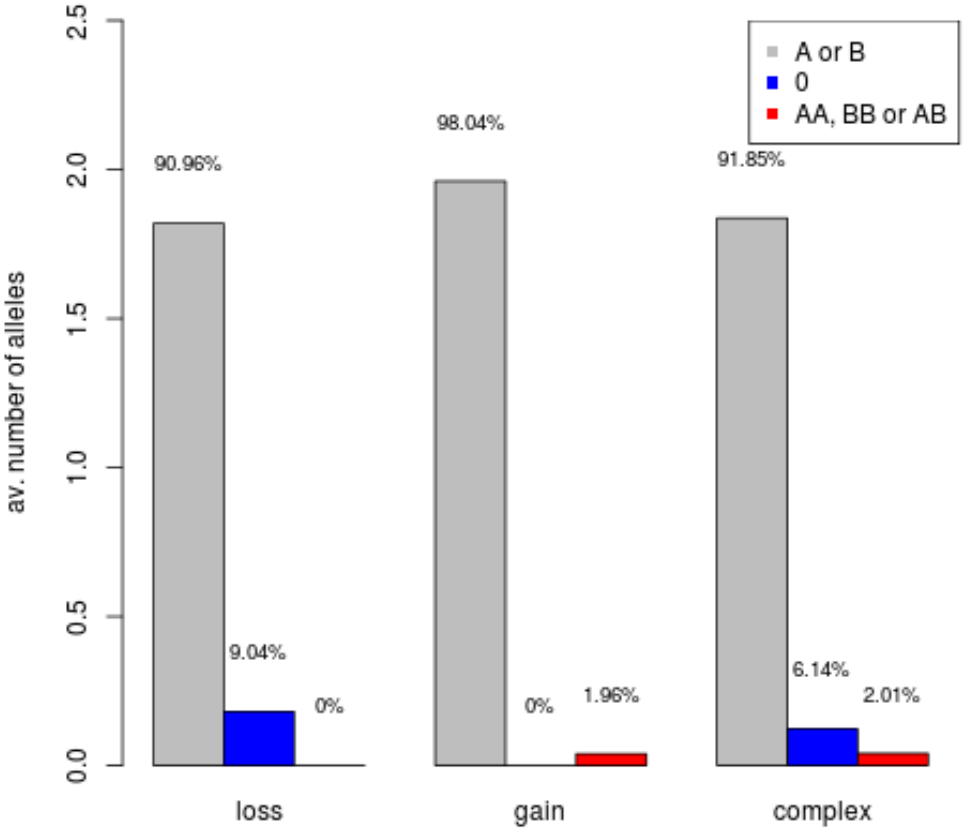
1 Cramer (1946) Mathematical Models of Statistics
2 Zhao (2007) Gap: Genetic Analysis Package. Journal of Statistical Software 23 (8):1-18

