Illumina Seminar - Milan June 18, 2009

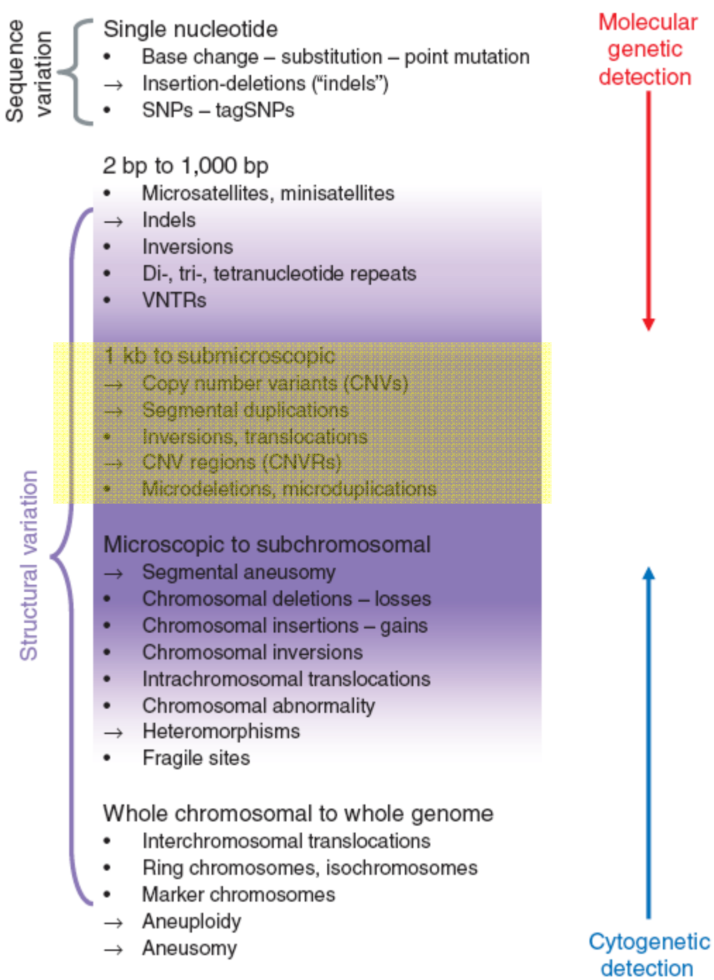
Copy Number Variations

Untangling the complexity of mendelian and complex diseases

Federica Torri
Dept of Science & Biomedical Technologies
Fondazione Filarete, University of Milan

What are CNVs?

Stretches of DNA larger than 1 kb that display copy number differences in comparison to a reference genome



- 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)

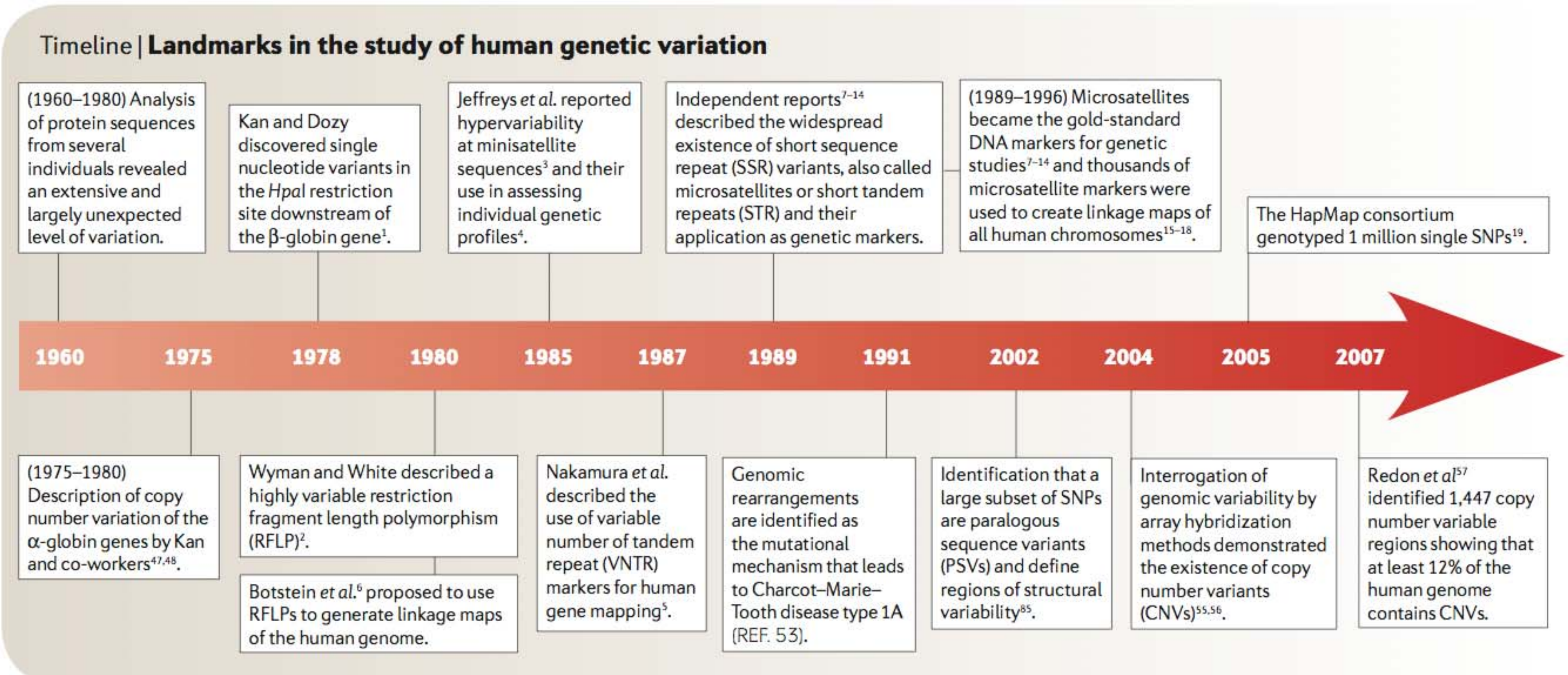
- Copy number **polymorphism** (relatively common CNV, with a fixed starting/ending position)

- Copy number **difference** (between-species copy number differences, e.g chimpanzees and humans)

OPINION
Nat Rev Genet. 2007;8:639-46.

Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability

Jacques S. Beckmann, Xavier Estivill and Stylianos E. Antonarakis



CNVs in α-globins

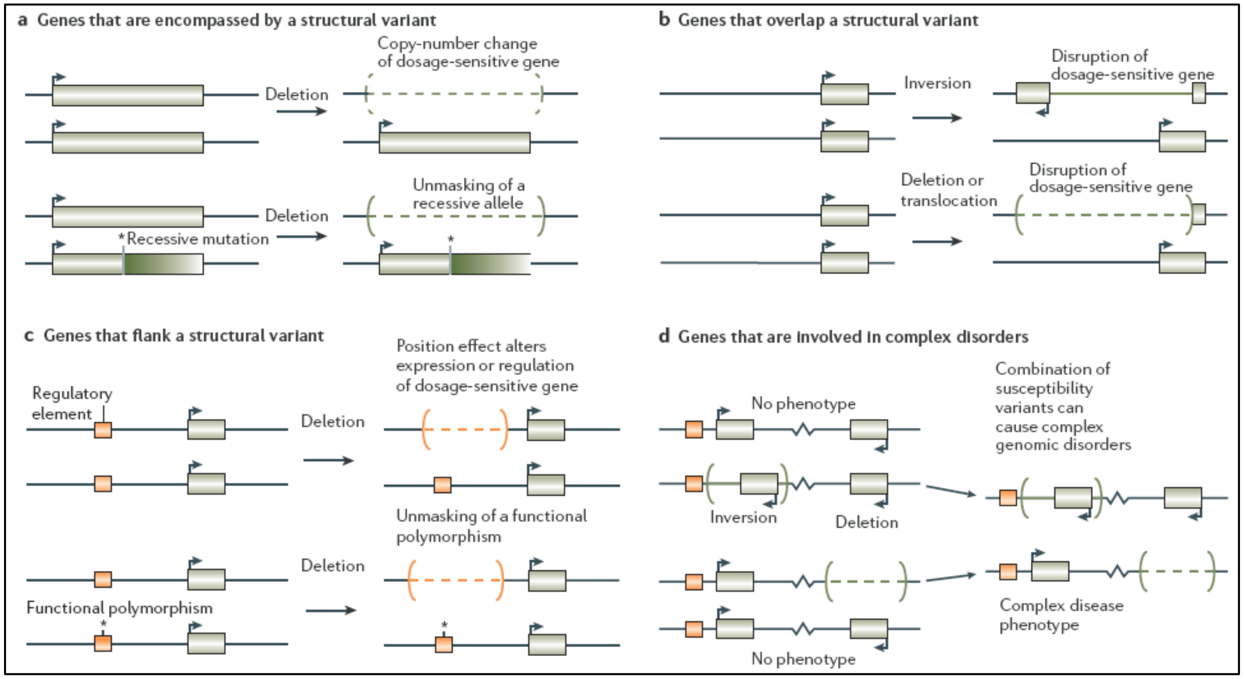
CNVs in CMT

CNVs as a general feature

Structural variation in the human genome

Structural variation in the human genome **Nat Rev Genet 7: 85-97, 2006**

Lars Feuk, Andrew R. Carson and Stephen W. Scherer



Millions of nucleotides of heterogeneity within every genome likely to make an important contribution to human diversity and disease susceptibility.

nature Vol 444|23 November 2006|doi:10.1038/nature05329

ARTICLES

Global variation in copy number in the human genome

Richard Redon[1], Shumpei Ishikawa[2,3], Karen R. Fitch[4], Lars Feuk[5,6], George H. Perry[7], T. Daniel Andrews[1], Heike Fiegler[1], Michael H. Shapero[4], Andrew R. Carson[5,6], Wenwei Chen[4], Eun Kyung Cho[7], Stephanie Dallaire[7], Jennifer L. Freeman[7], Juan R. González[8], Mònica Gratacòs[8], Jing Huang[4], Dimitrios Kalaitzopoulos[1], Daisuke Komura[3], Jeffrey R. MacDonald[5], Christian R. Marshall[5,6], Rui Mei[4], Lyndal Montgomery[1], Kunihiro Nishimura[2], Kohji Okamura[5,6], Fan Shen[4], Martin J. Somerville[9], Joelle Tchinda[7], Armand Valsesia[1], Cara Woodwark[1], Fengtang Yang[1], Junjun Zhang[5], Tatiana Zerjal[1], Jane Zhang[4], Lluis Armengol[8], Donald F. Conrad[10], Xavier Estivill[8,11], Chris Tyler-Smith[1], Nigel P. Carter[1], Hiroyuki Aburatani[2,12], Charles Lee[7,13], Keith W. Jones[4], Stephen W. Scherer[5,6] & Matthew E. Hurles[1]

~12 % or 4.8%?

Detection of CNV on NG42M

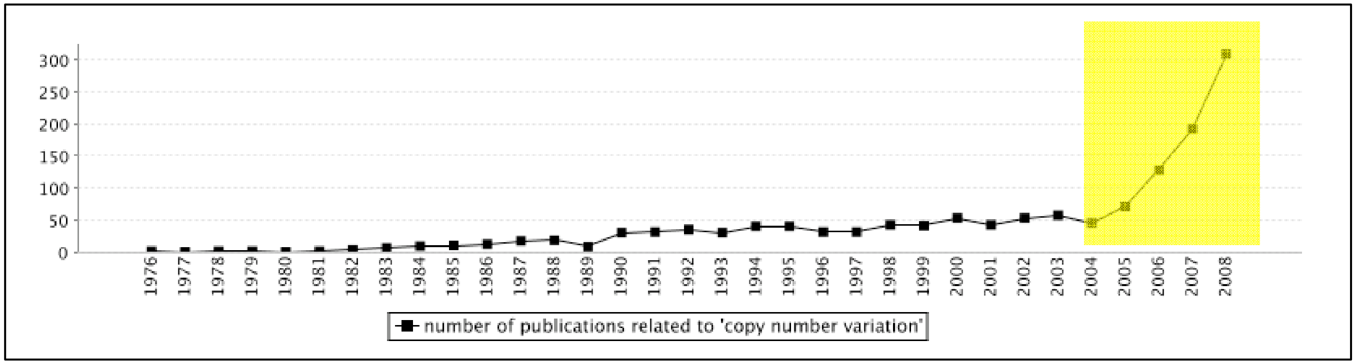
Quality control: 20 arrays * 40 individuals = 800 hybridisations
=> 81 had to be repeated

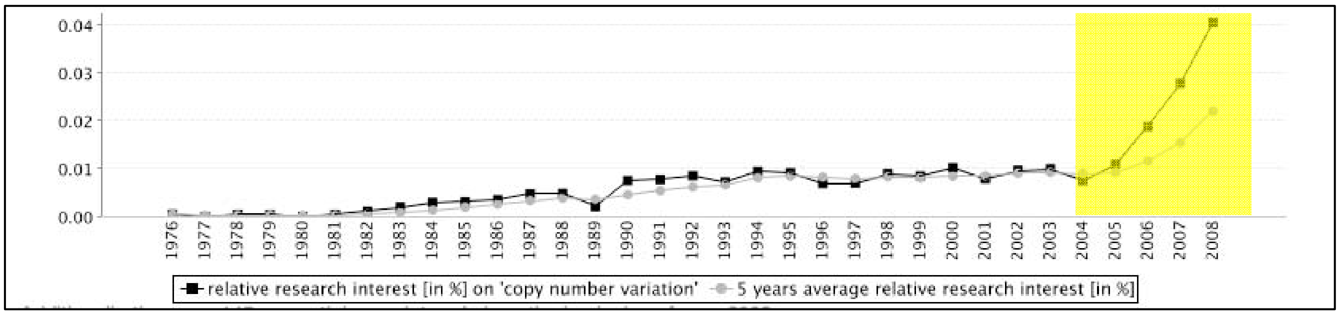
Normalisation: Equalization of intensities (qspline)
GC correction (linear regression)
Wave normalisation (loess)

Preliminary results: **51,981 CNV calls ***
=> on average: **1,300 CNVs per comparison** (18-30 Mb)

corresponding to **9,299 discrete CNV loci**
which cover **4.8% of the human genome** (135.6 Mb)

CNV: the hottest topic?





Top 20 journals

Journal title	Publications
genetics	48
cytogenet genome res	45
genome res	41
nat genet	33
p natl acad sci usa	30
am j hum genet	29
nucleic acids res	28
bmc genomics	26
mol biol evol	26
hum mol genet	23
j mol evol	22
gene	22
hum mutat	21
genome	20
plos genet	18
hum genet	17
j bacteriol	17
nature	15
immunogenetics	15
j clin microbiol	15

Last 20 years

Year	Publications
2009	145
2008	309
2007	192
2006	127
2005	71
2004	45
2003	57
2002	53
2001	42
2000	53
1999	41
1998	42
1997	31
1996	31
1995	40
1994	40
1993	30
1992	35
1991	31
1990	30

The current map of human structural variation is far from complete….

