Testing Structural Variants

by Next-Generation Sequencing

Richard Redon

richard.redon@inserm.fr

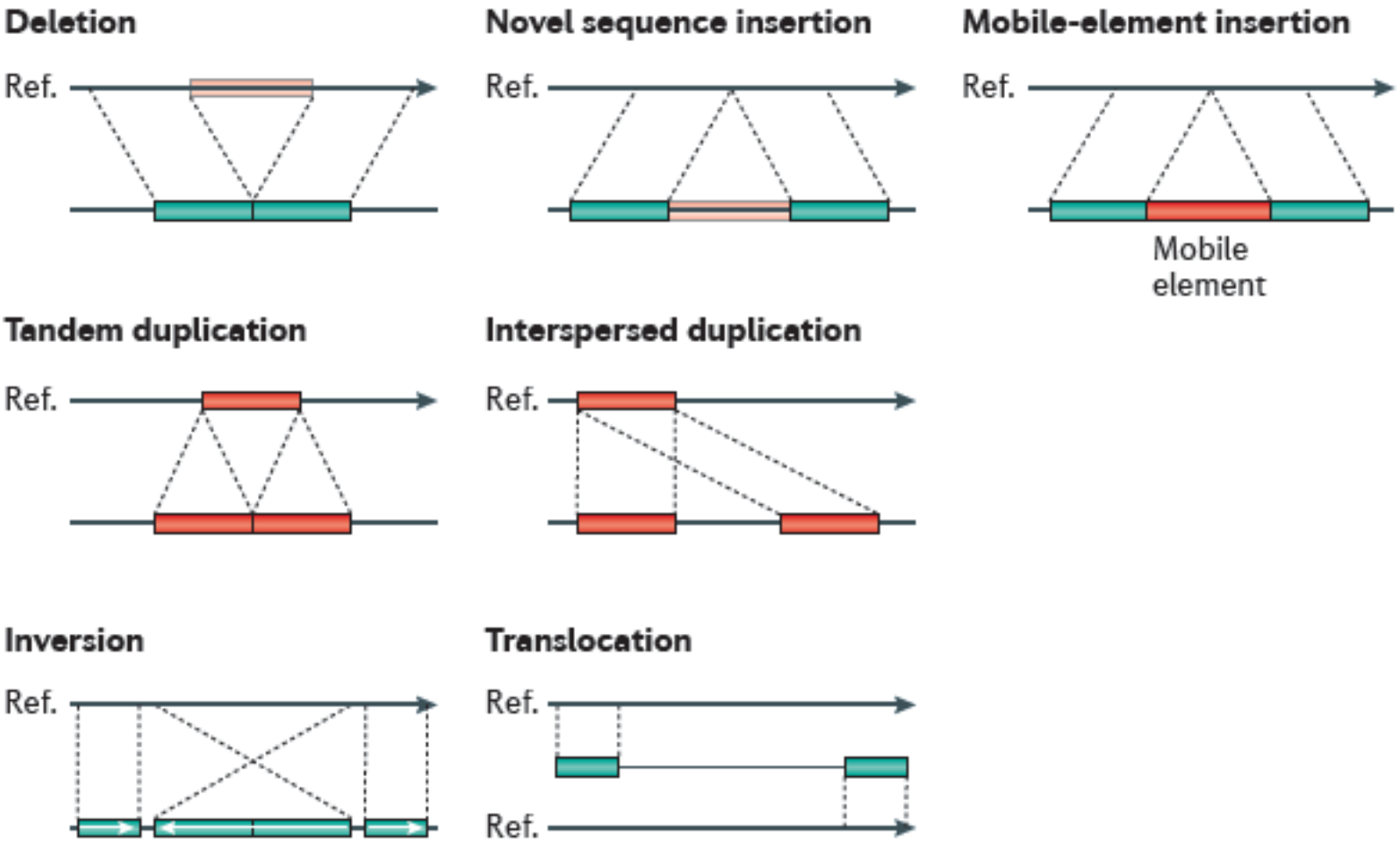
UMR Inserm 1087 CNRS 6291
Nantes

Inserm cnrs CENTRE HOSPITALIER UNIVERSITAIRE DE NANTES UNIVERSITÉ DE NANTES

Genome structural variation discovery and genotyping

Can Alkan‡, Bradley P. Coe* and Evan E. Eichler*‡*

Classes of structural variation



Structural Variation:

another source of genetic diversity

identified using microarray-based technologies

Detection of large-scale variation in the human genome

A John Iafrate[1,2], Lars Feuk[3], Miguel N Rivera[1,2], Marc L Listewnik[1], Patricia K Donahoe[2,4], Ying Qi[3], Stephen W Scherer[3,5] & Charles Lee[1,2,5]

2004

Large-Scale Copy Number Polymorphism in the Human Genome

Jonathan Sebat,[1] B. Lakshmi,[1] Jennifer Troge,[1] Joan Alexander,[1] Janet Young,[2] Pär Lundin,[3] Susanne Månér,[3] Hillary Massa,[2] Megan Walker,[2] Maoyen Chi,[1] Nicholas Navin,[1] Robert Lucito,[1] John Healy,[1] James Hicks,[1] Kenny Ye,[4] Andrew Reiner,[1] T. Conrad Gilliam,[5] Barbara Trask,[2] Nick Patterson,[6] Anders Zetterberg,[3] Michael Wigler[1]*

and high-throughput sequencing strategies

Fine-scale structural variation of the human genome

Eray Tuzun[1,5], Andrew J Sharp[1,5], Jeffrey A Bailey[2,5], Rajinder Kaul[3], V Anne Morrison[1], Lisa M Pertz[2], Eric Haugen[3], Hillary Hayden[3], Donna Albertson[4], Daniel Pinkel[4], Maynard V Olson[3] & Evan E Eichler[1]

2005

Detection of large-scale variation in the human genome

A John Iafrate[1,2], Lars Feuk[3], Miguel N Rivera[1,2], Marc L Listewnik[1], Patricia K Donahoe[2,4], Ying Qi[3], Stephen W Scherer[3,5] & Charles Lee[1,2,5]

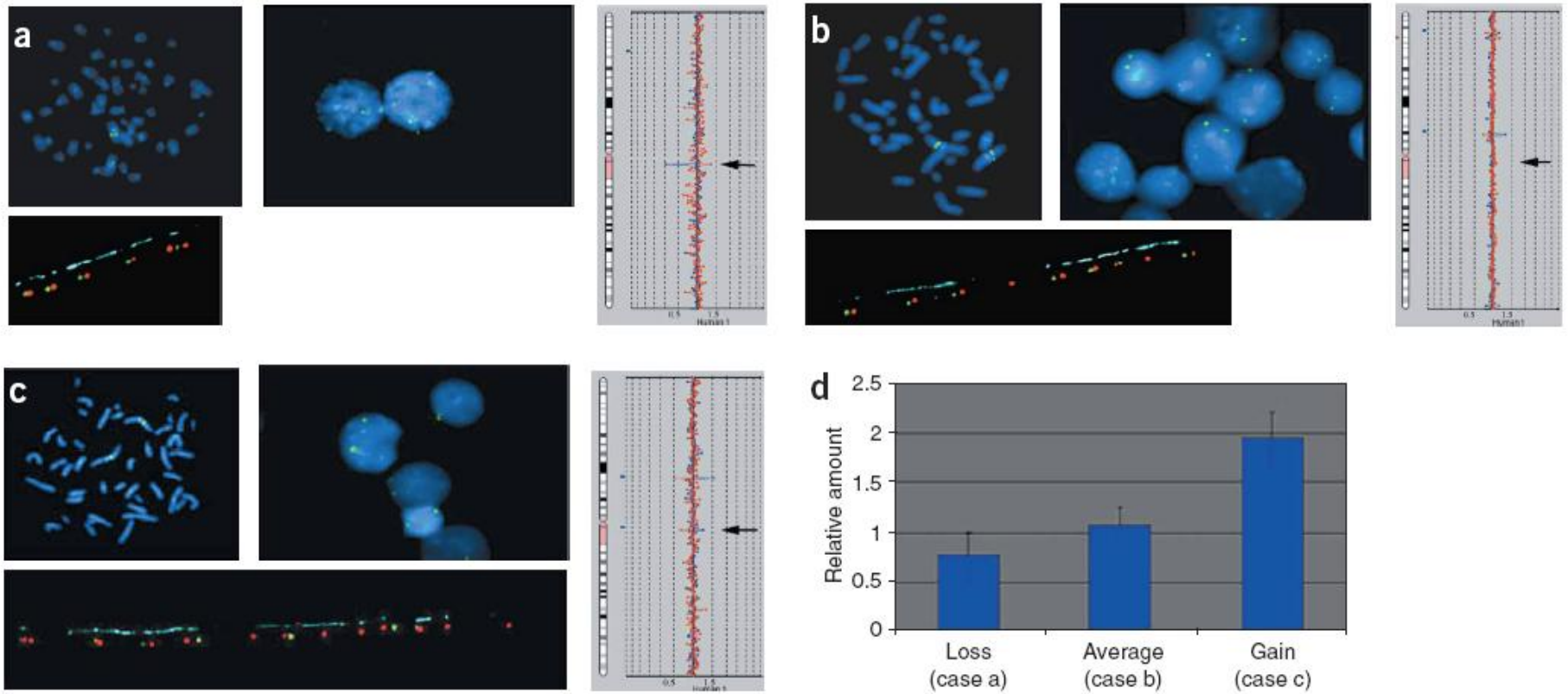
Analysis of 55 individuals using array-CGH at 1-Mb resolution

CNVs detected in 255 unique regions

24 of them are found in >10% individuals

12.4 CNVs per individual

The most common CNV:
variable number of tandem repeats in amylase genes

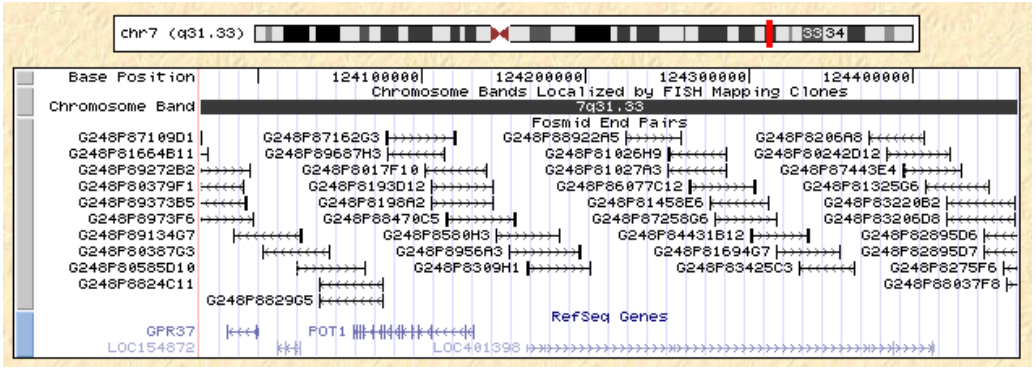


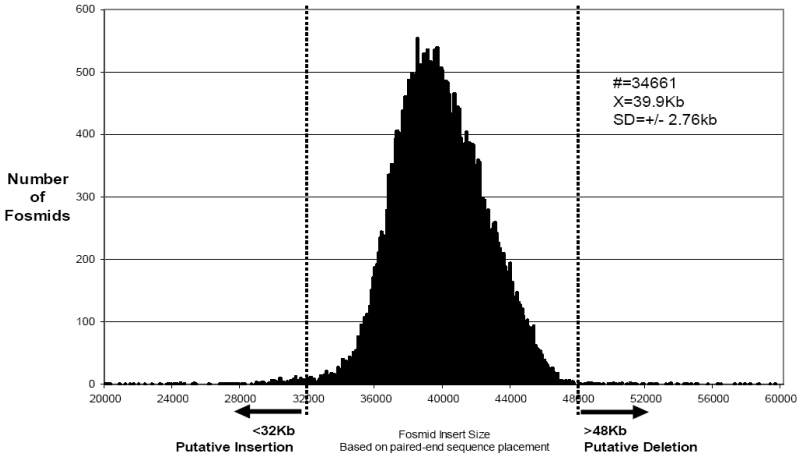
Fine-scale structural variation of the human genome

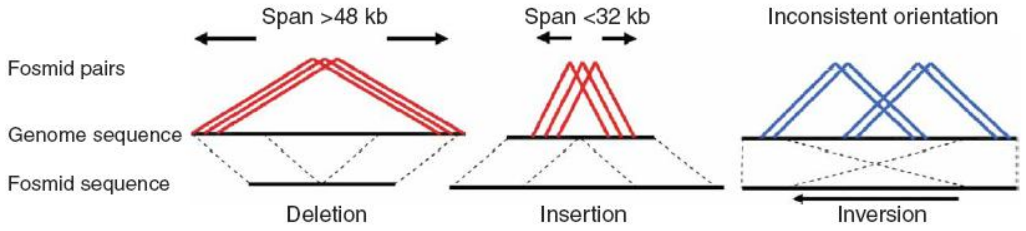
Eray Tuzun[1,5], Andrew J Sharp[1,5], Jeffrey A Bailey[2,5], Rajinder Kaul[3], V Anne Morrison[1], Lisa M Pertz[2], Eric Haugen[3], Hillary Hayden[3], Donna Albertson[4], Daniel Pinkel[4], Maynard V Olson[3] & Evan E Eichler[1]

297 sites of structural variation identified
in the genome of one single individual
when compared to the reference human genome sequence

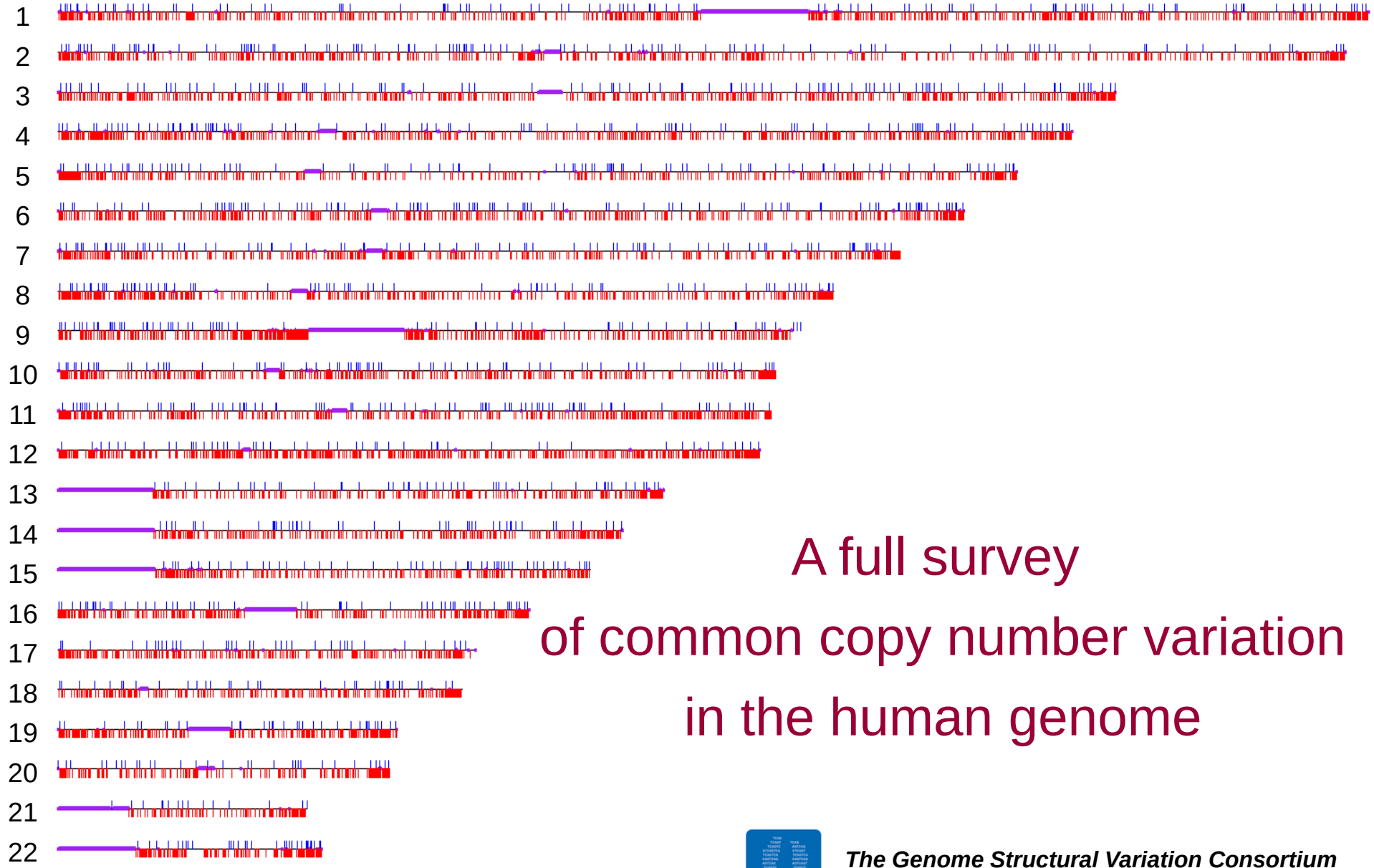
(size: order of 10 kb in size)







