

Genomic Copy-Number Anomalies in Cancer



Their study through microarray-based and next-generation sequencing technologies

Introduction

First things first

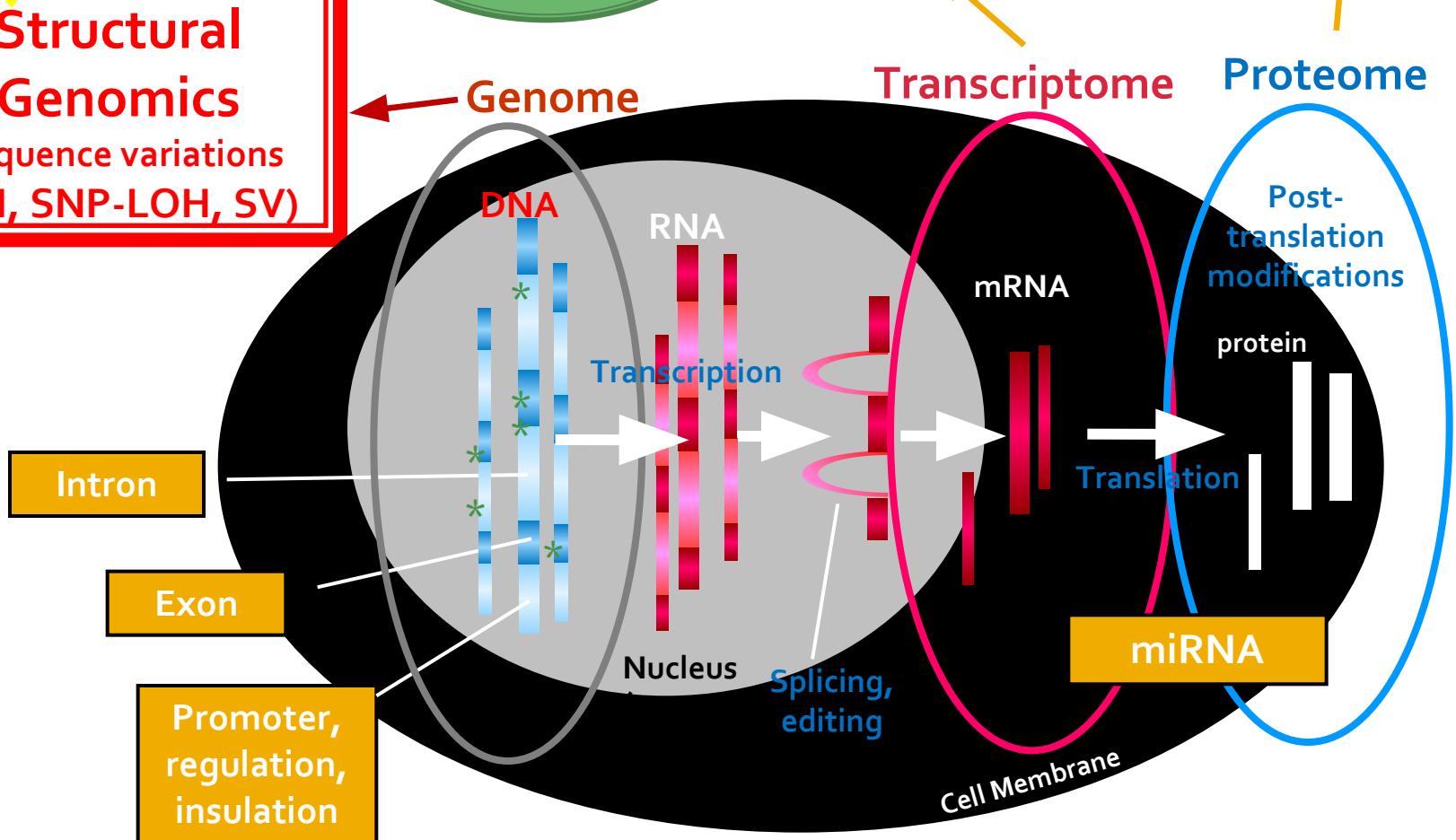
**YOU ARE
HERE**

Structural Genomics
Sequence variations
(CN, SNP-LOH, SV)

Regulatory Genomics
Methylation /
Chromatin state ...
(CH₃, HiC...)

Fonctional Genomics
Gene expression /
splicing...
(GEXa, Q-PCR,
RNAseq...)

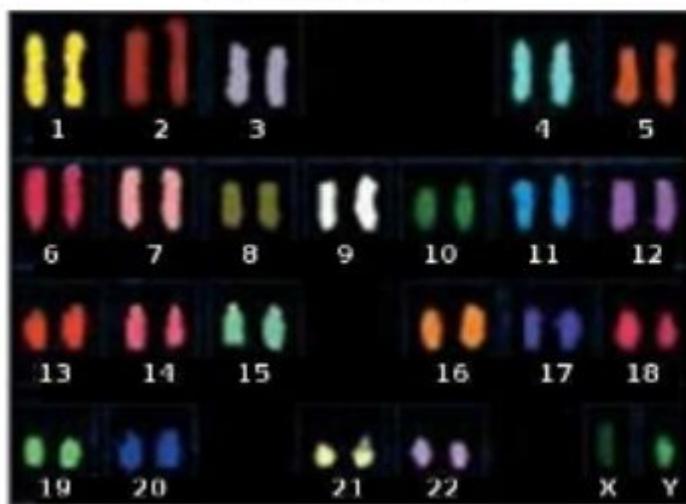
Proteomics
(Antibody arrays,
2D EP +MS/MS,
HPLC+MS / MS, ...)



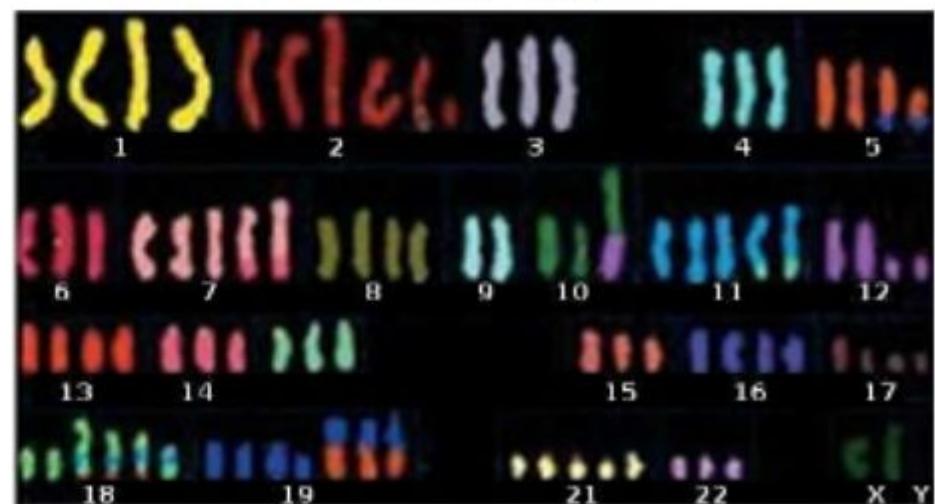
Copy number anomalies and cancer

Credit : Philippe HUPPÉ (2008)

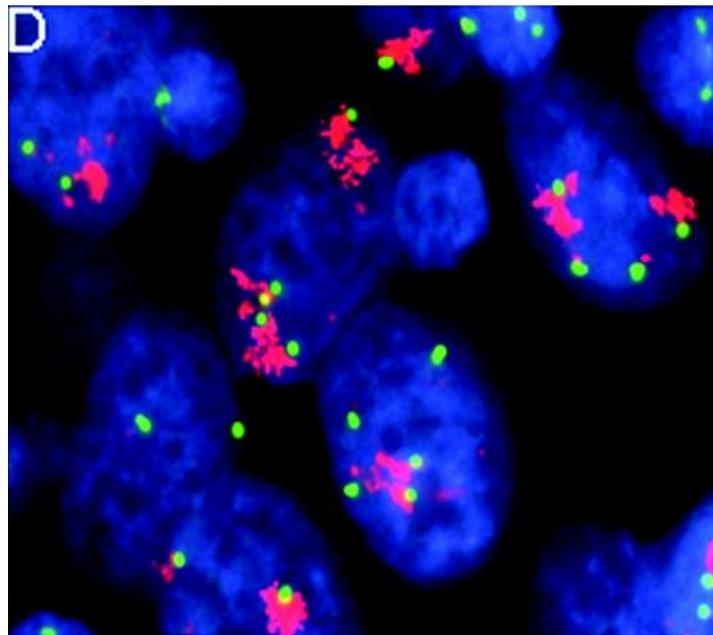
Normal cell



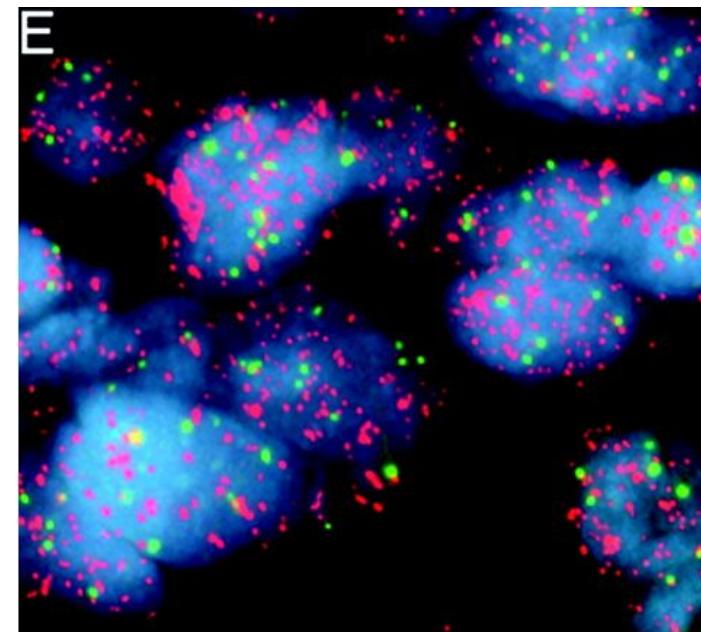
Tumor cell



Chromosomal amplifications

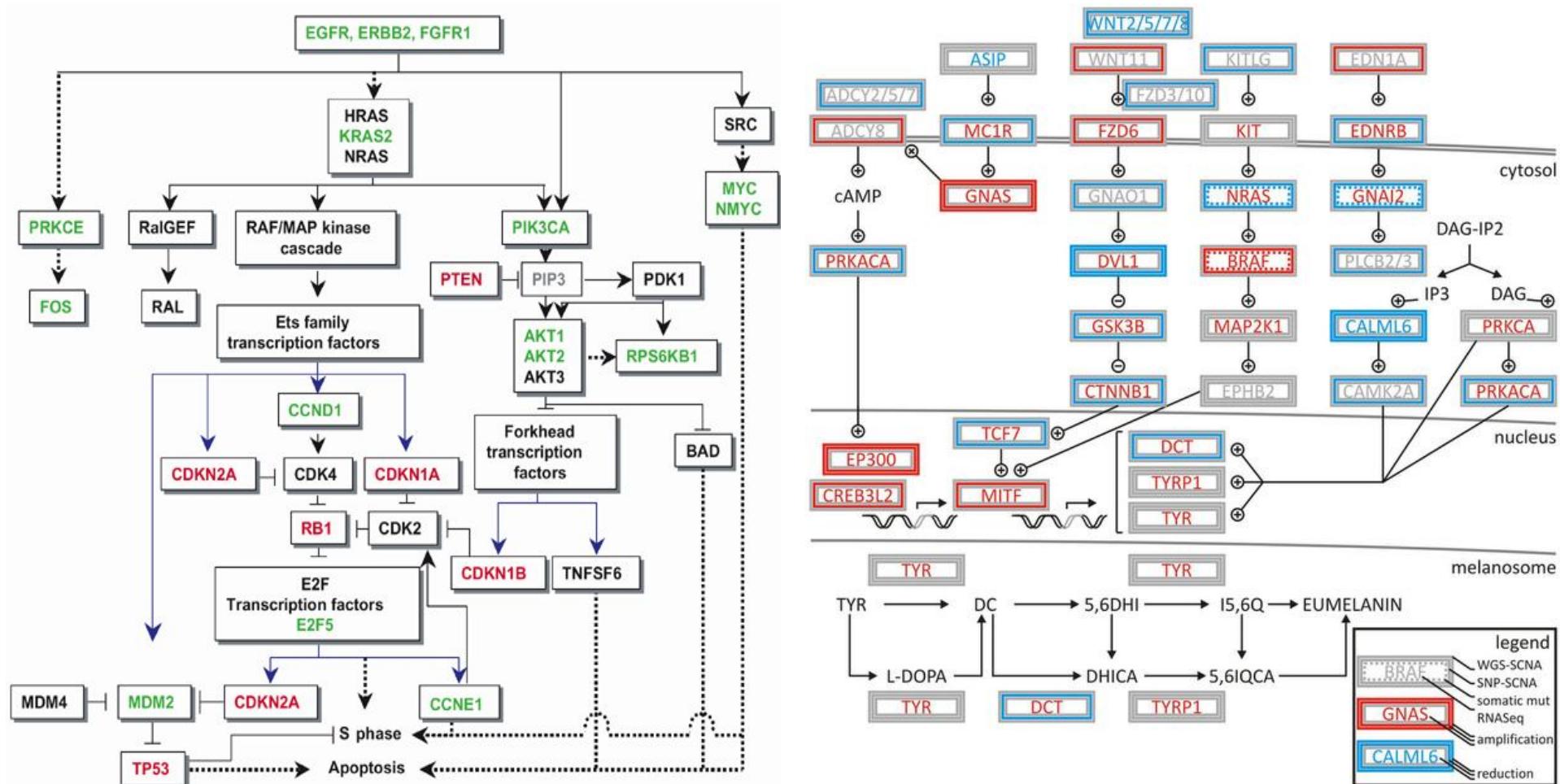


EGFR amplification in lung cancer as **HSR**
(*homogeneously stained region*)

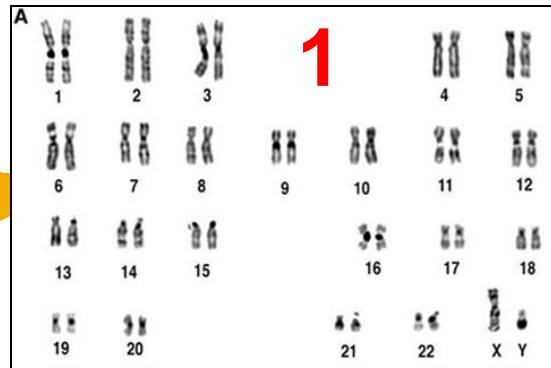


EGFR amplification in lung cancer as
double-minutes

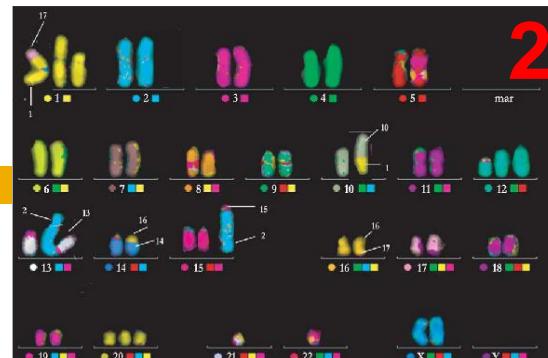
Copy number anomalies and cancer : “Target” genes



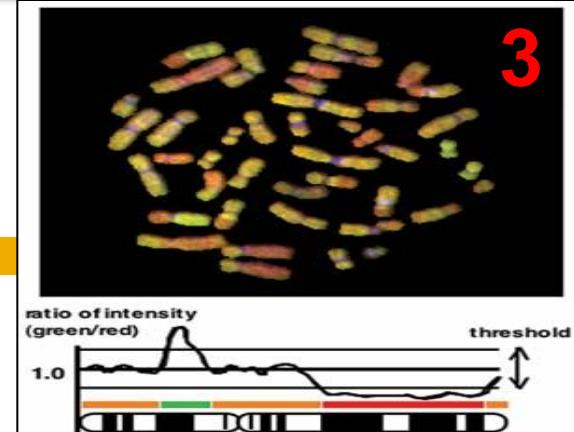
History



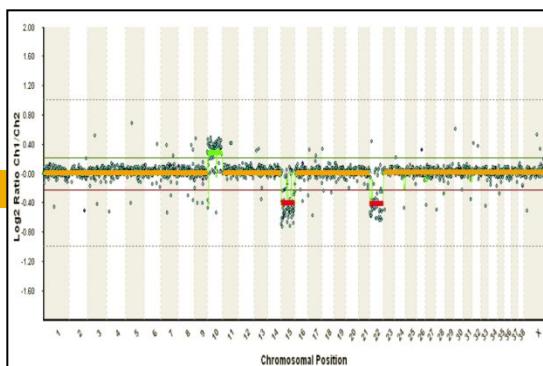
196x : Karyotype



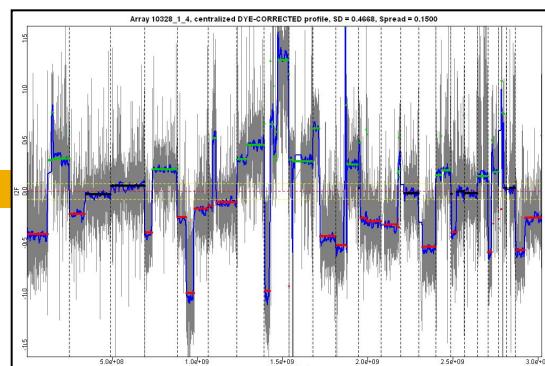
1993 : Spectral
karyotyping (SKY)



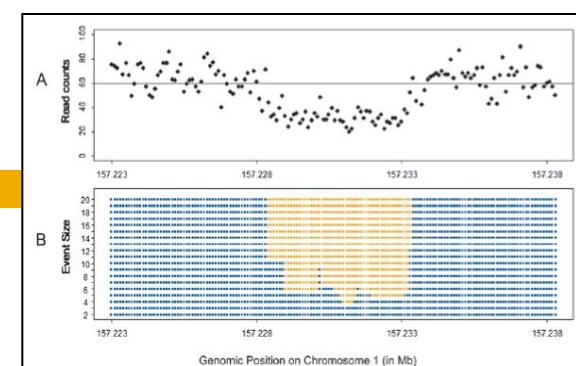
199x : CGH on
chromosomes



200x :
cDNA/BAC-based

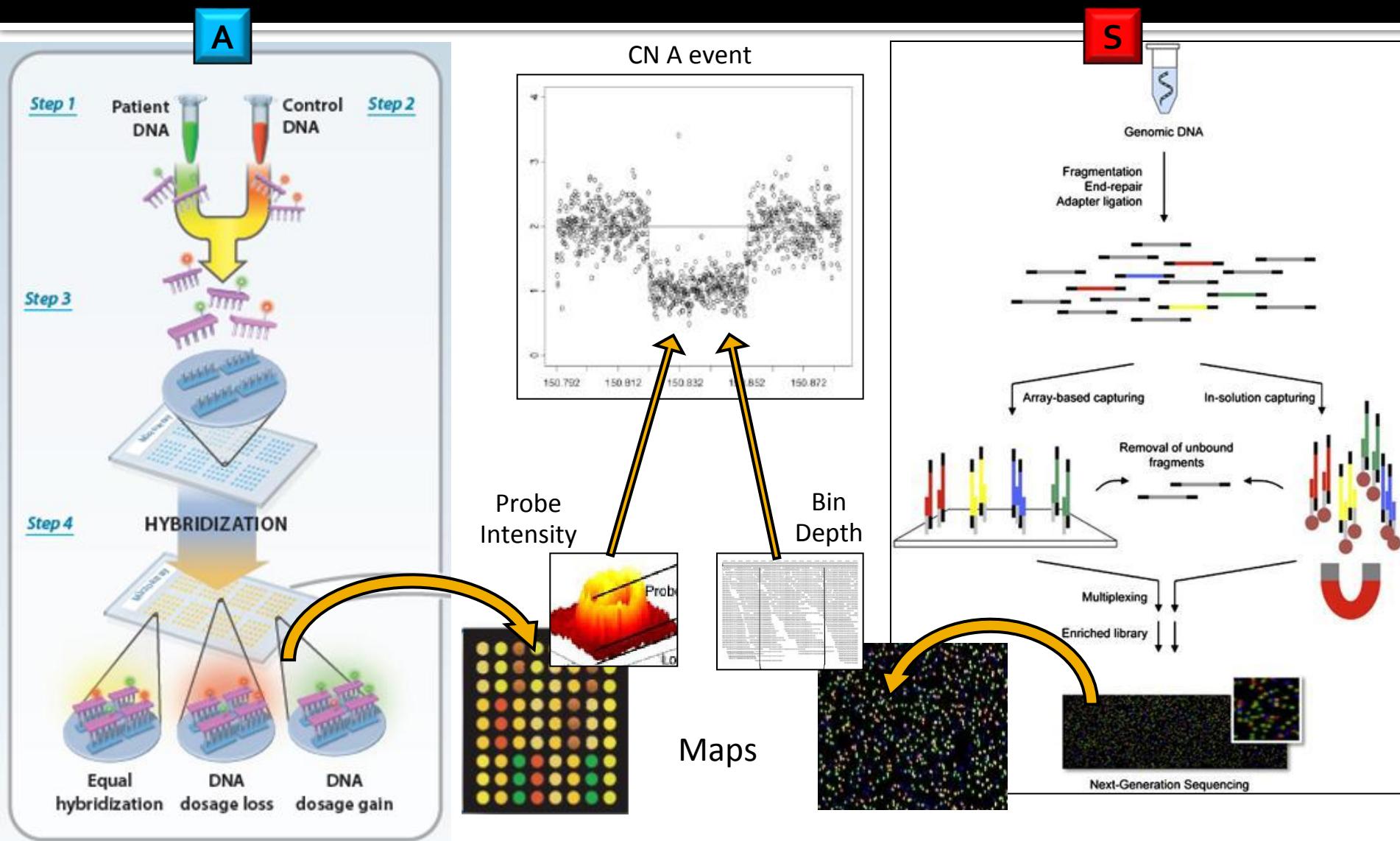


2005 : oligo-based
CGH array



201x : Read-depth
from NGS (WGS / WES)

Technical principles

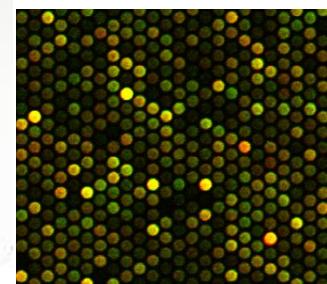


A

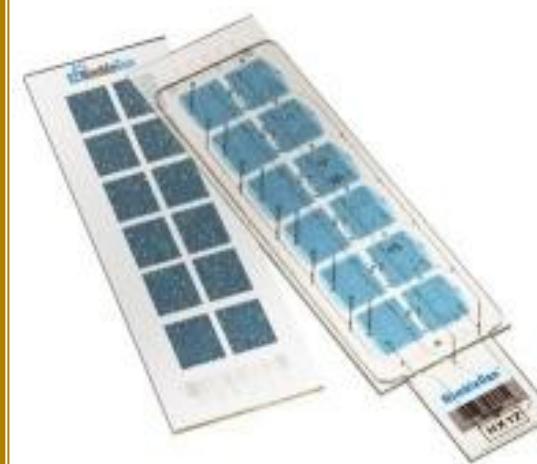
AGILENT



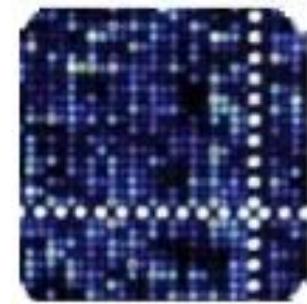
In-situ
printed,
Dual color



ROCHE-NIMBLEGENE



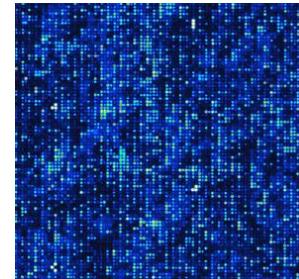
In-situ
printed,
Dual color



AFFYMETRIX



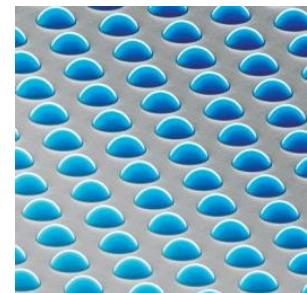
Photolithographic
synthesis,
Single color



ILLUMINA

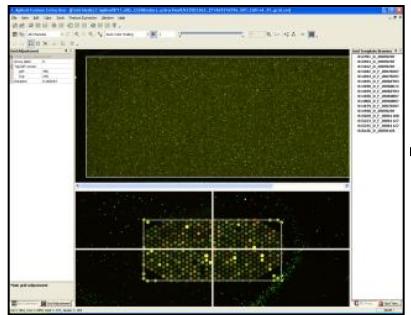


Coated
on beads,
Dual color

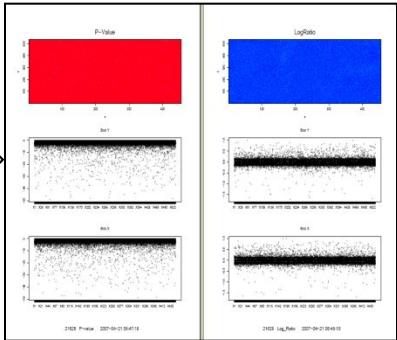


Bioinformatics analysis workflow

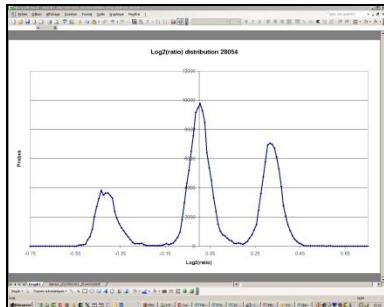
Signals acquisition



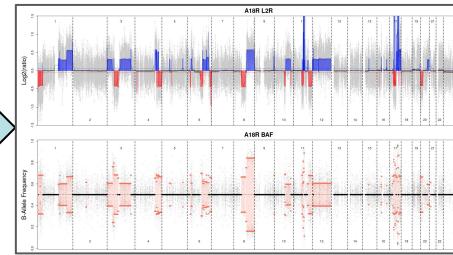
Quality controls



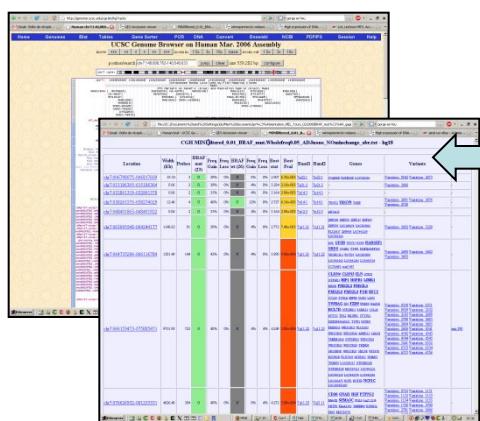
Normalization, centralization



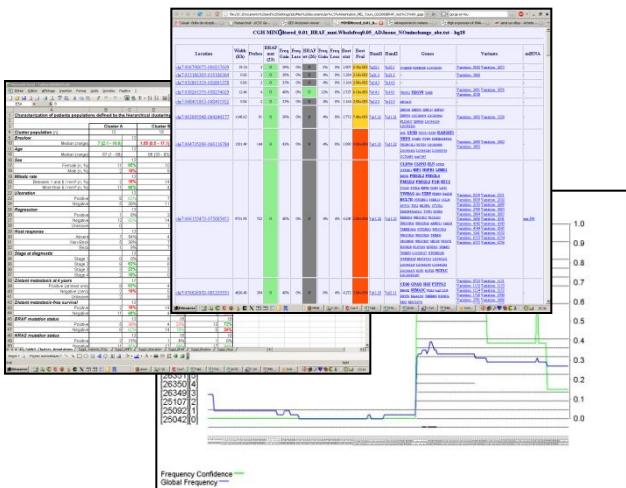
Segmentation, calling



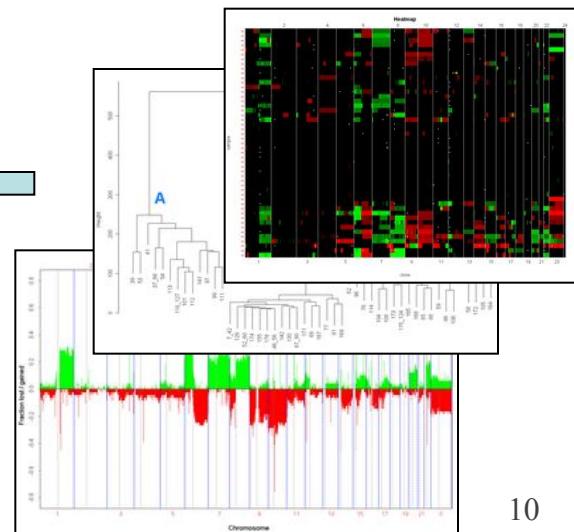
Annotation



Identification of genomic regions of interest



Analysis



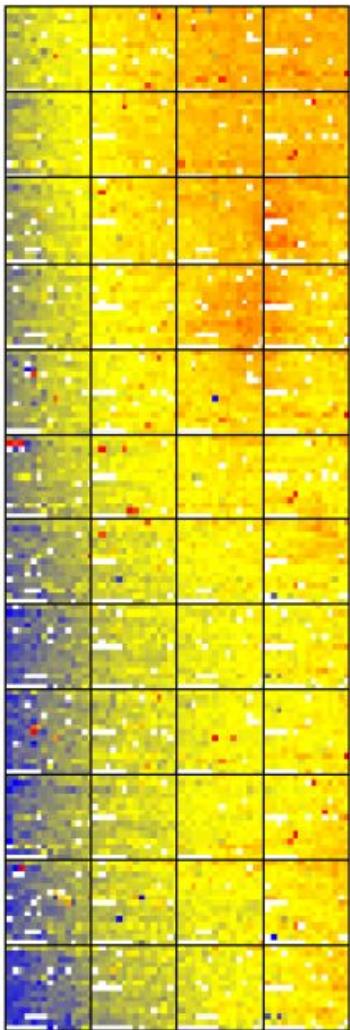
Normalization

Removing / reducing sources of bias

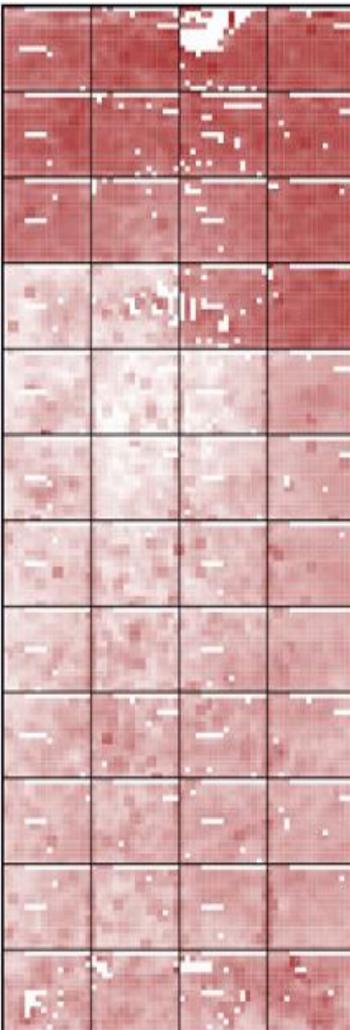
Spatial biases (legacy)

