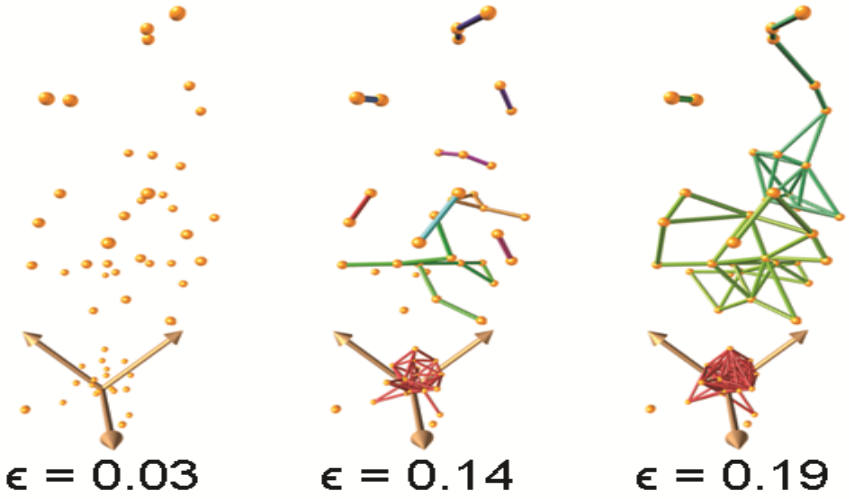
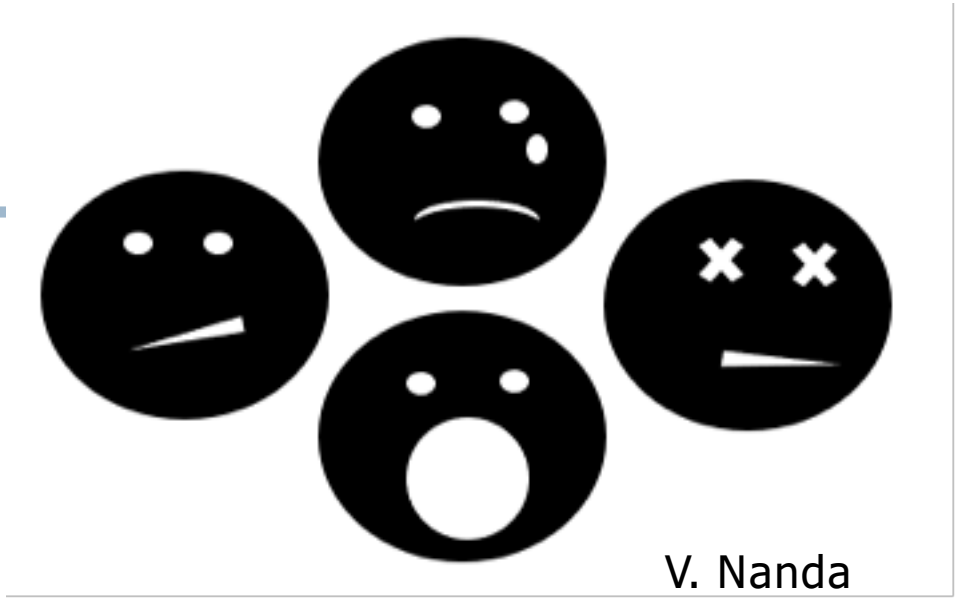


Identification of genetic interactions using computational homology

Javier Arsuaga
Mathematics
Molecular and Cellular Biology
University of California, Davis



Identification of genetic interactions using computational homology

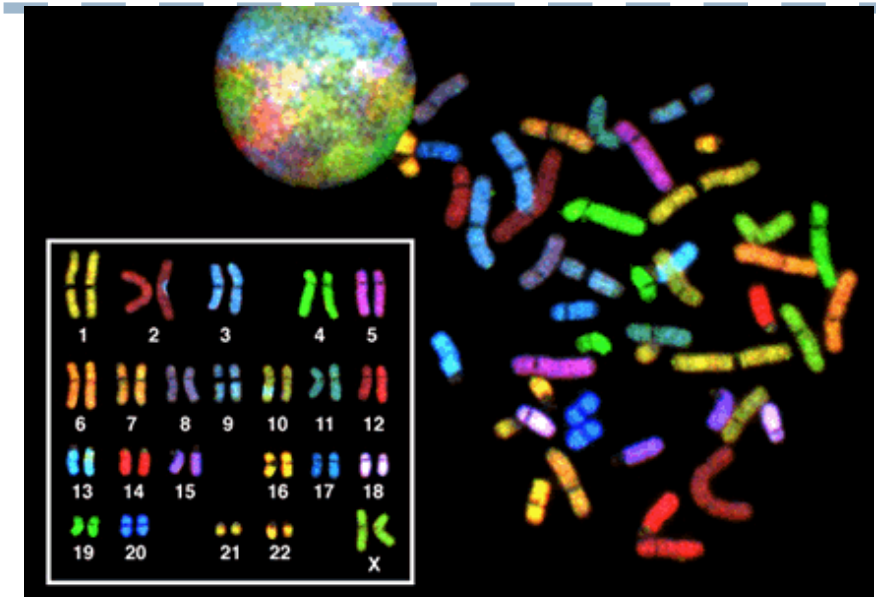
Javier Arsuaga
Mathematics
Molecular and Cellular Biology
University of California, Davis