Genome Sciences

UW

Model Organism Genetics Human & Medical Genetics Genomics & Proteomics Computational Biology

http://humanparalogy.gs.washington.edu/structuralvariation/

Database of Genomic Variants

A curated catalogue of structural variation in the human genome

Hosted by:
The Centre for
Applied
Genomics

http://projects.tcag.ca/variation/

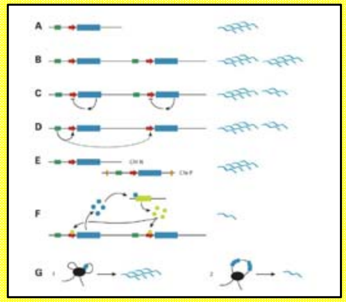
Summary Statistics

Total entries: 38406 (hg18)
CNVs: 21178
Inversions: 499
InDels (100bp-1Kb): 16729
Total CNV loci: 6558
Articles cited: 31

Last updated: Mar 11, 2009
Join our mailing list

CNVs in health…

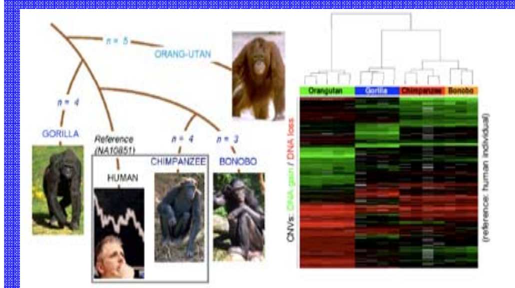
Overlap and regulate the expression of ~ 18% of genes



Account for more nucleotide variation than SNPs in the human genome

Non-random distribution across the genome: "hot spots"

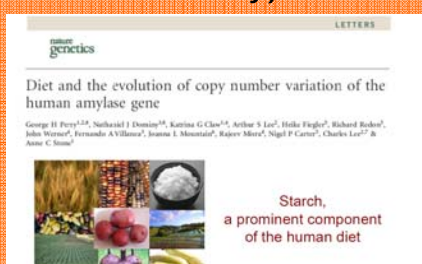
Important in primate evolution



Nearly 50% map in SD's regions

Mechanisms: NAHR, NHEJ, retrotransposition, FoSTeS

Underwent selective pressure in the human lineage (shaped by mutation, selection, demographic history)


LETTERS

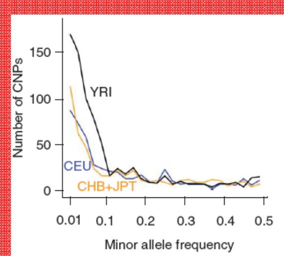
nature genetics

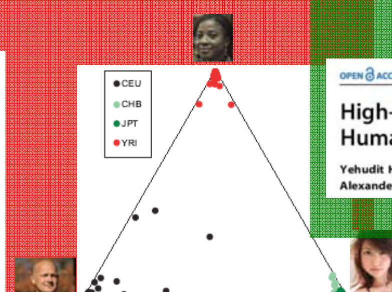
Diet and the evolution of copy number variation of the human amylase gene

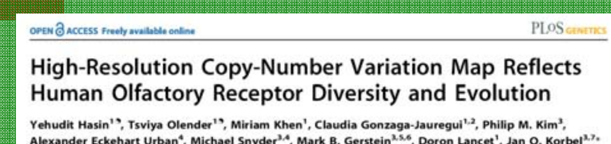
AMY1 gene paradigm

Enriched towards "environmental sensor genes" (sensory perception and defense response functions)

Population-specific variation





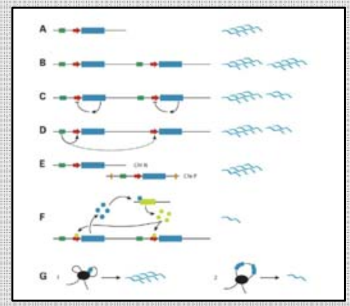

OPEN ACCESS Freely available online — PLoS GENETICS

High-Resolution Copy-Number Variation Map Reflects Human Olfactory Receptor Diversity and Evolution

Yehudit Hasin, Tsviya Olender, Miriam Khen, Claudia Gonzaga-Jauregui, Philip M. Kim, Alexander Eckehart Urban, Michael Snyder, Mark B. Gerstein, Doron Lancet, Jan O. Korbel

CNVs in health…

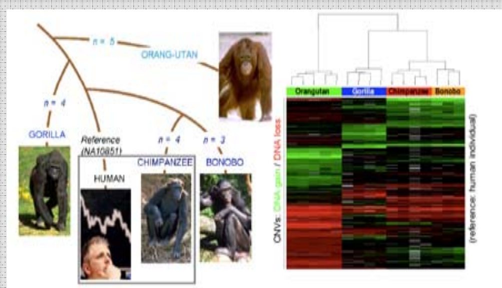
Overlap and regulate the expression of ~ 18% of genes



Account for more nucleotide variation than SNPs in the human genome

Non-random distribution across the genome: "hot spots"

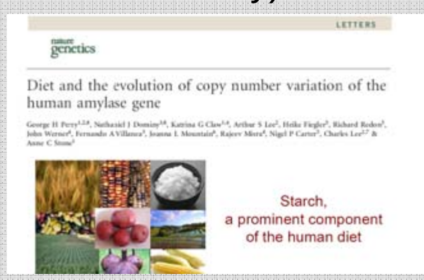
Important in primate evolution



Nearly 50% map in SD's regions

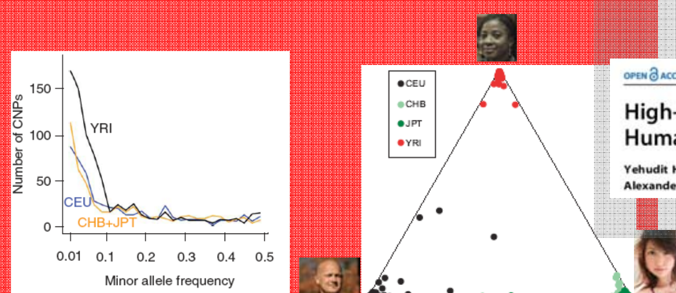
Mechanisms: NAHR, NHEJ, retrotransposition, FoSTeS

Underwent selective pressure in the human lineage (shaped by mutation, selection, demographic history)



AMY1 gene paradigm

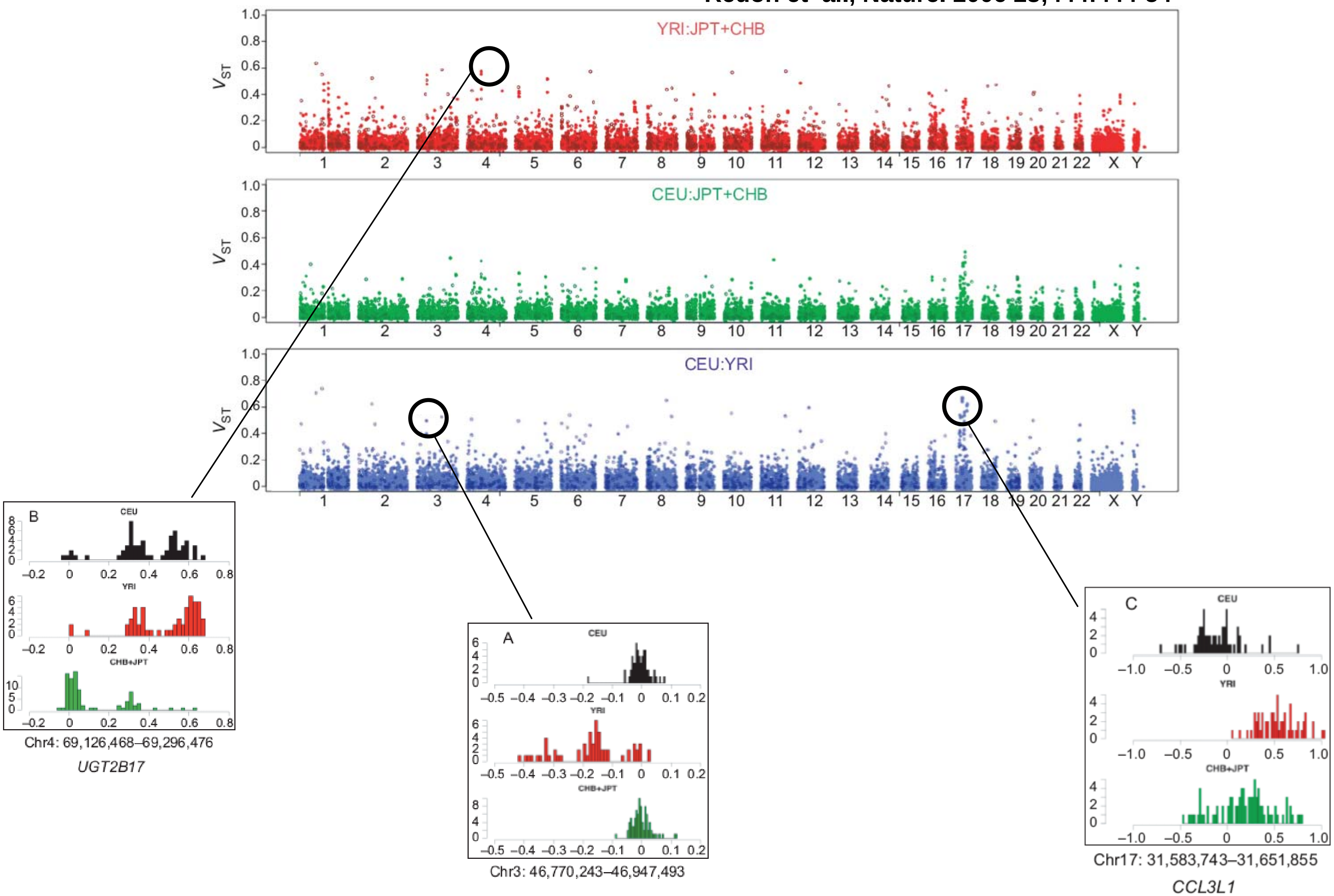
Population-specific variation



Enriched towards "environmental sensor genes" (sensory perception and defense response functions)

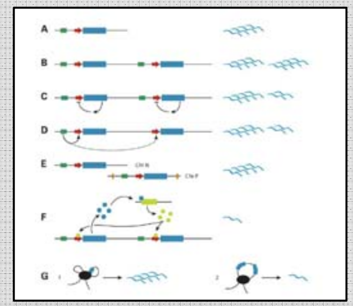
Population differentiation

Redon et al., Nature. 2006 23;444:444-54



CNVs in health…

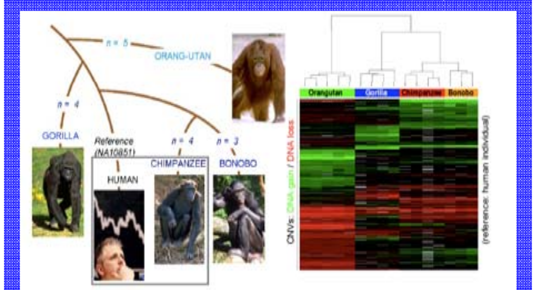
Overlap and regulate the expression of ~ 18% of genes



Account for more nucleotide variation than SNPs in the human genome

Non-random distribution across the genome: “hot spots”

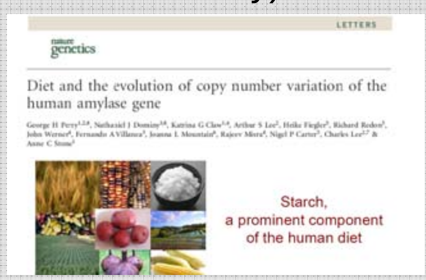
Important in primate evolution



Nearly 50% map in SD’s regions

Mechanisms: NAHR, NHEJ, retrotransposition, FoSTeS

Underwent selective pressure in the human lineage (shaped by mutation, selection, demographic history)


LETTERS

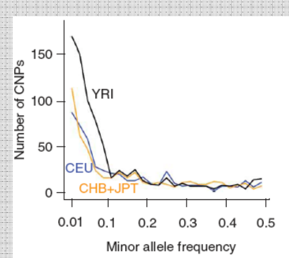
nature genetics

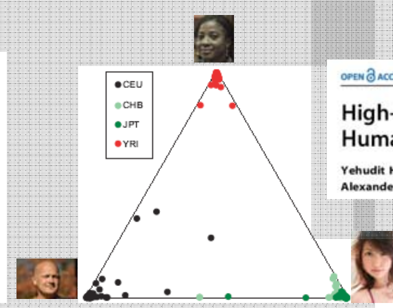
Diet and the evolution of copy number variation of the human amylase gene

Starch, a prominent component of the human diet

AMY1 gene paradigm

Population-specific variation





Enriched towards “environmental sensor genes” (sensory perception and defense response functions)


OPEN ACCESS Freely available online

PLoS GENETICS

High-Resolution Copy-Number Variation Map Reflects Human Olfactory Receptor Diversity and Evolution

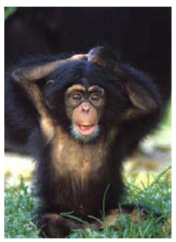
CNVs and evolution: how has CNVs shaped our genomes?

Article Genome Res. 2008, 18:1698-710

Copy number variation and evolution in humans and chimpanzees

George H. Perry,[1,2,6] Fengtang Yang,[3] Tomas Marques-Bonet,[4] Carly Murphy,[2] Tomas Fitzgerald,[3] Arthur S. Lee,[2] Courtney Hyland,[2] Anne C. Stone,[1] Matthew E. Hurles,[3] Chris Tyler-Smith,[3] Evan E. Eichler,[4] Nigel P. Carter,[3] Charles Lee,[2,5] and Richard Redon[3,6,7]

CNV in 30 chimpanzees

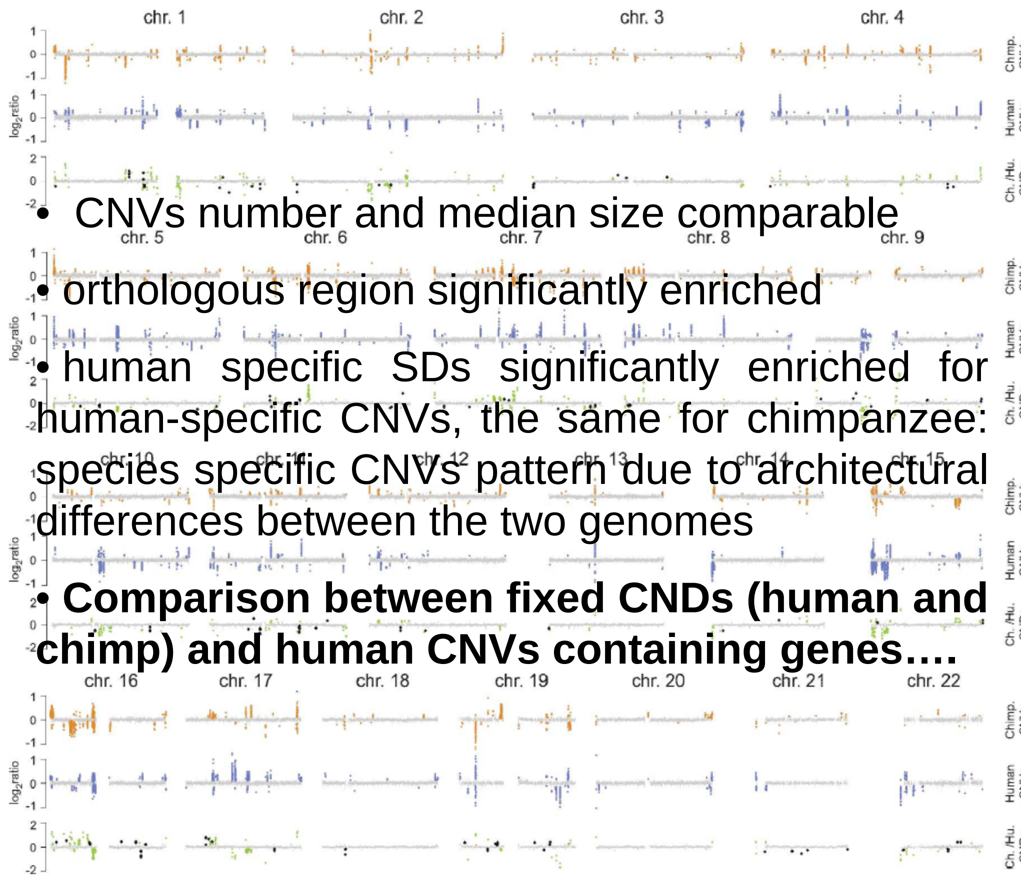


reference: Clint

Copy Number Differences

reference: NA10851

CNV in 30 humans



- CNVs number and median size comparable
- orthologous region significantly enriched
- human specific SDs significantly enriched for human-specific CNVs, the same for chimpanzee: species specific CNVs pattern due to architectural differences between the two genomes
- **Comparison between fixed CNDs (human and chimp) and human CNVs containing genes….**

