生物信息学

基因组变异分析

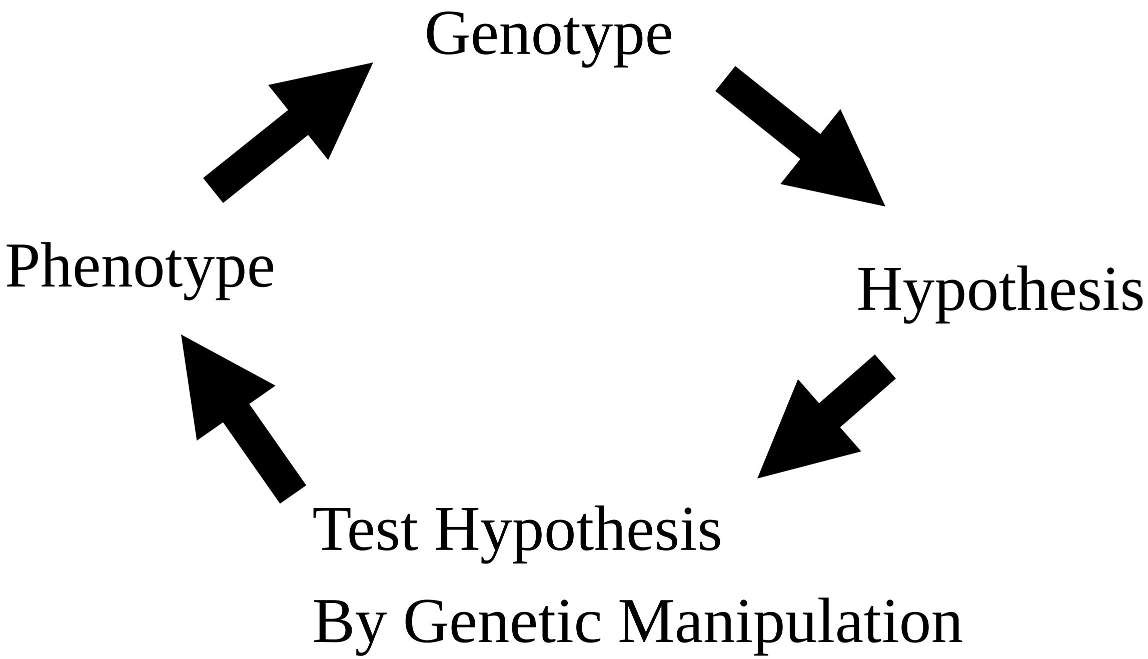
吴凌云

中国科学院数学与系统科学研究院

http://zhangroup.aporc.org

Chinese Academy of Sciences

Genetics Study


Genotype

Hypothesis

Test Hypothesis

By Genetic Manipulation

Phenotype

Genetics Study

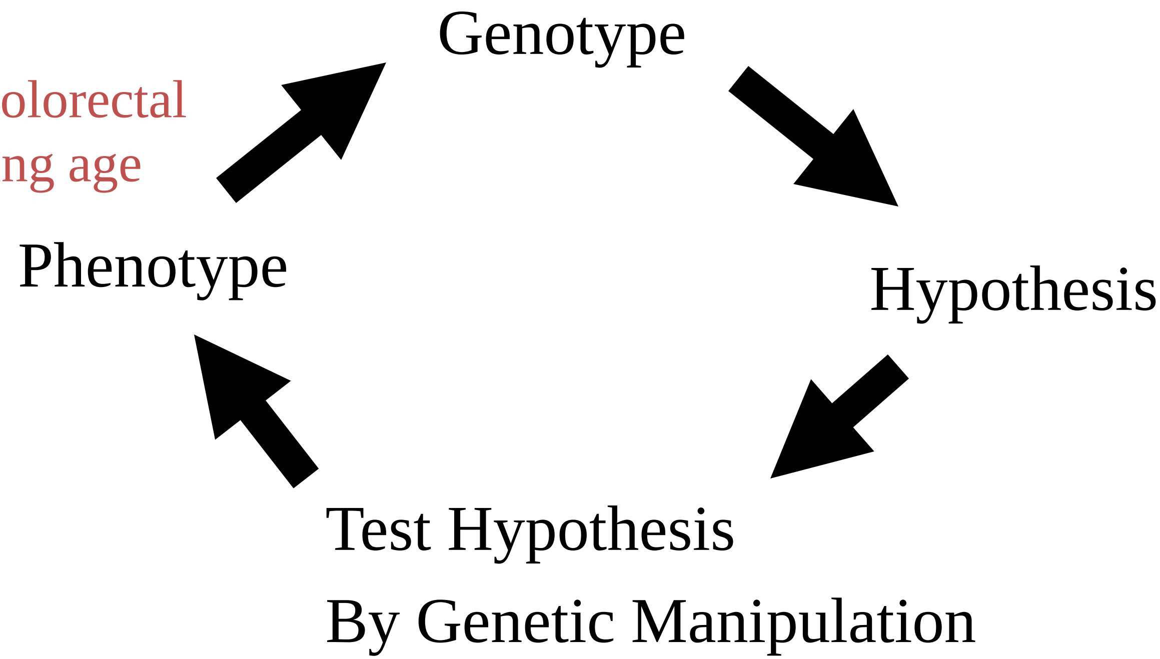
Mutation in APC

Gene

Two groups:

1. Develop Colorectal cancer at young age

2. Do not



Genotype

Hypothesis

Test Hypothesis

By Genetic Manipulation

Phenotype

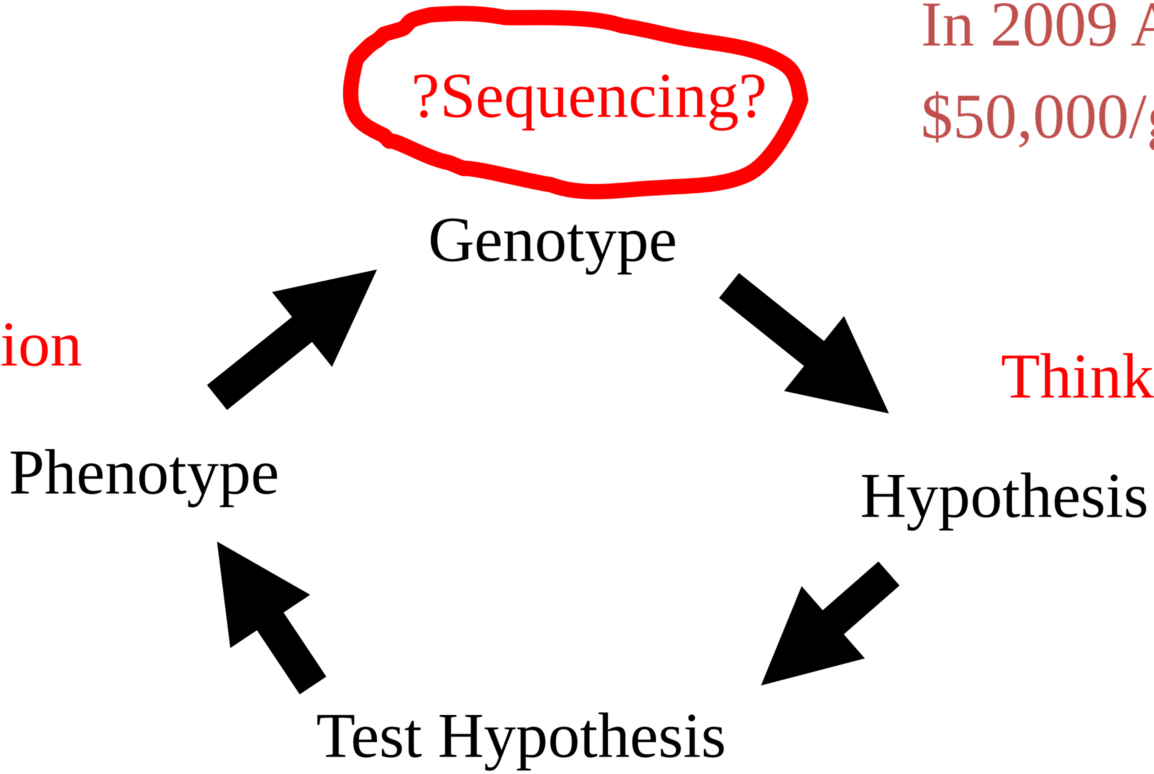
APC is a Tumor

Suppressor Gene

Delete APC in Mouse

Control: Isogenic APC+

The Cycle of Genetics Study



In 2009 Aug.

$50,000/genome

?Sequencing?

Genotype

Observation

Thinking

Phenotype

Hypothesis

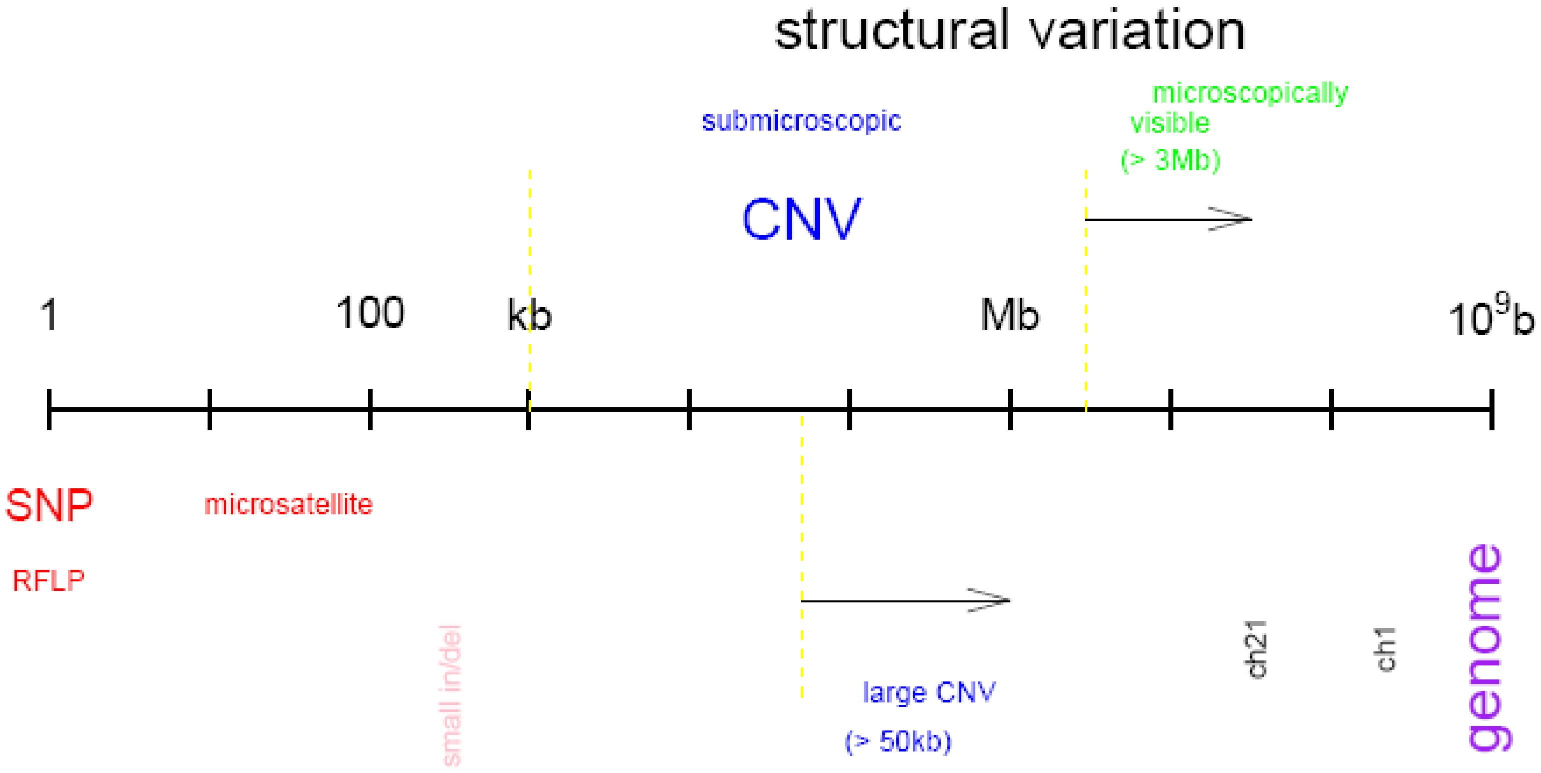
Test Hypothesis

By Genetic Manipulation

Gene Deletion/Replacement

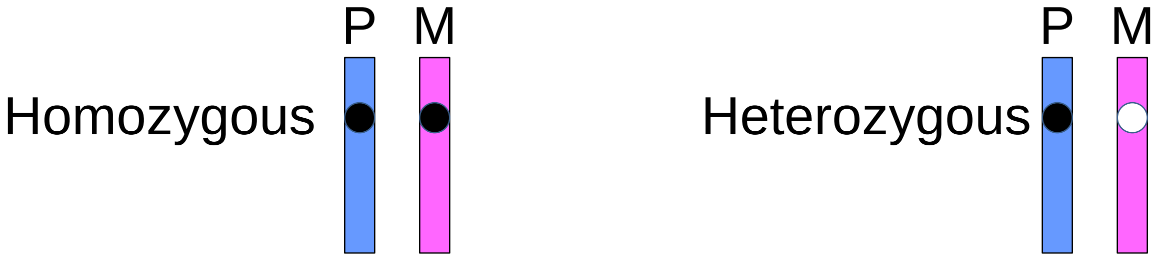
Recombinant Technology

Genome Variation



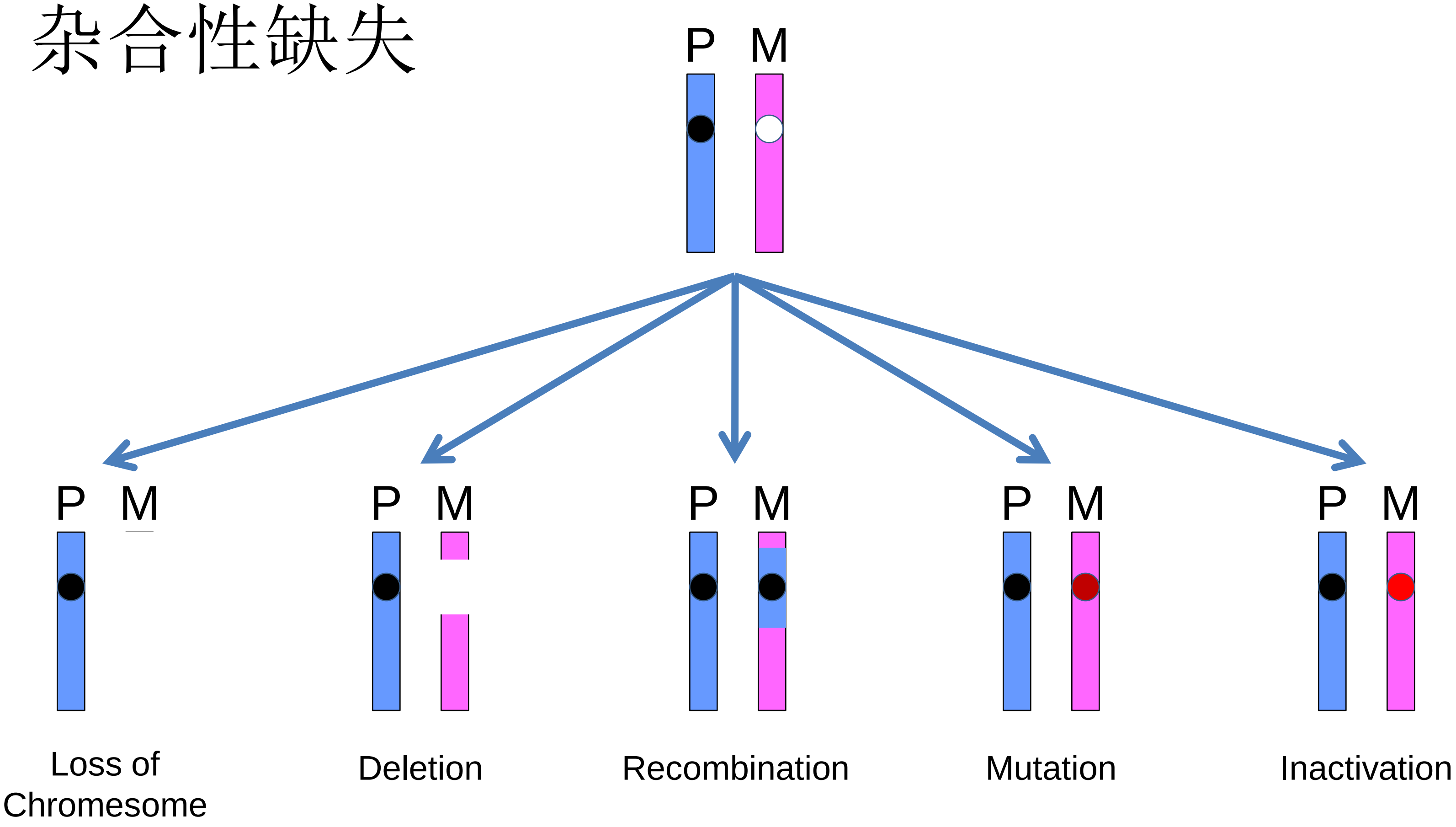
杂合性（Heterozygosity）

- Human is diploid organism
- Chromosome pair: Paternal, Maternal
- Two DNA sequences are almost identical except some mutated or polymorphic sites



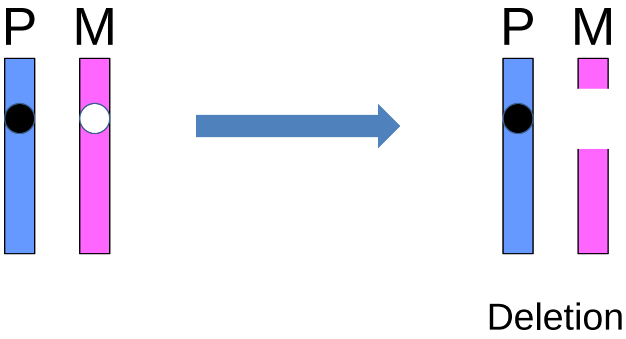
Loss of Heterozygosity (LOH)

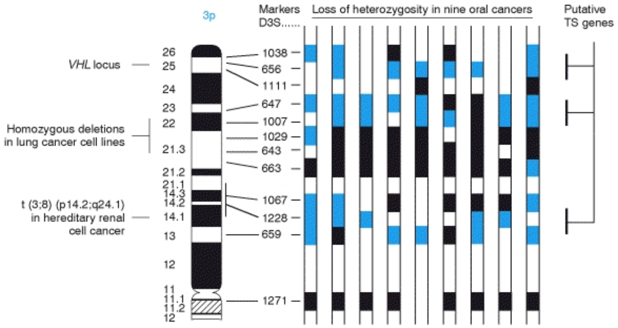
- 杂合性缺失



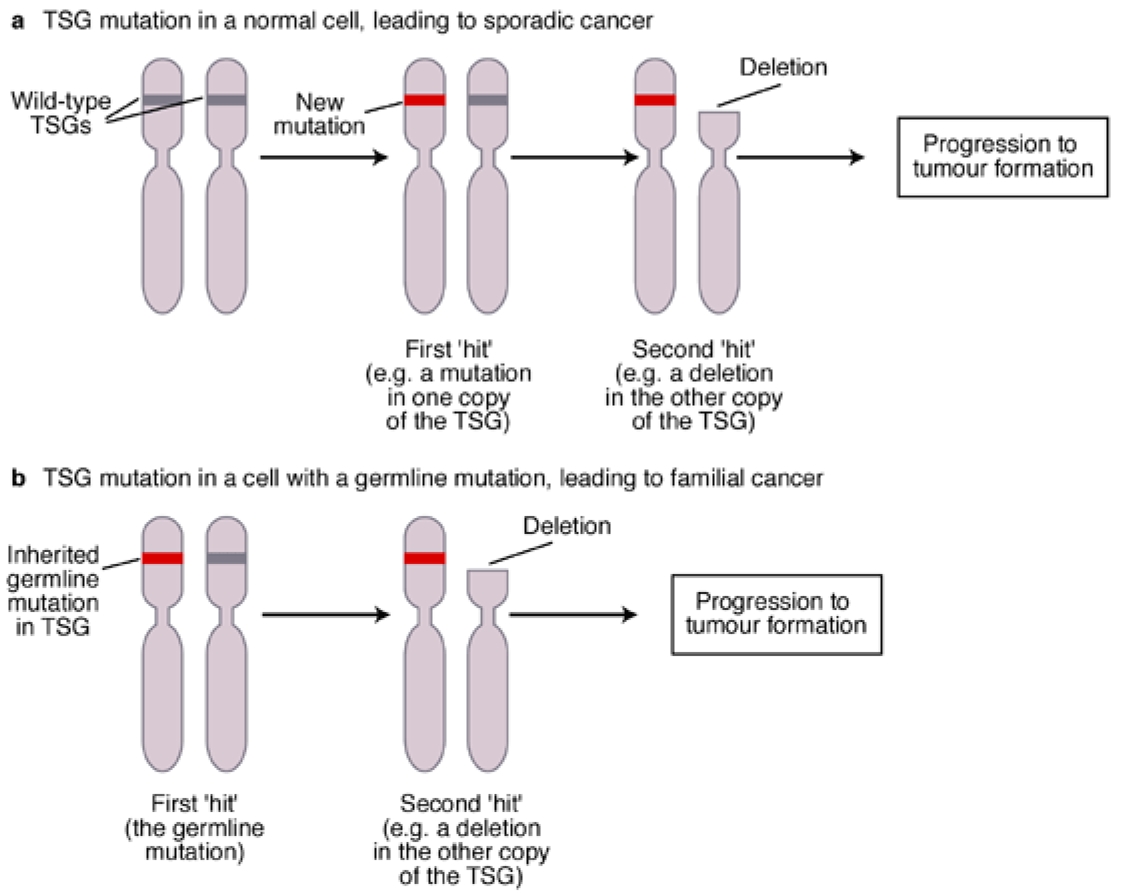
LOH and Cancer

- LOH of chromosomal regions with tumor suppressors is one of the key mechanisms in the tumor evolution.
- Identification of LOH regions will facilitate mapping susceptibility loci for cancers and disorders.





