VUmc Cancer Center Amsterdam

Structural variant detection in colorectal cancer

E van den Broek[1], JC Haan[1], MH Jansen[1], B Carvalho[1], MA van de Wiel[2], ID Nagtegaal[3], CJA Punt[4], B Ylstra[1], S Abeln[5], GA Meijer[1], RJA Fijneman[1]

1 Dept. of Pathology, VU University Medical Center, Amsterdam, The Netherlands
2 Dept. of Epidemiology & Biostatistics, VU University Medical Center, Amsterdam, The Netherlands
3 Dept. of Pathology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
4 Dept. of Medical Oncology, Academic Medical Center, University of Amsterdam, The Netherlands
5 Dept. of Computer Science, VU University, Amsterdam, The Netherlands

VUmc Cancer Center Amsterdam

decode a ctmm project

trait a ctmm project

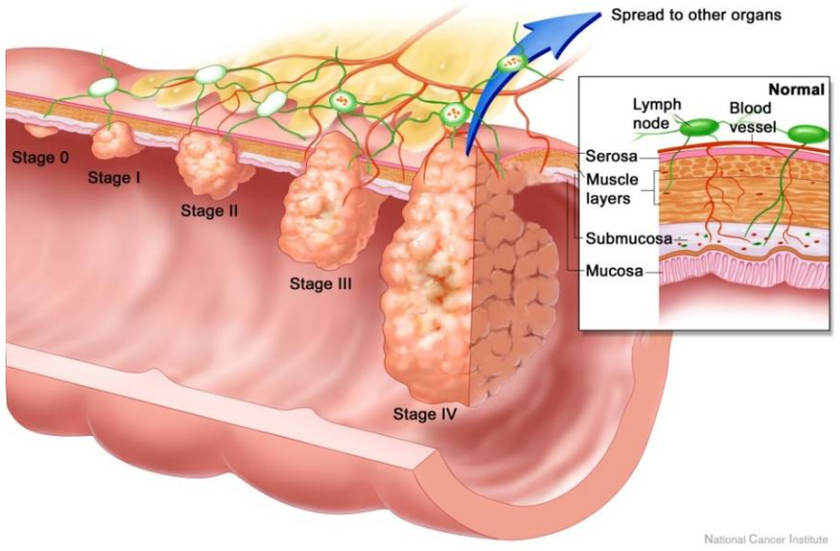
Colorectal cancer (CRC)

- Colorectal cancer is a major health concern worldwide

- Second cause of cancer related death
  - The incidence worldwide is 1,200,000
  - The incidence in the US is 144,000

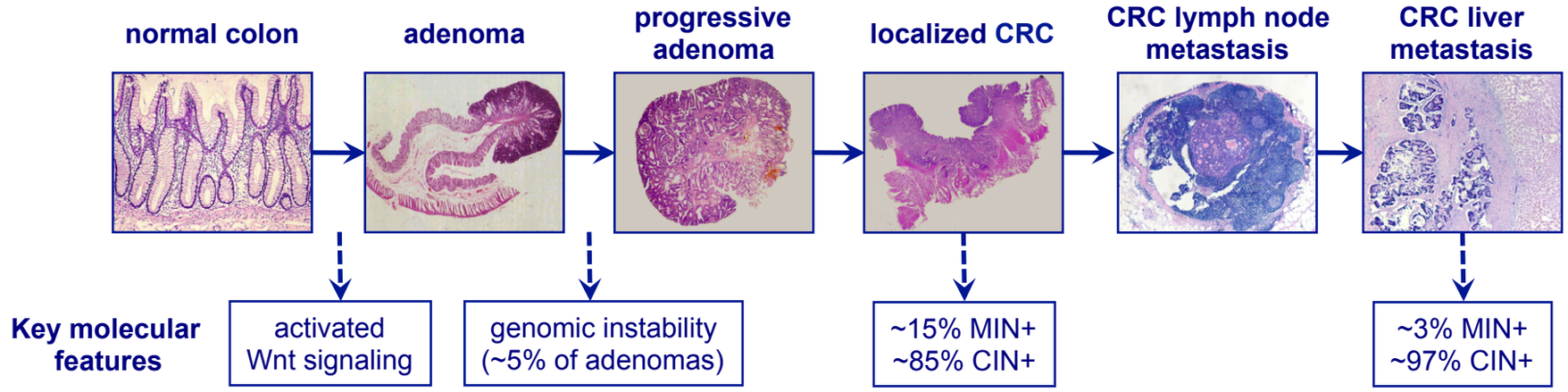
- Mortality rates

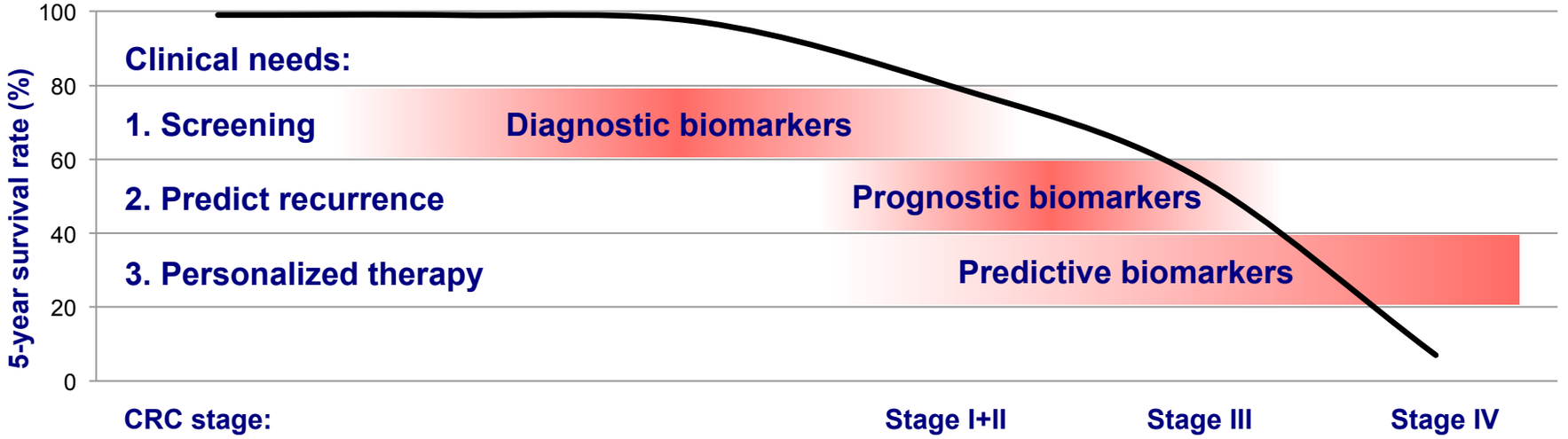
Stage 1	< 10 %
Stage 2	25 - 30 %
Stage 3	45 – 50 %
Stage 4	> 90 %