The relations between CNVs affecting genes and fixation

Table 3. Rates of copy number fixation and polymorphism by gene functional categories

GO categories[a]	Description	Fixed CNDs[b]	Total CNVRs[c]	Ratio *F*/*T*	Score[d]	*P*-value[e]
—	No gene (intergenic)	18	117	0.15	1.00	NA
—	One or more gene(s)	74	518	0.14	0.93	0.886
Lowest scores						
GO:0008233	Peptidase activity	0	25	0.00	0.04	0.048
GO:0048503	GPI anchor binding	0	23	0.00	0.04	0.077
GO:0016301	Kinase activity	0	18	0.00	0.05	0.132
GO:0006811	Ion transport	1	48	0.02	0.15	0.027
GO:0005215	Transporter activity	1	45	0.02	0.16	0.029
Other scores (discussed in text)						
GO:0006955	Immune response	3	35	0.09	0.57	0.420
GO:0004984	Olfactory receptor activity	5	20	0.25	1.60	0.534
Highest scores						
GO:0005506	Iron ion binding	8	28	0.29	1.83	0.197
GO:0051301	Cell division	5	15	0.33	2.09	0.182
GO:0007067	Mitosis	5	15	0.33	2.09	0.182
GO:0008283	Cell proliferation	6	15	0.40	2.50	0.099
GO:0006954	Inflammatory response	5	12	0.42	2.58	0.141

[a]Gene Ontology (GO) categories were included in the analysis only if $F + T > 16$, where F is the number of fixed CNDs and T the total number of CNVRs with one or more genes from the GO category.
[b]The number of CNDs between the human and chimpanzee reference individuals that did not overlap any within-species human or chimpanzee CNVR, that overlap one or more genes assigned to a given GO category.
[c]The number of total CNVRs (human-only CNVRs + chimpanzee-only CNVRs + CNVRs observed in the same regions in both species; i.e., no CNVR regions are counted twice) that overlap one or more genes assigned to a given GO category.
[d]The score is a normalized F/T ratio for each GO category. It was calculated using the formula $(1 + F/A)/(1 + T)$, where A is the ratio F/T for all CNDs/CNVRs that do not contain genes (intergenic variants). The GO categories with the five lowest and five highest scores are listed, as well as two categories discussed in the text: see Supplemental Table 4 for complete data set.
[e]Two-tailed Fisher's exact tests for each GO category versus the intergenic F/T ratio (CNDs/CNVRs). *P*-values are not corrected for multiple tests.

…future studies aiming to characterize genetic basis of adaptive phenotypic differences between humans and chimpanzees….

…latest news on CNVs differences between any two individuals

➢ 99% of the copy number differences between any two individuals is explained by the simple mendelian inheritance of the same allele from a parent

COMMON VARIANTS

-Only few >100kb

-Copy-number differences between two individuals seems to be due to CNPs: **90% from CNPs with MAF > 1%, 80% from CNPs with MAF > 5%**

-mostly inherited

-patterns of allele frequency, LD, population differentiation that mirror SNPs

-90% biallelic

-10% multiallelic

RARE VARIANTS

-Mostly >100kb

-Affect less than the 5% of the human genome

- Rare CNVs sharing imply sharing of larger genomic region

-fewer allelic states

-highly penetrant and short lived in the population

-De novo/inherited persisting for only a few generations within a pedigree

…next frontier of human genetics research!

1000 genomes projects: a glimpse on lower frequency variants

-**AIM** : a nearly complete catalog of common human genetic variants (defined as frequency 1% or higher) by high-quality sequence data for >85% of the genome for three sets of 400-500 individuals

- Each set is sampled from broad geographic regions (tentatively, Bantu-speaking populations in Africa; populations in East Asia; and populations in Europe).

-This catalog will include SNPs, copy number variants, and short insertion and deletion polymorphisms.

1000 Genomes

A Deep Catalog of Human Genetic Variation

www.1000genomes.org

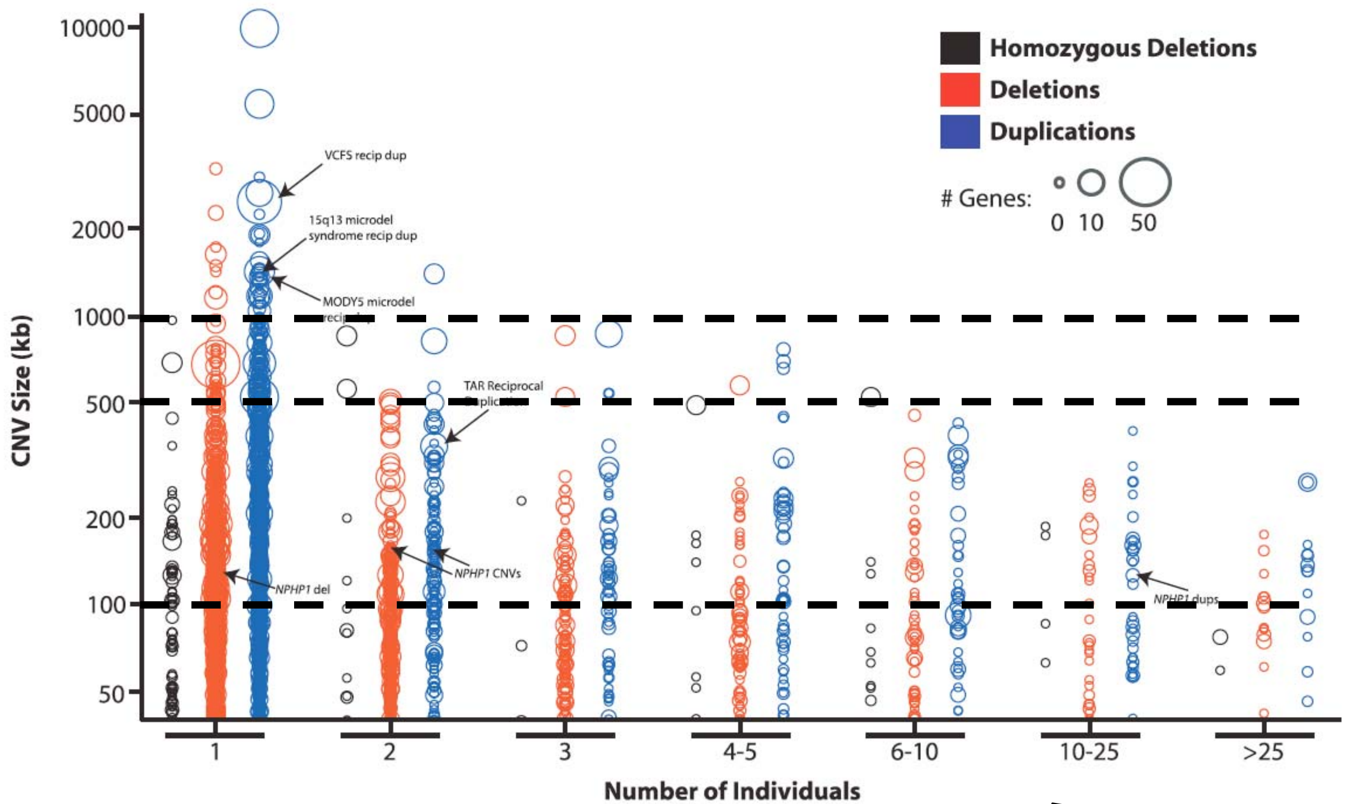
Filling the gap between rare and common variants

ARTICLE

Population Analysis of Large Copy Number Variants and Hotspots of Human Genetic Disease

Andy Itsara,[1,7] Gregory M. Cooper,[1,7] Carl Baker,[1,2] Santhosh Girirajan,[1,2] Jun Li,[2] Devin Absher,[3] Ronald M. Krauss,[4] Richard M. Myers,[3] Paul M. Ridker,[5] Daniel I. Chasman,[5] Heather Mefford,[1] Phyllis Ying,[1] Deborah A. Nickerson,[1] and Evan E. Eichler[1,6,*]

Am J Hum Genet. 2009;84:148-61

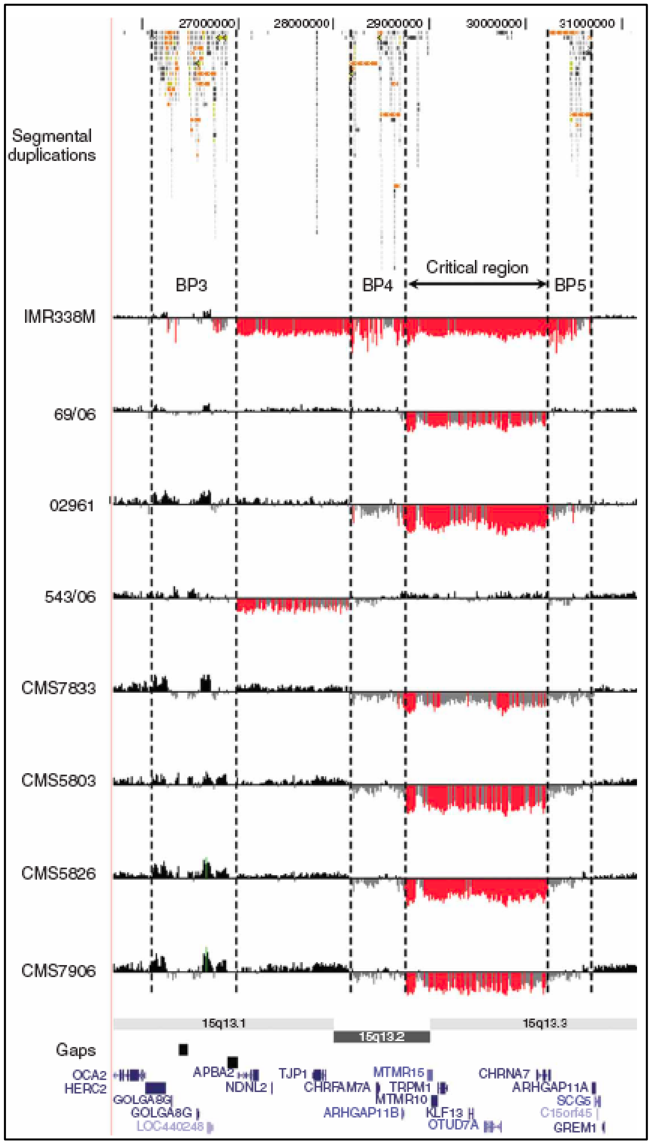


~16% of the genome

Low frequency

High frequency

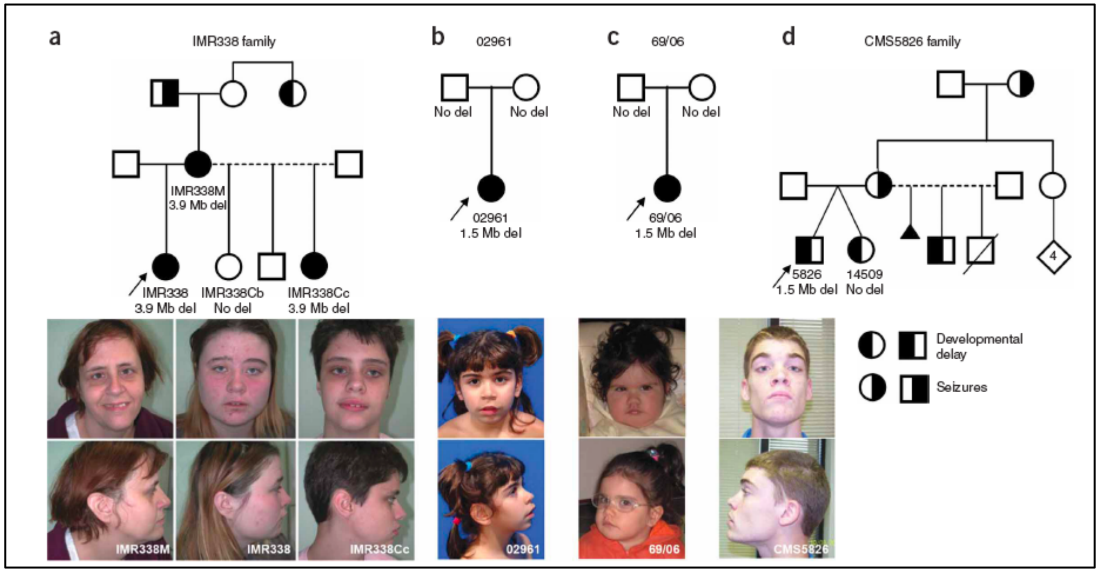
CNVs not only in simple (genomic) diseases



A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures

Andrew J Sharp[1,15], Heather C Mefford[1], Kelly Li[2], Carl Baker[1], Cindy Skinner[3], Roger E Stevenson[3], Richard J Schroer[3], Francesca Novara[4], Manuela De Gregori[4], Roberto Ciccone[4], Adam Broomer[2], Iris Casuga[2], Yu Wang[2], Chunlin Xiao[2], Catalin Barbacioru[2], Giorgio Gimelli[5], Bernardo Dalla Bernardina[6], Claudia Torniero[6], Roberto Giorda[7], Regina Regan[8], Victoria Murday[9], Sahar Mansour[10], Marco Fichera[11], Lucia Castiglia[11], Pinella Failla[11], Mario Ventura[12], Zhaoshi Jiang[1], Gregory M Cooper[1], Samantha J L Knight[8], Corrado Romano[11], Orsetta Zuffardi[4,13], Caifu Chen[2], Charles E Schwartz[3] & Evan E Eichler[1,14]

Nat Genet 40: 322-8, 2008



GENOMIC DISORDERS: sporadic diseases caused by recurrent *de novo* structural mutations

…but also in complex traits

Table 1. Extensions of GWAS to discover CNV–disease associations

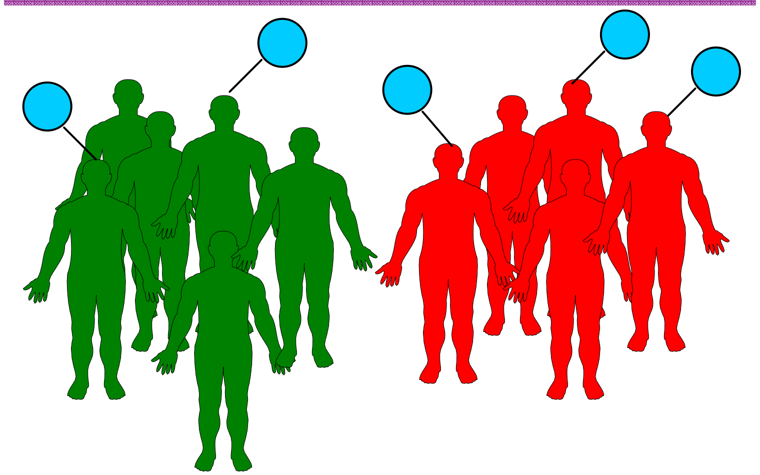
Disease	Analysis approach	Locus	Type of CNV	Size (kb)	Frequency in population	Frequency in Cases	Effect size (OR)	References
Autism	Copy-number analysis of SNP array data	16p11.2	*De novo* deletion	593	1×10^{-4}	1%	100	(27)
Autism	Copy-number analysis of SNP array data	16p11.2	*De novo* duplication	593	3×10^{-4}	0.5%	16	(27)
Schizophrenia	Copy-number analysis of SNP array data	1q21.1	*De novo* deletion	1350	2×10^{-4}	0.3%	15	(25,26)
Schizophrenia	Copy-number analysis of SNP array data	15q13.3	*De novo* deletion	1580	2×10^{-4}	0.2%	12	(25,26)
Schizophrenia	Copy-number analysis of SNP array data	15q11.1	*De novo* deletion	470	0.2%	0.5%	2.7	(25)
Crohn's disease	SNP GWAS + SNP-CNP LD	*IRGM*	Inherited deletion polymorphism	20	7%	10%	1.5	(32)

Note that the *de novo* deletions and duplications above may also be inherited, though *de novo* mutation appeared to explain most or all of the cases in which inheritance could be evaluated. For the schizophrenia findings, frequency and effect size are estimated from the data in Ref. (25); for the autism findings, frequency and effect size are estimated from the replication cohorts in Ref. (27).