n=139 n=102 n=56



A full survey of common copy number variation in the human genome

gaps

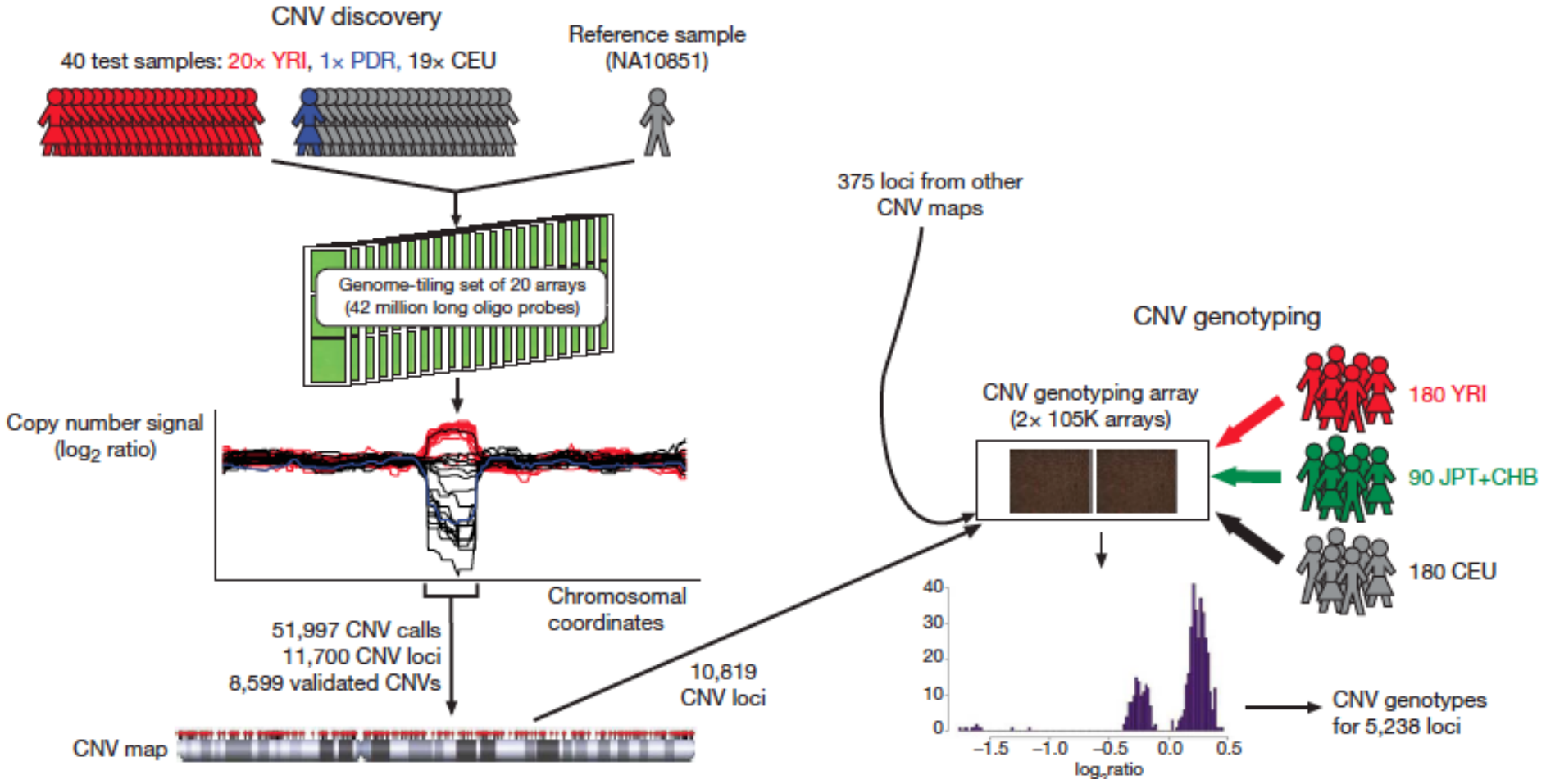
Redon *et al.*, *Nature*, Nov. 2006

Conrad et al., *Nature*, April 2010

The Genome Structural Variation Consortium

wellcome trust sanger institute

Origins and functional impact of copy number variation in the human genome



Nature 464:704-12, 1 April 2010

Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls

Nature 464:713-20, 1 April 2010

The Wellcome Trust Case Control Consortium*

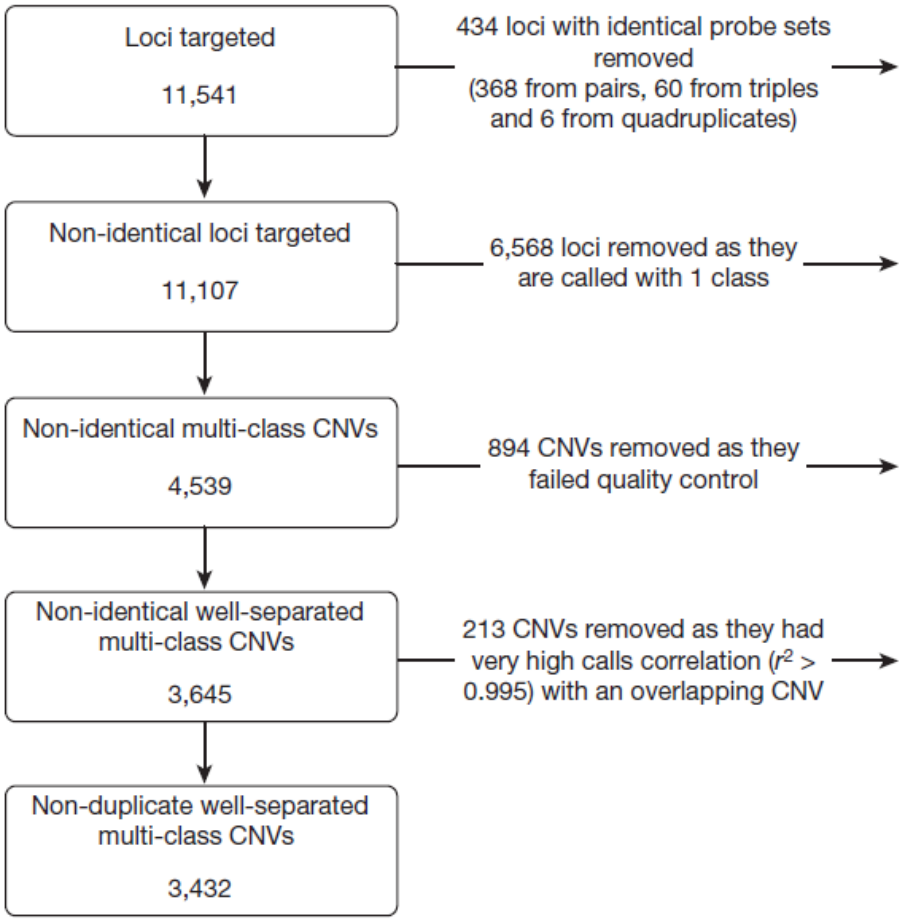


Figure 1 | Flowchart showing which CNVs are included on the array. The chart shows the reasons for CNVs being removed from consideration (the column of arrows and text to the right of the figure) from those originally targeted on the array, and the number of CNVs remaining at each stage of filtering.

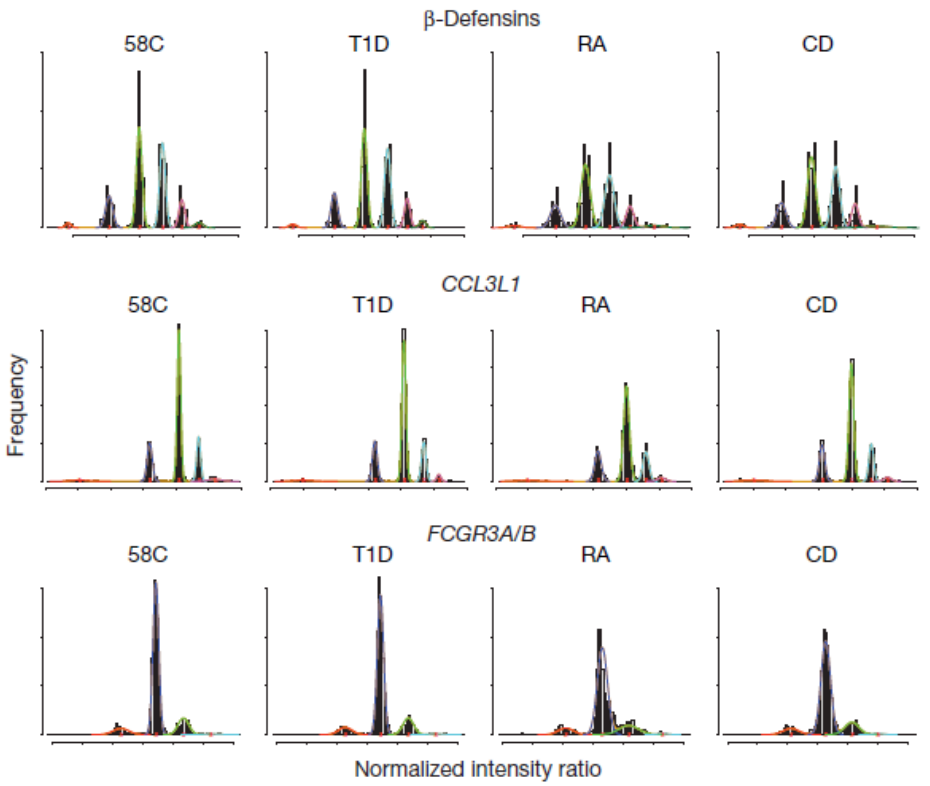
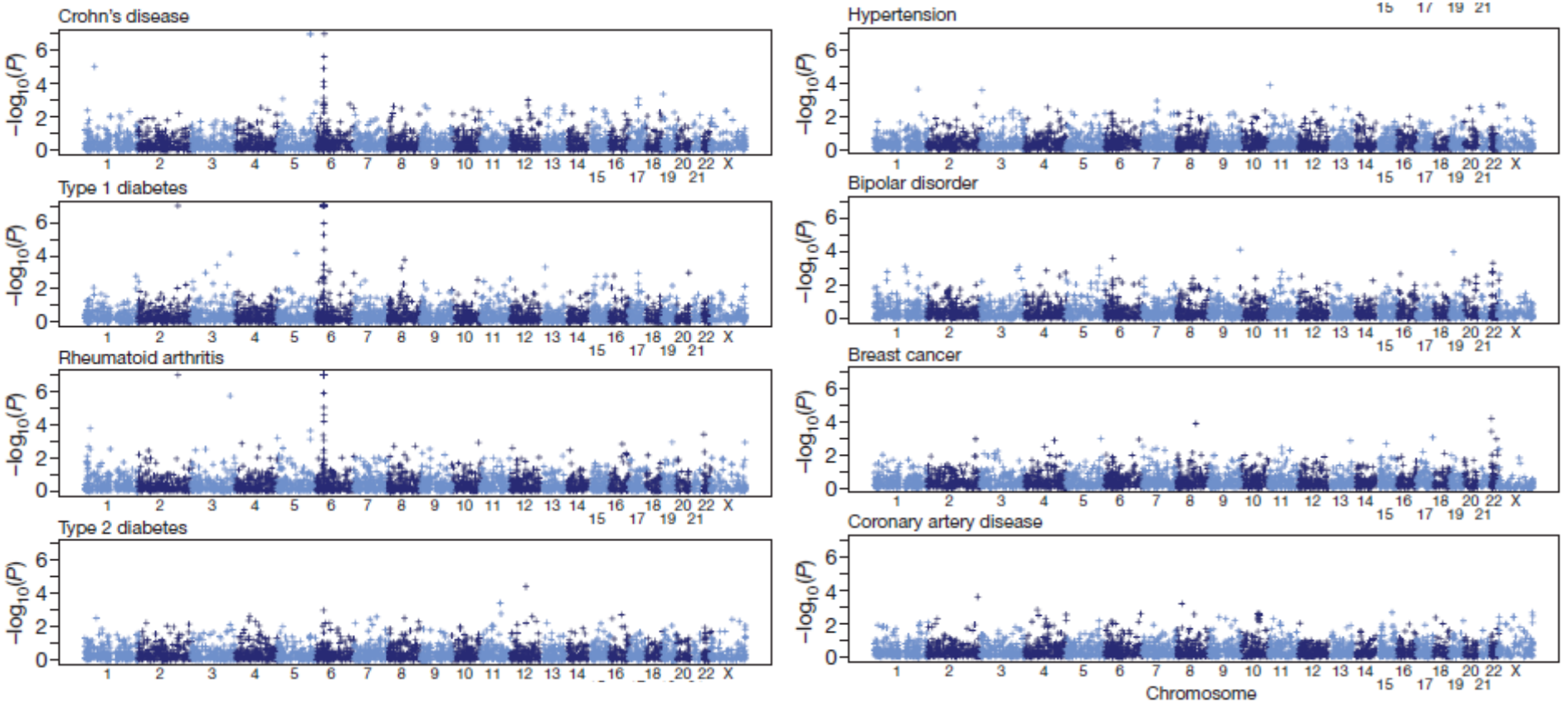


Figure 2 | Illustrative CNVs. Histograms of three multi-allelic CNVs (one per row) previously reported to be associated with autoimmune diseases: β-defensin (CNVR3771.10), *CCL3L1* (CNVR7077.12) and *FCGR3A/B* (CNVR383.1), showing 6, 5 and 4 fitted copy number classes, respectively. The histogram of normalized intensity ratios is shown for one control and the three autoimmune collections. Histograms are overlaid by the fitted distribution used to model each class (variously the red, blue, light-green, cyan, magenta and dark green curves). In all such figures, the area under the fitted curve of a particular colour is the same for all collections at the same CNV. 58C, 1958 British Birth Cohort; CD, Crohn's disease; RA, rheumatoid arthritis; T1D, type 1 diabetes.

Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls

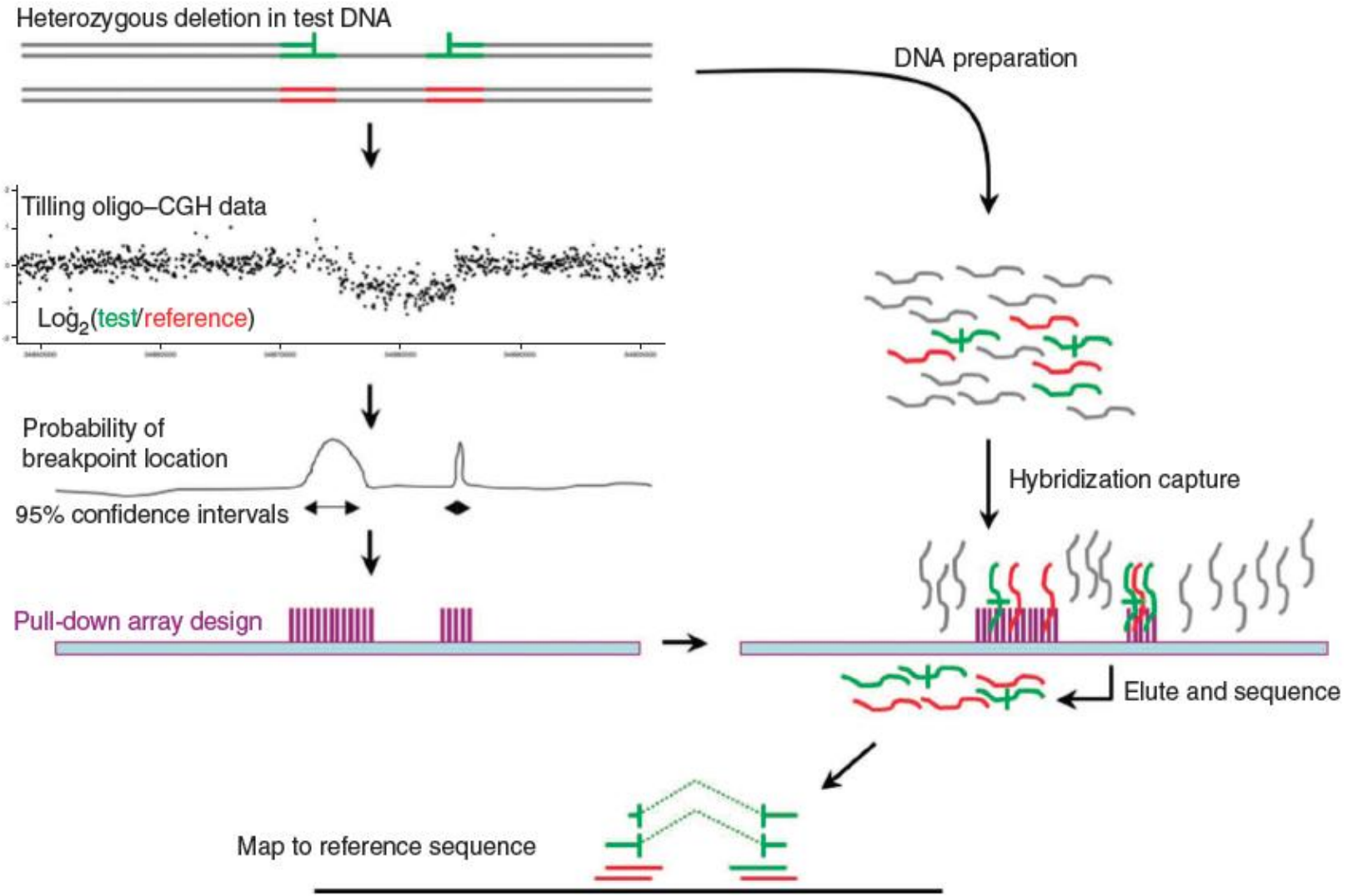
Nature 464:713-20, 1 April 2010

The Wellcome Trust Case Control Consortium*

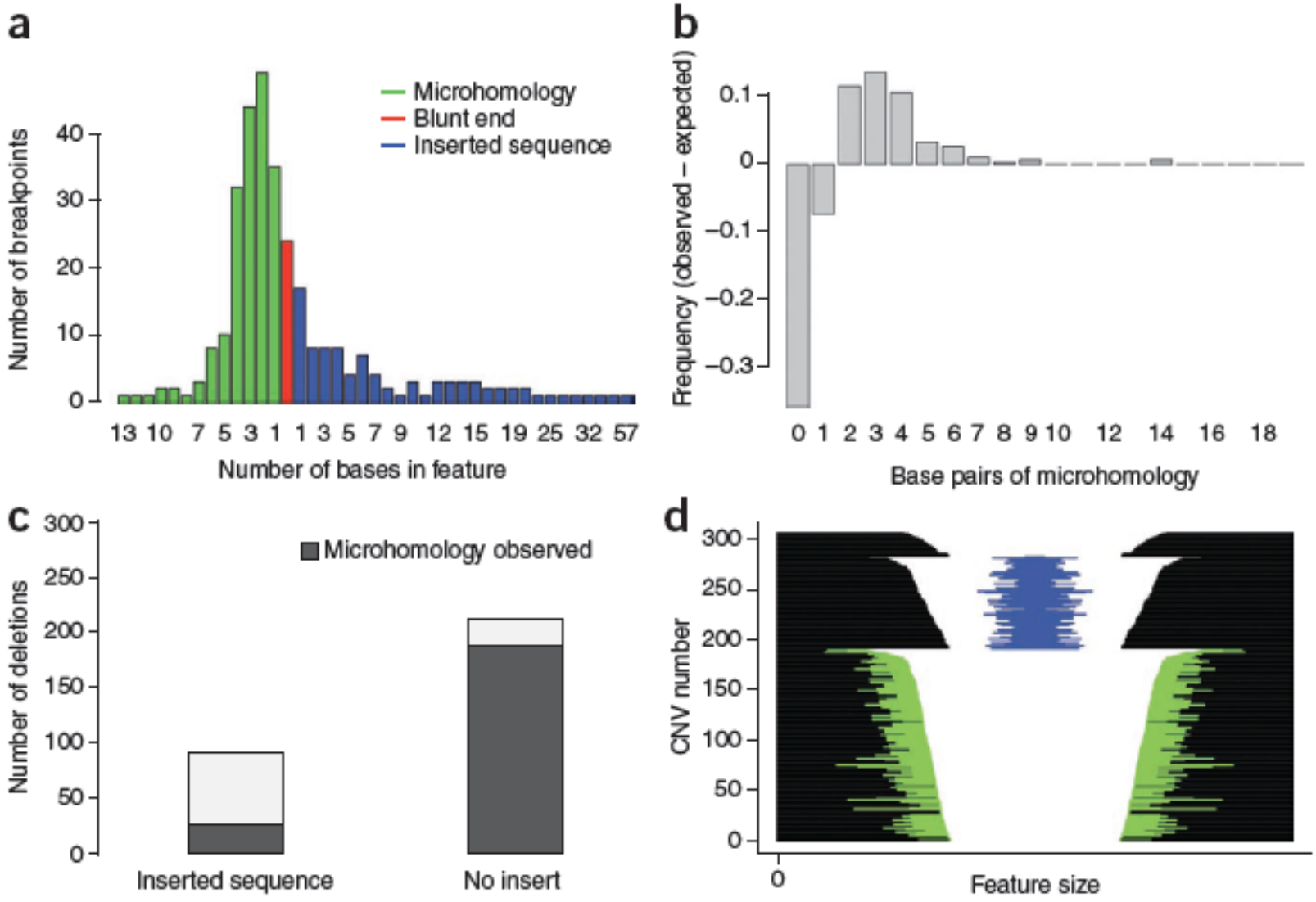


"…most common CNVs that are well-typed on our array are well tagged by SNPs and so have been indirectly explored through SNP studies. We conclude that common CNVs that can be typed on existing platforms are unlikely to contribute greatly to the genetic basis of common human diseases."