A

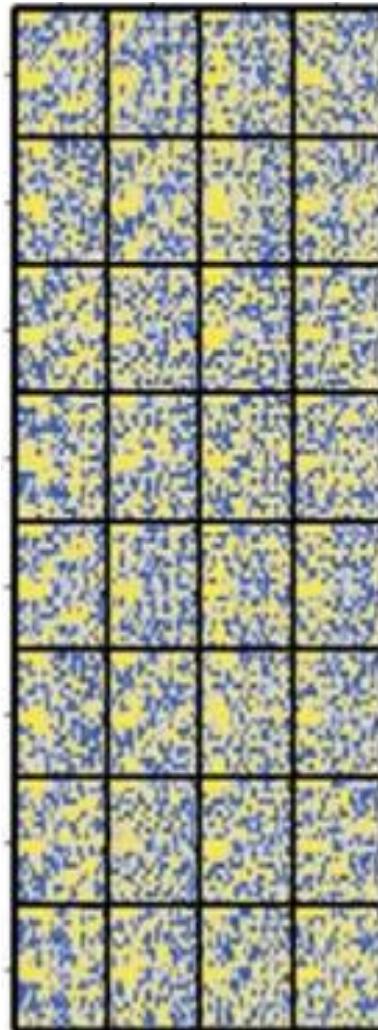
Gradient



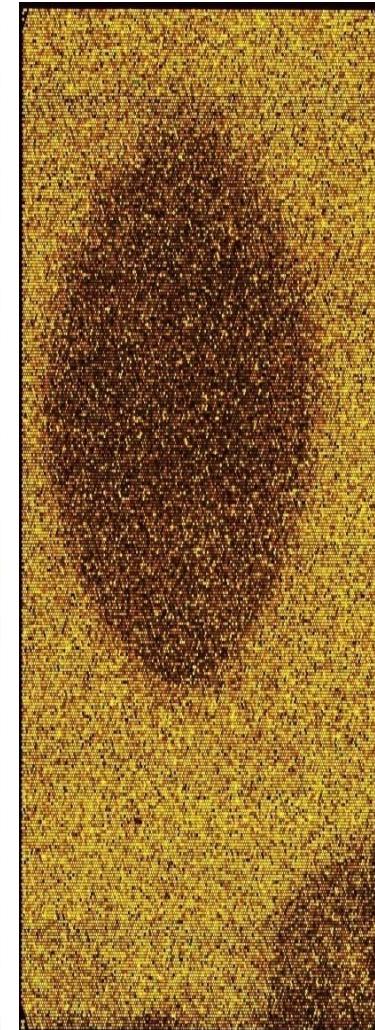
Spotter



Print-tip

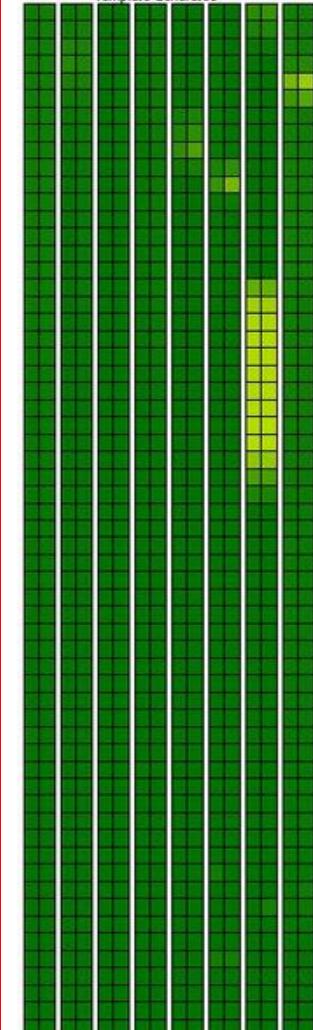


Leak

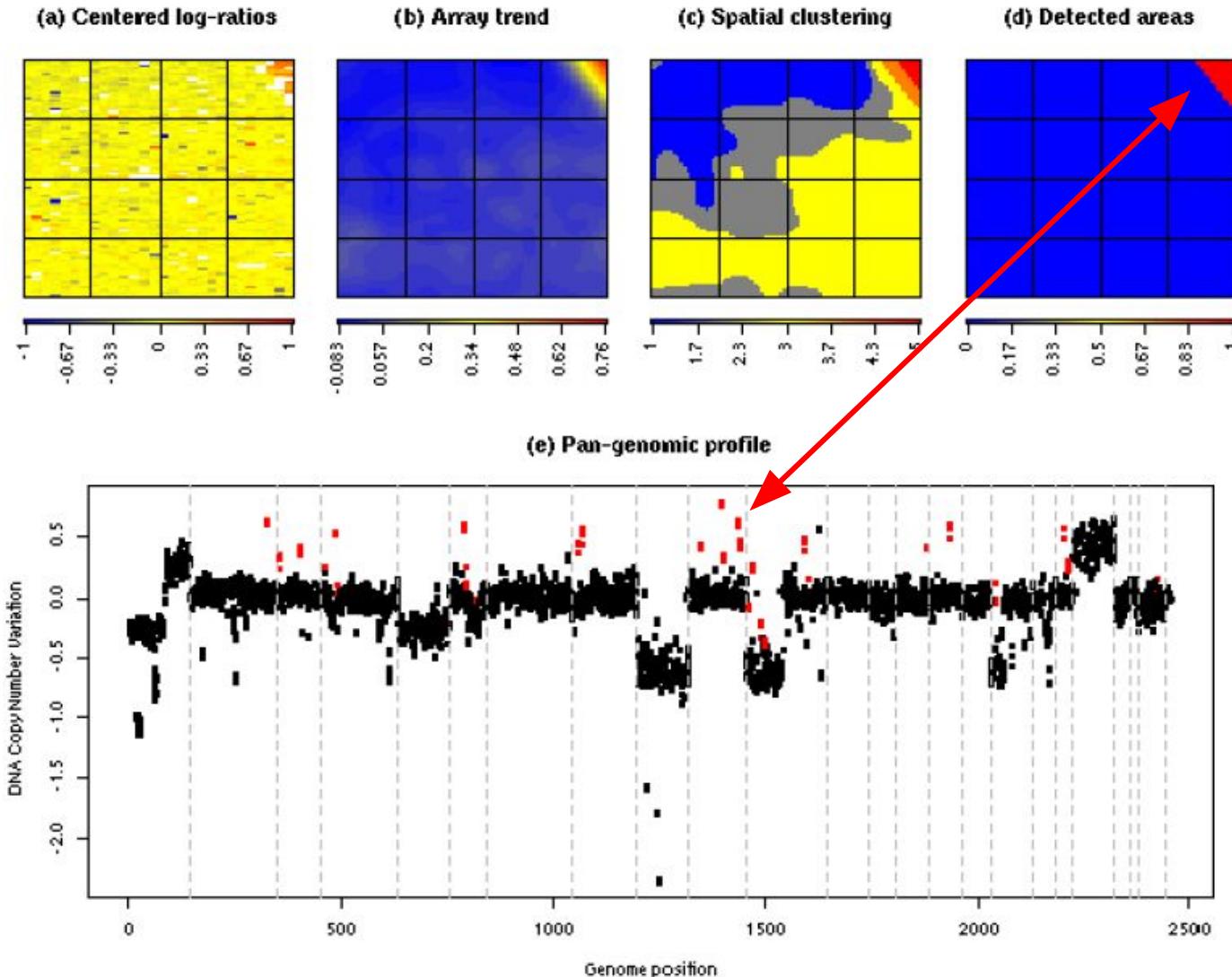


S

Density



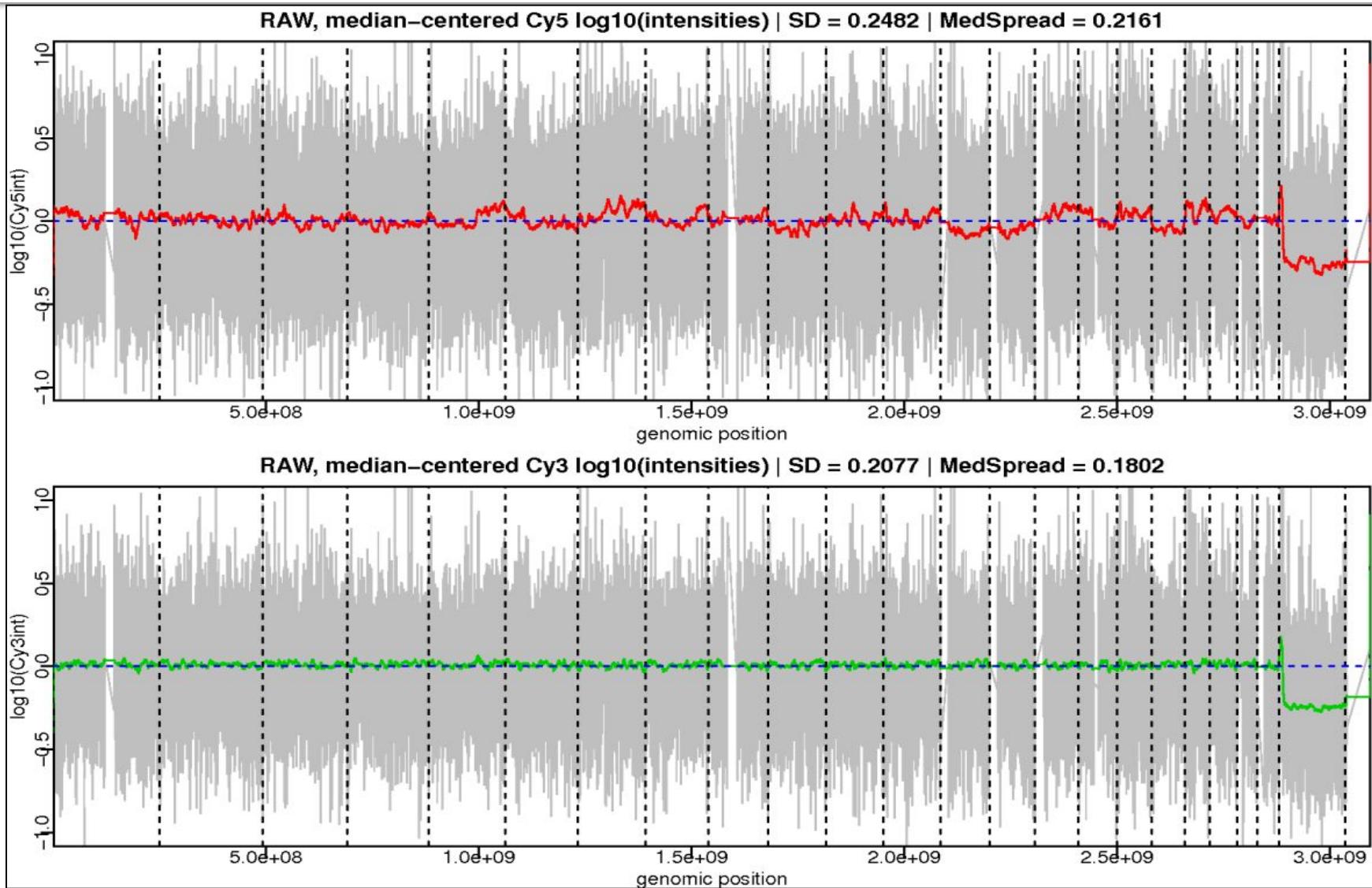
Spatial biases correction



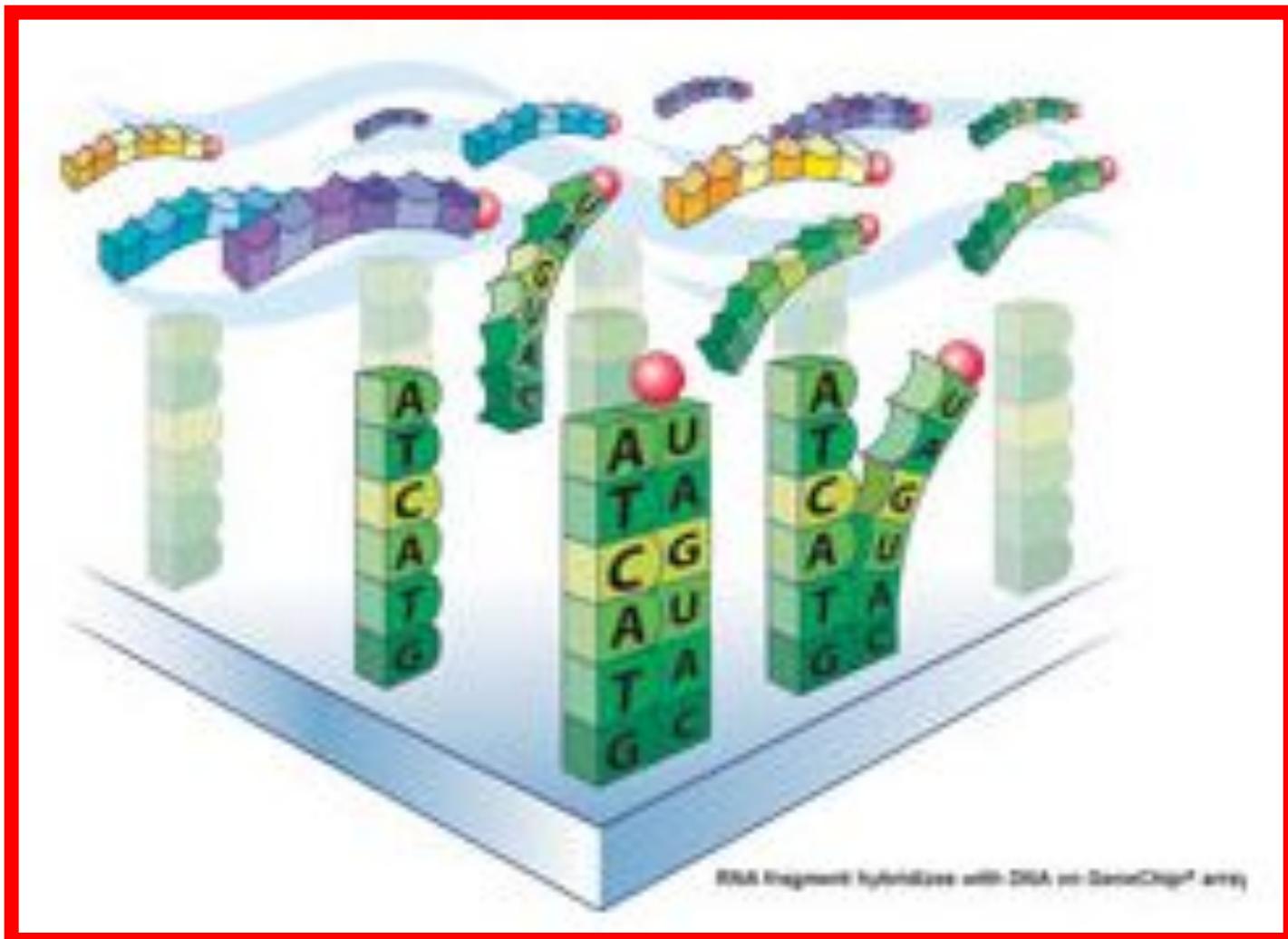
A

S

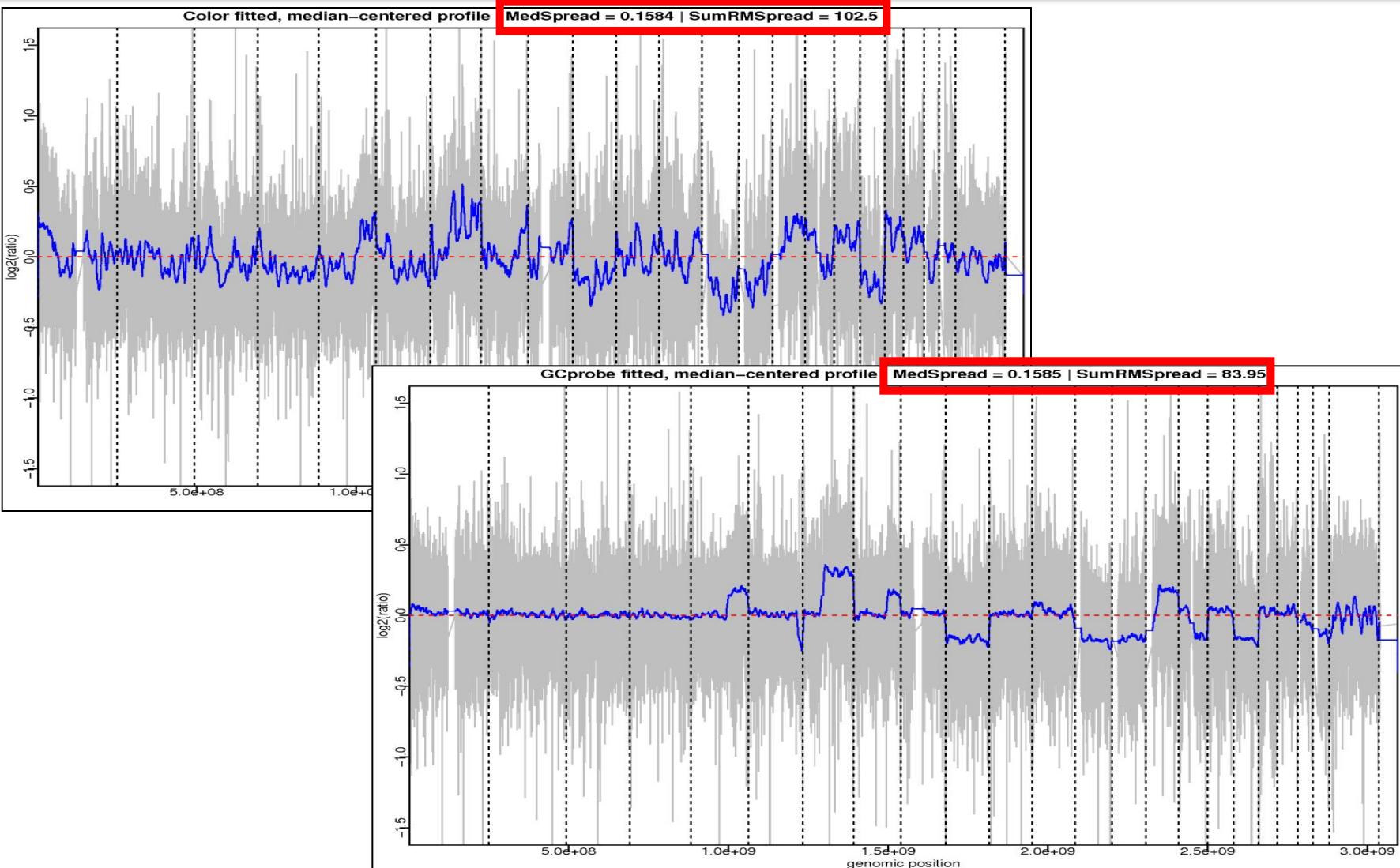
Dye bias / Library bias



GC-content bias

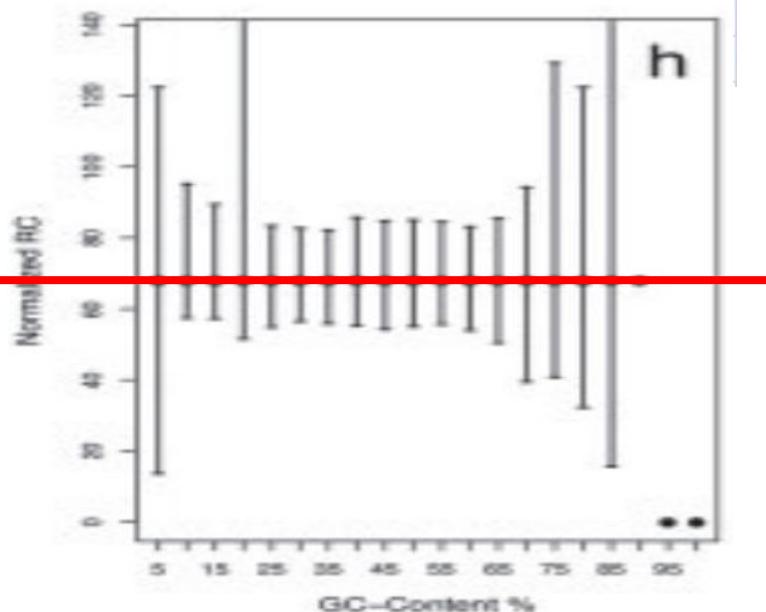
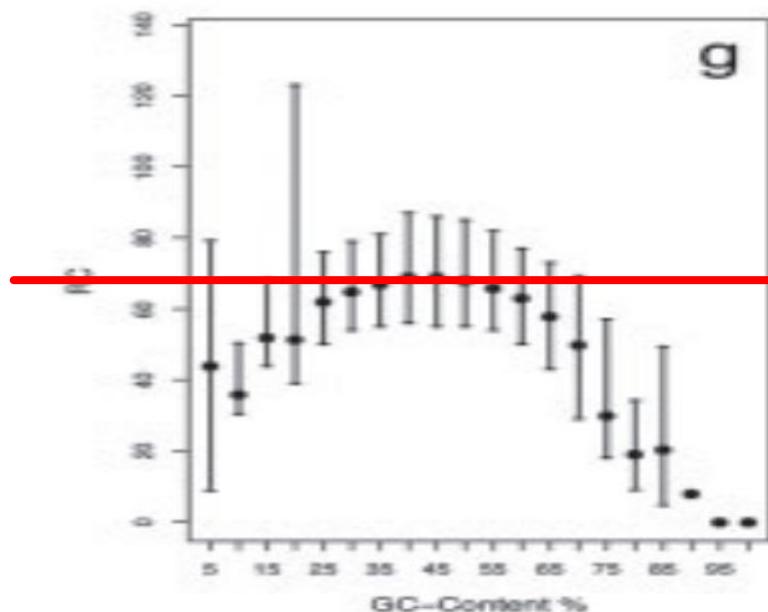


GC-content bias correction (lowess)

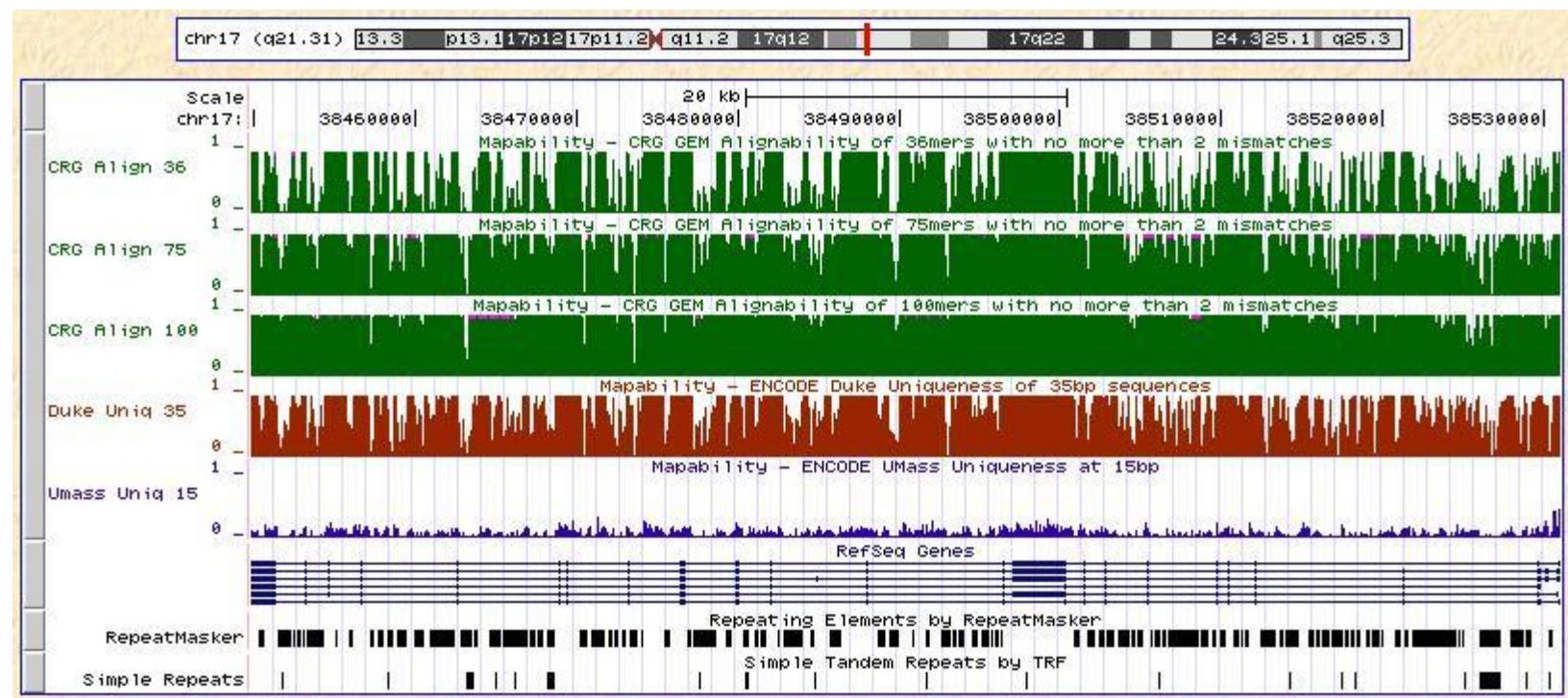


GC-content bias correction (per-centile shift)

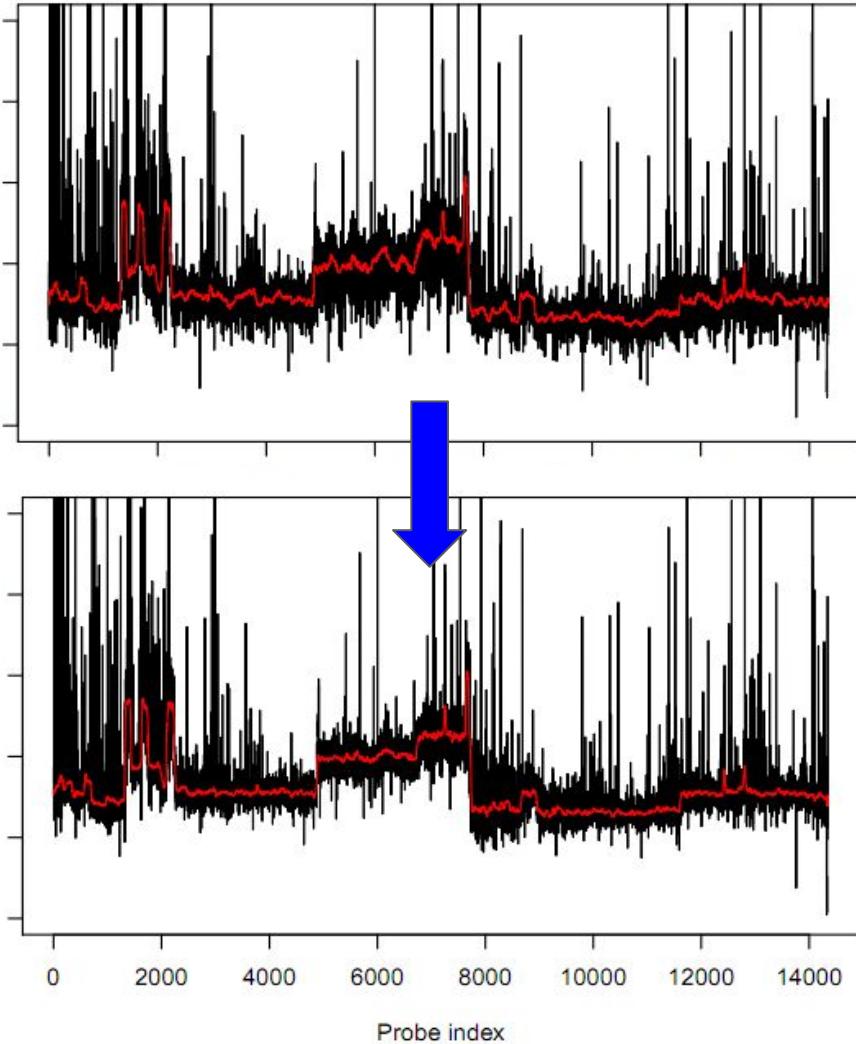
| chr | start | end | L2R | GC0b | GC50b | GC100b | GC250b | GC500b | GC1000b | GC2500b | GC5000b |
|------|-------|-------|-------|------|-------|--------|--------|--------|---------|---------|---------|
| chr1 | 65410 | 65461 | 0.015 | 0.44 | 0.36 | 0.37 | 0.33 | 0.34 | 0.31 | 0.32 | 0.33 |
| chr1 | 65462 | 65513 | -0.03 | 0.27 | 0.36 | 0.35 | 0.32 | 0.34 | 0.30 | 0.32 | 0.33 |
| chr1 | 65514 | 65565 | 0.013 | 0.35 | 0.32 | 0.33 | 0.34 | 0.34 | 0.28 | 0.32 | 0.33 |
| chr1 | 65566 | 65617 | 0.011 | 0.31 | 0.32 | 0.31 | 0.34 | 0.34 | 0.27 | 0.32 | 0.33 |
| chr1 | 65618 | 65669 | -0.04 | 0.29 | 0.30 | 0.30 | 0.35 | 0.35 | 0.27 | 0.32 | 0.33 |
| chr1 | 65670 | 65725 | 0.01 | 0.34 | 0.27 | 0.31 | 0.35 | 0.33 | 0.27 | 0.32 | 0.33 |
| chr1 | 65732 | 65781 | 0.07 | 0.24 | 0.33 | 0.33 | 0.34 | 0.32 | 0.27 | 0.32 | 0.33 |