McCarroll SA, Hum Mol Genet. 2008 15;17:R135-42

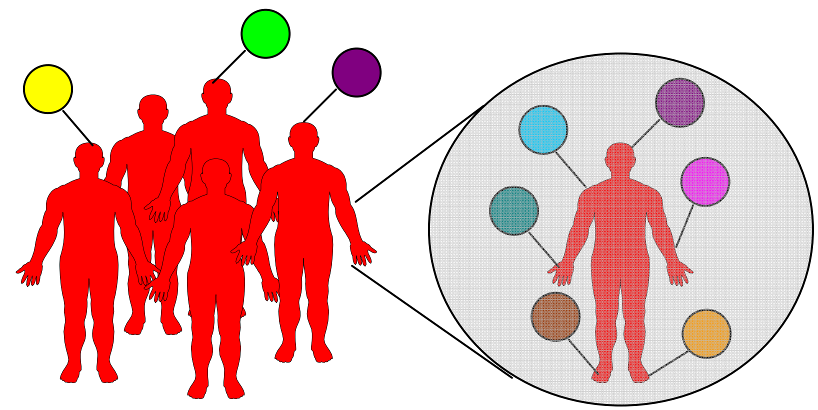
The common versus rare hypothesis

Common variant common disease scenario



VS

Rare variant common disease scenario



Case-control CNVs mapping in shared regions

AIM : to find CNVs segments that are associated to a complex trait in our cohort, looking to replicate this findings in other samples.

SUGGESTED ASSOCIATION ANALYSIS: difference in allele frequency between cases and controls

Rare CNVs mapping in multiple regions defining an individual pattern of variation

AIM : to find CNVs patterns that may be differentially associated to a complex traits (looking for a shared biological function)

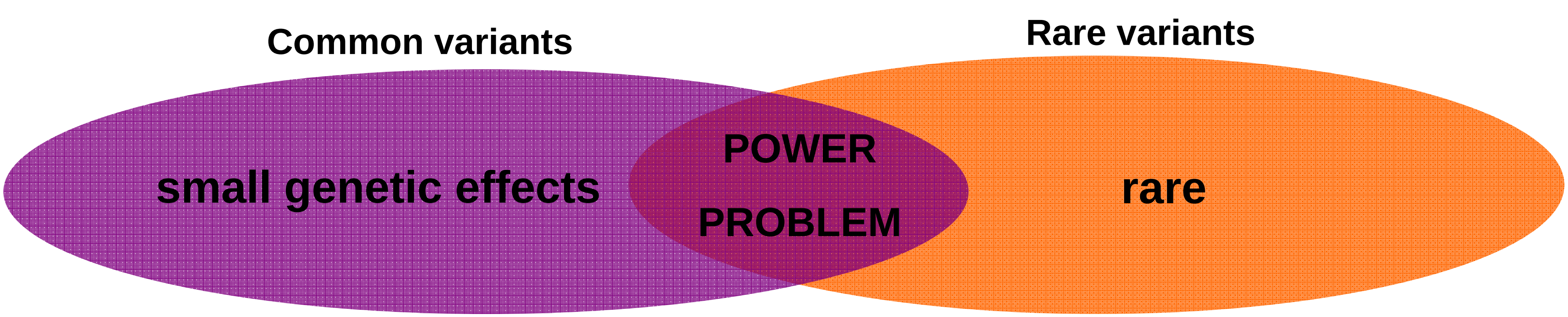
SUGGESTED ASSOCIATION ANALYSIS: enrichment of a collection of rare variants in cases or controls

The schizophrenia paradigm

- Mixture of common and rare variants likely
- Mendelising pedigrees uncommon
  - no common risk variants found to date
  - mathematical modelling of familial risk is inconsistent with single genes of large effects

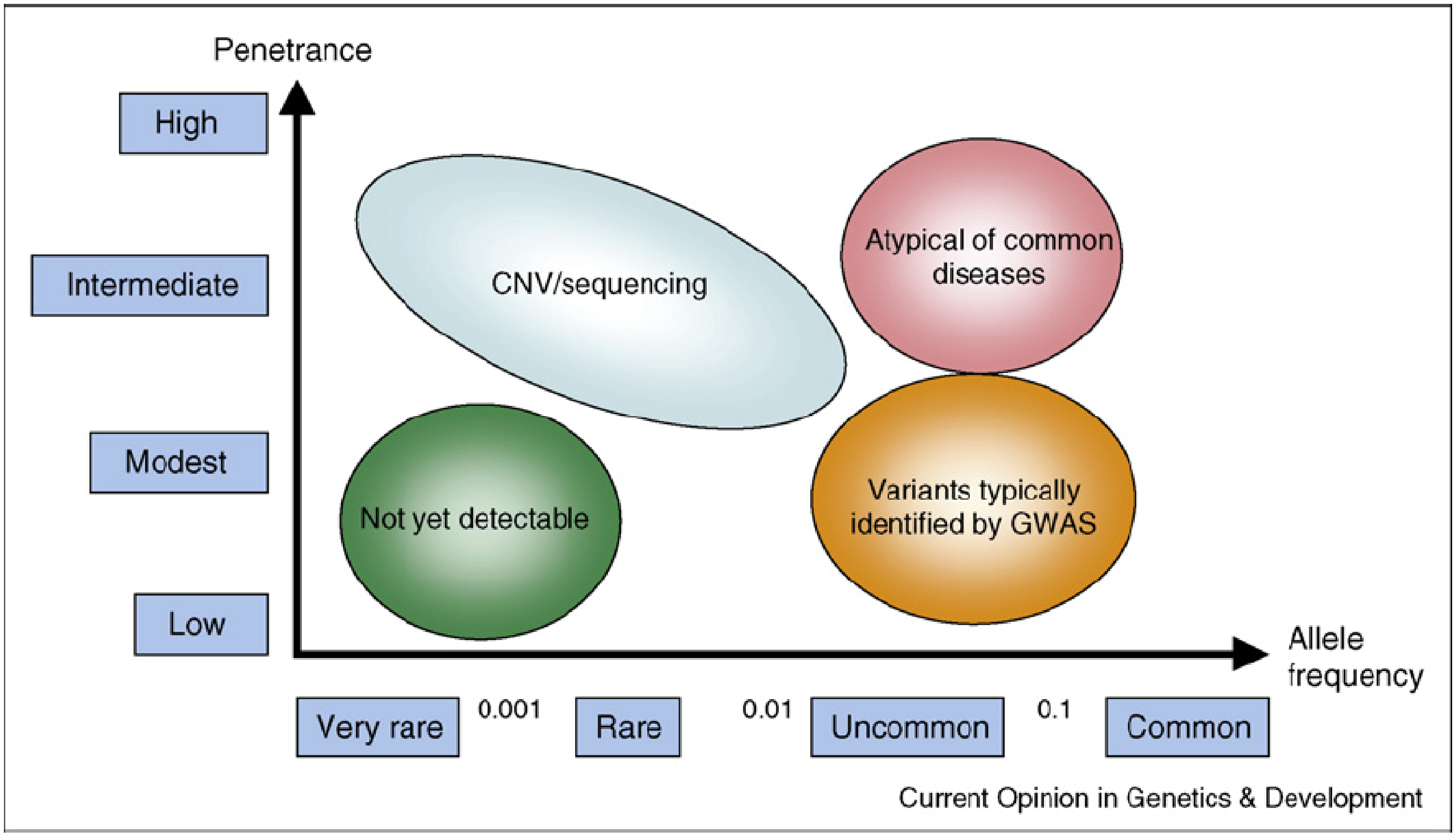
- Apparently stable incidence and prevalence
- Reduced fecundity

Multiple rare recurrent CNVs that underlie strong negative selection may account for a larger fraction of overall genetic risk



…sequencing needed at the single nucleotide/small structural level

…walking through complexity



Owen MJ et al., Curr Opin Genet Dev. 2009

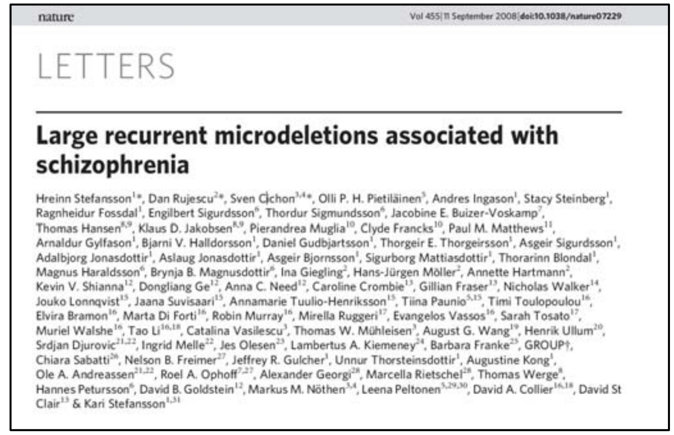
…..converging evidences on loci

Vol 455|11 September 2008|doi:10.1038/nature07239 nature

LETTERS

Rare chromosomal deletions and duplications increase risk of schizophrenia

The International Schizophrenia Consortium*



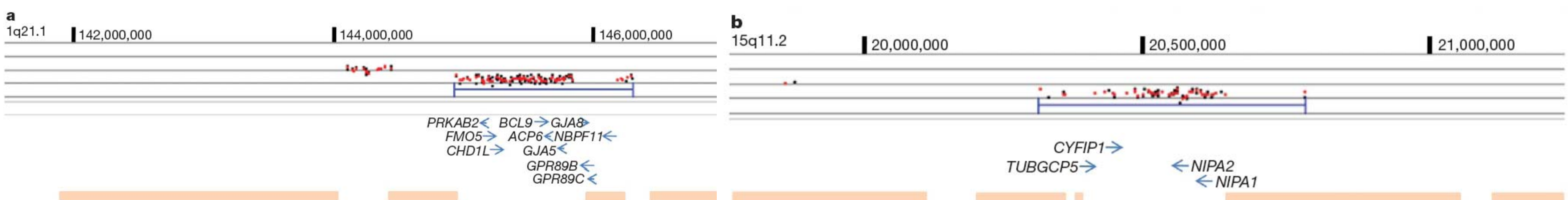
nature Vol 455|11 September 2008|doi:10.1038/nature07229

LETTERS

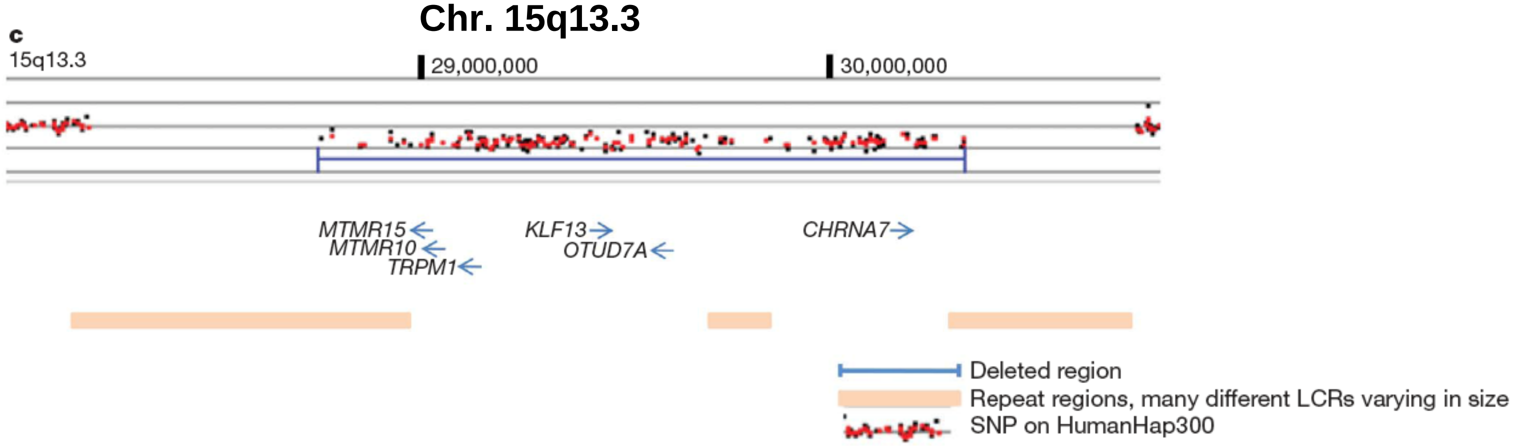
Large recurrent microdeletions associated with schizophrenia

Hreinn Stefansson[1*], Dan Rujescu[2*], Sven Cichon[3,4*], Olli P. H. Pietiläinen[5], Andres Ingason[1], Stacy Steinberg[1], Ragnheidur Fossdal[1], Engilbert Sigurdsson[6], Thordur Sigmundsson[6], Jacobine E. Buizer-Voskamp[7], Thomas Hansen[8,9], Klaus D. Jakobsen[8,9], Pierandrea Muglia[10], Clyde Francks[10], Paul M. Matthews[11], Arnaldur Gylfason[1], Bjarni V. Halldorsson[1], Daniel Gudbjartsson[1], Thorgeir E. Thorgeirsson[1], Asgeir Sigurdsson[1], Adalbjorg Jonasdottir[1], Aslaug Jonasdottir[1], Asgeir Bjornsson[1], Sigurborg Mattiasdottir[1], Thorarinn Blondal[1], Magnus Haraldsson[6], Brynja B. Magnusdottir[6], Ina Giegling[2], Hans-Jürgen Möller[2], Annette Hartmann[2], Kevin V. Shianna[12], Dongliang Ge[12], Anna C. Need[12], Caroline Crombie[13], Gillian Fraser[13], Nicholas Walker[14], Jouko Lonnqvist[15], Jaana Suvisaari[15], Annamarie Tuulio-Henriksson[15], Tiina Paunio[5,15], Timi Toulopoulou[16], Elvira Bramon[16], Marta Di Forti[16], Robin Murray[16], Mirella Ruggeri[17], Evangelos Vassos[16], Sarah Tosato[17], Muriel Walshe[16], Tao Li[16,18], Catalina Vasilescu[3], Thomas W. Mühleisen[3], August G. Wang[19], Henrik Ullum[20], Srdjan Djurovic[21,22], Ingrid Melle[22], Jes Olesen[23], Lambertus A. Kiemeney[24], Barbara Franke[25], GROUP†, Chiara Sabatti[26], Nelson B. Freimer[27], Jeffrey R. Gulcher[1], Unnur Thorsteinsdottir[1], Augustine Kong[1], Ole A. Andreassen[21,22], Roel A. Ophoff[7,27], Alexander Georgi[28], Marcella Rietschel[28], Thomas Werge[8], Hannes Petursson[6], David B. Goldstein[12], Markus M. Nöthen[3,4], Leena Peltonen[5,29,30], David A. Collier[16,18], David St Clair[13] & Kari Stefansson[1,31]