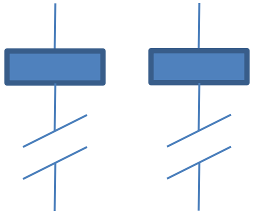
Two-Hit Hypothesis


a TSG mutation in a normal cell, leading to sporadic cancer
Wild-type TSGs
New mutation
Deletion
Progression to tumour formation
First 'hit' (e.g. a mutation in one copy of the TSG)
Second 'hit' (e.g. a deletion in the other copy of the TSG)
b TSG mutation in a cell with a germline mutation, leading to familial cancer
Inherited germline mutation in TSG
Deletion
Progression to tumour formation
First 'hit' (the germline mutation)
Second 'hit' (e.g. a deletion in the other copy of the TSG)

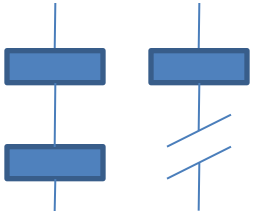
Knudson's two-hit hypothesis for tumourigenesis involving a tumour suppressor gene (TSG)

Expert Reviews in Molecular Medicine ©2001 Cambridge University Press

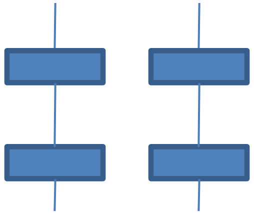
Copy Number Variation



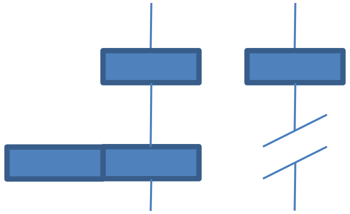
Homozygous deletion
Copy number 0



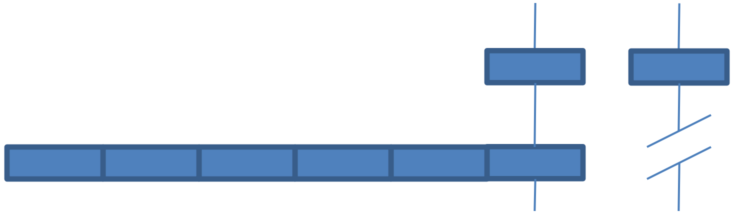
Hemizygous deletion
Copy number 1



Normal
Copy number 2



Copy neutral LOH
Copy number 2



Amplication
Copy number 6

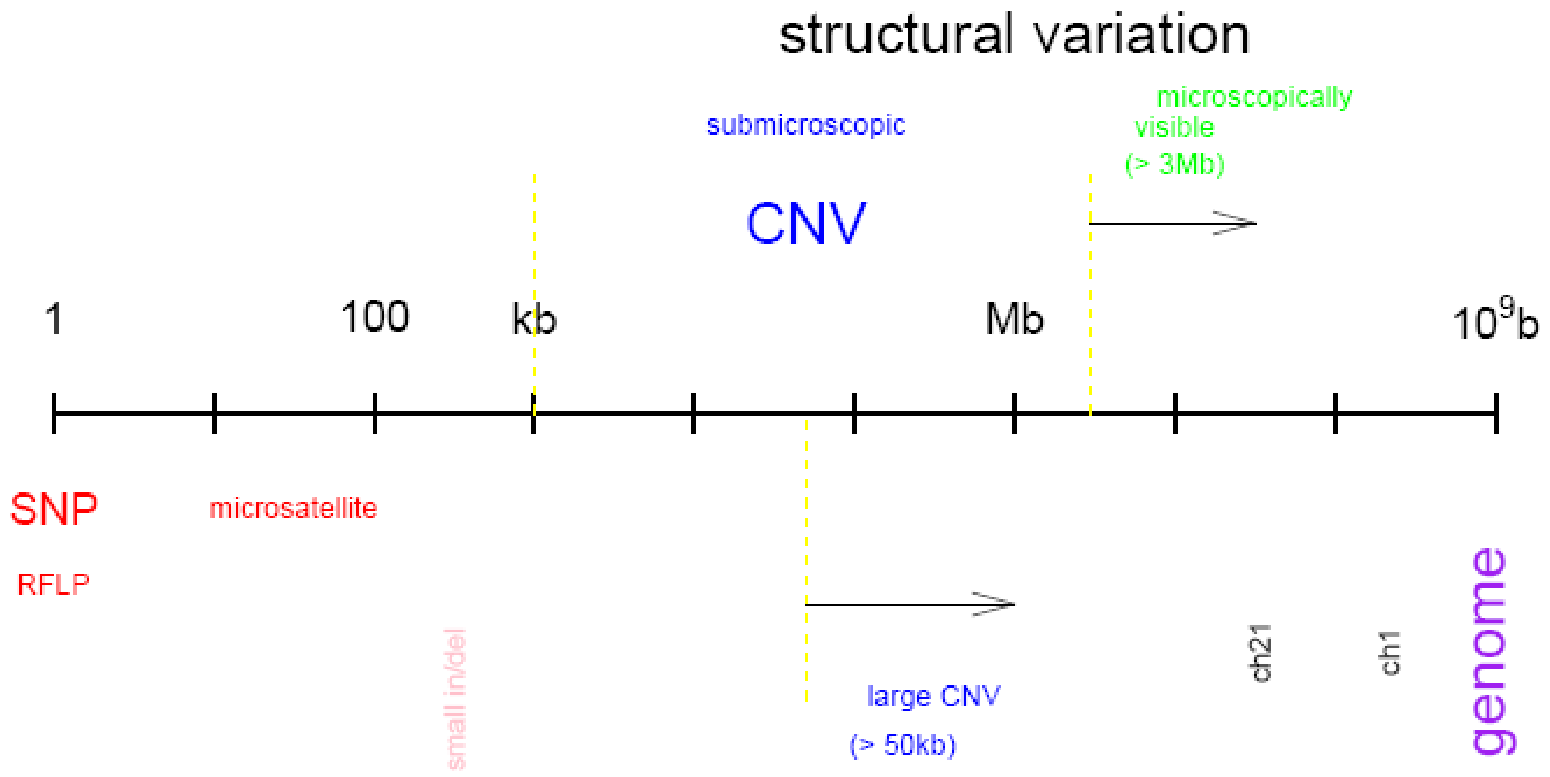
CNV & LOH

- Detection of CNV can reveal LOH due to hemizygous deletion
- Copy neutral LOH due to duplication
- LOH needs paired normal tissue from same patient, but CNV does not

CNV Terminology

- Copy number variation (germline, inherited)
  - – inherited: also present in parents’ genome
  - – de novo: absent in parents’ genome
- Copy number alteration (somatic, e.g. in cancer cells)
- Copy number polymorphism (relatively common CNV, with a fixed starting/ending position)

Genome Variation



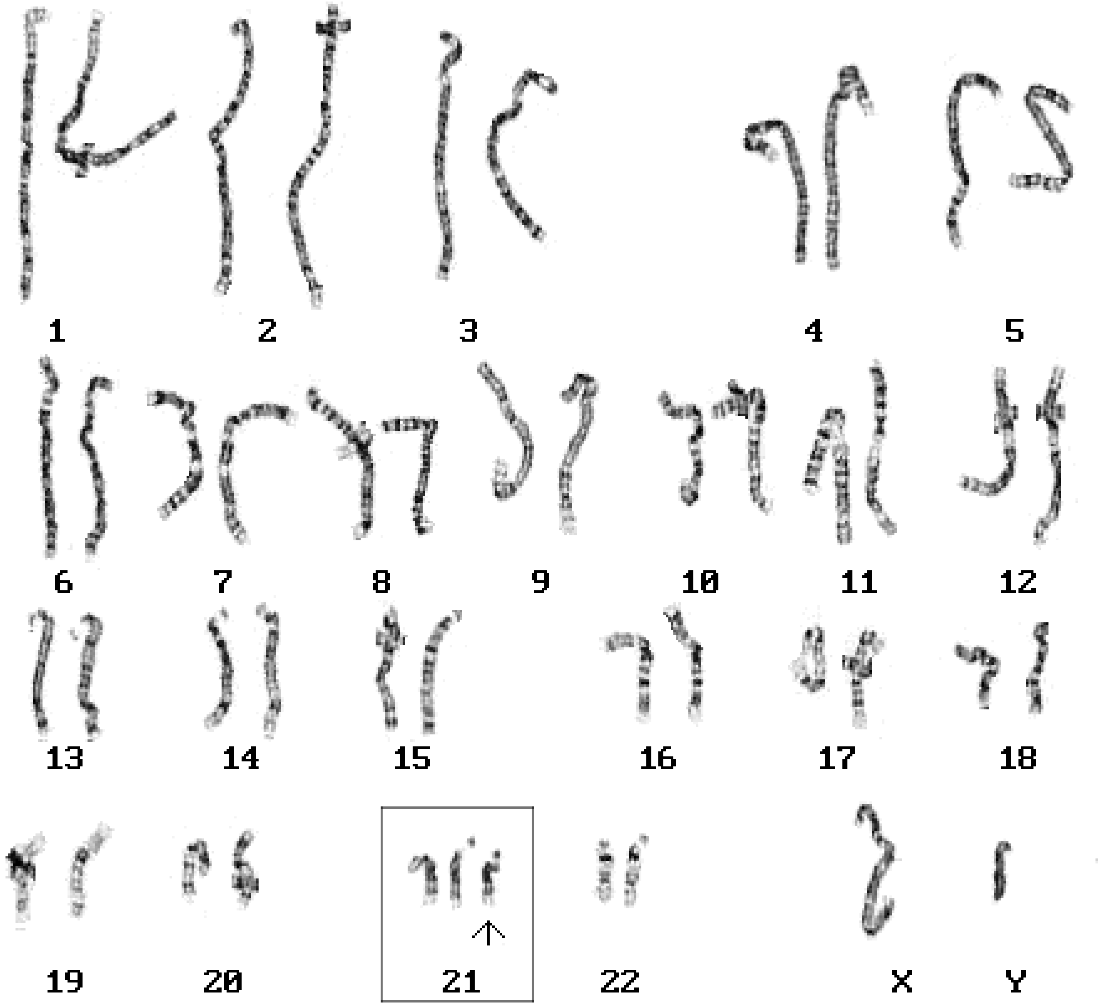
Structural Variation (1)

- Whole Genome Duplication
  - Polyploidy is common in plants (Rare in animals).
  - Survival rate after WGD may be very low. Major genomic instability would follow including massive gene losses.
  - In vertebrates, WGD is thought to occur twice around 500 million years ago (2R hypothesis).

Structural Variation (2)

- Gain or Loss of Certain Chromosomes
  - Aneuploidy (非整倍体): monosomy[1], trisomy[3], tetrasomy[4]
  - Either fatal (spontaneous abortion) or responsible for abnormal phenotypes
  - Chromosome-specific aneuploidy rate? less number of chiasmata -- shorter chromosomes: ch21, ch22
  - Down syndrome (唐氏综合症): trisomy 21

Down Syndrome

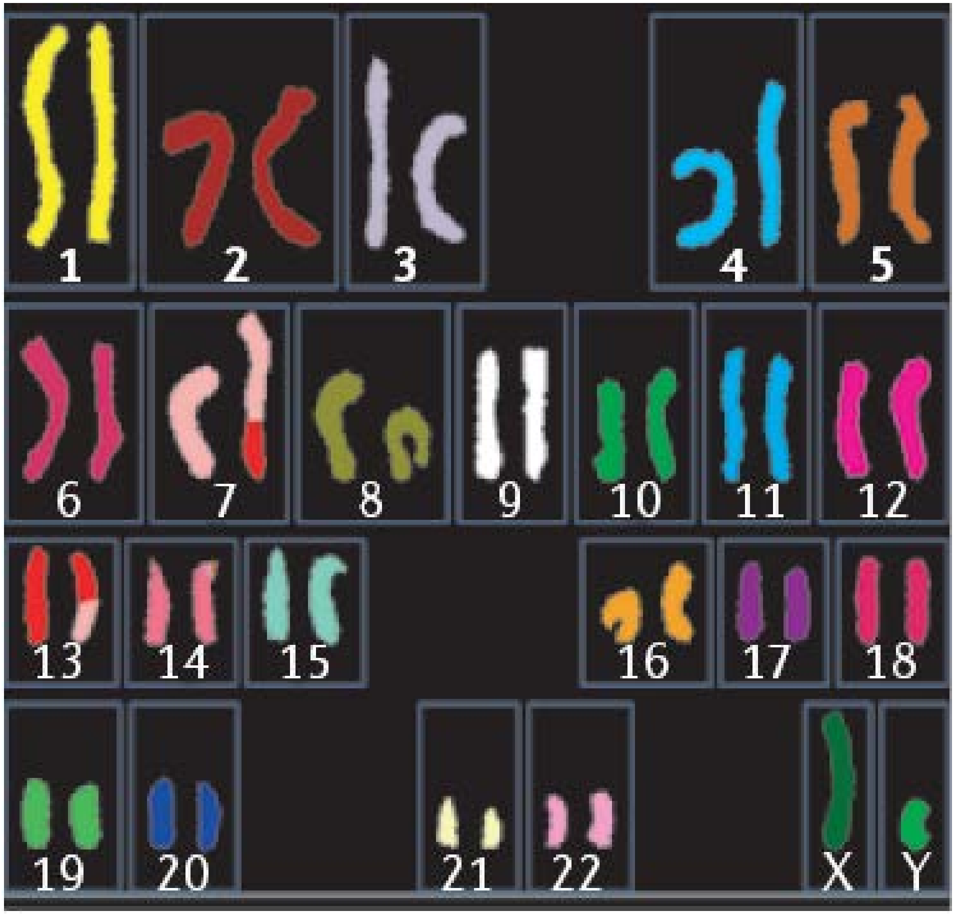


Karyotype:47,XY,+21

Structural Variation (3)

- microscopically-visible aberrations
  - Breaks
  - Double-breaks (inversion, translocation)
  - Deletions (4p, 5p, 9p, 11p/11q, 13q, 18p/18q). deletion syndromes
  - Duplications (inverted 15p). Iso-chromosomes are inverted duplications of the whole arm.
  - “balanced” vs. “unbalanced” (deletion/loss, duplication/gain)

Translocation

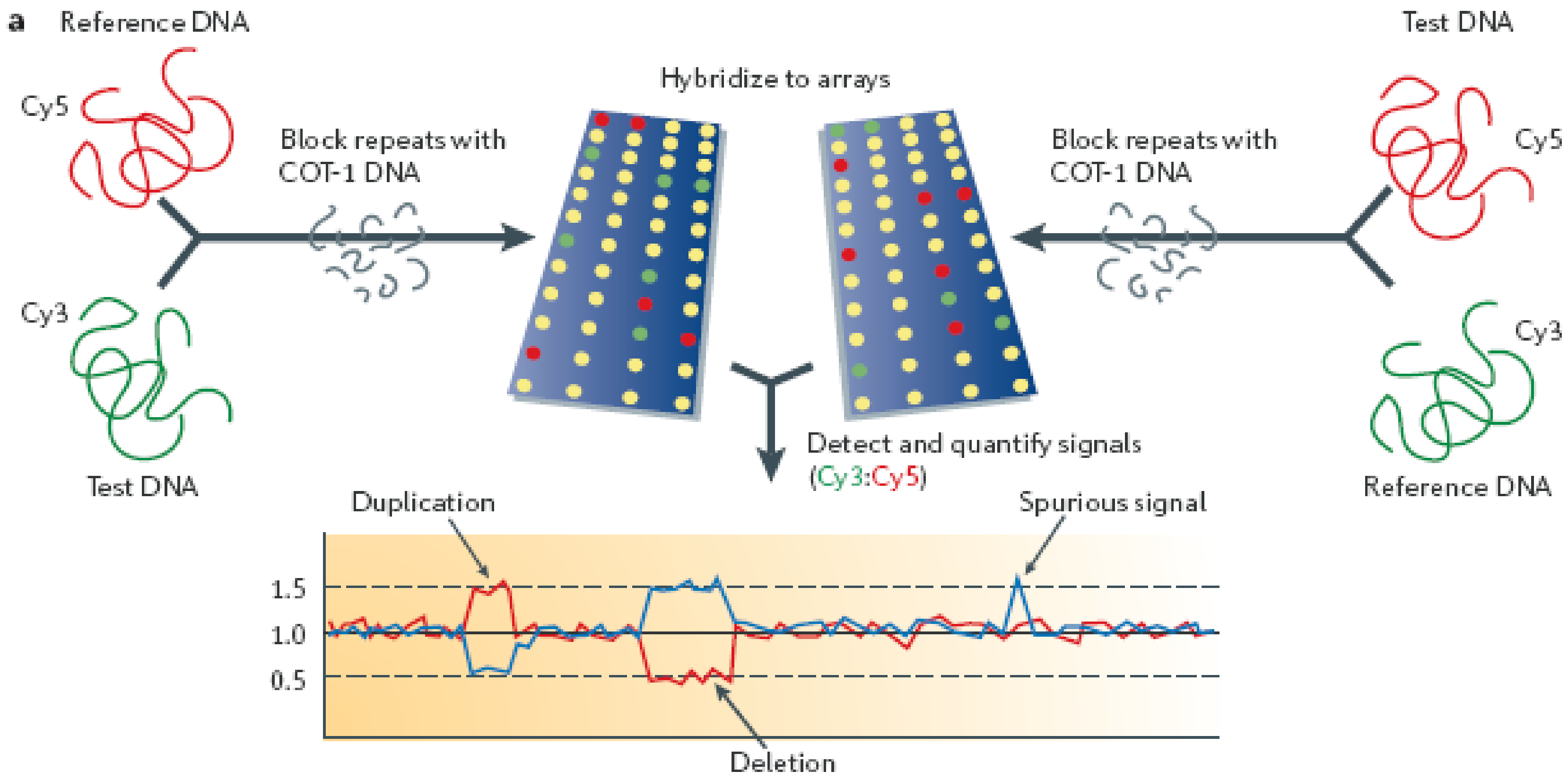


Karyotyping with each chromosome stained with a different color (Iafrate et al. 2004)

CNV的检测

- Clone-based comparative genomic hybridization (Array CGH)
  - Test and reference DNA are differentially fluorescent labeled and hybridized to the array.
  - Cons: low resolution (cannot find small CNV region)
- SNP genotyping array
  - Pros: higher resolution
  - Cons: poor signal-to-noise ratio of hybridization

CGH Array



1. Array can be spotted by any DNA sources: BAC clone, oligonucleotide…
2. “Swap” in a second hybridization to remove artifact