CRC research

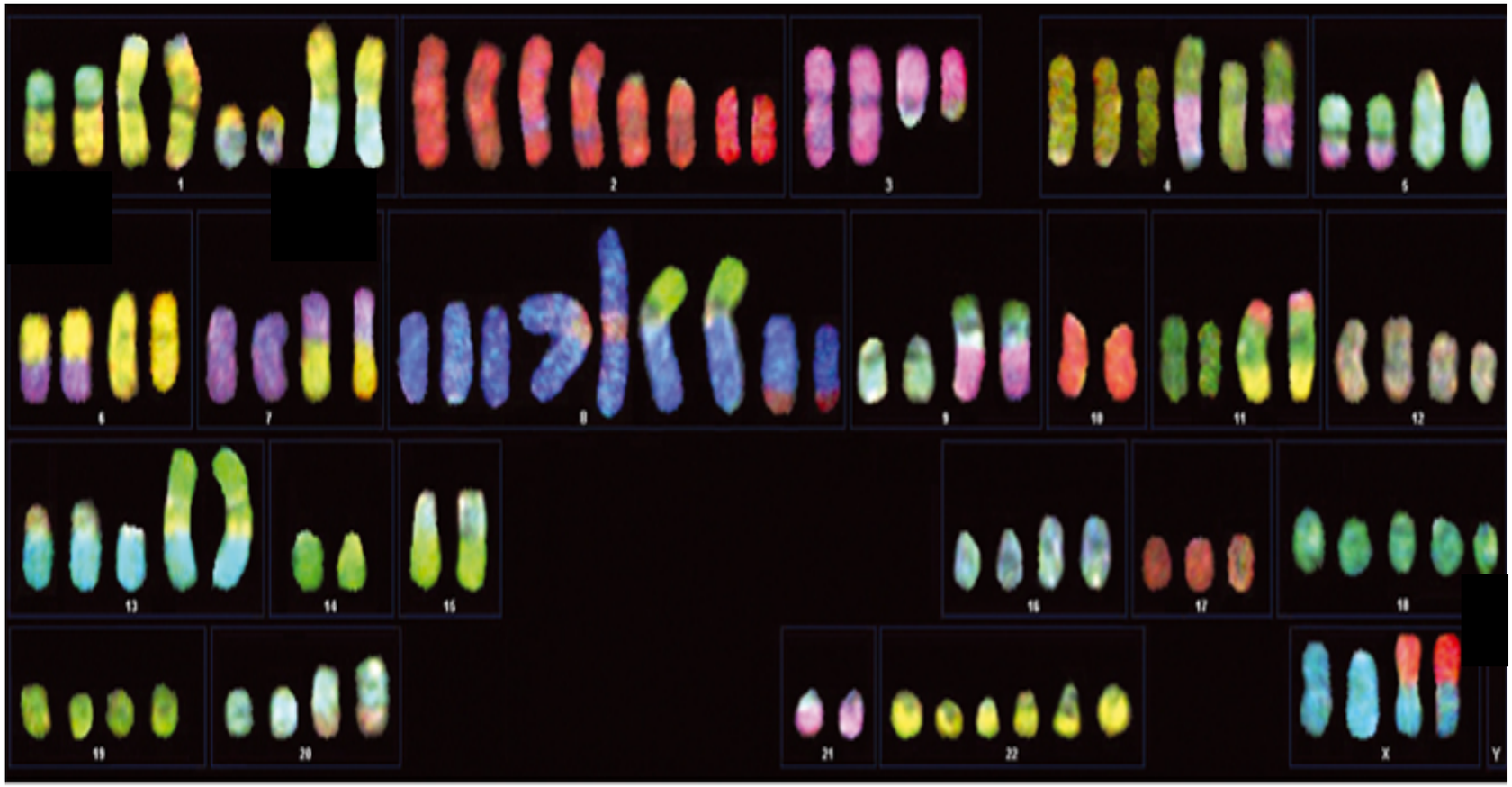
Clinical needs for biomarker discovery





Chromosomal Instability
a hallmark of CRC

SKY: numerical & structural aberrations



M Hermsen et al., Oncogene 2005

CAIRO & CAIRO2 studies

Phase III clinical trials
In total 1575 patients were included

CApecitabine, IRinotecan, Oxaliplatin
in advanced colorectal cancer

CAIRO: *Koopman et al. Lancet 2007*
CAIRO2: *Tol et al. N Engl J Med 2009*

DNA from 356 patients: primary tumor and matched normal
- Representative group
- Isolated from FFPE

Comparative Genomic Hybridization (CGH)