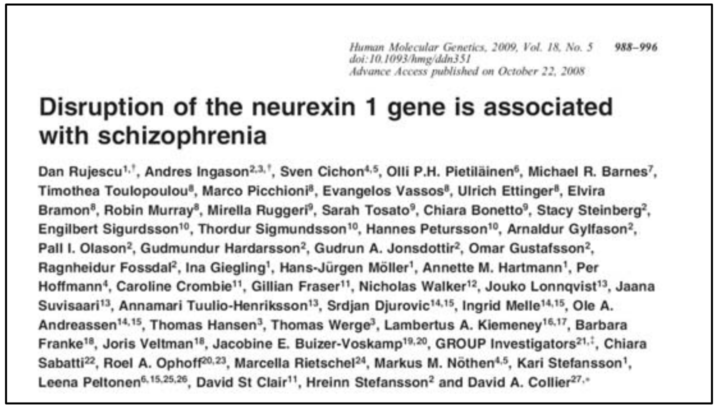
Chr. 1q21.1 **Chr. 15q11.2**



Chr. 15q13.3



…..converging evidences on genes

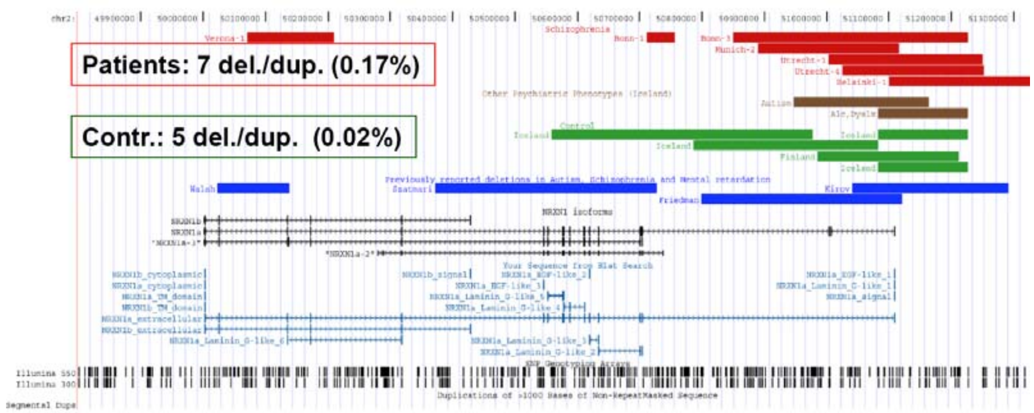


Human Molecular Genetics, 2009, Vol. 18, No. 5 988–996
doi:10.1093/hmg/ddn351
Advance Access published on October 22, 2008

Disruption of the neurexin 1 gene is associated with schizophrenia

Dan Rujescu[1,†], Andres Ingason[2,3,†], Sven Cichon[4,5], Olli P.H. Pietiläinen[6], Michael R. Barnes[7], Timothea Toulopoulou[8], Marco Picchioni[8], Evangelos Vassos[8], Ulrich Ettinger[8], Elvira Bramon[8], Robin Murray[8], Mirella Ruggeri[9], Sarah Tosato[9], Chiara Bonetto[9], Stacy Steinberg[2], Engilbert Sigurdsson[10], Thordur Sigmundsson[10], Hannes Petursson[10], Arnaldur Gylfason[2], Pall I. Olason[2], Gudmundur Hardarsson[2], Gudrun A. Jonsdottir[2], Omar Gustafsson[2], Ragnheidur Fossdal[2], Ina Giegling[1], Hans-Jürgen Möller[1], Annette M. Hartmann[1], Per Hoffmann[4], Caroline Crombie[11], Gillian Fraser[11], Nicholas Walker[12], Jouko Lonnqvist[13], Jaana Suvisaari[13], Annamari Tuulio-Henriksson[13], Srdjan Djurovic[14,15], Ingrid Melle[14,15], Ole A. Andreassen[14,15], Thomas Hansen[3], Thomas Werge[3], Lambertus A. Kiemeney[16,17], Barbara Franke[18], Joris Veltman[18], Jacobine E. Buizer-Voskamp[19,20], GROUP Investigators[21,‡], Chiara Sabatti[22], Roel A. Ophoff[20,23], Marcella Rietschel[24], Markus M. Nöthen[4,5], Kari Stefansson[1], Leena Peltonen[6,15,25,26], David St Clair[11], Hreinn Stefansson[2] and David A. Collier[27,*]

Analysis only with CNVs that disrupt *NRXN1* exons



Cochran-Mantel-Haenszel test: p=0.0027; OR=8.97 [1.8-51.9]

Rujescu et al., Hum. Mol. Genet. 2008

… genomic burden of rare variants

Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia

Tom Walsh,[1]* Jon M. McClellan,[2]*† Shane E. McCarthy,[3]* Anjené M. Addington,[4]* Sarah B. Pierce,[1] Greg M. Cooper,[5] Alex S. Nord,[5] Mary Kusenda,[3,6] Dheeraj Malhotra,[3] Abhishek Bhandari,[3] Sunday M. Stray,[1] Caitlin F. Rippey,[5] Patricia Roccanova,[3] Vlad Makarov,[3] B. Lakshmi,[3] Robert L. Findling,[7] Linmarie Sikich,[8] Thomas Stromberg,[4] Barry Merriman,[9] Nitin Gogtay,[4] Philip Butler,[4] Kristen Eckstrand,[4] Laila Noory,[4] Peter Gochman,[4] Robert Long,[4] Zugen Chen,[9] Sean Davis,[10] Carl Baker,[5] Evan E. Eichler,[5] Paul S. Meltzer,[10] Stanley F. Nelson,[9] Andrew B. Singleton,[11] Ming K. Lee,[1] Judith L. Rapoport,[4] Mary-Claire King,[1,5] Jonathan Sebat[3]

Science 320: 539-543, 2008

Genes disrupted by SV breakpoints

NBPF10
SLC35F3, TARBP1
STON1-GTF2A1L
ERBB4
GRM7
PRKCD
-
SKP2, SLC1A3
MAGI2, PHTF2
SLC12A9, CAV1
PRKAG2, MLL3
PTK2
SMARCA2
-
-
HIPK3, C11orf41
DLG2
-
LAMA1, PTPRM
TMC4
LARGE

-
-
FRRS1
-
ROBO1
MANEA
-
FLJ31818
SND1
CTSB
MPDZ
SOX5, LYRM5
TMTC1
HYDIN
BPIL2

LETTERS

nature genetics

Strong association of *de novo* copy number mutations with sporadic schizophrenia

Bin Xu[1,2], J Louw Roos[3], Shawn Levy[4], E J van Rensburg[5], Joseph A Gogos[1,6] & Maria Karayiorgou[2]

Nat Rev Genet 40: 8881-885, 2008

Detecting CNVs is not a trivial issue

- **Circular Binary Segmentation** (Olshen et al., 2004)

  - cnvPartition (Illumina Inc., San Diego, CA)

- **Circular Binary Segmentation Modified**

  -Nexus Copy Number (Biodiscovery Inc., El Segundo, CA)

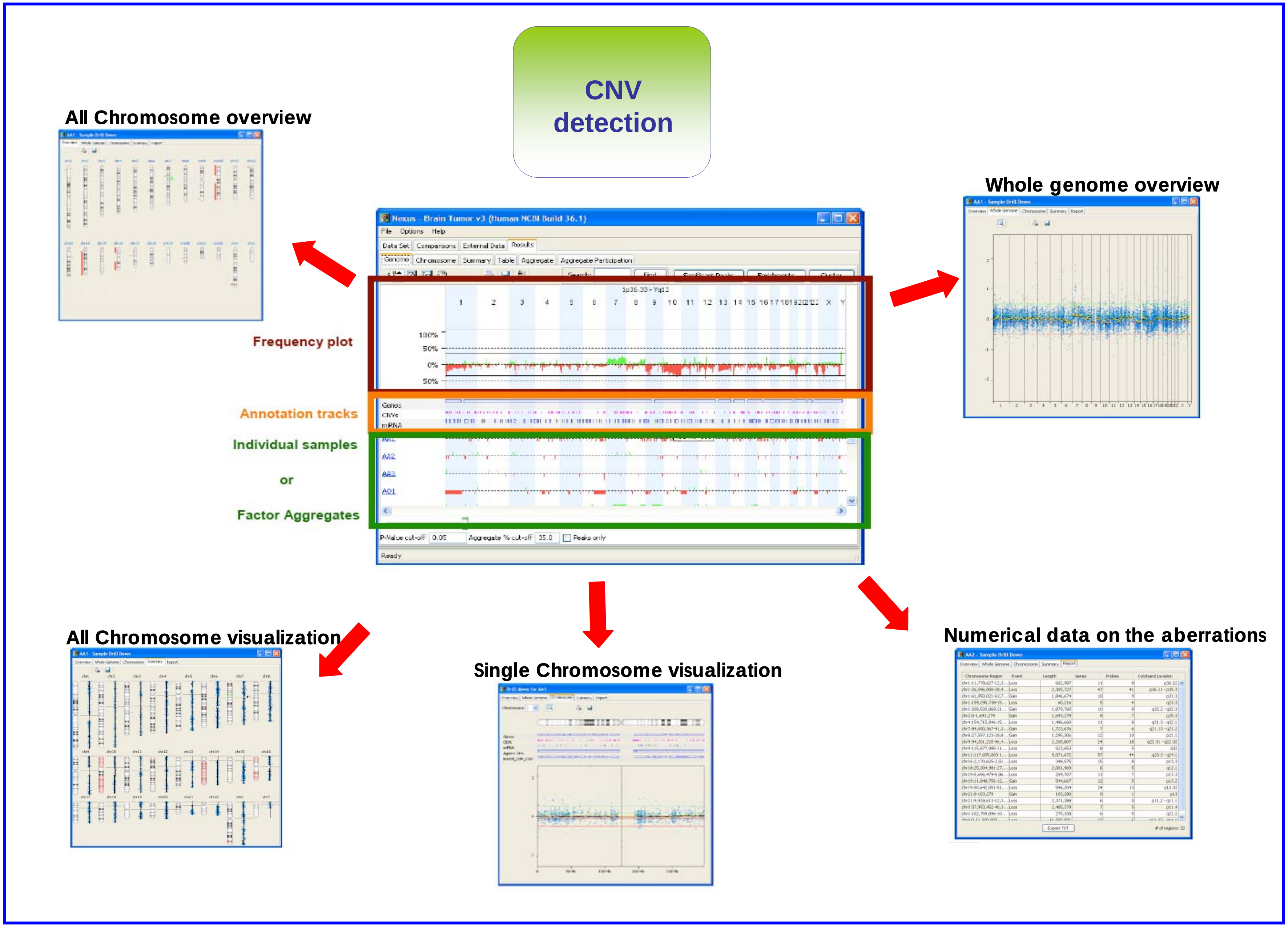
- **Hidden Markov Models** (HMMs)

  - QuantiSNP (Wellcome Trust for Human Genetics, University of Oxford)

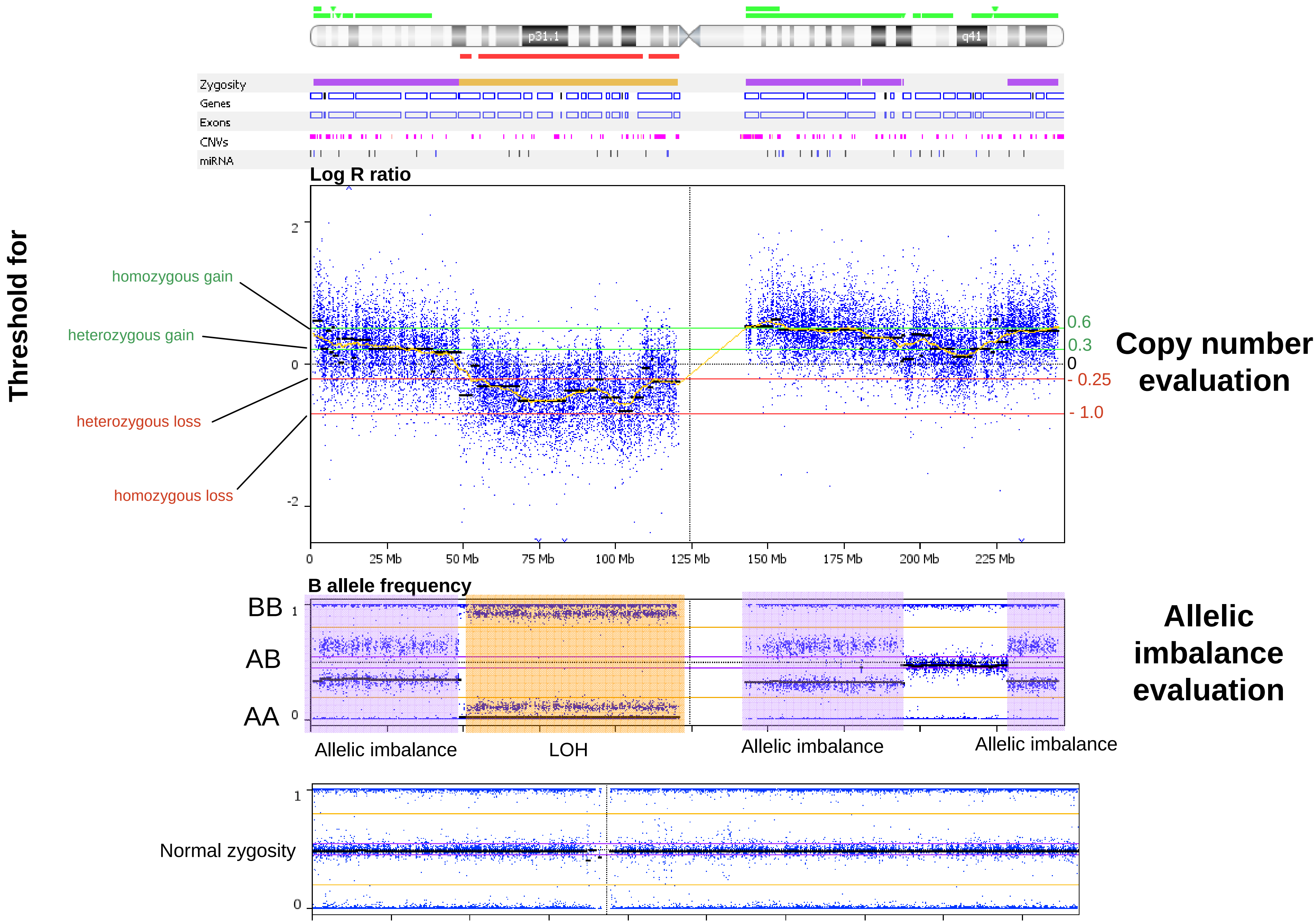
  - PennCNV (University of Pennsylvania)

- **Golden Helix Dynamic Optimal Segmentation** (Golden Helix Inc.)

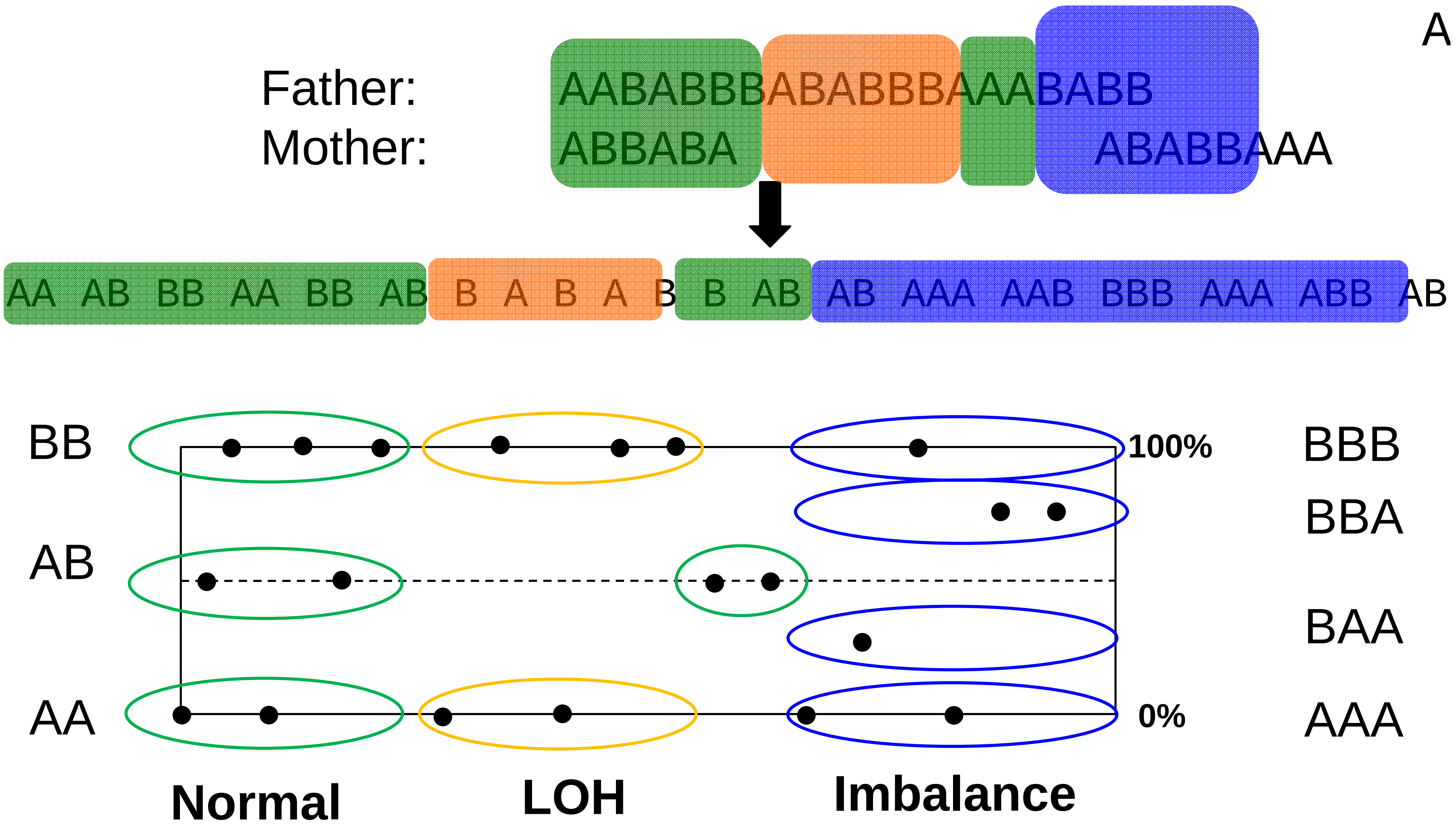
Nexus v4 (Biodiscovery Inc.)