SNP Array

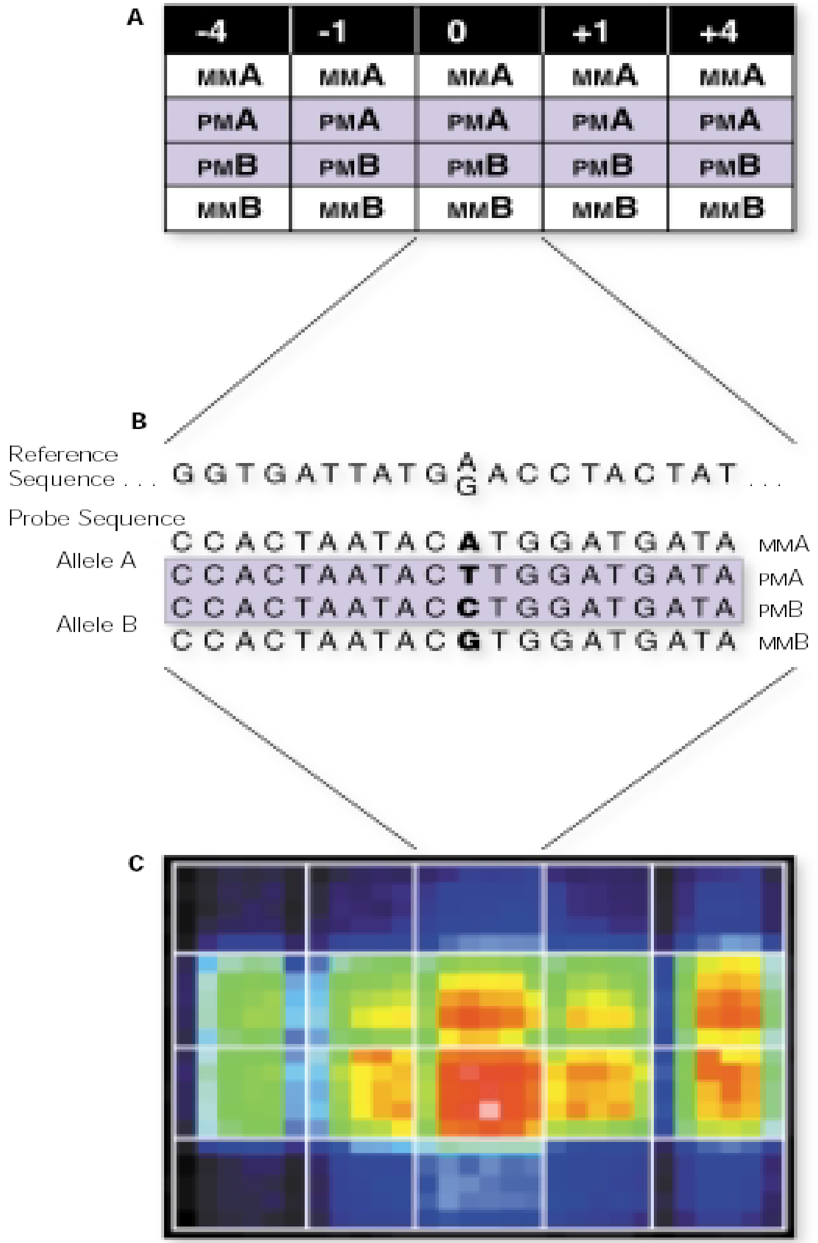
- Ilummina Bead Array
  - Human-1 Beadchip (100,000)
  - 240,000 BeadArray
  - 300,000
  - 550,000
  - 650,000
  - 1 Million (human1M)
- Affymetrix SNP array
  - 10,000 (Mapping 10K array)
  - 100,000 (Mapping 100K array)
  - 500,000 (Mapping 500K array)
  - 1 Million (Genome-wide Human SNP Array 6.0)

SNP Array

AFFYMETRIX®

illumina®

How the GeneChip® HuSNP™ Array Calls Genotypes



A

-4	-1	0	+1	+4
MMA	MMA	MMA	MMA	MMA
PMA	PMA	PMA	PMA	PMA
PMB	PMB	PMB	PMB	PMB
MMB	MMB	MMB	MMB	MMB

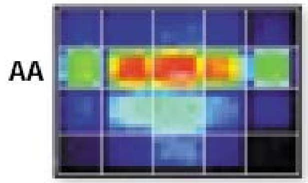
B

Reference Sequence . . . GGTGATTATG$\overset{A}{G}$ACCTACTAT . . .

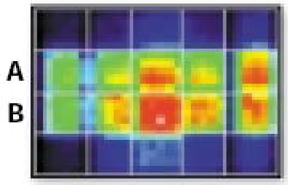
Probe Sequence

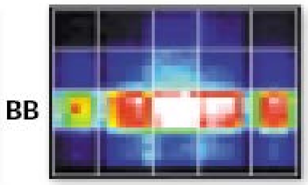
Allele A	CCACTAATAC**A**TGGATGATA	MMA
	CCACTAATAC**T**TGGATGATA	PMA
	CCACTAATAC**C**TGGATGATA	PMB
Allele B	CCACTAATAC**G**TGGATGATA	MMB

C



AA





BB

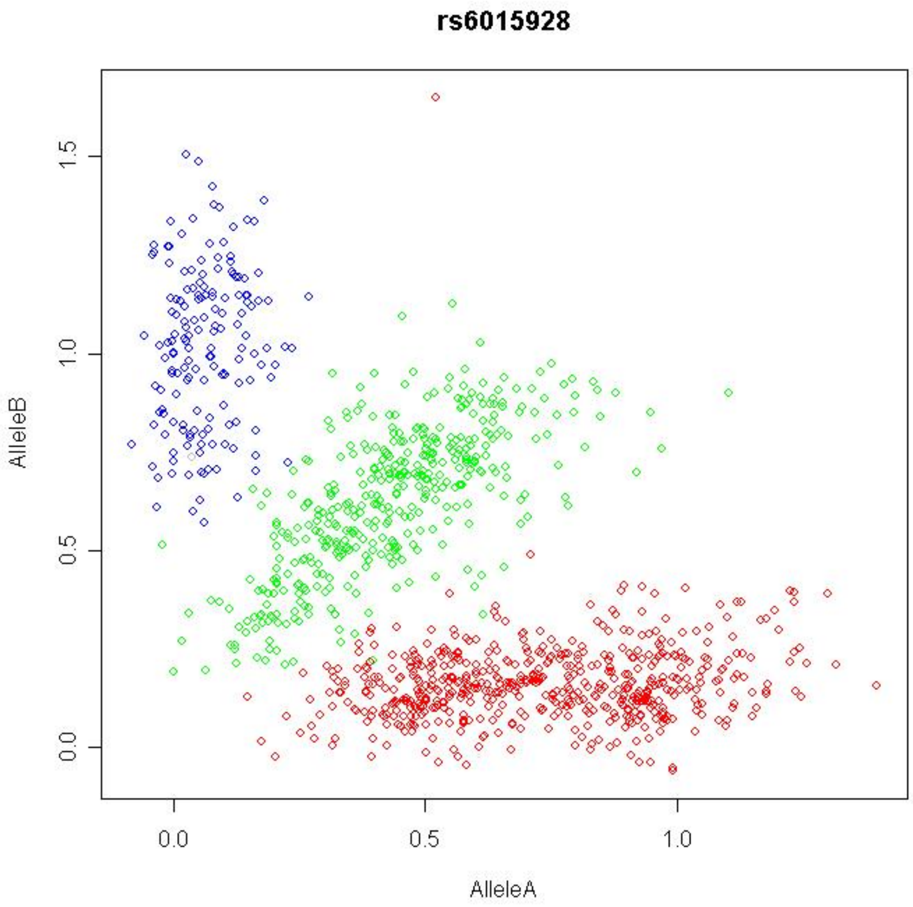
Probe Set

- Mapping 100K/500K:
  - 1 probe set: 40 probes (20 PM, 20 MM), 25 bp/each
- SNP Array 6.0:
  - 906,600 SNPs, 946,000 CNV probes
  - 1 SNP probe set: 6~8 probes (all PM), 25 bp/each
  - CNV probe (1 probe/probe set) : 202,000 probes targeting 5,677 known regions of copy number variation, 3,182 distinct, nonoverlapping segments, each interrogated with an average of 61 probes. In addition, more than 744,000 probes were chosen evenly spaced along the genome to find novel CNVs.

SNP Array Analysis

- Pre-processing
  - Normalization
  - Summarization
- SNP Genotyping
- CNV Inference
- LOH Inference

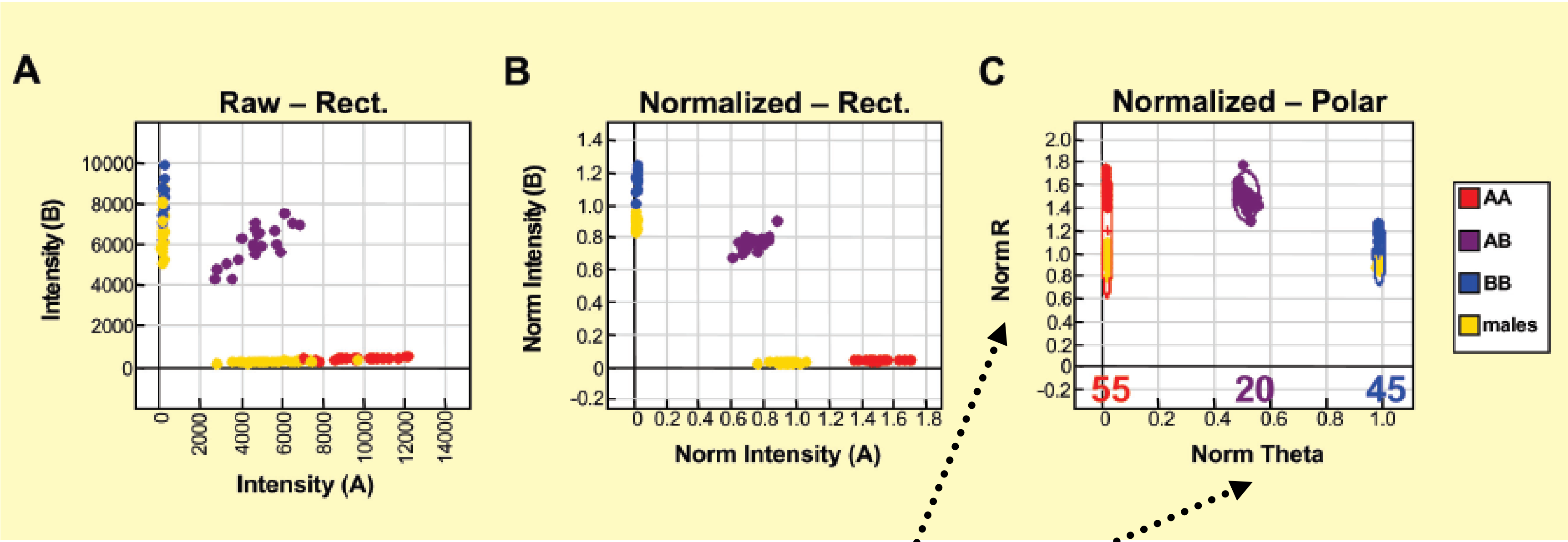
SNP Genotyping



CNV & LOH Inference Algorithm

- dChipSNP (Lin et al., *Bioinformatics 2004)*
- CNAT (Bignell et al., *Genome Research 2004)*
- GIM (Ishikawa et al., *Bioc. Biophys. Res. Comm. 2005)*
- CNAG (Nannya et al., *Cancer Research 2005)*
- PLASQ (LaFramboise et al., *PLoS Comp. Bio. 2005, Biostatistics 2007)*
- CARAT (Huang et al., *BMC Bioinformatics 2006)*
- PennCNV (Wang et al., Genome Research 2007)
- QuantiSNP (Colella et al., *Nucleic Acids Research 2007)*

CNV Inference

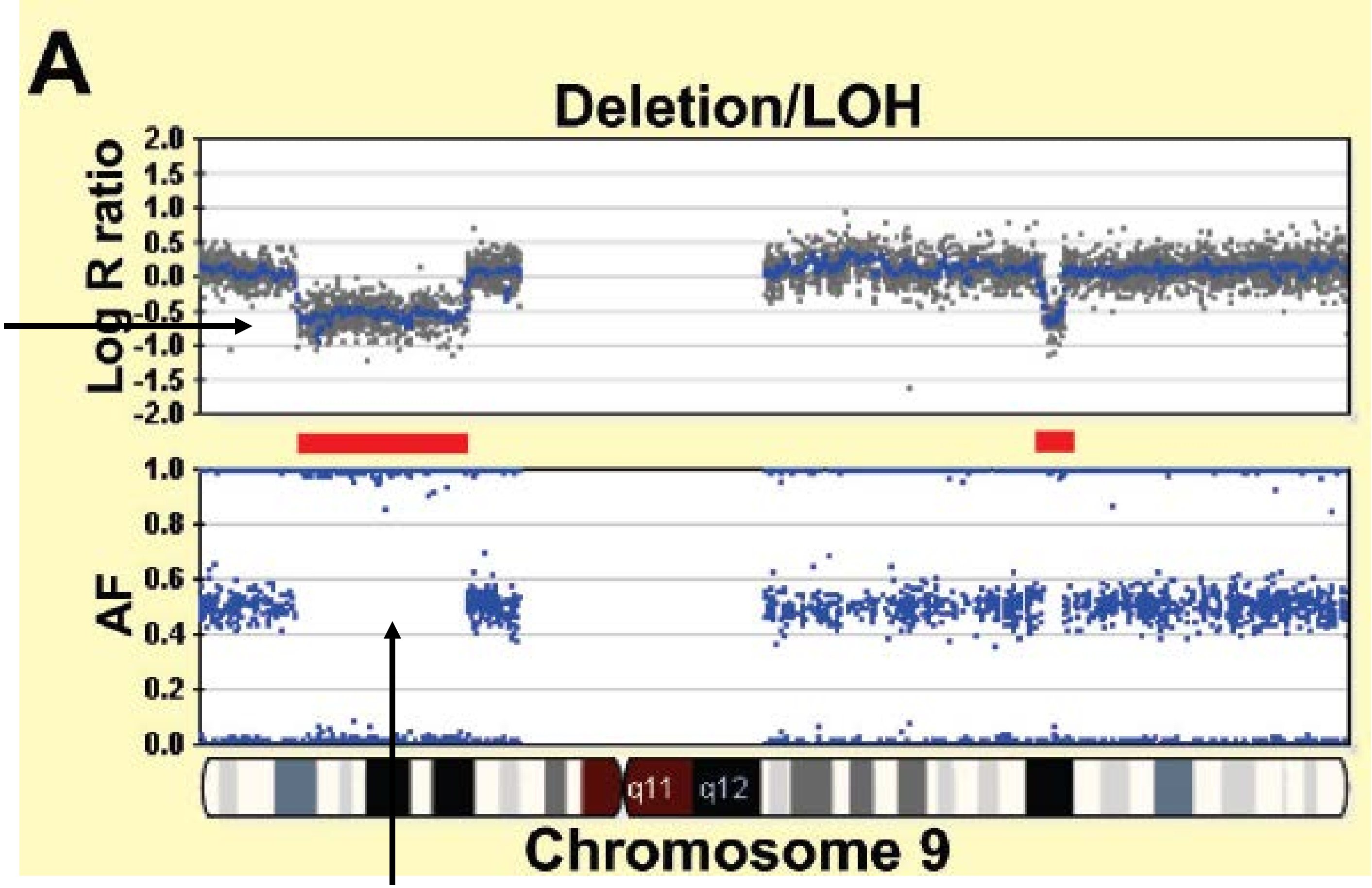


(A) Two-channel (two-allele) intensities (x and y)

(B) normalizing x,y with a reference value (based on ~100 controls, provided by the company)

(C) derive angle (theta) and radius (R) from x,y

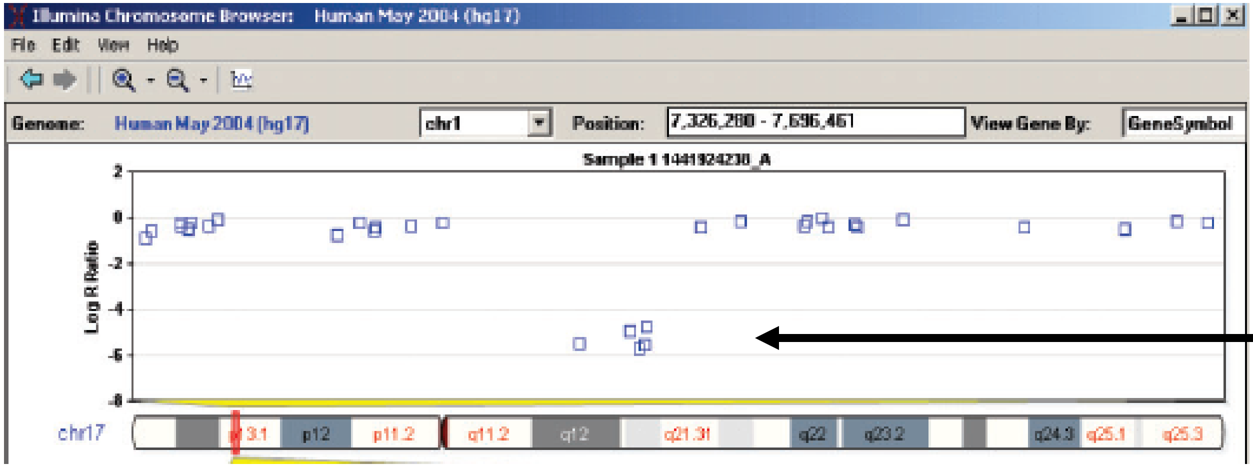
Hemizygous Deletion (CN=1)



Log(1/2)

No heterozygote (loss of heterozygosity)

Homozygous Deletion (CN=2)



Illumina Chromosome Browser: Human May 2004 (hg17)

Genome: Human May 2004 (hg17)

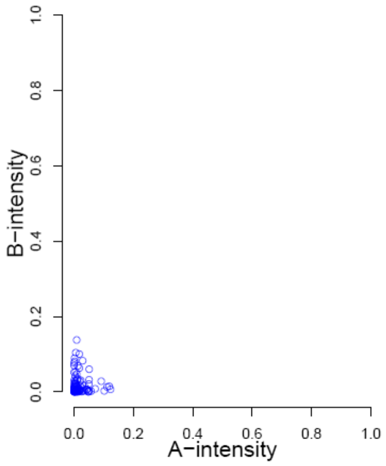
chr1

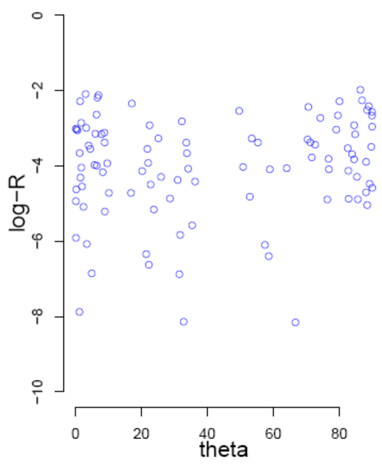
Position: 7,326,280 - 7,696,461

View Gene By: GeneSymbol

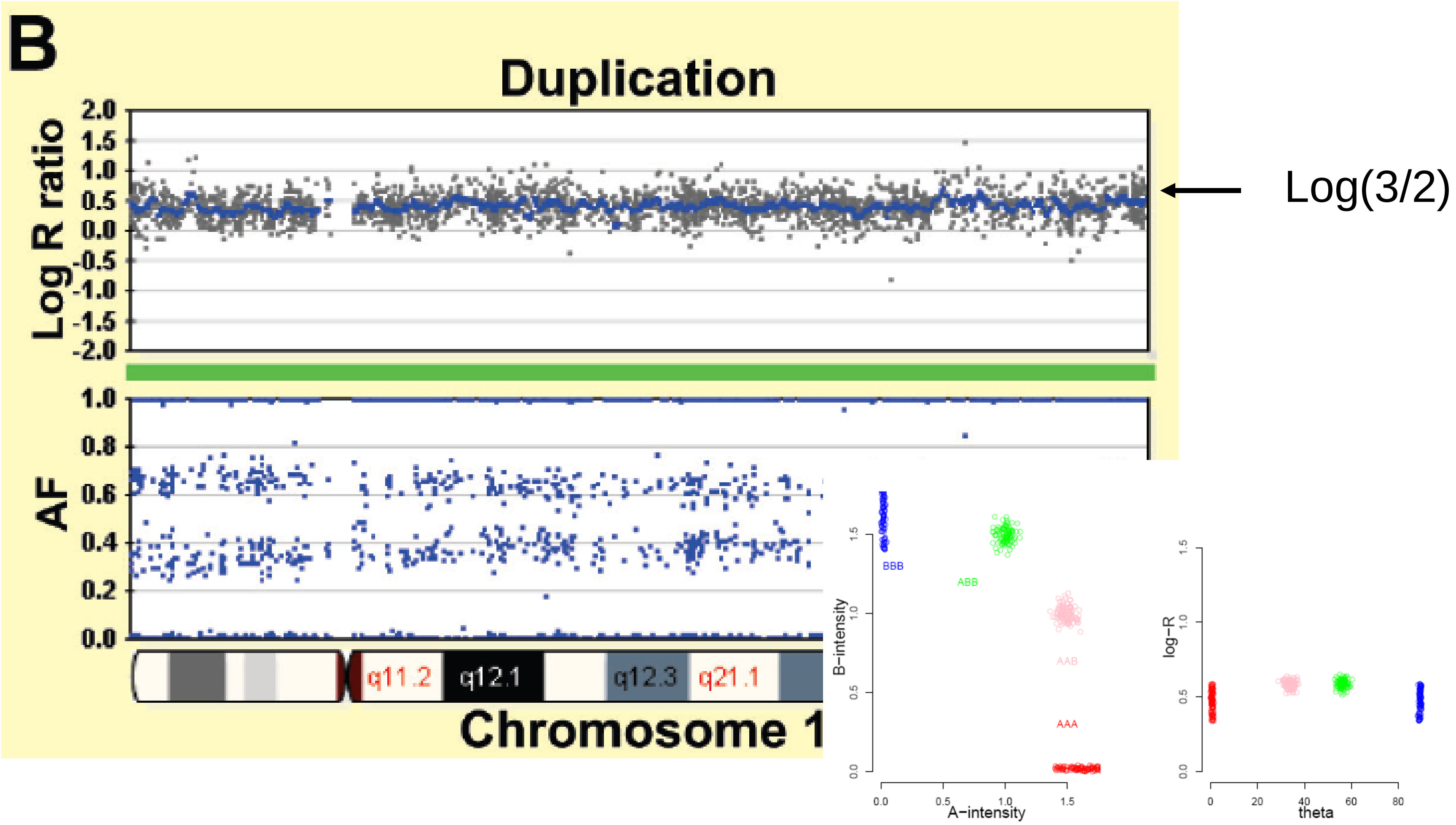
Sample 1 1441924238_A

Log(0/2)