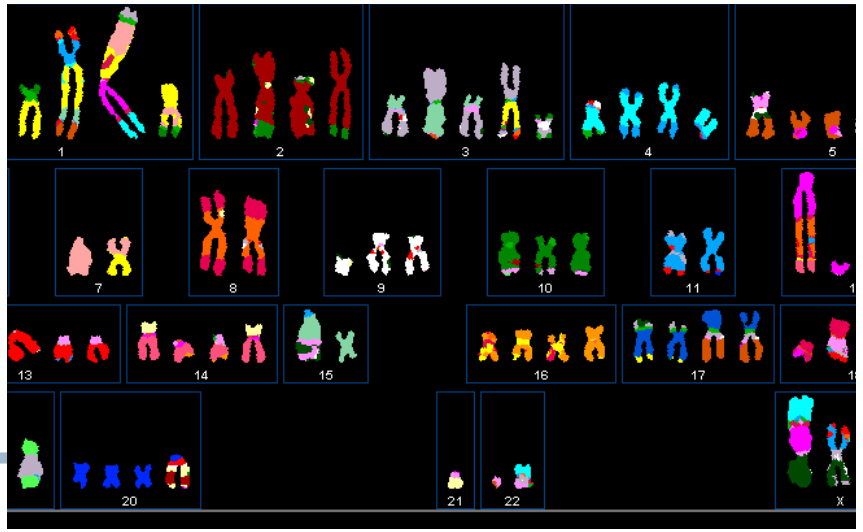
Topological Molecular Biology Lab

UC DAVIS
UNIVERSITY OF CALIFORNIA

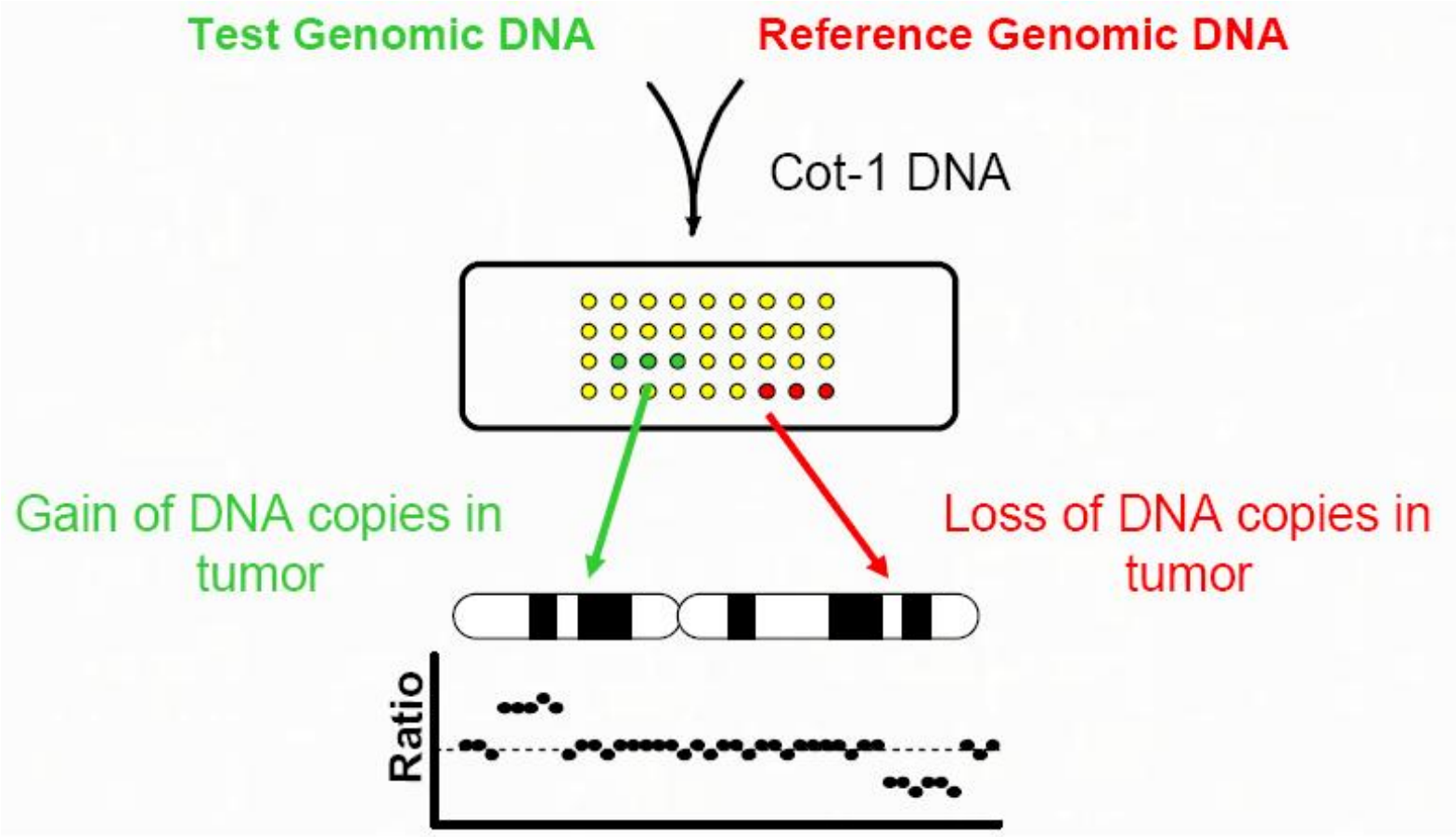
NATIONAL INSTITUTES OF HEALTH

In cancer the structure of the genome can be heavily disrupted





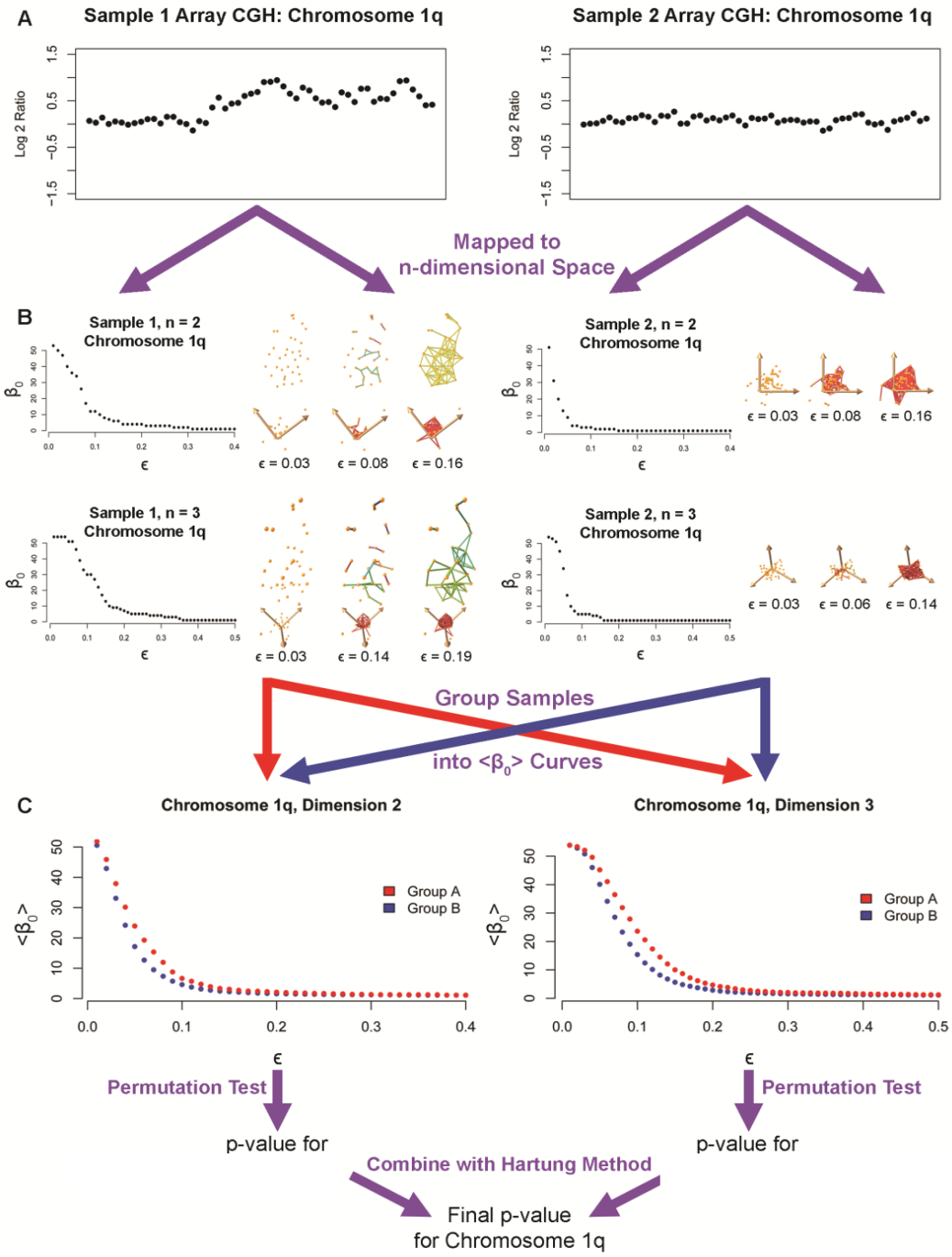
Amplifications and Deletions across the entire genome can be detected using array CGH



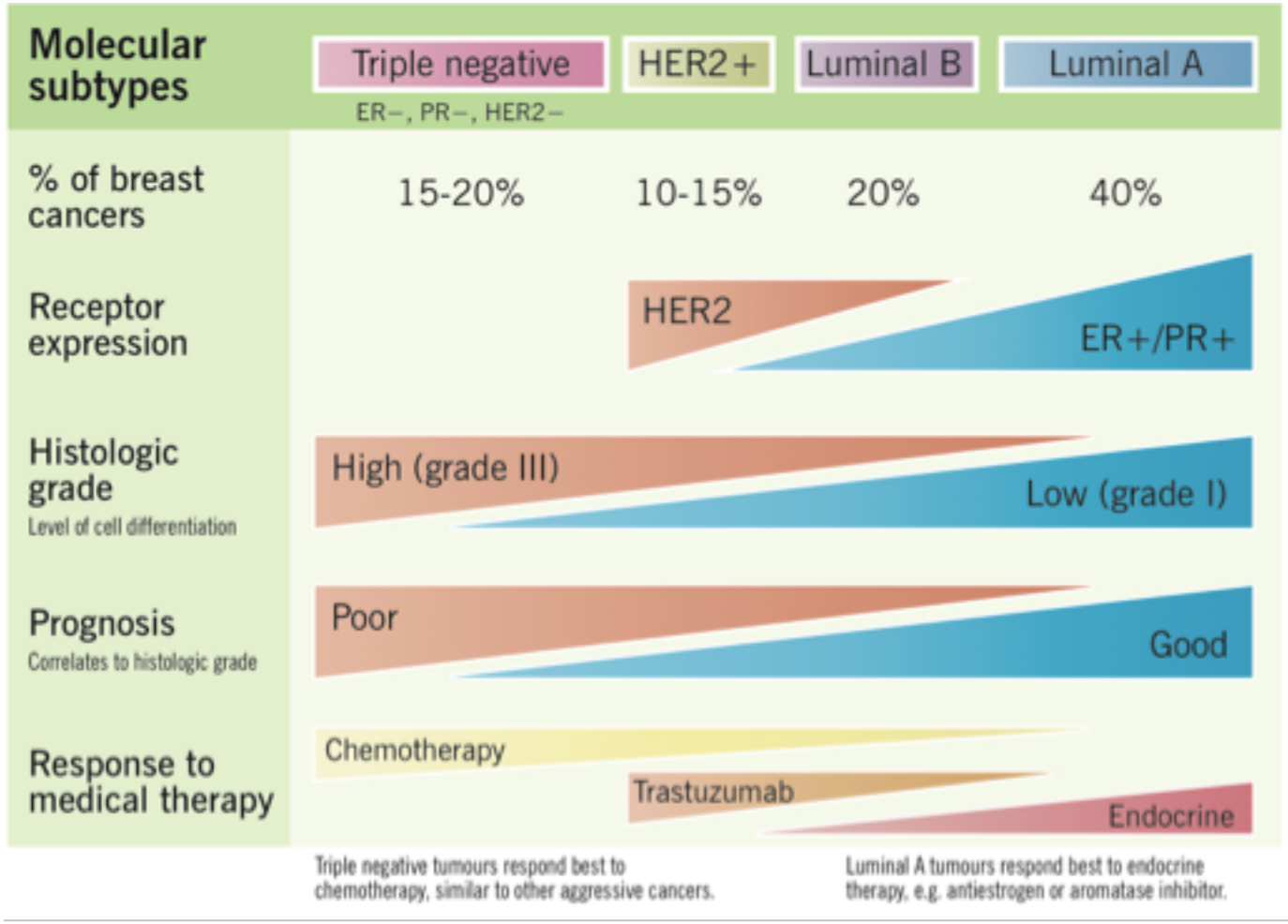
Topological Analysis of array CGH

(TAaCGH)

DeWoskin et al. 2009
DeWoskin et al. 2010
Arsuaga et al. 2012
Arsuaga et al. 2015



We analyzed four breast cancer subtypes



Clin Cancer Res; 16(2) January 15, 2010 663

Clinical Cancer Research

Imaging, Diagnosis, Prognosis

Integration of DNA Copy Number Alterations and Prognostic Gene Expression Signatures in Breast Cancer Patients

Hugo M. Horlings[1], Carmen Lai[2,6], Dimitry S.A. Nuyten[3], Hans Halfwerk[1], Petra Kristel[1], Erik van Beers[1], Simon A. Joosse[1], Christiaan Klijn[2], Petra M. Nederlof[4], Marcel J.T. Reinders[6], Lodewyk F.A. Wessels[2,6], and Marc J. van de Vijver[5]

GENES, CHROMOSOMES & CANCER 45:1033–1040 (2006)

Distinct Patterns of DNA Copy N... Alteration Are Associated with Different Clinicopathological Features and Gene-Expression Subtypes of Breast Cancer

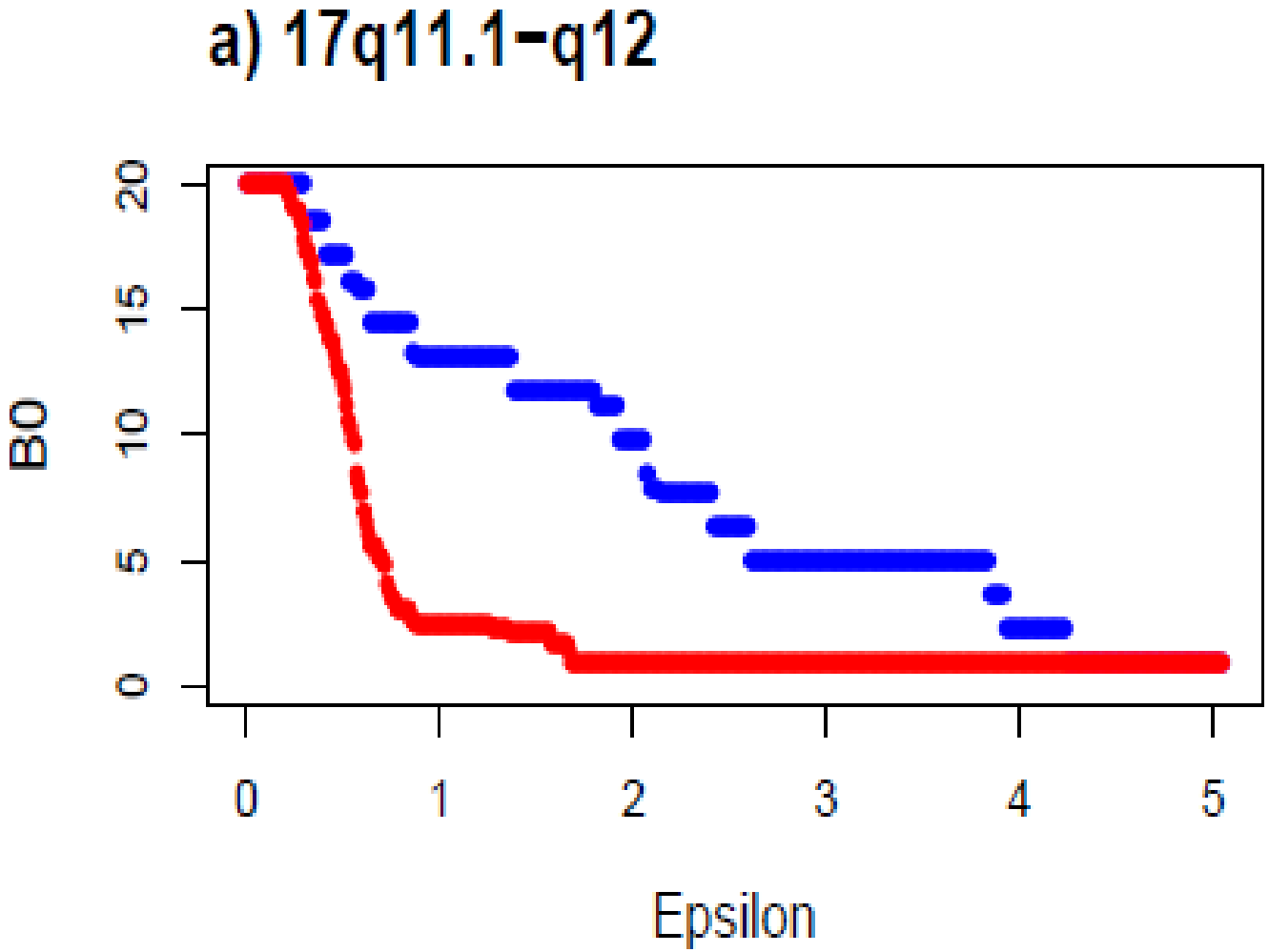
Anna Bergamaschi,[1,2] Young H. Kim,[2] Pei Wang,[3†] Therese Sørlie,[1] Tina Hernandez-Boussard,[4] Per E. Lonning,[5] Robert Tibshirani,[3,6] Anne-Lise Børresen-Dale,[1,7] and Jonathan R. Pollack[2*]