Nexus parameters



B-allele frequency possible scenarios



Our approaches to CNVs topic

1) **Schizophrenia** : exploring the two different sides

2) **Validation of a *de novo* genomic loss of heterozygosity *(LOH)*** for HLA haplotype as a mechanism of in vivo leukemia immune escape in patients with relapse after haploidentical transplantation and adoptive transfer of donor T cells

3) **Novel gene discovery** in monogenic cancer predisposition syndromes: a first proof of concept of our analysis pipeline

Our approaches to CNVs topic

1) **Schizophrenia** : exploring the two different sides

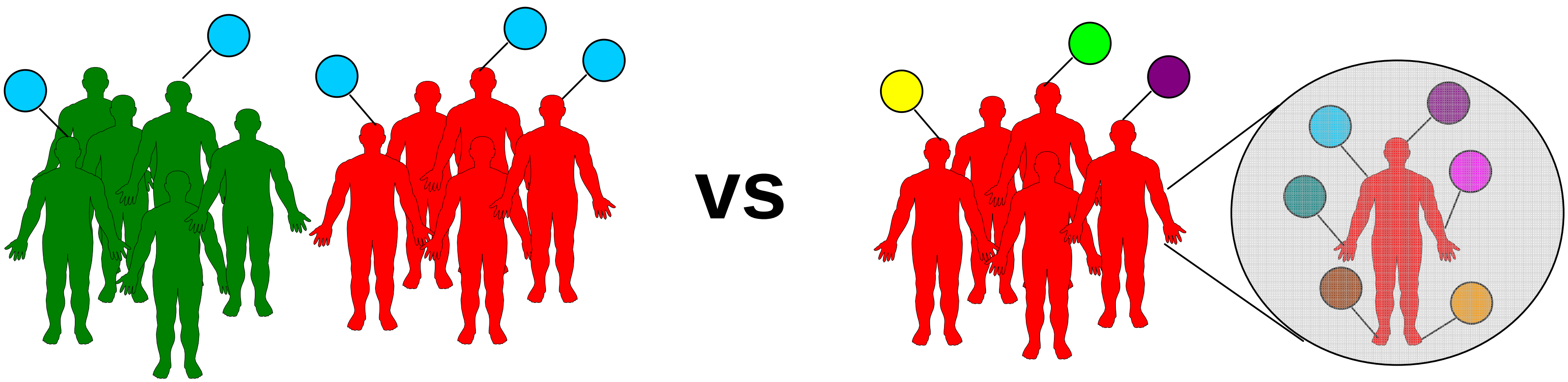
2) **Validation of a *de novo* genomic loss of heterozygosity *(LOH)*** for HLA haplotype as a mechanism of in vivo leukemia immune escape in patients with relapse after haploidentical transplantation and adoptive transfer of donor T cells

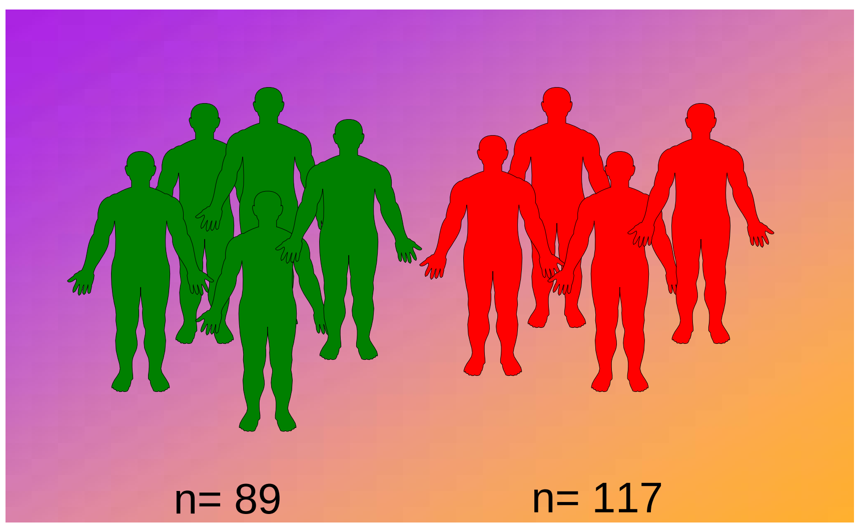
3) **Novel gene discovery** in monogenic cancer predisposition syndromes: a first proof of concept of our analysis pipeline

Schizophrenia : exploring the two different sides

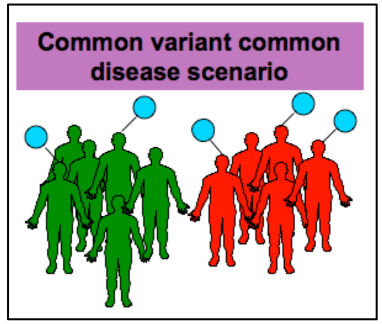
Common variant common disease scenario

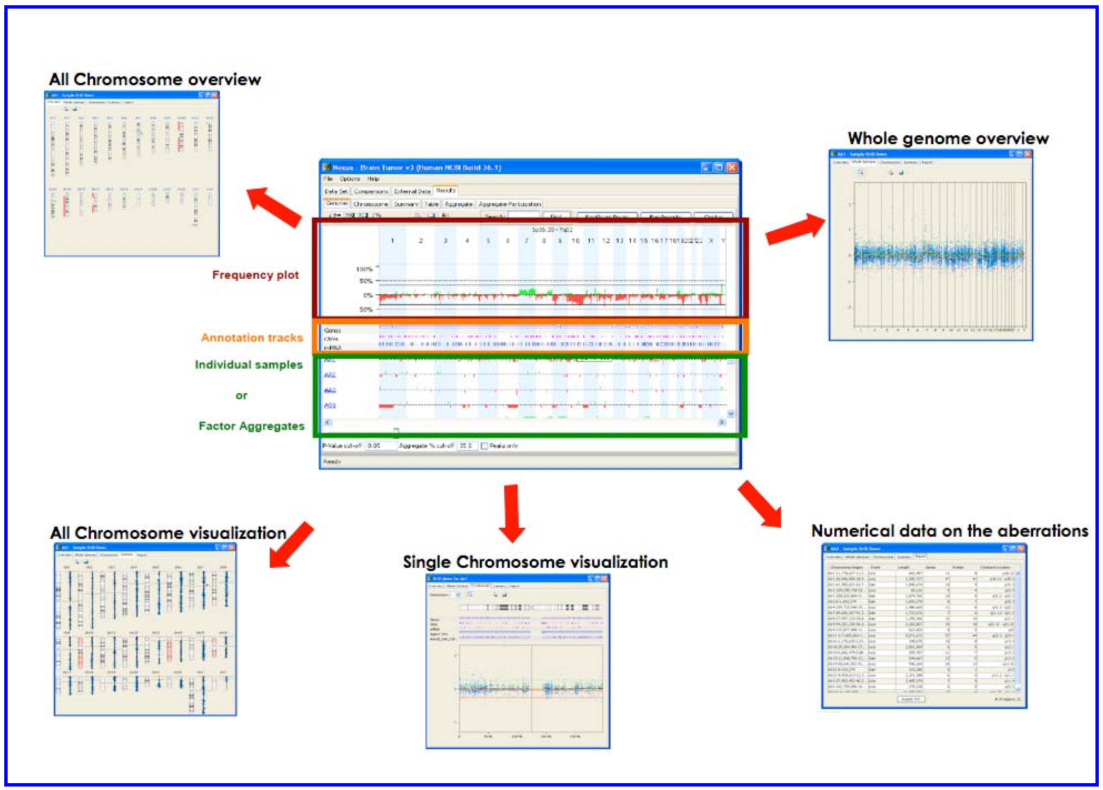
Rare variant common disease scenario



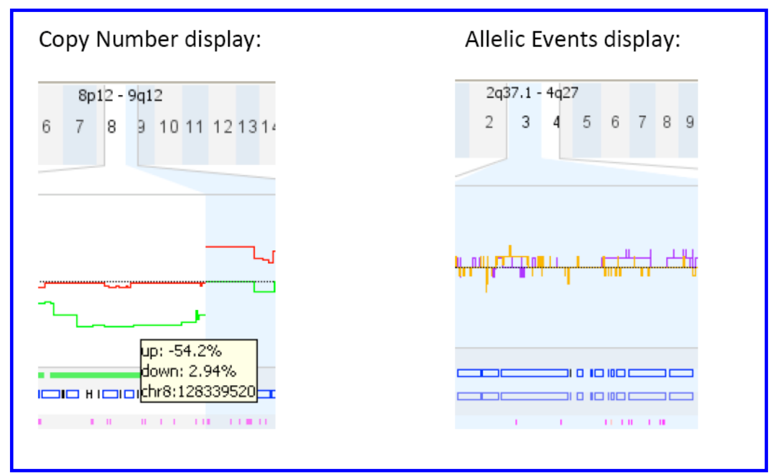


Nexus v4 (Biodiscovery Inc.): the comparison feature





CNV detection → Frequency Comparison



COMPARISON FEATURE: comparison in CNVs frequency between groups assessing which CNVs are significantly enriched

…are we moving in the right direction?

ORIGINAL ARTICLE

High-density SNP association study and copy number variation analysis of the *AUTS1* and *AUTS5* loci implicate the *IMMP2L–DOCK4* gene region in autism susceptibility

E Maestrini[1,11], AT Pagnamenta[2,11], JA Lamb[2,3,11], E Bacchelli[1], NH Sykes[2], I Sousa[2], C Toma[1], G Barnby[2], H Butler[2], L Winchester[2], TS Scerri[2], F Minopoli[1], J Reichert[4], G Cai[4], JD Buxbaum[4], O Korvatska[5], GD Schellenberg[6], G Dawson[7,8], A de Bildt[9], RB Minderaa[9], EJ Mulder[9], AP Morris[2], AJ Bailey[10] and AP Monaco[2], IMGSAC[12]

Mol Psychiatry. 2009

- Only SNP replicated maps in DOCK4
- DOCK4 activates Rac GTPase
- Predominant expression in hyppocampus
- Upregulated at the same time dendrites start growing
- Knockdown results in impaired dendritic morphogenesis

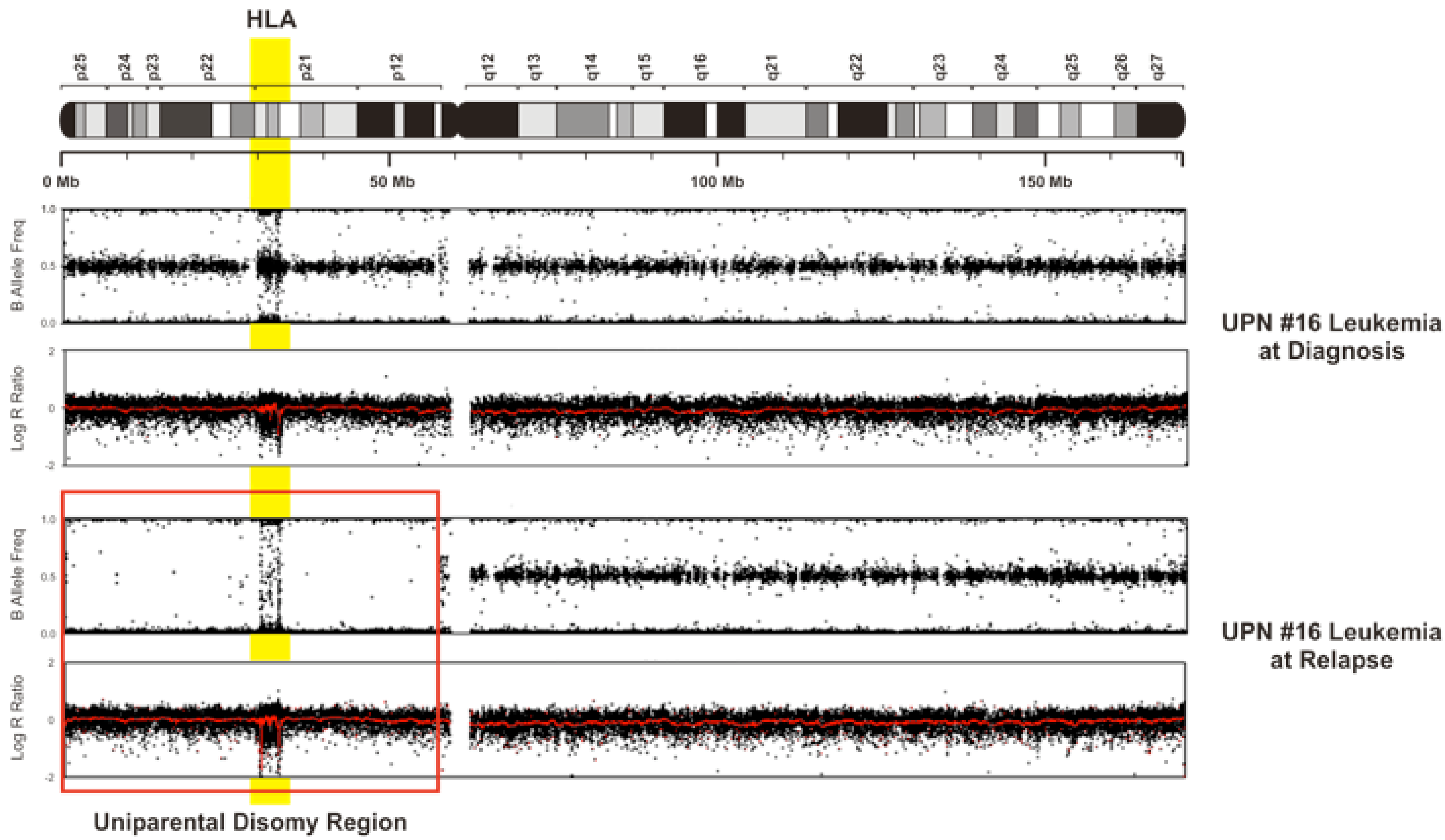
Our approaches to CNVs topic

1) **Schizophrenia** : exploring the two different sides

2) **Validation of a *de novo* genomic loss of heterozygosity *(LOH)*** for HLA haplotype as a mechanism of in vivo leukemia immune escape in patients with relapse after haploidentical transplanation and adoptive transfer of donor T cells

3) **Novel gene discovery** in monogenic cancer predisposition syndromes: a first proof of concept of our analysis pipeline