Mutation spectrum revealed by breakpoint sequencing of human germline CNVs

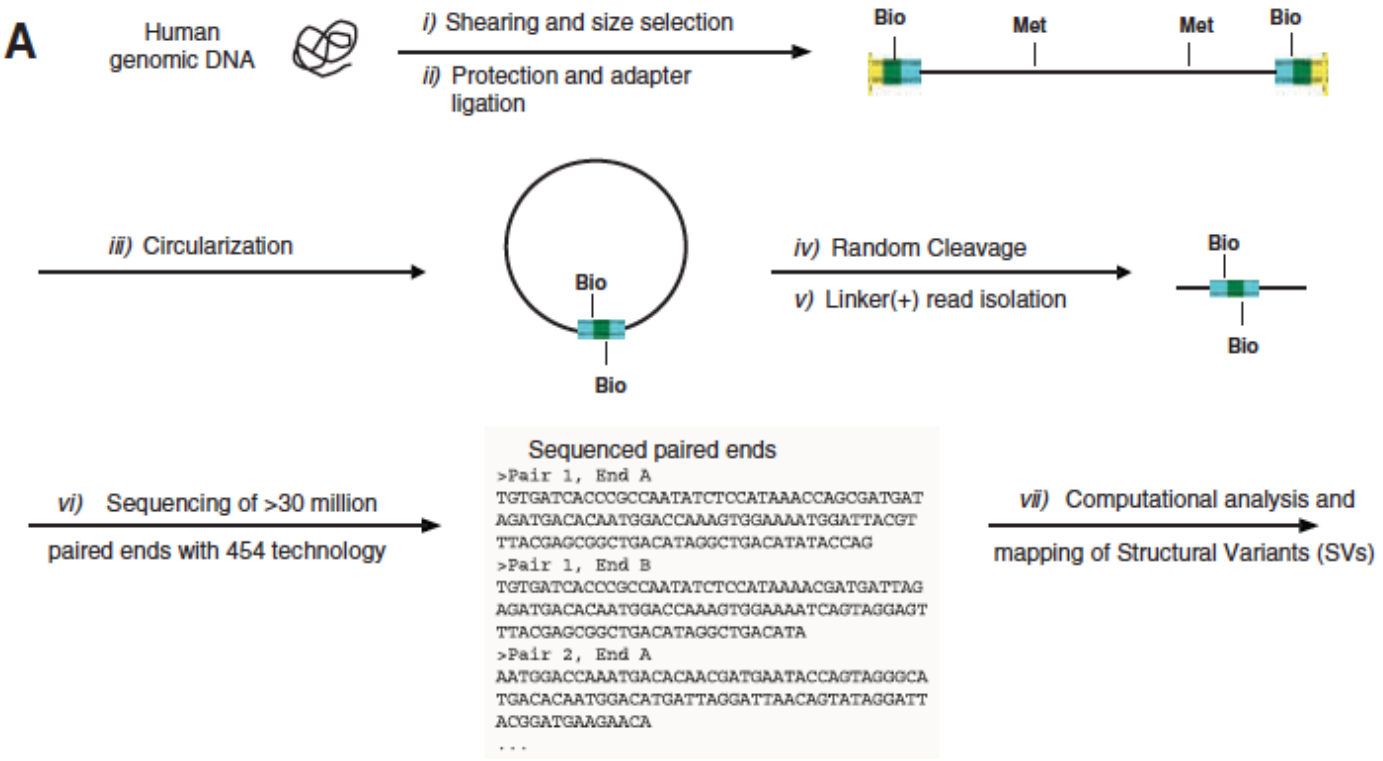


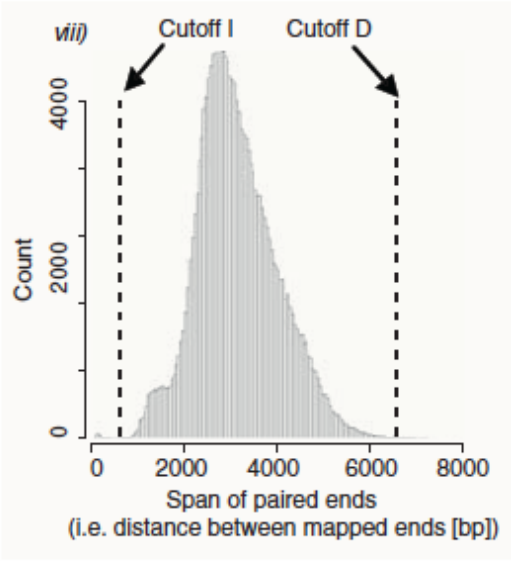
Mutation spectrum revealed by breakpoint sequencing of human germline CNVs



Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome

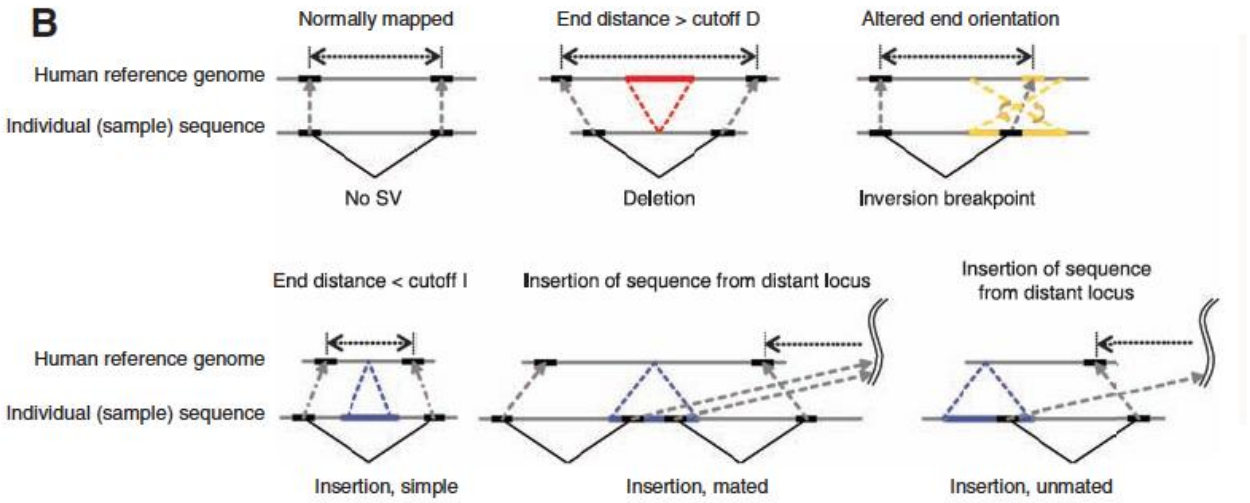
19 OCTOBER 2007 VOL 318 SCIENCE

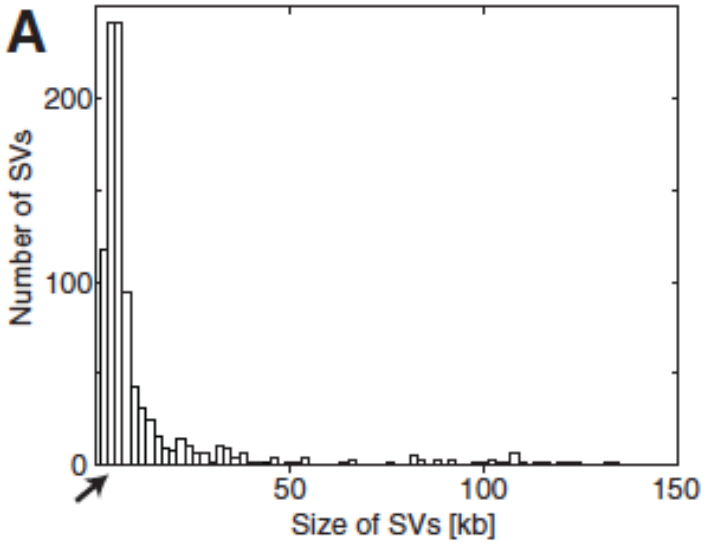




Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome

19 OCTOBER 2007 VOL 318 SCIENCE



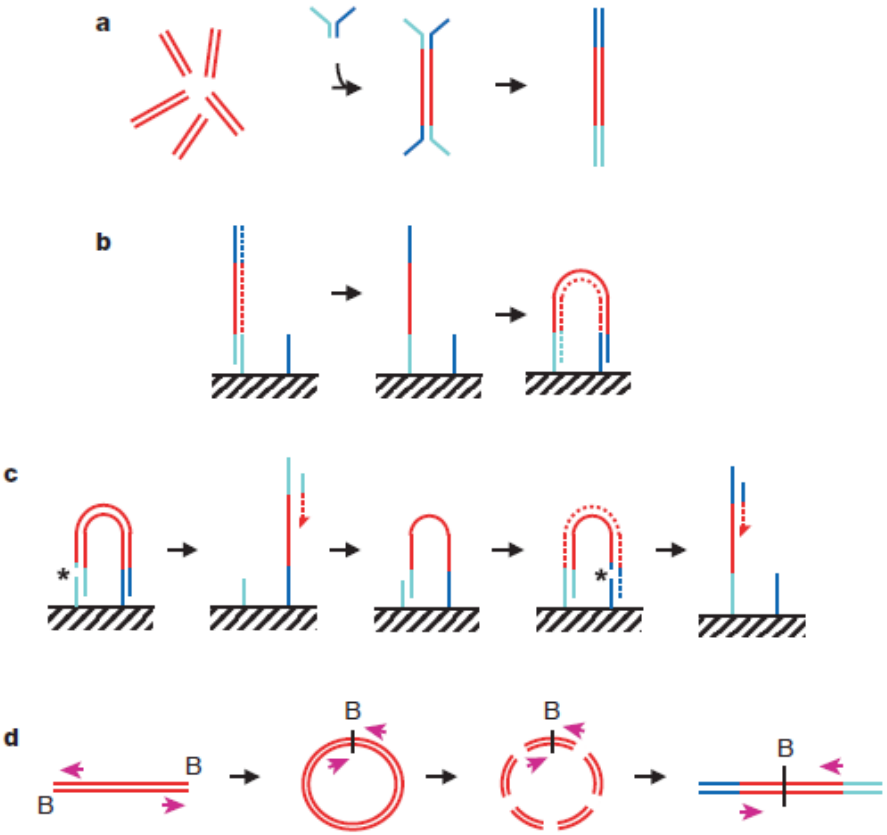


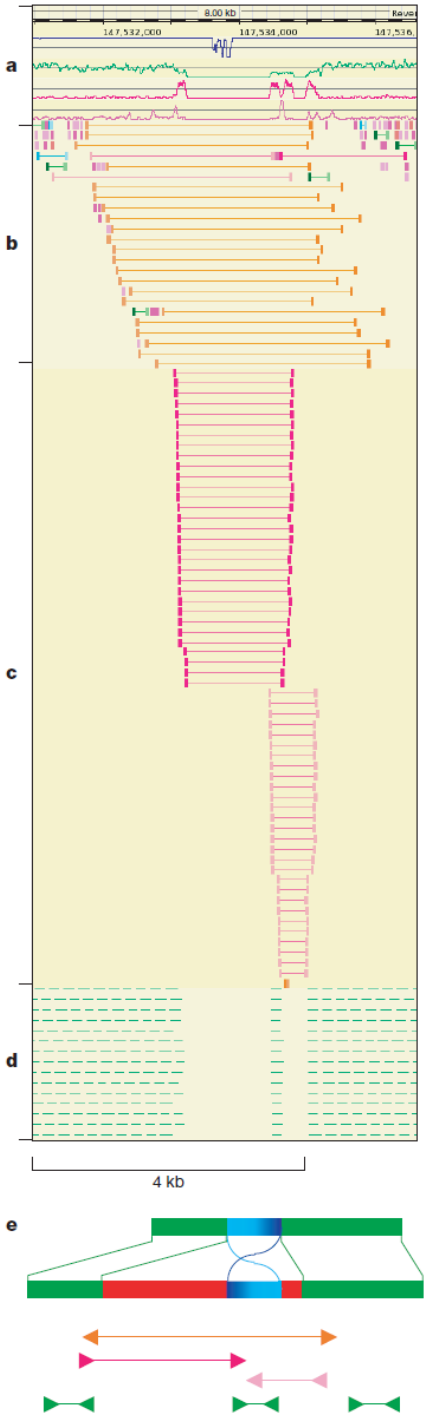
~1300 SVs identified in two individuals (mostly CNVs)

Vol 456 | 6 November 2008 | doi:10.1038/nature07517

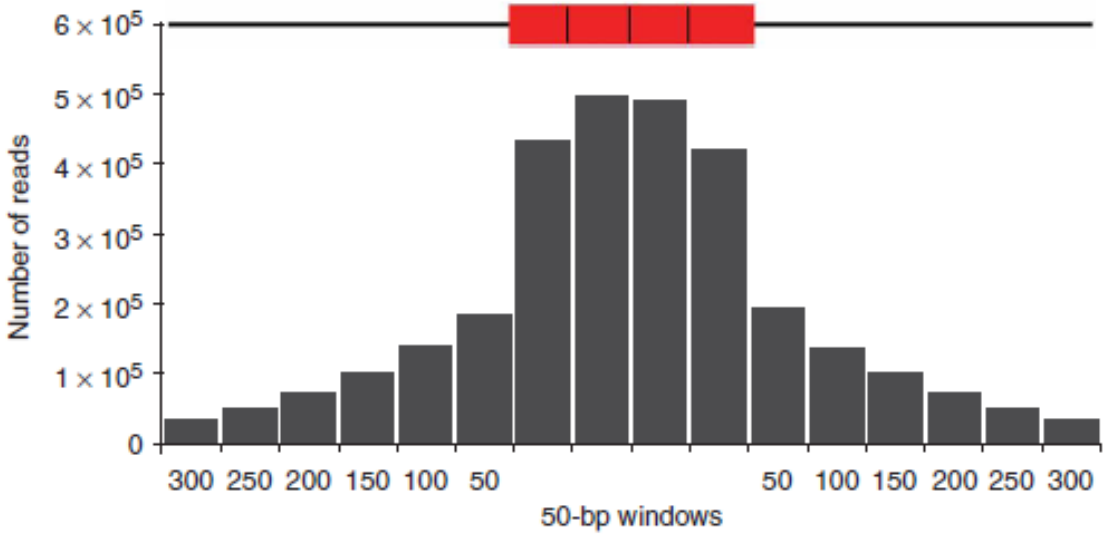
nature

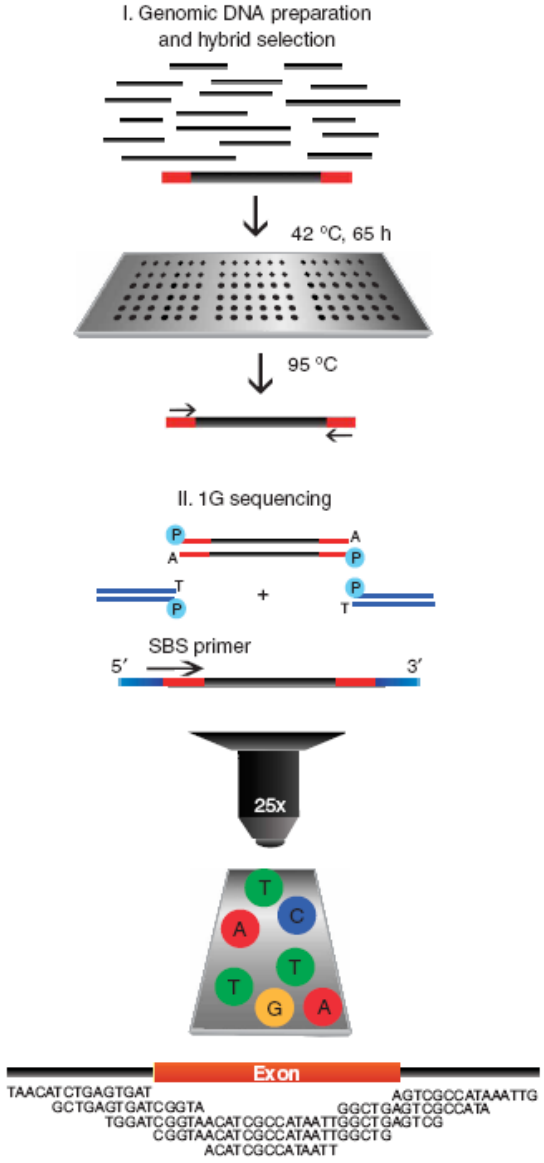
Accurate whole human genome sequencing using reversible terminator chemistry





Targeted sequencing: NGS at 'low-cost'