Reads mappability



Wave effect

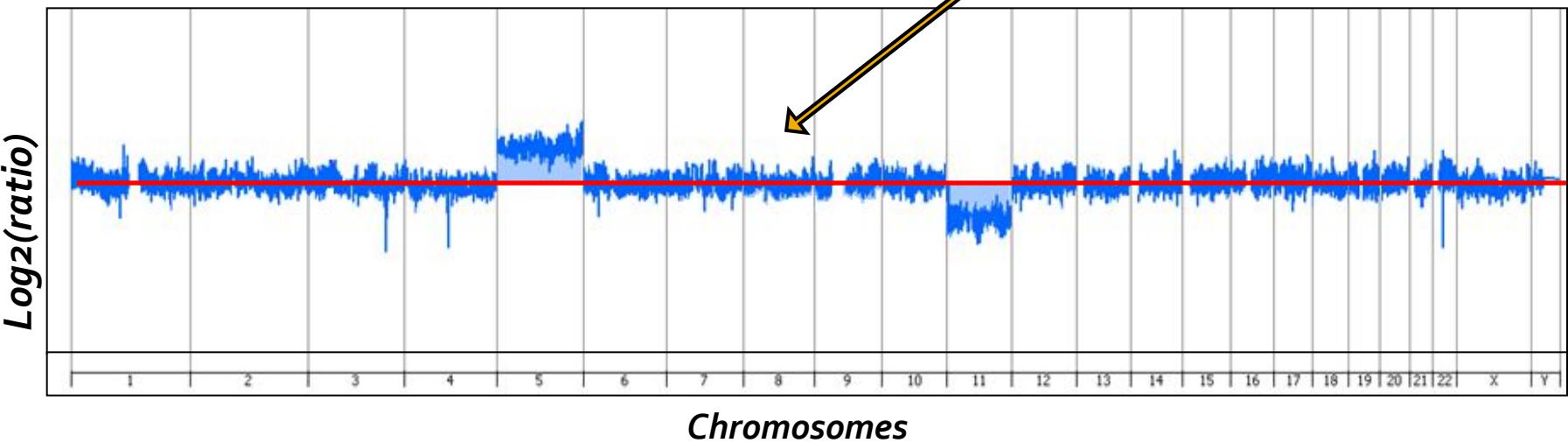
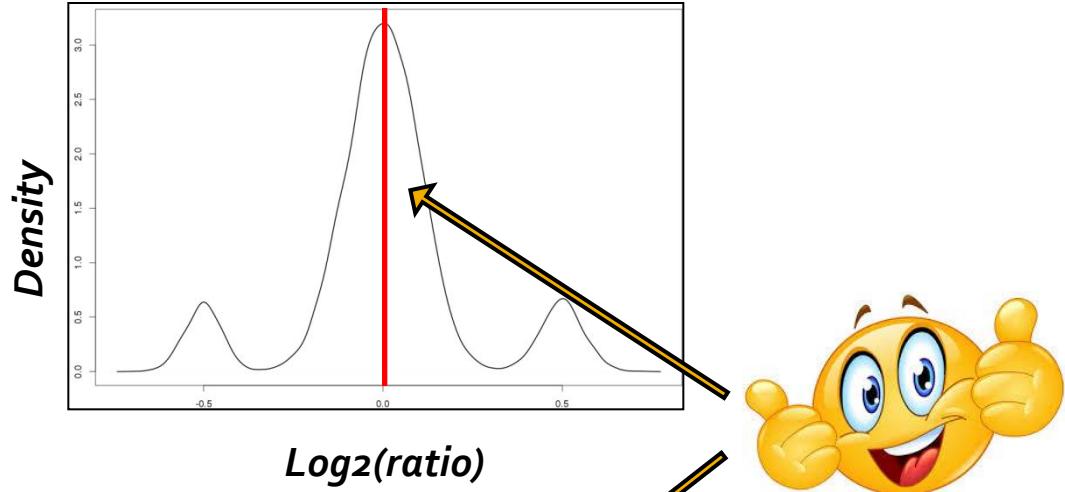


- *Joint normalization*
- *Spline smoothing using spatial correlation*
- *Cghseg R package*
- *Efficient with few samples (~5 - 10)*
- *Slow...*

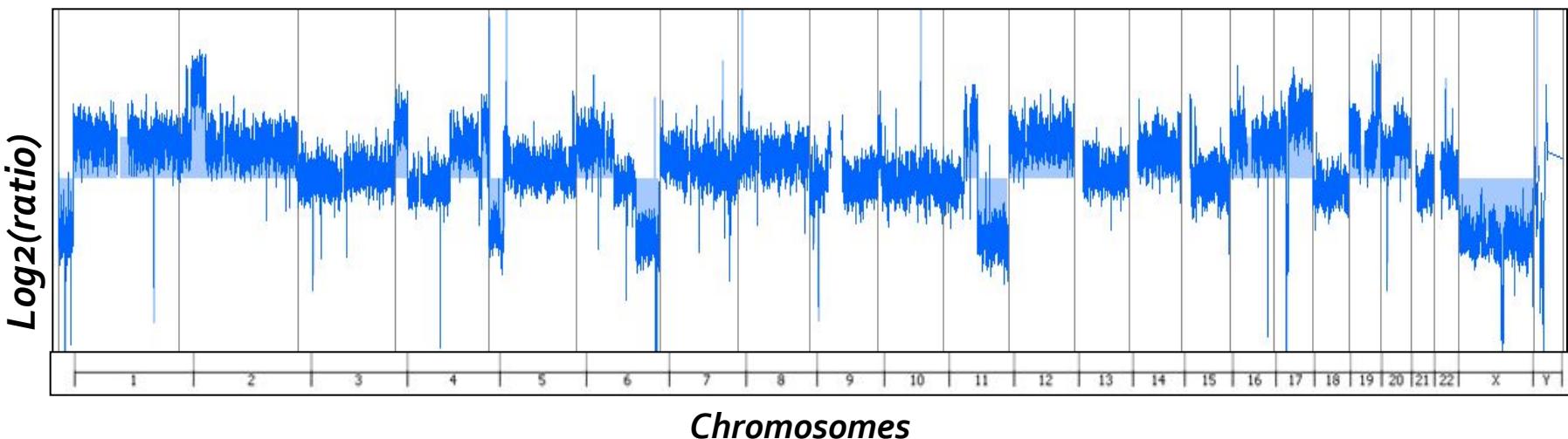
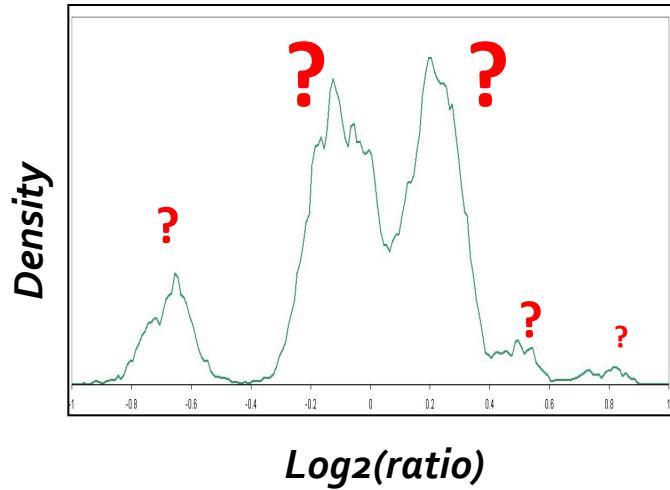
Centralization

Finding a basis

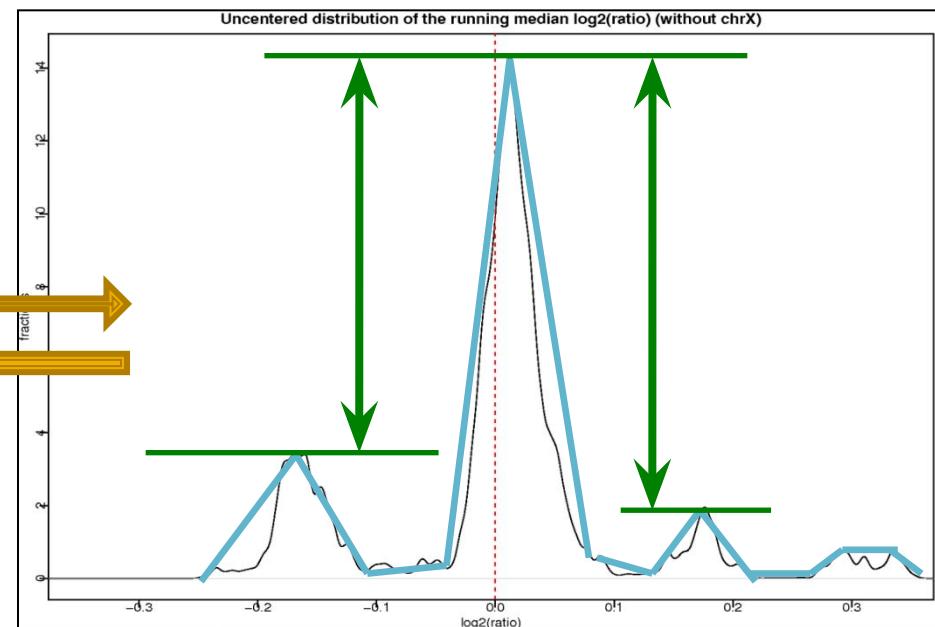
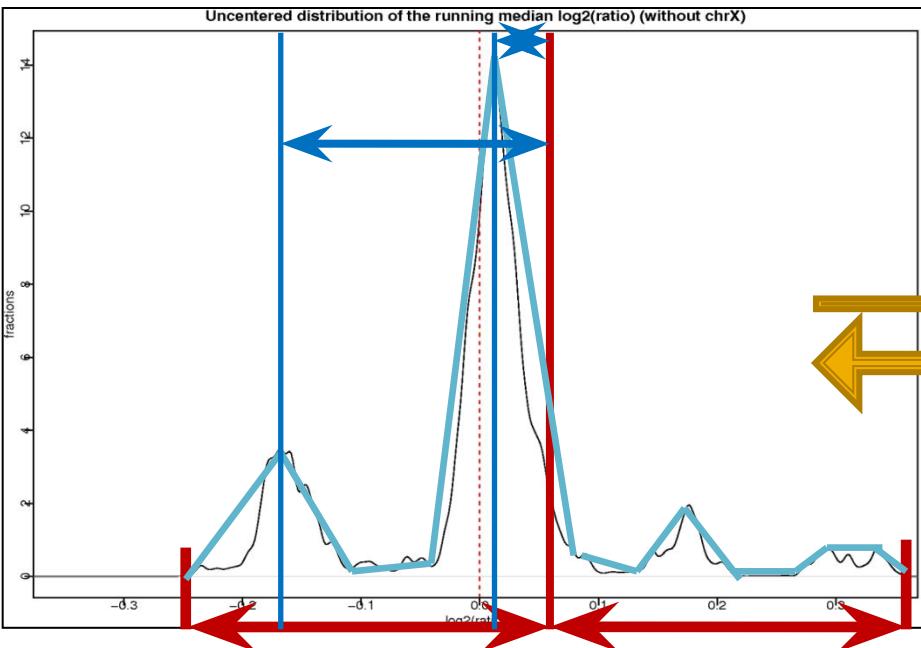
Centralization : An (synthetic) obvious example



Centralization : A typical cancer example



Centralization : Basic considerations



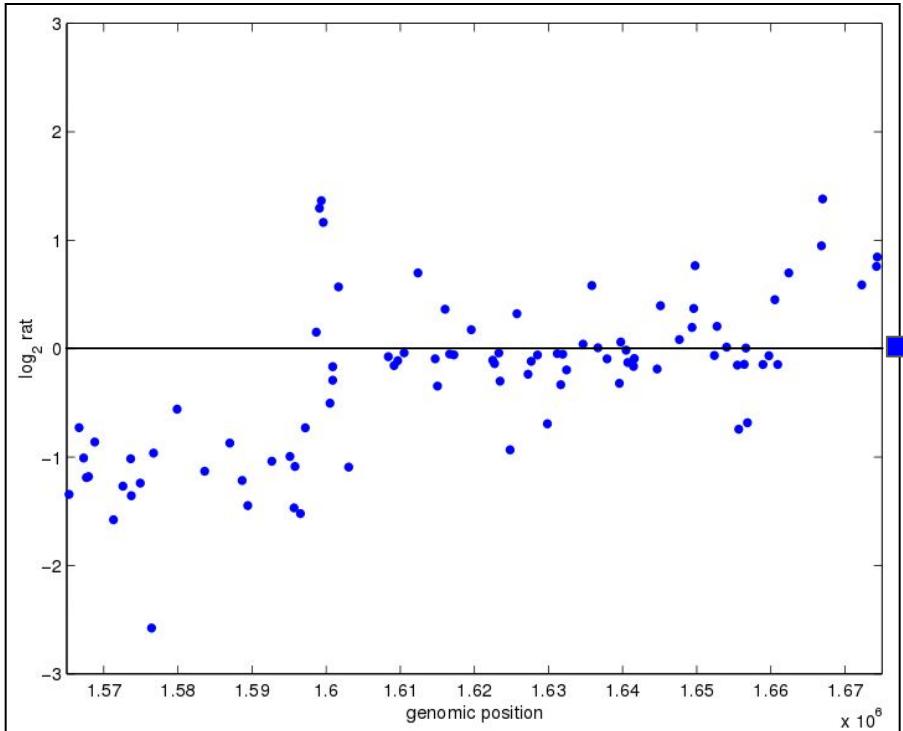
Segmentation

Data reduction

Segmentation : Data reduction

Credit : Stéphane ROBIN

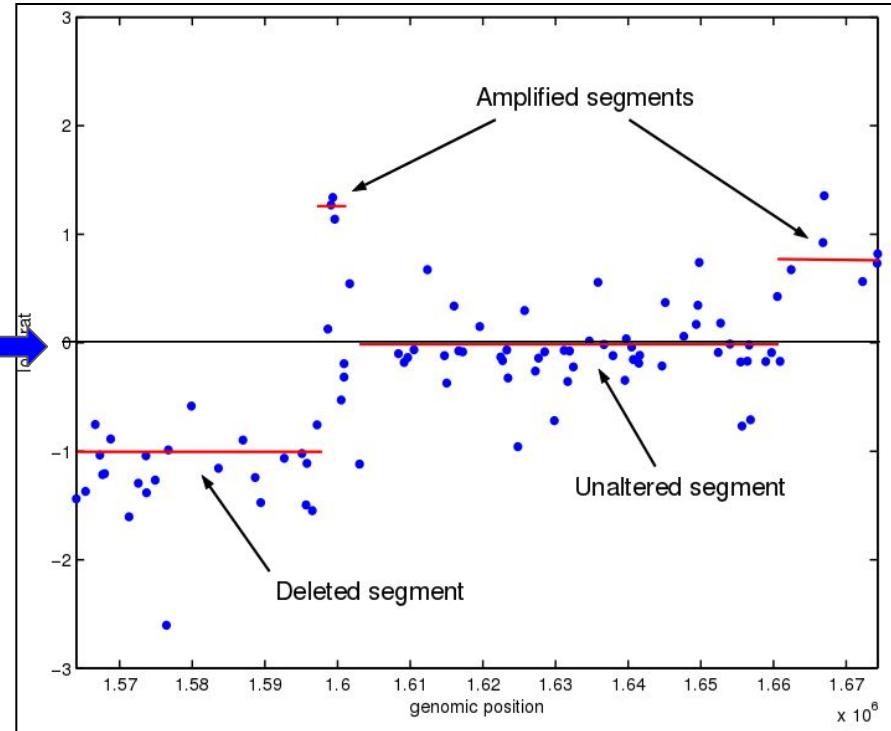
Raw profile



Numerous, noisy **positional measurements**

N = 100

Segmented profile



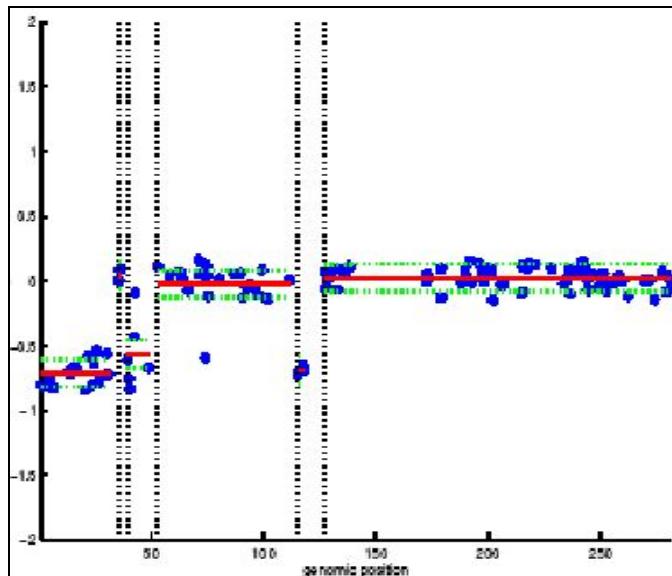
Reduced, denoised **genomic intervals**

S = 4

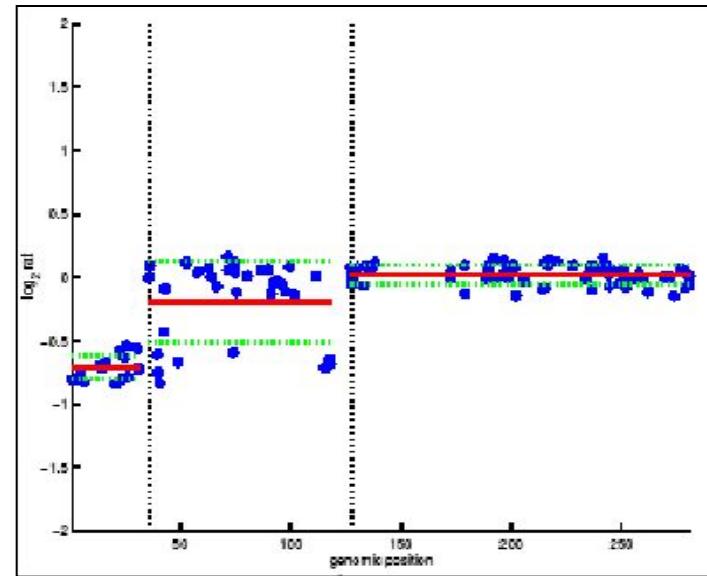
Segmentation : Challenging breakpoints detection

- Two unknowns for breakpoints :
 - Localization
 - Quantity
- Three families of algorithms :
 - Smoothers (wavelet)
 - Change-point
 - Binary segmentation (CBS)
 - Optimal partitionning (PELT)
 - HMM modeling (bioHMM)

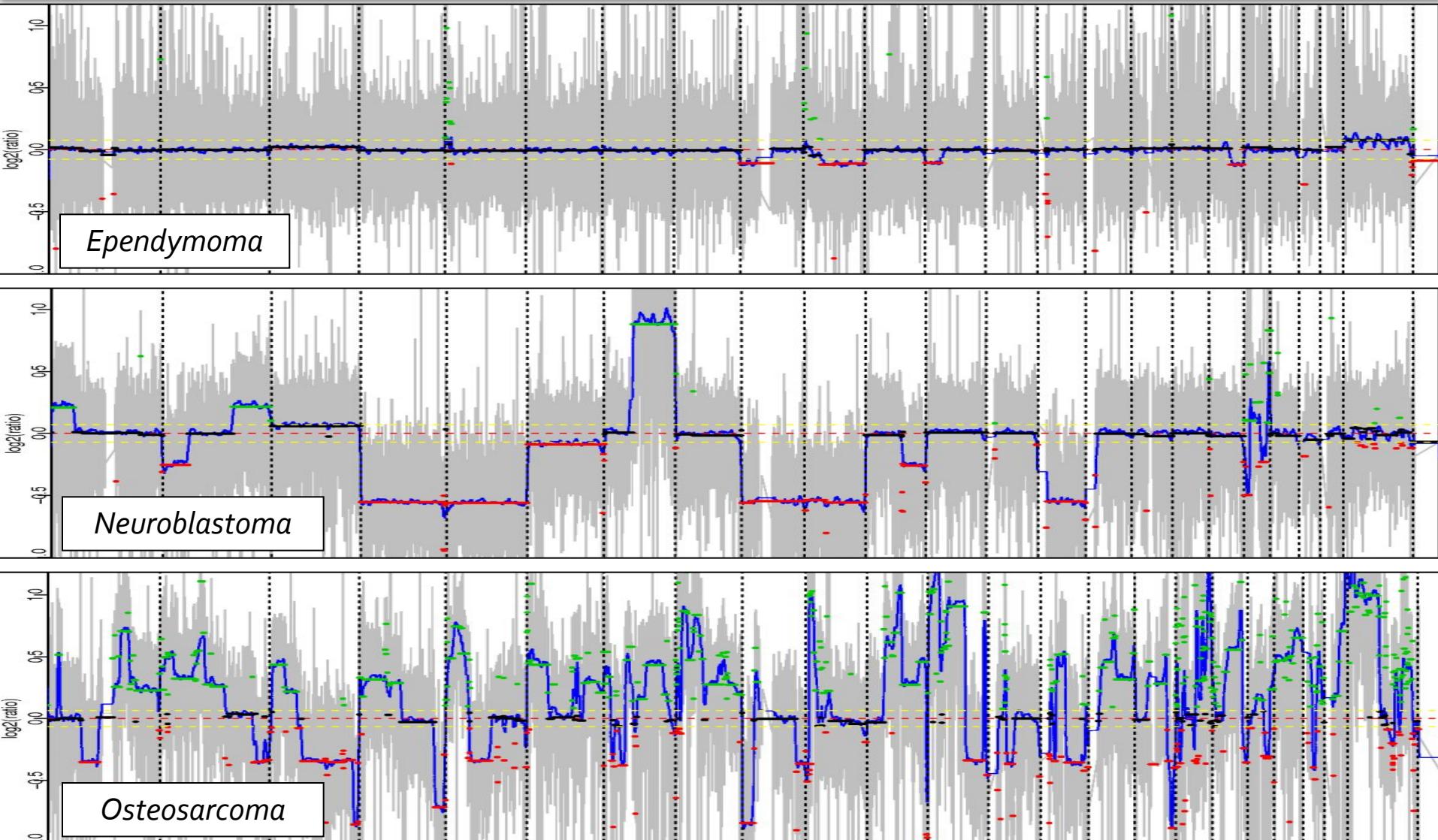
Homoscedastic (m)



Heteroscedastic (m, V)



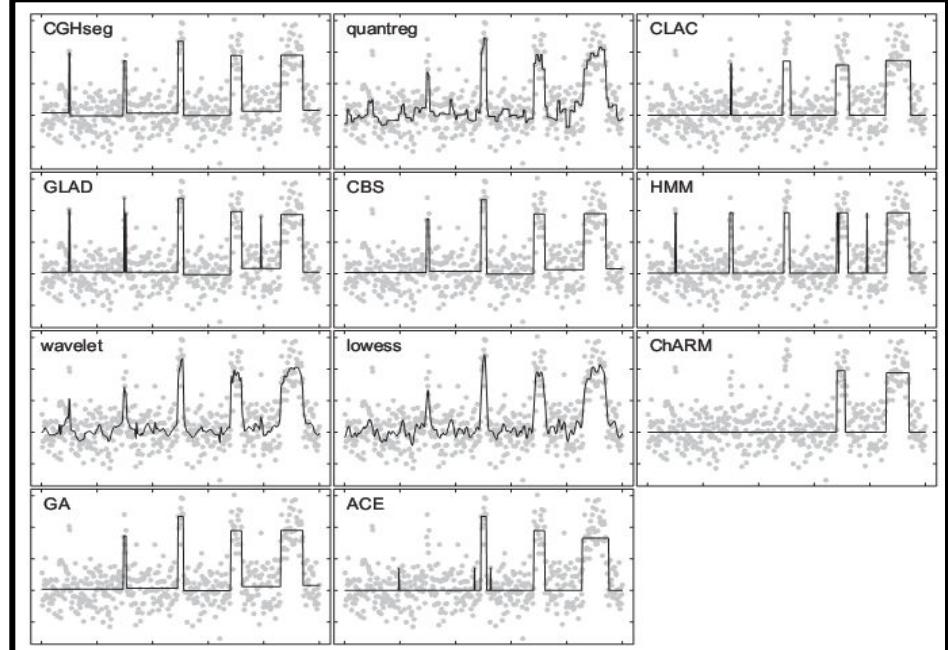
Segmentation : Tumoral complexity



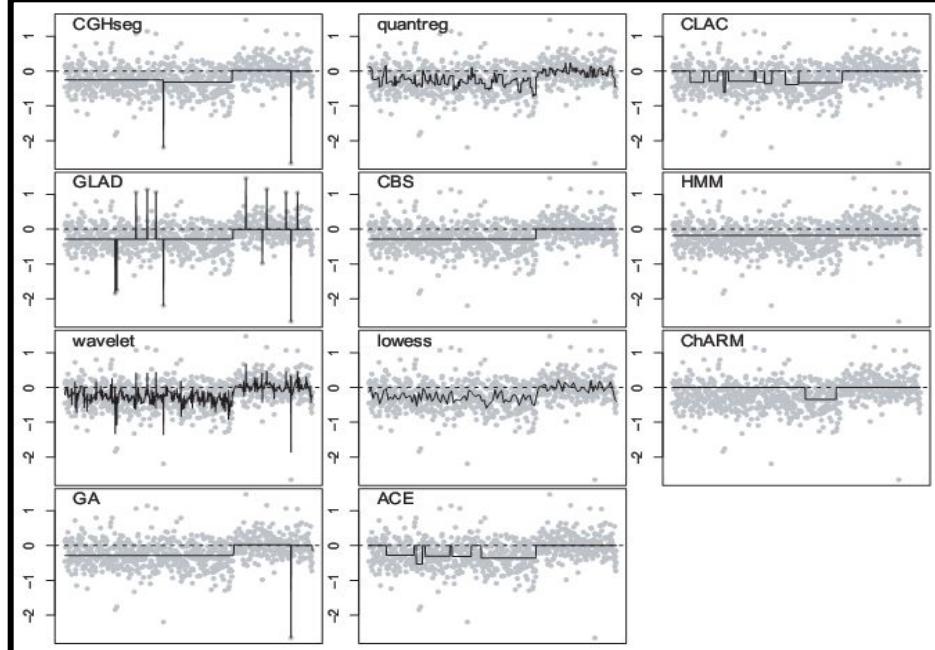
Segmentation : Several methods available

Lai et al. 2005

Synthetic data



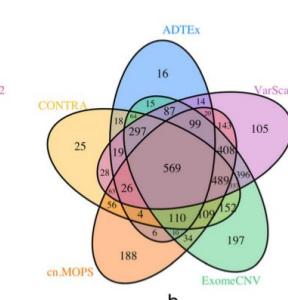
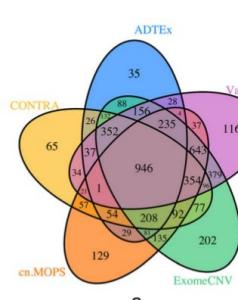
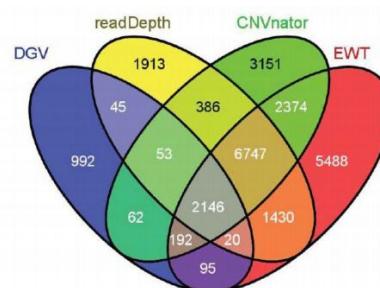
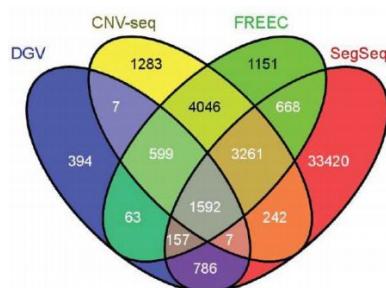
Tumoral profile



| Algorithm | Tool | Autor | Publication Year | Model type | Criterion | K choice |
|----------------------|-------------|-----------|------------------|-----------------|----------------------|-----------|
| Dynamic programming | CGH-plotter | Autio | 2003 | Other | Least-square | Ad hoc |
| Genetic algorithm | NA | Jong | 2003 | Heteroscedastic | Max likelihood | Penalized |
| EM | aCGH-HMM | Fridlyand | 2004 | HMM | Max likelihood | Penalized |
| Adaptative smoothing | GLAD | Hupé | 2004 | Homoscedastic | Max likelihood | Penalized |
| CBS | DNAcopy | Olshen | 2004 | Homoscedastic | Partial sums / perm° | Ad hoc |
| Dynamic programming | CGHseg | Picard | 2005 | Homoscedastic | Max likelihood | Penalized |

Segmentation : Several methods available

| Method | Reference | Language | Control required? | Input format | GC correction | single-end/pair-end | Methodology characteristics |
|------------------|-----------|------------|-------------------|--------------------|---------------|---------------------|--------------------------------------|
| CNV-seq | [15] | R, perl | Yes | hits | No | single-end | statistical testing |
| FREEC | [21] | C | Optional | SAM,BAM,bed,etc. | Optional | both | LASSO regression |
| readDepth | [22] | R | No | bed | Yes | both | CBS, LOESS regression |
| CNVnator | [23] | C | No | BAM | Yes | both | mean shift algorithm |
| SegSeq | [14] | Matlab | Yes | bed | No | single-end | statistical testing,CBS |
| EWT (RDExplorer) | [11] | R, python | No | BAM | Yes | single-end | statistical testing |
| cnD | [16] | D | No | SAM,BAM | No | both | HMM, Viterbi algorithm |
| CNVer | [17] | C | No | BAM | Yes | pair-end | maximum-likelihood, graphic flow |
| CopySeq | [18] | Java | No | BAM | Yes | pair-end | MAP estimator |
| rSW-seq | [19] | NA | Yes | NA | Yes | single-end | Smith-Waterman algorithm |
| CNAseg | [20] | R | Yes | BAM | No | pair-end | wavelet transform and HMM |
| CNAnorm | [24] | R | Yes | SAM,BAM | Yes | both | linear regression or CBS |
| cn.MOPS | [26] | R, C++ | multiple samples | BAM or data matrix | No | both | mixture of Poissons, MAP, EM, CBS |
| JointSLM | [27] | R, Fortran | multiple samples | data matrix | Yes | both | HMM, ML estimator, Viterbi algorithm |