Genomic loss of mismatched HLA in leukemia is a major mechanism of in vivo escape from T cell immunosurveillance following haploidentical HSCT



Vago L. et al., Genomic Loss Of Mismatched HLA In Leukemia is a Major Mechanism of in Vivo Escape from T Cell Immunosurveillance Following Haploidentical HSCT, NEJM, in press

Our approaches to CNVs topic

1) **Schizophrenia** : exploring the two different sides

2) **Validation of a *de novo* genomic loss of heterozygosity *(LOH)*** for HLA haplotype as a mechanism of in vivo leukemia immune escapein patients with relapse after haploidentical transplanation and adoptive transfer of donor T cells

3) **Novel gene discovery** in monogenic cancer predisposition syndromes: a first proof of concept of our analysis pipeline

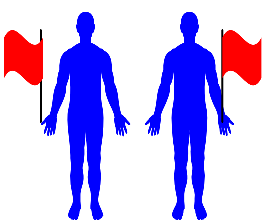
Gene discovery: looking for a needle in a haystack…

… knowing the needle to search for

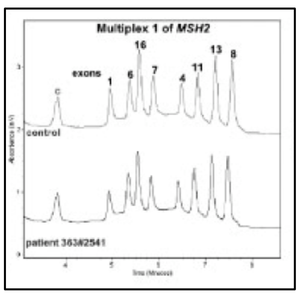
Reporter pathogenic CNV in a monogenic cancer predisposition syndrome (HNPCC)

=

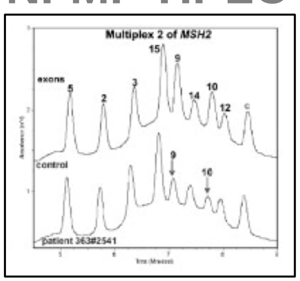
20-55 % of patients are negative for mutation/rearrangements in candidate genes and display a phenotype superimposable to known monogenic cancer predisposition syndromes



MLPA



NFMP-HPLC

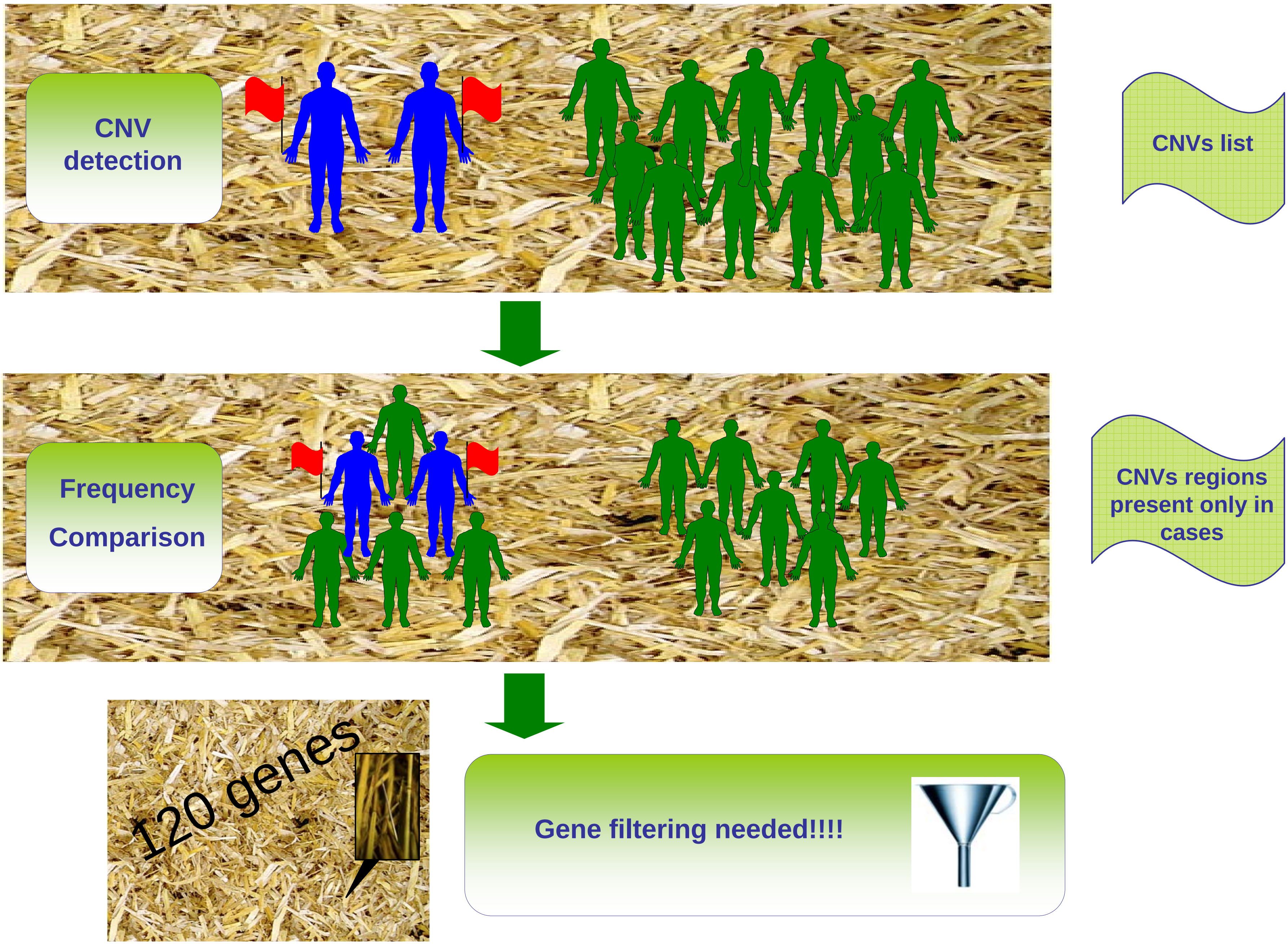


\+

Human1m-duo BeadChips

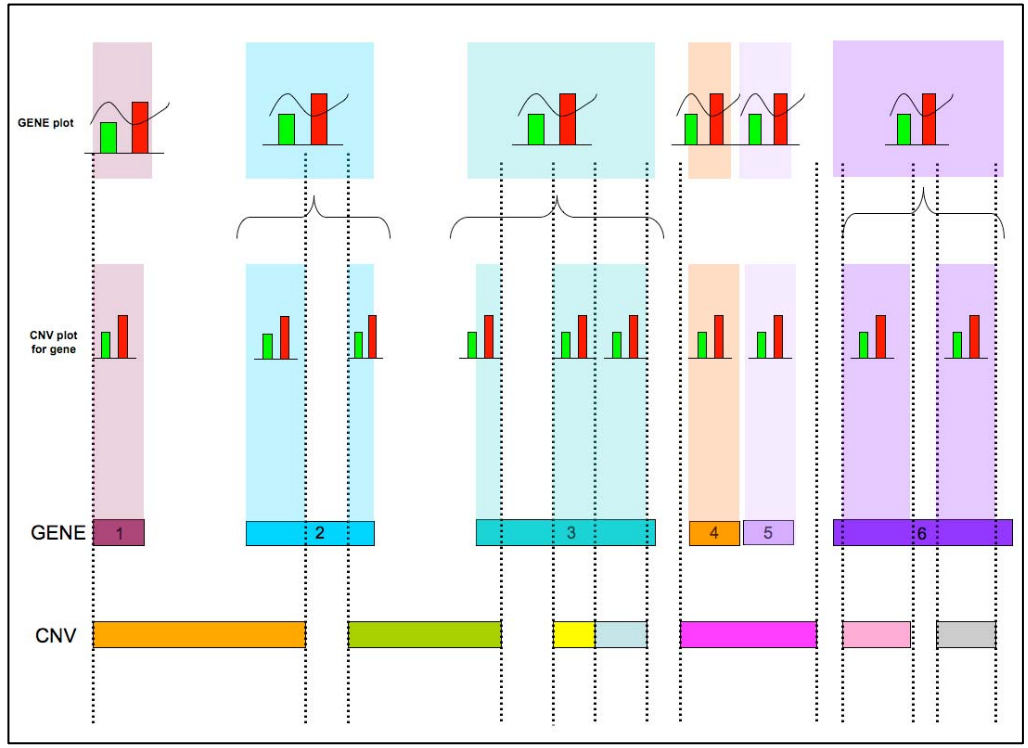
- pathogenic CNV boundaries ortogonal validation
- empirical testing of theoretical parameters to optimize analysis conditions

Gene discovery : towards the gene point of view

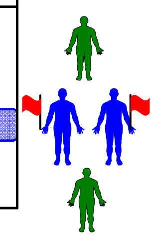


Gene discovery : towards the gene point of view

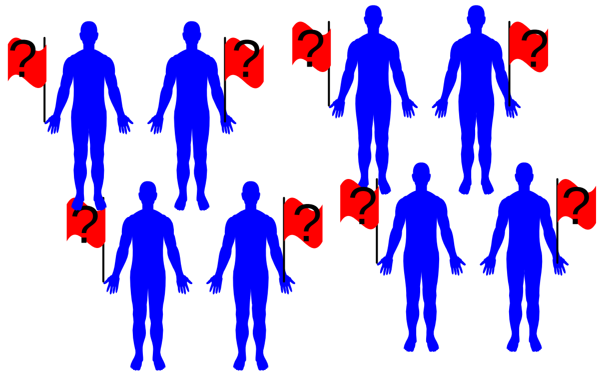
Gene filtering



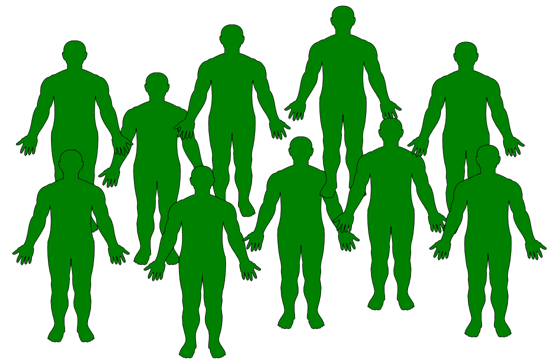
Gene	Region IDs	Region Code Explicited	Cumulative Freq Cases	Cumulative Freq Controls
GKAP1	0	chr9:85,616,440-85,623,279	7,5	0
CKMT1B	5,031,166	chr15:41,676,268-41,682,935 <-> chr15:41,660,781-41,676,268	5	0
ANO3	0	chr11:26,565,073-26,567,094	5	0
MSH2	702,0	chr2:47,460,187-47,511,188 <-> chr2:47,539,880-47,547,541	5	0
LOC728855	0	chr1:147,779,233-147,849,072	5	0
PPIAL4C	0	chr1:147,779,233-147,849,072	5	0



Gene discovery : identification of novel cancer predisposition genes



Mutation negative in candidate genes patients cohort (20-55 %)



CNV detection

Frequency Comparison

Gene filtering

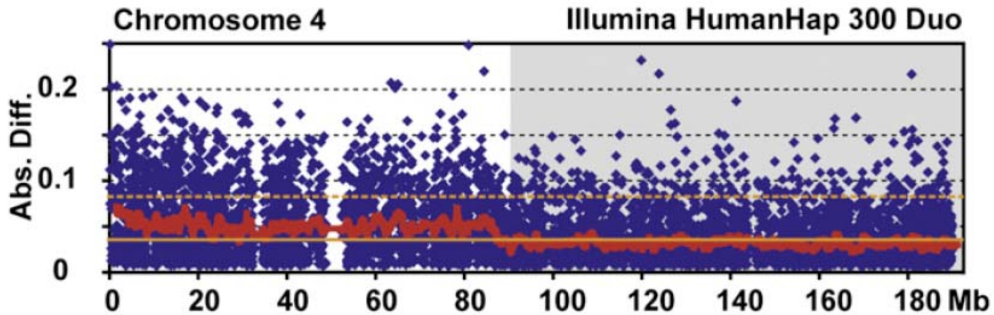
CNVs genomic content is different between…

- Monozygotic twins: frequency of differences in genetic makeup up to 10 %!!!

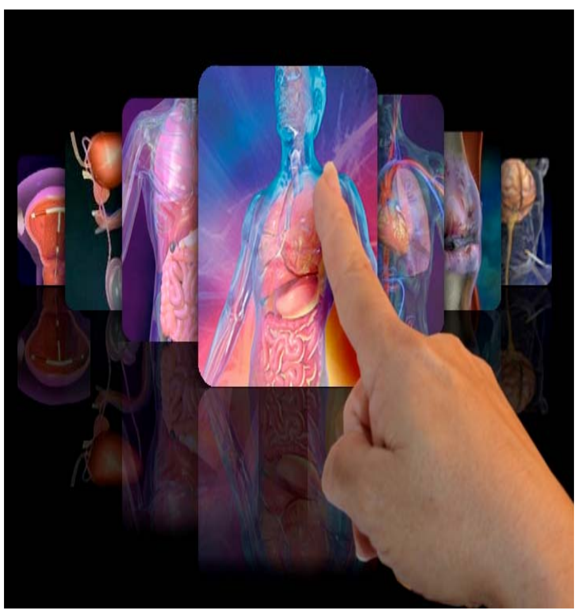
Am J Hum Genet. 2008;82:763-71 REPORT

Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy-Number-Variation Profiles

Carl E.G. Bruder,[1,*] Arkadiusz Piotrowski,[1] Antoinet A.C.J. Gijsbers,[2,3] Robin Andersson,[4] Stephen Erickson,[5] Teresita Diaz de Ståhl,[6] Uwe Menzel,[6] Johanna Sandgren,[7] Desiree von Tell,[1] Andrzej Poplawski,[1] Michael Crowley,[1] Chiquito Crasto,[1] E. Christopher Partridge,[1] Hemant Tiwari,[5] David B. Allison,[1,5] Jan Komorowski,[4] Gert-Jan B. van Ommen,[2,3] Dorret I. Boomsma,[8] Nancy L. Pedersen,[9] Johan T. den Dunnen,[2,3] Karin Wirdefeldt,[9] and Jan P. Dumanski[1,6]



- Different tissues within an individual (mosaicism)



Gene	Locus	Disease	Reference
COL3A1	2q31	Ehlers Danlos Syndrome Type IV	Milewicz et al., 1993
VHL	3p25	von-Hippel-Lindau Disease	Sgambati et al., 2000
D4Z4 repeat	4q35	Facioscapulohumeral Muscular Dystrophy	Upadhyaya et al., 1995
RB	13q14	Retinoblastoma	Sippel et al., 1998
TSC2	16p13	Tuberous Sclerosis	Kozlowski et al., 2007
CREBBP	16p13	Rubinstein-Taybi Syndrome	Gervasini et al., 2007
NF1	17q11	Neurofibromatosis Type 1	Lazaro et al., 1994
SOX9	17q24	Campomelic Dysplasia	Smyk et al., 2007
DMD	Xp21	Duchenne Muscular Dystrophy	Bakker et al., 1987
CYBB	Xp21	Chronic Granulomatous Disease	Faizunnessa et al., 1997
F8C	Xq28	Hemophilia A	Higuchi et al., 1988
DKC1	Xq28	X-linked Dyskeratosis Congenita	Higuchi et al., 1988
IKBKG	Xq28	Incontinentia Pigmenti	Kenwrick et al., 2001

Notini AJ et al., Cyt Genome Res 2008 123:270-277

Somatic Mosaicism for Copy Number Variation in Differentiated Human Tissues

Arkadiusz Piotrowski,[1,4] Carl E.G. Bruder,[1] Robin Andersson,[2] Teresita Diaz de Ståhl,[3] Uwe Menzel,[3] Johanna Sandgren,[3] Andrzej Poplawski,[1] Desiree von Tell,[1] Chiquito Crasto,[1] Adam Bogdan,[4] Rafal Bartoszewski,[5] Zsuzsa Bebok,[5] Maciej Krzyzanowski,[6] Zbigniew Jankowski,[6] E. Christopher Partridge,[1] Jan Komorowski,[2] and Jan P. Dumanski[1,3*]

[1]Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama; [2]Linnaeus Centre for Bioinformatics, Uppsala University, Uppsala, Sweden; [3]Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden; [4]Department of Biology and Pharmaceutical Botany, Medical University of Gdansk, Gdansk, Poland; [5]Department of Cell Biology, University of Alabama at Birmingham, Birmingham, Alabama; [6]Department of Forensic Medicine, Medical University of Gdansk, Gdansk, Poland

Piotrowski A. et al., . Hum Mutat. 2008;29:1118-24.