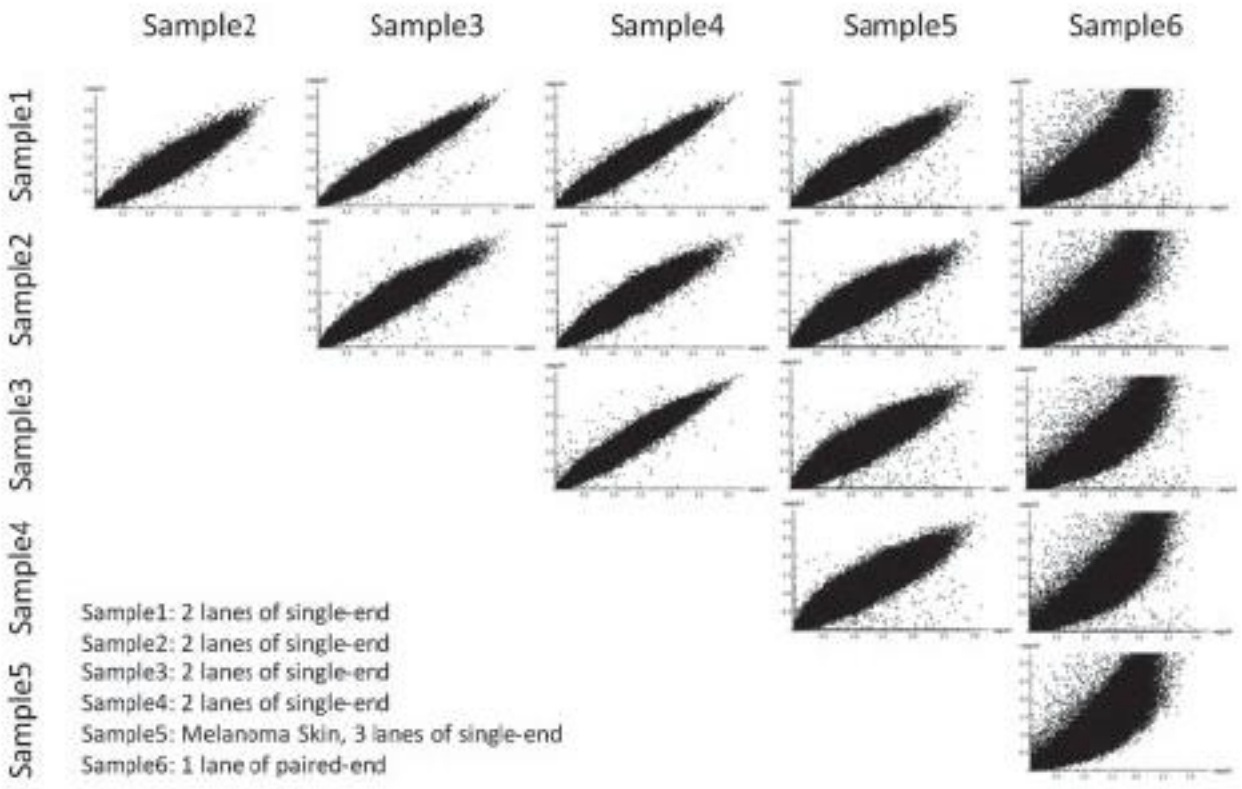
Nat Genet 2007;39:1522-7

Genome-wide *in situ* exon capture for selective resequencing

Emily Hodges[1,4], Zhenyu Xuan[1,2,4], Vivekanand Balija[2], Melissa Kramer[2], Michael N Molla[3], Steven W Smith[3], Christina M Middle[3], Matthew J Rodesch[3], Thomas J Albert[3], Gregory J Hannon[1] & W Richard McCombie[2]

BIOINFORMATICS ORIGINAL PAPER
Vol. 27 no. 19 2011, pages 2648–2654
doi:10.1093/bioinformatics/btr462

Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV

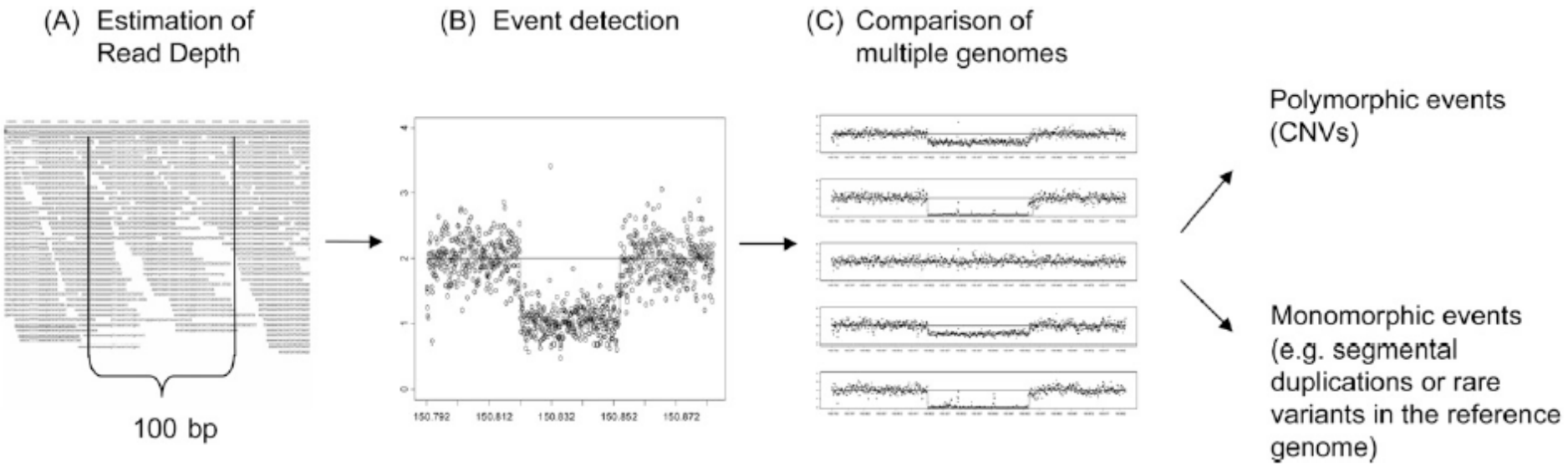


Correlation of depth-of-coverage across exome sequencing samples.

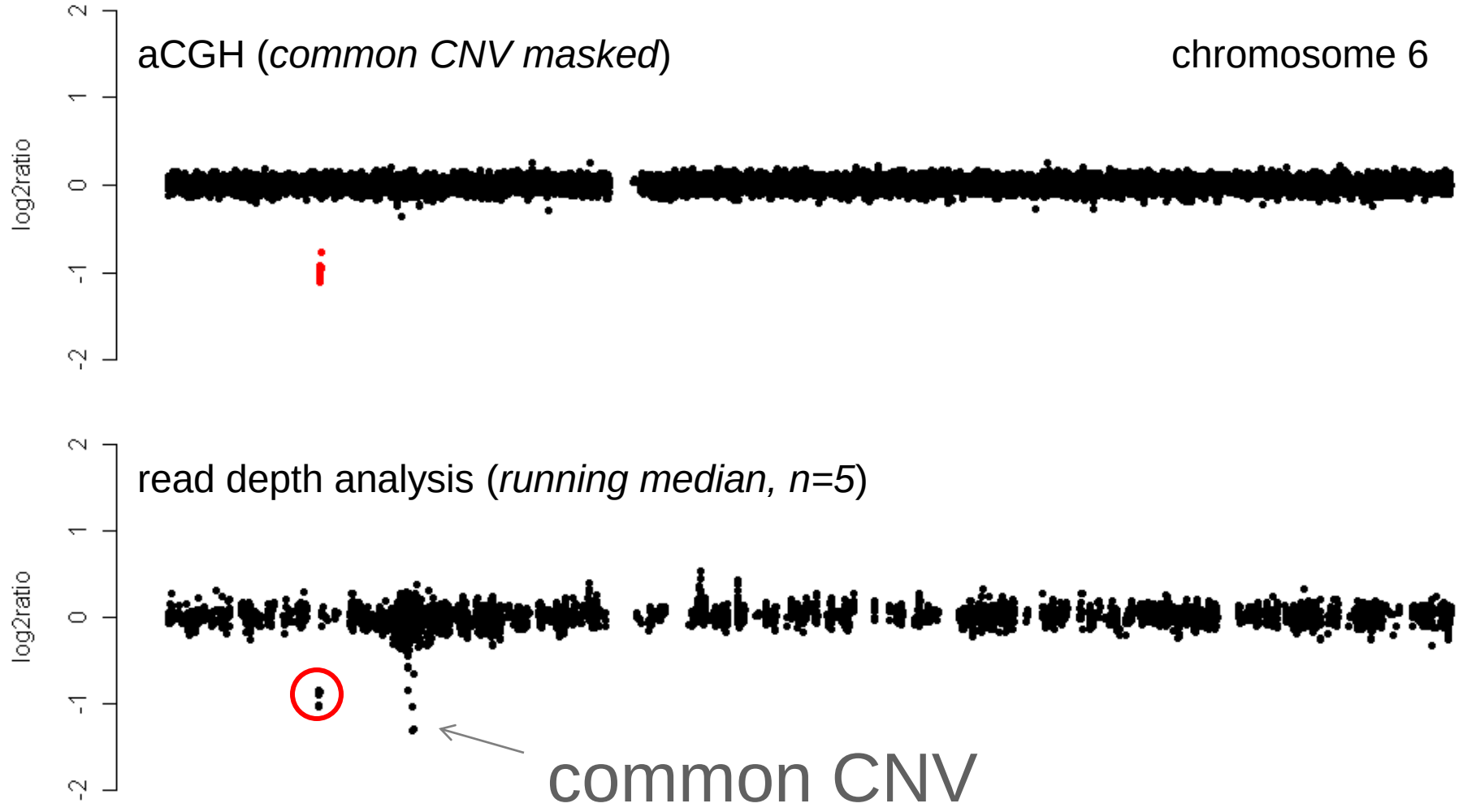
Sensitive and accurate detection of copy number variants using read depth of coverage

Seungtai Yoon,[1] Zhenyu Xuan,[1] Vladimir Makarov,[1] Kenny Ye,[2,3] and Jonathan Sebat[1]

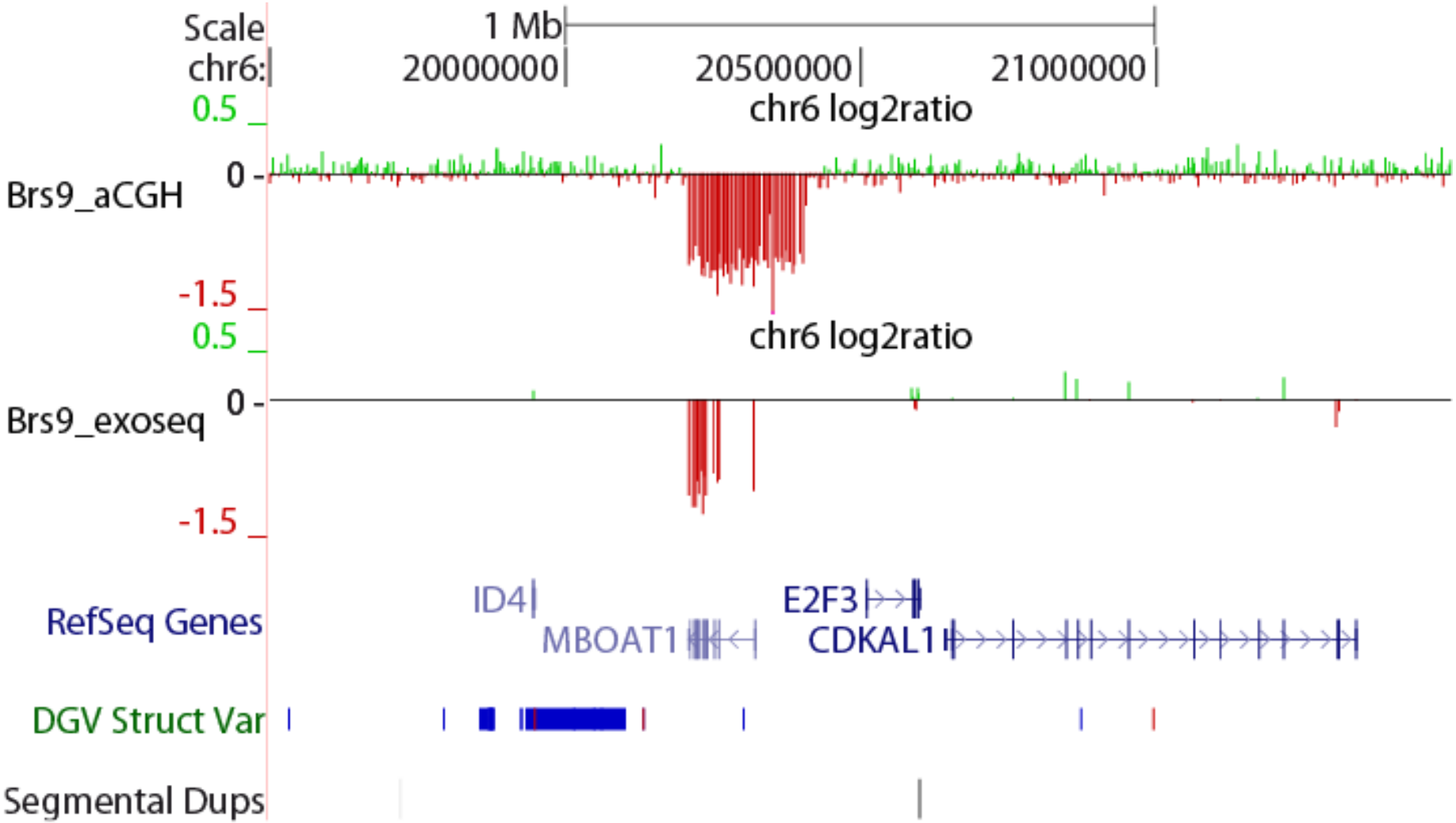
Genome Research
www.genome.org
19:1586–1592 © 2009



CNV detection by exome sequencing



CNV detection by exome sequencing



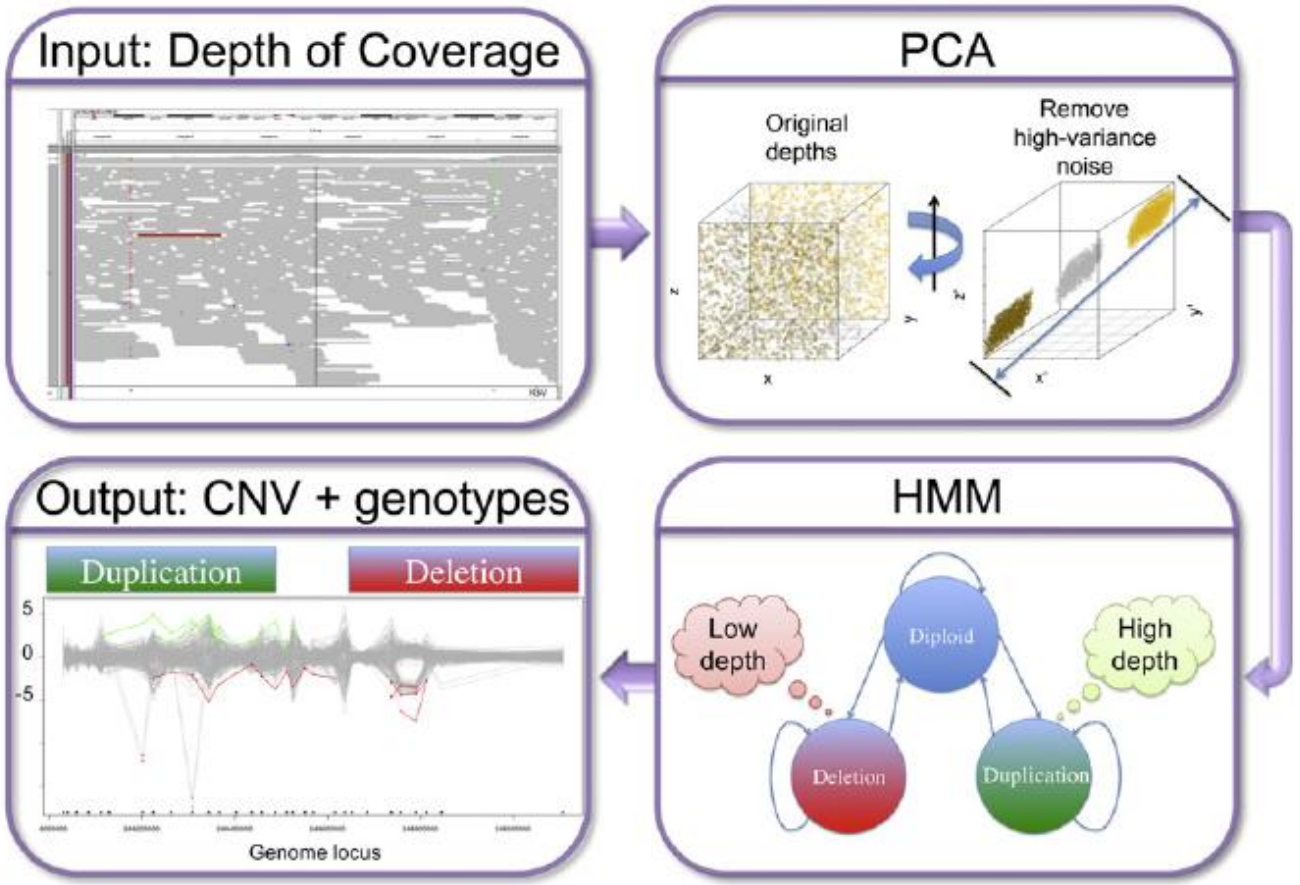
l'institut du thorax

Inserm cnrs CENTRE HOSPITALIER UNIVERSITAIRE DE NANTES UNIVERSITÉ DE NANTES

Discovery and Statistical Genotyping of Copy-Number Variation from Whole-Exome Sequencing Depth

Menachem Fromer,[1,2,3,4,5,*] Jennifer L. Moran,[2] Kimberly Chambert,[2] Eric Banks,[3] Sarah E. Bergen,[2,5] Douglas M. Ruderfer,[1,2,4,5] Robert E. Handsaker,[3,6] Steven A. McCarroll,[2,3,6] Michael C. O'Donovan,[7] Michael J. Owen,[7] George Kirov,[7] Patrick F. Sullivan,[8,9] Christina M. Hultman,[9] Pamela Sklar,[1] and Shaun M. Purcell[1,2,3,4,5,*]

The American Journal of Human Genetics *91*, 597–607, October 5, 2012 597



BIOINFORMATICS

Vol. 28 ECCB 2012, pages i370–i374
doi:10.1093/bioinformatics/bts379

An exome sequencing pipeline for identifying and genotyping common CNVs associated with disease with application to psoriasis