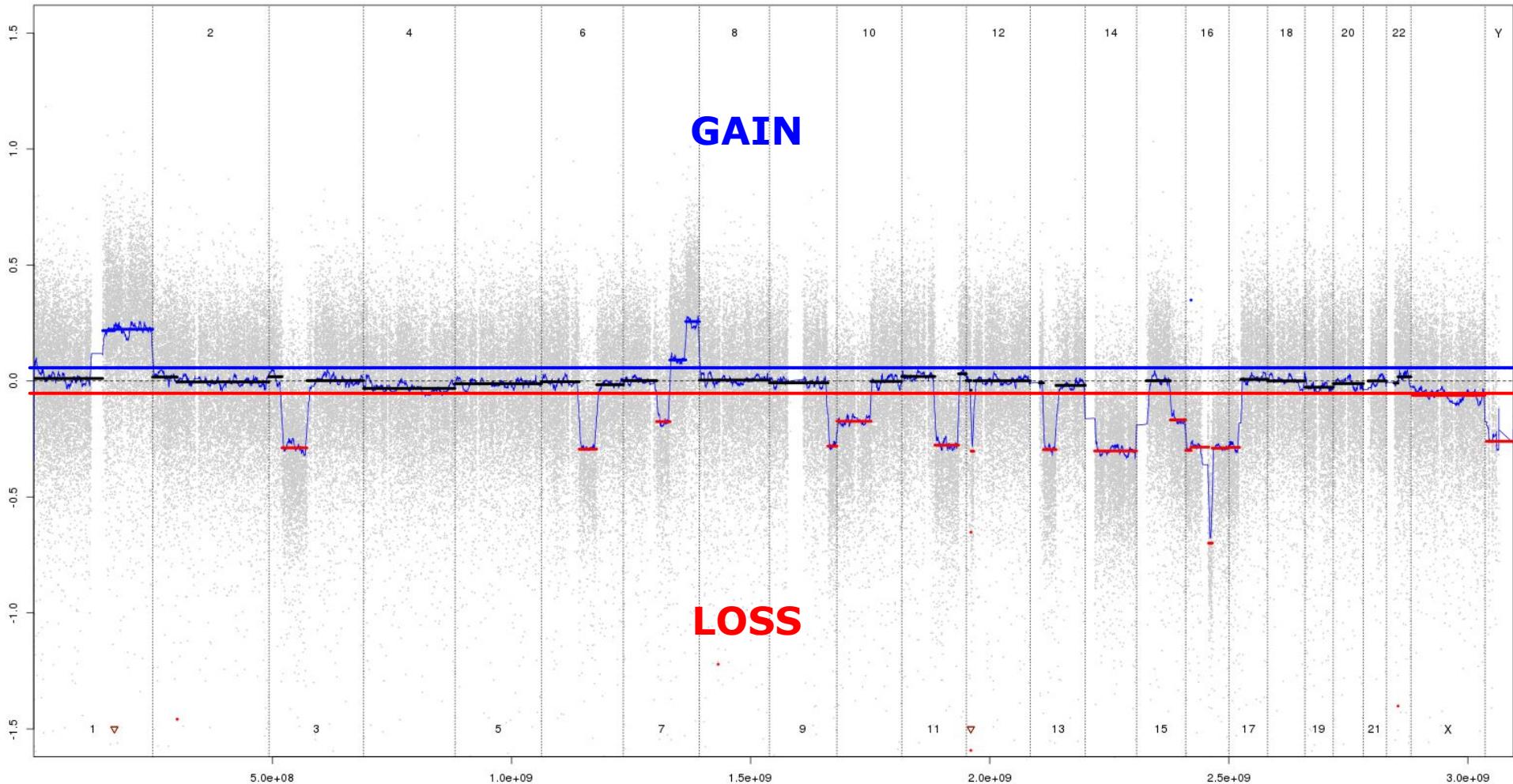
Duan et al.
Plos One
2013

Zare et al.
BMC Bioinformatics
2017

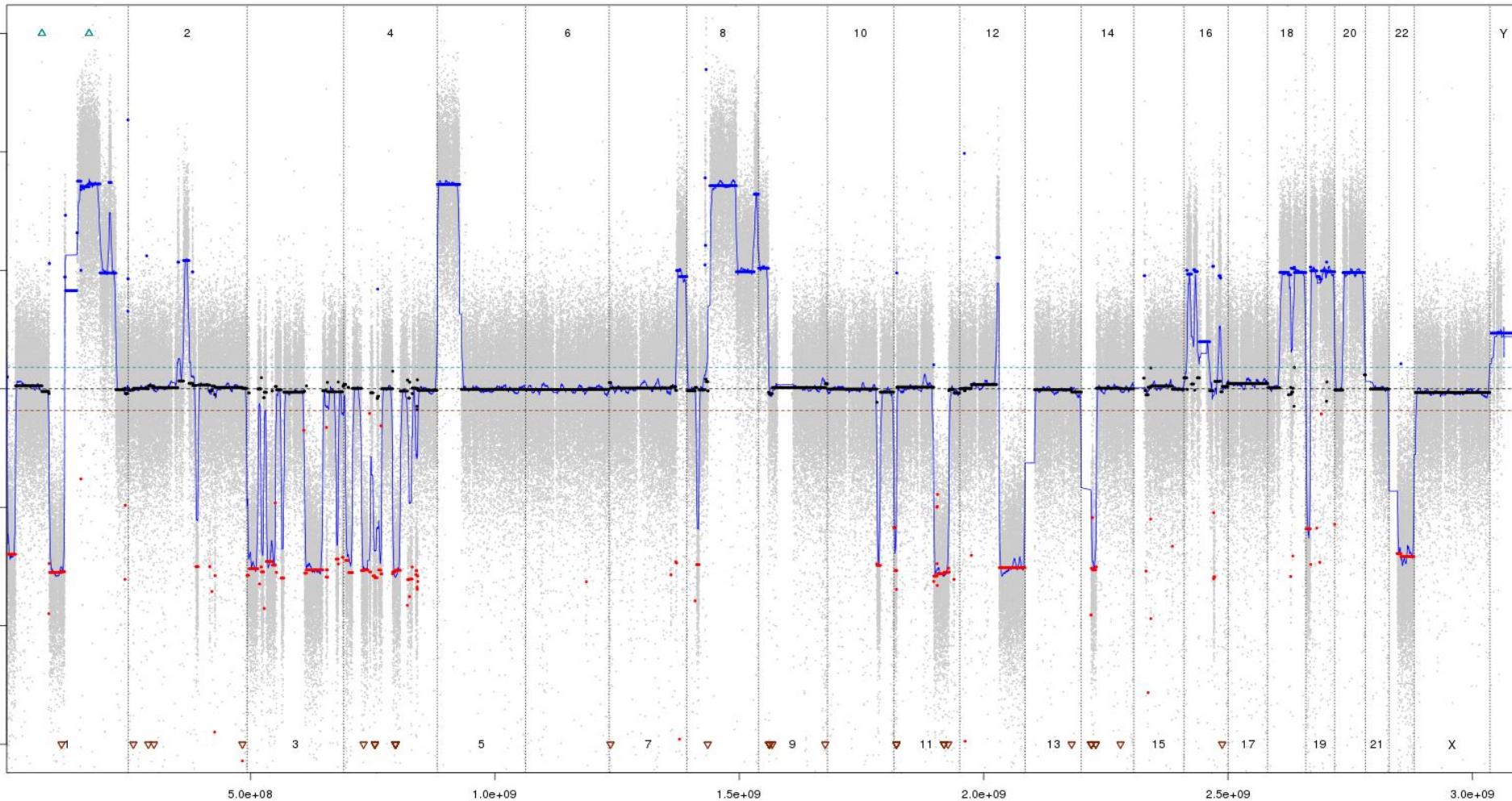
Calling

Who's who

Aberration calling

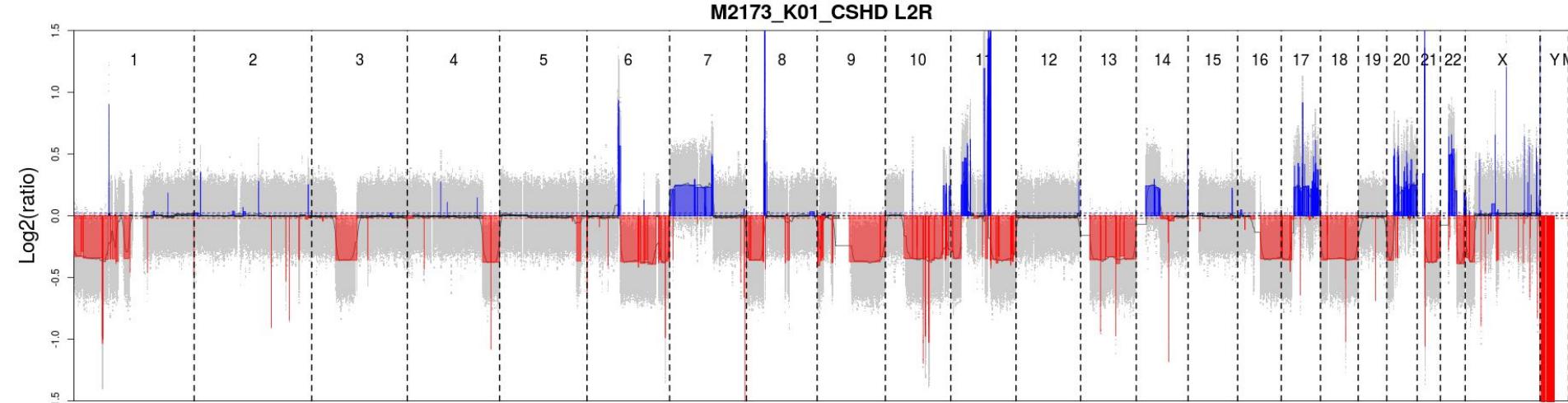


Sample profile

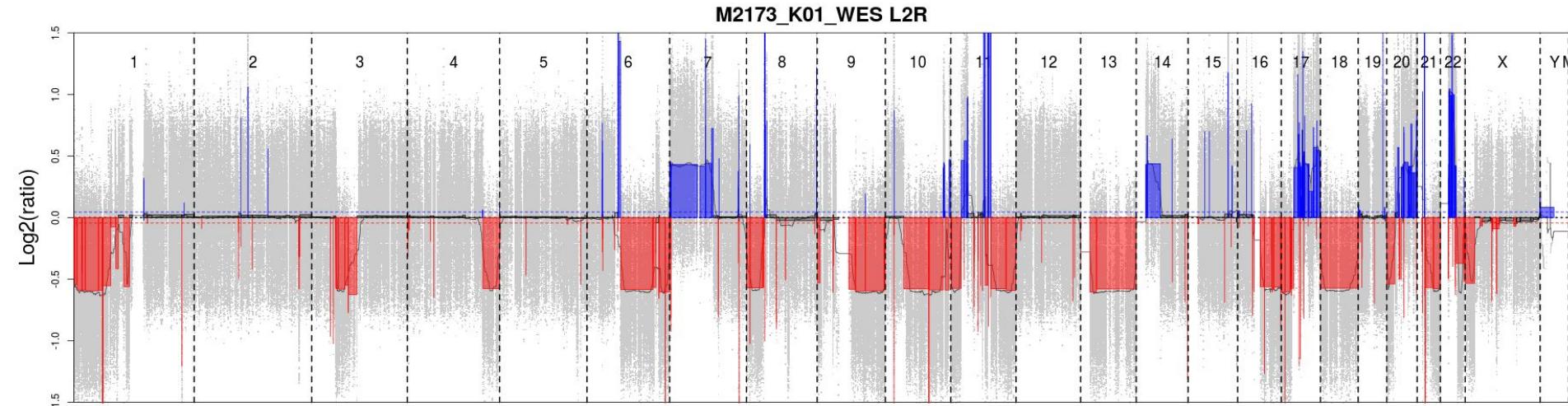


NGS or microarray ?

NGS versus microarrays



MICROARRAY (Affymetrix CytoScan HD)

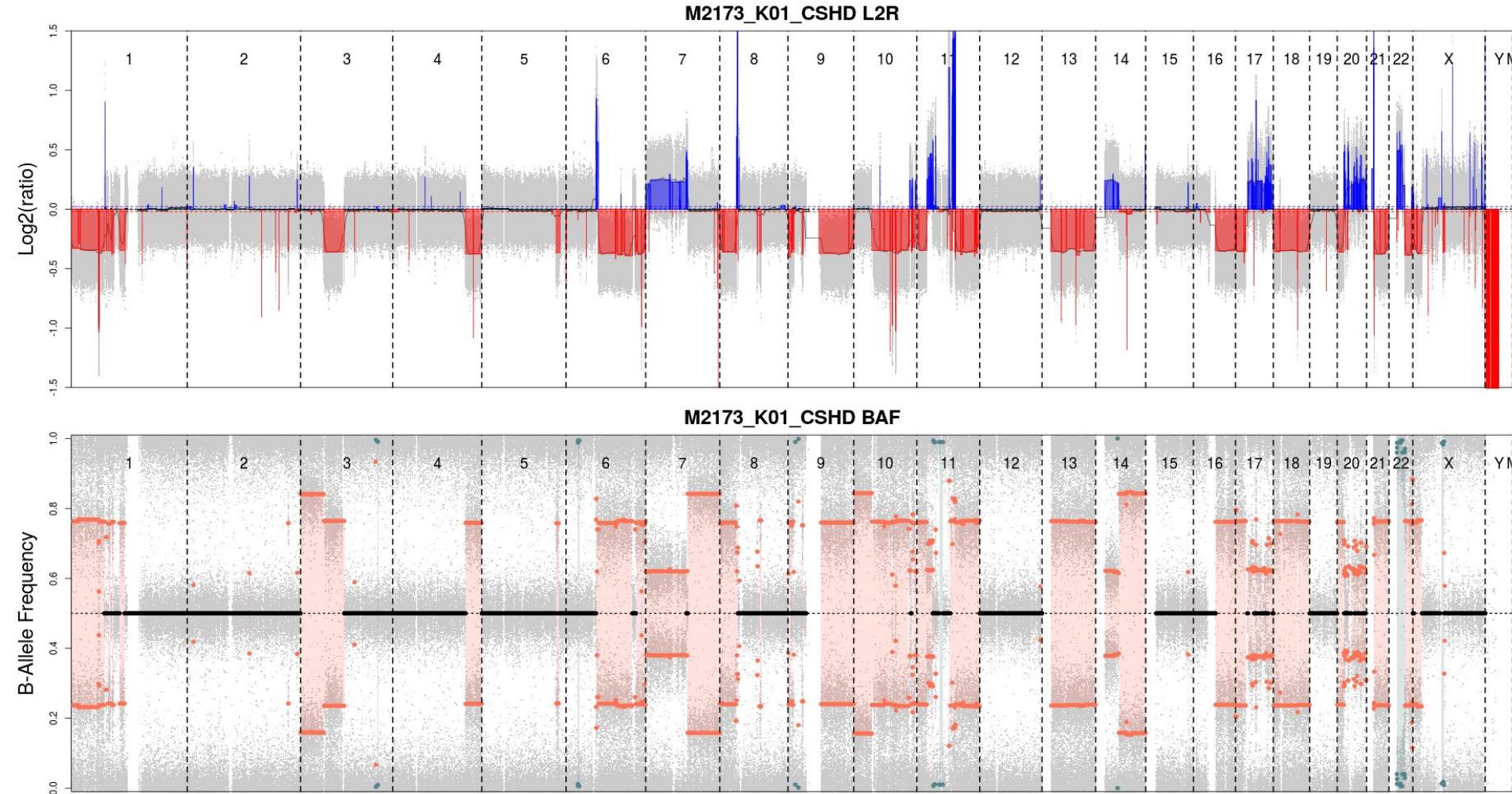


WES (Agilent SureSelect v5 capture sequenced on Illumina HiSeq)

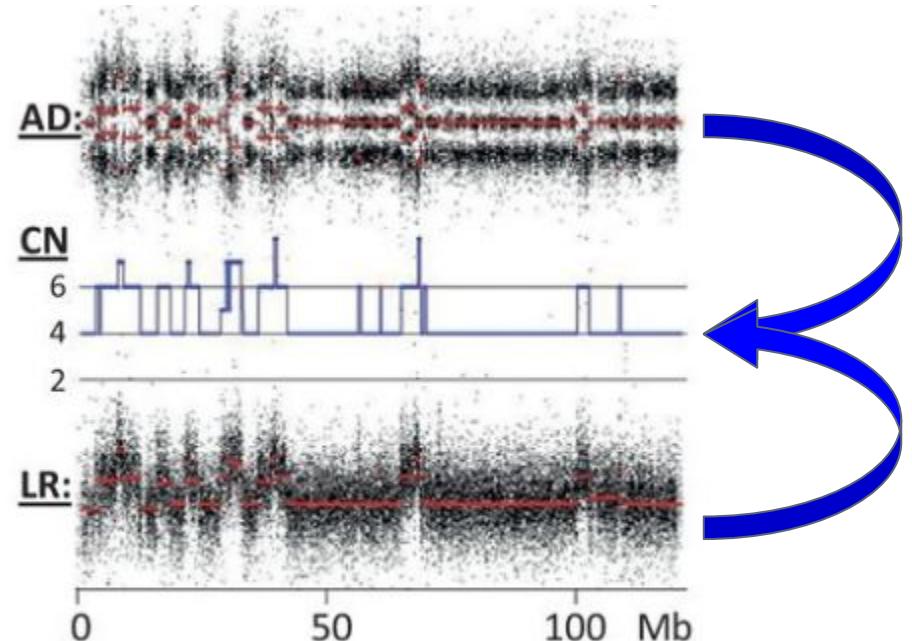
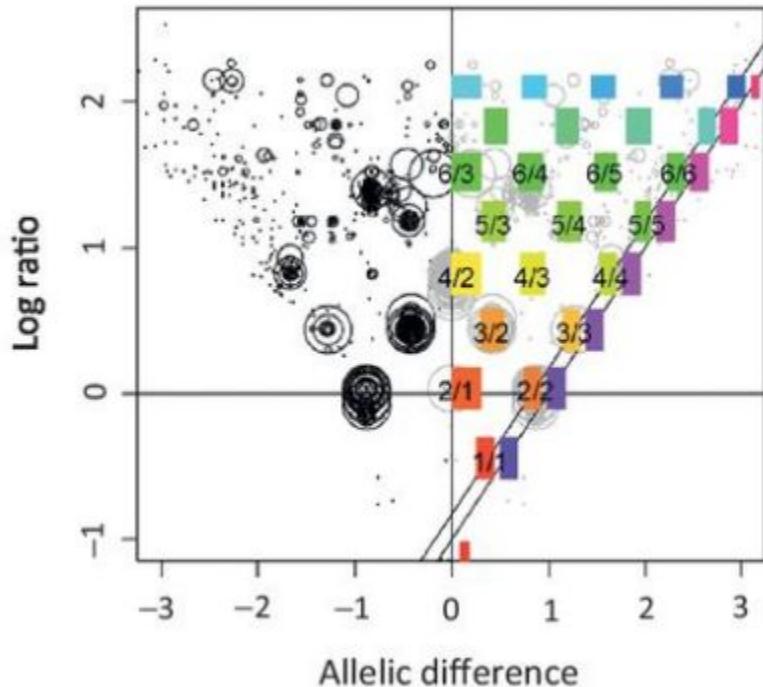
Beyond canonical CGH

From relativity to absoluteness

Up to TCN & ASCN : L₂R + BAF



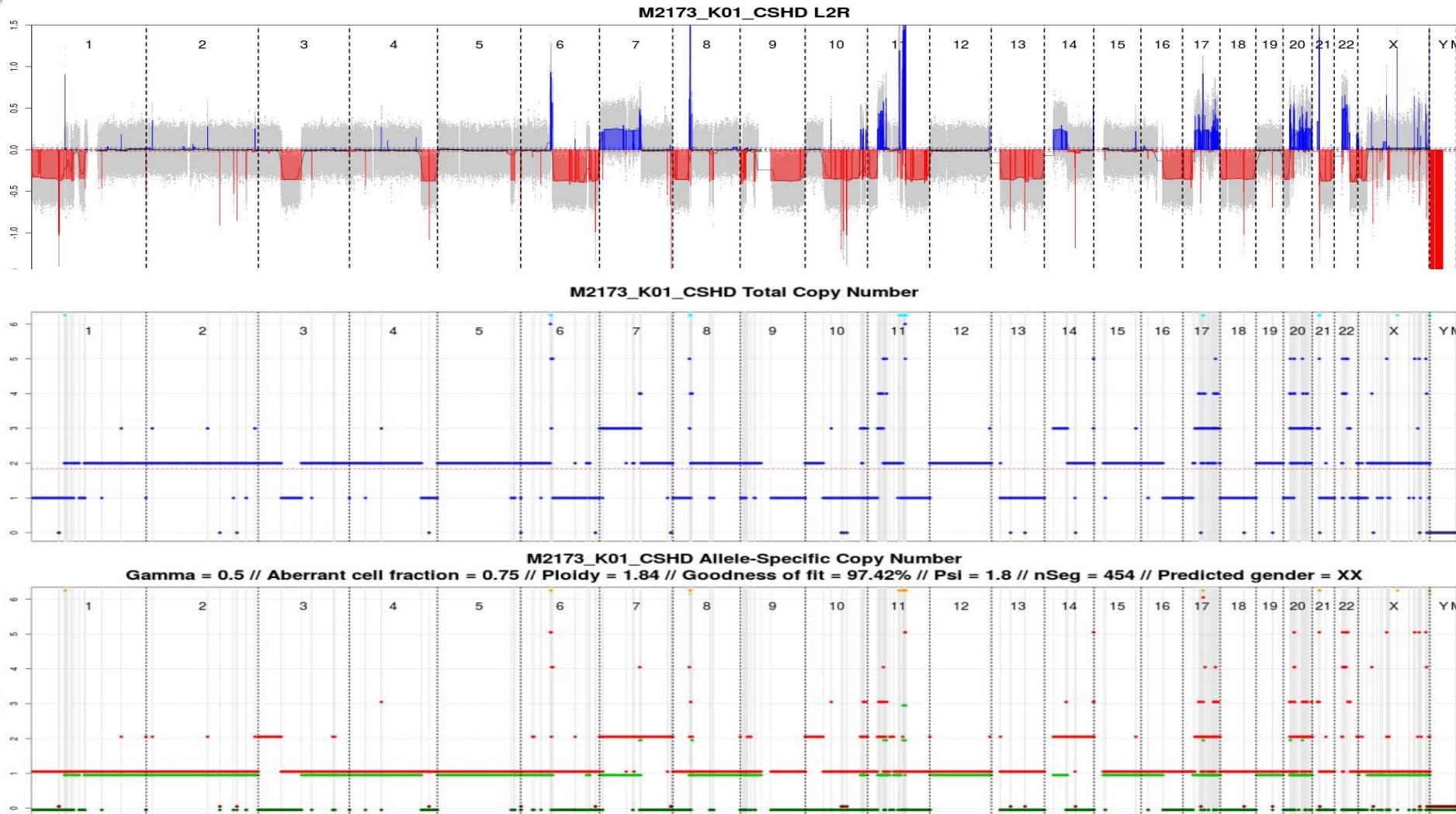
Up to TCN & ASCN : Modeling



Popova et al., Cancer Res, 2016

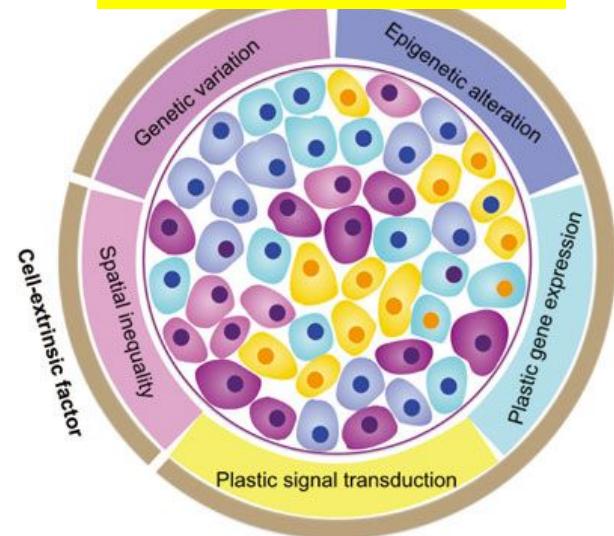
- Functional combination of L₂R & BAF (or AD)
- Also modeling **global ploidy**
- Derives putative **tumor cellularity**
- Tools : ASCAT, facets, sequenza, ControlFREEC, ...

Up to TCN & ASCN : Results



Up to TCN & ASCN : Clonality

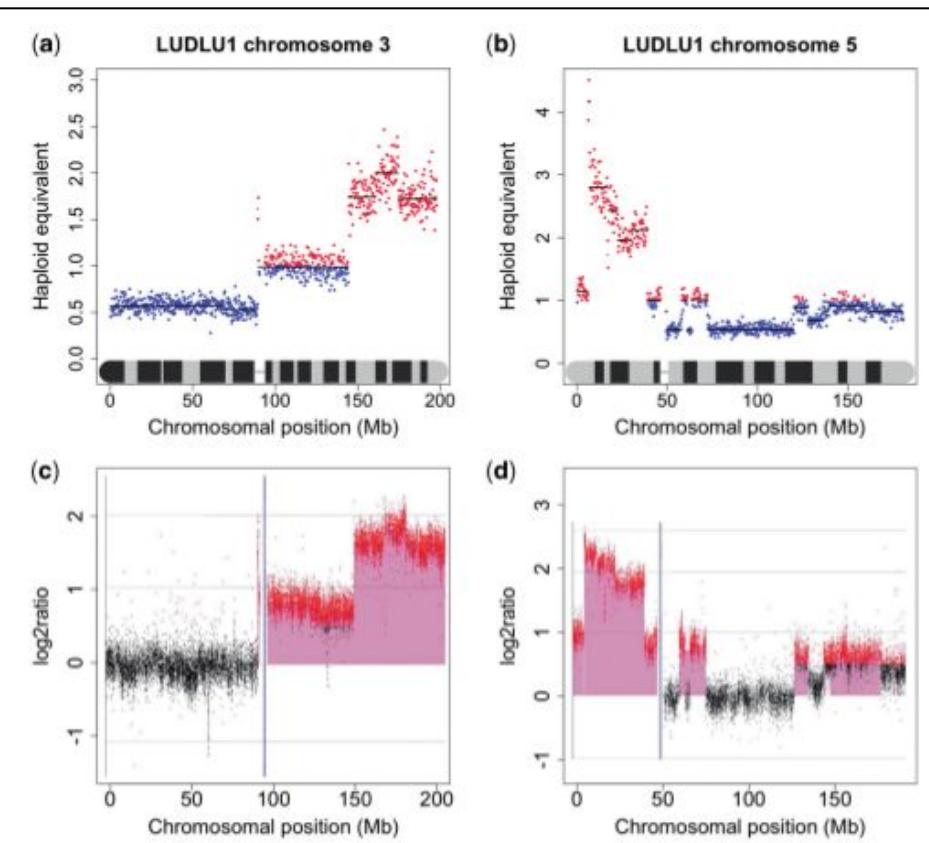
- Bulk WGS = mix of several (different ?) cells
- Current algorithm infer a **single, major** clone
- In case of polyclonality, risks are :
 - *Over*complexification (ploidy doubled, tripled, +)
 - *Under*complexification (true aberrations ignored)
- Polyclonality **deconvolution** algorithms exist (Battenberg, cloneCNA, ...) but complex (many parameters) and with low efficiency
- Long live (soon?) **single cell** ! (despite actually limited resolution)



NGS : beyond microarrays

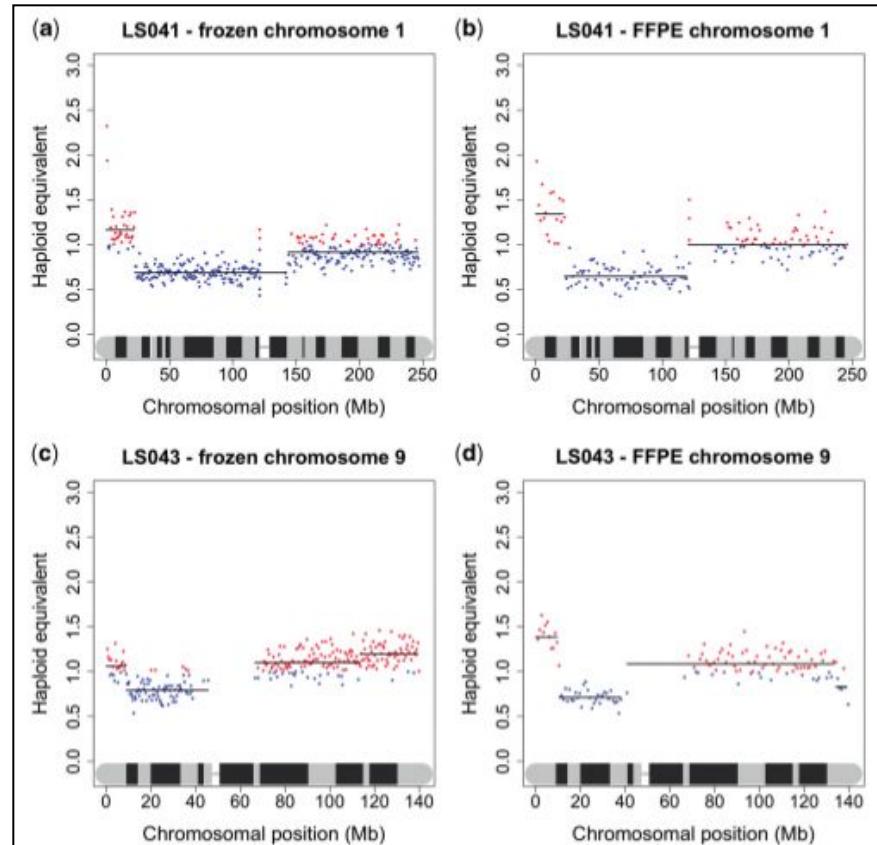
NGS : low input, FFPE

Wood, 2010



- 2 to 5 ng of DNA

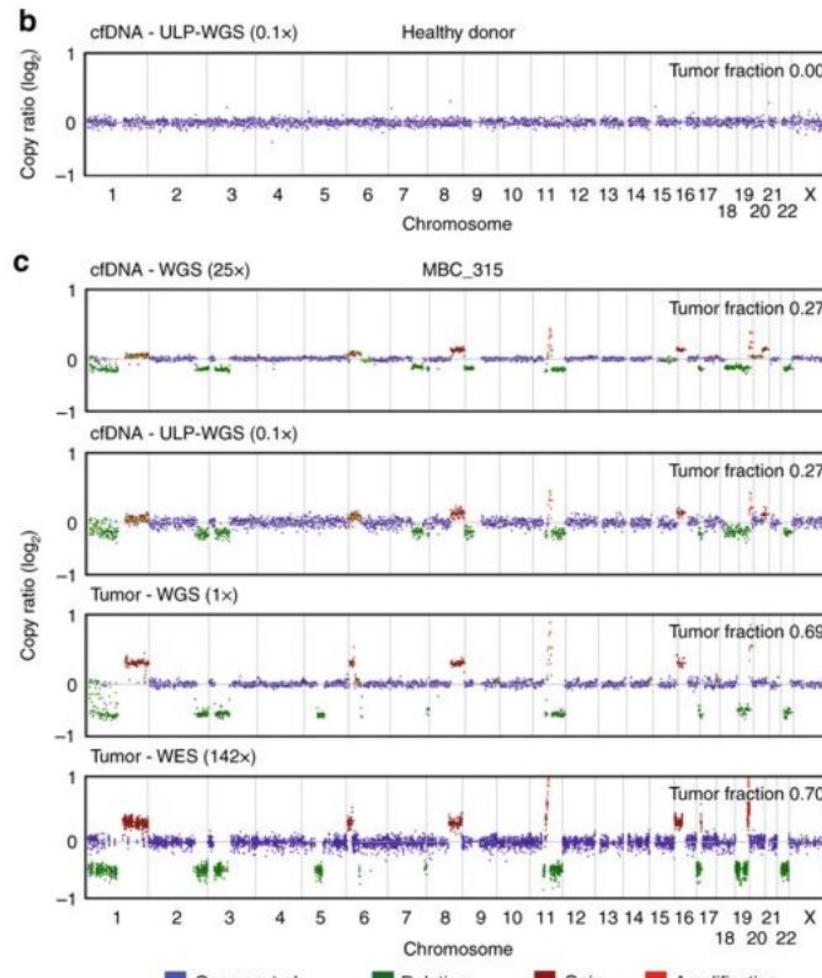
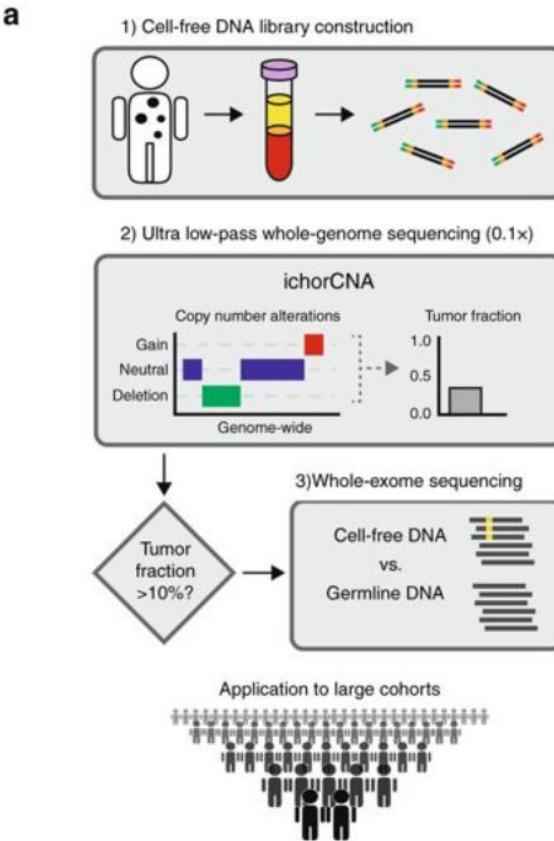
- FFPE (Formalin-fixed paraffin-embedded) samples