Agilent, 180k array CGH

Array CGH: 356 CAIRO & CAIRO2 samples

1 Segmentation

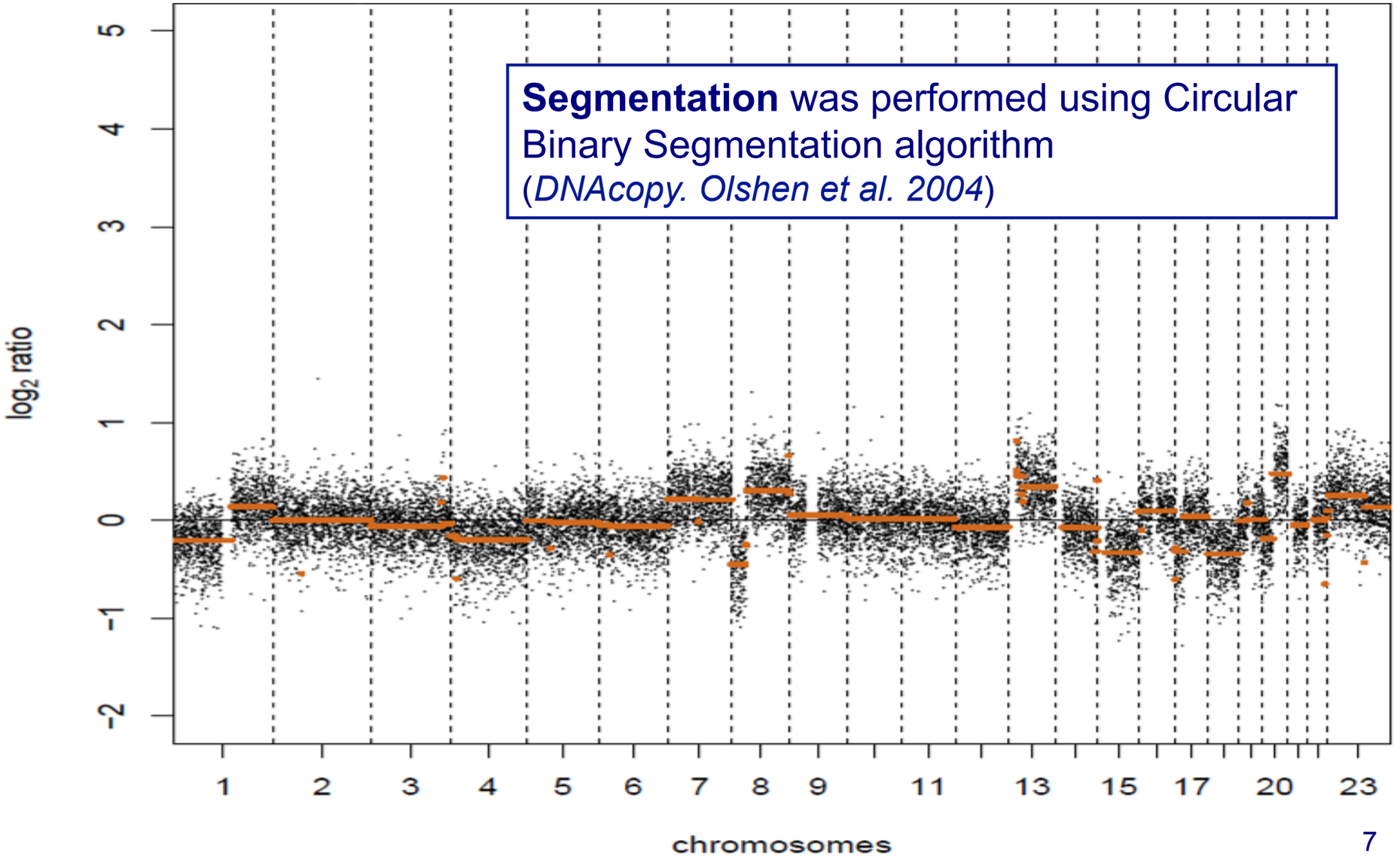
2 Calling

3 Copy numbers

Numerical aberrations

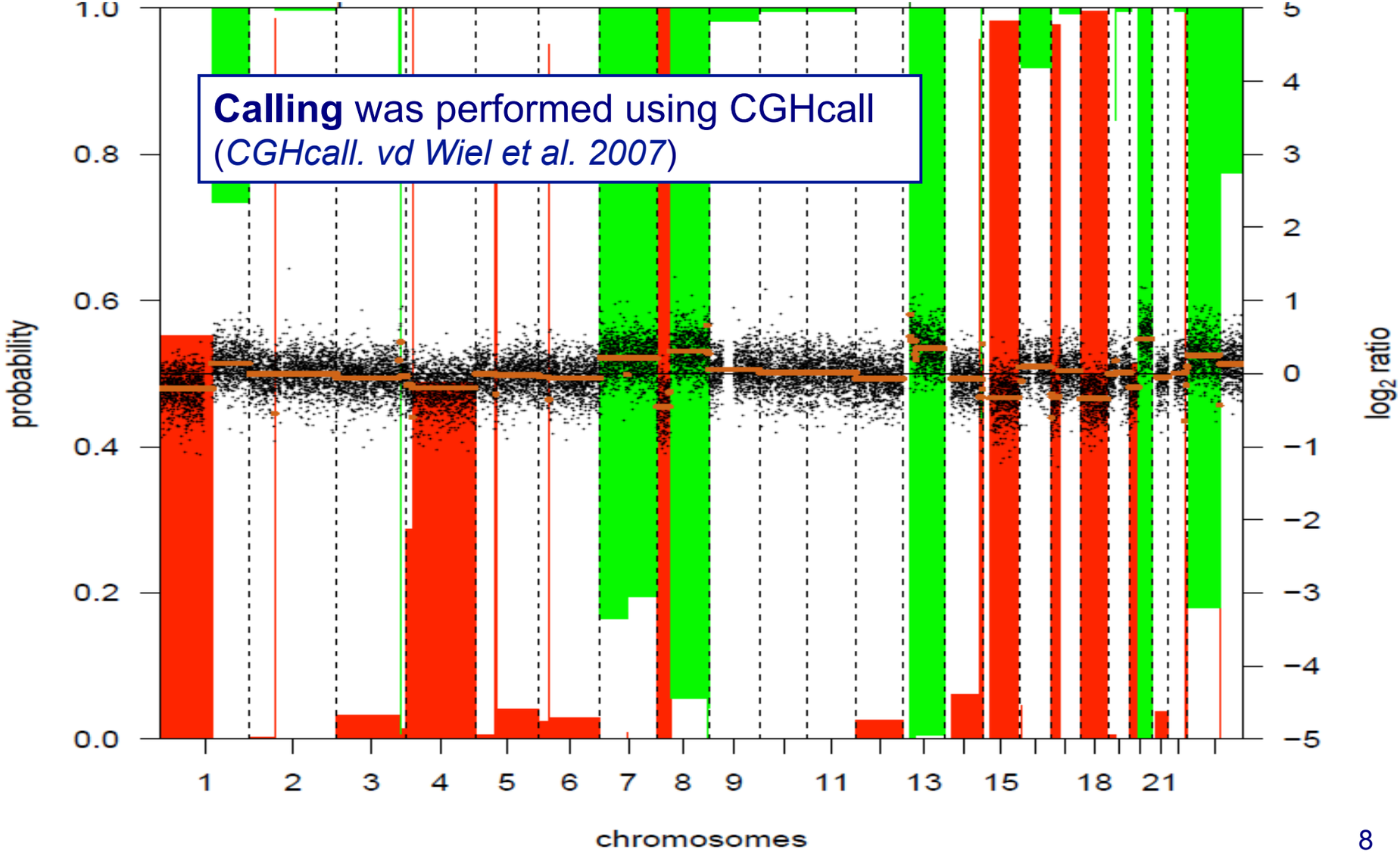
Segmentation - array CGH

Profile of one tumor with 180k probes



Calling - array CGH

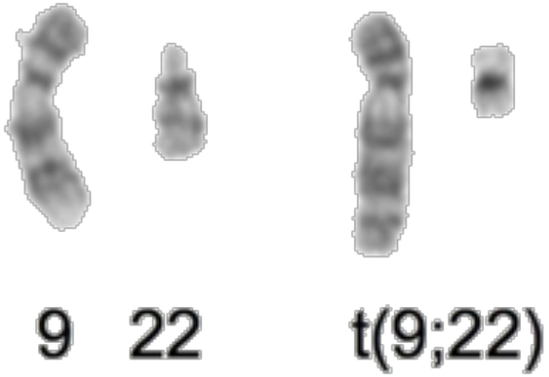
Profile of one tumor with 180k probes



Structural Variants (SV) in cancer

Hematological disorders

- Philadelphia chromosome
  - t(9;22)
  - Fusion gene: BCR-ABL
  - Drug: Imatinib / Gleevec



Epithelial cancers

- TMPRSS2-ERG in prostate cancers
- VTI1A-TCF7L2 is confirmed in 3% of 97 CRCs
  - *Bass et al., Nature Genetics 2011*