Lachlan J.M. Coin[1,2,*], Dandan Cao[1], Jingjing Ren[1], Xianbo Zuo[3,4], Liangdan Sun[3,4], Sen Yang[3,4], Xuejun Zhang[3,4], Yong Cui[3,4], Yingrui Li[1], Xin Jin[1,5] and Jun Wang[1,*]

Table 3. CNV genotyping results at LCE3B_LCE3C locus

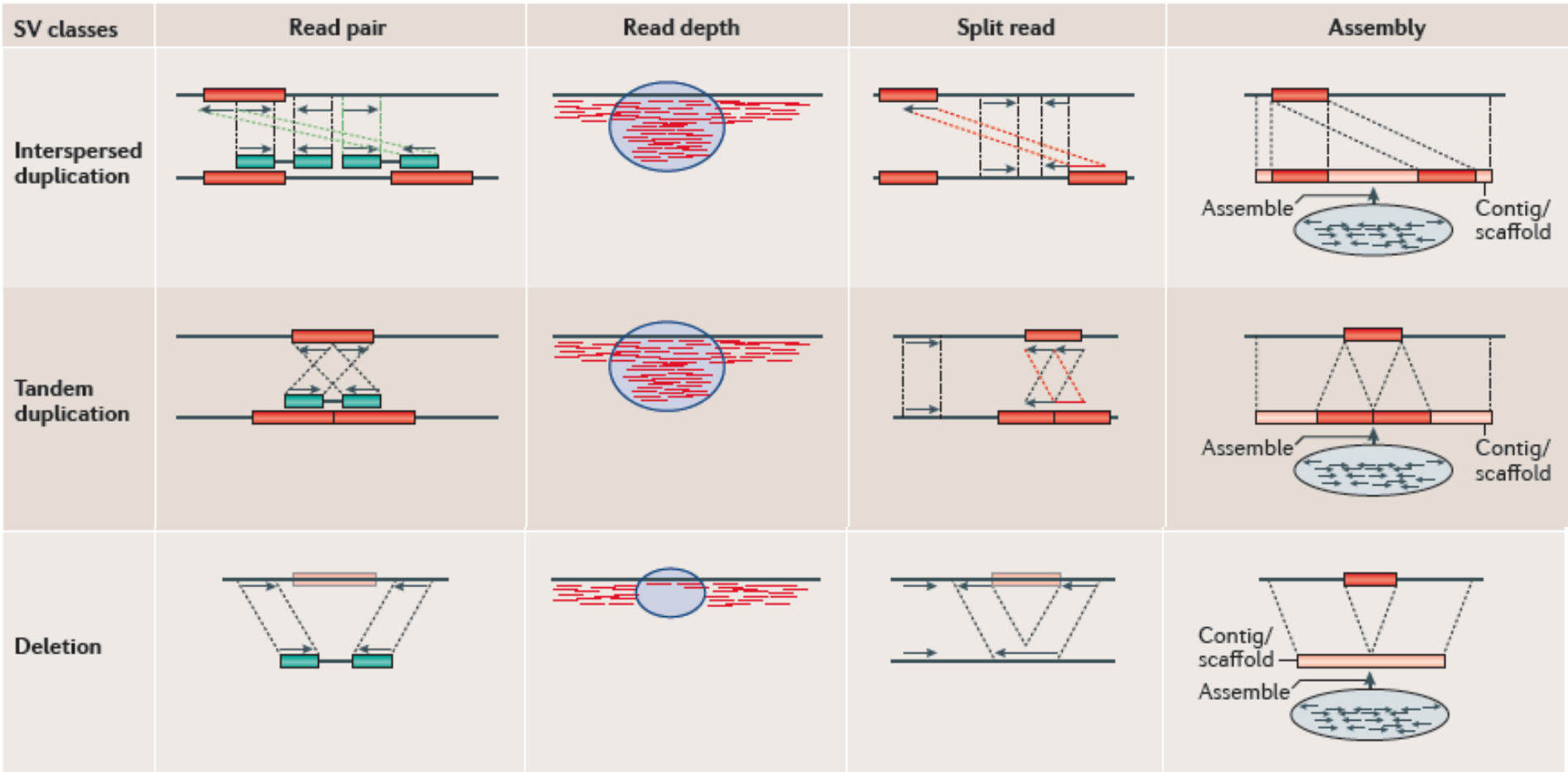
# GPCS	CNV genotype		
	P-value	Accuracy (%)	Missing rate (%)
0	$1.4e-4$	95.8	15.7
5	$1.1e-5$	97.1	14.4
20	$2.7e-6$	97.7	13.3
40	$5.0e-6$	97.4	14.7
50	$1.7e-5$	95.8	14.3

CNV genotyping accuracy, missing rates and association with psoriasis status after correction for GPCs.

Genome structural variation discovery and genotyping

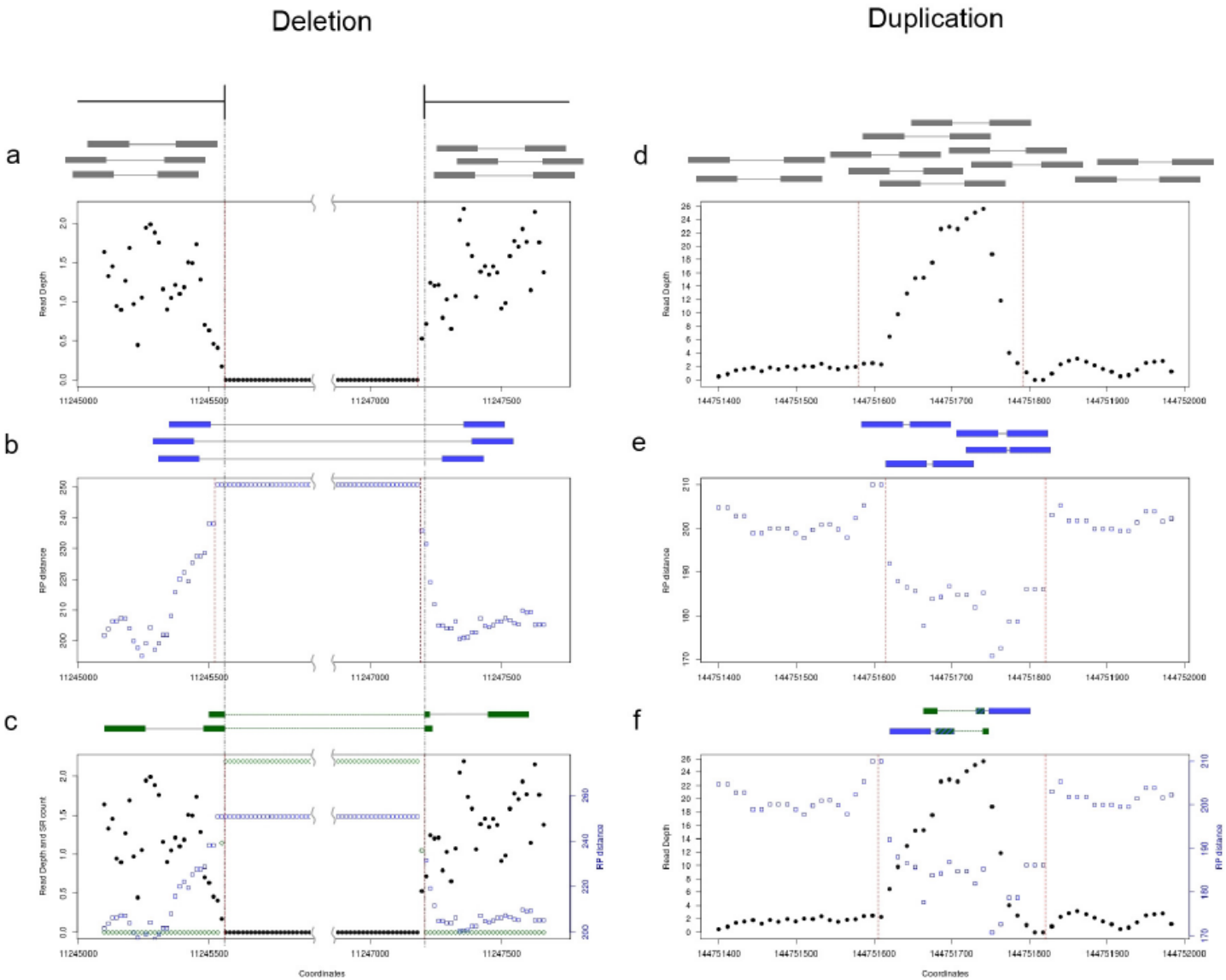
Can Alkan‡, Bradley P. Coe* and Evan E. Eichler*‡*

Classes of structural variation => Copy-Number Variants



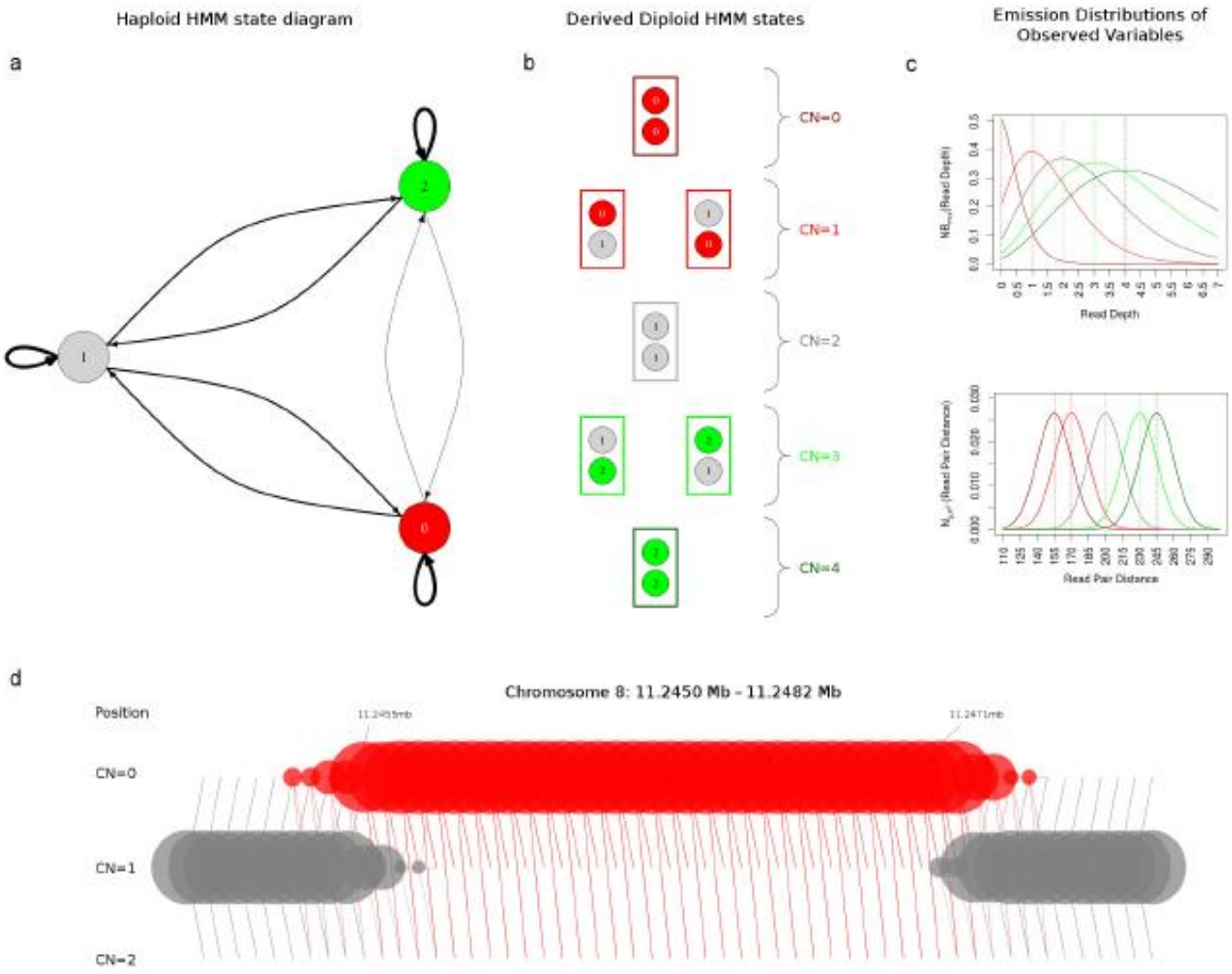
cnvHiTSeq: integrative models for high-resolution copy number variation detection and genotyping using population sequencing data

Bellos *et al. Genome Biology* 2012, **13**:R120
http://genomebiology.com/2013/13/12/R120

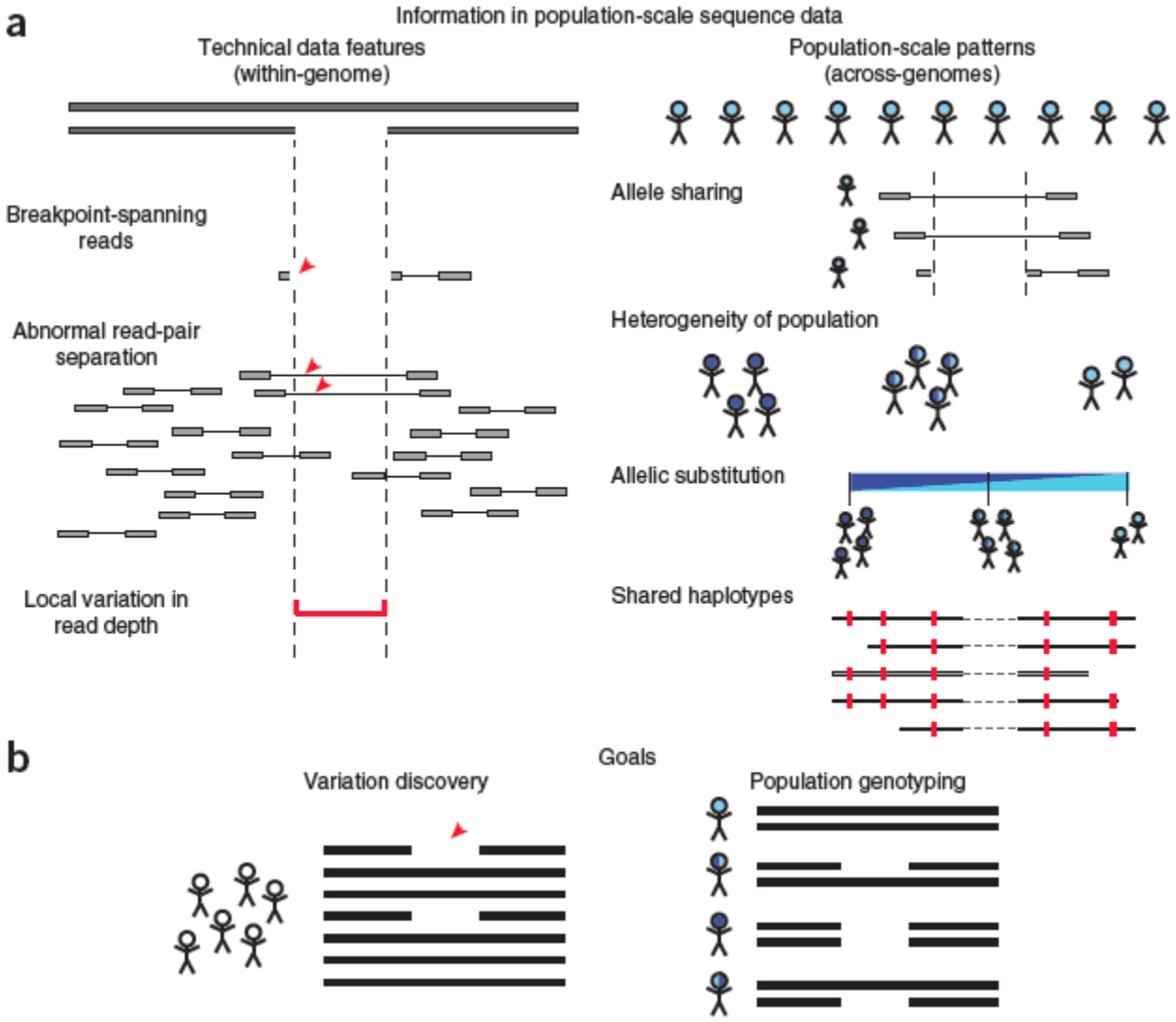


cnvHiTSeq: integrative models for high-resolution copy number variation detection and genotyping using population sequencing data

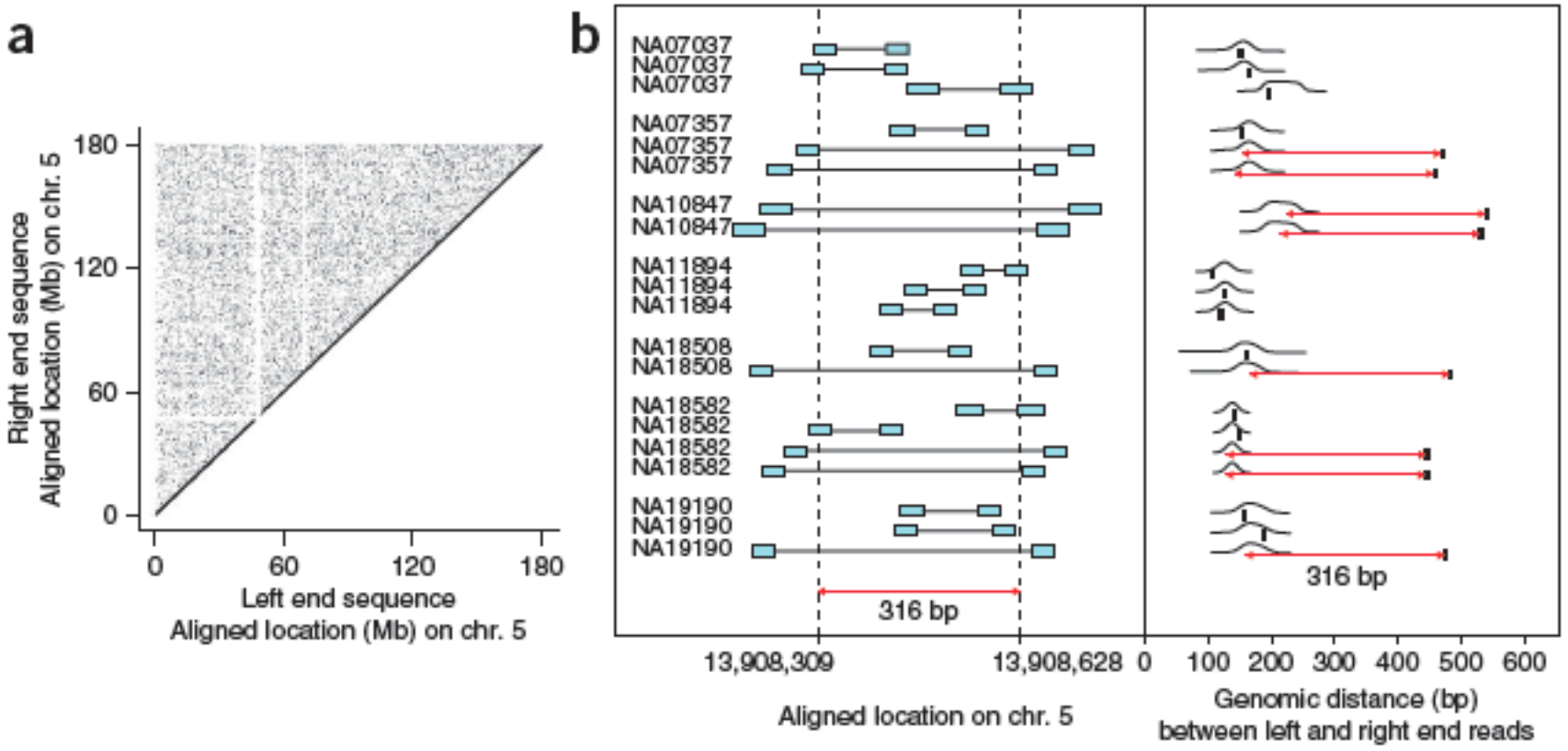
Bellos *et al. Genome Biology* 2012, **13**:R120
http://genomebiology.com/2013/13/12/R120