NGS : cell-free DNA WGS

Adalsteisson, Nature Com, 2017

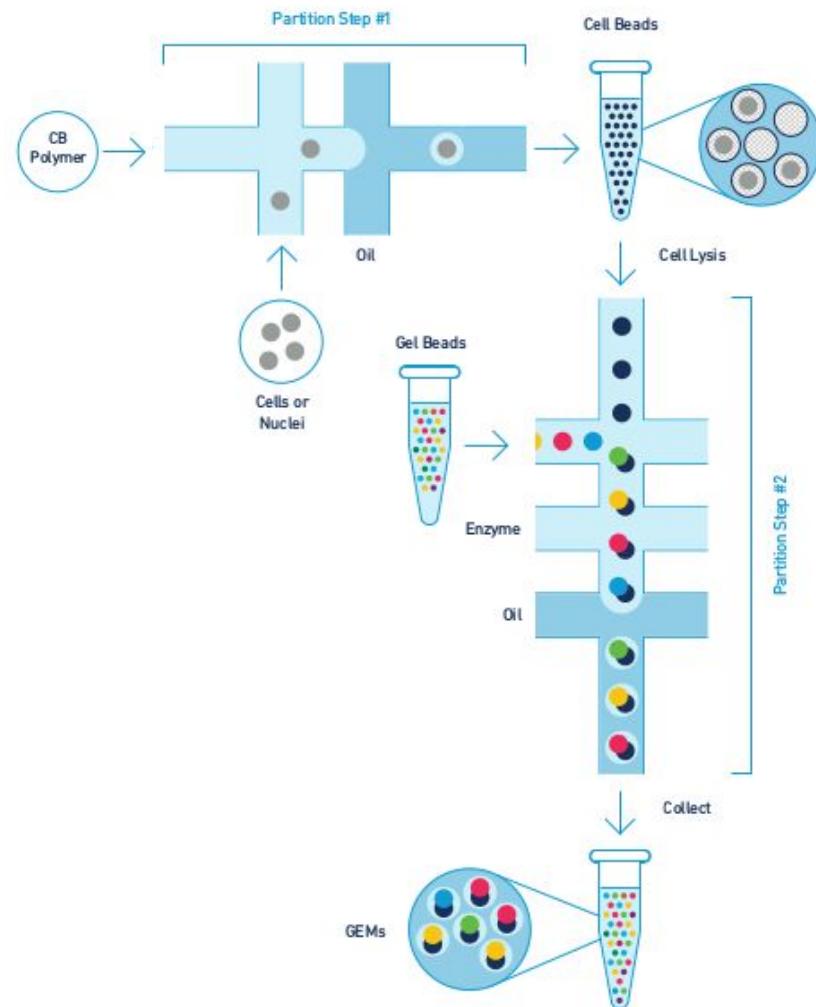
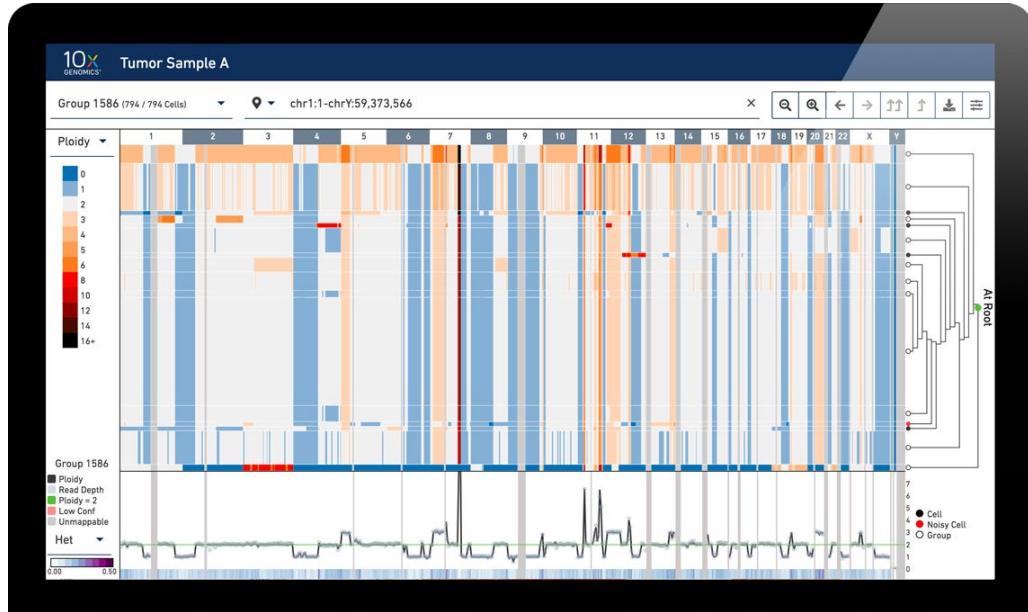
From: Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors



- Cell-free DNA
- ULP-WGS
- 0.1X coverage

NGS : Single Cell CNA

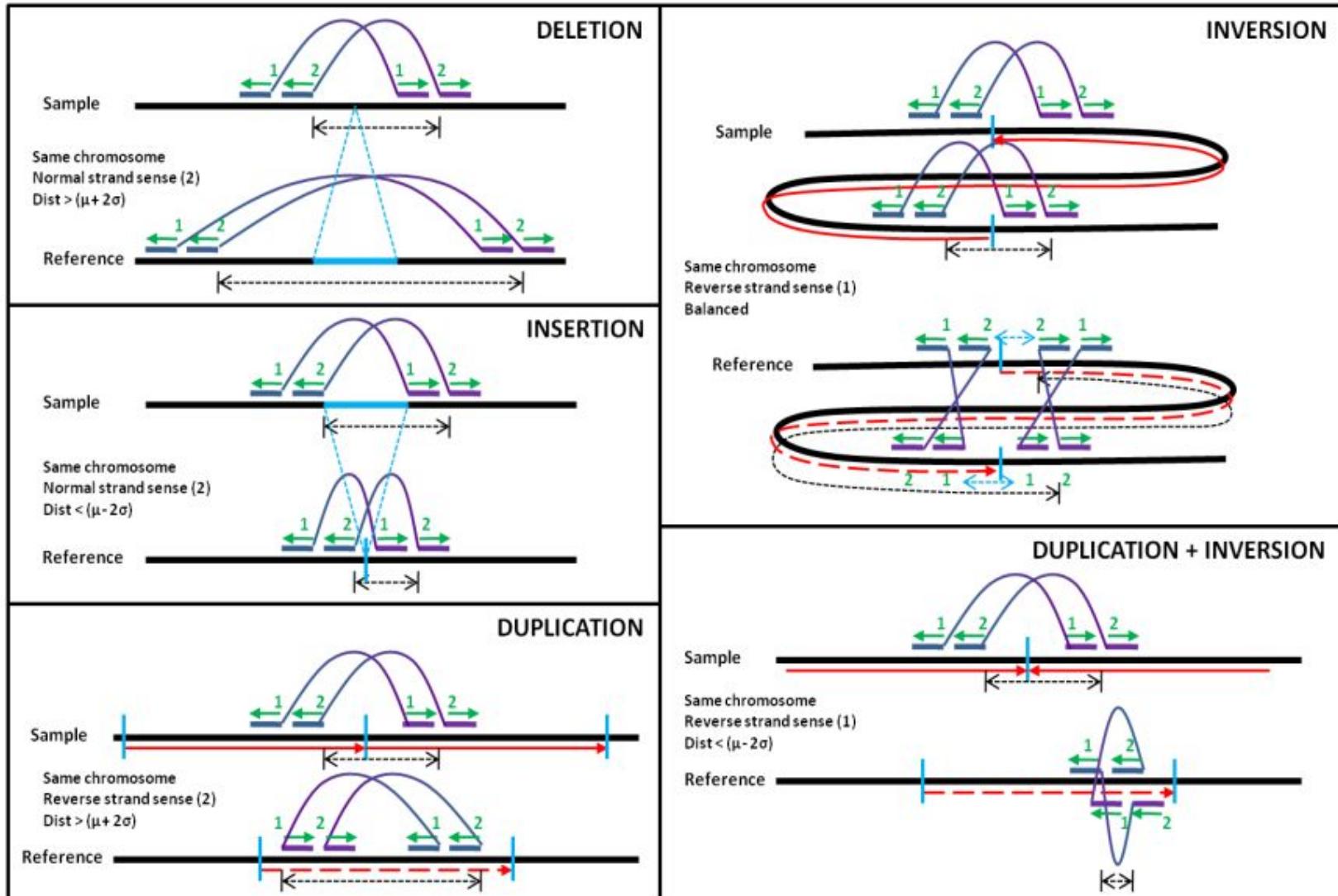
10x Technologies, 2018



- “CNA at the cell level”
- Actually, *cell patch* level
- Bioinfo enhancement needed

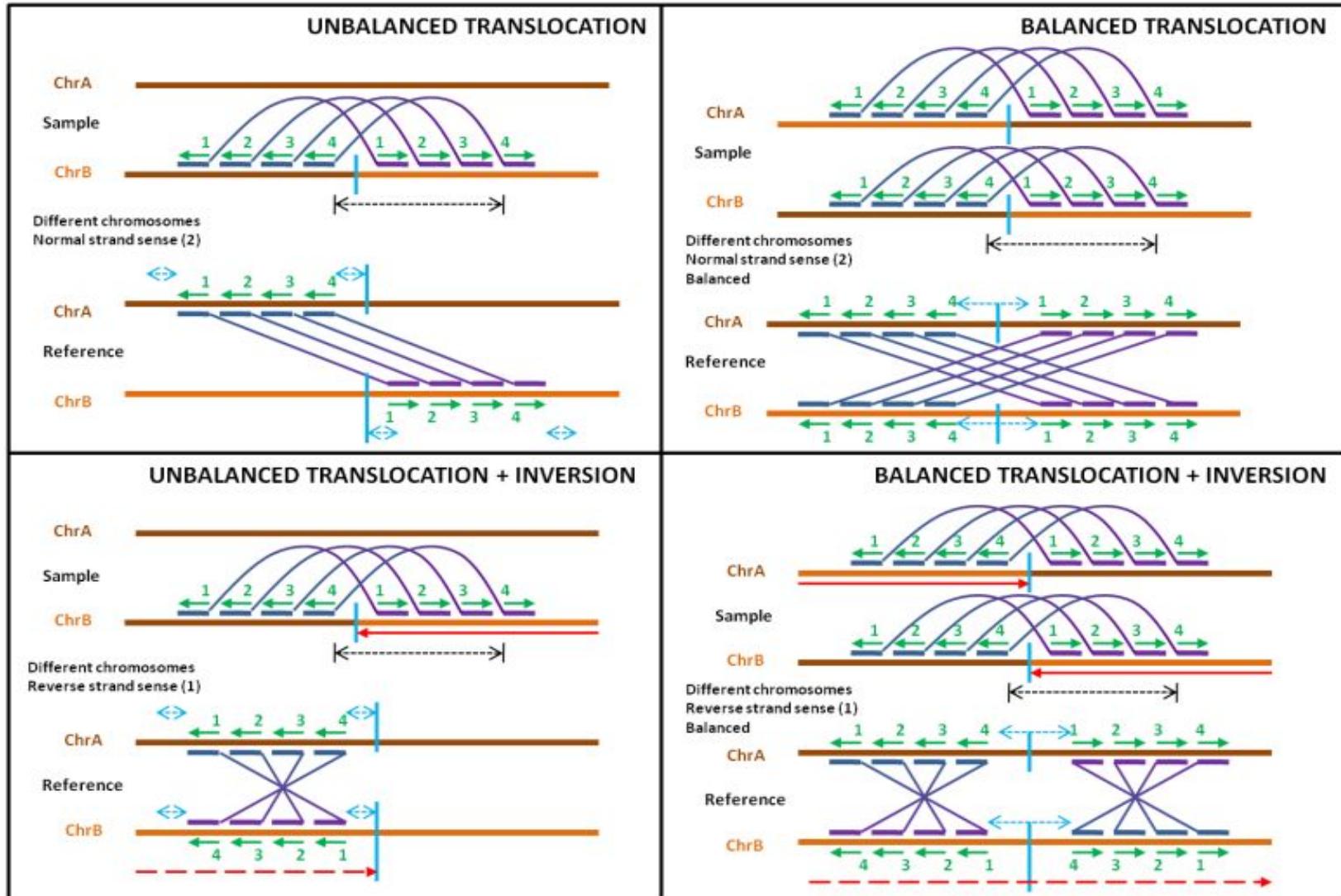
WGS : Intra-chromosomal structural variations

Courtesy of Bruno Zeitouni

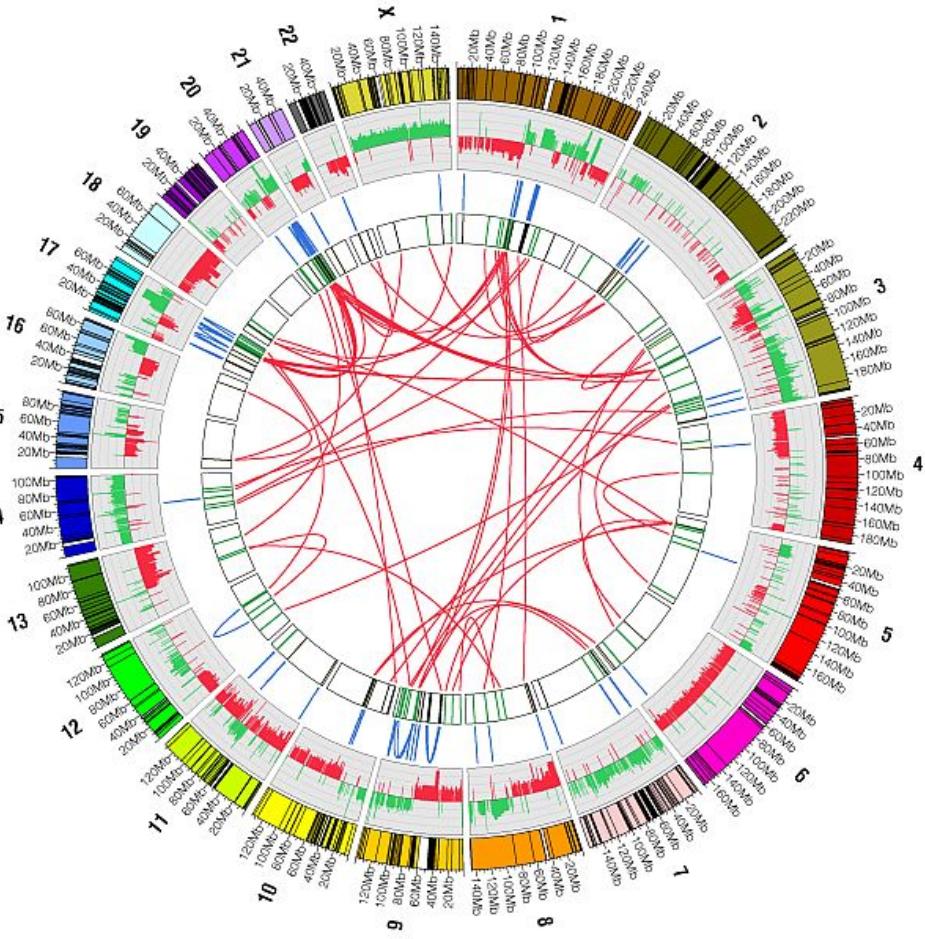
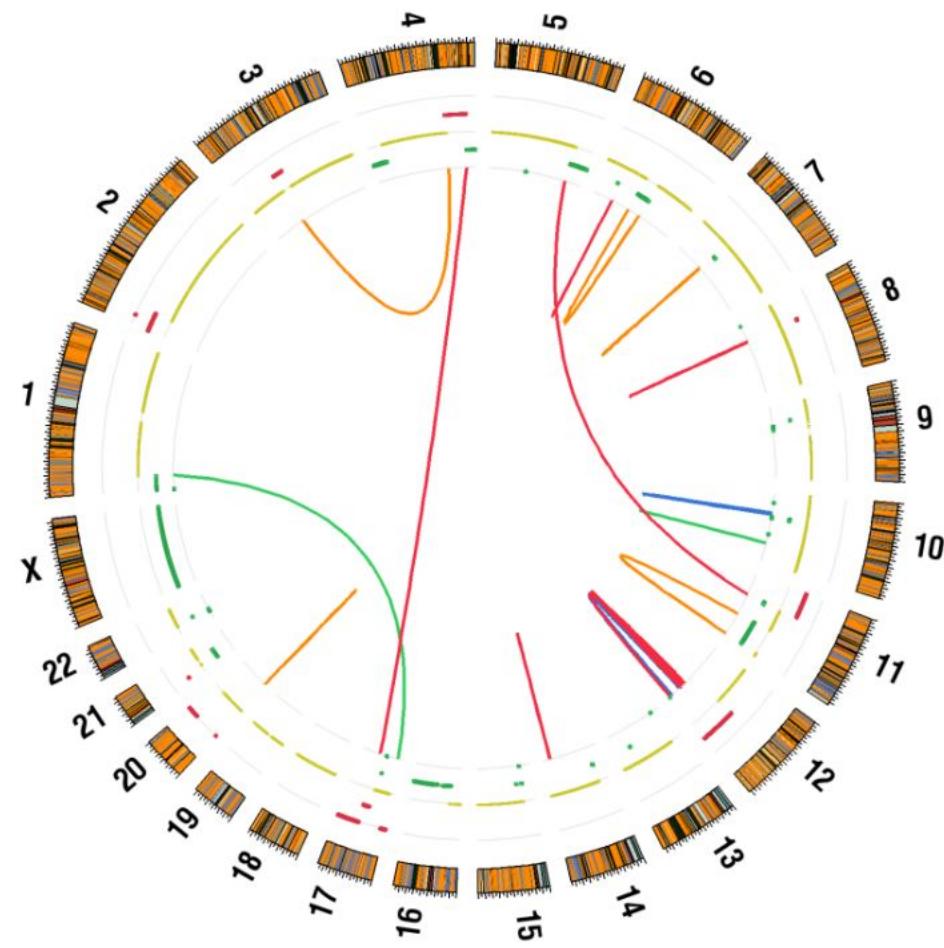


WGS : Inter-chromosomal structural variations

Courtesy of Bruno Zeitouni



Visualizing structural variations



NGS versus microarrays

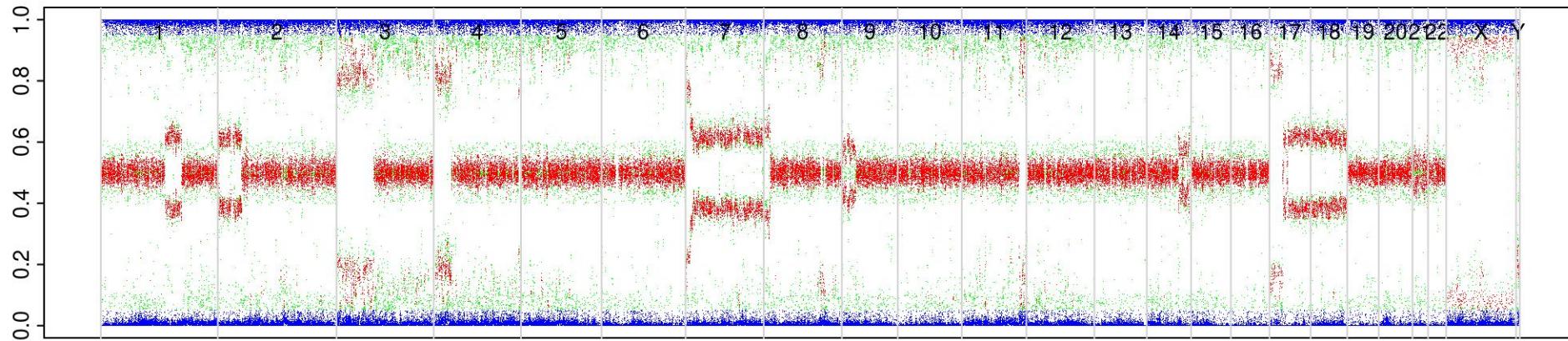
| | Microarray | NGS (WES / WGS) |
|----------------------------|---|--|
| Physical entity | Array on a glass slide | Lane in a flowcell |
| Measurement entity | Spot of probes | Cluster of fragments |
| Measurement unit | Luminous intensity per genomic position | Read depth per genomic bin |
| Data distribution | Log-normal | Negative binomial |
| Data transformation | Log ratio of intensities Test / Ref | Log ratio of depths Test / Ref |
| Bias main sources | Spatial effects, dye, GC-content | Library effects, spatial effects, coverage, GC-content, mappability |
| CNV information | Normality, gains and losses relative to the reference | Normality, gains and losses relative to the reference, absolute and allele-specific copy number levels |
| CNV event precision | Up to ~3 Kb | ~50 b |
| Structural information | Large-scale deletions | Insertions, deletions, inversions, balanced translocations |
| SNP information | Known SNPs (if specific probes) | All kinds of SNPs, position and allele frequency |
| SNV (mutation) information | No / some* | All SNVs |
| Sequence information | No | Full covered sequence |

Microarrays are still alive

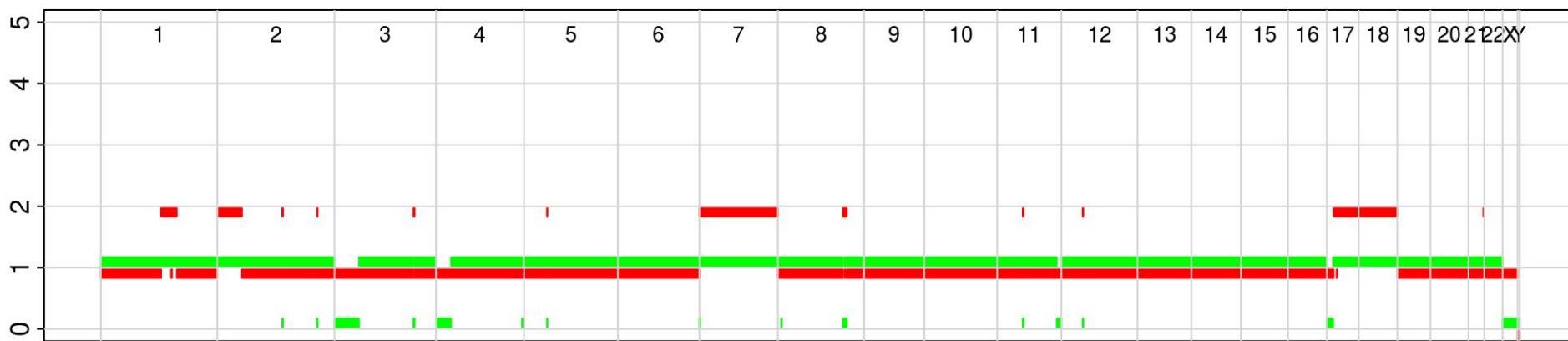
A

Works on FFPE samples !

M1084_PED 58370 129246



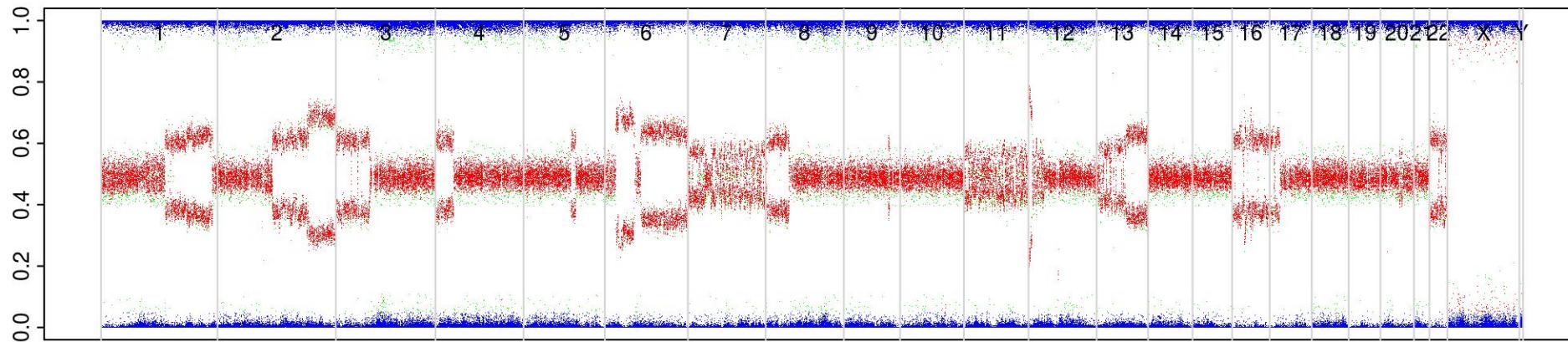
Ploidy: 2.17, aberrant cell fraction: 82%, goodness of fit: 94.7%



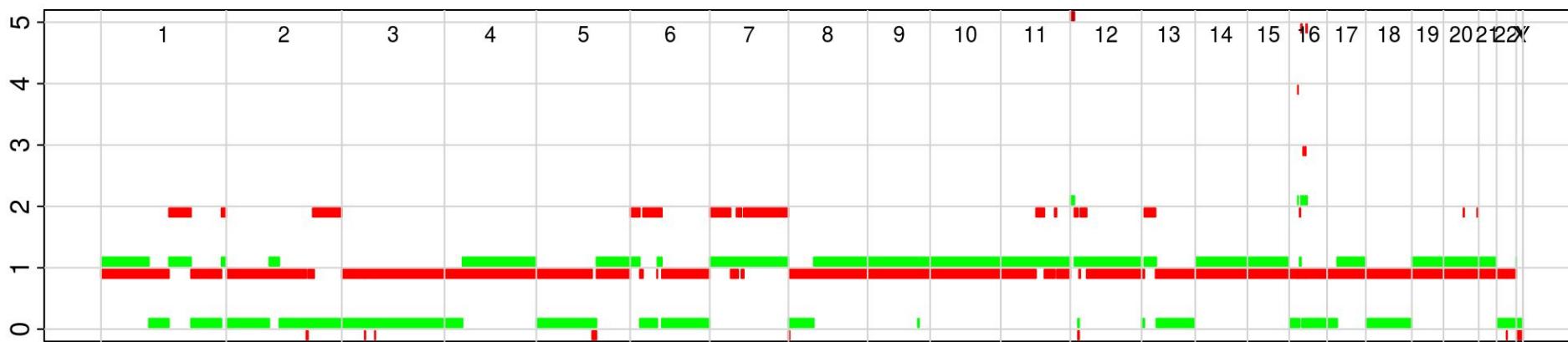
A

Works on ctDNA !

M782_circ 54608 156269



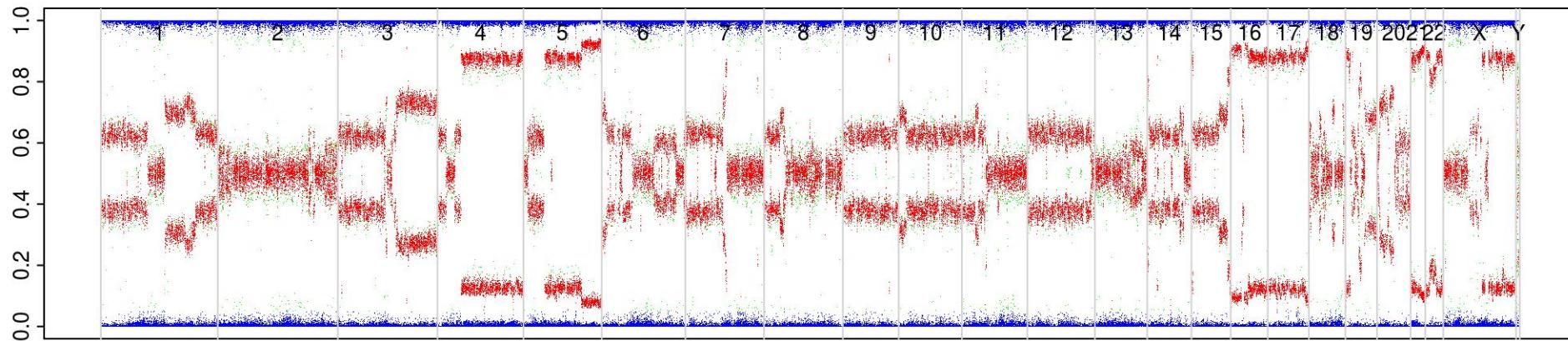
Ploidy: 1.61, aberrant cell fraction: 47%, goodness of fit: 89.8%



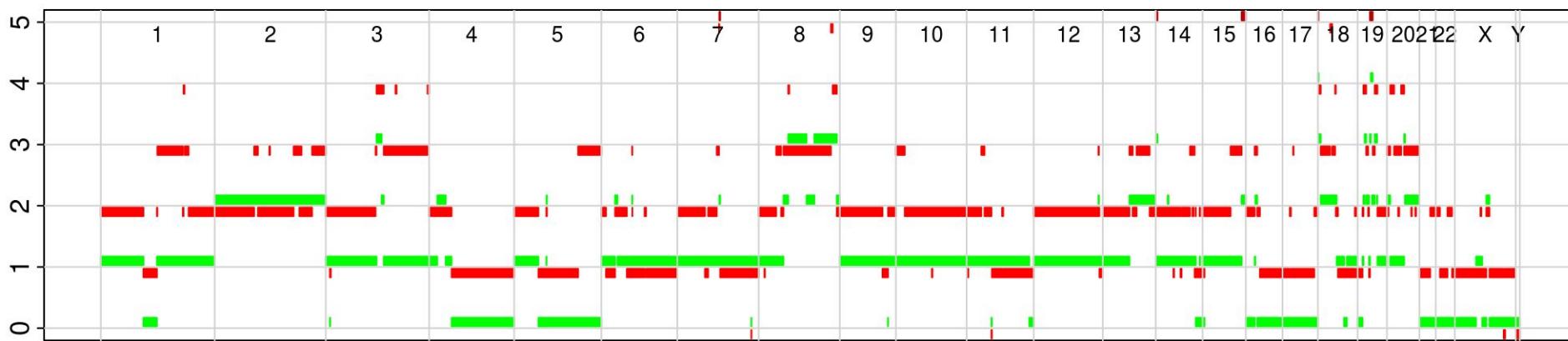
A

Works on ascites DNA !