Duplication (CN=3)

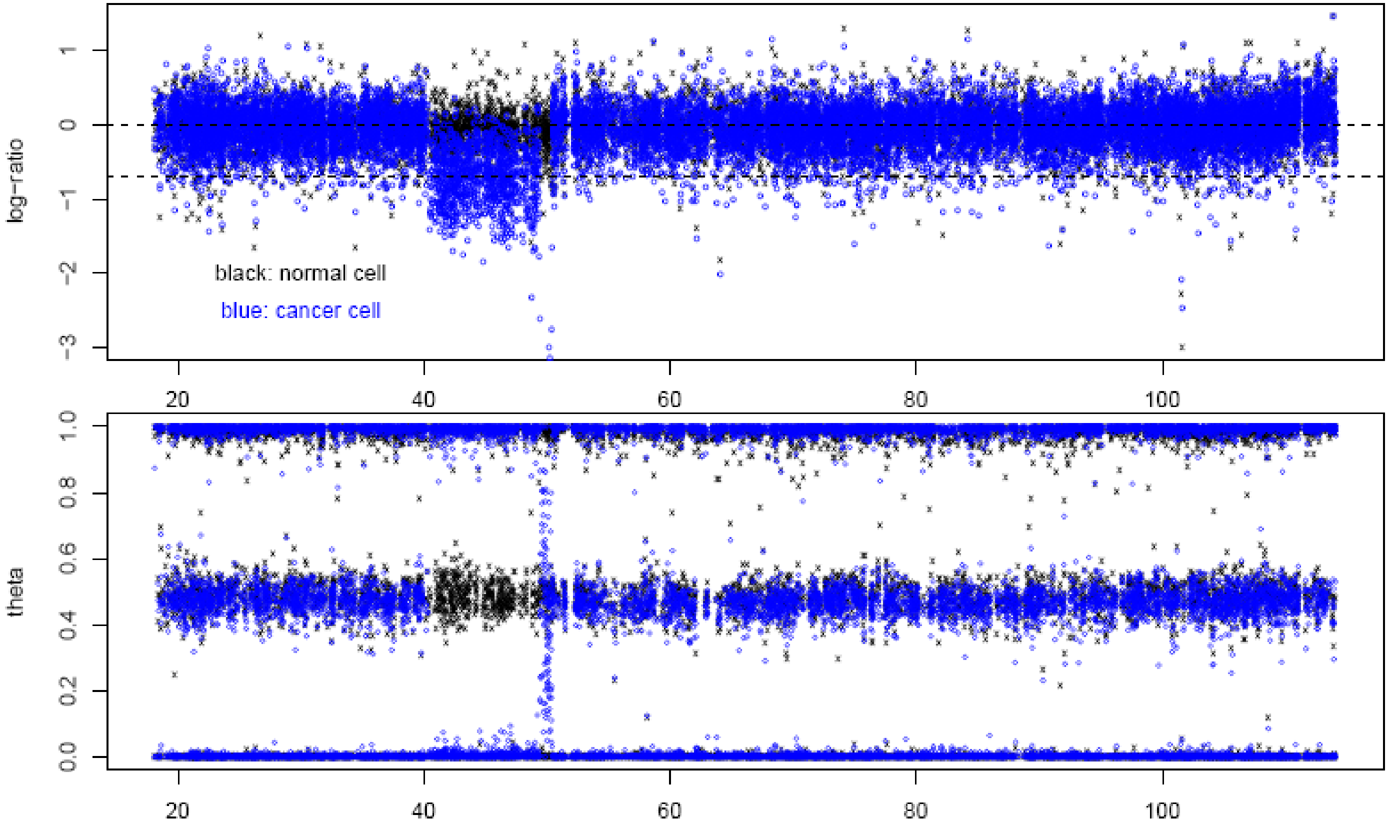


Log(3/2)

Delineate CNV Regions

- Eyeballing the theta and R-ratio plots (for large CNV regions)
- Cumulative plots
- Hidden Markov Model

CNV in Cancer Cell

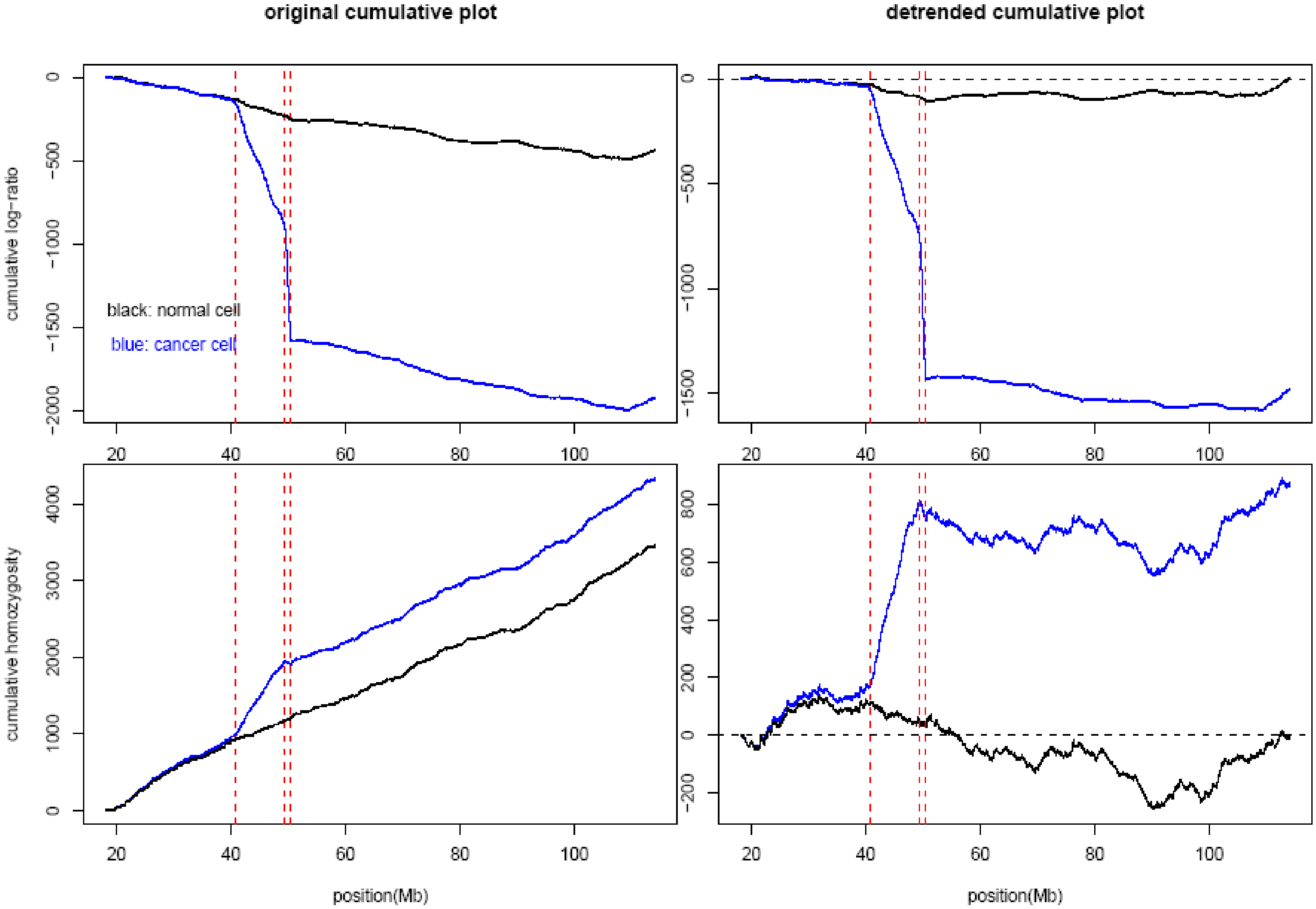


CNV in cancer cell: chronic lymphocytic leukemia (black: normal, blue: cancer cell) [ch13]

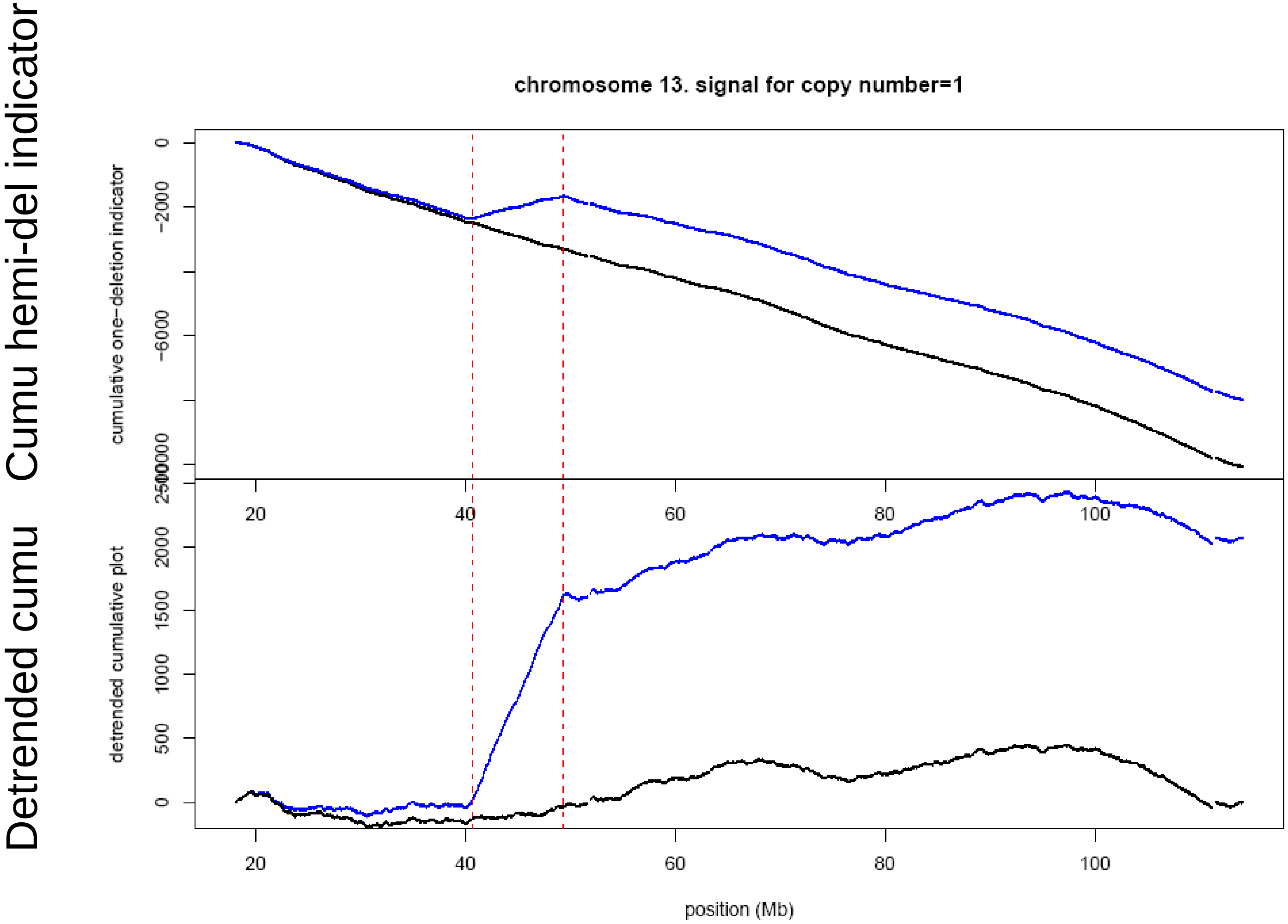
Cumulative Plots

Cumu Log (R-ratio)

Cumu homozygosity



Hemizygous Deletion Indicator

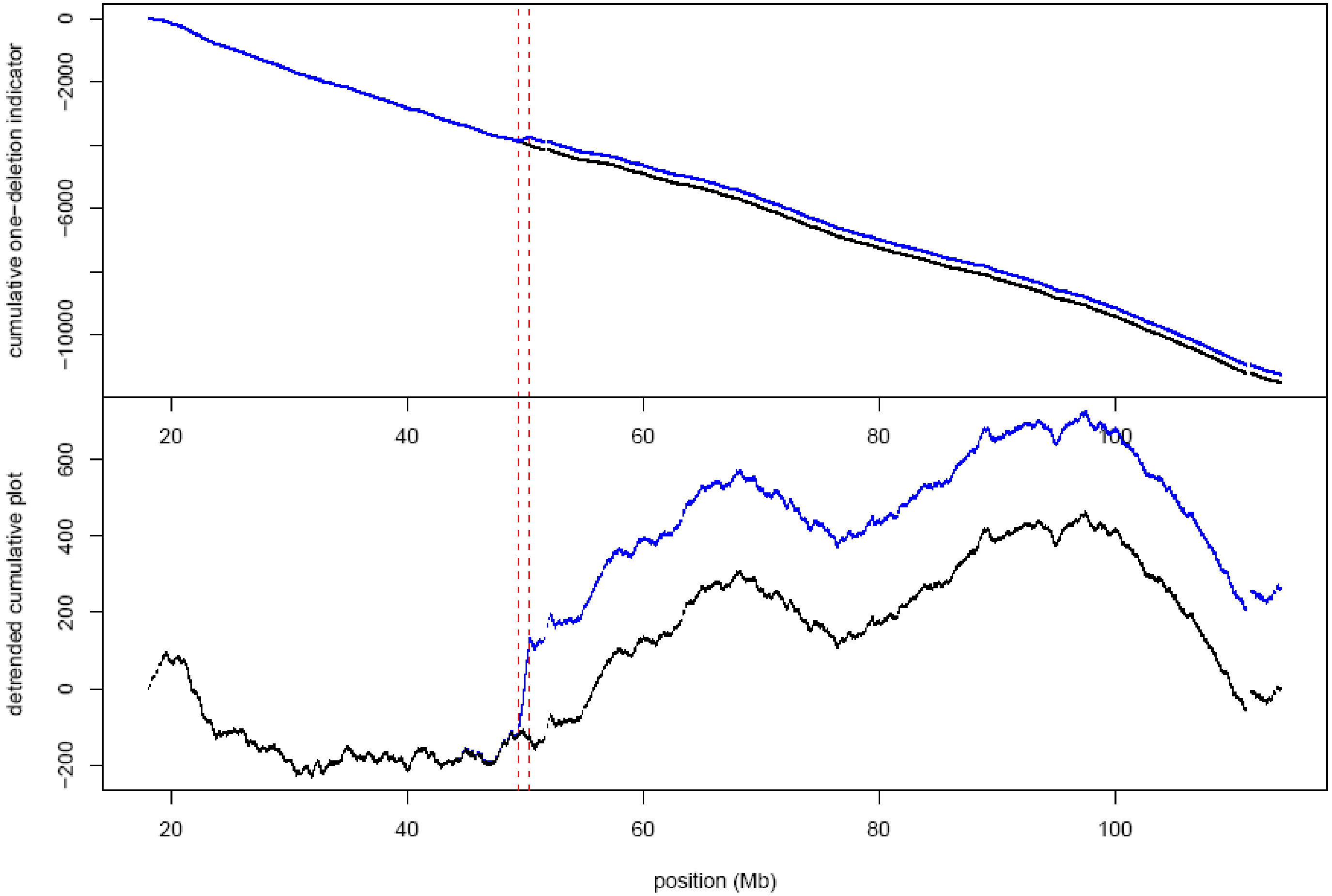


Hemizygous deletion indicator variable: 1 if logR is bw -2 and -0.346 AND homozygosity=1; -1 otherwise

Homozygous Deletion Indicator

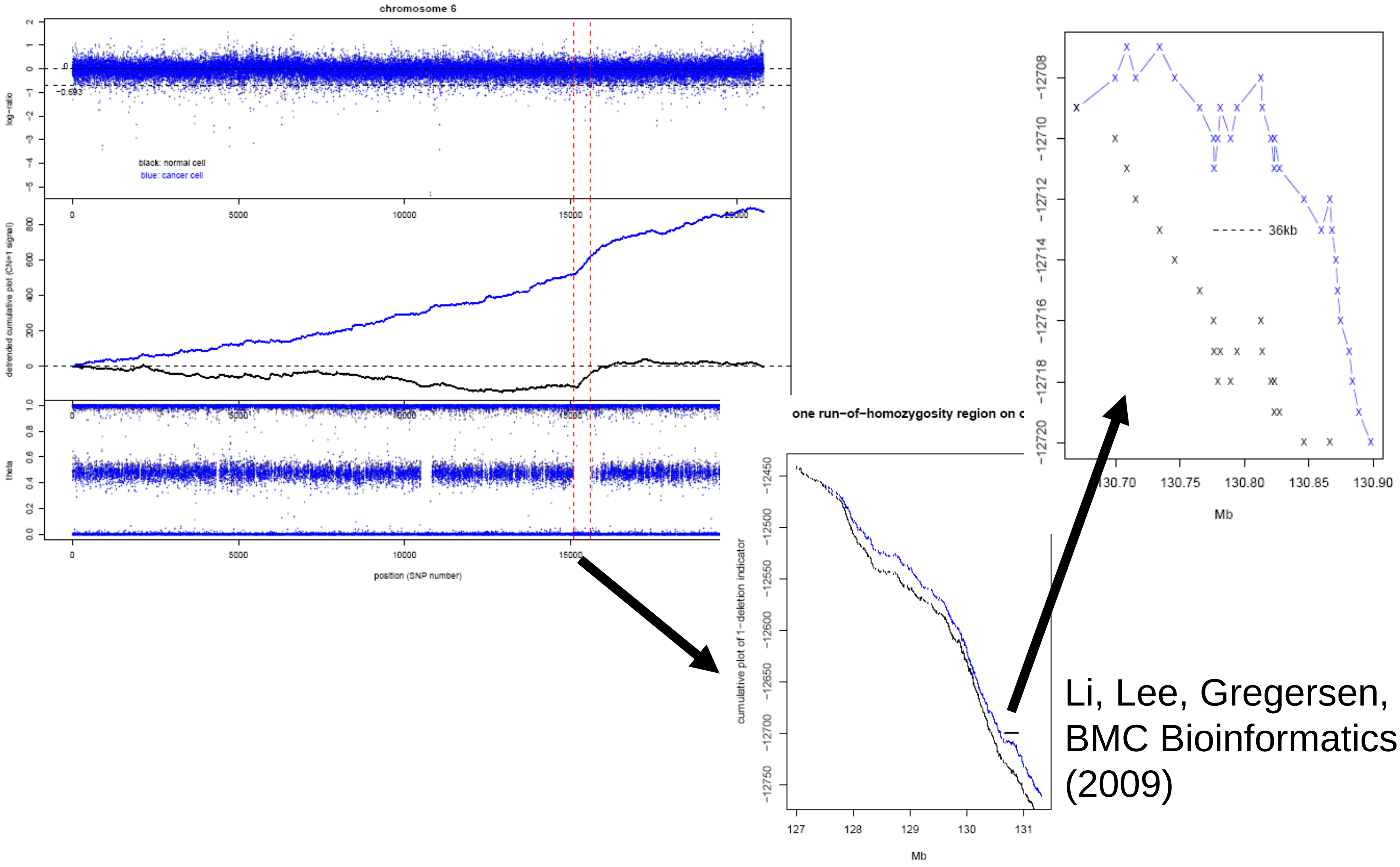
cumu homo-del indicator var

detrended



Homozygous deletion indicator variable: 1 if log(R-ratio) < -2; -1 otherwise

Zoom In of Smaller Regions



Li, Lee, Gregersen, BMC Bioinformatics (2009)