SEVENTH FRAMEWORK PROGRAMME

Quantomics

Results -1
number of alleles in CNVRs

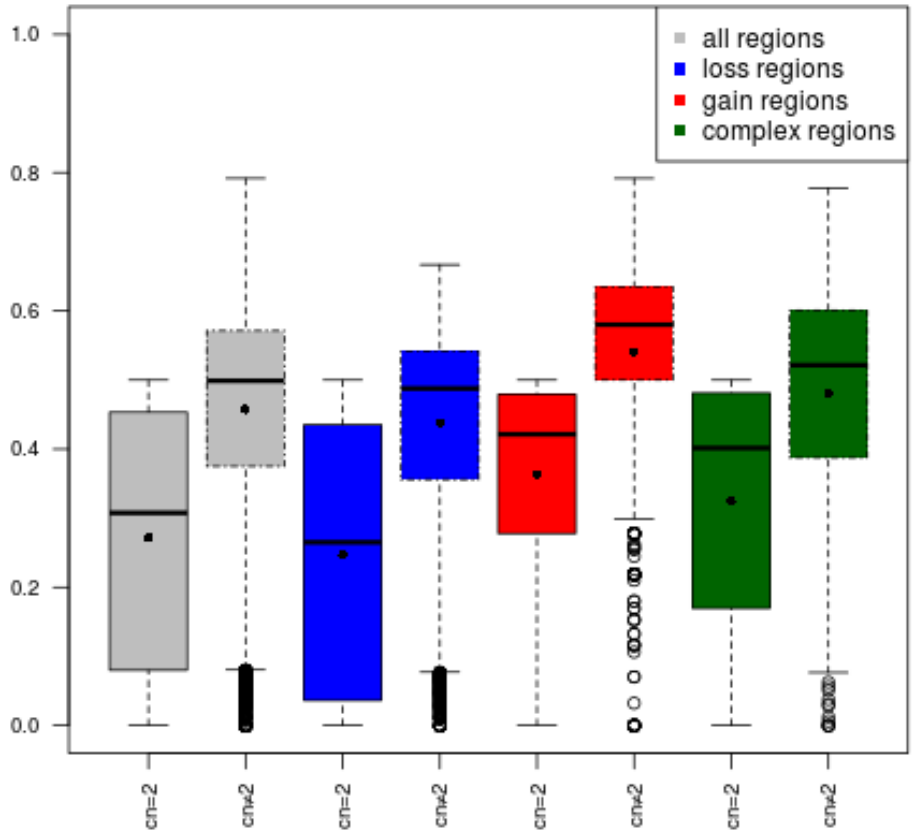




Quantomics

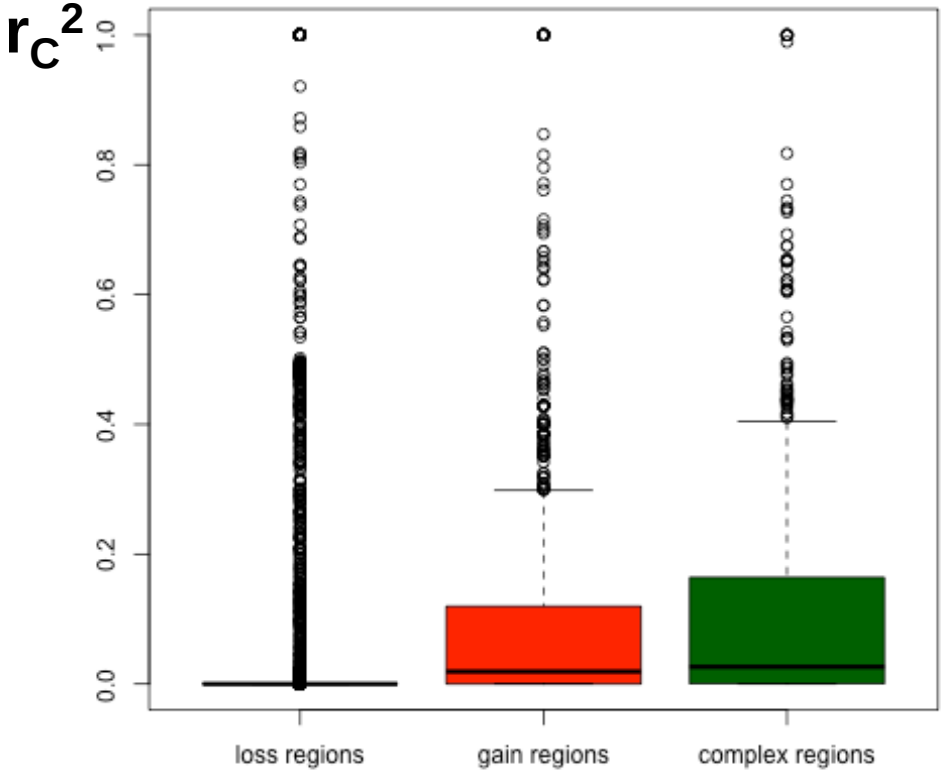
Results - 2
Polymorphic Information Content