A26 67028 140247

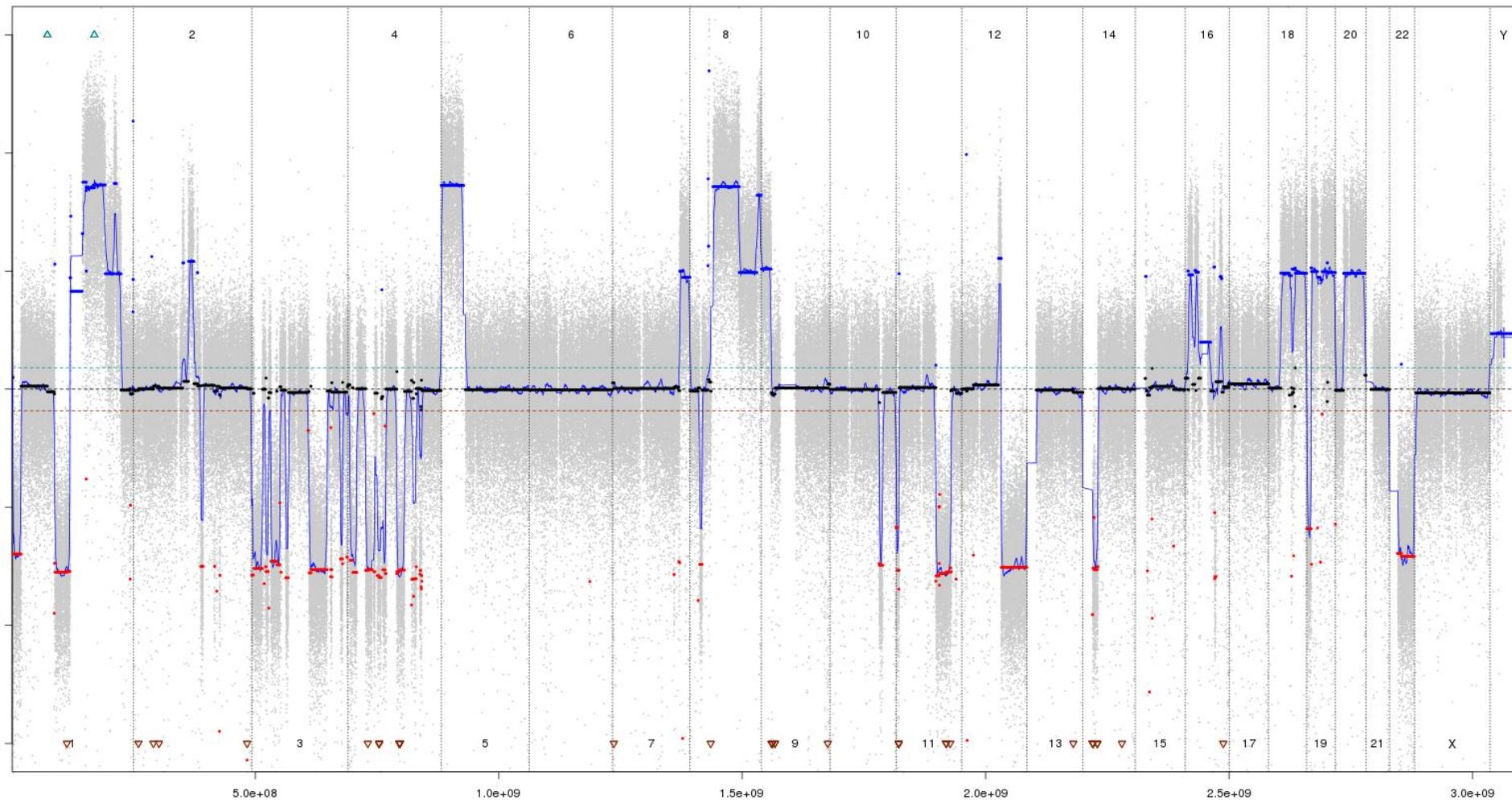


Ploidy: 2.98, aberrant cell fraction: 86%, goodness of fit: 93.5%

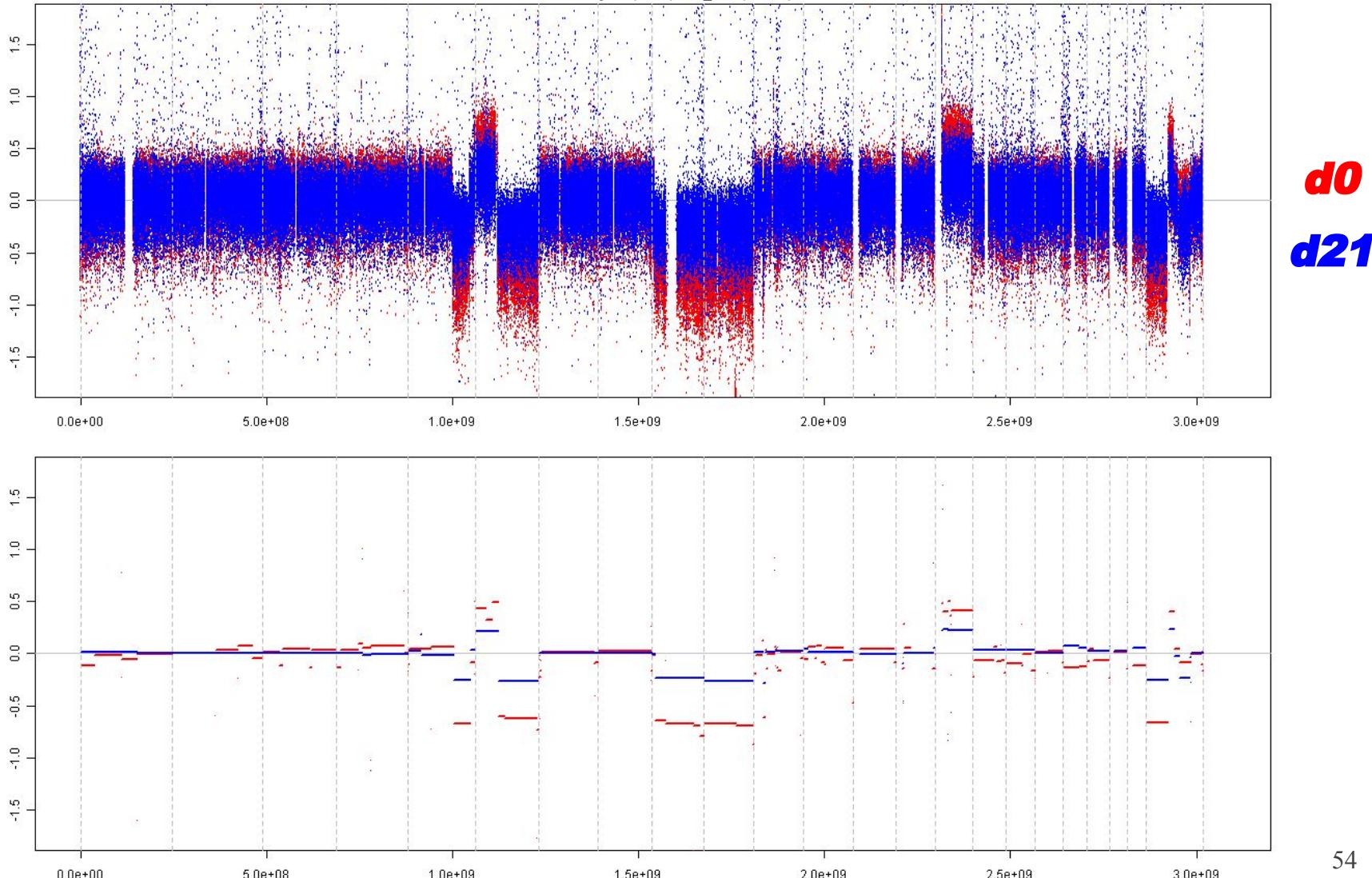


Further analyses

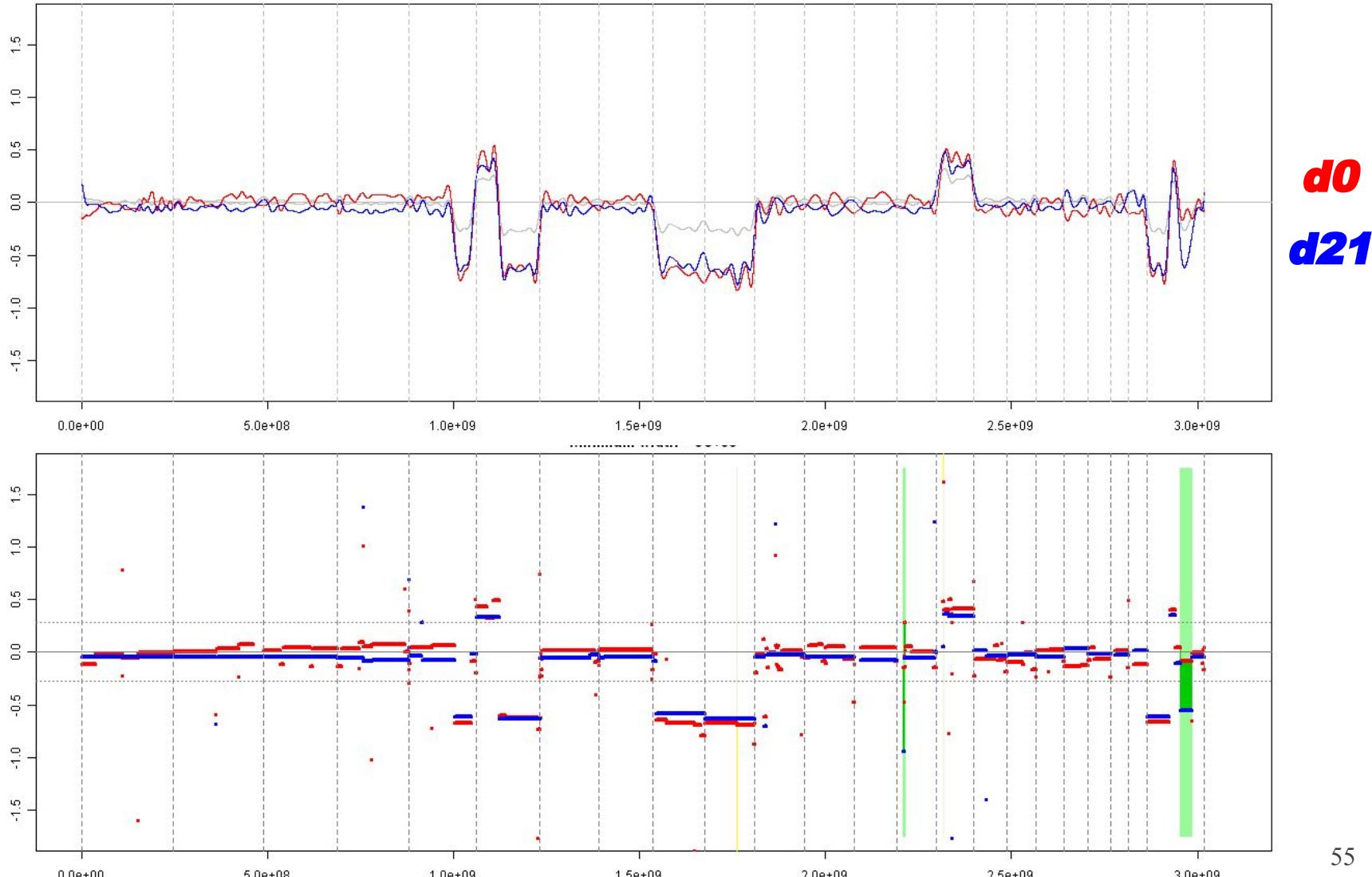
Sample profile



Comparison of pairs (scaling)



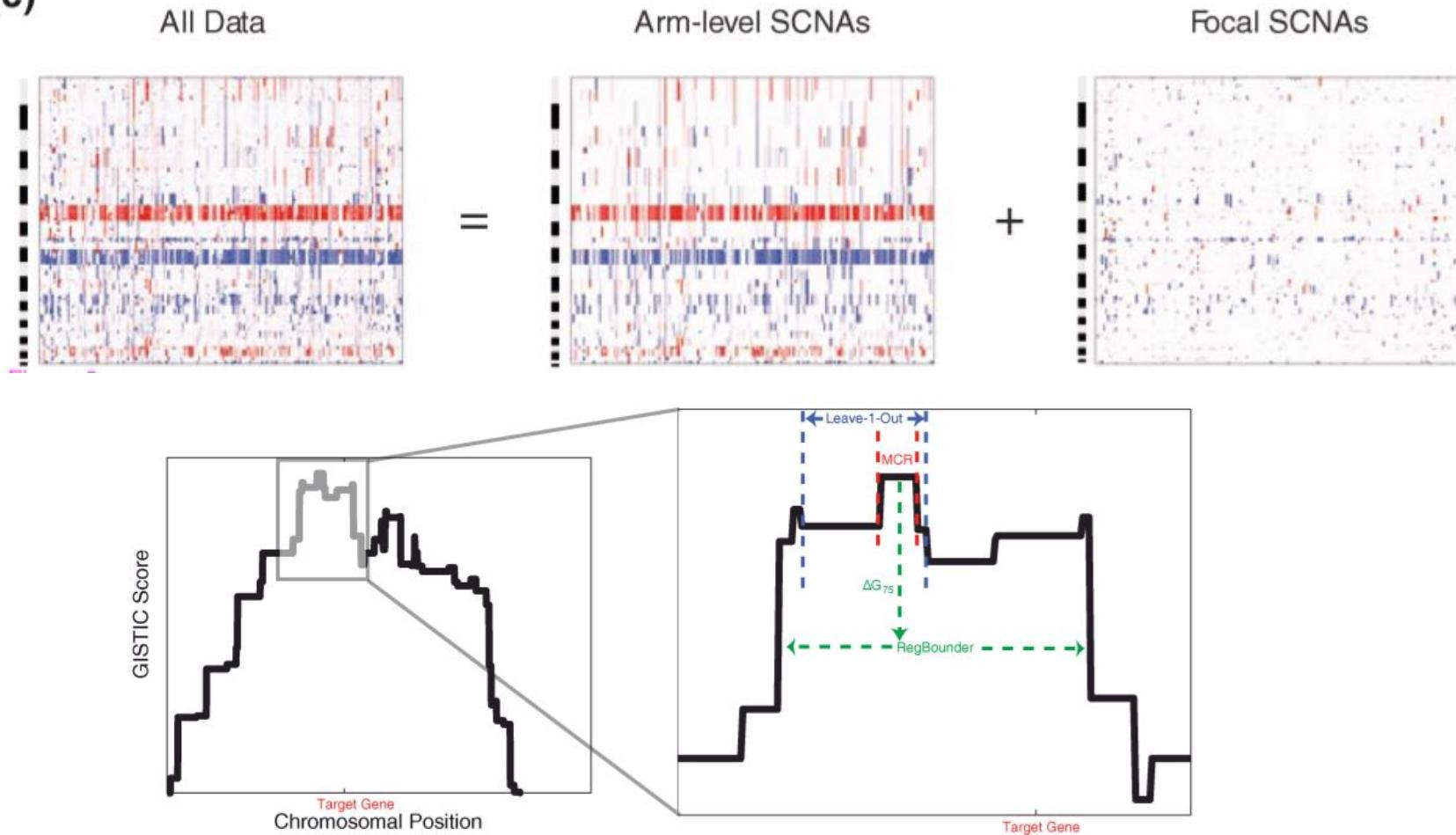
Comparison of pairs (calling)



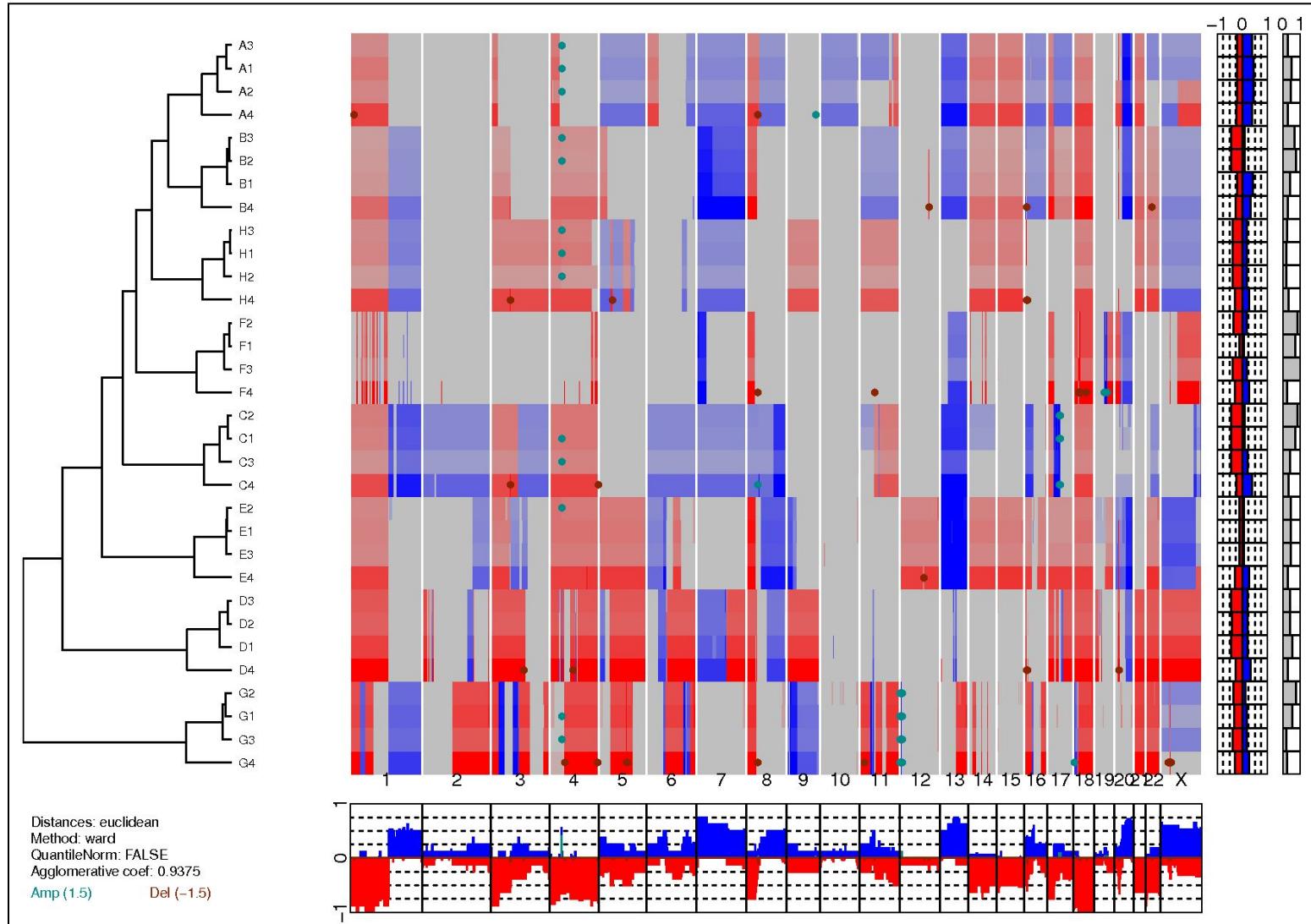
Minimal common regions (GISTIC₂)

Mermel et al, *Genome Biology*, 2011

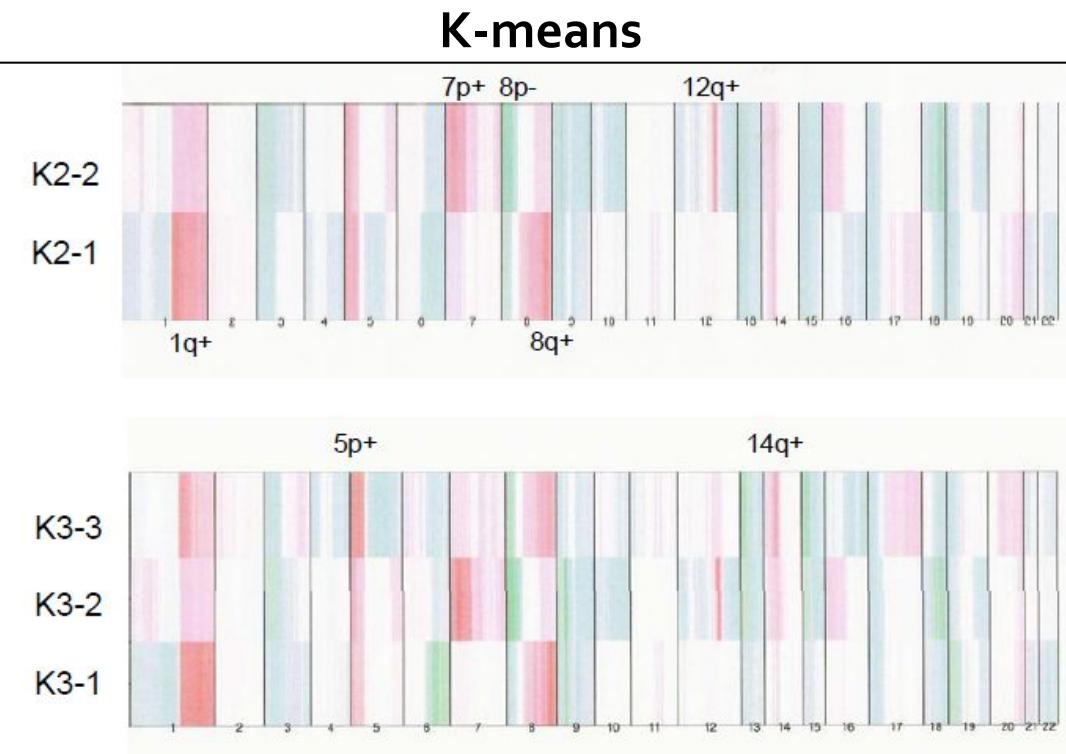
(c)



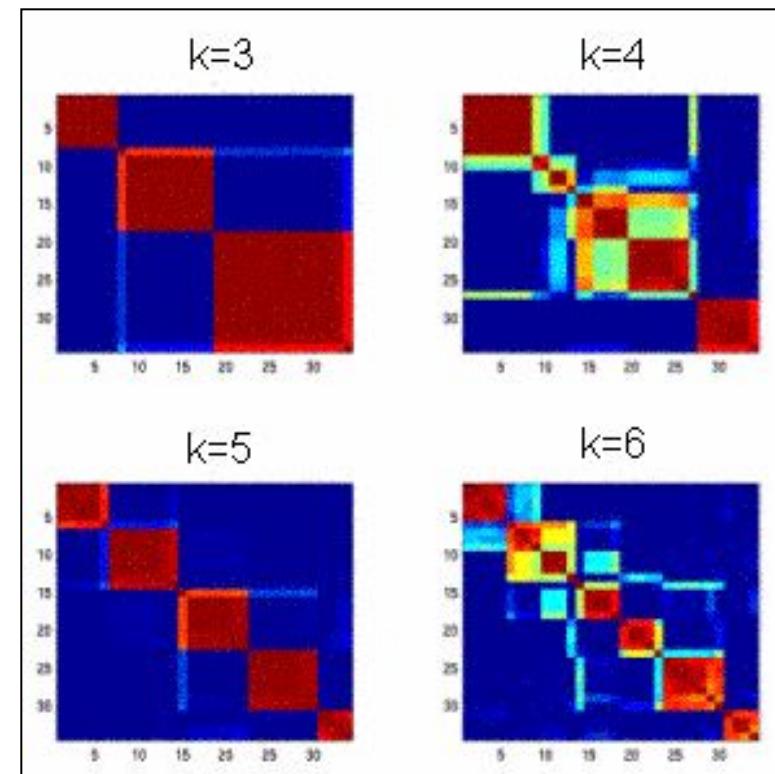
Hierarchical clustering, heatmap frequency of aberrations, genomic instability



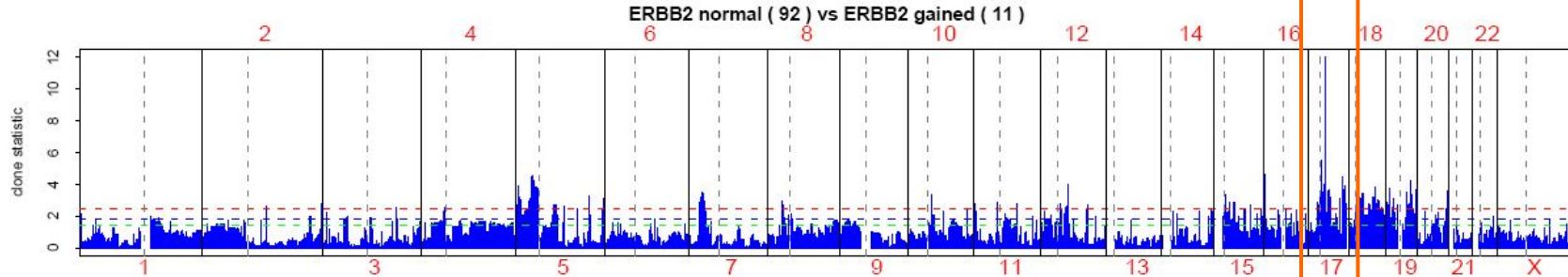
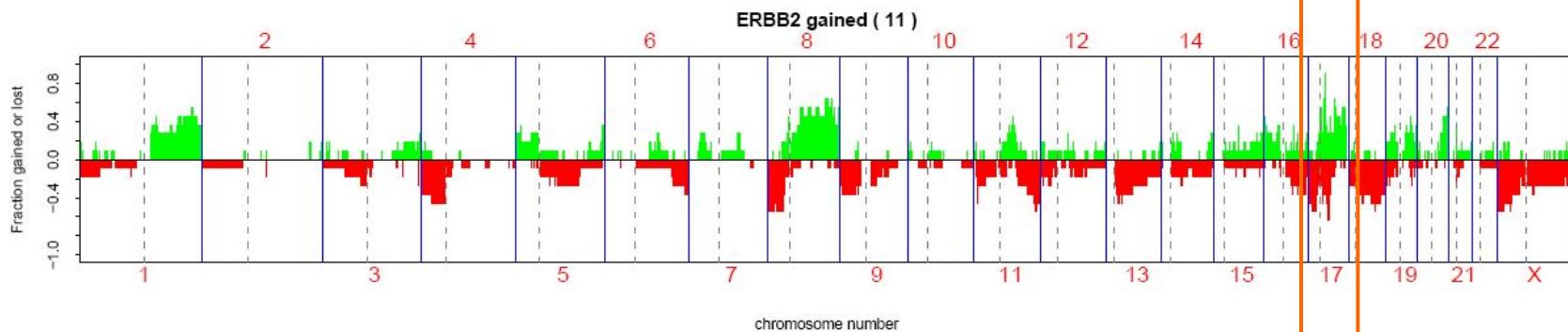
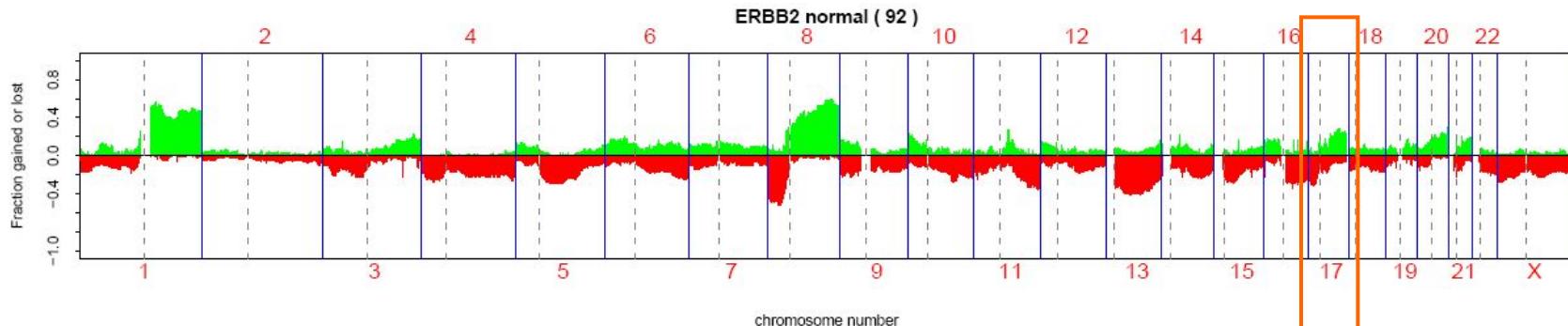
Other clustering methods



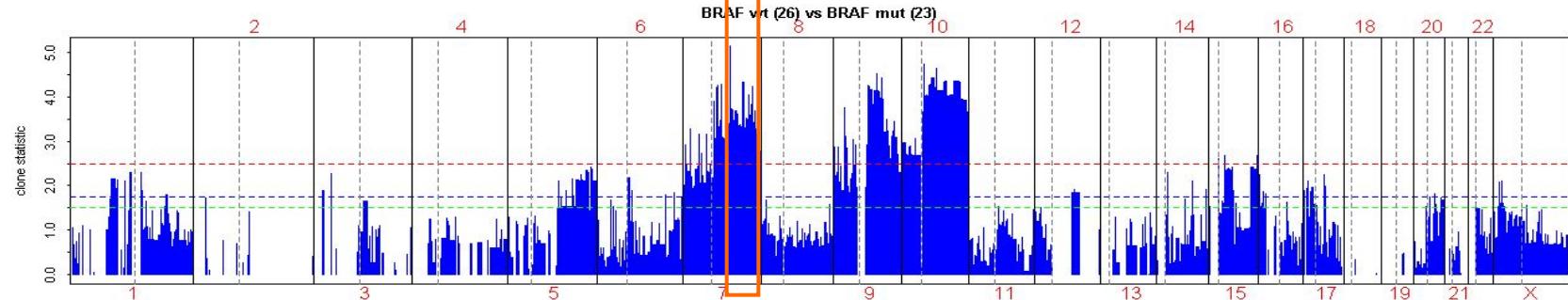
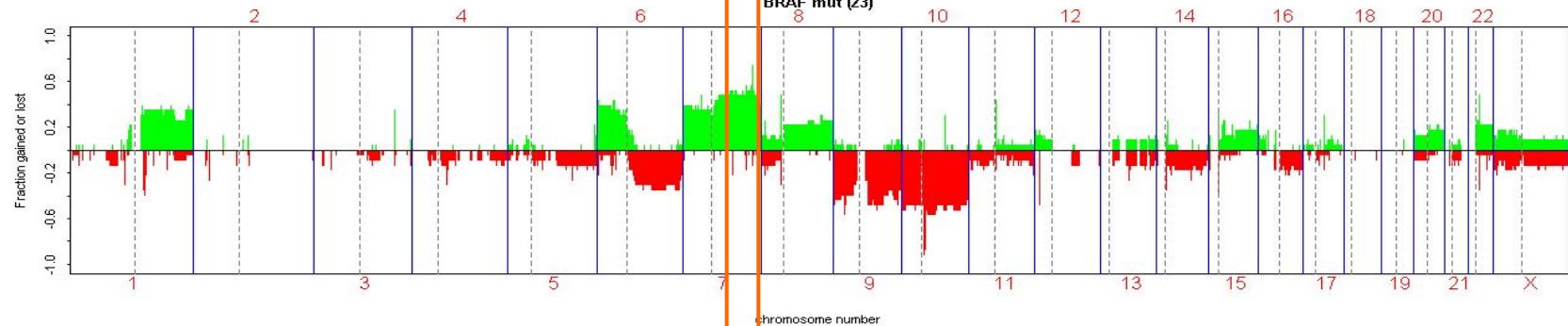
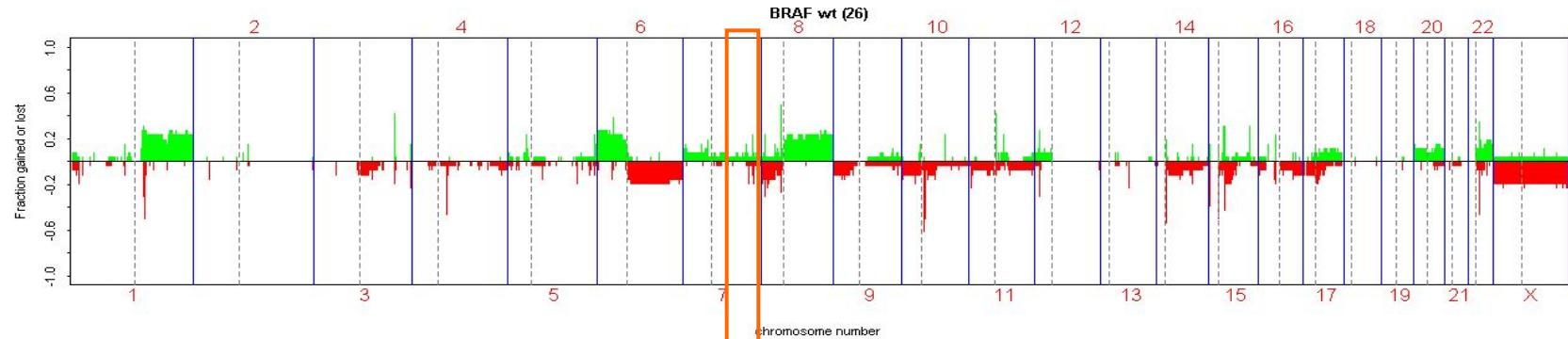
Non-negative Matrix Factorization