AIM

TO IDENTIFY RECURRENT SOMATIC STRUCTURAL GENOMIC VARIANTS THAT CAUSE CRC

Breakpoint (BP) detection

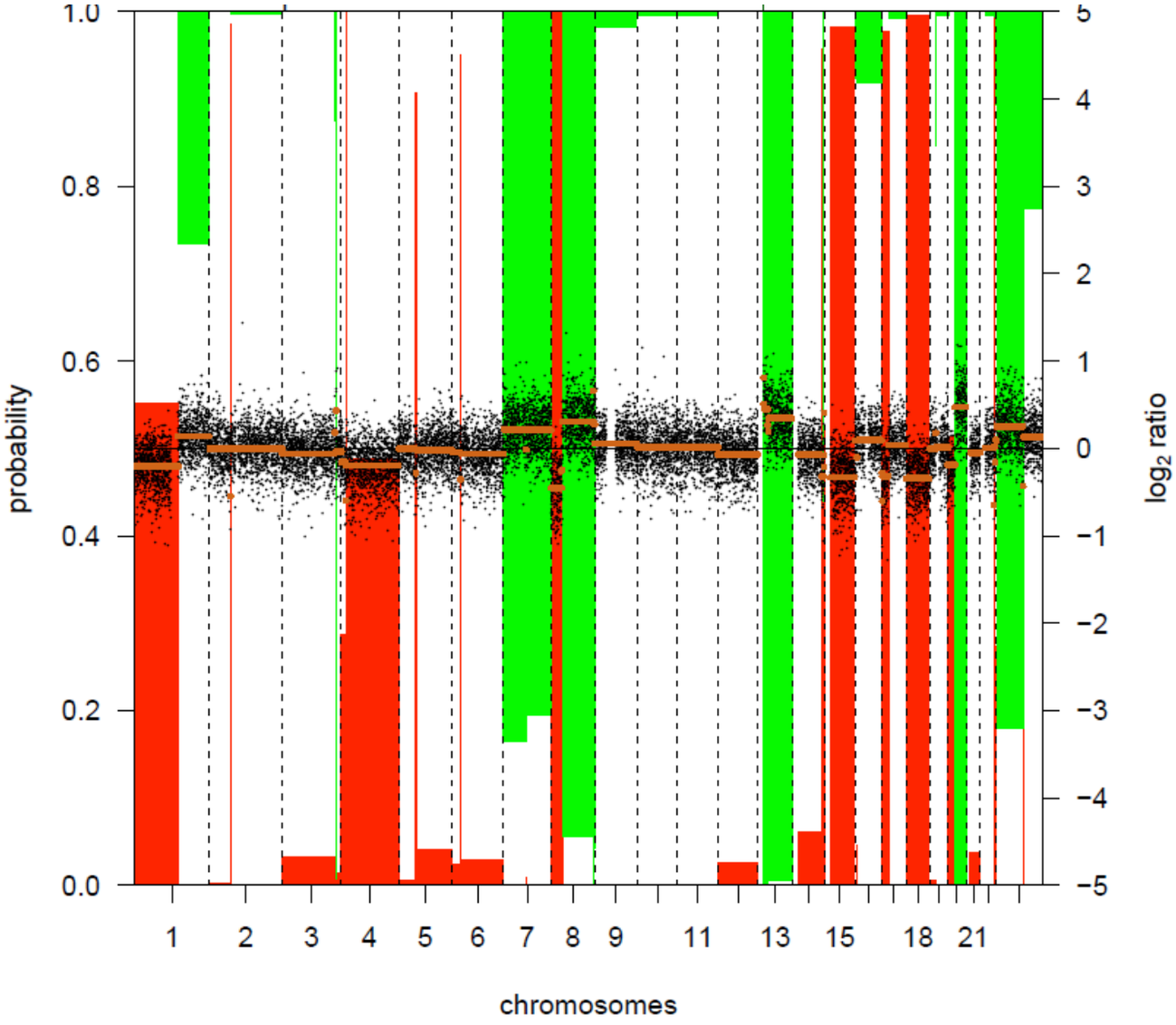
Based on array CGH

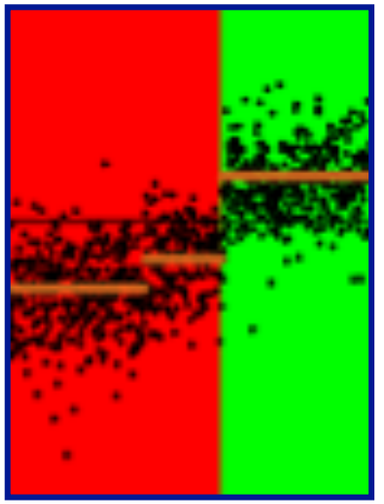
Array CGH: 356 CAIRO & CAIRO2 samples

1	Segmentation	1	Segmentation
2	Calling	2	Breakpoint detection
3	Copy numbers	3	Candidate genes
	Numerical aberrations		**Structural variants**

BP detection in array CGH

Profile of one tumor with 180k probes





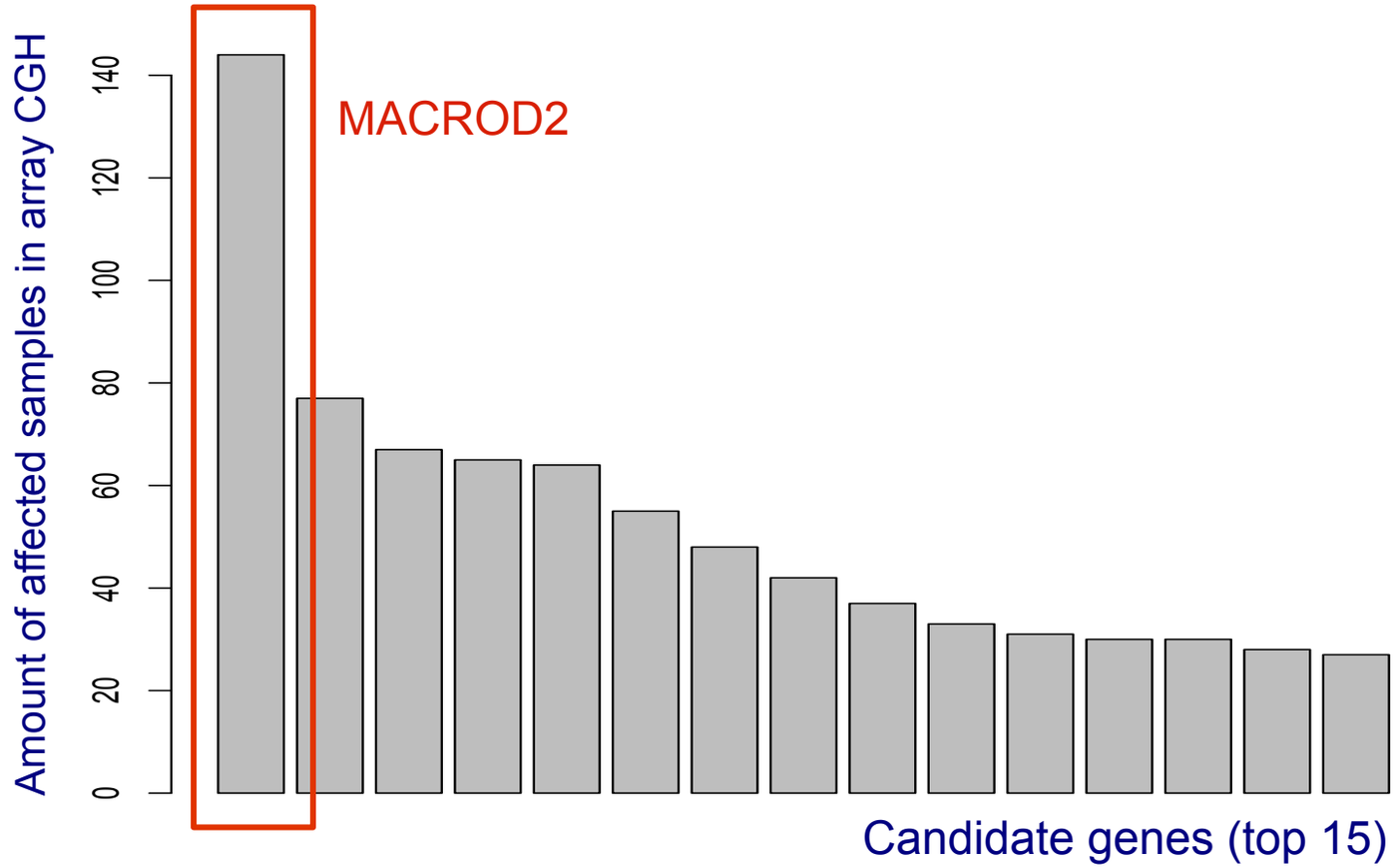
Breakpoints are defined by the start position of the first probe of each segment