PIC



cn=2: copy number normal
cn≠2: copy number variable

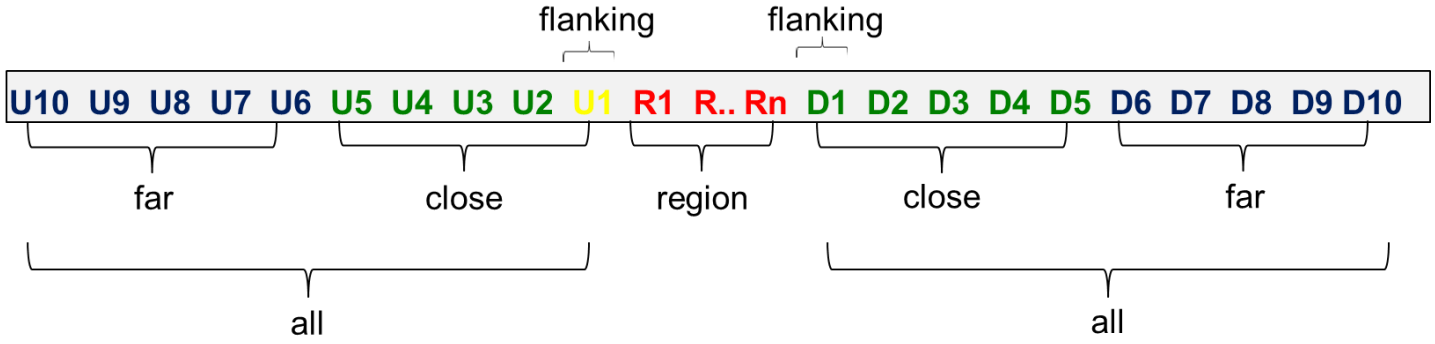
Results - 3
LD between SNPs in CNVRs & underlying CNV