Subpopulations and annotations (Continuous data)



Subpopulations and annotations (Continuous data)



Genomic regions annotations

| | Loc | Width | Band1 | Band2 | Num.probes | Status | log2(ratio) | Ratio | Genes | CNV | miRNA | CpGisI |
|----|--------------------------|----------|------------|------------|------------|--------|---|-------|---|------|-------|--------|
| 1 | 8:161471-6914076 | 6.75Mb | 8p23.3 | 8p23.1 | 409 | L | -1.13 | 0.46 | 2 12 32 | 1949 | 15 | 120 |
| 2 | 8:6939250-7786708 | 847.46Kb | 8p23.1 | 8p23.1 | 10 | L | -0.34 | 0.79 | 3 23 | 135 | - | 11 |
| 3 | 8:8100383-22878739 | 14.78Mb | 8p23.1 | 8p21.3 | 879 | L | -1.09 | 0.47 | 1 15 39 109 | 1560 | 28 | 120 |
| 4 | 8:22888308-24757290 | 1.87Mb | 8p21.3 | 8p21.2 | 102 | L | -0.45 | 0.73 | 3 7 15 | 105 | - | 16 |
| 5 | 8:24773594-26994749 | 2.22Mb | 8p21.2 | 8p21.2 | 160 | L | -1.09 | 0.47 | 1 9 12 | 190 | 9 | 17 |
| 6 | 8:27015529-27667961 | 652.43Kb | 8p21.2 | 8p21.1 | 53 | L | -0.41 | 0.75 | 2 5 12 | 34 | 11 | 10 |
| 7 | 8:27678088-33627376 | 5.95Mb | 8p21.1 | 8p12 | 369 | L | -1.07 | 0.48 | 2 1 16 37 | 284 | 6 | 34 |
| 8 | 8:33665709-34086359 | 420.65Kb | 8p12 | 8p12 | 16 | G | 0.58 | 1.49 | - | 11 | - | - |
| 9 | 8:34129287-34595586 | 466.30Kb | 8p12 | 8p12 | 17 | G | 1.38 | 2.61 | - | 35 | - | - |
| 10 | 8:34615562-35126922 | 511.36Kb | 8p12 | 8p12 | 22 | G | 2.23 | 4.70 | 1 1 | 27 | - | 2 |
| 11 | 8:35137186-37228379 | 2.09Mb | 8p12 | 8p11.23 | 94 | L | -1.07 | 0.47 | 2 2 | 75 | - | - |
| 12 | 8:37281736-38008581 | 726.85Kb | 8p11.23 | 8p11.23 | 49 | G | 2.13 | 4.38 | 9 11 | 35 | - | 11 |
| 13 | 8:38021058-39195522 | 1.17Mb | 8p11.23 | 8p11.22 | 91 | G | 2.52 | 5.72 | 1 1 1 7 16 | 56 | - | 14 |
| 14 | | | | | | | | | | | | |
| 15 | | | | | | | | | | | | |
| 16 | | | | | | | | | | | | |
| 17 | | | | | | | | | | | | |
| 18 | | | | | | | | | | | | |
| | Gene | Chr | Start | End | Width | | Description | | Pathways | | CTD | CNV |
| 19 | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | |
| 21 | LSM1 | 8 | 38,020,838 | 38,034,248 | 13.41Kb | | LSM1, U6 small nuclear RNA associated | | Gene Expression Metabolism of RNA RNA degradation | | 6 | 1 |
| 22 | | | | | | | | | | | | |
| X | BAG4 | 8 | 38,034,105 | 38,070,819 | 36.72Kb | | BCL2-associated athanogene 4 | | | | 11 | 1 |
| Y | DDHD2 | 8 | 38,089,008 | 38,120,287 | 31.28Kb | | DDHD domain containing 2 | | | | 4 | 2 |
| T | PPAPDC1B | 8 | 38,120,649 | 38,126,738 | 6.09Kb | | phosphatidic acid phosphatase type 2 domain containing 1B | | Immune System | | 14 | - |
| B | WHSC1L1 | 8 | 38,132,560 | 38,239,790 | 107.23Kb | | Wolf-Hirschhorn syndrome candidate 1-like 1 | | Lysine degradation | | 9 | 5 |
| | LETM2 | 8 | 38,243,958 | 38,266,062 | 22.11Kb | | leucine zipper-EF-hand containing transmembrane protein 2 | | | | 7 | - |
| | FGFR1 | 8 | 38,268,655 | 38,326,352 | 57.70Kb | | fibroblast growth factor receptor 1 | | Adherens junction Developmental Biology Disease Immune System MAPK signalling pathway Melanoma Pathways in cancer Prostate cancer Regulation of actin cytoskeleton Signal Transduction | | 51 | - |

Genomic regions annotations

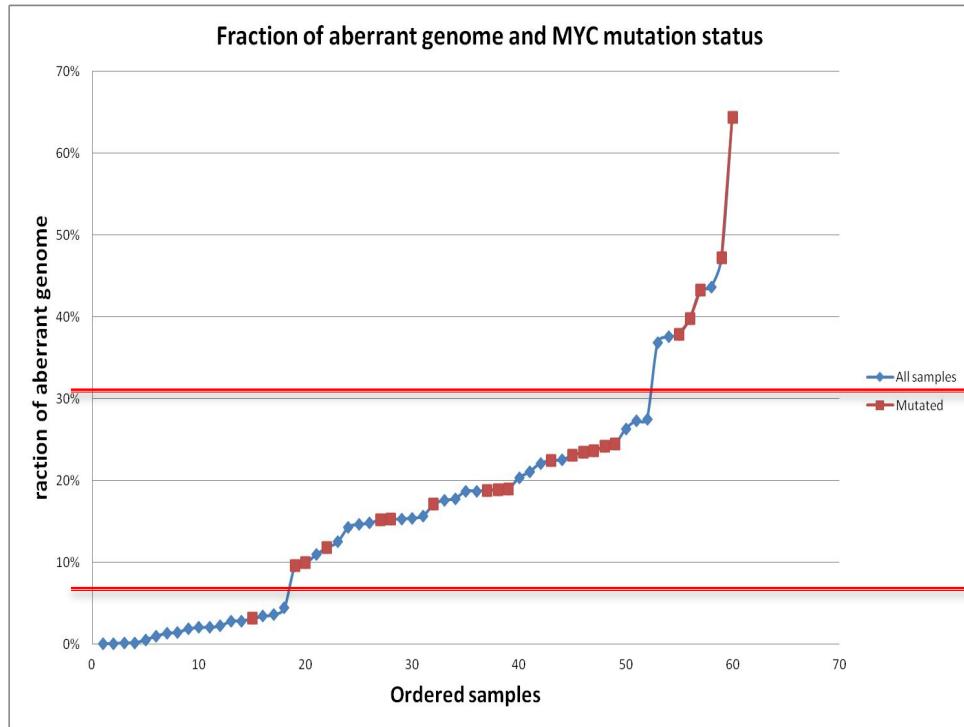
| | Loc | Width | Band1 | Band2 | Num.probes | Status | log2(ratio) | Ratio | Genes | CNV | miRNA | CpGis! | |
|------|-------------------------|----------|-----------|-----------|---|---|-------------|-------|---|---|-------|--------|----|
| 2 | 5:26141-182627 | 156.49Kb | 5p15.33 | 5p15.33 | 10 | G | 1.51 | 2.85 | 1 1 | 58 | - | 7 | |
| 3 | 5:195334-440625 | 245.29Kb | 5p15.33 | 5p15.33 | 21 | L | -0.61 | 0.65 | 2 2 5 | 88 | - | 16 | |
| 4 | 5:449198-514289 | 65.09Kb | 5p15.33 | 5p15.33 | 6 | G | 1.10 | 2.14 | 2 2 | 79 | - | 14 | |
| 5 | 5:527551-667003 | 139.45Kb | 5p15.33 | 5p15.33 | 12 | L | -2.54 | 0.17 | 1 3 | 127 | 2 | 11 | |
| 6 | 5:684828-1188561 | 503.73Kb | 5p15.33 | 5p15.33 | 31 | L | -0.38 | 0.77 | 5 7 | 286 | 2 | 41 | |
| 7 | 5:1207958-2991547 | 1.78Mb | 5p15.33 | 5p15.33 | 107 | G | 1.41 | 2.66 | 1 8 14 | 433 | 6 | 66 | |
| Gene | Chr | Start | End | Width | Description | | | | Pathways | | CTD | CNV | |
| | | | | | solute carrier family 6 (neutral amino acid transporter), member 19 | | | | Amine compound SLC transporters Amino acid and oligopeptide SLC transporters Mineral absorption Protein digestion and absorption Transmembrane transport of small molecules | | | 23 | |
| | SLC6A19 | 5 | 1,201,709 | 1,225,230 | 23.52Kb | solute carrier family 6 (neutral amino acid transporter), member 19 | | | | Amine compound SLC transporters Amino acid and oligopeptide SLC transporters Mineral absorption Protein digestion and absorption Transmembrane transport of small molecules | | | |
| | SLC6A18 | 5 | 1,225,469 | 1,246,304 | 20.84Kb | solute carrier family 6 (neutral amino acid transporter), member 18 | | | | Amine compound SLC transporters Amino acid and oligopeptide SLC transporters Transmembrane transport of small molecules | | 5 | 20 |
| | TERT | 5 | 1,253,286 | 1,295,162 | 41.88Kb | telomerase reverse transcriptase | | | | Cell Cycle DNA replication and repair Disease Signal Transduction | | 96 | 40 |
| | MIR4457 | 5 | 1,309,424 | 1,309,492 | 69b | microRNA 4457 | | | | | | | 7 |
| | CLPTM1L | 5 | 1,317,999 | 1,345,002 | 27.00Kb | CLPTM1-like | | | | | | 6 | 24 |
| | SLC6A3 | 5 | 1,392,904 | 1,445,543 | 52.64Kb | solute carrier family 6 (neurotransmitter transporter), member 3 | | | | Amine compound SLC transporters Neuronal System Parkinson's disease Transmembrane transport of small molecules | | 48 | 22 |

| | Loc | Width | Band1 | Band2 | Num.probes | Status | log2(ratio) | Ratio | Genes | CNV | miRNA | CpGis! | |
|------|----------------------|----------|------------|------------|-------------|--|-------------|-------|----------------|------|-------|--------|----|
| 2 | 16:106270-7487257 | 7.38Mb | 16p13.3 | 16p13.3 | 560 | G | 0.30 | 1.23 | 2 4 1 74 215 | 1313 | 43 | 450 | |
| 3 | 16:8038064-12214112 | 4.18Mb | 16p13.2 | 16p13.13 | 252 | G | 0.32 | 1.25 | 1 1 1 1 9 33 | 322 | 2 | 39 | |
| 4 | 16:13609462-22718338 | 9.11Mb | 16p13.12 | 16p12.2 | 432 | G | 0.35 | 1.28 | 1 1 1 26 83 | 724 | 55 | 102 | |
| 5 | 16:22985536-30510531 | 7.52Mb | 16p12.2 | 16p11.2 | 422 | G | 0.35 | 1.27 | 2 1 46 105 | 496 | 16 | 114 | |
| 6 | 16:31016726-31157900 | 141.18Kb | 16p11.2 | 16p11.2 | 13 | L | -0.39 | 0.76 | 5 10 | 14 | - | 11 | |
| 7 | 16:31934833-35147508 | 3.21Mb | 16p11.2 | 16p11.1 | 70 | G | 0.30 | 1.23 | 11 21 | 651 | - | 59 | |
| 8 | 16:46500740-72458109 | 25.96Mb | 16q11.2 | 16q22.2 | 1456 | L | -0.38 | 0.77 | 1 10 1 115 251 | 1377 | 28 | 268 | |
| 9 | 16:72499170-72960504 | 461.33Kb | 16q22.2 | 16q22.3 | 25 | L | -2.41 | 0.19 | 1 1 | 18 | - | 2 | |
| Gene | Chr | Start | End | Width | Description | | | | Pathways | | CTD | CNV | |
| | ZHX3 | 16 | 72,816,785 | 73,092,534 | 275.75Kb | zinc finger homeobox 3 | | | | | | 17 | 12 |

Subpopulations and annotations (Discrete data)

| | Hclust A | | | Hclust B | | | Hclust C |
|---|----------|-----|----|----------|----|-----|----------|
| Stage at diagnostic | | 13 | | 16 | | 20 | |
| Stage 1 | 0 | 0% | 9 | 56% | 4 | 20% | |
| Stage 2 | 8 | 62% | 4 | 25% | 13 | 65% | |
| Stage 3 | 3 | 23% | 3 | 19% | 3 | 15% | |
| Stage 4 | 2 | 15% | 0 | 0% | 0 | 0% | |
| Distant metastasis at 4 years | | 11 | | 8 | | 11 | |
| Positive (at least one) | 9 | 82% | 4 | 50% | 3 | 27% | |
| Negative (zero) | 2 | 18% | 4 | 50% | 8 | 73% | |
| Distant metastasis-free survival | | 13 | | 16 | | 20 | |
| Positive | 2 | 15% | 12 | 75% | 15 | 75% | |
| Negative | 11 | 85% | 4 | 25% | 5 | 25% | |
| Death | | 13 | | 15 | | 19 | |
| Positive | 10 | 77% | 6 | 40% | 9 | 47% | |
| Negative | 3 | 23% | 9 | 60% | 10 | 53% | |

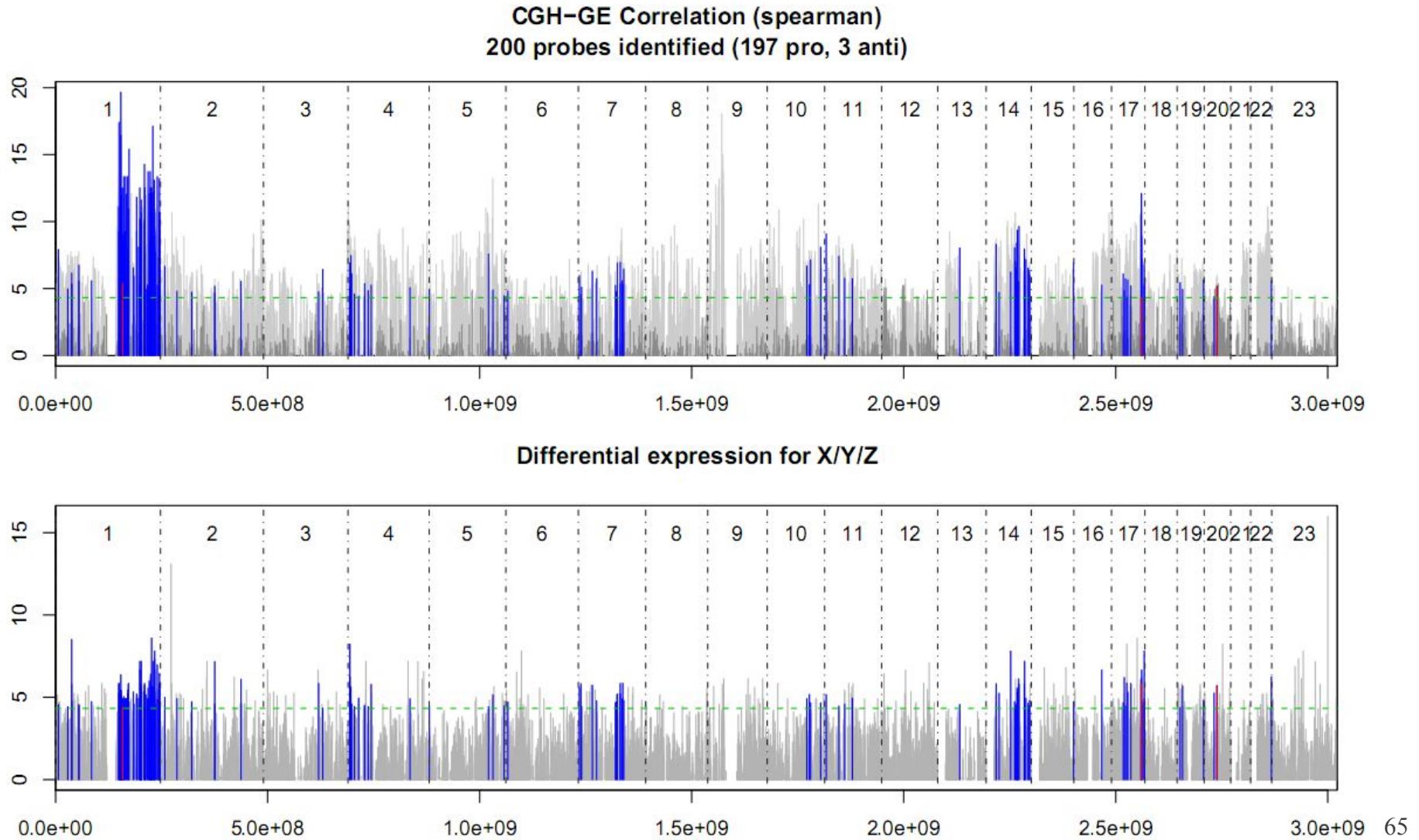
Subpopulations and annotations (Discrete data)



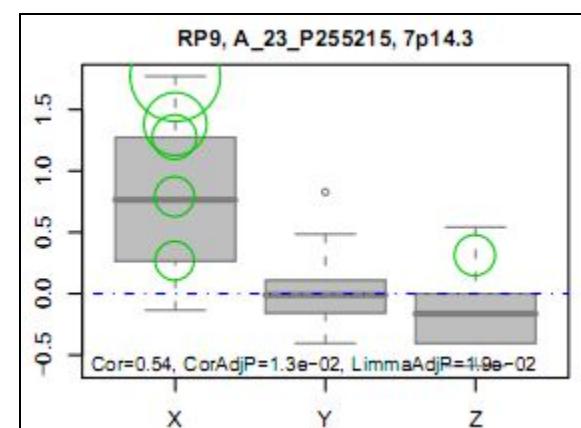
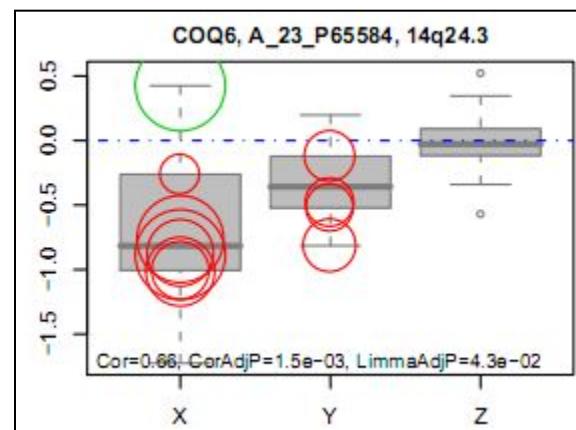
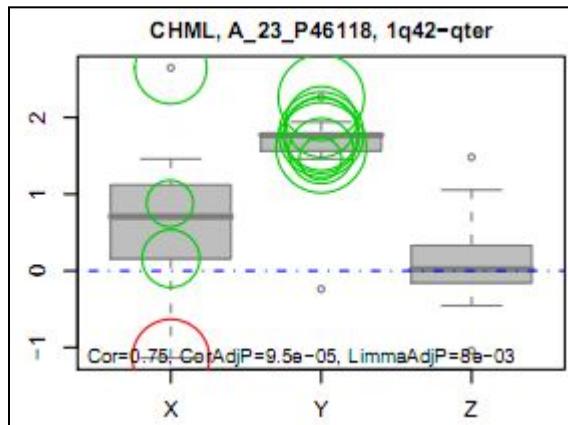
Fisher's exact test, adjusted P-values

| F-tests | Aberrant_Genome | Hclust_2branches |
|-----------------|-----------------|------------------|
| Age_median_68.5 | 8.19E-01 | 1.00E+00 |
| Age_mean_67.45 | 4.11E-01 | 7.97E-01 |
| Sex | 1.06E-01 | 4.32E-01 |
| TNM_Simplified | 4.32E-01 | 3.75E-01 |
| TNM_1 | 7.11E-01 | 1.00E+00 |
| TNM_2 | 2.48E-01 | 3.29E-01 |
| TNM_3+4 | 9.31E-01 | 5.92E-01 |
| Differentiation | 4.46E-01 | 3.44E-01 |
| Fibrosis | 2.37E-01 | 2.59E-01 |
| Tissue_Invasion | 5.88E-01 | 7.81E-01 |
| Mutation_EGFR | 9.25E-06 | 8.60E-05 |
| Mutation_KRAS | 2.44E-01 | 8.30E-02 |
| EGFRwt_KRASwt | 2.59E-05 | 6.87E-03 |

CNA+GEX integration : Correlation analysis

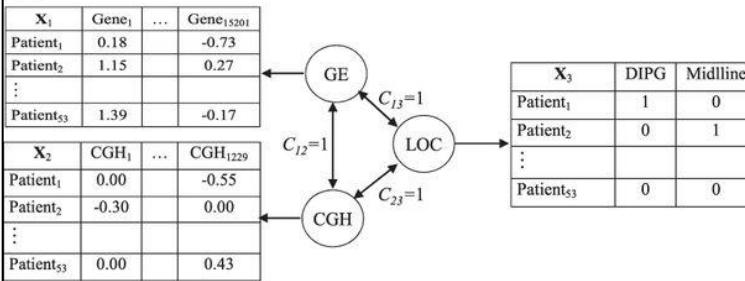


CGH+GEX integration : Correlation analysis

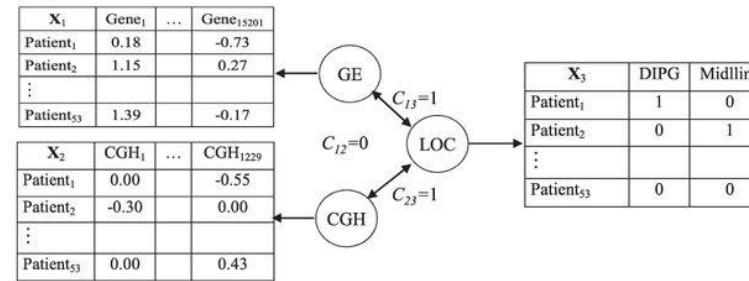


CNA+GEX integration : Multiblock analysis

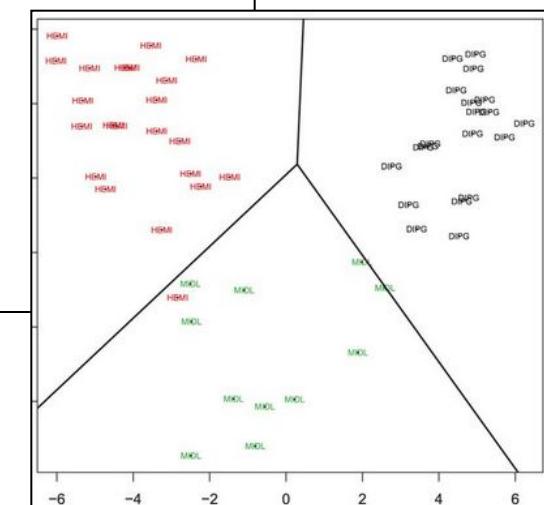
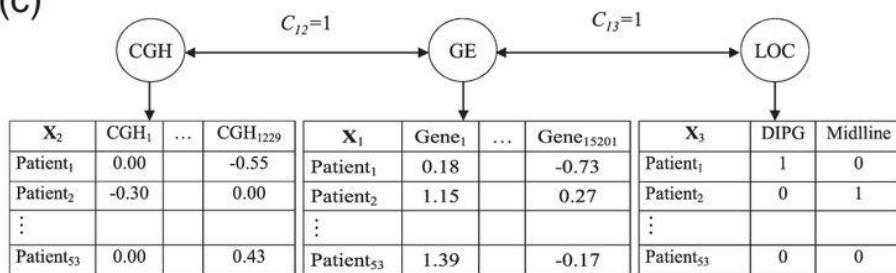
(a)



(b)

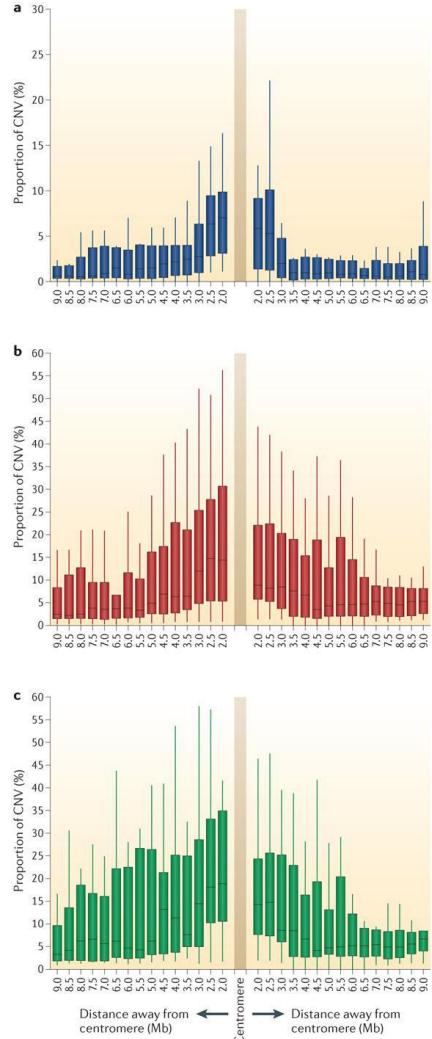


(c)

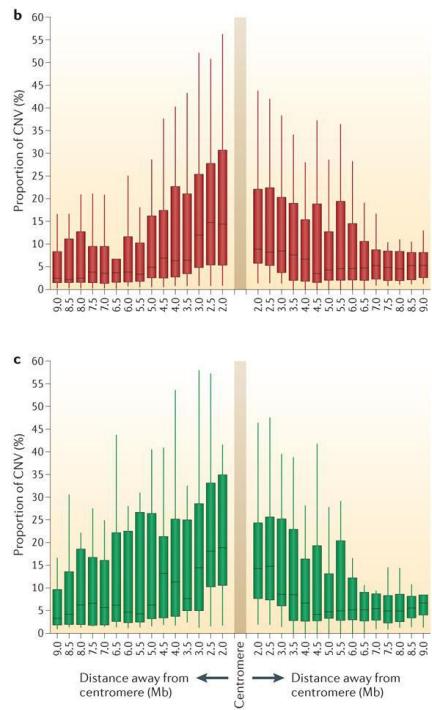


Interesting findings : CNAs and chromosomal locations

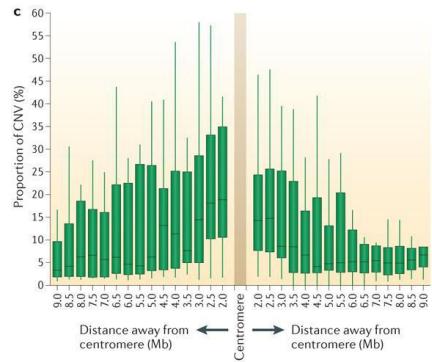
Frequency of CNA and distance to centromere



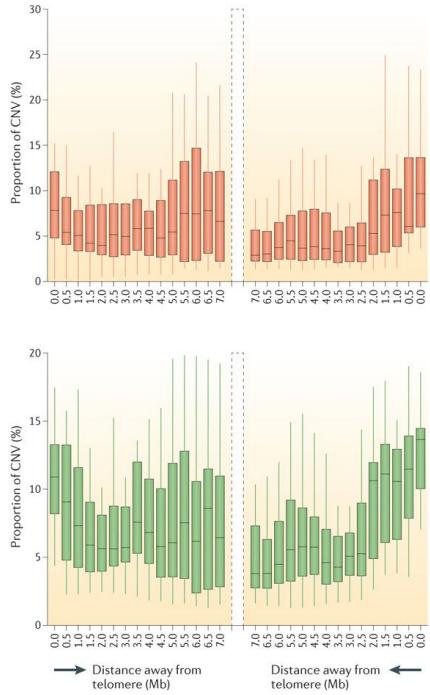
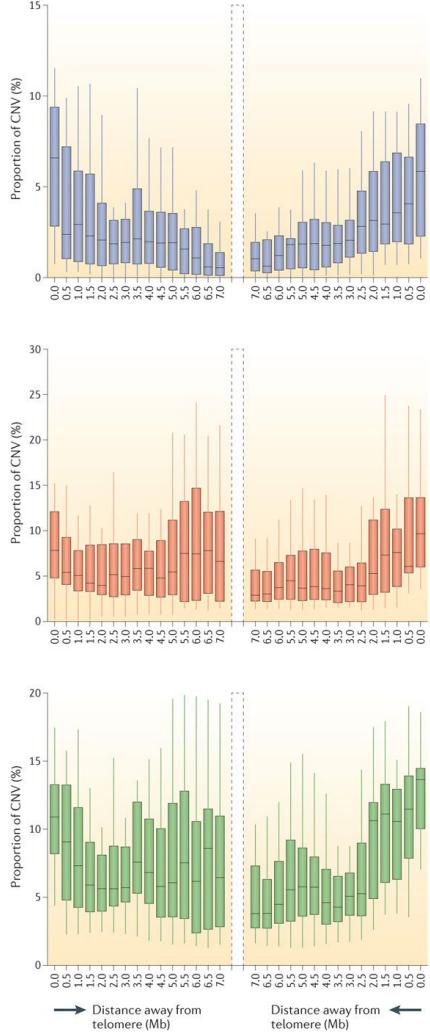
GAIN