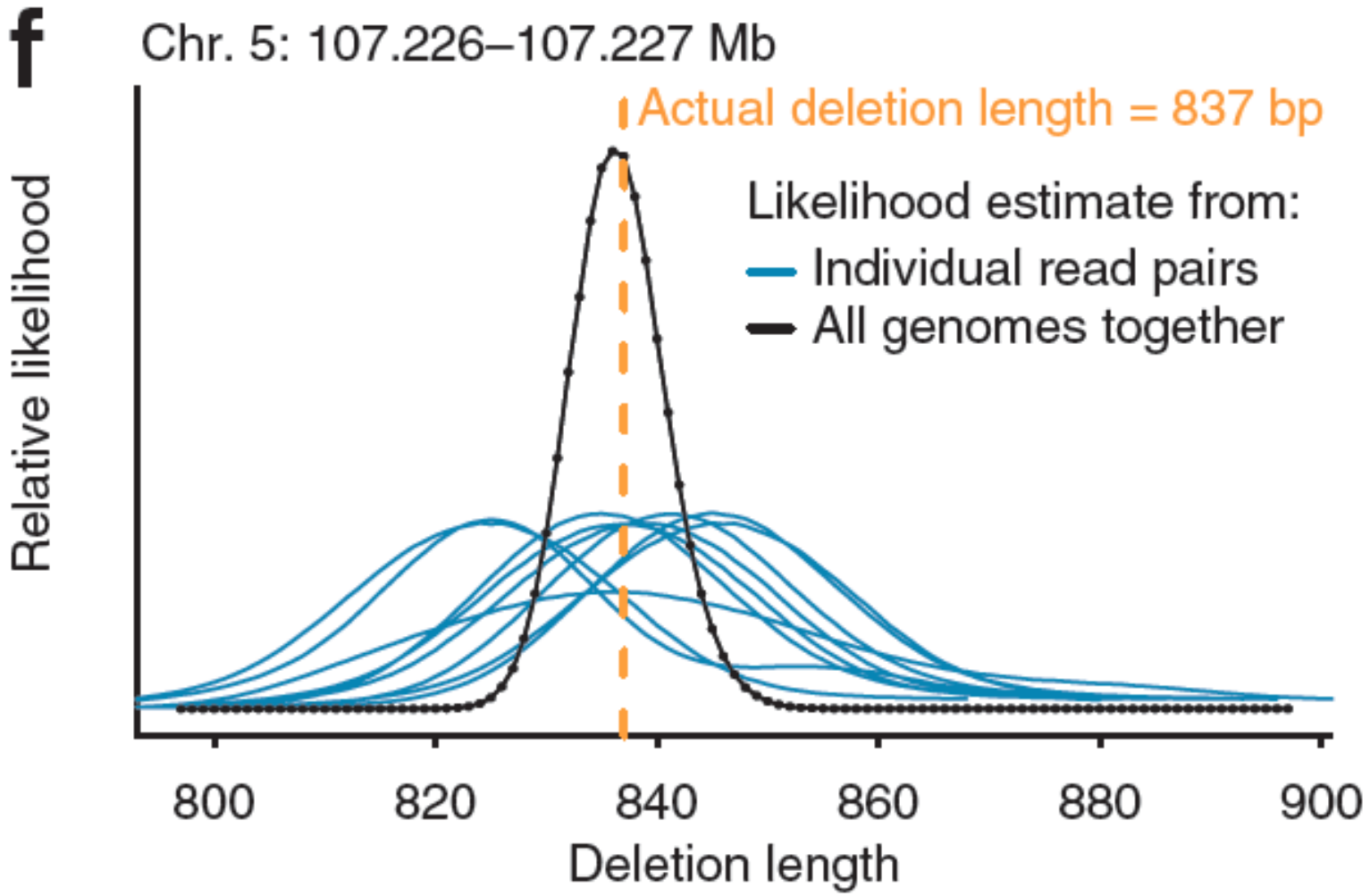
Discovery and genotyping of genome structural polymorphism by sequencing on a population scale



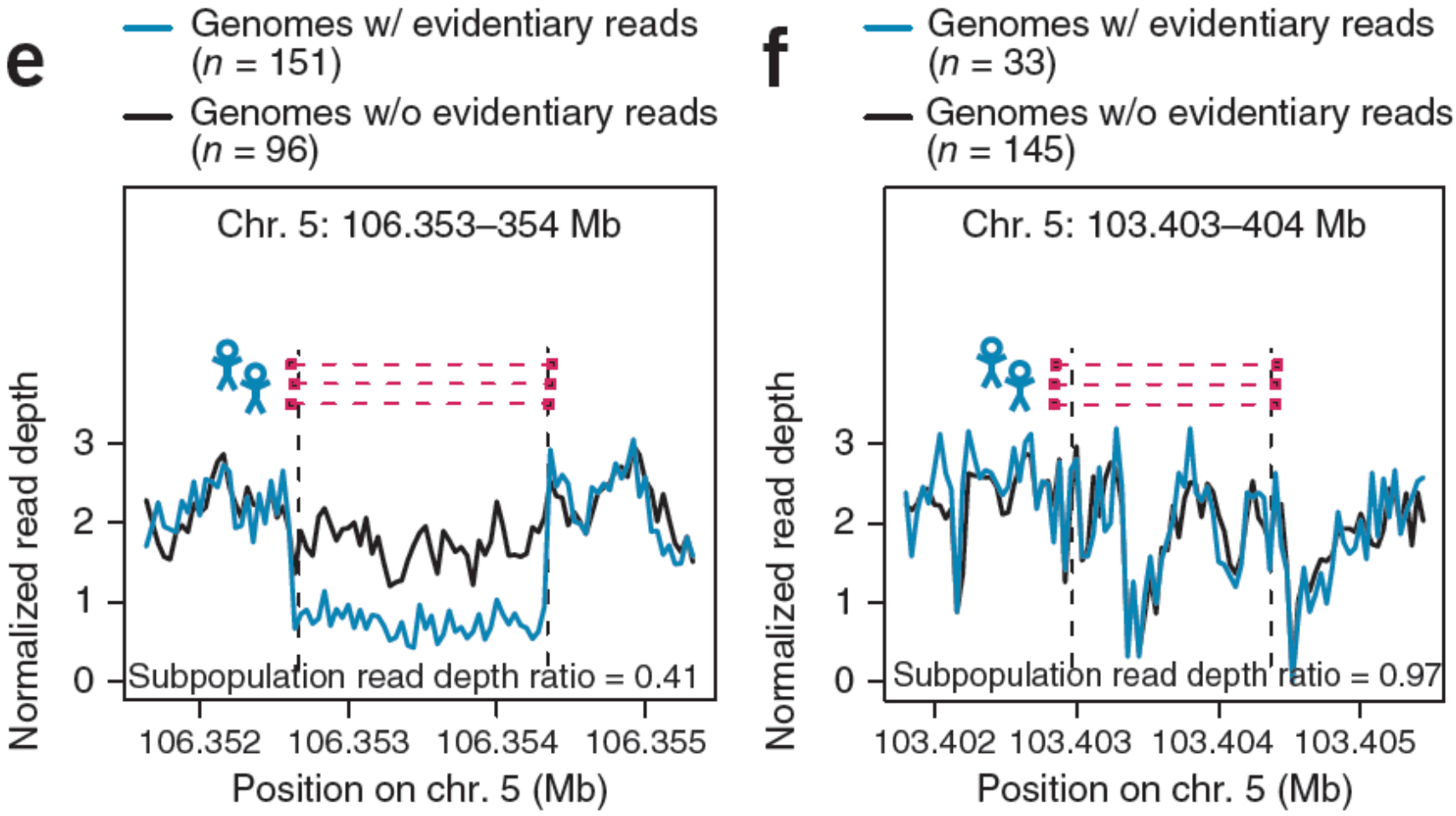
Discovery and genotyping of genome structural polymorphism by sequencing on a population scale



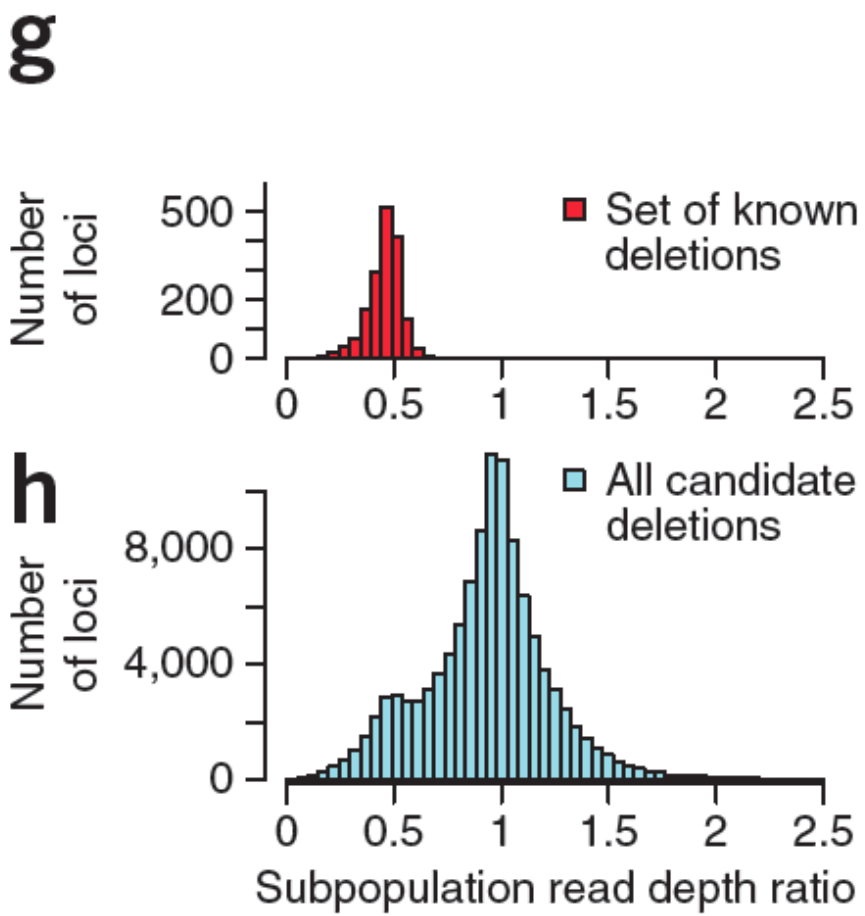
Discovery and genotyping of genome structural polymorphism by sequencing on a population scale



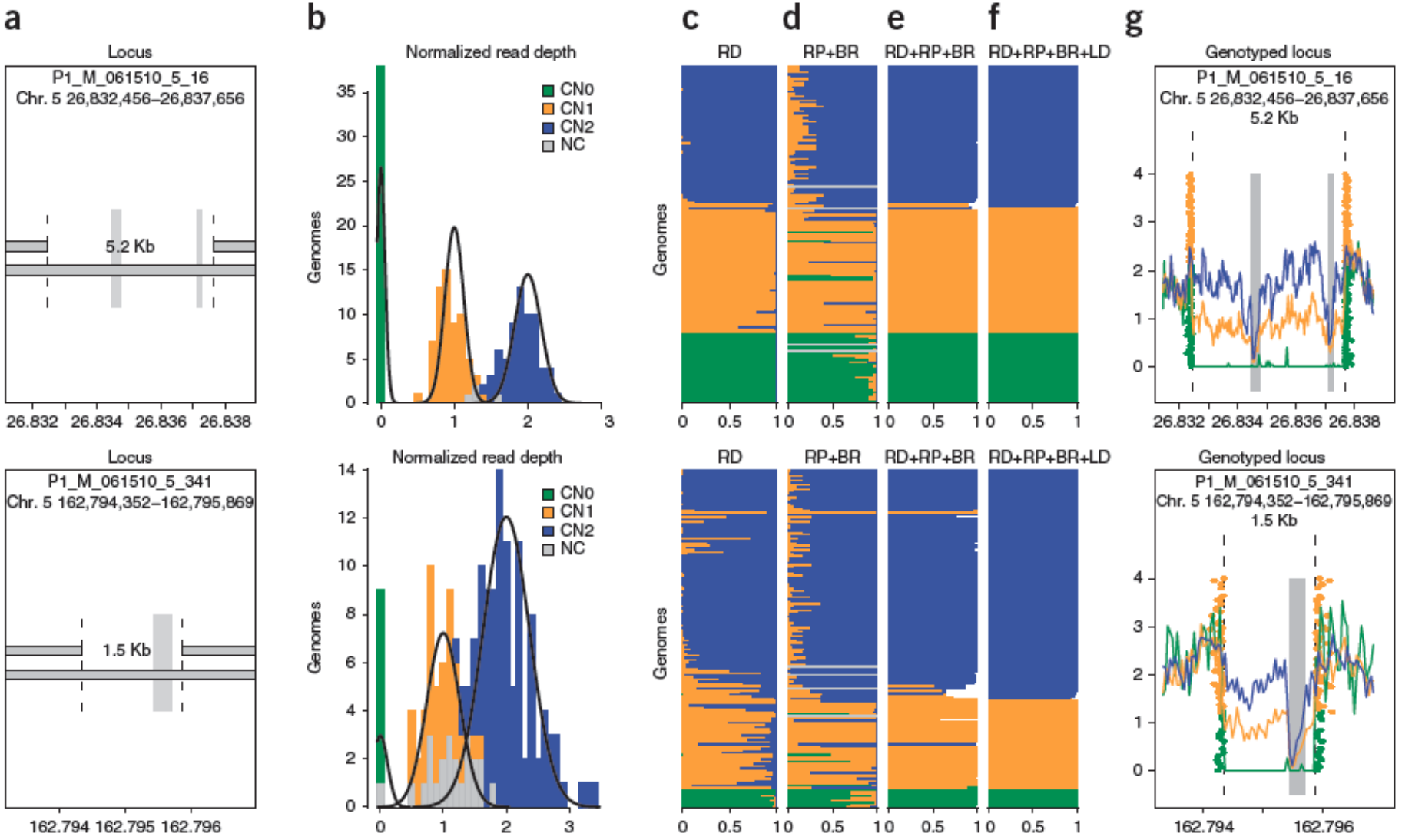
Discovery and genotyping of genome structural polymorphism by sequencing on a population scale



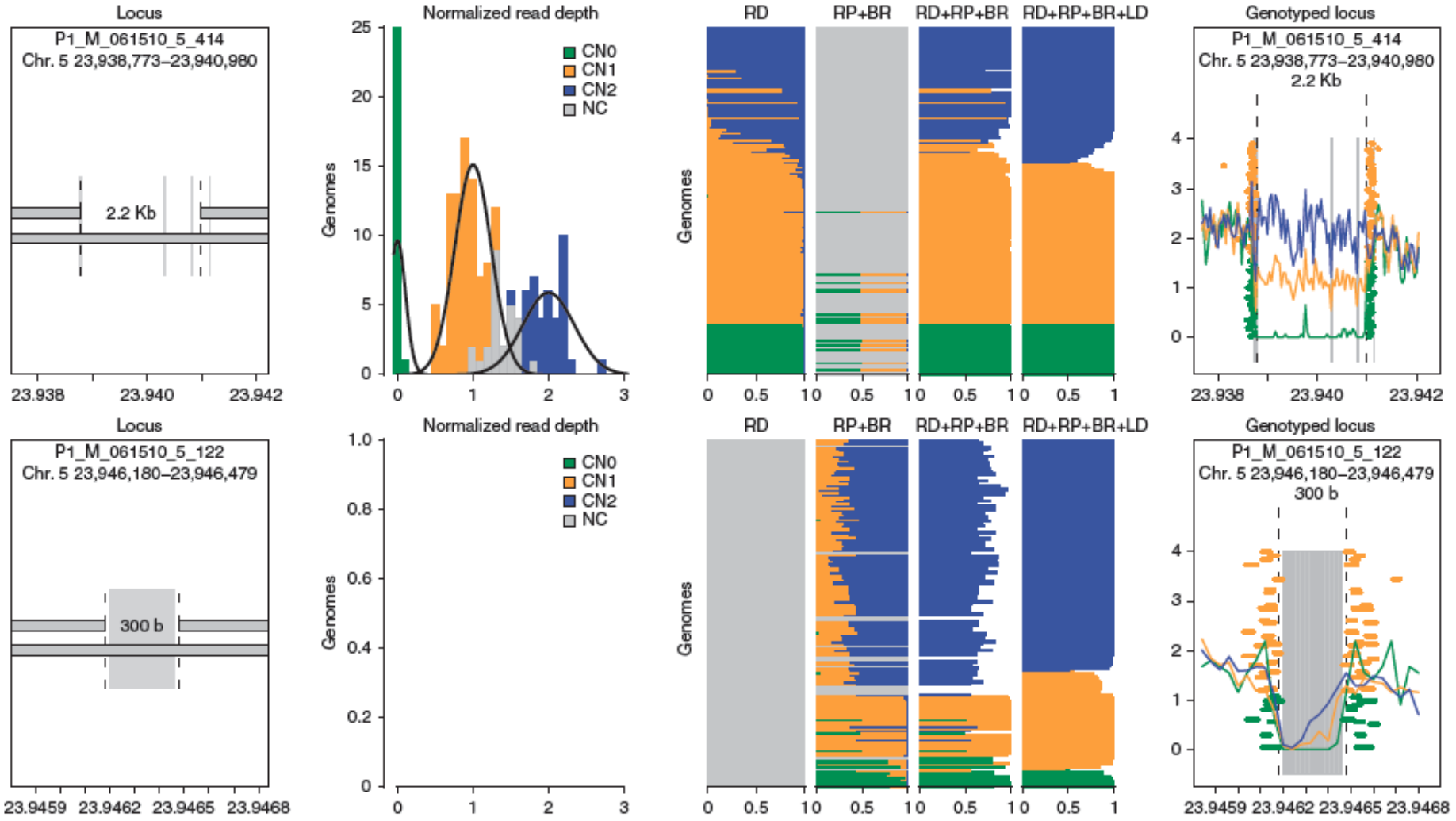
Discovery and genotyping of genome structural polymorphism by sequencing on a population scale