SEVENTH FRAMEWORK PROGRAMME

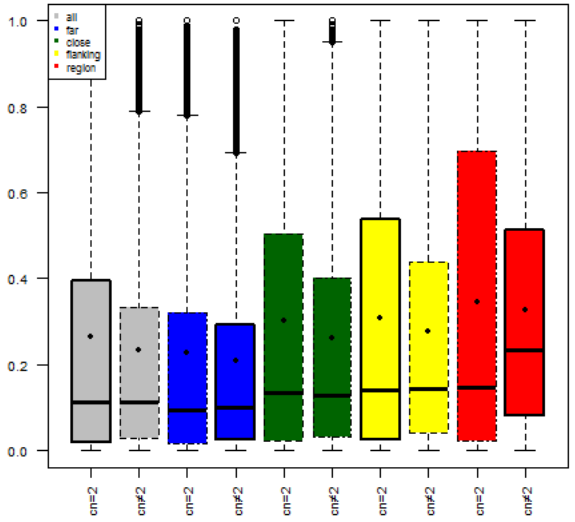
Quantomics

Results – 4 Global LD between neighbouring SNPs and SNPs in CNVRs

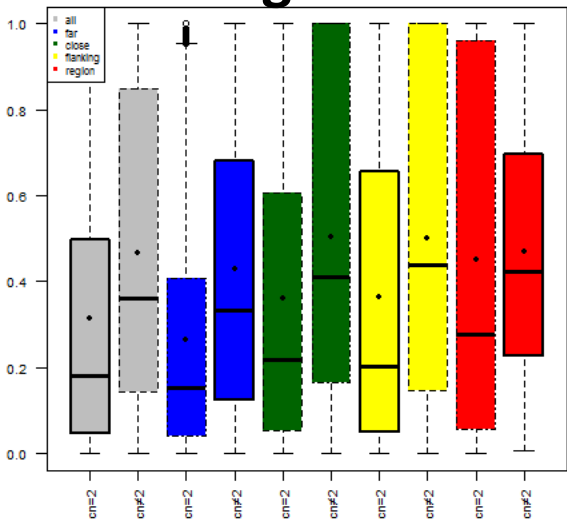


Wn

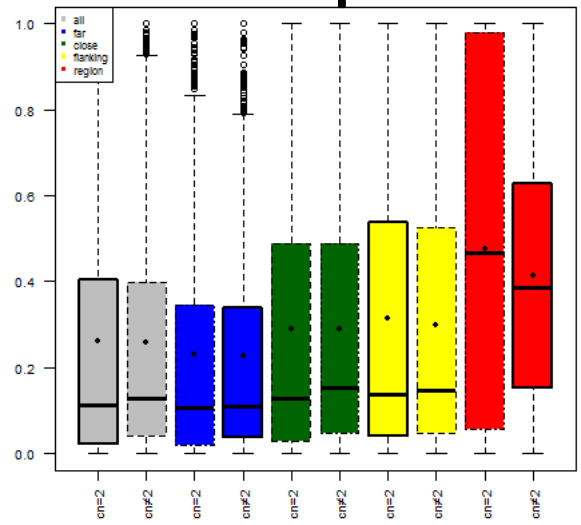
loss



gain



complex



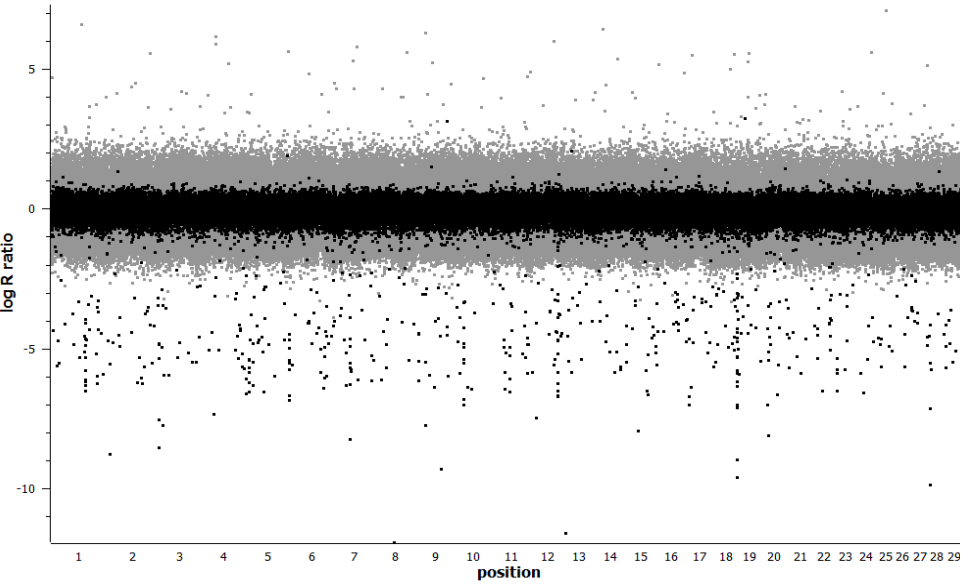
Conclusions

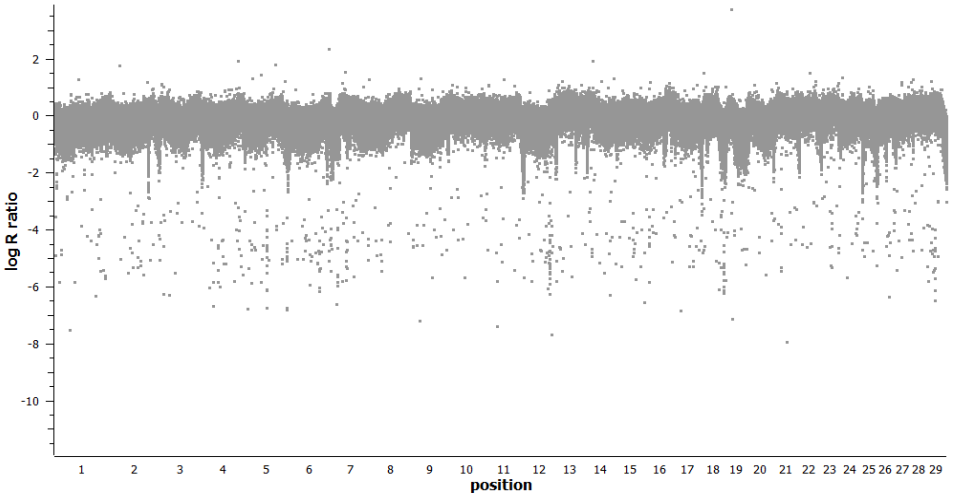
- CNVs are valuable genetic markers
  - high PIC
  - not sufficiently tagged by SNPs on HD chip
- Thank you for your attention!