ARTICLE

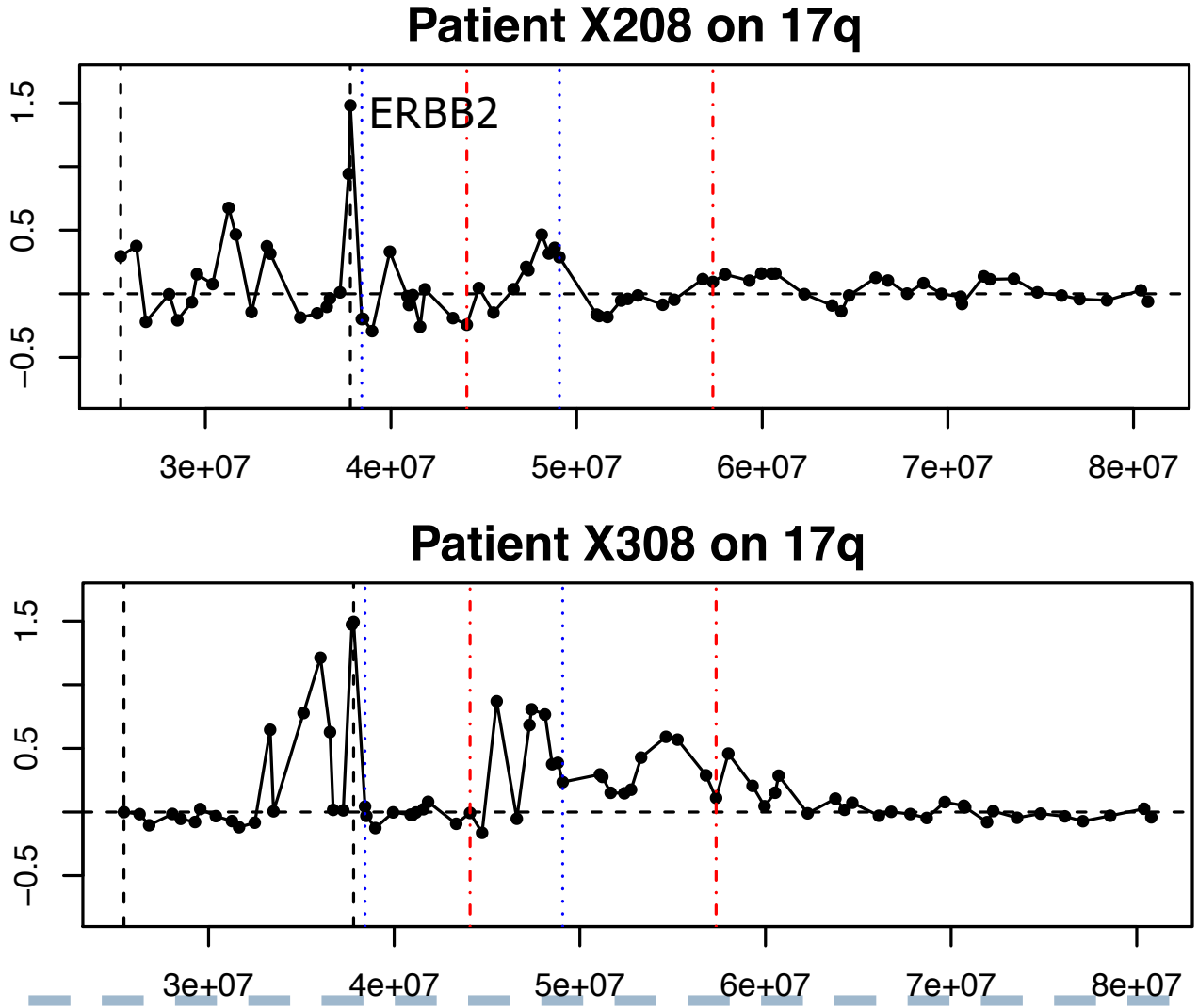
Comprehensive molecular portraits of human breast tumours

The Cancer Genome Atlas Network*

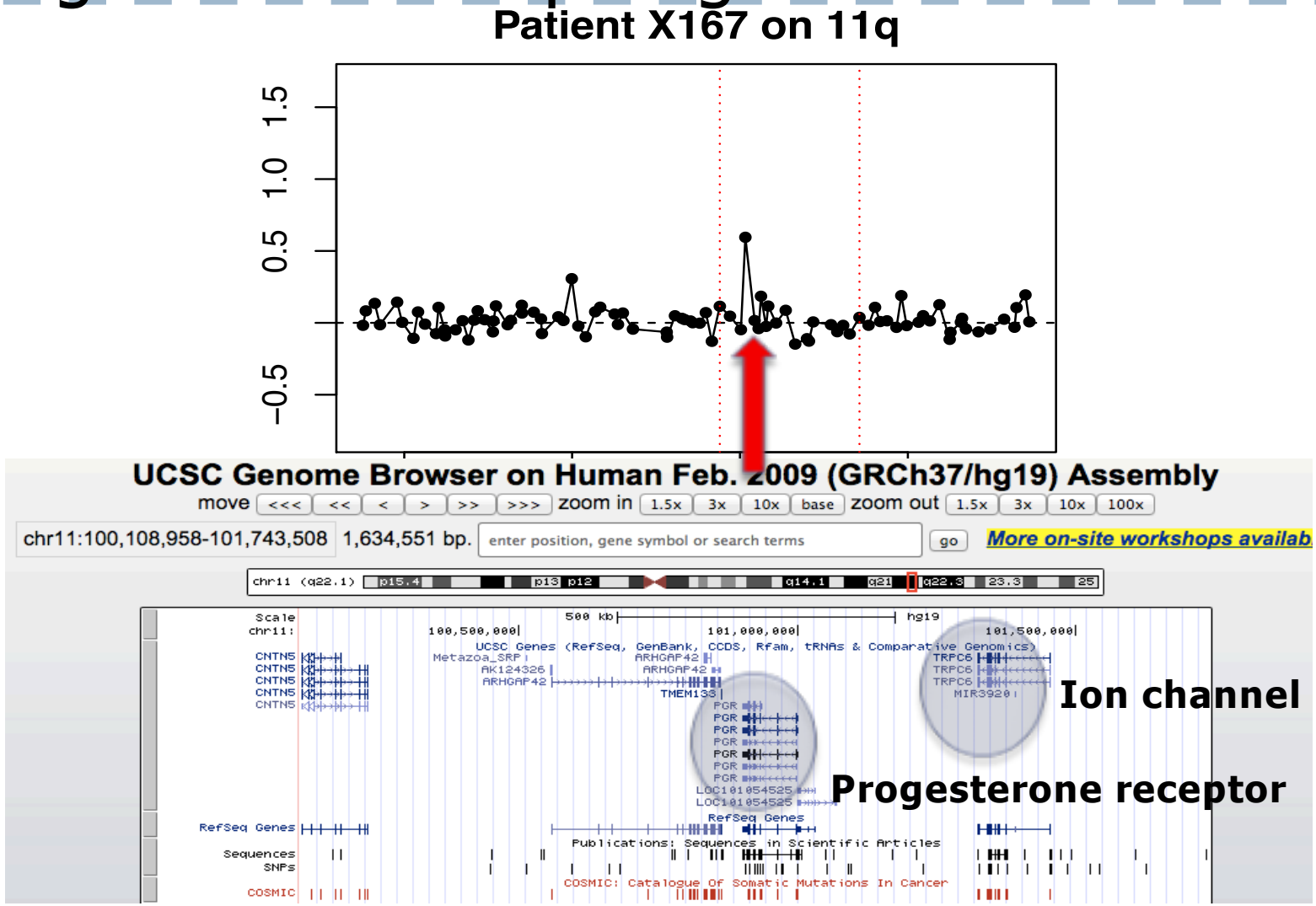
Our method identified the region of ERBB2 (Her2+ shown in blue)