LOSS

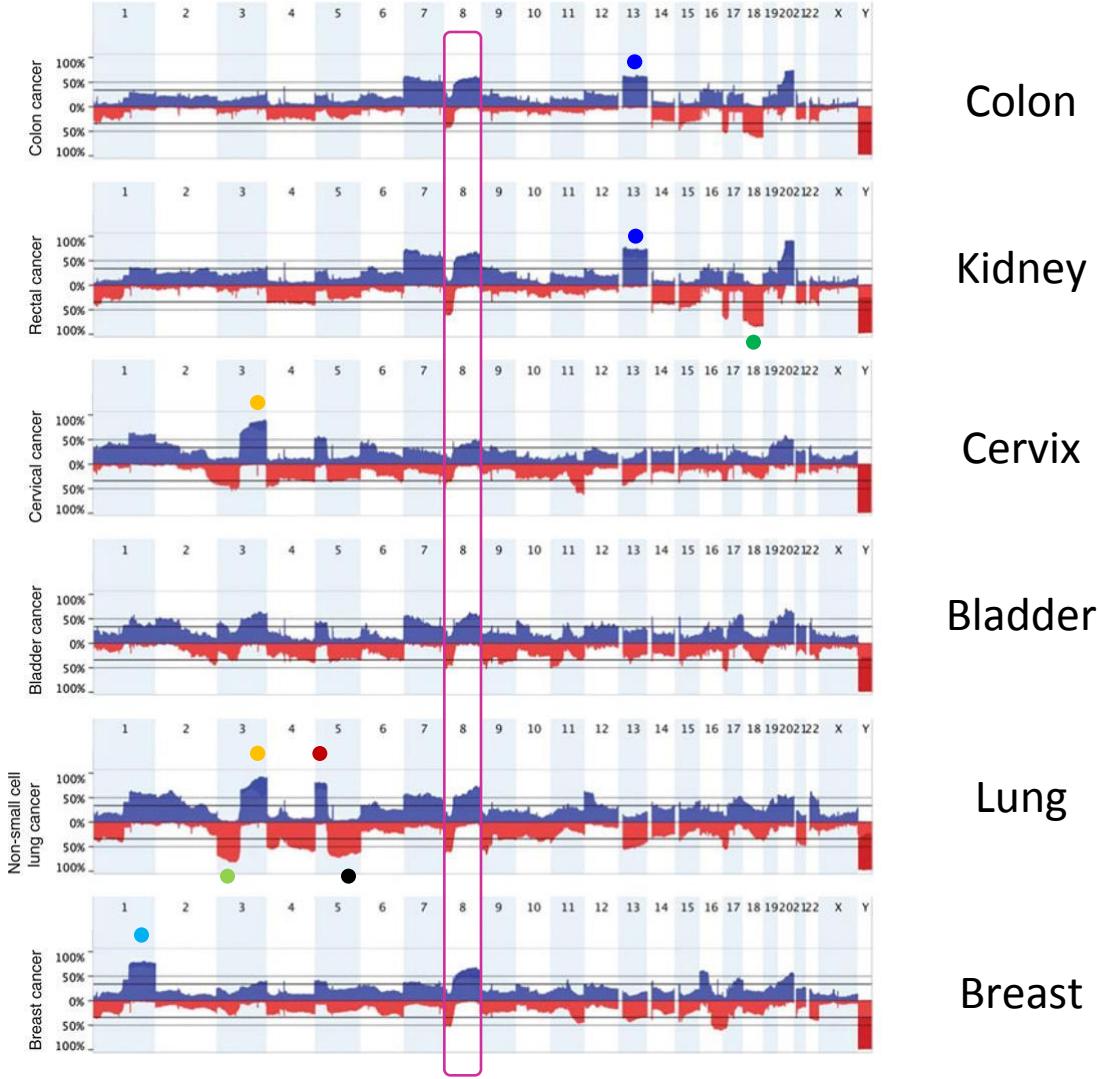
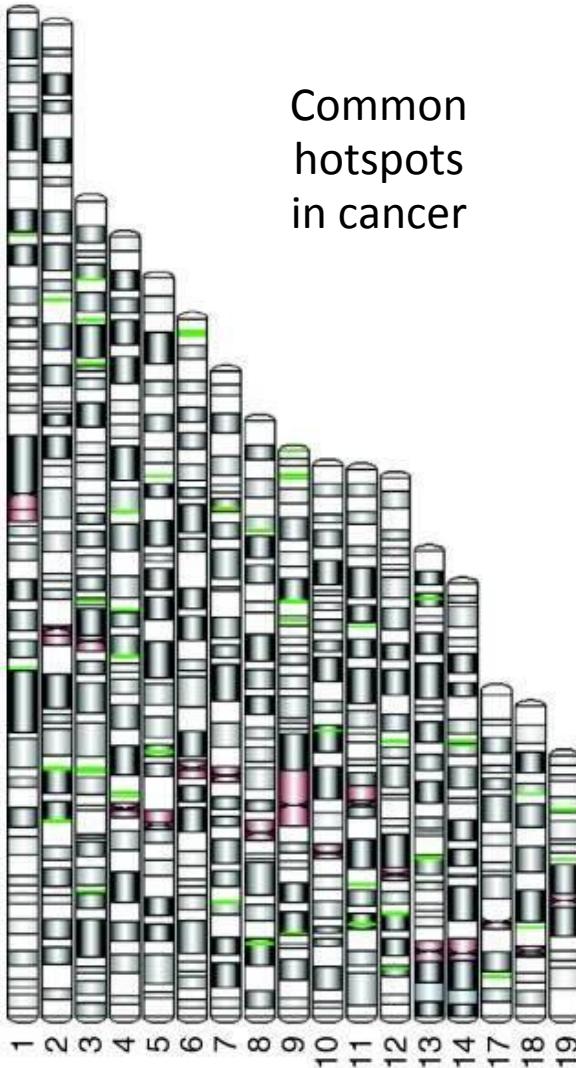


OVERALL

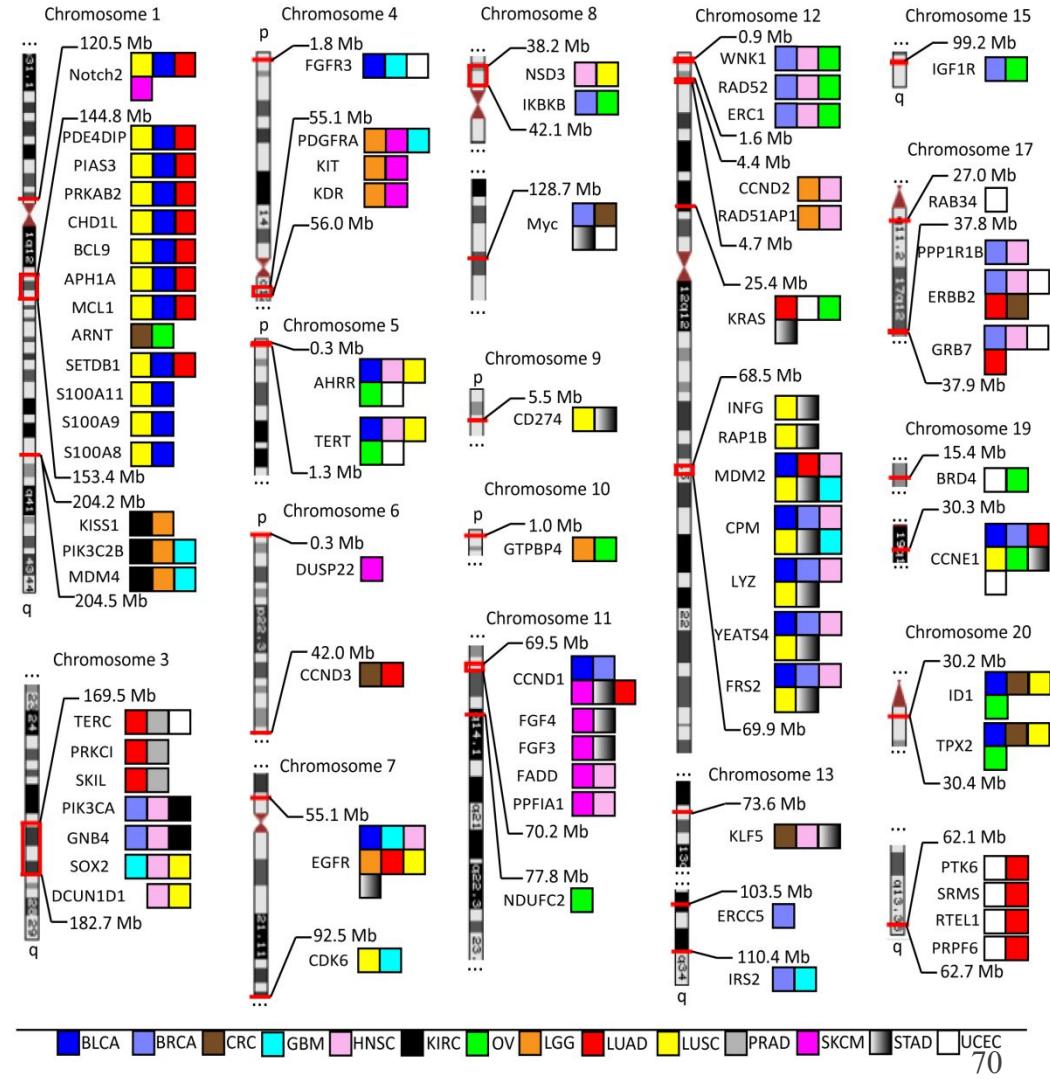
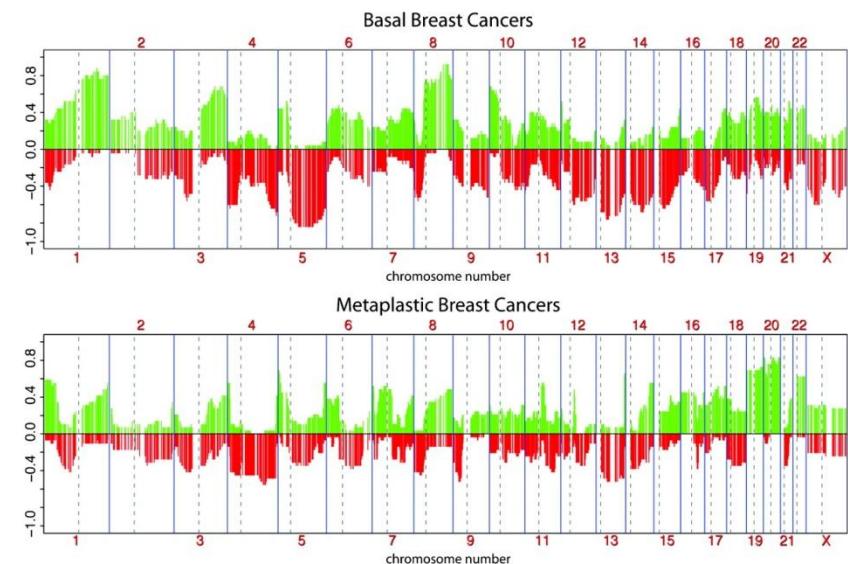


Frequency of CNA and distance to telomere

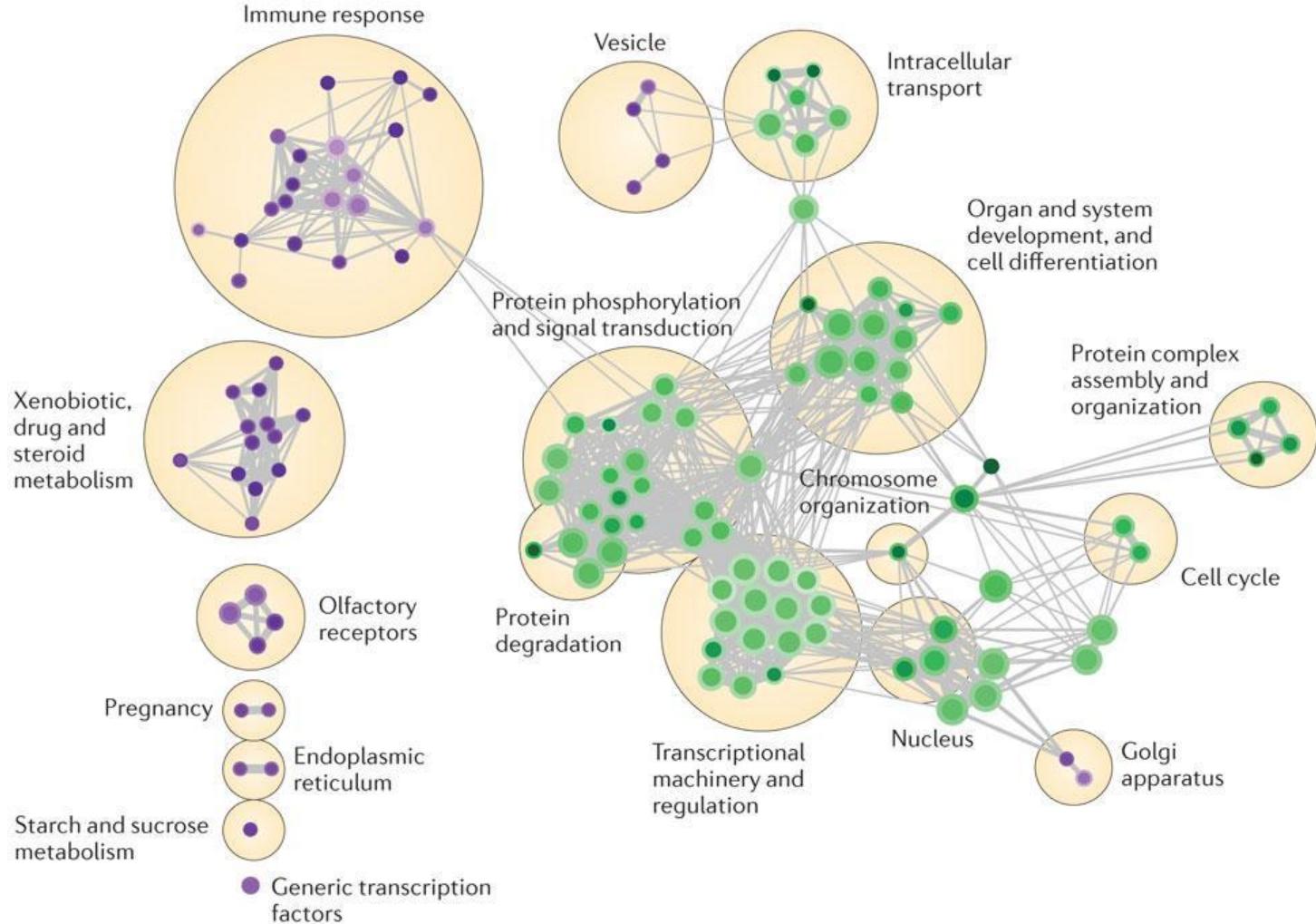
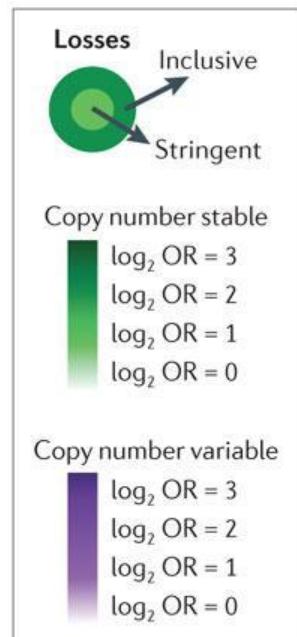
Interesting findings : As (globally) close as (locally) different ...



Interesting findings : As (globally) close as (locally) different ...



Interesting findings : Genomic losses and cell function



APPENDIX

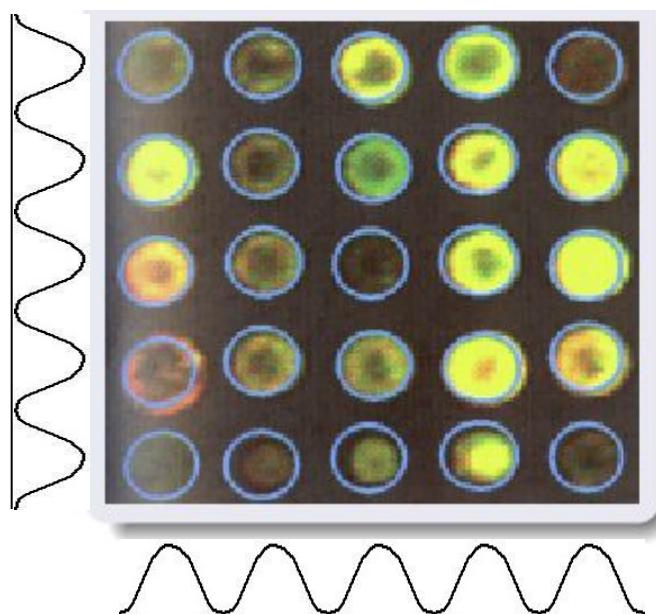
APPENDIX :

Signal acquisition & QC

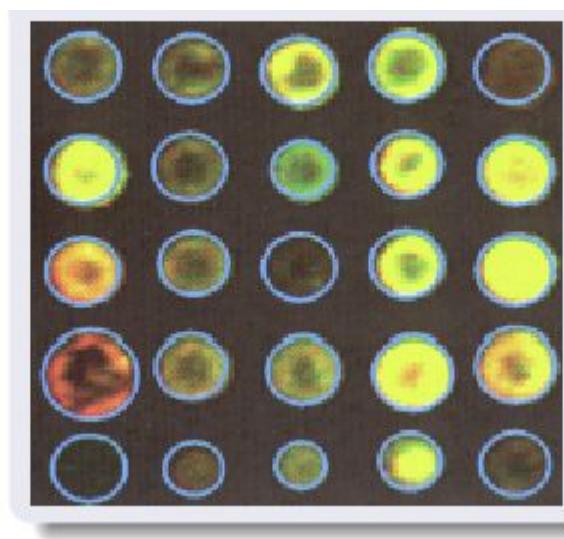
Spot identification

Credits : Pierre NEUVIAL (ENSAE)

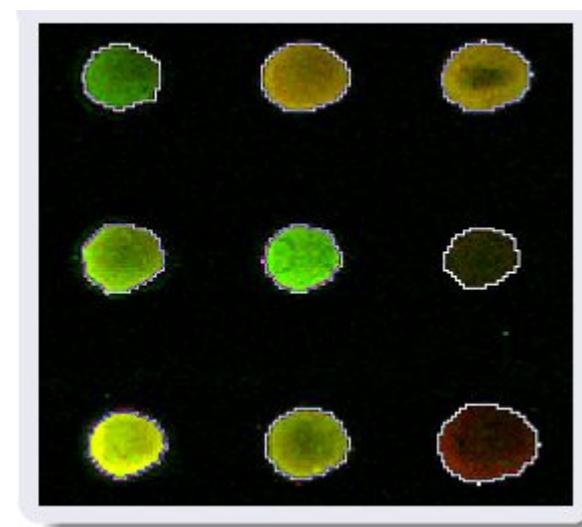
Fixed radius



Adaptative radius



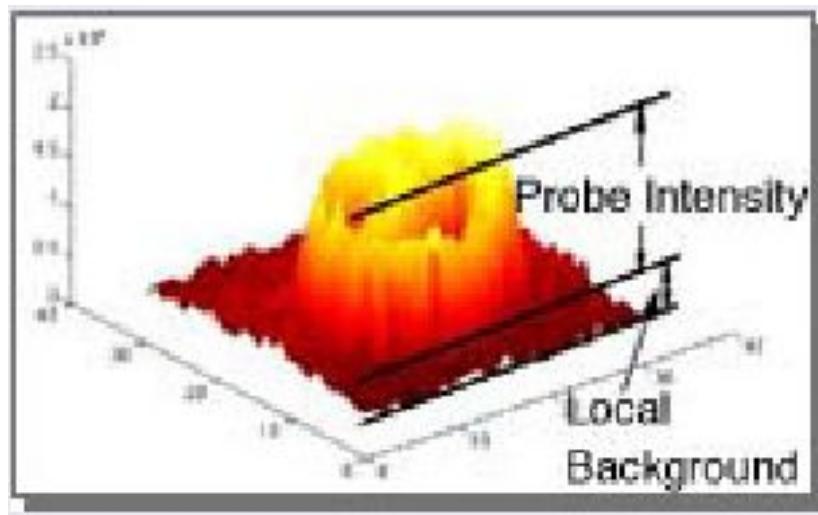
Pixel seeding



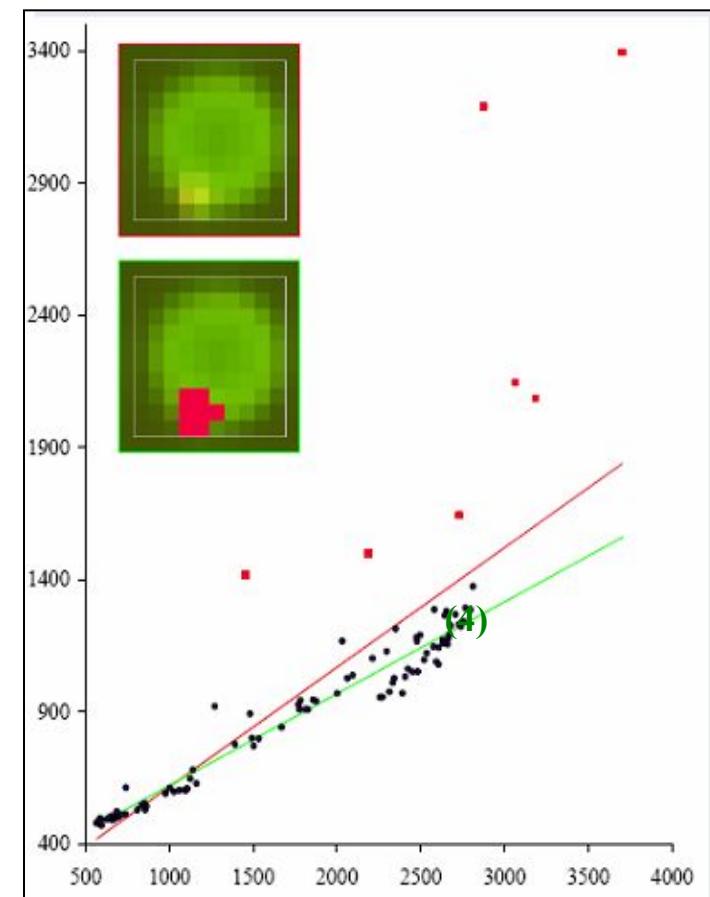
NGS clusters

Signals acquisition

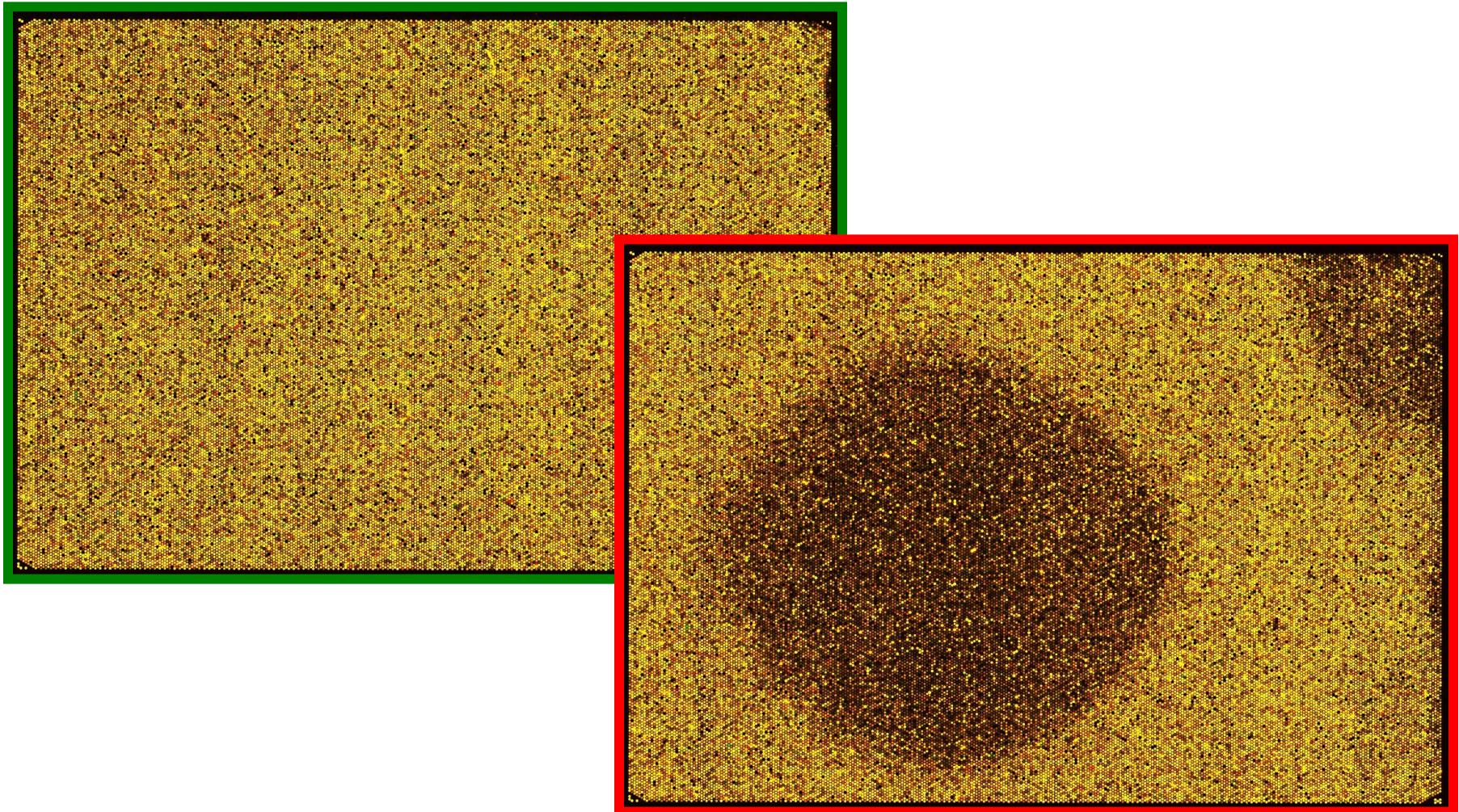
Intensity segmentation



Linear regression w/
outlier detection



QC : Scan observation

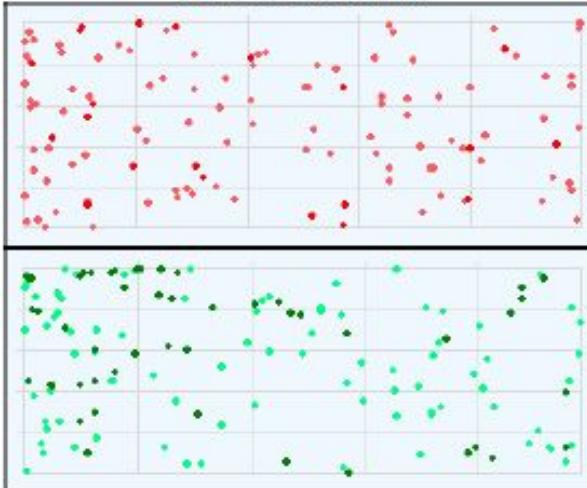


QC : Reports

QC Report - Agilent Technologies : 2 Color CGH

| | | | |
|------------------|------------------------------|------------------------|-------------------|
| Date | Monday, May 03, 2010 - 11:19 | Sample(red/green) | |
| User Name | ugf | FE Version | 10.7.3.1 |
| Image | US82900153_251469363399_S01 | BG Method | Detrend on (NegC) |
| Protocol | CGH_107_Sep09 (Read Only) | Multiplicative Detrend | True |
| Grid | 014693_D_F_20090929 | Dye Norm | Linear |
| Saturation Value | 65528 (r), 65528 (g) | | |

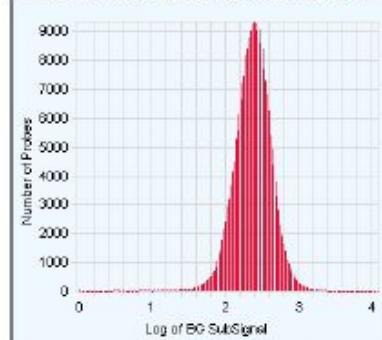
Outlier Numbers with Spatial Distribution
534 rows x 456 columns



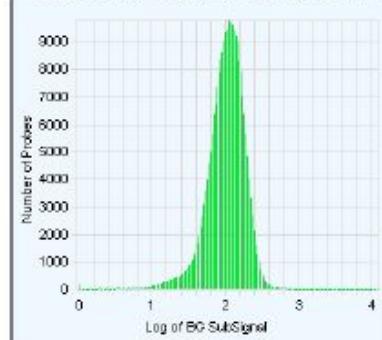
Red FeaturePopulation Red Feature NonUniform
Green FeaturePopulation Green Feature NonUniform

| Feature | Red | Green | Any | % Outlier |
|------------------------|-----|-------|-----|-----------|
| Non Uniform Population | 22 | 53 | 60 | 0.02 |
| Population | 101 | 84 | 169 | 0.07 |

Histogram of Signals Plot (Red)



Histogram of Signals Plot (Green)

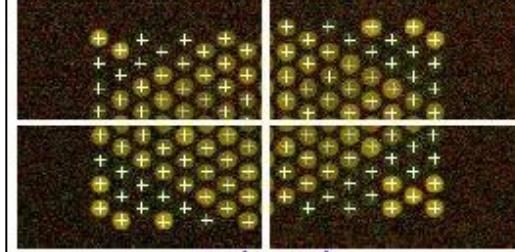


Evaluation Metrics for CGH_QCMT_Sep09 :
Excellent (5) ; Good (6)

| Metric Name | Value | Excellent | Good | Evaluate |
|--------------------------|--------|-----------|--------------|-------------|
| IsGoodGrid | 1.00 | >1 | NA | <1 |
| AnyColorPrntFeatNonUn... | 0.02 | <1 | 1 to 5 | >5 |
| DerivativeLR_Spread | 0.21 | <0.20 | 0.20 to 0.30 | >0.30 |
| g_Repro | 0.08 | 0 to 0.05 | 0.05 to 0.20 | <0 or >0.20 |
| g_BGNoise | 1.34 | <5 | 5 to 10 | >10 |
| g_Signal2Noise | 71.43 | >100 | 30 to 100 | <30 |
| g_SignalIntensity | 95.85 | >150 | 50 to 150 | <50 |
| r_Repro | 0.09 | 0 to 0.05 | 0.05 to 0.20 | <0 or >0.20 |
| r_BGNoise | 2.26 | <5 | 5 to 10 | >10 |
| r_Signal2Noise | 93.53 | >100 | 30 to 100 | <30 |
| r_SignalIntensity | 211.46 | >150 | 50 to 150 | <50 |

♦ Excellent ♦ Good ♦ Evaluate

Spot Finding of the Four Corners of the Array



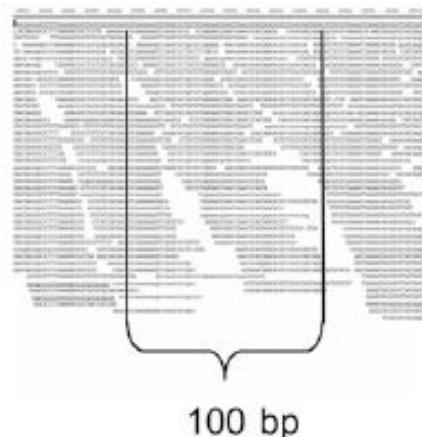
Grid Normal

APPENDIX :

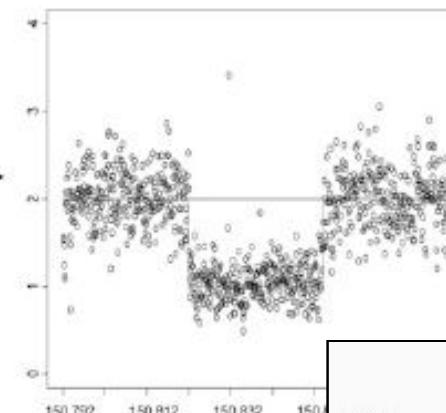
WES – Read depth

NGS : Read depth

(A) Estimation of Read Depth



(B) Event detection



Yoon, 2009