SEVENTH FRAMEWORK PROGRAMME

Quantomics

Quality filtering

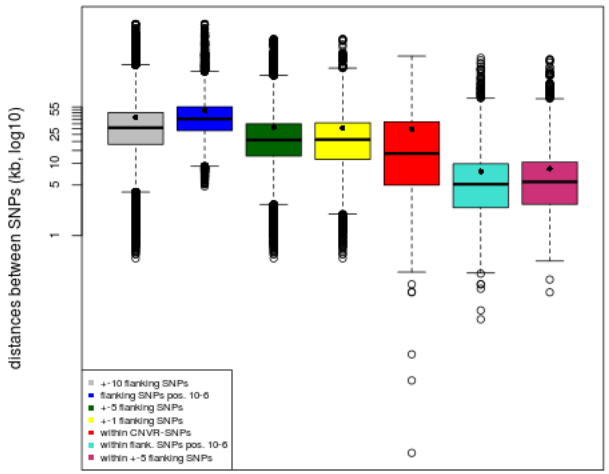


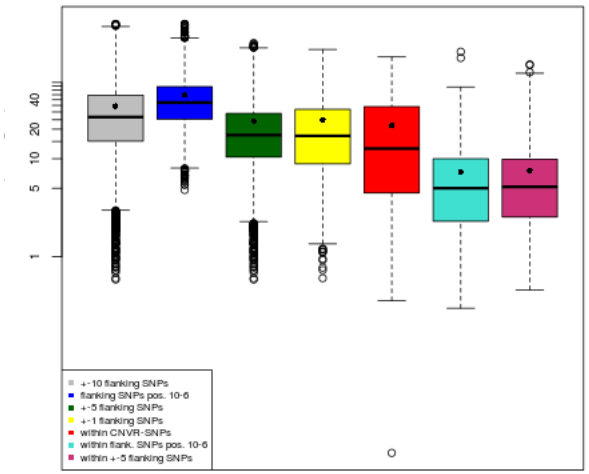


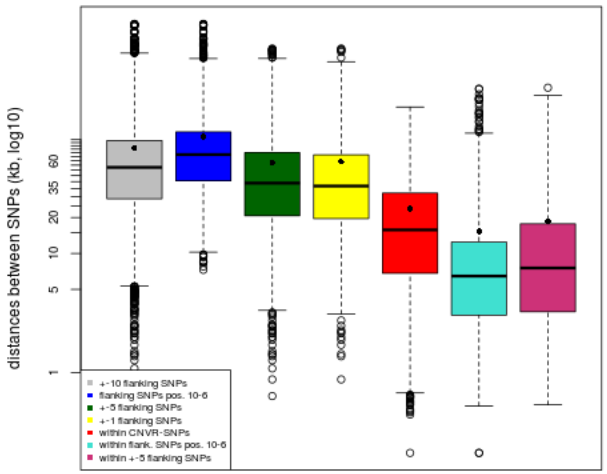
SEVENTH FRAMEWORK PROGRAMME

Quantomics

Distances

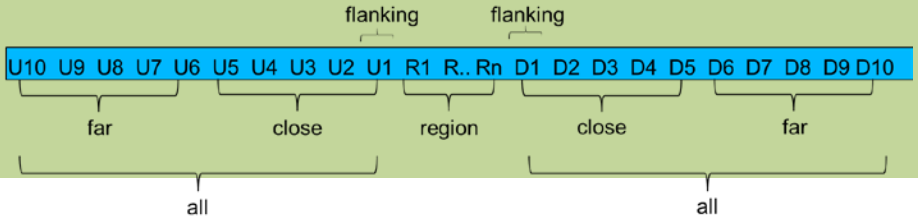




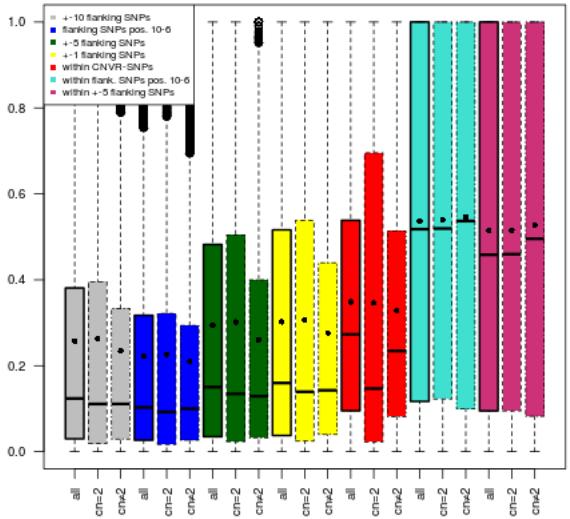


Quantomics

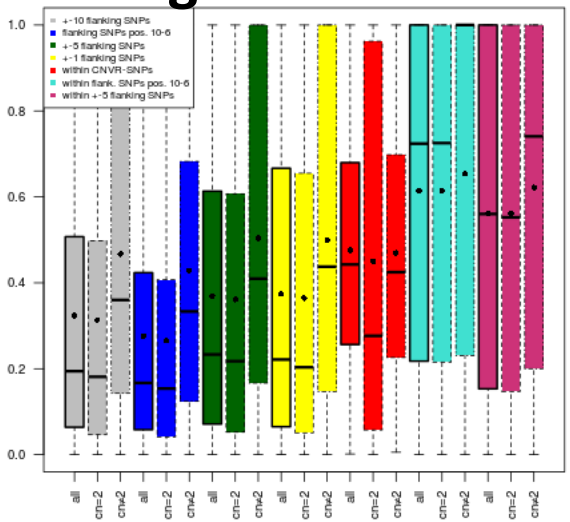