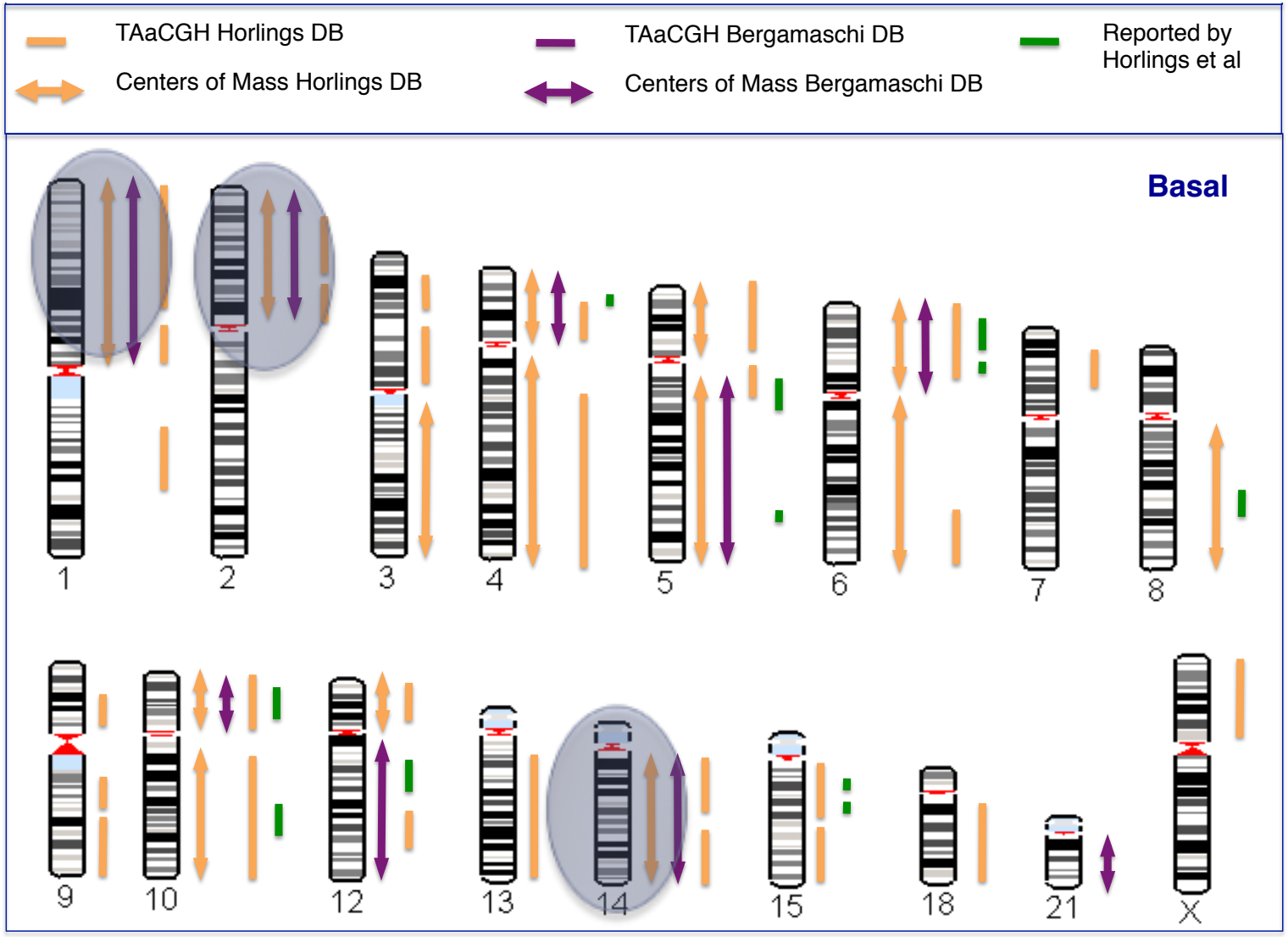
Examples of ERBB2/Her2 patient profiles



Luminal A: An amplification at the site of the Progesterone Receptor gene

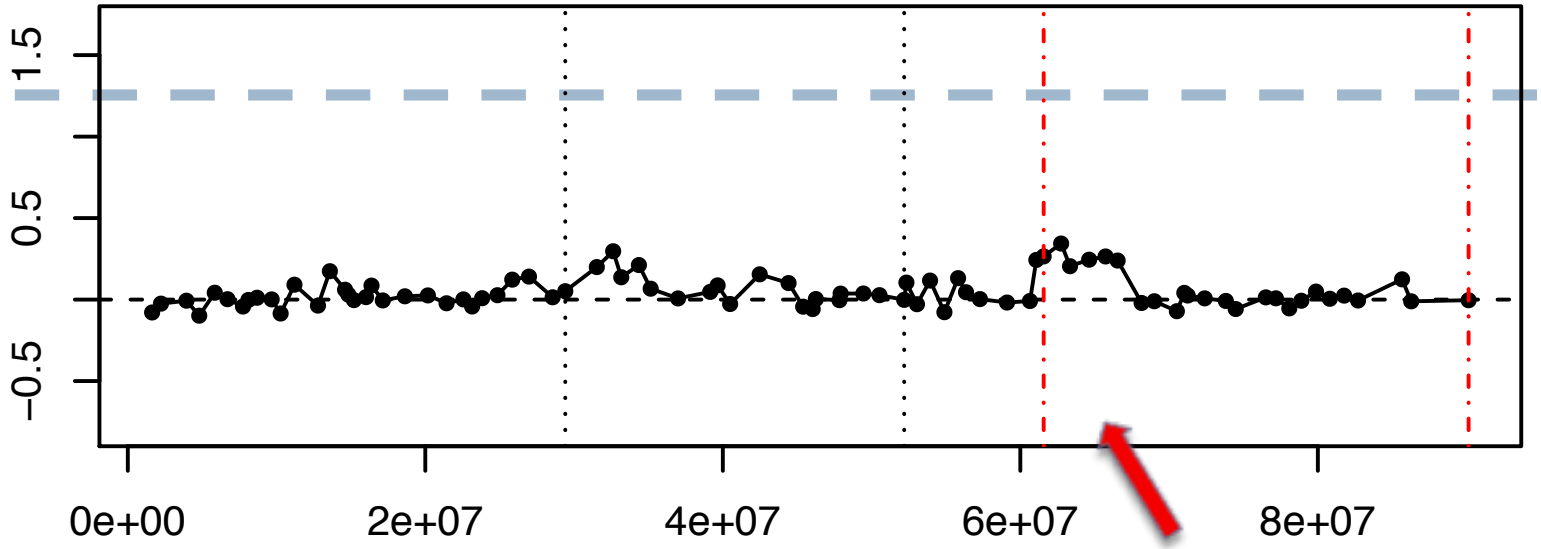


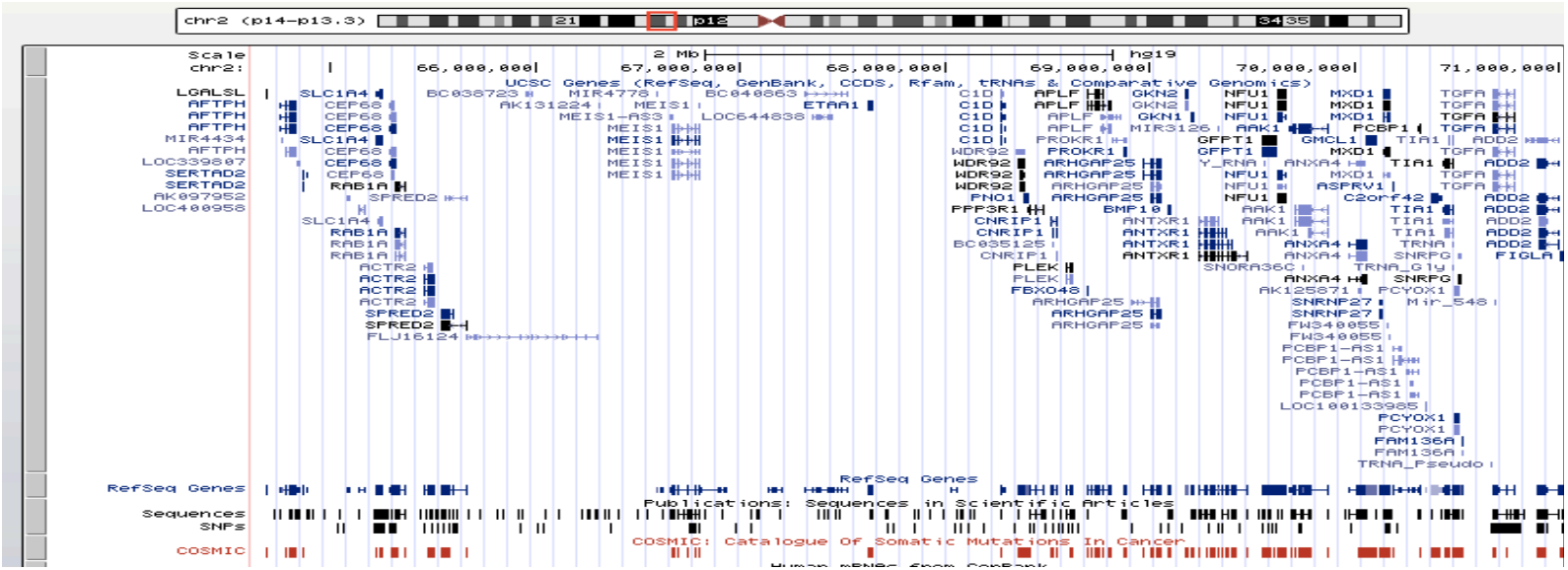
Regions detected in Basal Patients



Gain in 2p

Patient X330 on 2p





Luminal A

Horlings et al. dataset			Bergamaschi et al. dataset (Validation)
TAaCGH	SIRAC	GISTIC	TAaCGH
1q			
Full Arm	q21.3-q44	q23.3 (76%) q41 (81%)	Full Arm
8p			
		p23.2 (38%)	
8q			
		q24.11 (62%)	
11q			
q22.1-q23.2		q13.4 (29%)	Full Arm
13q			
		q14.11 (48%)	
16p			
Full Arm	p13-p12.3		Full Arm
16q			
Full Arm	q11.2-q13 q22.1-q24.1		Full Arm

HER2+

Horlings et al. dataset			Bergamaschi et al. dataset (Validation)
TAaCGH	SIRAC	GISTIC	TAaCGH
1q			
		q23.3 (57%) q41 (64%)	
3p			
		p14.3 (43%)	
8p			
		p23.2 (43%)	
13q			
		q14.11 (36%)	
17q			
q11.1-q12 q12-q21.31 q21.31-q22	q11.1-q12 q21.31-q23.2 q21.31-q23.2	q23.1 (50%)	q12.q21.2 q12.q21.2

Luminal B

Horlings et al. dataset			Bergamaschi et al. dataset (Validation)
TAaCGH	SIRAC	GISTIC	TAaCGH
1p			
p32.3-p31.1 p35.1-p33 p36.32-p34.2	p31.3		p34.2-p31.1 p35.1-p34.3 p35.3-p35.2 p36.21-p36.11 p36.32-p36.22
1q			
		q23.3 (67%) q41 (58%)	
3p			
		p14.3 (50%)	
3q			
		q27.2 (42%)	
4q			
q24-q27			q23-q27
6q			
		q26 (50%)	
8p			
p23.3-p11.1 p22-p11.1	p23.1-p21.2	p23.2 (75%)	p21.2-p11.21 p23.3-p21.3
8q			
q24.11-q24.3		q24.11 (83%)	q24-13-q24.3
9p			
Full Arm p23-p21.1 p24.3-p22.3			p22.1-p21.1 p24.2-p22.3
9q			
q13-q22.32 q31.1-q33.1			q12-q22.33 q22.33-q31.1 q31.2-q33.2
11q			
		q24.3 (50%)	
13q			
q12.2-q21.1 q31.1-q32.2		q14.11 (92%)	q13.1-q14.3 q31.1 q32.2
17q			
	q23.2	q23.1 (67%) q24.3 (50%)	
18q			
		q12.2 (42%)	
21q			
q11.2-q22.3			q11.2-q22.3

results from Arsuaga et al 2015

Basal			
Horlings et al. dataset			Bergamaschi et al. dataset (Validation)
TAaCGH	SIRAC	GISTIC	TAaCGH
1p			
Full Arm p22.2-p12 p32.1-p31.1 p32.3-p31.1 p35.1-p33 p36.22-p35.1			Full Arm
1q			
q23.1-q31.1		q23.3 (76%) q41 (52%)	
2p			
Full Arm p15-p11.2			Full Arm
3p			
p22.1-p11.2 p25.1-p23 p26.3-p25.1		p14.3 (67%)	
3q			
Full Arm		q27.2 (57%)	
4p			
Full Arm p15.1-p11	 p15.31	 p15.2 (62%)	Full Arm
5q			
Full Arm q11.1-q13.1	 q12.3-q13.2 q33.1	 q32 (57%)	Full Arm
6p			
Full Arm q11.2-q25.3 q11.2-q25.3	 q12.3 q21.1-q23		Full Arm
6q			
Full Arm q24.1-q27		q23.3 (43%)	
7q			
		q34 (62%)	