Breakpoint annotation per gene

Results based on array CGH

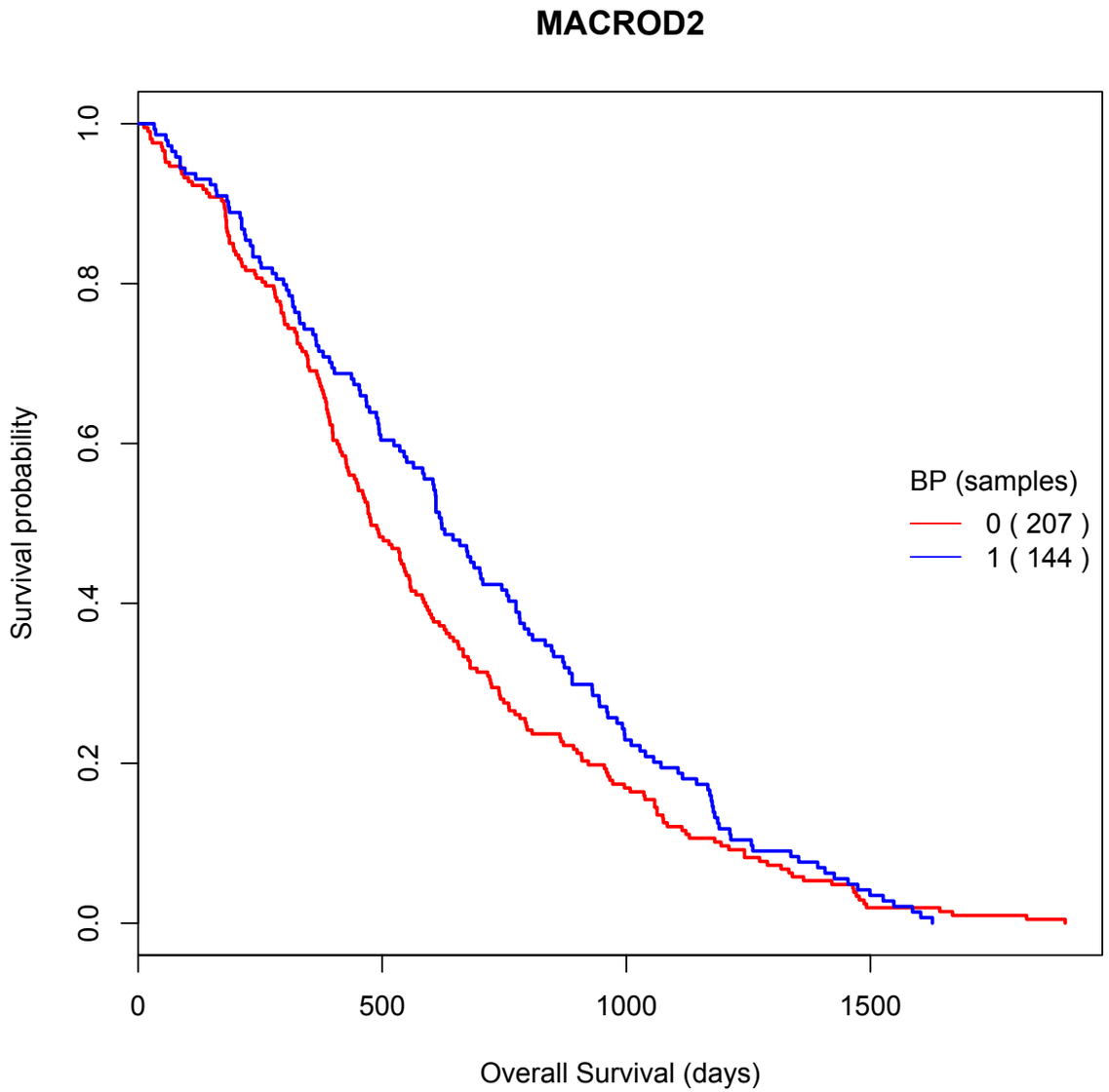
BP detection

- Total number of genes with BPs: 5,737 genes
- 482 candidate genes were identified with recurrent BP (FDR < 0.1)



Overall survival: MACROD2

Recurrent BP (1) versus no-BP (0)



Log rank P= 0.08

Results based on array CGH

BP detection

- Total number of genes with BPs: 5,737 genes
- 482 candidate genes were identified with recurrent BP (FDR < 0.1)

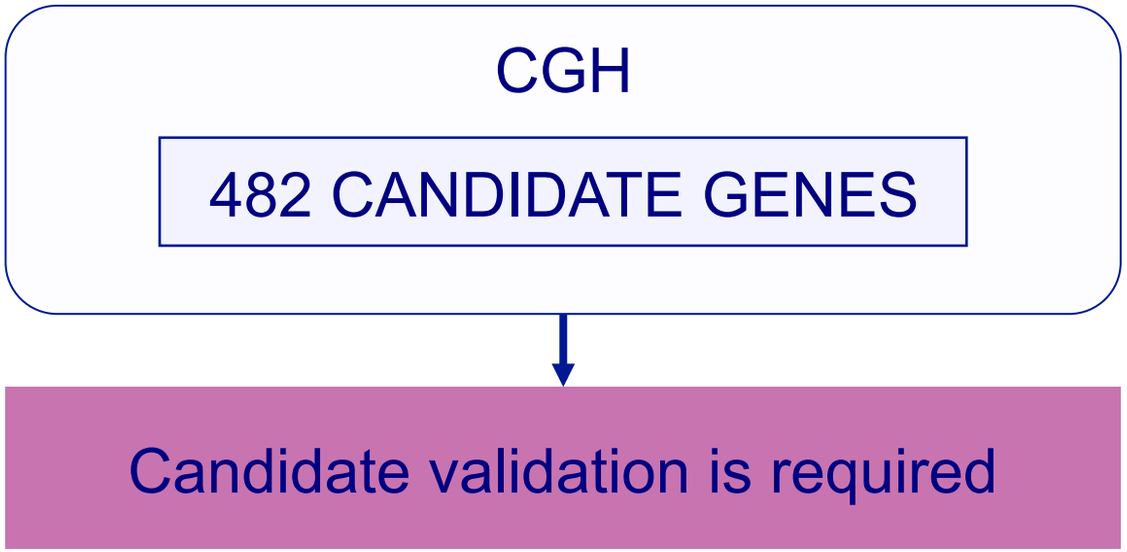
Limitations breakpoint determination using array CGH:

– Location BP is estimation (average probe distance is ~17 kb)
– DNA structure is unknown
– Balanced events will be missed

Validation array CGH BPs

NGS data from TCGA

- 482 candidate genes were identified with recurrent BP (FDR < 0.1)