Improvement

- Consider the linkage between neighboring SNPs
- Adjusted cumulative plots based on Haldane's map

$$R = \frac{1 - exp(-2M)}{2}$$

$$\alpha = p_{same} / \overline{p}_{same} = e^{-2(M - \bar{M})}$$

PLASQ

- Generalized linear model based CNV detection algorithm

$$Y^{(ijk)} = \log(\gamma^{(j)}_{O_{jk}} + \alpha^{(j)}_{A_{jk}O_{jk}} C^{(ij)}_{A} + \beta^{(j)}_{B_{jk}O_{jk}} C^{(ij)}_{B}) + e_{ijk}$$

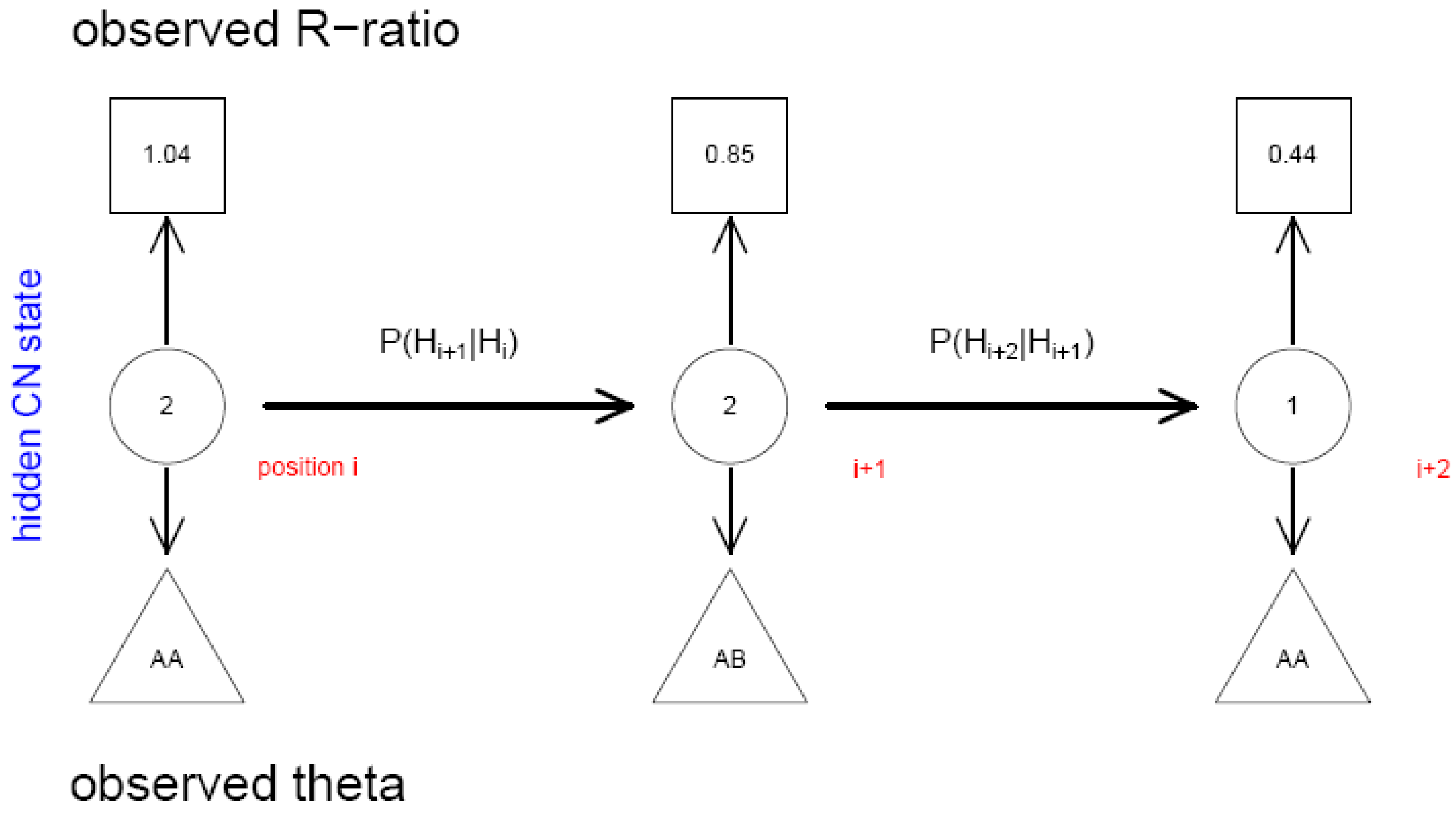
$Y^{(ijk)}$ = log probe intensity of probe *k* for SNP *j* in sample *i*

O_{jk} = F or R (orientation)

A_{jk} , B_{jk} = 0,1, or 2 from above

Parameters: $\gamma^{(j)}_{F}$, $\gamma^{(j)}_{R}$, $\alpha^{(j)}_{0F}$, $\alpha^{(j)}_{0R}$, $\alpha^{(j)}_{1F}$, $\alpha^{(j)}_{1R}$, $\beta^{(j)}_{0F}$, $\beta^{(j)}_{0R}$, $\beta^{(j)}_{1F}$, and $\beta^{(j)}_{1R}$

HMM Model of CNV



HMM Based CNV Software

- QuantiCNV
  - – http://www.well.ox.ac.uk/QuantiSNP/
- PennCNV
  - – http://www.neurogenome.org/cnv/penncnv/
- dChip
  - – http://biosun1.harvard.edu/complab/dchip/

PennCNV

- Hidden Markov Model designed for high resolution CNV detection in whole genome SNP genotyping data

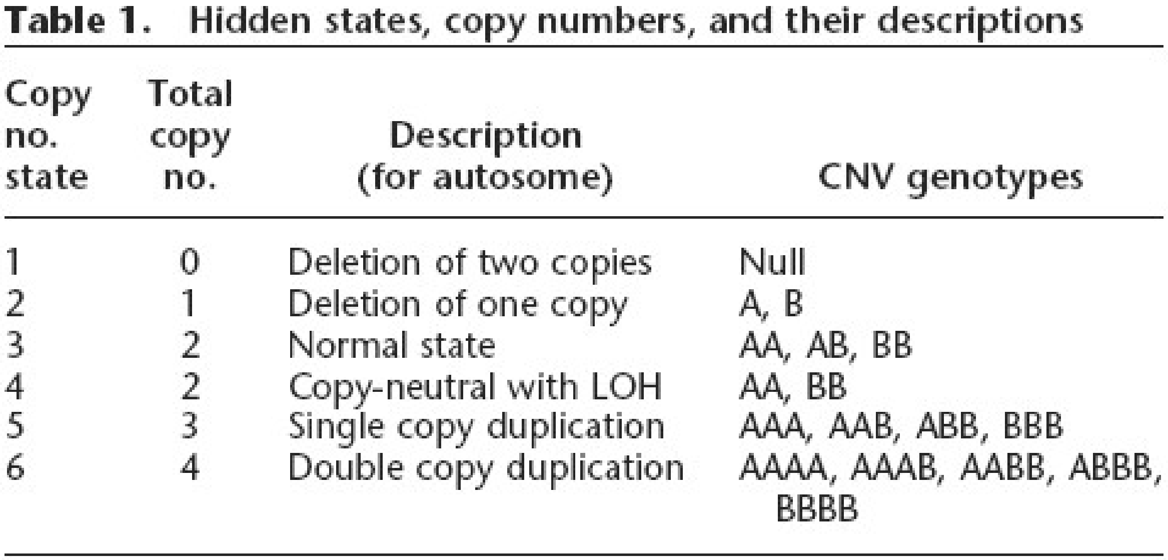


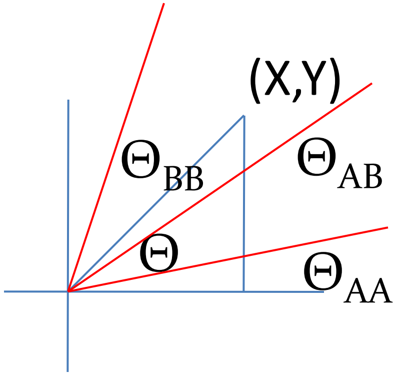
Table 1. Hidden states, copy numbers, and their descriptions

Copy no. state	Total copy no.	Description (for autosome)	CNV genotypes
1	0	Deletion of two copies	Null
2	1	Deletion of one copy	A, B
3	2	Normal state	AA, AB, BB
4	2	Copy-neutral with LOH	AA, BB
5	3	Single copy duplication	AAA, AAB, ABB, BBB
6	4	Double copy duplication	AAAA, AAAB, AABB, ABBB, BBBB

- **Log R ratio (LRR):** total fluorescent intensity signals from both sets of probe/allele at each SNP
- **B Allelle Frequence (BAF):** relative ratio of the intensity signals between two probes/allele at each SNP
- Accurate model for log R ratio and B Allele Frequency
- + Population allele frequency + distance between adjacent SNPs + family information

LRR and BAF

- X, Y : normalized signal intensity
- R = X+Y : total signal intensity
- Θ = arctan(Y/X)/(π/2)



$$\mathrm{LRR} = \log_2(\mathrm{R}_{\mathrm{observed}}/\mathrm{R}_{\mathrm{expected}})$$

$$\mathrm{BAF} = \begin{cases} 0,\ if\ \ \theta < \theta_{AA} \\ 0.5(\theta - \theta_{AA})/(\theta_{AB} - \theta_{AA}),\ if\ \ \theta_{AA} \le \theta < \theta_{AB} \\ 0.5 + 0.5(\theta - \theta_{AB})/(\theta_{BB} - \theta_{AB}),\ if\ \ \theta_{AB} \le \theta < \theta_{BB} \\ 1,\ if\ \ \theta \ge \theta_{BB} \end{cases}$$

HMM Model

- First order HMM assumes that the hidden copy number state at each SNP depends only the copy number state of the most preceding SNP.
- $\{r_i, b_i, z_i\}$: log R ratio, B allele Frequency, Copy number state at SNP *i* ($1 \le i < M$)

$$P(r_1,\ldots,r_M,b_1,\ldots,b_M) = \sum_{z1}\cdots\sum_{zM} P(r_1,\ldots,r_M,b_1,\ldots,b_M \mid z_1,\ldots,z_M)P(z_1,\ldots,z_M)$$

$$= \sum_{z1}\cdots\sum_{zM}\left\{\left(\prod_{i=1}^{M} P(r_i \mid z_i)P(b_i \mid z_i)\right)\left(P(z_1)\prod_{i=2}^{M} P(z_i \mid z_{i-1})\right.\right\}$$

Emission Probability

- Emission probability of log R ratio

$$P(r \mid z) = \pi_r + (1-\pi_r)\phi(r; \mu_{r,z}, s_{r,z})$$

- Emission probability of B allele Frequency

$$P(b \mid z) = \pi_b + (1-\pi_b)\sum_{g=2}^{K(z)-1} BN[g-1; K(z)-1, p_B]\varphi(b; \mu_{b,g}, s_{b,g})$$

$$+(1-\pi_b)BN[0; K(z)-1, p_B][I_{\{b=0\}}M_0 + I_{\{0<b<1\}}\varphi(b; \mu_{b,1}, s_{b,1})]$$

$$+(1-\pi_b)BN[K(z)-1; K(z)-1, p_B][I_{\{b=1\}}M_1 + I_{\{0<b<1\}}\varphi(b; \mu_{b,K(z)}, s_{b,K(z)})]$$

$$\text{where } BN[g-1; K(z)-1, p_B] = \binom{K(z)-1}{g-1} p_B{}^{g-1}(1-p_B)^{K(z)-g}$$

Transition Probability

- Probability of having a copy number state change between two adjacent SNPs.
- Intuition: The copy number state is unlikely to change for SNPs that are nearby but is more likely to change for SNPs that are far apart.

$$P(z_i = l \mid z_{i-1} = j) = \begin{cases} 1 - \sum_{k=2}^{6} P_{j,k-1}(1 - e^{-d_i/D}), \; if \;\; l = j \\ P_{j,l-1}(1 - e^{-d_i/D}), \; if \;\; l \neq j \end{cases}$$

- D is constant number. 100MB for state4 and 100KB for others
- Value p are treated as unknown parameter and estimated in the Baum-Welch algorithm

Model Training and CNV Calling

- Baum-Welch algorithm for training model to maximize the likelihood of the observed data of each individual
- Viterbi algorithm to infer most likely path.
- CNV is called most likely state sequence whenever a stretch of states that is different from normal state is observed.

PennCNV

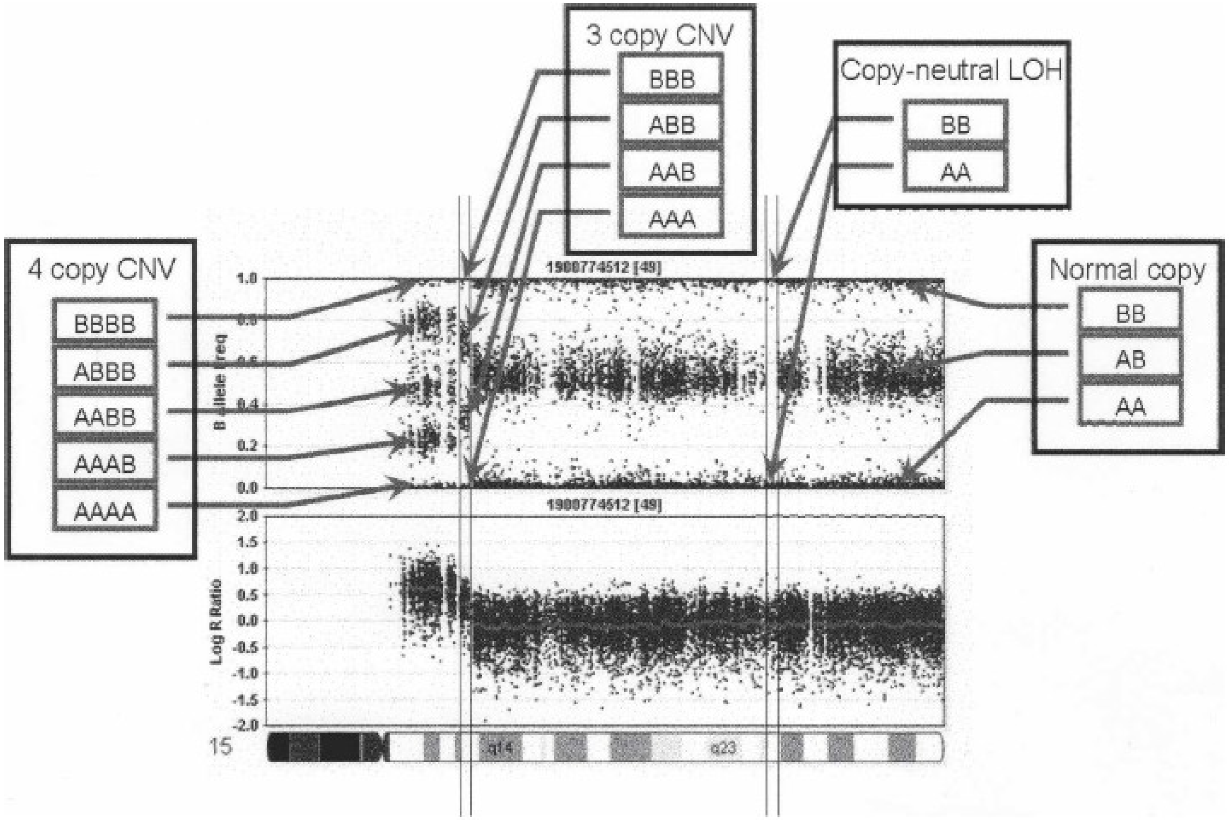


Figure 1. An illustration of log R Ratio (LRR) and B Allele Freq (BAF) values for the chromosome 15 q-arm of an individual. A normal chromosome region has three BAF genotype clusters, as represented as AA, AB, and BB genotypes in boxes, and with LRR values centered around zero. The copy-neutral LOH region has normal LRR values, but without the AB genotype cluster. The increased copy number for a CNV region can be detected based on an increased number of peaks in the BAF distribution, as well as increased LRR values. The patterns of LRR and BAF for different CNV regions, normal regions, and copy-neutral LOH regions are distinct from each other, thus the combination of LRR and BAF can be used to generate CNV calls.