Basal (...continue)			
Horlings et al. dataset			Bergamaschi et al. dataset (Validation)
TAaCGH	SIRAC	GISTIC	SIRAC
8p			
		p23.2 (57%)	
8q*			
Full Arm	 q22-q24.21	q24.11 (76%)	
10p			
Full Arm p15.3-p11.1	 p14-p12.33	p14 (62%)	Full Arm
10q			
Full Arm q11.21-q23.1 q23.1-q24.2 q24.2-q26.3	 q23.33	 q23.32 (52%)	
11q			
		q24.3 (38%)	
12p			
Full Arm p13.33-p11.21		p13.33 (48%)	
12q*			
 q22-q24.11	q13.13-q13.3		Full Arm
13q			
q12.2-q31.2 q31.2-q34		q14.11 (81%)	
14q			
Full Arm q12-q21.3 q24.3-q32.13 q31.3-q32.33			Full Arm
15q			
q11.2-q22.31 q11.2-q22.31 q23-q26.3	q15.1 q21.1		
17q			
		q24.3 (48%)	
18q			

TAaCGH is a method within statistical genetics: topological genetics

If $E(y|g)$ is the average phenotype for individuals with genotype $g = (g_1, \ldots g_p)$ then we can hypothesize that

$$E(y|g) = \mu + \sum_{j=1}^{p} \alpha_j z_j$$

Where z_j are the genetic loci and α_i are the effects

What we have done is to substitute the elements z_j by $\beta_0(z_j)$ when performing statistical tests.

$$E(y|g) = \mu + \sum_{j=1}^{p} \alpha_j \beta_0(z_j)$$

Phenotype

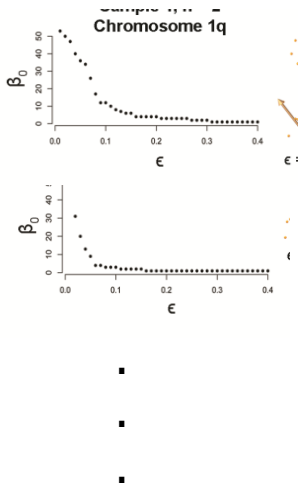
1
0
.
.
.

Markers: z_j

A	B	B
A	A	A
	.	
	.	
	.	

Phenotype

1
0
.
.
.

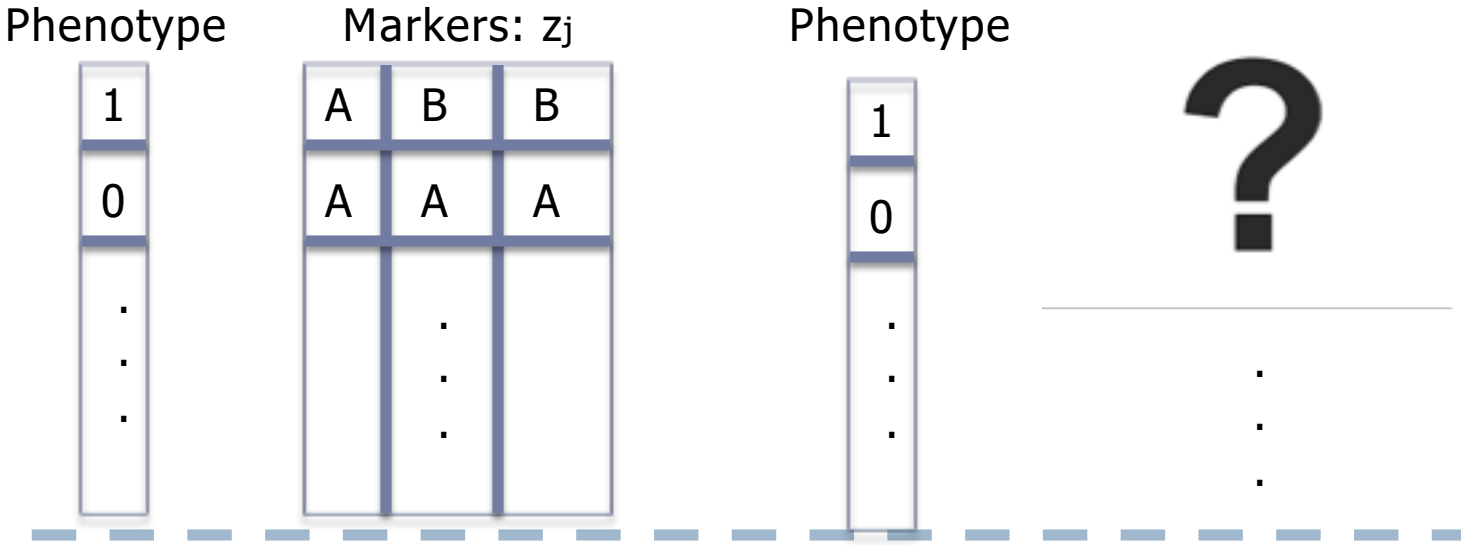


Can we include genetic interactions?

And a more general model in which pairwise interactions are considered

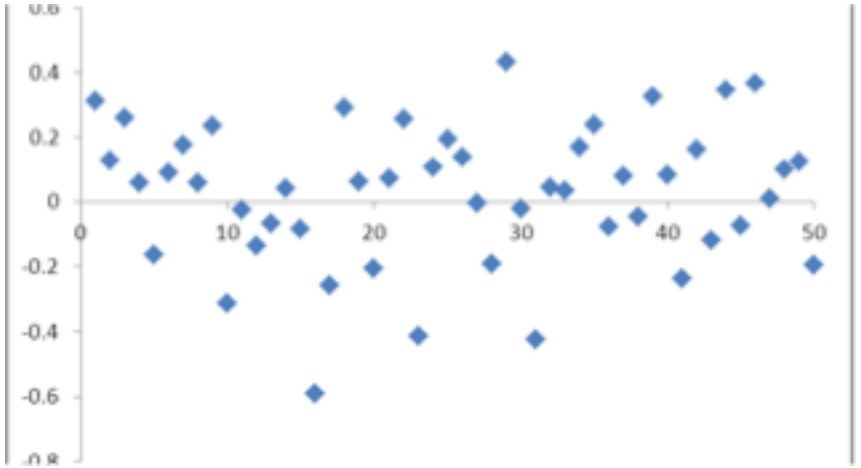
$$E(y|g) = \mu + \sum_{j=1}^{p} \beta_j z_j + \sum_{i,j=1}^{p} \gamma_{ij} z_i z_j$$

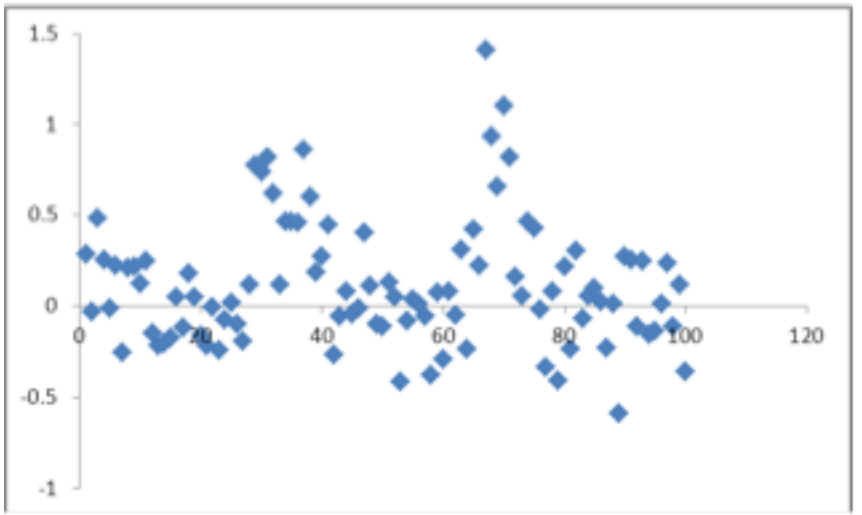
Where as before z_j are the genetic loci and β_i are the effects and γ_{ij} is a binary variable

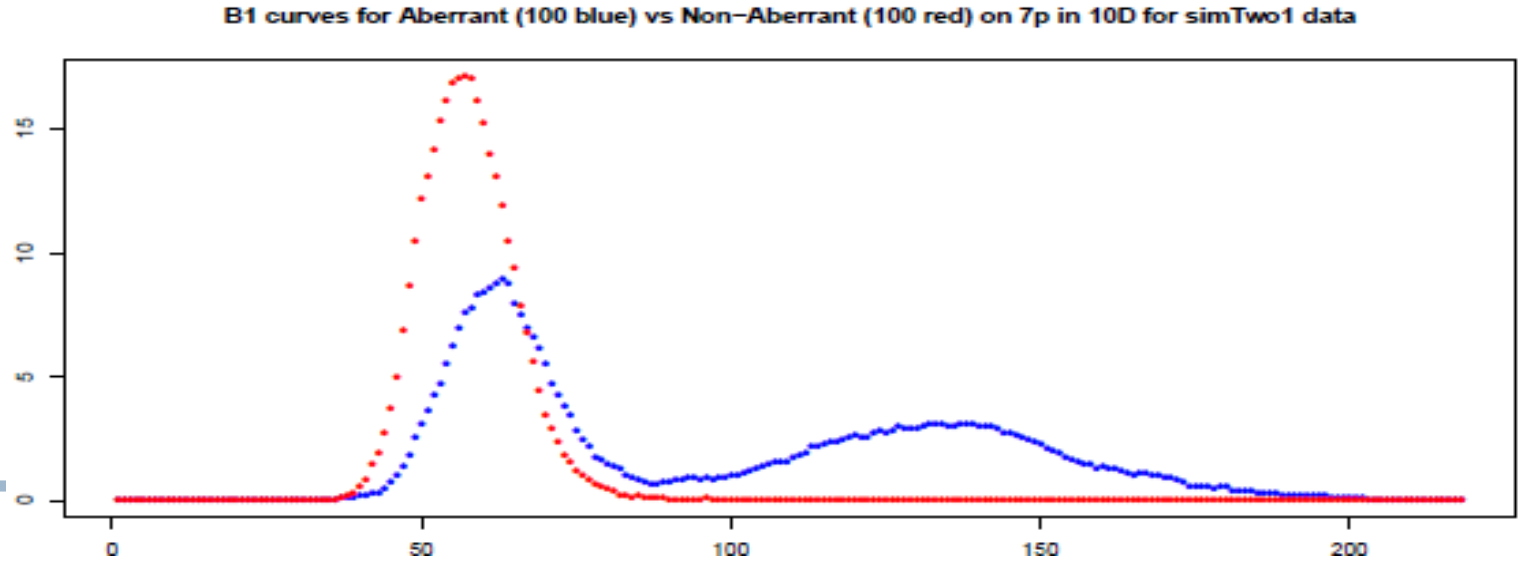


Hypothesis: Genetic interactions in the form of co-amplifications/deletions can be detected by β_1