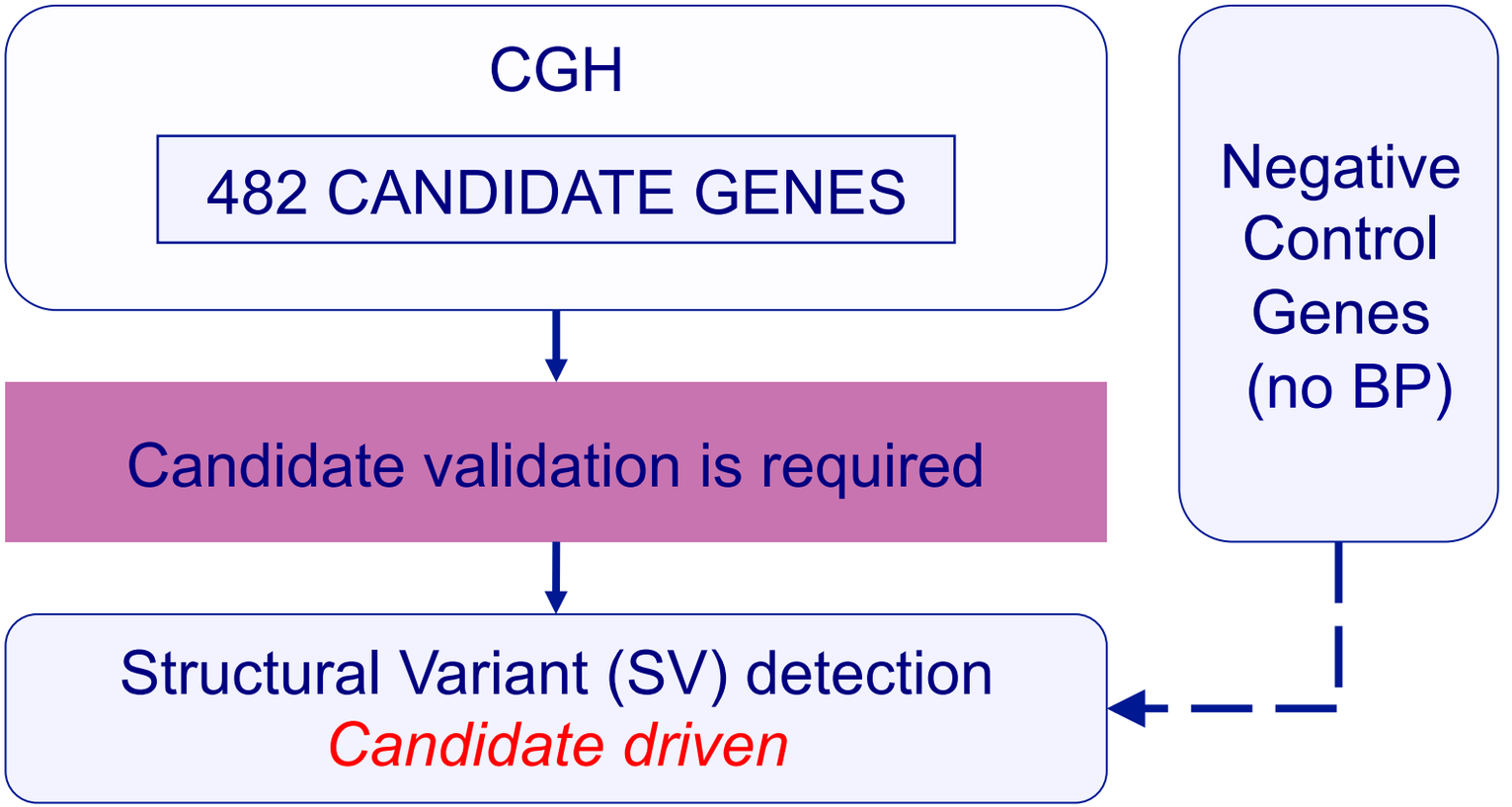
THE CANCER GENOME ATLAS **CRC samples (COAD & READ)**

Whole Genome DNA Seq from paired tumor-normal samples

Validation array CGH BPs

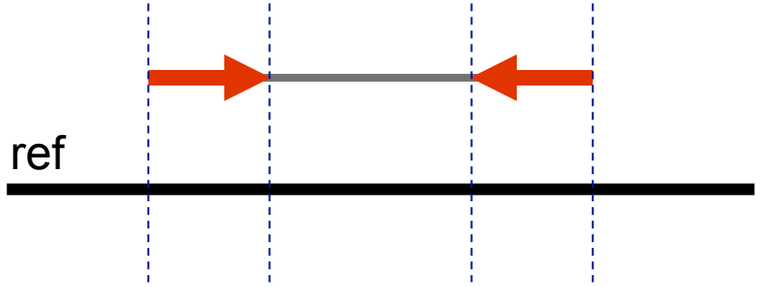
NGS data from TCGA

- 482 candidate genes were identified with recurrent BP (FDR < 0.1)



Computational methods

Focus on candidate genes



Based on paired-end NGS data

- Read-pair approach
  - Discordance: location / bridge length / orientation reads

Discordant pairs (DP) types

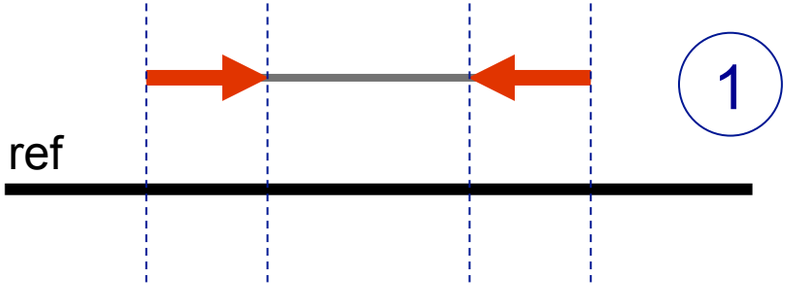
- Translocation > *different chromosomes*
- Insertion > *bridge length*
- Deletion > *bridge length*
- Inversion > *orientation*
- Eversion > *orientation*
- Single mapped *could indicate a breakpoint*

Computational methods

Focus on candidate genes

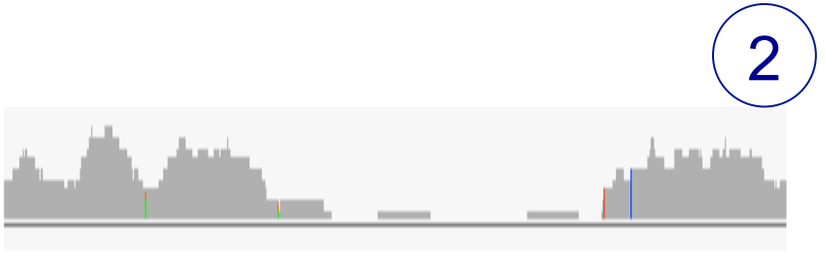
Based on paired-end NGS data

1. Read-pair approach

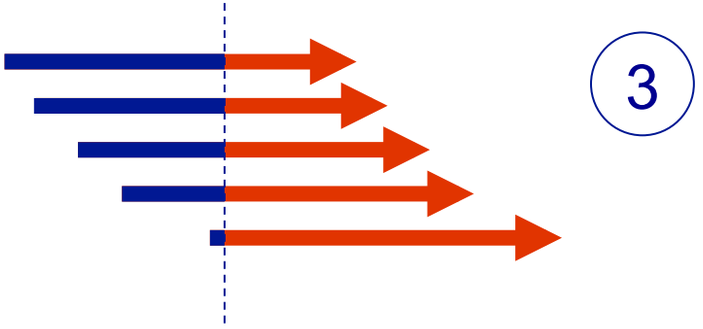


Combined with:

2. Read-depth



3. Define breakpoint location



4. Determine tumor specific events

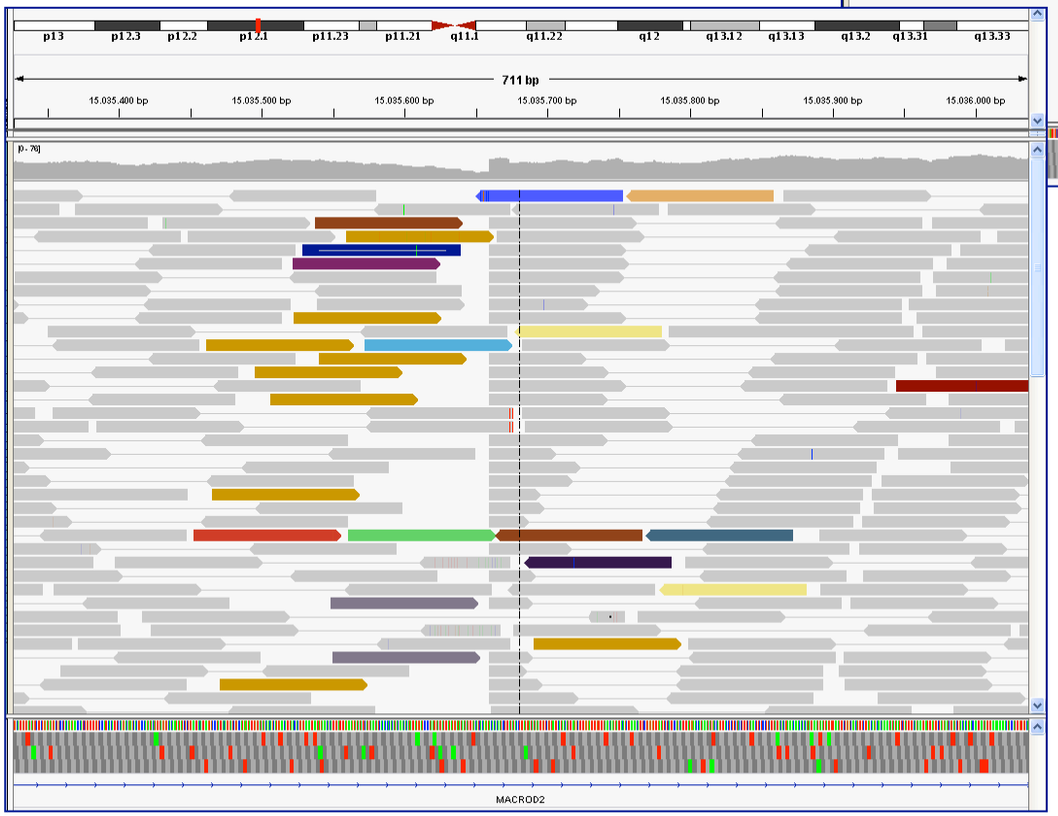
Translocation

IGV

MACROD2

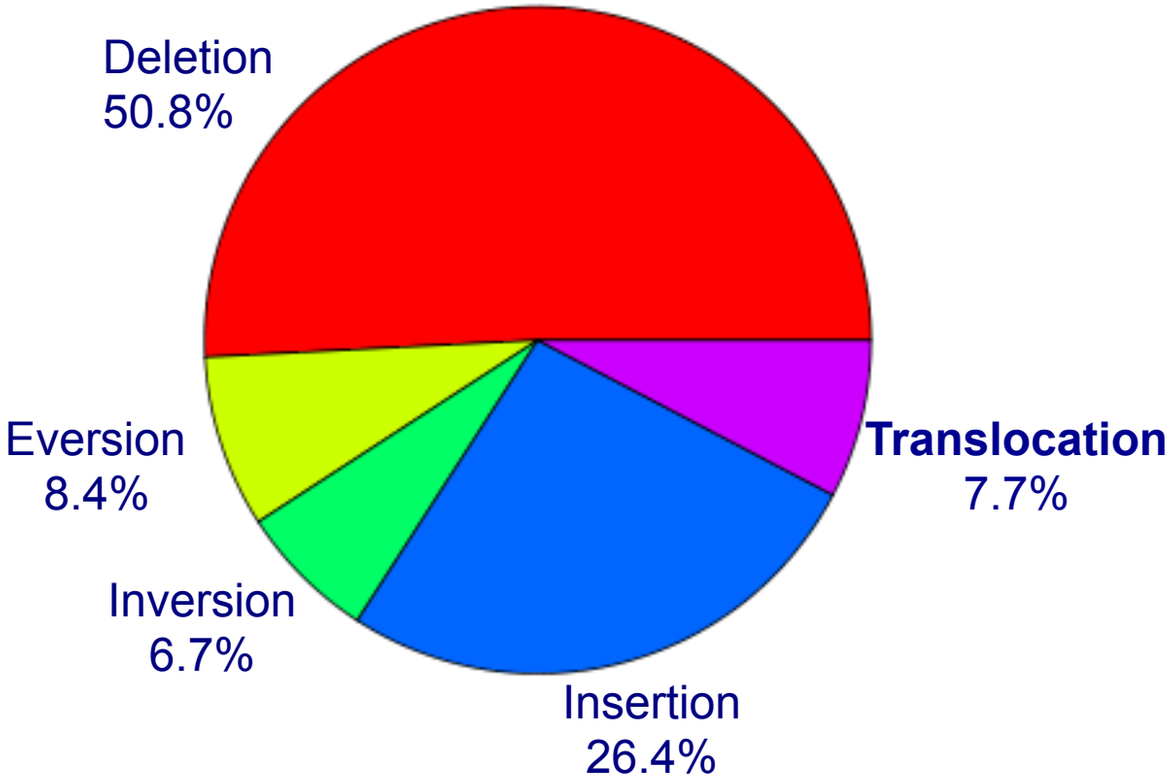
- Discordant pairs
- Breakpoints

Fusion partner

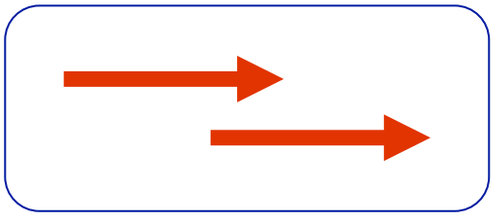


Distribution DP groups per type

Preliminary results candidate genes in TCGA data



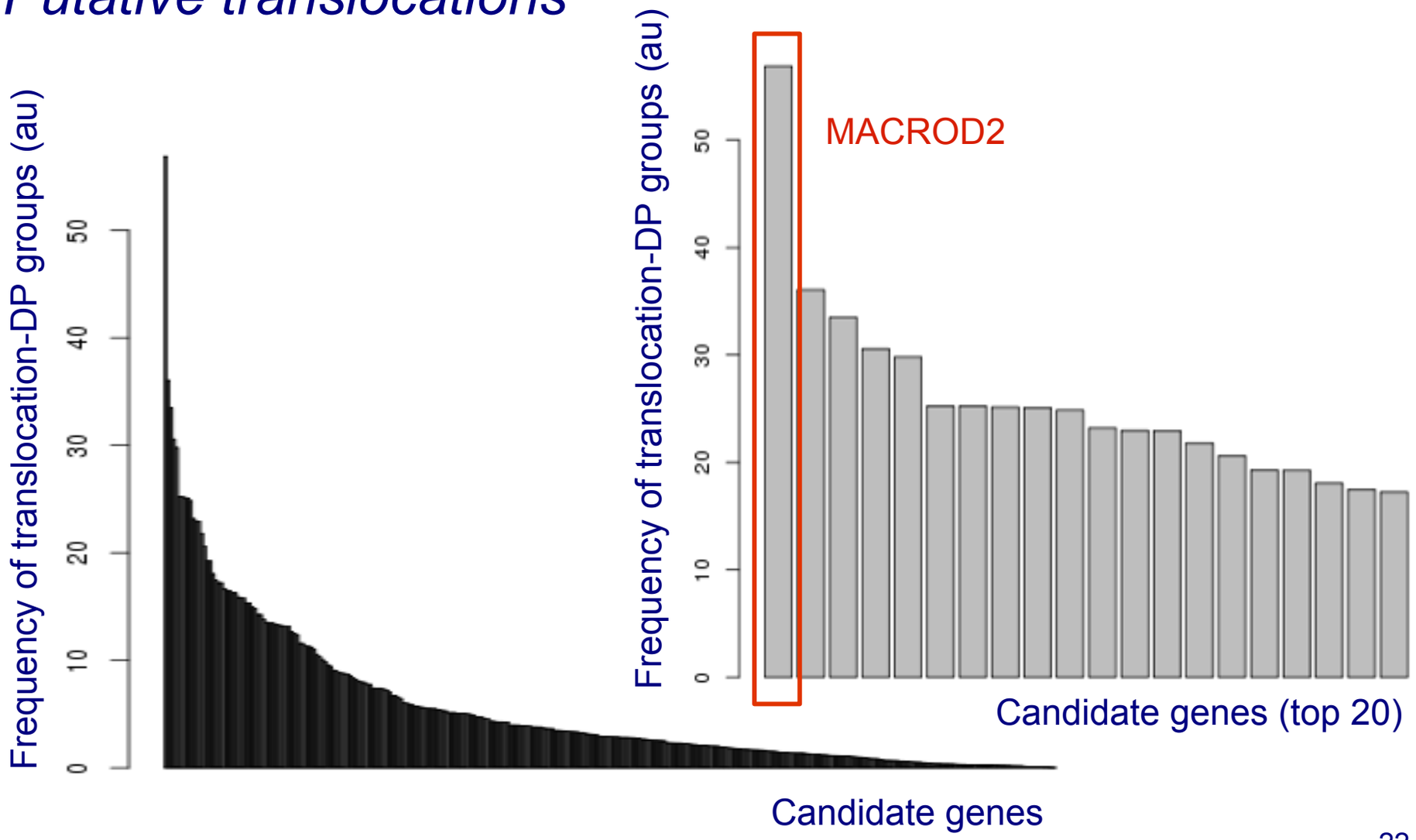
Based on DP groups:



Approximately 5 fold higher number of translocation-DP groups for candidate genes compared to control genes

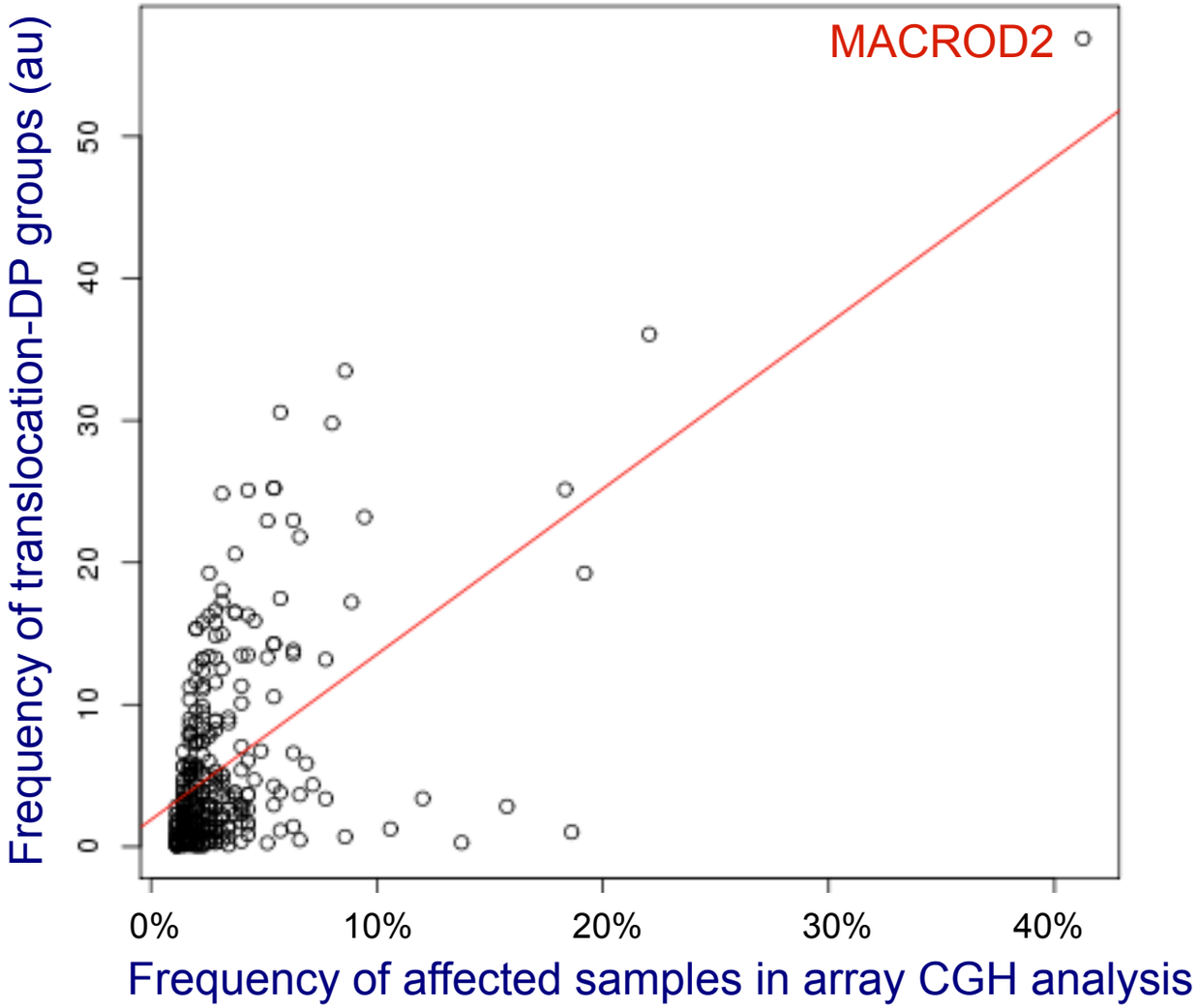
Translocation-DP groups per candidate gene in TCGA samples

Putative translocations



Correlation per candidate gene

- *Frequency of samples with BP based on array CGH*
- *Frequency of translocation-DP groups in TCGA data*



Conclusions

- 482 candidate genes with recurrent breakpoints were identified in a large cohort of 356 CRC samples, based on array CGH analysis
- 

provided an essential CRC reference dataset (COAD, READ) to validate Structural Variants in candidate genes with recurrent breakpoints
- Identification of BPs based on array CGH is correlated with SV detection in TCGA CRC NGS data
- Further studies will be performed to investigate clinical and functional significance of validated candidate genes