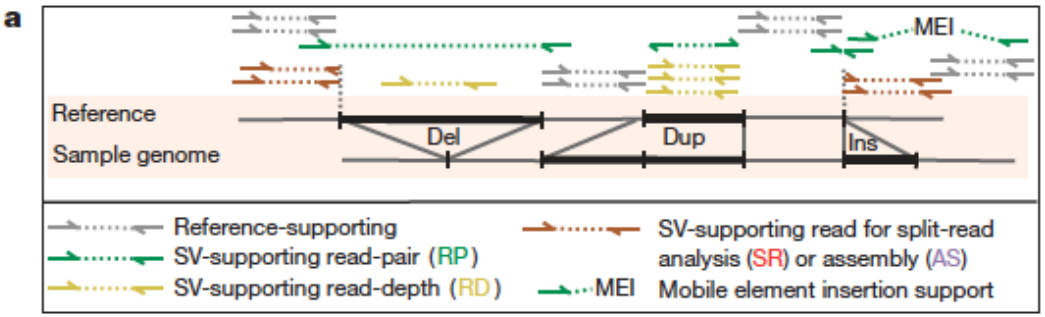
Discovery and genotyping of genome structural polymorphism by sequencing on a population scale

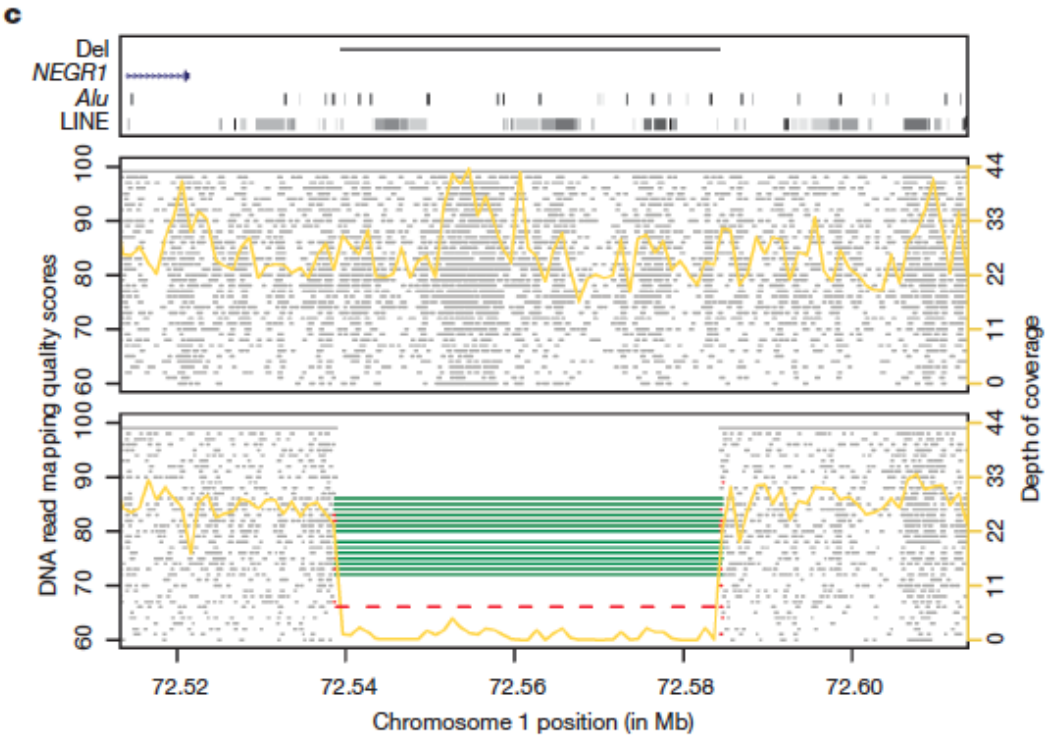


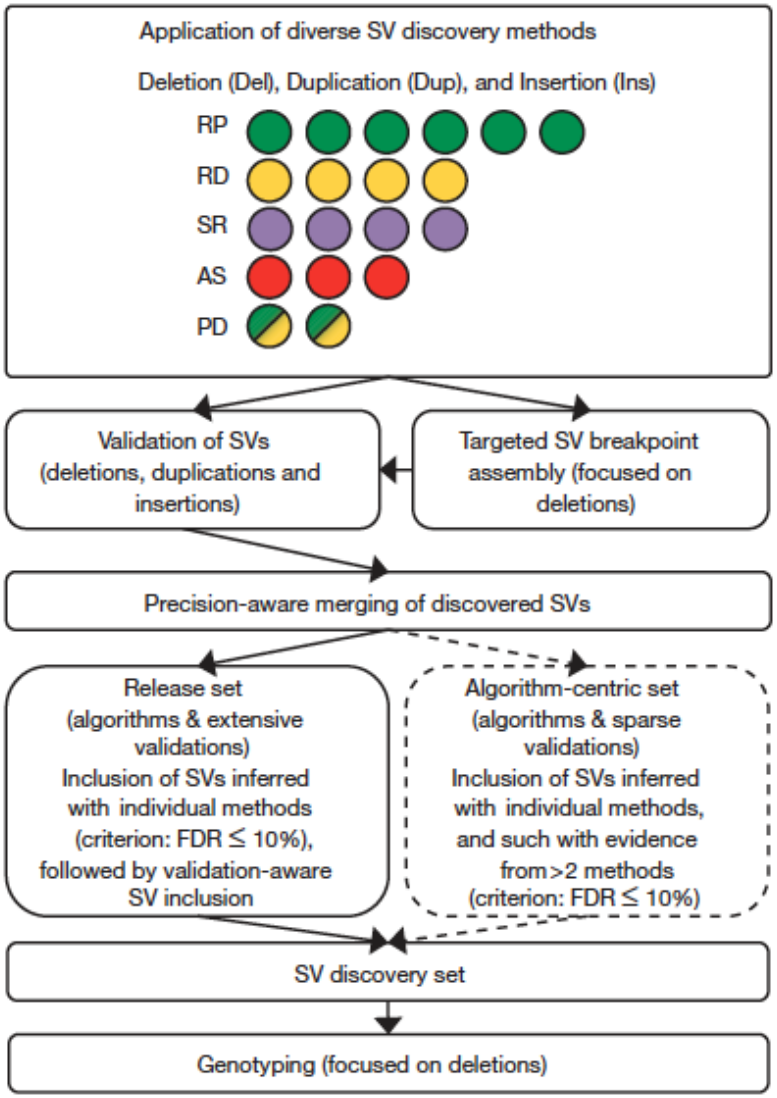
Discovery and genotyping of genome structural polymorphism by sequencing on a population scale



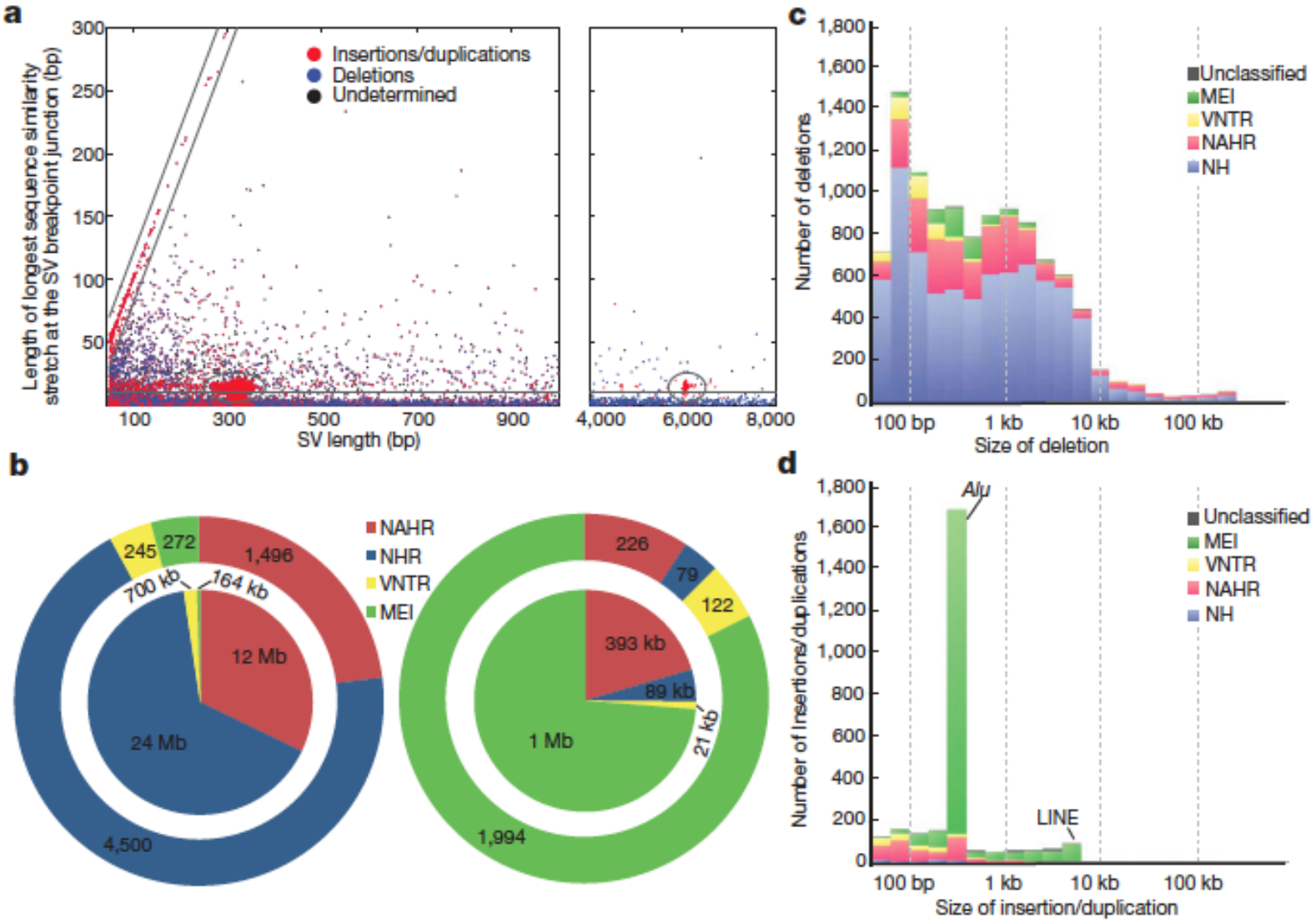
Mapping copy number variation by population-scale genome sequencing



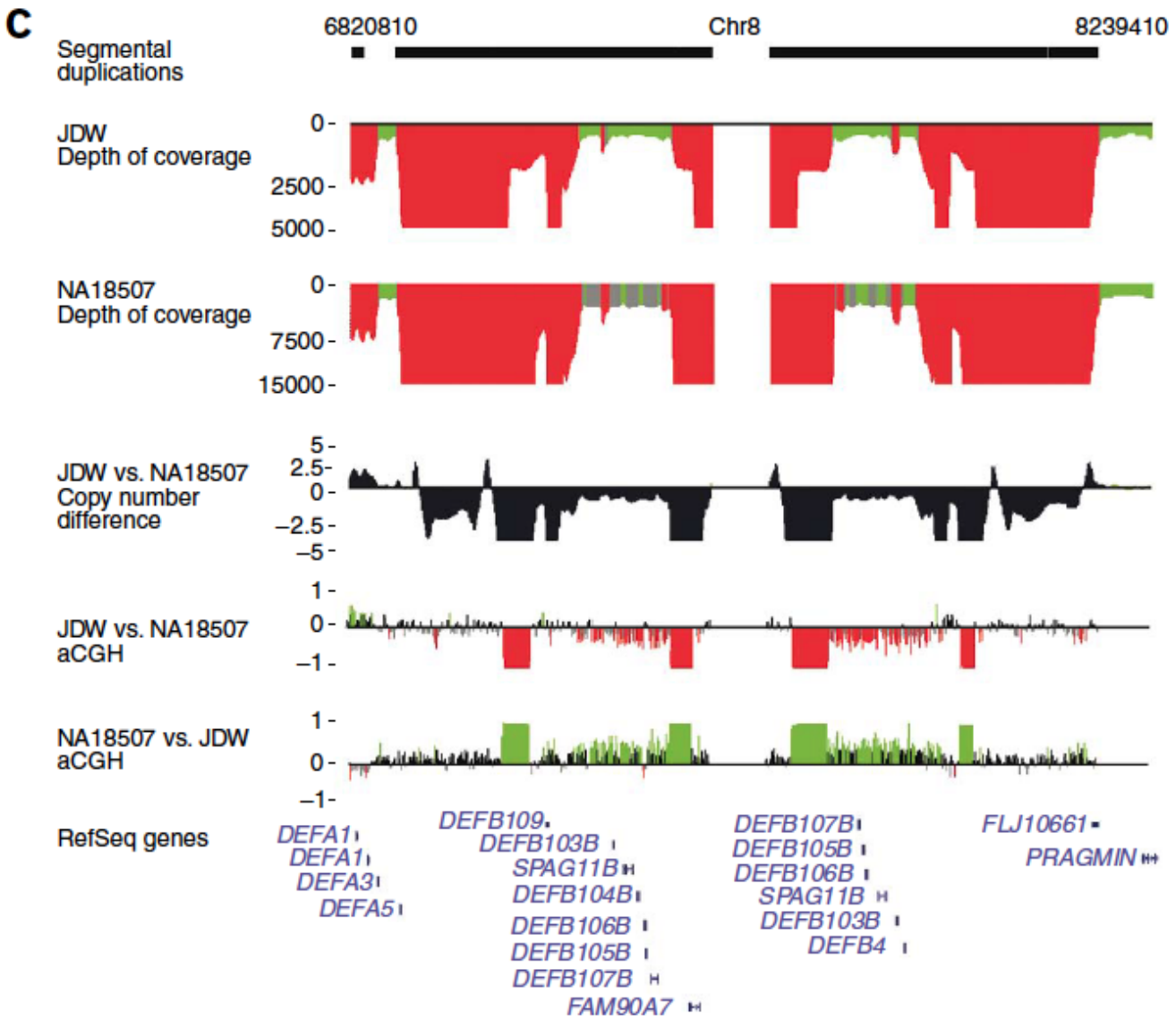




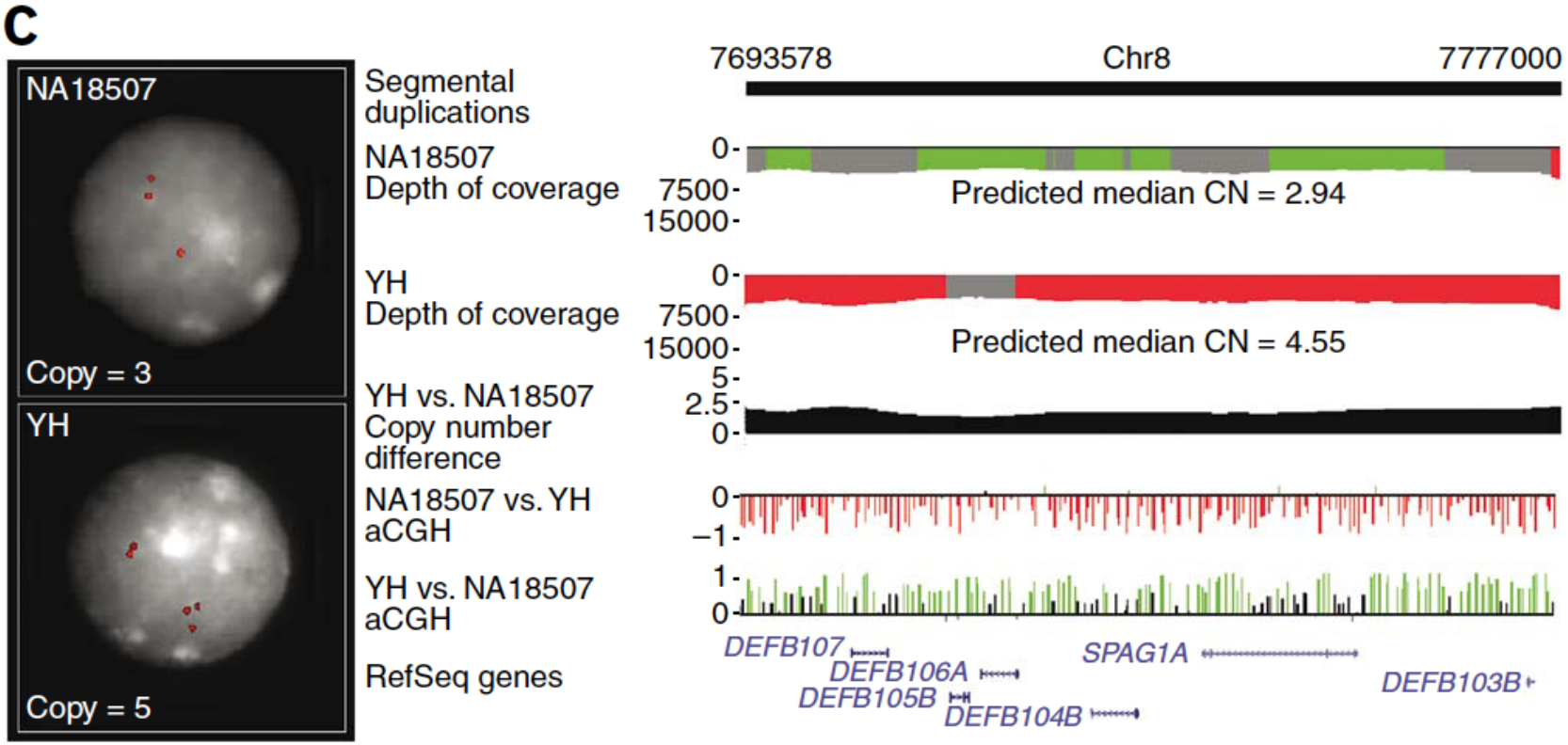
Mapping copy number variation by population-scale genome sequencing