Computer simulations suggest that β_1 curves can detect co-occurring copy number changes





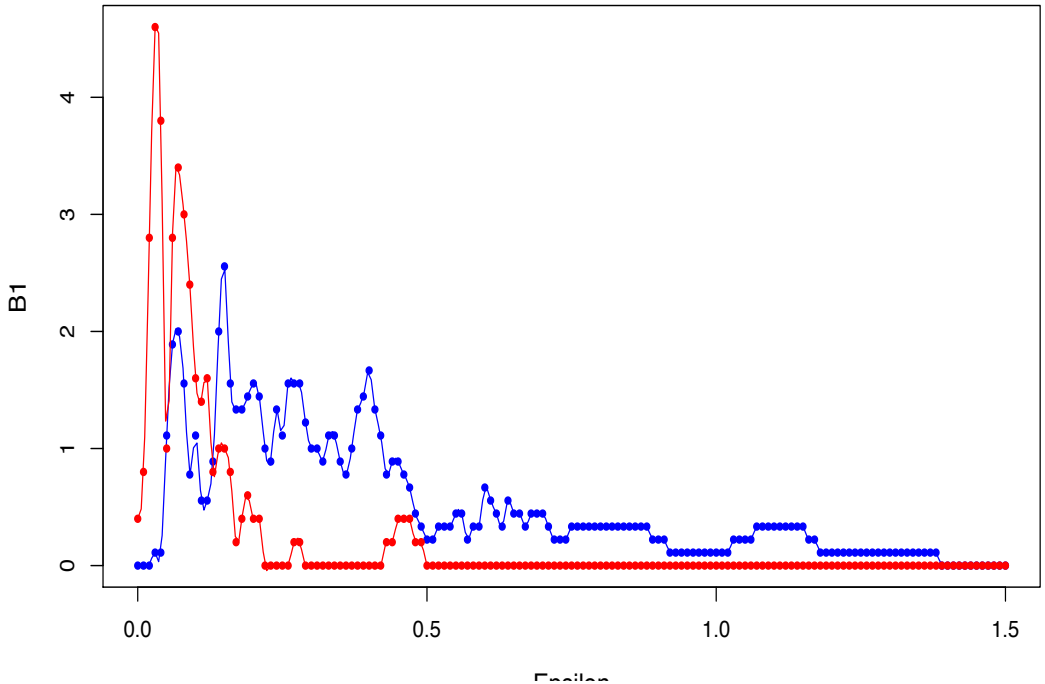


Oncogene. 2009 April 30; 28(17): 1892–1903. doi:10.1038/onc.2009.34.

Co-amplified genes at 8p12 and 11q13 in breast tumors cooperate with two major pathways in oncogenesis

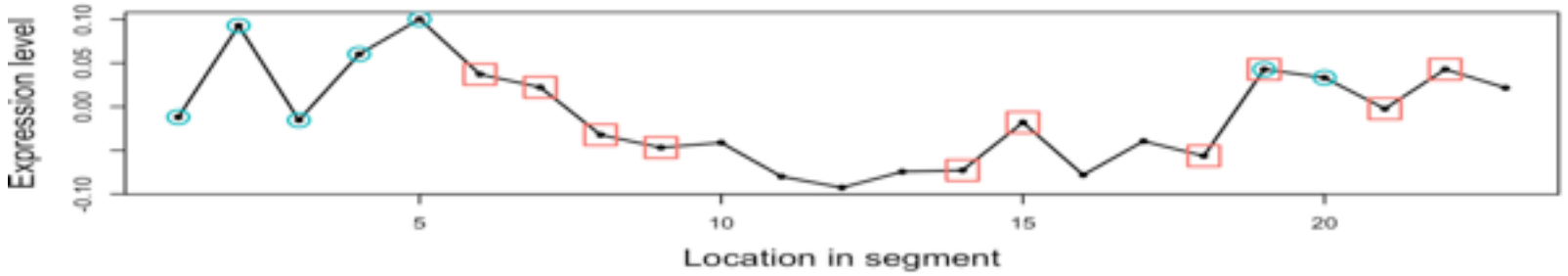
Serena S. Kwek[1,*], Ritu Roy[2], Hua Zhou[1,§], Joan Climent[1,3], Jose A. Martinez-Climent[4], Jane Fridlyand[2,5,¶], and Donna G. Albertson[1,2,6]

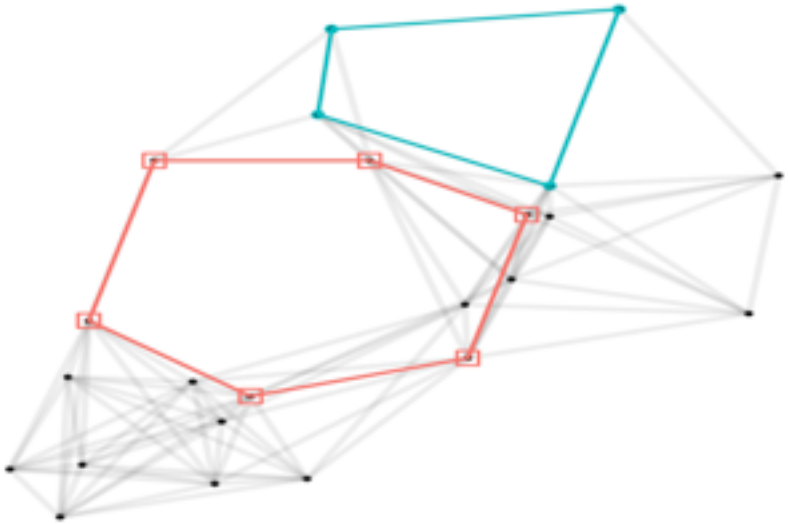
[1] Cancer Research Institute, University of California San Francisco, San Francisco, CA 94143-0808



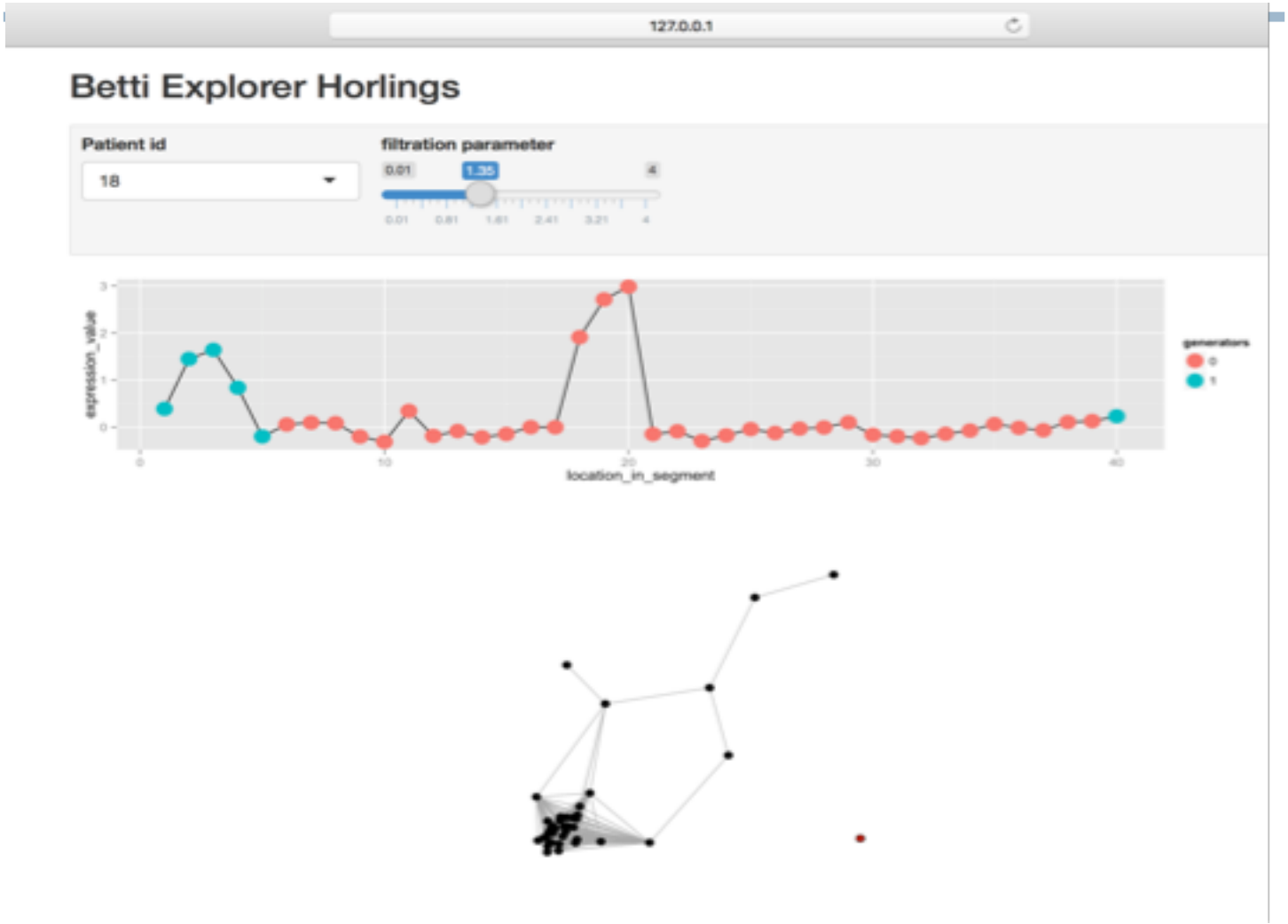
- 60% probes are in 8p/11q
- Test set: 9 patients contain the co-amplification (by inspection),
- Control set: no aberration set was generated artificially by mixing data
- Significant p=0.045