Two Ways for LOH Inference

- Unpaired samples
  - Use only the tumor samples
  - LOH is inferred from the decreased heterozygous rate in certain regions of the tumor samples
- Paired samples
  - Use both tumor and normal samples from the same individual
  - LOH is inferred by comparing the genotypes of the tumor sample and its normal counterpart

Single Loci LOH

Genotypes		Tumor			
		A	H	B	NoCall
Normal	A	No-info	Mutation	Mutation	No-info
	H	LOH	RET	LOH	No-info
	B	Mutation	Mutation	No-info	No-info
	NoCall	No-info	RET	No-info	No-info

- LOH: Loss of Heterozygosity
- RET: Retention
- No-info: Non-informative

Example of LOH

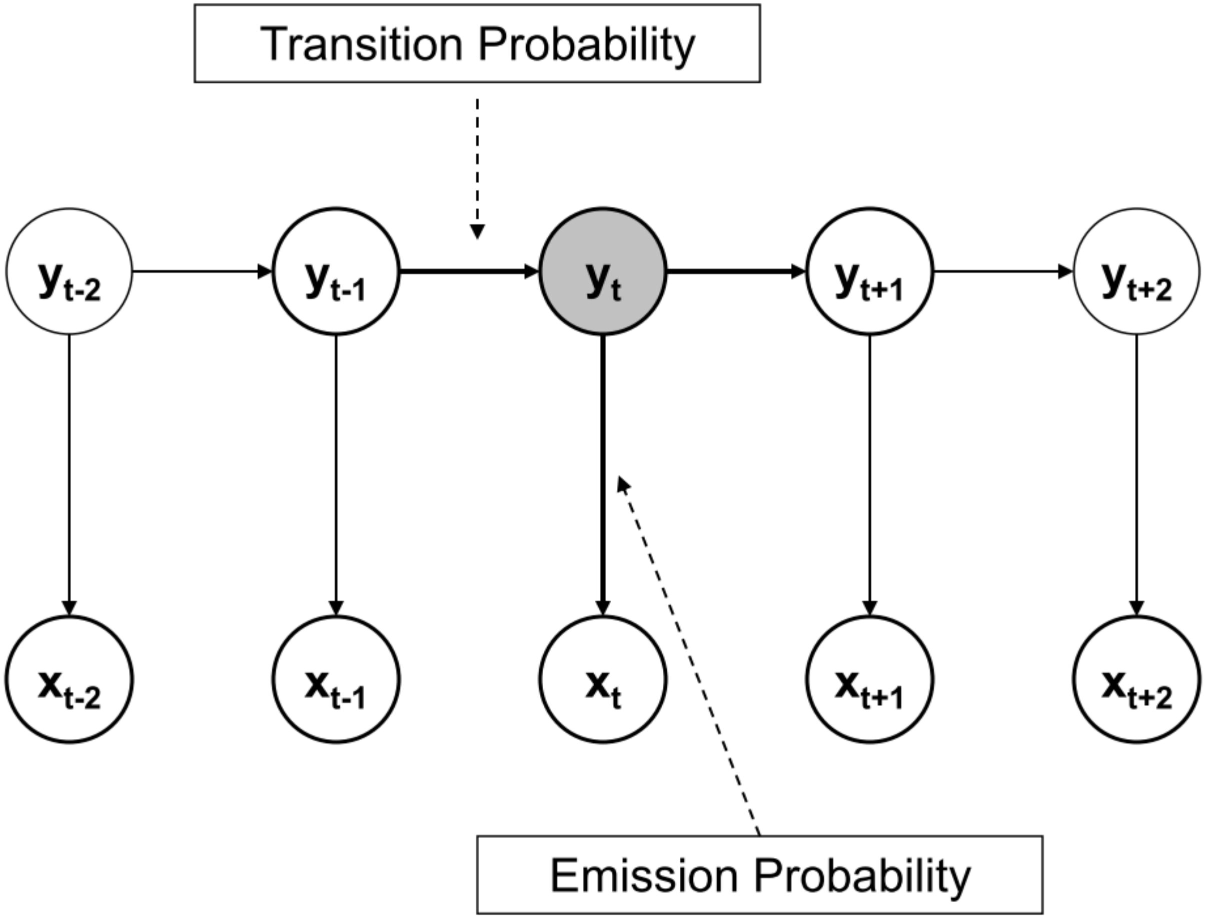
Ch 1:	A	B	B	A	B	A	A	A	A
Ch 2:	B	B	B	A	A	A	A	B	B
Genotypes:	**H**	**B**	**B**	**A**	**H**	**A**	**A**	**H**	**H**

Ch 1:	A	B	B	A	B	A	A	A	A
Ch 2:	B	B							B
Genotypes:	**H**	**B**	**B**	**A**	**B**	**A**	**A**	**A**	**H**

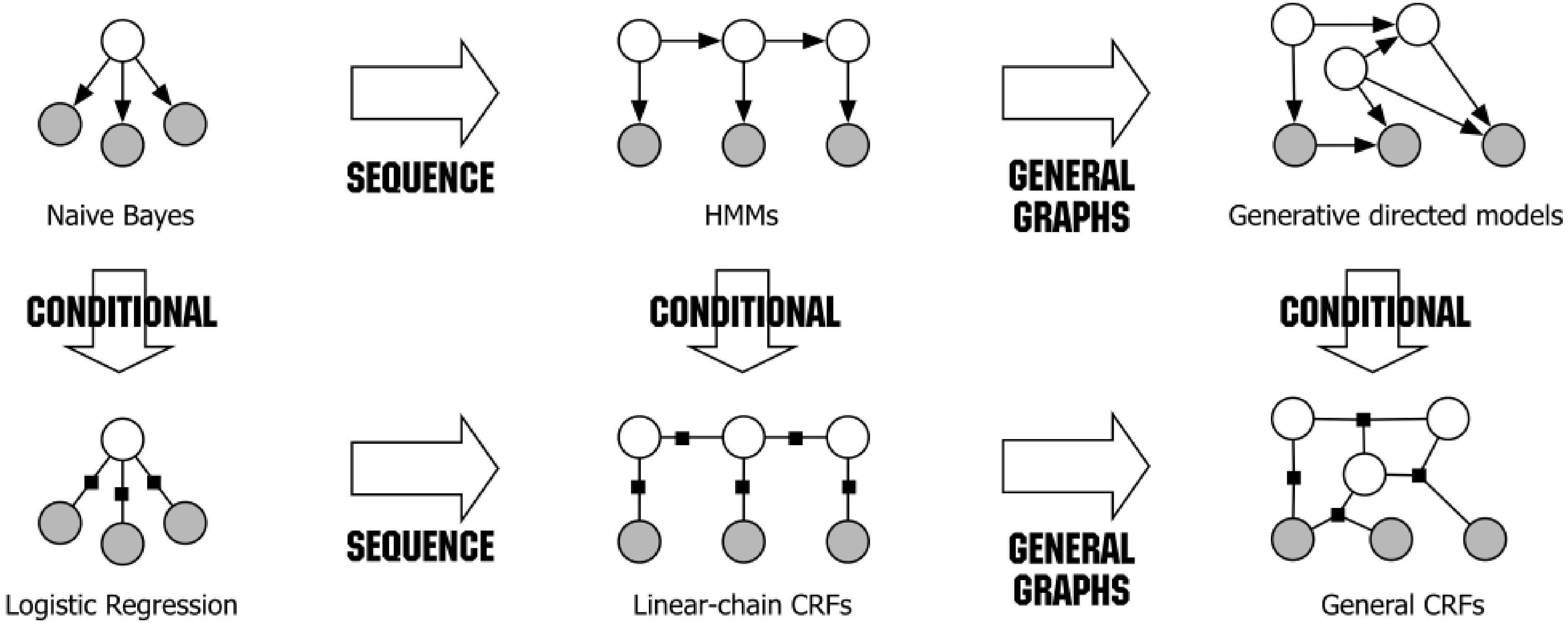
Motivation

- Difficulties
  - Genotyping errors
  - Non-informative SNPs
- Motivation
  - Two SNPs that are close in chromesome tend to be in same status
  - Borrow the information from neighboring SNPs to reduce the false positive

HMM Approach

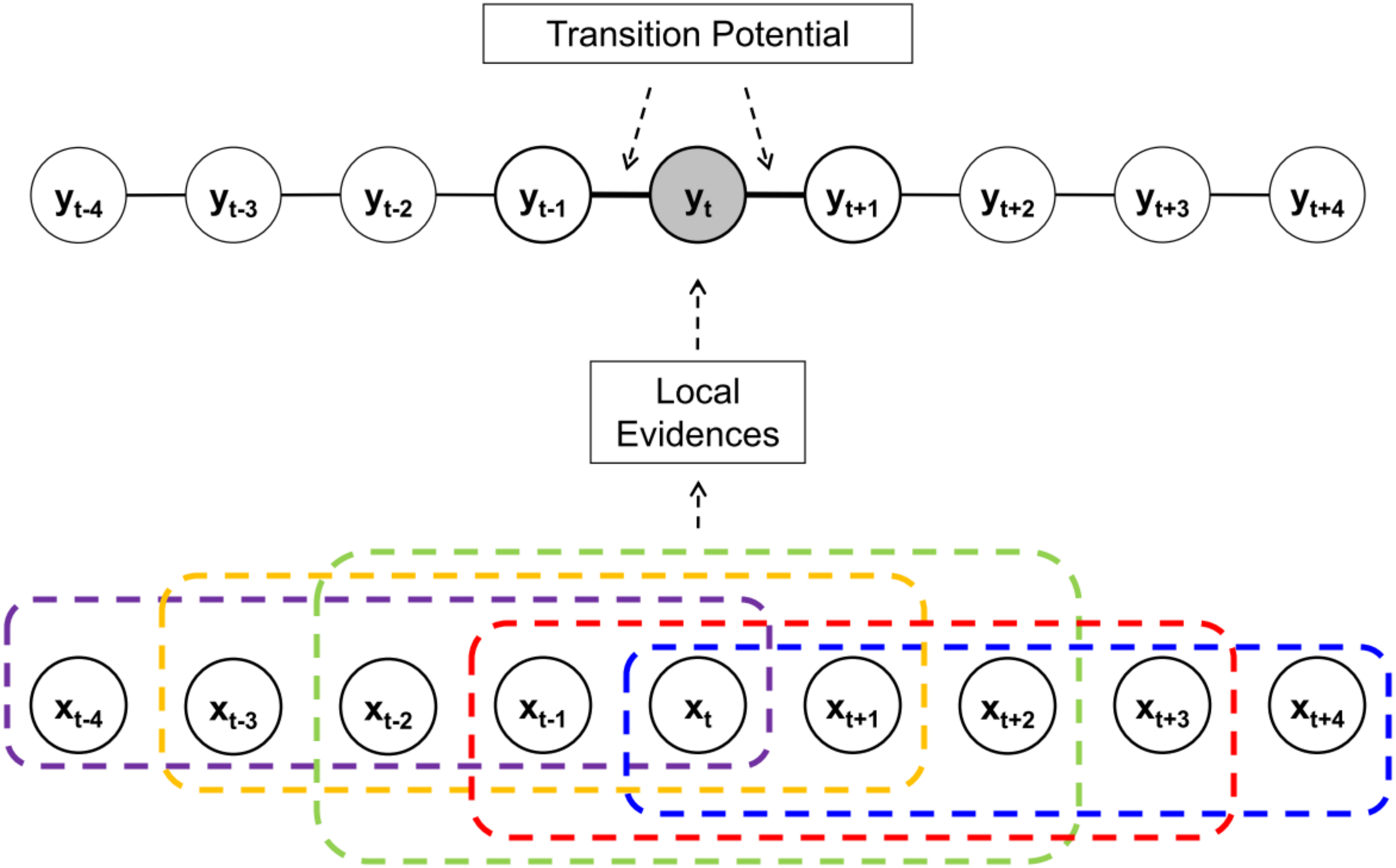


Conditional Random Fields



An Introduction to Conditional Random Fields for Relational Learning.
Charles Sutton, Andrew McCallum. In Lise Getoor and Ben Taskar, editors. *Introduction to Statistical Relational Learning*. MIT Press. 2007.

CRF Model for LOH Inference



Conditional Probability

$$p(y \mid x) = \frac{e^{\psi(y,x)}}{\sum_{z \in S} e^{\psi(z,x)}}$$

$$\psi(y,x) = \sum_{t=1}^{T-1} f_{TP}(y_t, y_{t+1}) + \sum_{t=1}^{T} f_{LE}(y_t, x)$$

Potential Functions (1)

- Transition Potentials

$$f_{TP}(y_t, y_{t+1}) = \begin{cases} (1-\theta)+\theta\rho & y_t = y_{t+1} = \mathrm{LOSS}, \\ (1-\theta)+\theta(1-\rho) & y_t = y_{t+1} = \mathrm{RET}, \\ \theta(1-\rho) & y_t = \mathrm{LOSS}, y_{t+1} = \mathrm{RET} \\ \theta\rho & y_t = \mathrm{RET}, y_{t+1} = \mathrm{LOSS} \end{cases}$$

- where $\theta = 1 - e^{-2d/\beta}$ is the probability that two neighboring SNPs are independent
  - d is the distance between two SNPs, beta is the transition decay parameter, rho is the estimated LOH rate

Potential Functions (2)

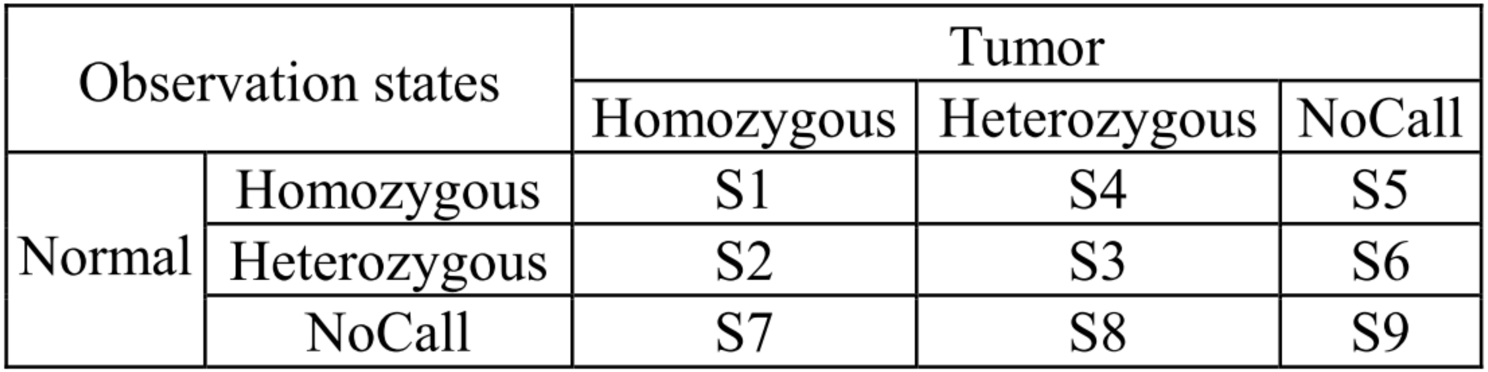
- Emission Potentials

$$f_{LE}(y_t, x) = \max_{i=1}^{K} \left\{ \left(\prod_{j=1}^{K} p(x_{t-i+j} \mid y_t) \right)^{1/K} \right\}$$

- where $p(x_j|y_t)$ is the emission probability that we observe the x_j at locus j while the hidden state in locus j is y_t

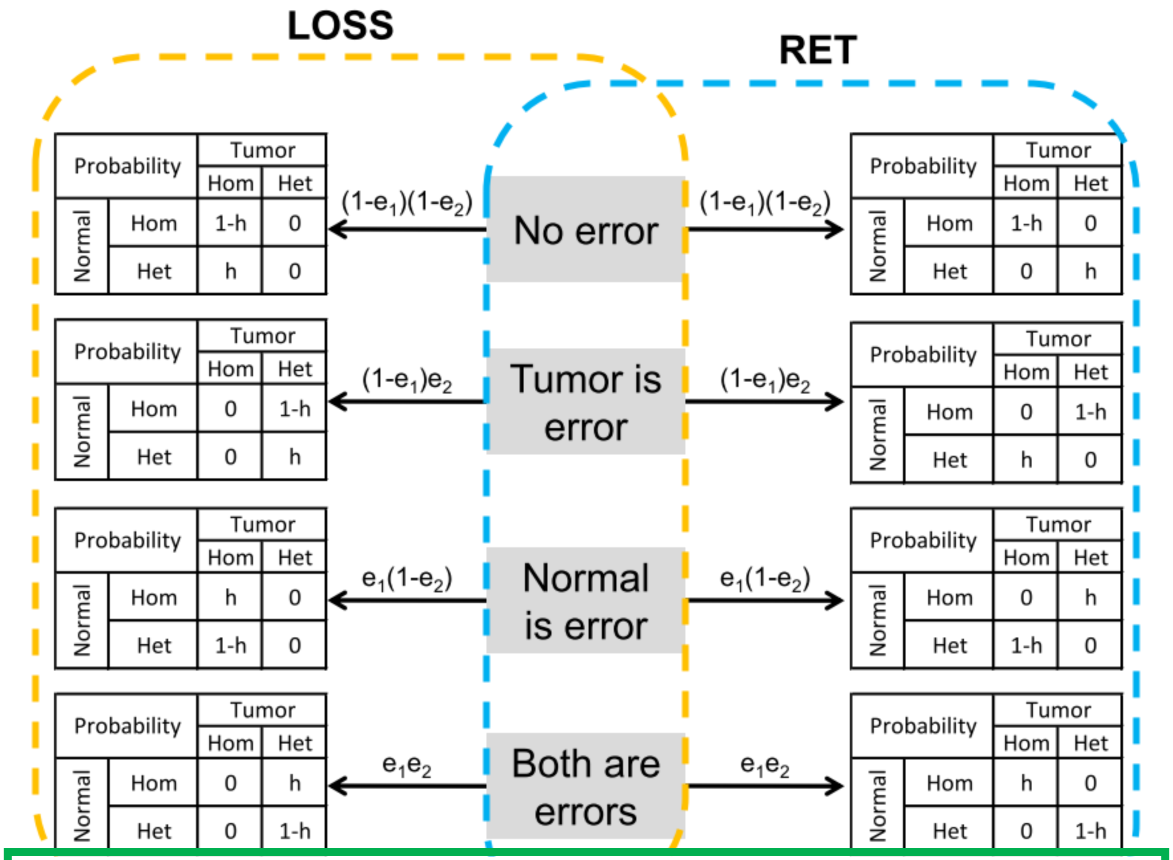
Hidden States and Observations

- Hidden States:
  - – LOSS (Loss of Heterozygosity)
  - – RET (Retention)
- Observation States:



Observation states		Tumor		
		Homozygous	Heterozygous	NoCall
Normal	Homozygous	S1	S4	S5
	Heterozygous	S2	S3	S6
	NoCall	S7	S8	S9

Emission Probability Model



LOSS

No error: $(1-e_1)(1-e_2)$

Probability		Tumor	
		Hom	Het
Normal	Hom	1-h	0
	Het	h	0

Tumor is error: $(1-e_1)e_2$

Probability		Tumor	
		Hom	Het
Normal	Hom	0	1-h
	Het	0	h

Normal is error: $e_1(1-e_2)$

Probability		Tumor	
		Hom	Het
Normal	Hom	h	0
	Het	1-h	0

Both are errors: e_1e_2

Probability		Tumor	
		Hom	Het
Normal	Hom	0	h
	Het	0	1-h

RET

No error: $(1-e_1)(1-e_2)$

Probability		Tumor	
		Hom	Het
Normal	Hom	1-h	0
	Het	0	h

Tumor is error: $(1-e_1)e_2$

Probability		Tumor	
		Hom	Het
Normal	Hom	0	1-h
	Het	h	0

Normal is error: $e_1(1-e_2)$

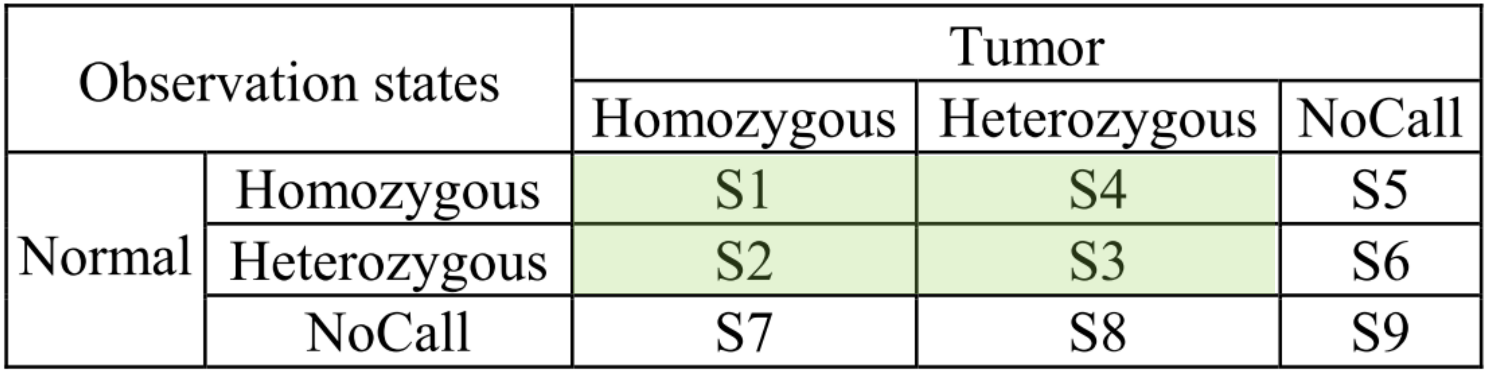
Probability		Tumor	
		Hom	Het
Normal	Hom	0	h
	Het	1-h	0

Both are errors: e_1e_2

Probability		Tumor	
		Hom	Het
Normal	Hom	h	0
	Het	0	1-h

$$p(S1 \mid LOSS) = (1-e_1)(1-e_2) \times (1-h) + (1-e_1)e_2 \times 0 + e_1(1-e_2) \times h + e_1e_2 \times 0$$