Results – 4
LD



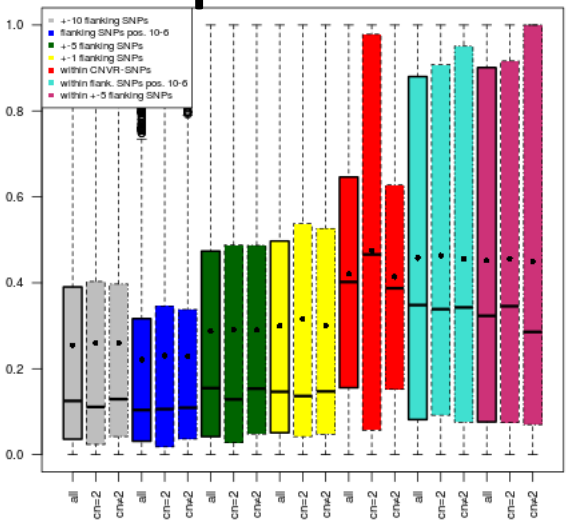
loss CNVRs

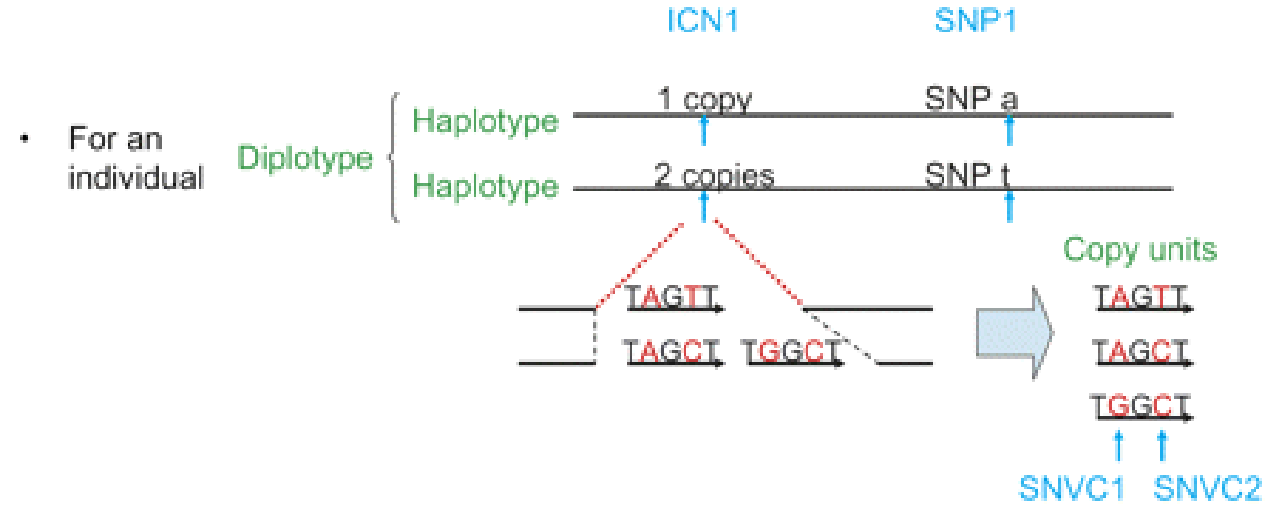


gain CNVRs



complex CNVRs





- Total copy number and SNP genotype at ICN1 and SNP1

ICN1	SNP1
3	at

- Unphased SNVC genotypes at SNVC1 and SNVC2

SNVC1	SNVC2
AAG	CCT

Figure 1 from Kato et al. (2011) doi: 10.1534/g3.111.000174