In order to detect co-occurring we need to be able to compute the cycles

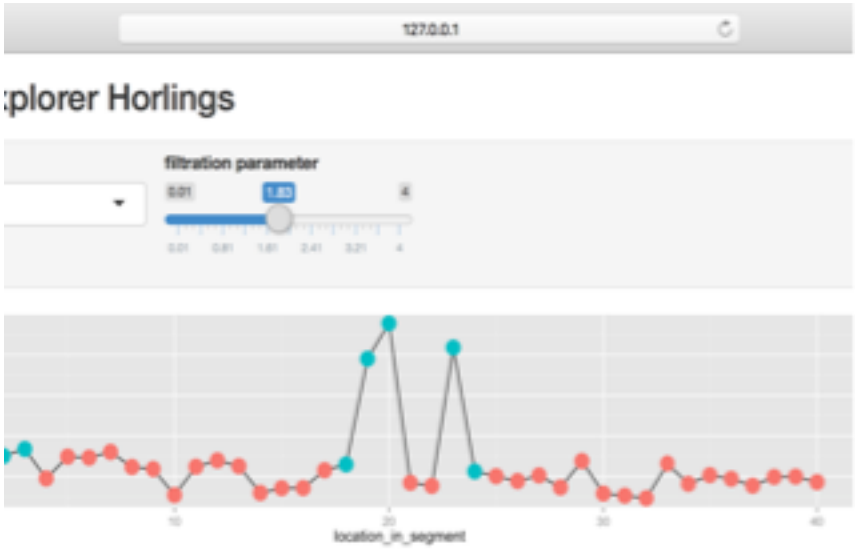


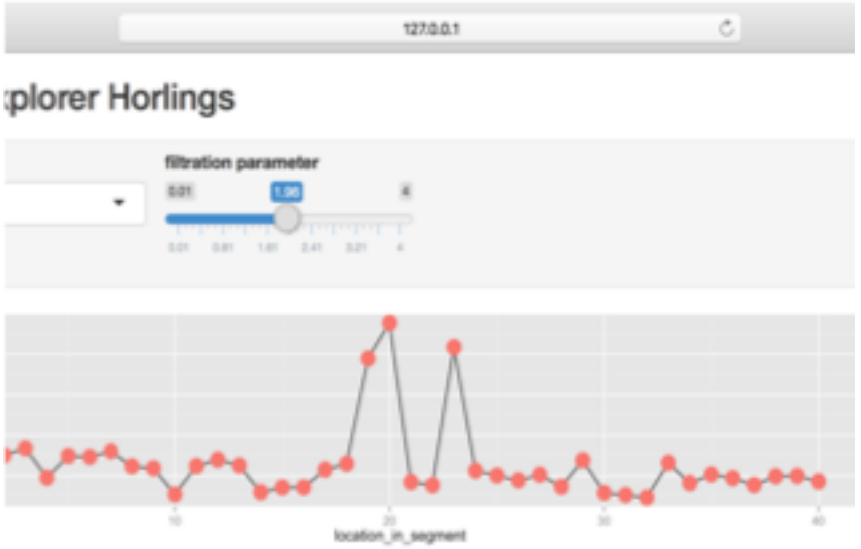


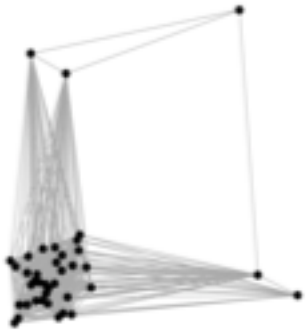
β_1 also detects single amplifications

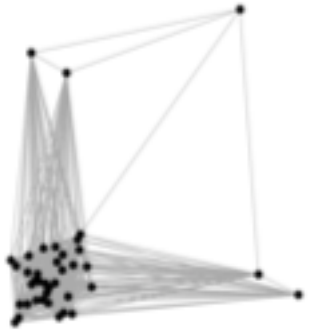


Peaks detected do not necessarily persist

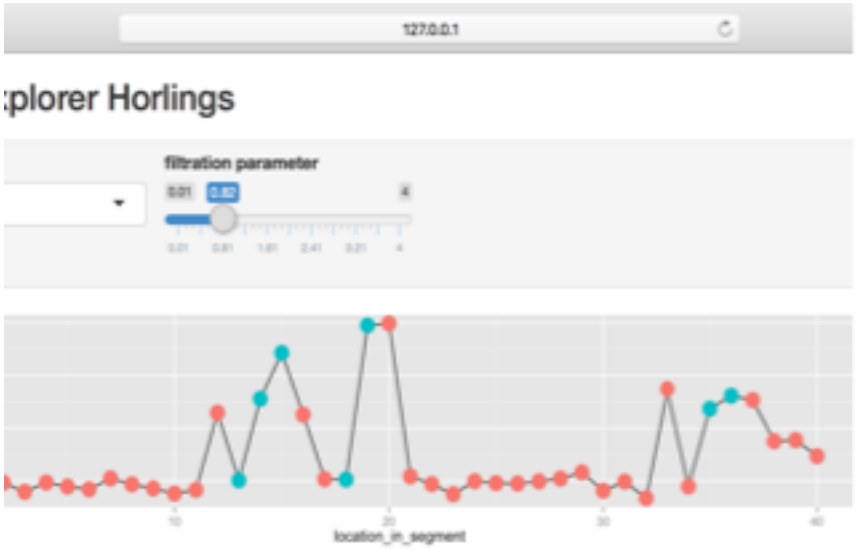


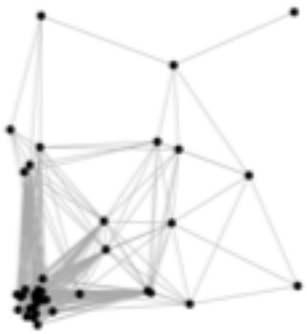


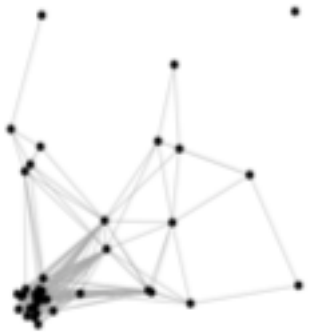




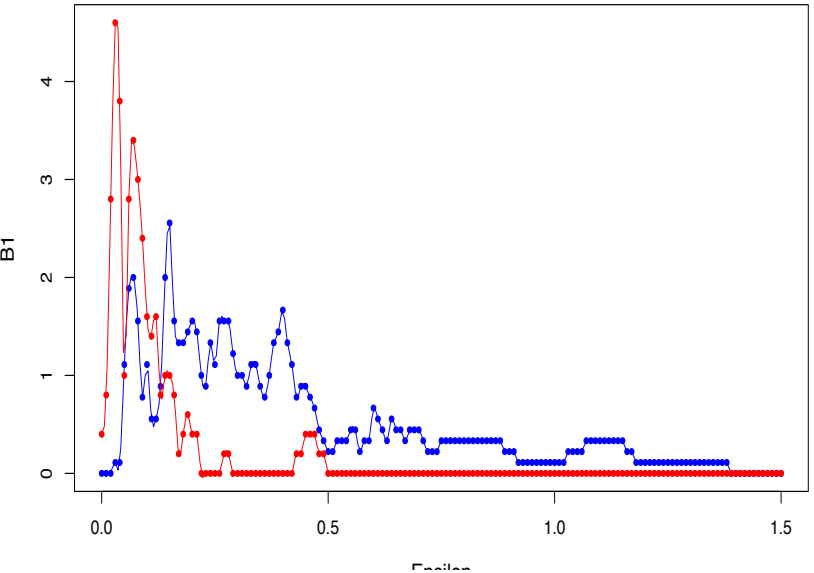
Patterns may change during filtration

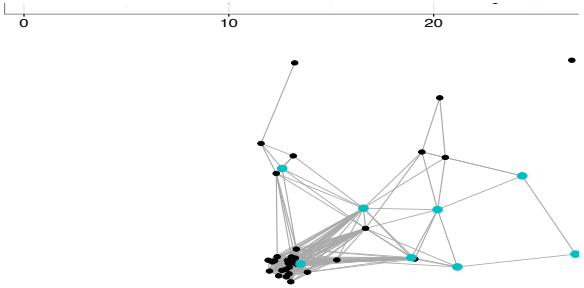


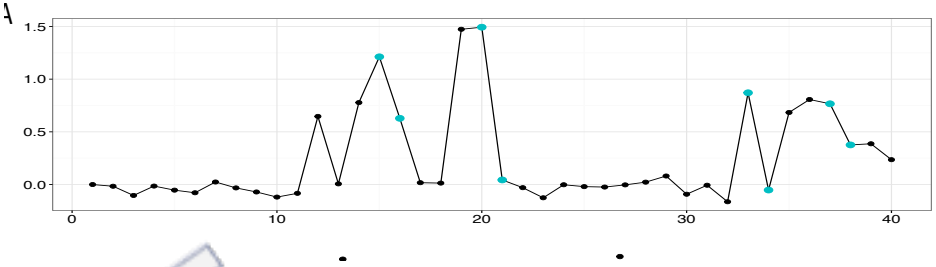


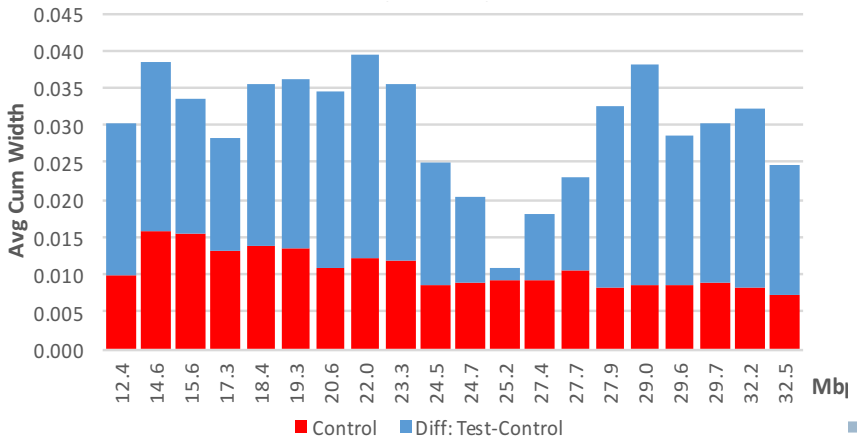


Proposed statistical method for inverse problem









Clin Cancer Res; 16(2) January 15, 2010 663

Clinical Cancer Research

Imaging, Diagnosis, Prognosis

Integration of DNA Copy Number Alterations and Prognostic Gene Expression Signatures in Breast Cancer Patients

Hugo M. Horlings[1], Carmen Lai[2,6], Dimitry S.A. Nuyten[3], Hans Halfwerk[1], Petra Kristel[1], Erik van Beers[1], Simon A. Joosse[1], Christiaan Klijn[2], Petra M. Nederlof[4], Marcel J.T. Reinders[6], Lodewyk F.A. Wessels[2,6], and Marc J. van de Vijver[5]

Research Article

Deletion of Chromosome 11q Predicts Response to Anthracycline-Based Chemotherapy in Early Breast Cancer

Joan Climent,[1,2,3] Peter Dimitrow,[4] Jane Fridlyand,[3] Jose Palacios,[5] Reiner Siebert,[7] Donna G. Albertson,[3] Joe W. Gray,[3,5] Daniel Pinkel,[3] Ana Lluch,[2] and Jose A. Martinez-Climent[1]