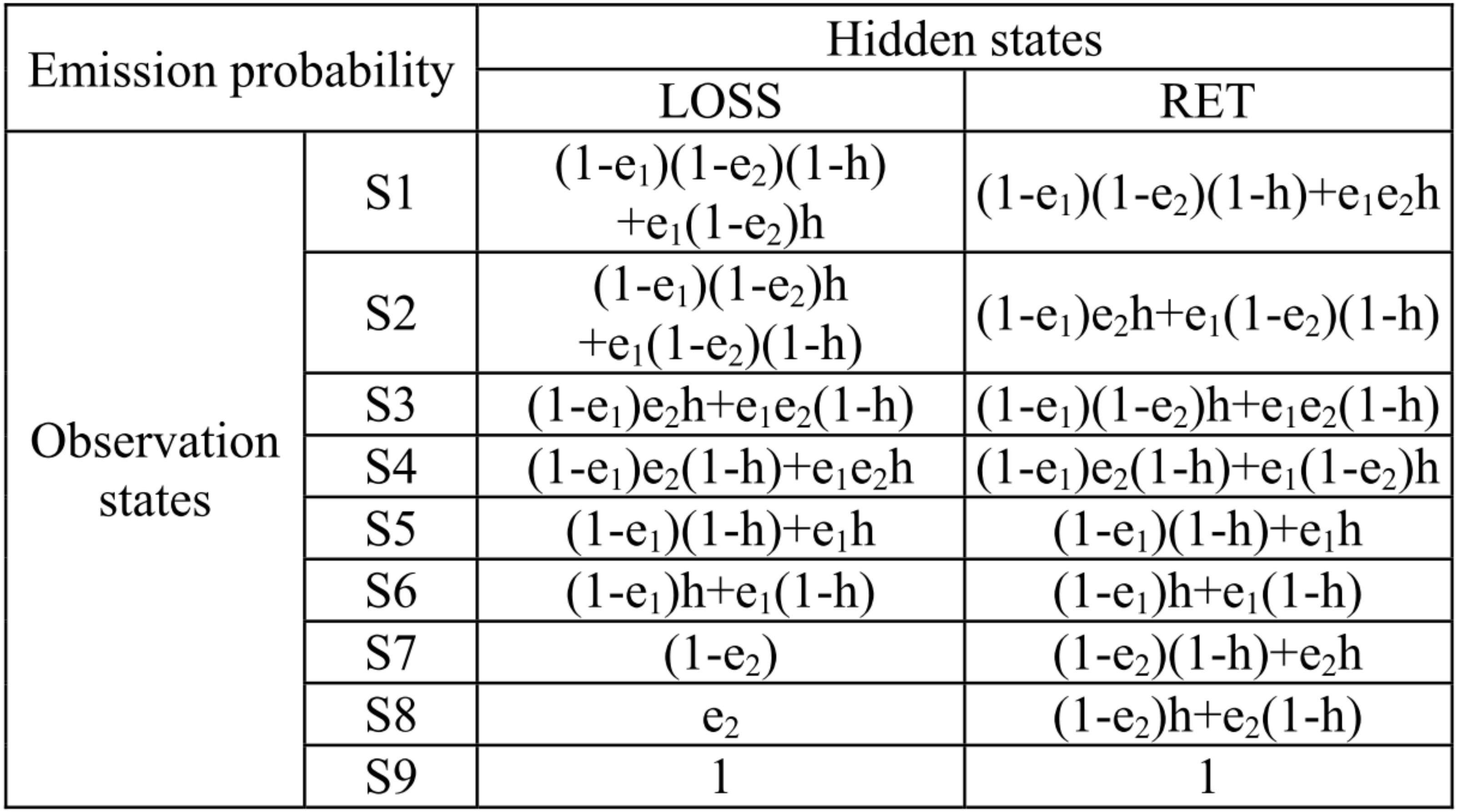
Emission Probability



Observation states		Tumor		
		Homozygous	Heterozygous	NoCall
Normal	Homozygous	S1	S4	S5
	Heterozygous	S2	S3	S6
	NoCall	S7	S8	S9

$$p(S7 \mid LOSS) = p(S1 \mid LOSS) + p(S2 \mid LOSS)$$

Emission Probability



Emission probability		Hidden states	
		LOSS	RET
Observation states	S1	$(1-e_1)(1-e_2)(1-h)$ $+e_1(1-e_2)h$	$(1-e_1)(1-e_2)(1-h)+e_1e_2h$
	S2	$(1-e_1)(1-e_2)h$ $+e_1(1-e_2)(1-h)$	$(1-e_1)e_2h+e_1(1-e_2)(1-h)$
	S3	$(1-e_1)e_2h+e_1e_2(1-h)$	$(1-e_1)(1-e_2)h+e_1e_2(1-h)$
	S4	$(1-e_1)e_2(1-h)+e_1e_2h$	$(1-e_1)e_2(1-h)+e_1(1-e_2)h$
	S5	$(1-e_1)(1-h)+e_1h$	$(1-e_1)(1-h)+e_1h$
	S6	$(1-e_1)h+e_1(1-h)$	$(1-e_1)h+e_1(1-h)$
	S7	$(1-e_2)$	$(1-e_2)(1-h)+e_2h$
	S8	e_2	$(1-e_2)h+e_2(1-h)$
	S9	1	1

LOH Inference

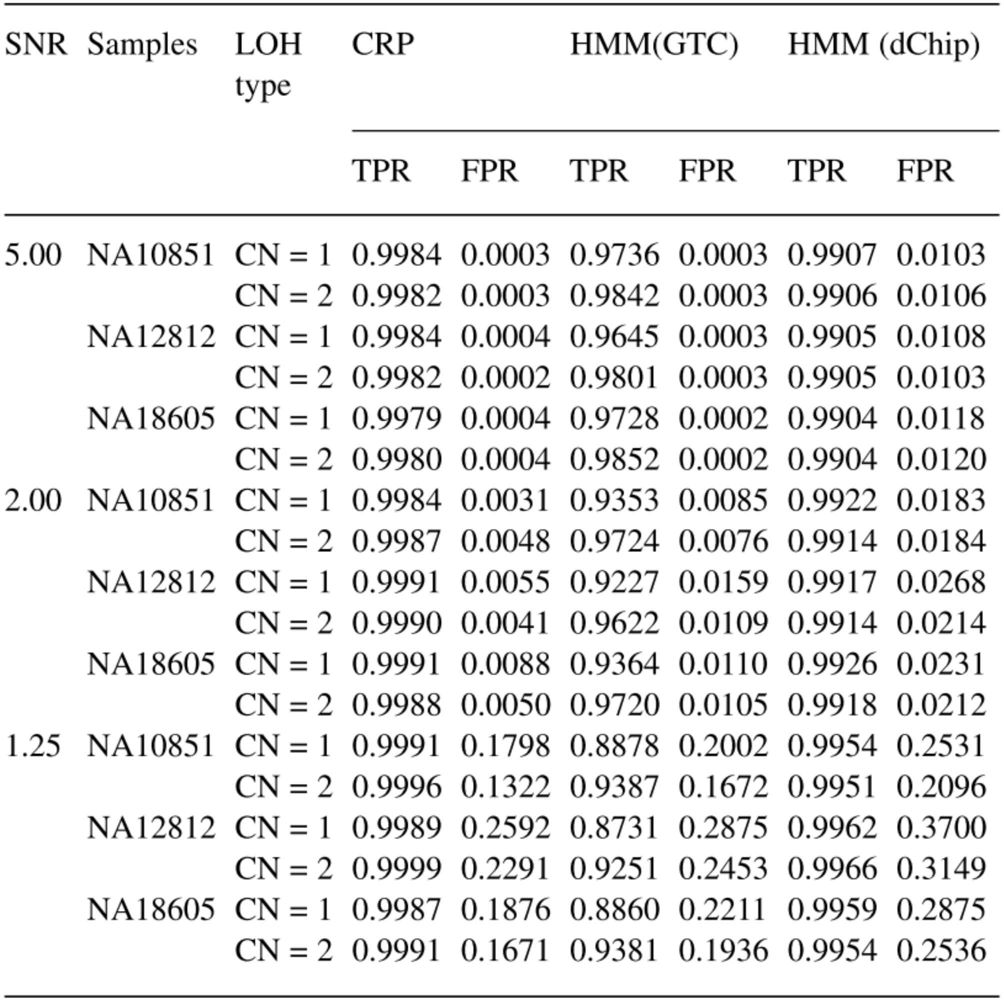
- Given observation sequence x, the hidden LOH status are inferred as:

$$\hat{y} = \arg\max_{y} p(y \mid x)$$

Simulated Data

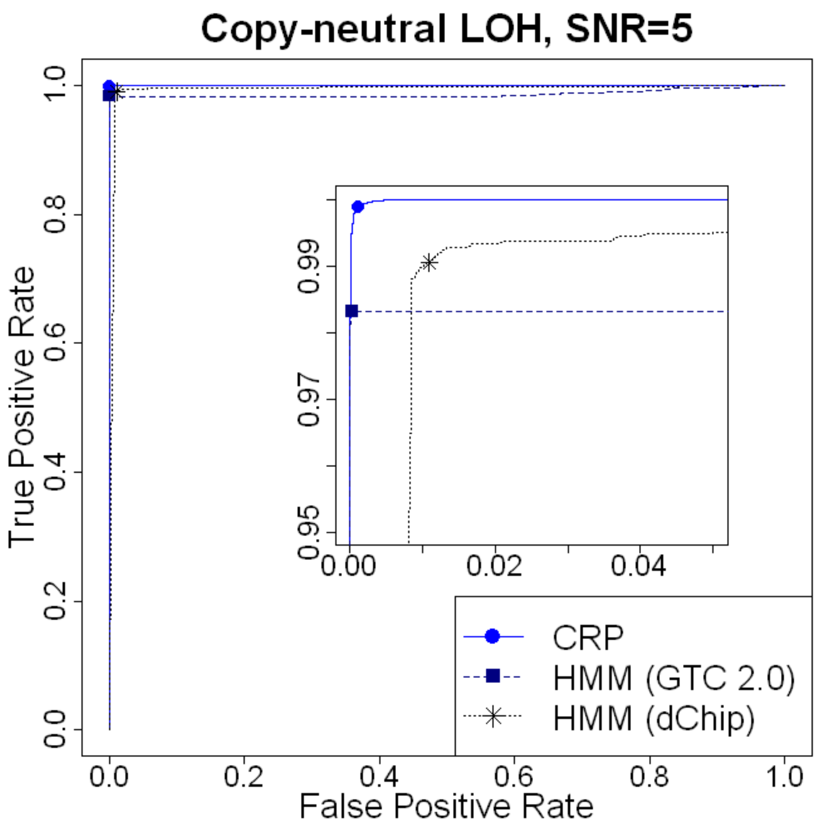
- Based on real Affymetrix’s 500K SNP arrays of HapMap samples
- Simulate LOH in the raw intensity level
  - Two types of LOH: copy-neutral and copy-less
  - Three levels of noise (error) following normal distribution: 20%, 50%, and 80% noise
    - SNR (signal to noise ratio) = 5, 2, and 1.25
- Process the simulated SNP arrays by Affymetrix’s official genotyping software

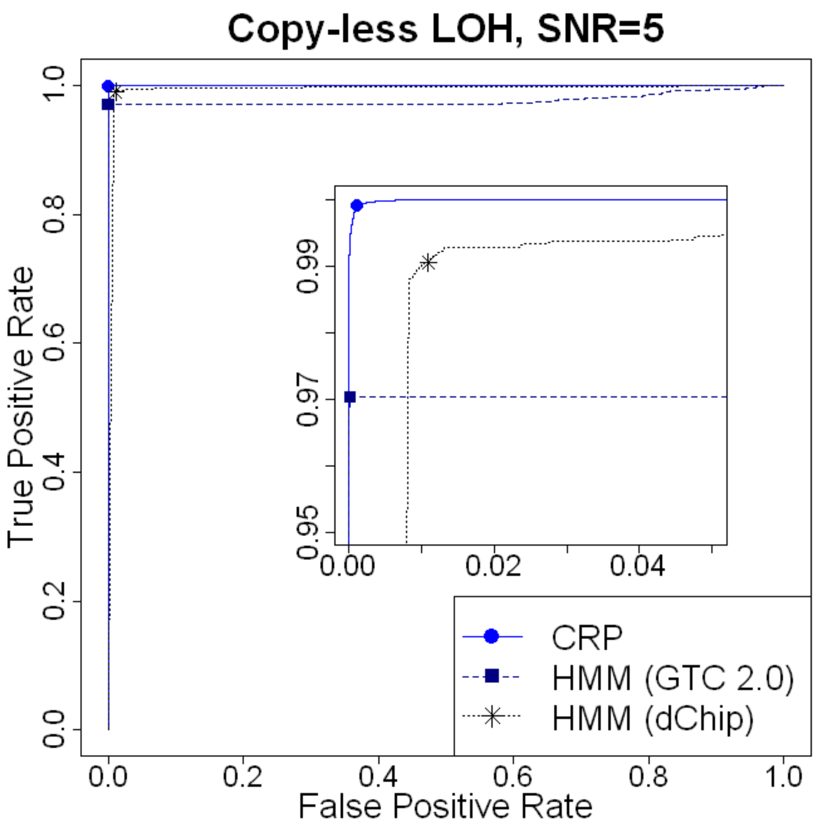
Informative SNPs



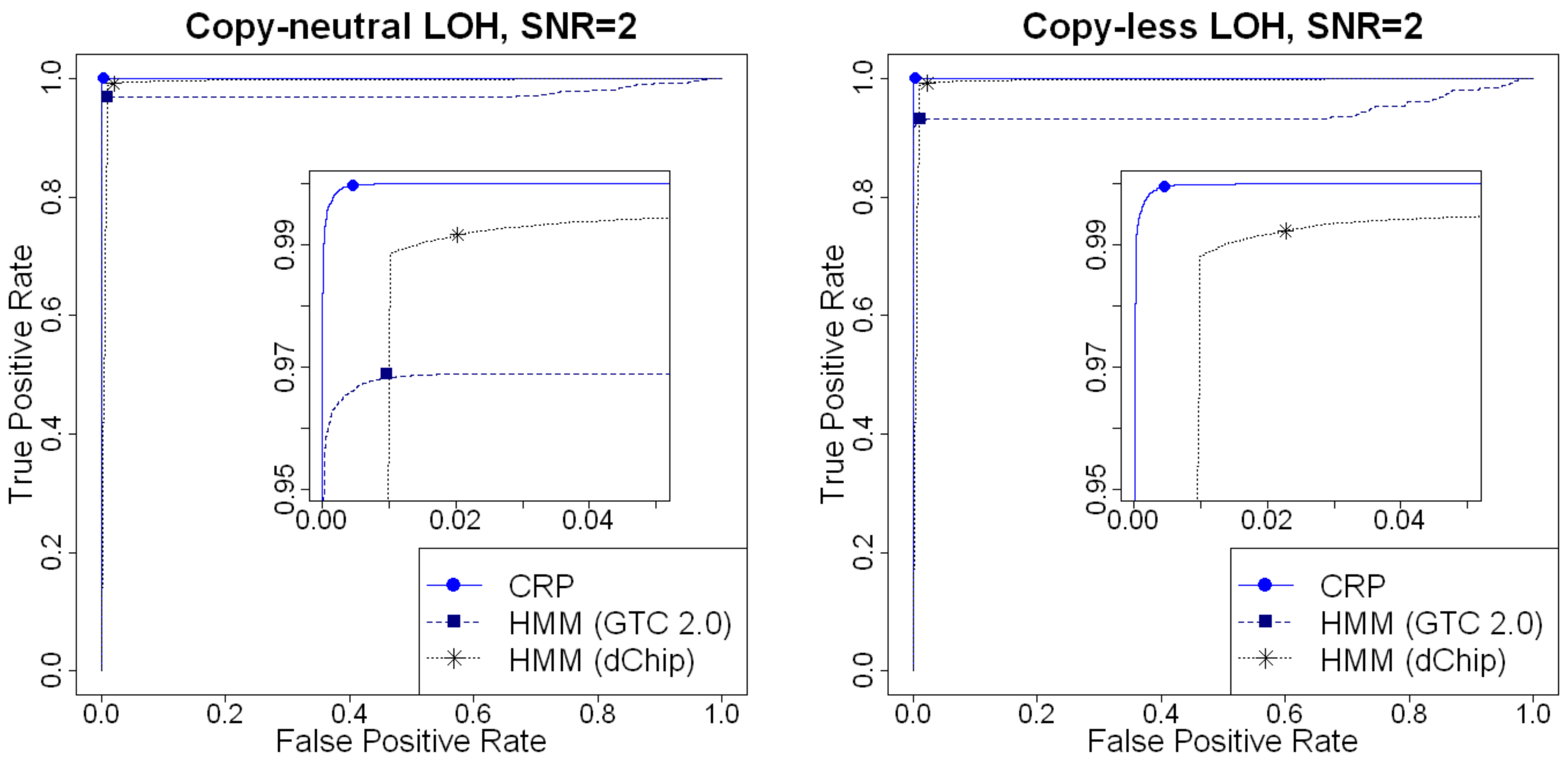
SNR	Samples	LOH type	CRP		HMM(GTC)		HMM (dChip)	
			TPR	FPR	TPR	FPR	TPR	FPR
5.00	NA10851	CN = 1	0.9984	0.0003	0.9736	0.0003	0.9907	0.0103
		CN = 2	0.9982	0.0003	0.9842	0.0003	0.9906	0.0106
	NA12812	CN = 1	0.9984	0.0004	0.9645	0.0003	0.9905	0.0108
		CN = 2	0.9982	0.0002	0.9801	0.0003	0.9905	0.0103
	NA18605	CN = 1	0.9979	0.0004	0.9728	0.0002	0.9904	0.0118
		CN = 2	0.9980	0.0004	0.9852	0.0002	0.9904	0.0120
2.00	NA10851	CN = 1	0.9984	0.0031	0.9353	0.0085	0.9922	0.0183
		CN = 2	0.9987	0.0048	0.9724	0.0076	0.9914	0.0184
	NA12812	CN = 1	0.9991	0.0055	0.9227	0.0159	0.9917	0.0268
		CN = 2	0.9990	0.0041	0.9622	0.0109	0.9914	0.0214
	NA18605	CN = 1	0.9991	0.0088	0.9364	0.0110	0.9926	0.0231
		CN = 2	0.9988	0.0050	0.9720	0.0105	0.9918	0.0212
1.25	NA10851	CN = 1	0.9991	0.1798	0.8878	0.2002	0.9954	0.2531
		CN = 2	0.9996	0.1322	0.9387	0.1672	0.9951	0.2096
	NA12812	CN = 1	0.9989	0.2592	0.8731	0.2875	0.9962	0.3700
		CN = 2	0.9999	0.2291	0.9251	0.2453	0.9966	0.3149
	NA18605	CN = 1	0.9987	0.1876	0.8860	0.2211	0.9959	0.2875
		CN = 2	0.9991	0.1671	0.9381	0.1936	0.9954	0.2536

ROC Curves (1)

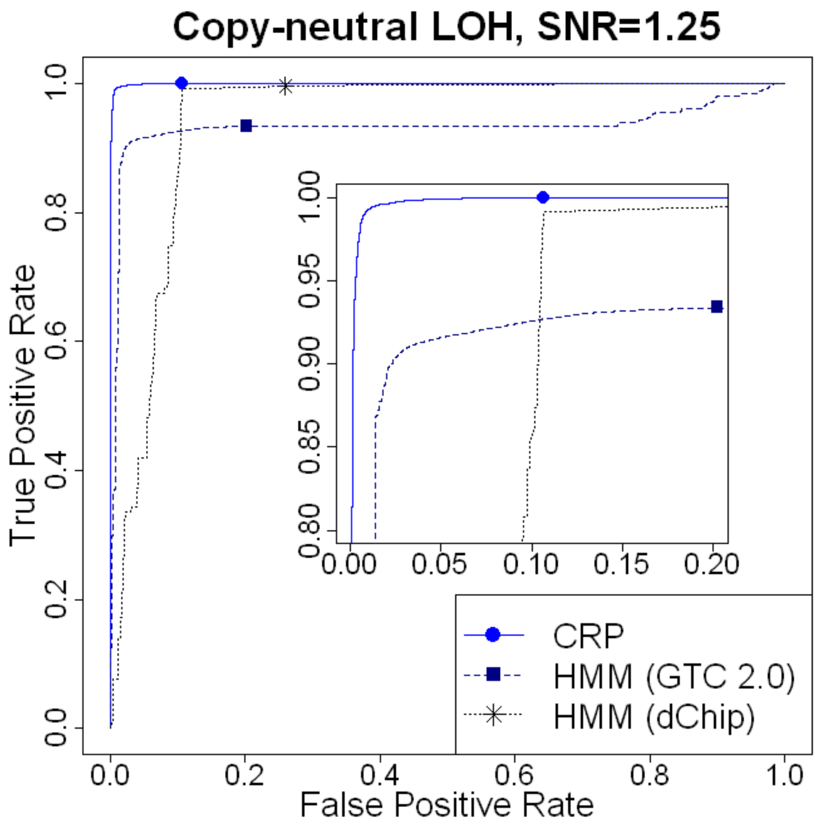

Copy-neutral LOH, SNR=5
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.95
0.97
0.99
0.00
0.02
0.04
CRP
HMM (GTC 2.0)
HMM (dChip)