Personalized copy number and segmental duplication maps using next-generation sequencing



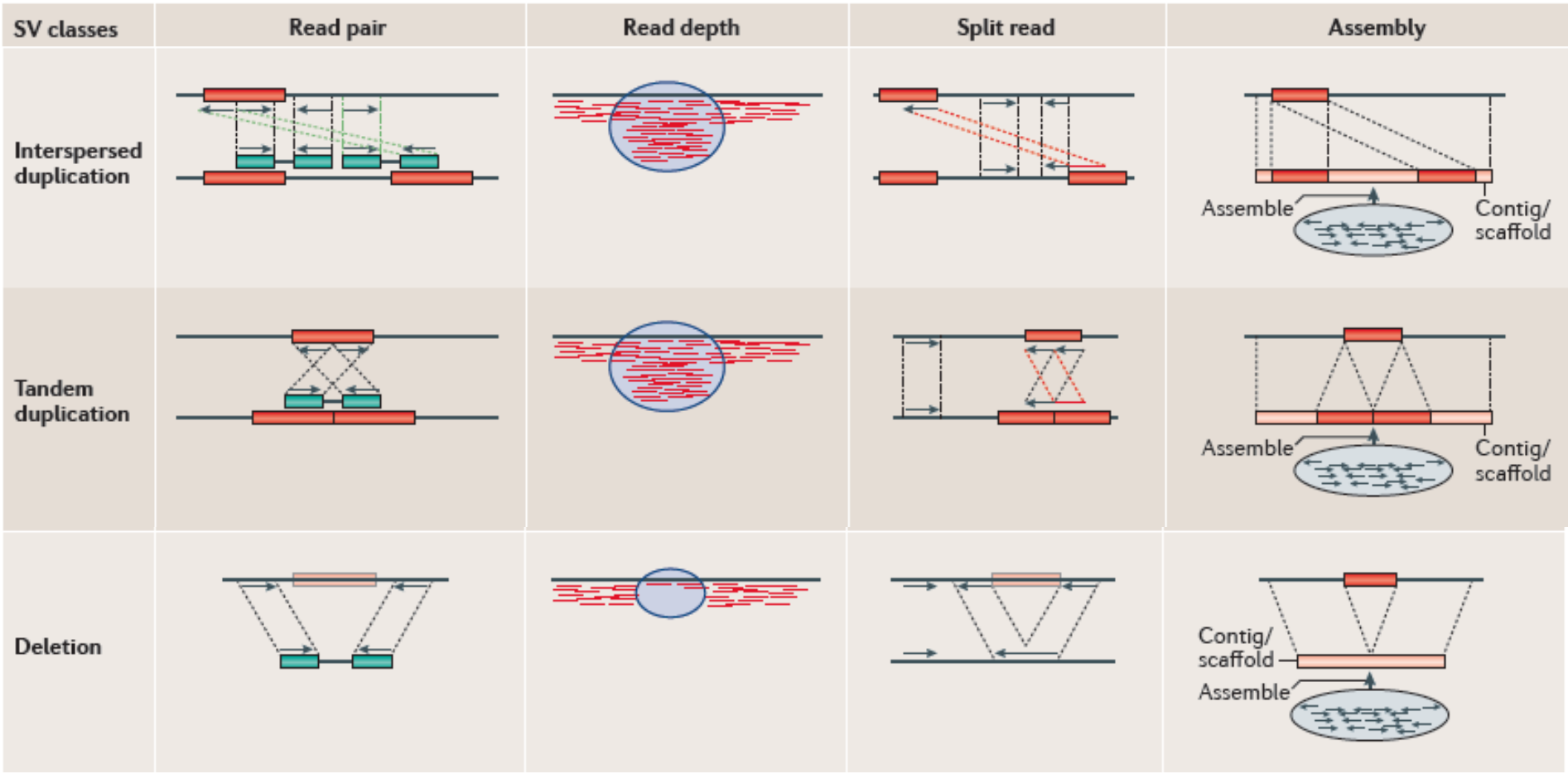
Personalized copy number and segmental duplication maps using next-generation sequencing



Genome structural variation discovery and genotyping

Can Alkan‡, Bradley P. Coe* and Evan E. Eichler*‡*

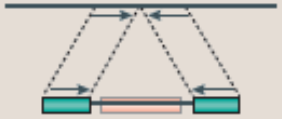
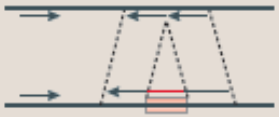
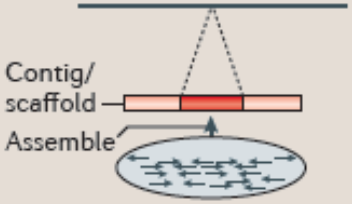
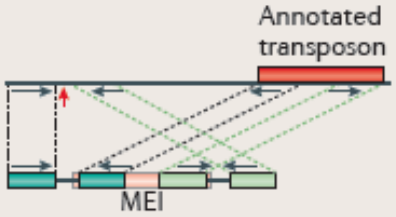
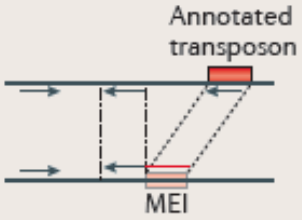
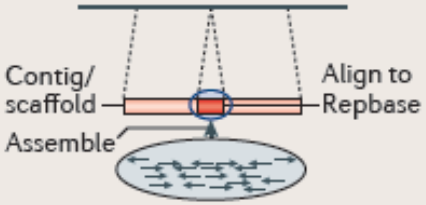
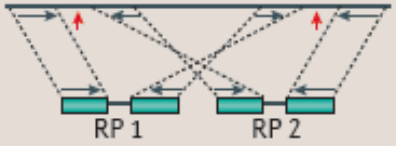
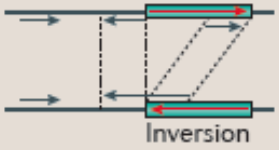
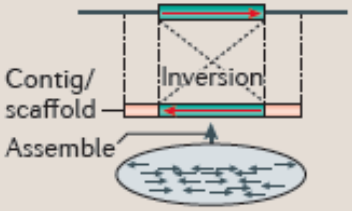
Classes of Structural Variation => Copy-Number Variants



Genome structural variation discovery and genotyping

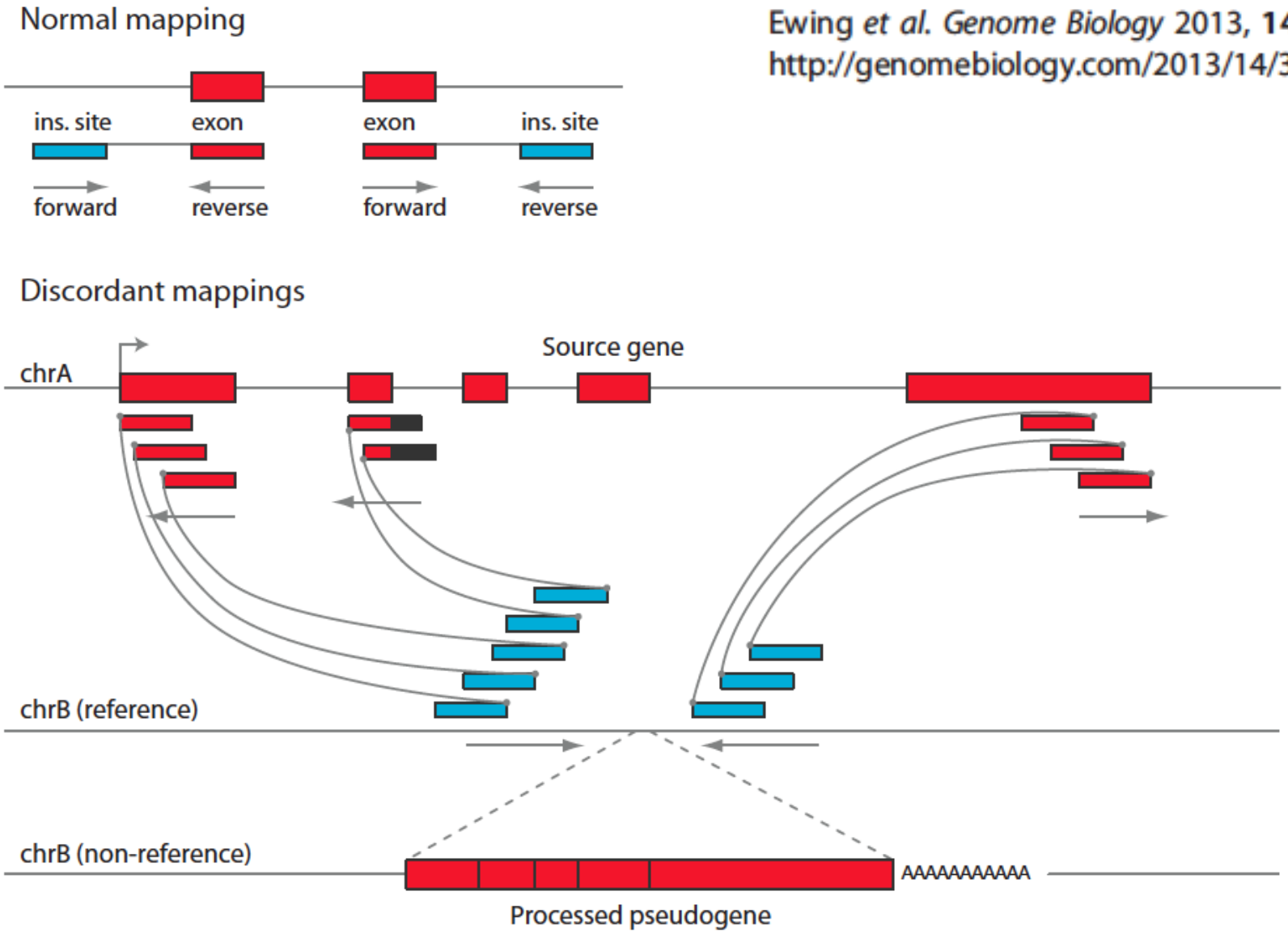
Can Alkan‡, Bradley P. Coe* and Evan E. Eichler*‡*

Classes of Structural Variation => 'Balanced' Variants

SV classes	Read pair	Read depth	Split read	Assembly
Novel sequence insertion		Not applicable		
Mobile-element insertion		Not applicable		
Inversion		Not applicable		

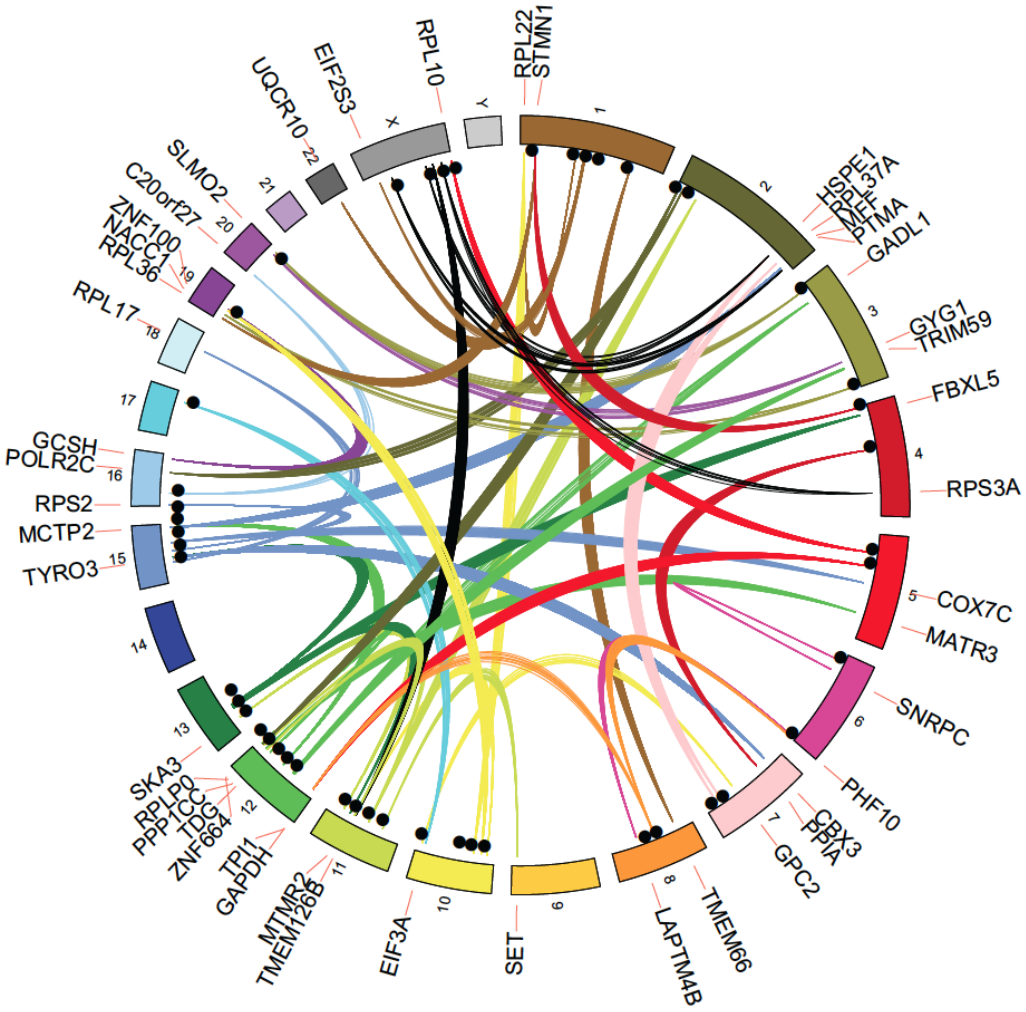
Retrotransposition of gene transcripts leads to structural variation in mammalian genomes

Ewing *et al. Genome Biology* 2013, **14**:R22
http://genomebiology.com/2013/14/3/R22



Retrotransposition of gene transcripts leads to structural variation in mammalian genomes

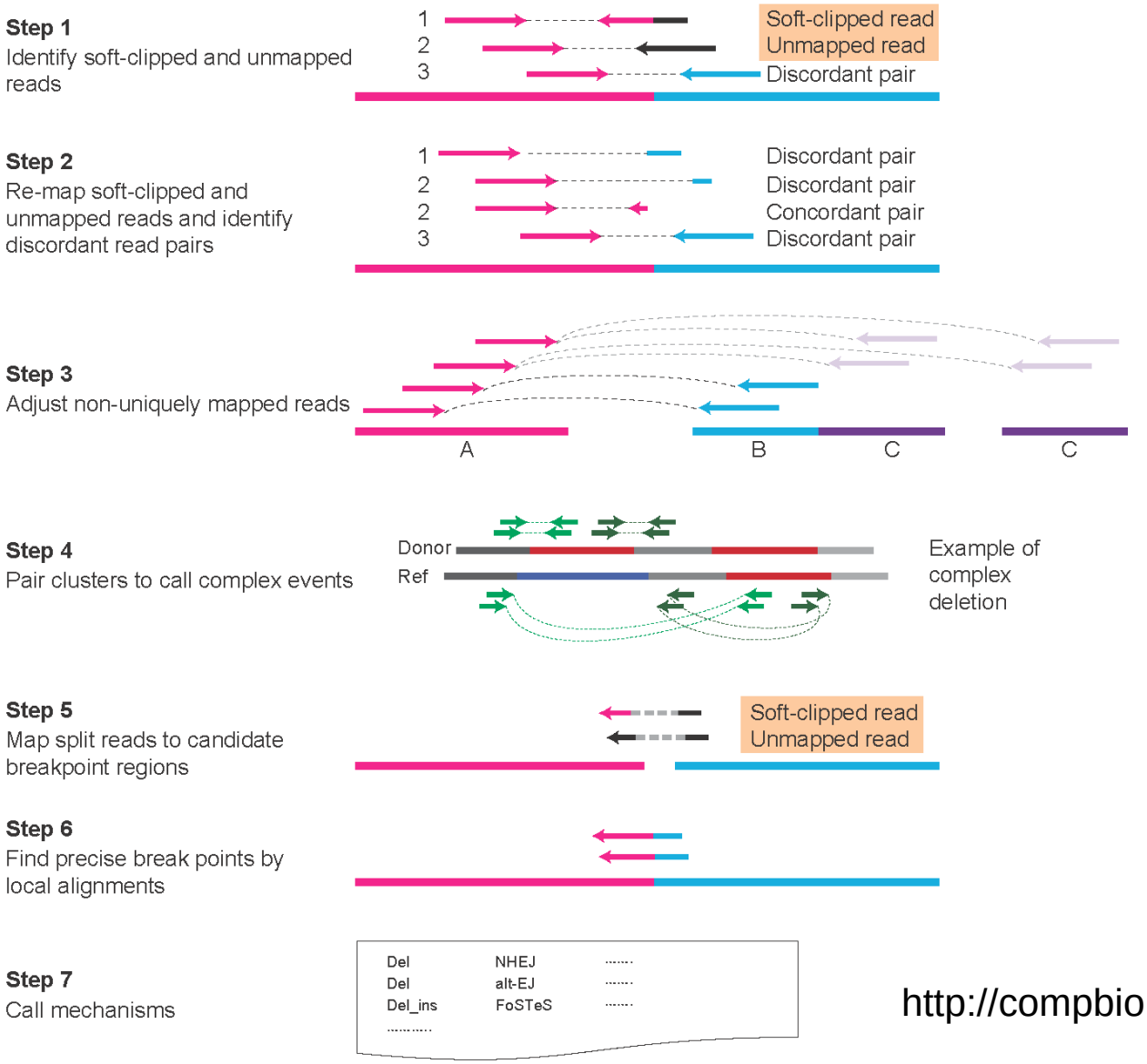
Ewing *et al. Genome Biology* 2013, **14**:R22
http://genomebiology.com/2013/14/3/R22



Locations of 48 non-reference gene retrocopy insertion sites in the human genome based on reads mapped to source genes.

Diverse Mechanisms of Somatic Structural Variations in Human Cancer Genomes

Cell *153*, 919–929, May 9, 2013



http://compbio.med.harvard.edu/Meerkat/