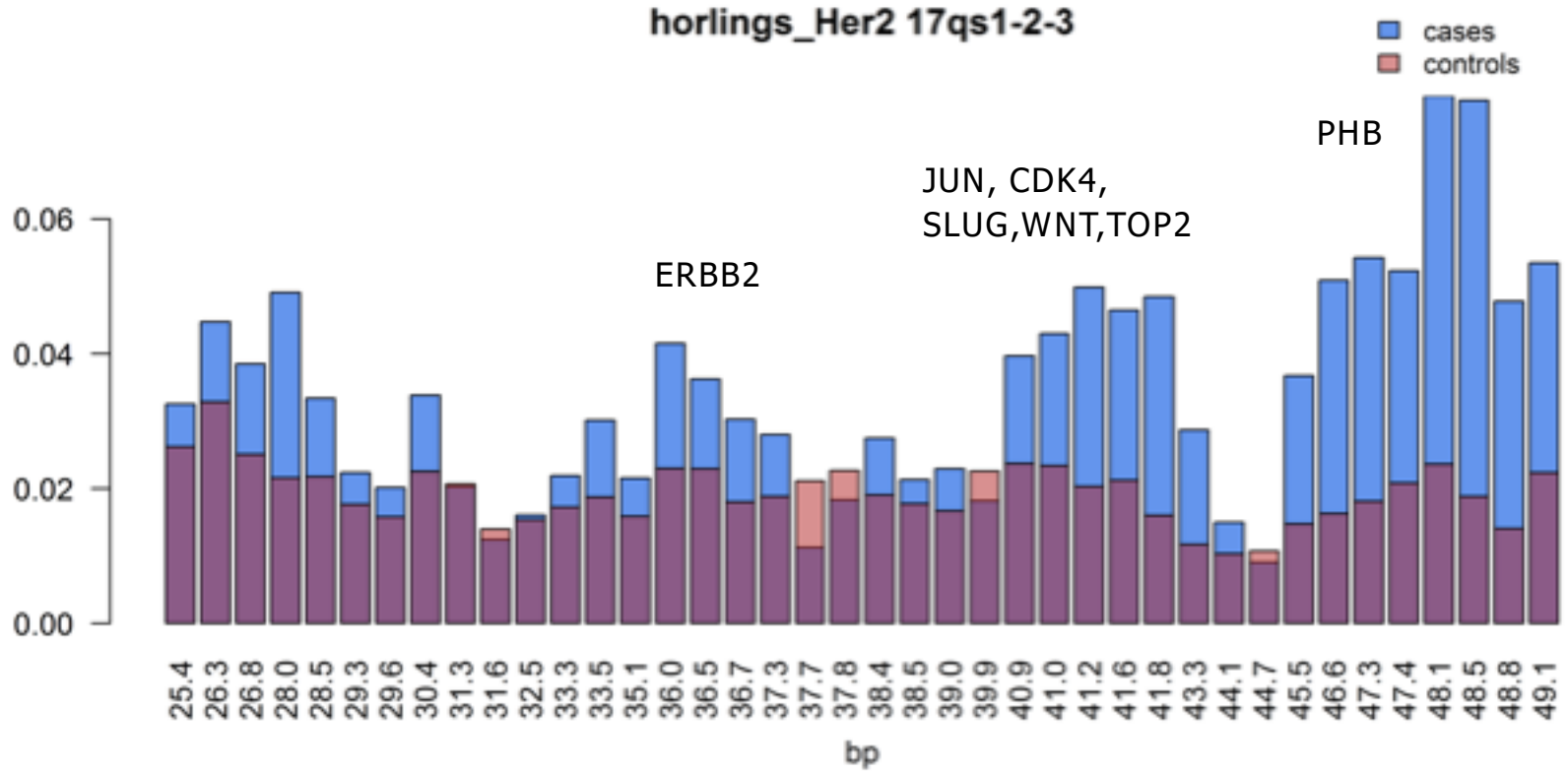
ARTICLE

doi:10.1038/nature11412

Comprehensive molecular portraits of human breast tumours

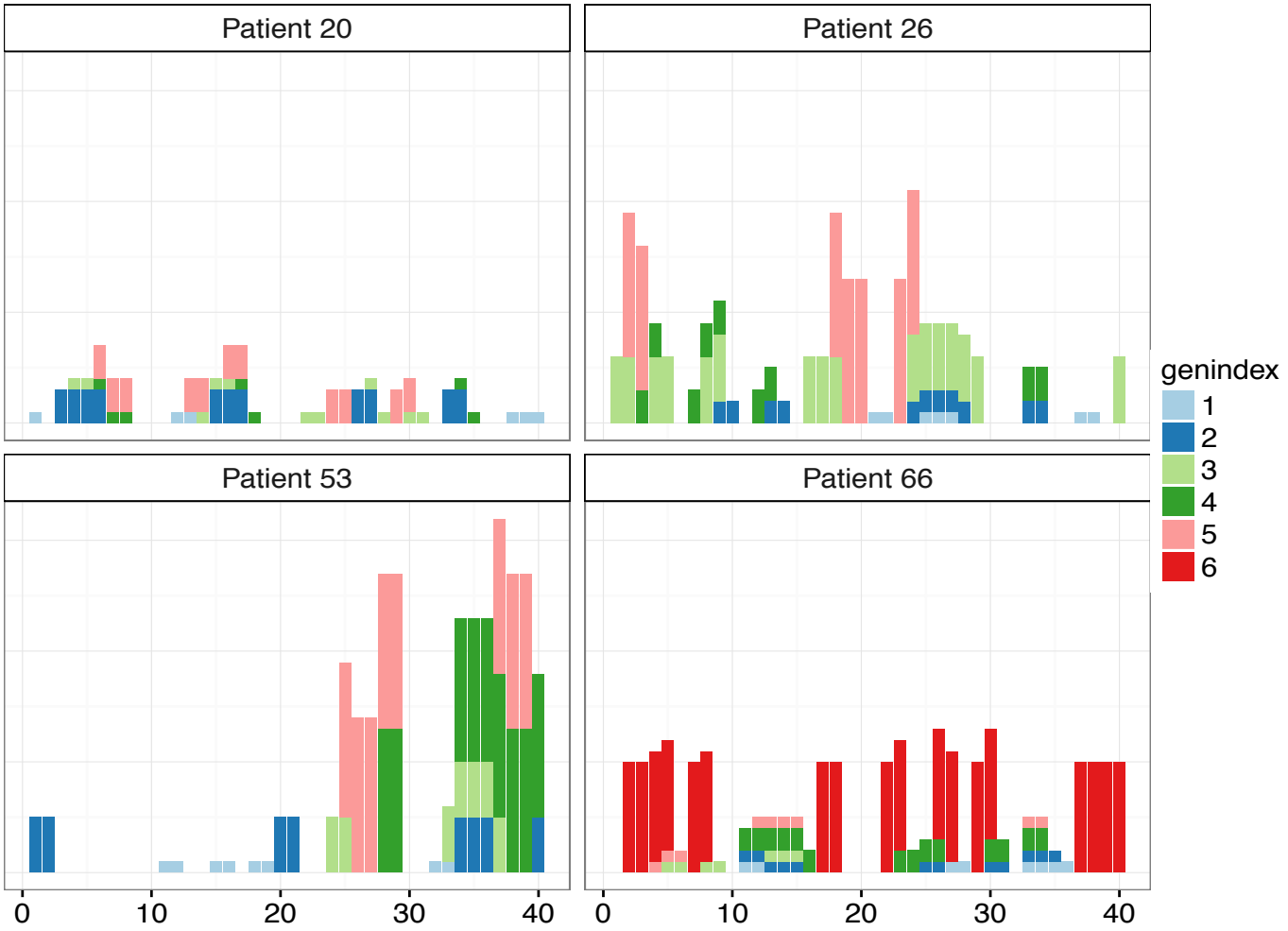
The Cancer Genome Atlas Network*

Co-amplifications detected in 17q across three different data sets

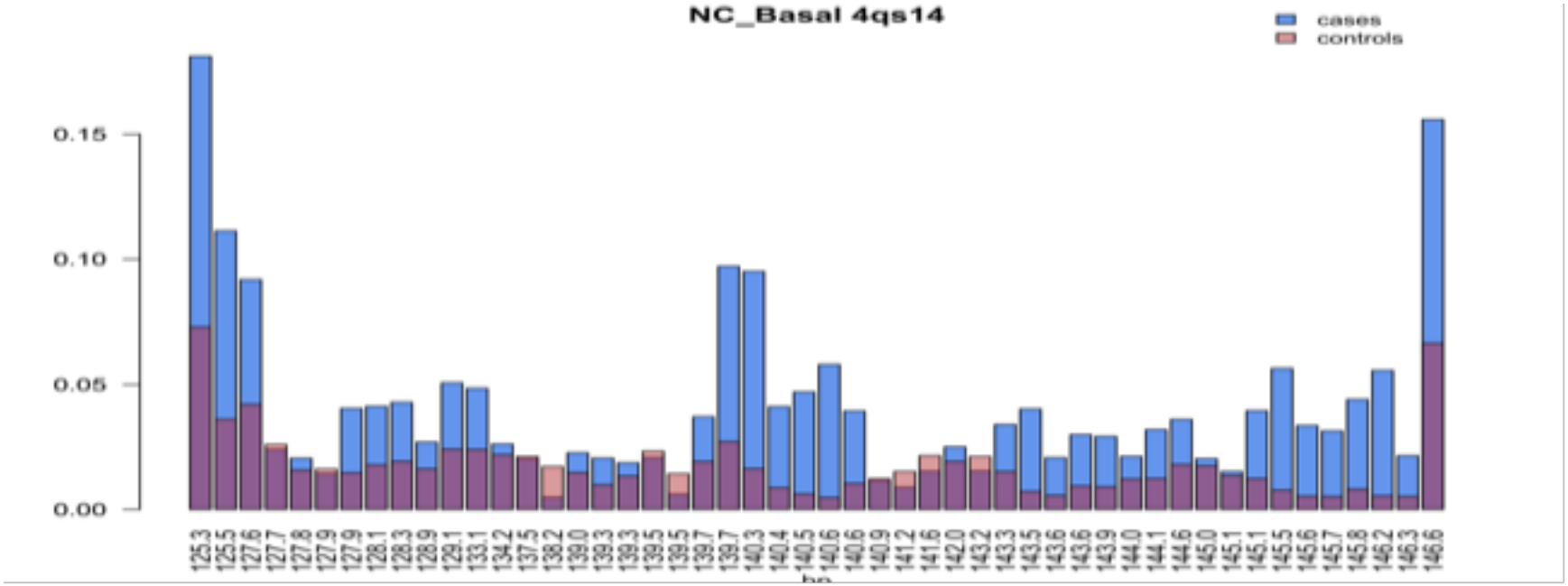


Ardanza et al. 2016

Generators are the product of co-amplifications not of single CNAs



TAaCGH suggests an interaction in 4q for basals



Conclusions and future research

- We are expanding TAaCGH to identify genetic interactions
- Genetic interactions are in the form of co-occurring copy number changes and/or the finer structure of copy number changes.
- In the ERBB2+ subtype we find co-expression of different regions of 17q. In Basal we find co-expression in 4q
- Next:
- Identify whether gene expression is regulated by these profiles
- Generalize to other situations in statistical genetics**: topological genetics**?

Thank you