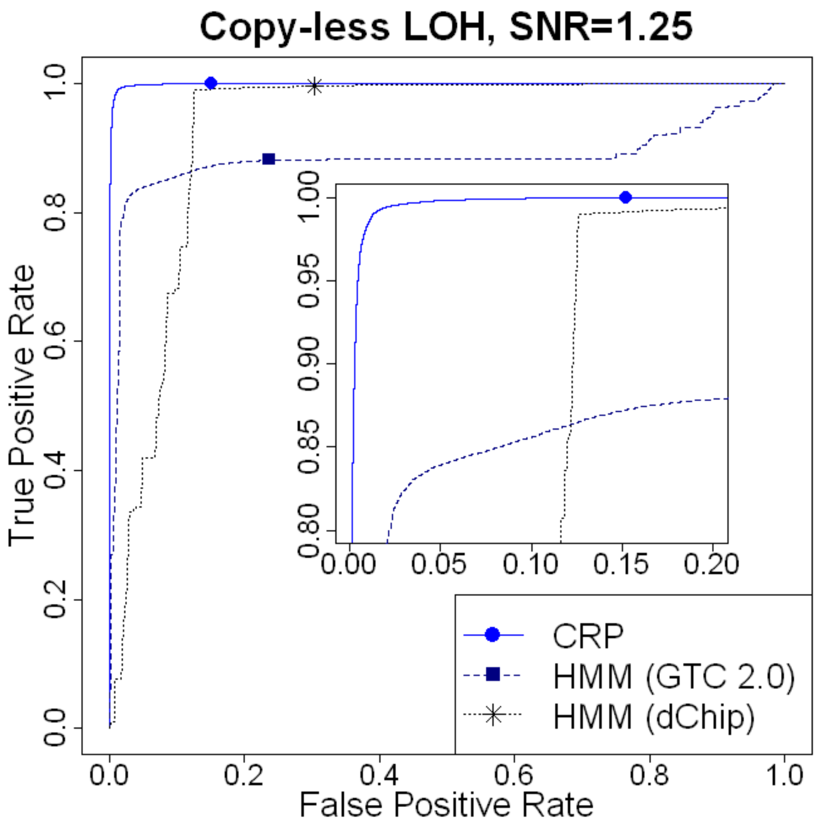

Copy-less LOH, SNR=5
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.95
0.97
0.99
0.00
0.02
0.04
CRP
HMM (GTC 2.0)
HMM (dChip)

ROC Curves (2)

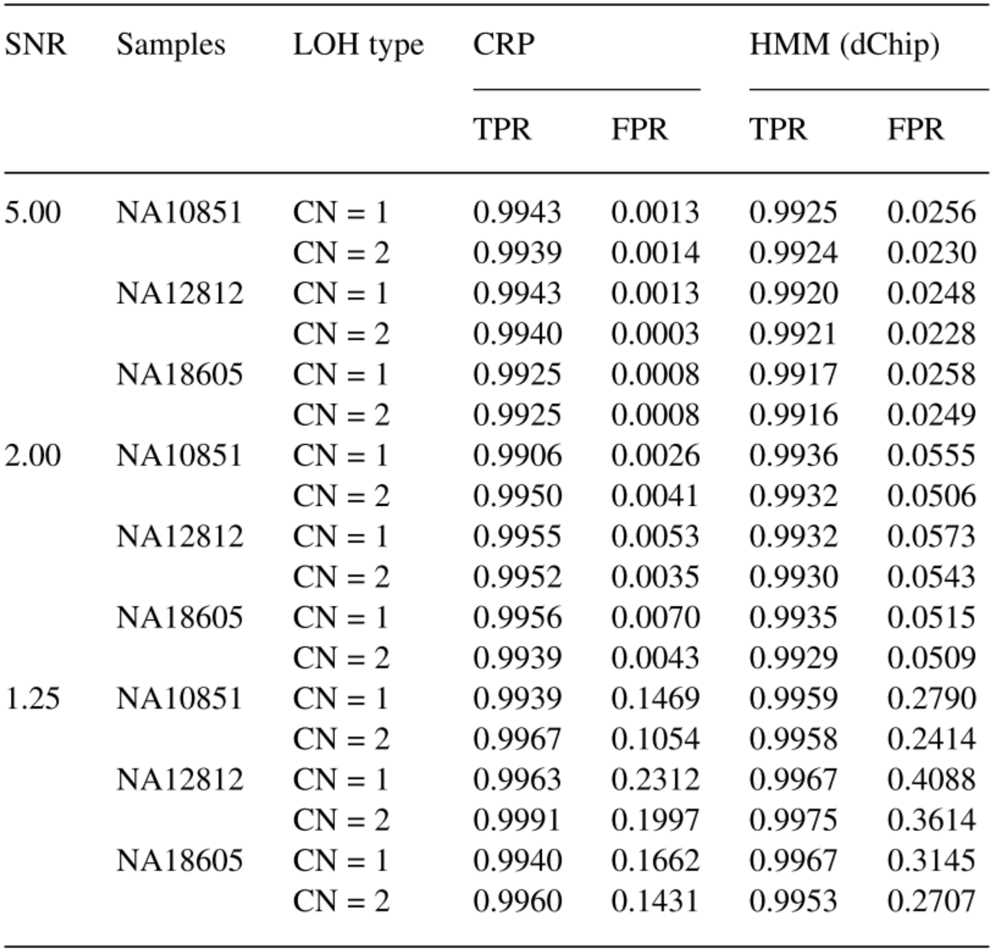

Copy-neutral LOH, SNR=2
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.95
0.97
0.99
0.00
0.02
0.04
CRP
HMM (GTC 2.0)
HMM (dChip)
Copy-less LOH, SNR=2
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.95
0.97
0.99
0.00
0.02
0.04
CRP
HMM (GTC 2.0)
HMM (dChip)

ROC Curves (3)


Copy-neutral LOH, SNR=1.25
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.80
0.85
0.90
0.95
1.00
0.00
0.05
0.10
0.15
0.20
CRP
HMM (GTC 2.0)
HMM (dChip)

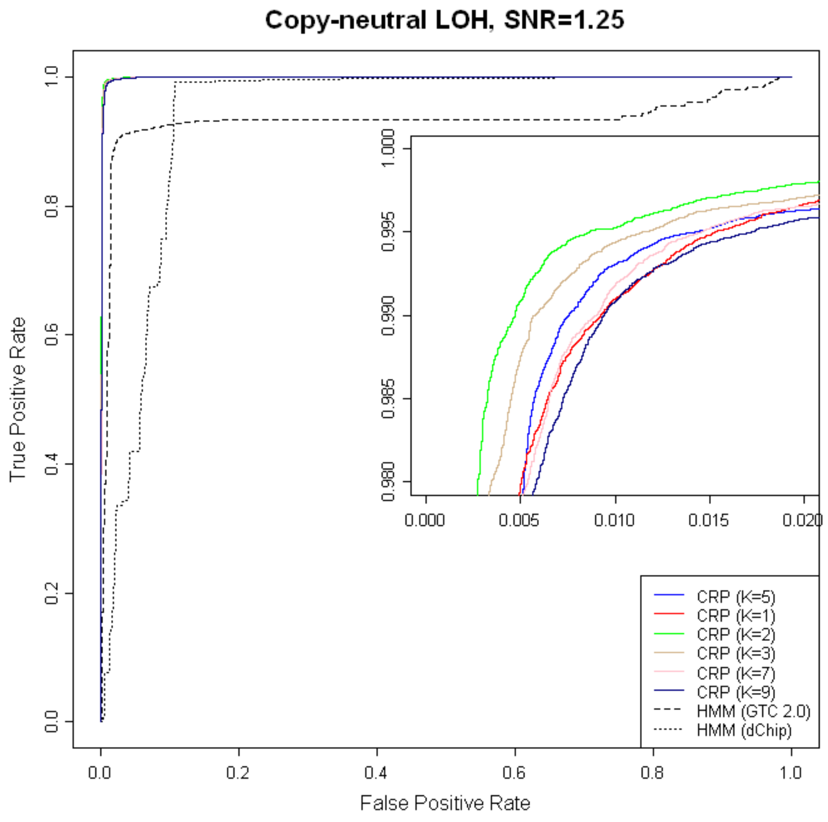

Copy-less LOH, SNR=1.25
True Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.80
0.85
0.90
0.95
1.00
0.00
0.05
0.10
0.15
0.20
CRP
HMM (GTC 2.0)
HMM (dChip)

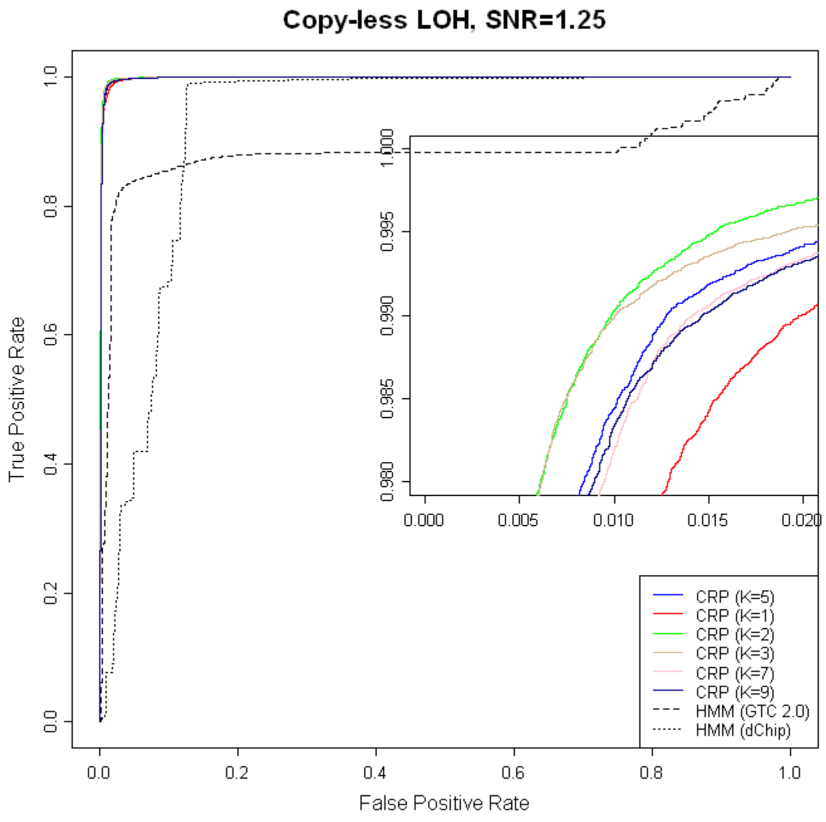
Non-informative SNPs



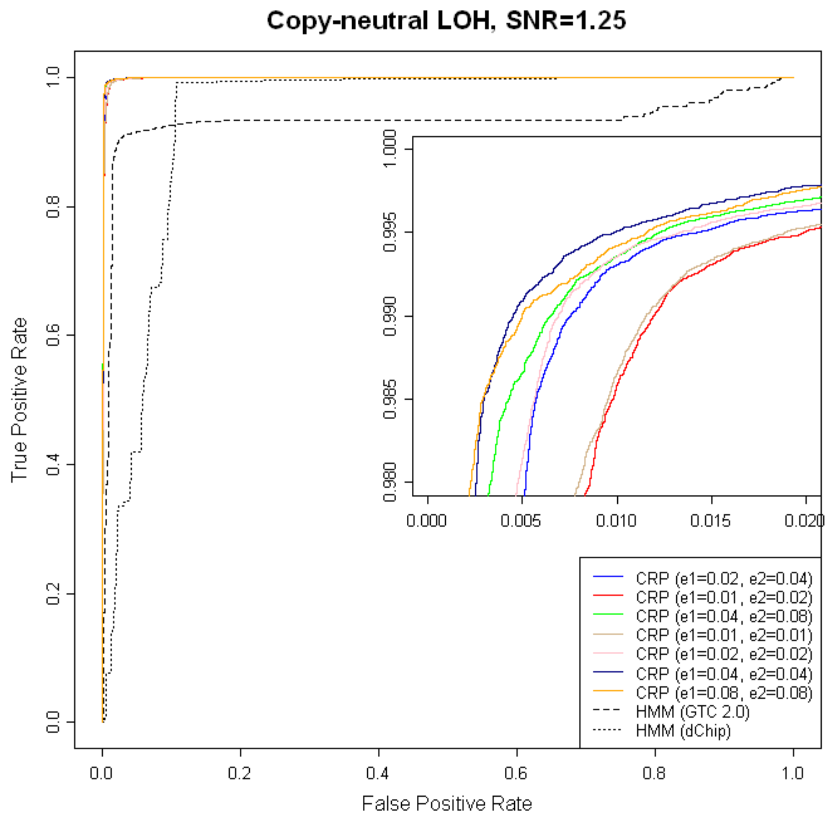
SNR	Samples	LOH type	CRP		HMM (dChip)	
			TPR	FPR	TPR	FPR
5.00	NA10851	CN = 1	0.9943	0.0013	0.9925	0.0256
		CN = 2	0.9939	0.0014	0.9924	0.0230
	NA12812	CN = 1	0.9943	0.0013	0.9920	0.0248
		CN = 2	0.9940	0.0003	0.9921	0.0228
	NA18605	CN = 1	0.9925	0.0008	0.9917	0.0258
		CN = 2	0.9925	0.0008	0.9916	0.0249
2.00	NA10851	CN = 1	0.9906	0.0026	0.9936	0.0555
		CN = 2	0.9950	0.0041	0.9932	0.0506
	NA12812	CN = 1	0.9955	0.0053	0.9932	0.0573
		CN = 2	0.9952	0.0035	0.9930	0.0543
	NA18605	CN = 1	0.9956	0.0070	0.9935	0.0515
		CN = 2	0.9939	0.0043	0.9929	0.0509
1.25	NA10851	CN = 1	0.9939	0.1469	0.9959	0.2790
		CN = 2	0.9967	0.1054	0.9958	0.2414
	NA12812	CN = 1	0.9963	0.2312	0.9967	0.4088
		CN = 2	0.9991	0.1997	0.9975	0.3614
	NA18605	CN = 1	0.9940	0.1662	0.9967	0.3145
		CN = 2	0.9960	0.1431	0.9953	0.2707

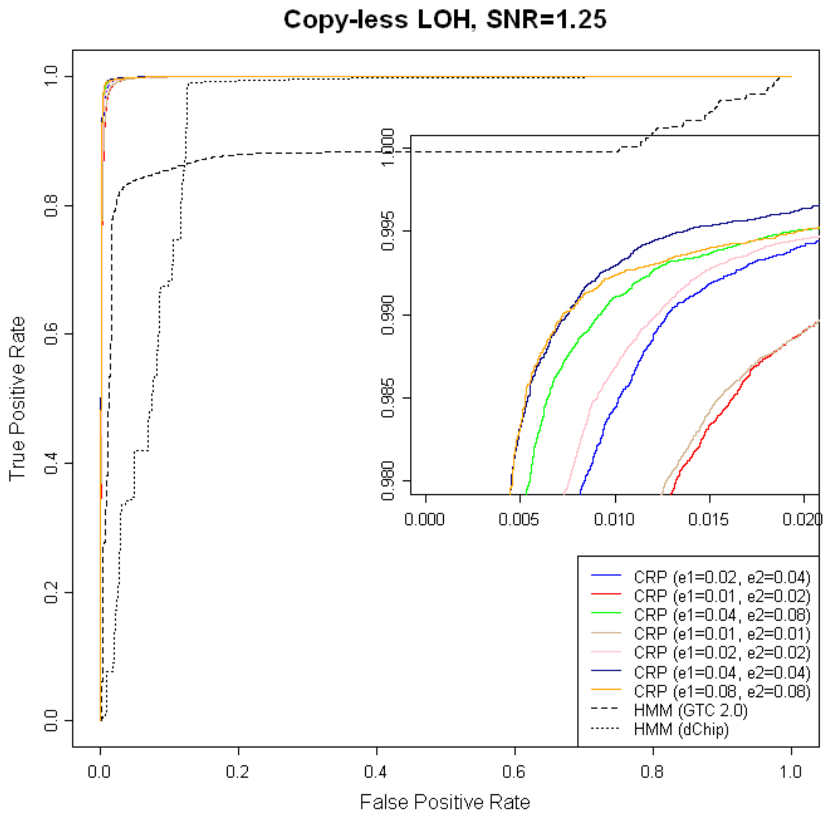
Parameters (1)



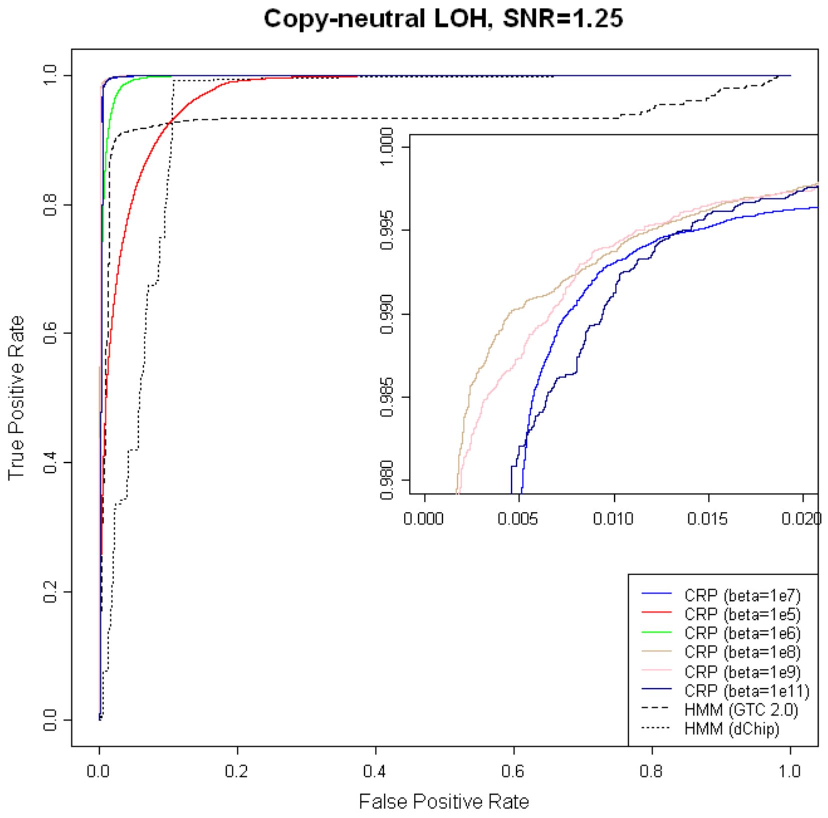


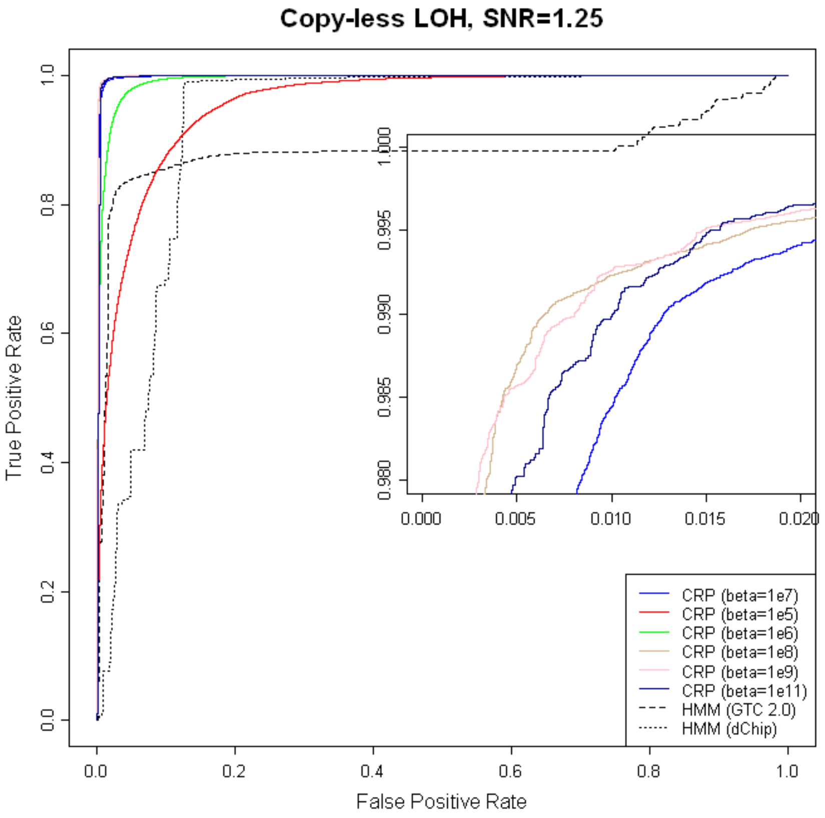
Parameters (2)


Copy-neutral LOH, SNR=1.25


Copy-less LOH, SNR=1.25

Parameters (3)


Copy-neutral LOH, SNR=1.25


Copy-less LOH, SNR=1.25