Many-but far from all-common CNVs can be interrogated with a tag SNP on genotyping platforms

Systematic assessment of copy number variant detection via genome-wide SNP genotyping

Gregory M Cooper[1,3], Troy Zerr[1,3], Jeffrey M Kidd[1], Evan E Eichler[1,2] & Deborah A Nickerson[1]

Nat Genet 40: 1199-1203, 2008

Integrated detection and population-genetic analysis of SNPs and copy number variation

Steven A McCarroll[1–4,10], Finny G Kuruvilla[1–4,10], Joshua M Korn[1–6], Simon Cawley[7], James Nemesh[1], Alec Wysoker[1], Michael H Shapero[7], Paul I W de Bakker[1,4,8], Julian B Maller[3], Andrew Kirby[3], Amanda L Elliott[1], Melissa Parkin[1], Earl Hubbell[7], Teresa Webster[7], Rui Mei[7], James Veitch[7], Patrick J Collins[7], Robert Handsaker[1], Steve Lincoln[7], Marcia Nizzari[1], John Blume[7], Keith W Jones[7], Rich Rava[7], Mark J Daly[1,3,4,9], Stacey B Gabriel[1] & David Altshuler[1–4,9]

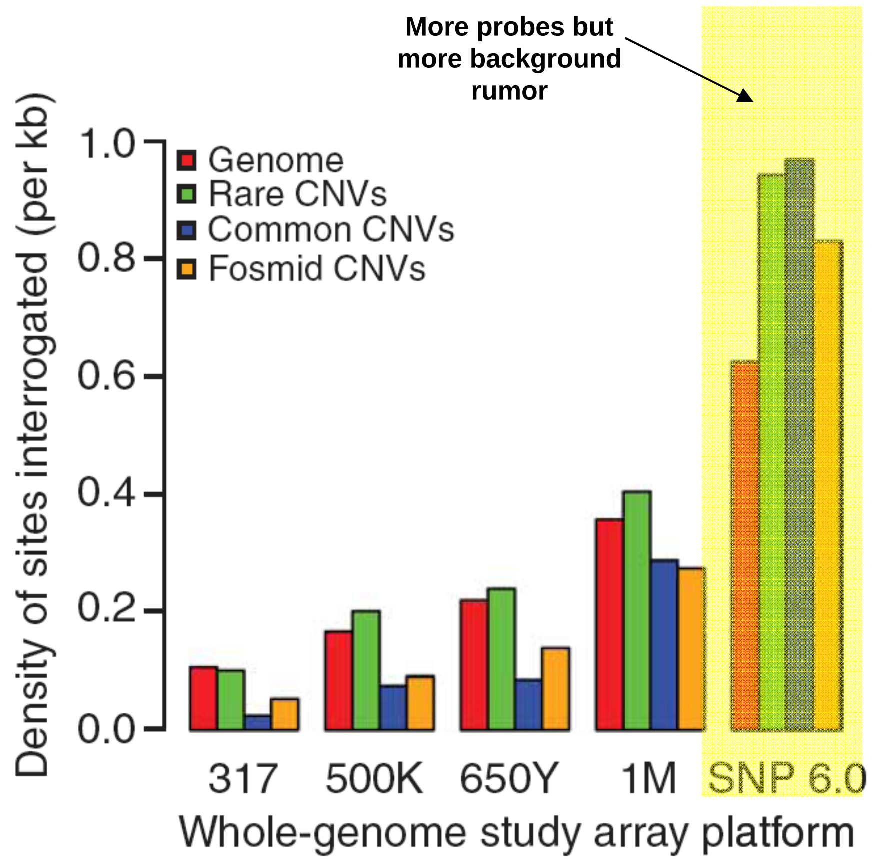
Nat Genet 40: 1166-74, 2008

- 82% of the common deletions (worldwide frequency >5%) strongly correlated to a HapMap SNP (worldwide $r^2 > 0.8$)

High density genome-wide SNP array tagged about ~50% of the common deletions

- Segmental duplication under-representation or cross-hybridization of paralogous sequences

NAHR-driven mutation less likely to be tagged LD with surrounding SNPs: underestimation

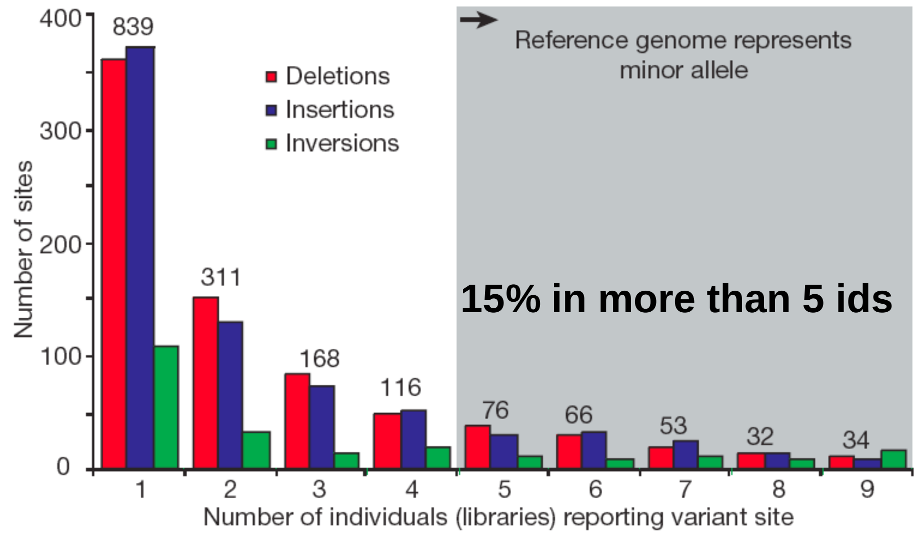


Towards deep sequencing….

Mapping and sequencing of structural variation from eight human genomes

Jeffrey M. Kidd[1], Gregory M. Cooper[1], William F. Donahue[2], Hillary S. Hayden[3], Nick Sampas[4], Tina Graves[5], Nancy Hansen[6], Brian Teague[7], Can Alkan[1], Francesca Antonacci[1], Eric Haugen[3], Troy Zerr[1], N. Alice Yamada[4], Peter Tsang[4], Tera L. Newman[1], Eray Tüzün[1], Ze Cheng[1], Heather M. Ebling[2], Nadeem Tusneem[2], Robert David[2], Will Gillett[3], Karen A. Phelps[3], Molly Weaver[1], David Saranga[2], Adrianne Brand[2], Wei Tao[2], Erik Gustafson[2], Kevin McKernan[2], Lin Chen[1], Maika Malig[1], Joshua D. Smith[1], Joshua M. Korn[8], Steven A. McCarroll[8], David A. Altshuler[8], Daniel A. Peiffer[9], Michael Dorschner[1], John Stamatoyannopoulos[1], David Schwartz[7], Deborah A. Nickerson[1], James C. Mullikin[6], Richard K. Wilson[5], Laurakay Bruhn[4], Maynard V. Olson[3], Rajinder Kaul[3], Douglas R. Smith[2] & Evan E. Eichler[1]

Nature 453: 56-64, 2008



• 1695 sites of structural variation: 49 % more than 1 ids, 15% more than 5

• 525 regions of novel insertion sequences not present in the reference genome: 40% are CNVs ranging (from few kbp up to 130 kb)

• nearly 50% lay outside previously reported CNVs region

• Comparison with the highest density SNP commercial platform (Illumina 1M and Affy6): 50% deletions cannot be detected!

DE NOVO SEQUENCING OF ADDITIONAL GENOMES NEEDED!

Looking to individual structural variants with sequencing technologies

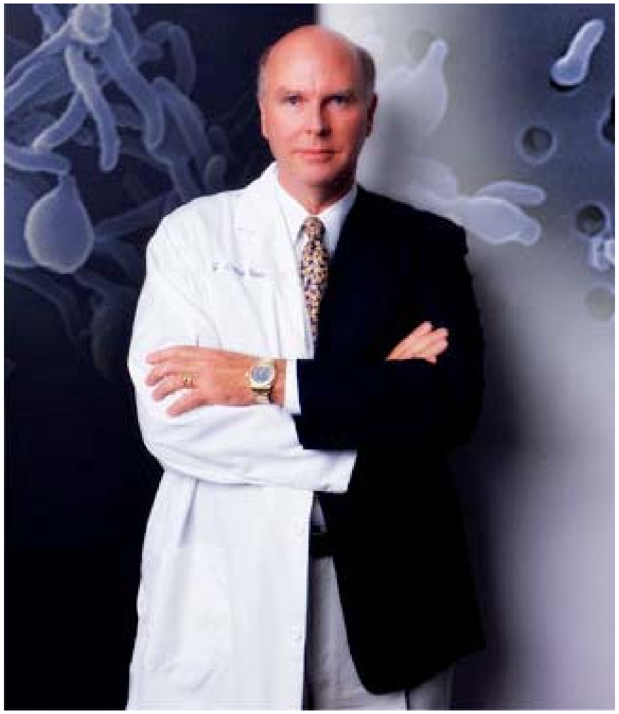
OPEN ACCESS Freely available online PLOS BIOLOGY

The Diploid Genome Sequence of an Individual Human

Samuel Levy[1*], Granger Sutton[1], Pauline C. Ng[1], Lars Feuk[2], Aaron L. Halpern[1], Brian P. Walenz[1], Nelson Axelrod[1], Jiaqi Huang[1], Ewen F. Kirkness[1], Gennady Denisov[1], Yuan Lin[1], Jeffrey R. MacDonald[2], Andy Wing Chun Pang[2], Mary Shago[2], Timothy B. Stockwell[1], Alexia Tsiamouri[1], Vineet Bafna[3], Vikas Bansal[3], Saul A. Kravitz[1], Dana A. Busam[1], Karen Y. Beeson[1], Tina C. McIntosh[1], Karin A. Remington[1], Josep F. Abril[4], John Gill[1], Jon Borman[1], Yu-Hui Rogers[1], Marvin E. Frazier[1], Stephen W. Scherer[2], Robert L. Strausberg[1], J. Craig Venter[1]

1 J. Craig Venter Institute, Rockville, Maryland, United States of America, 2 Program in Genetics and Genomic Biology, The Hospital for Sick Children, and Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada, 3 Department of Computer Science and Engineering, University of California San Diego, La Jolla, California, United States of America, 4 Genetics Department, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain

PLoS Biol. 2007 4;5:e254.

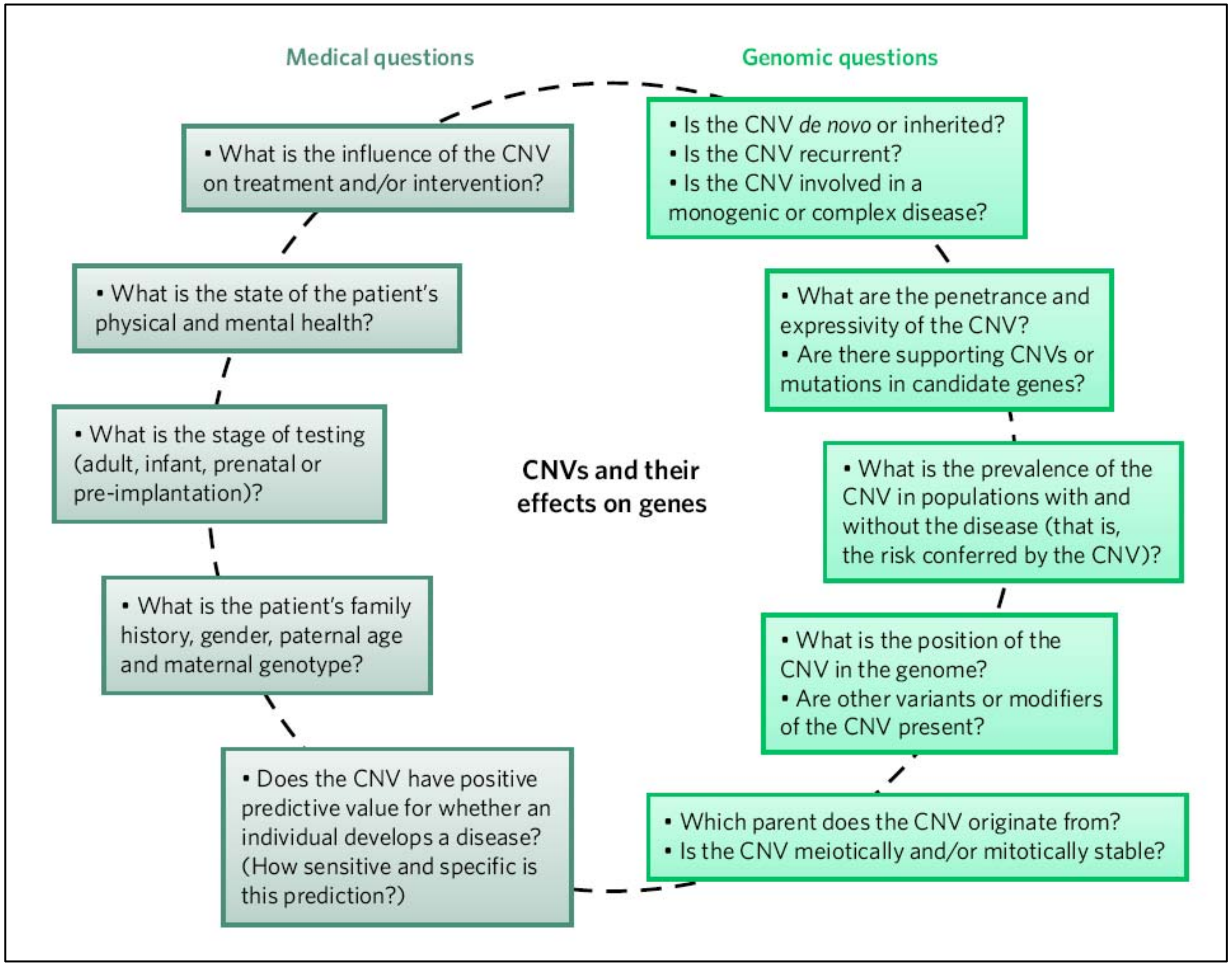


4.1 million DNA variants (~12.3 Mb)

1.288.319 (~30%) novel!

- 3,213,401 single nucleotide polymorphisms (SNPs)
- 53,823 block substitutions (2–206 bp)
- 559,473 homozygous indels (1–82,711 bp)
- 62 CNVs
- 292,102 heterozygous insertion/deletion events (indels)(1–571 bp)
- 90 inversions
- **Non-SNP DNA variation accounts for 22% of all events identified in the donor, however they involve 74% of all variant bases**

Assessing the relevance of CNVs in clinic



Cook et al., Nature. 2008 16;455:919-23

Aknowledgements

Laboratory of genetic of complex traits, University of Milan
Erika Salvi
Sara Lupoli
Cristina Barlassina
Luisa Strik Lievers
Valentina Tieran
Francesca Taddeo
Cristian Cosentino
Fausta Schiavini
Daniele Cusi
Fabio Macciardi

Istituto Tecnologie Biomediche, CNR
Alessandro Orro
Andrea Calabria
Paolo Cozzi

UC Irvine, CA
Steven G. Potkin
Jessica Turner
James Fallon
Guia Guffanti

Department of Oncology and Neurosciences, University ‘‘G. D’Annunzio,’’ Chieti
Laura De Lellis
Alessandro Cama

Istituto Nazionale dei Tumori, Milan
Paolo Radice

IRCCS San Raffaele
Luca Vago
Fabio Ciceri
Katharina Fleischauer

Illumina, Inc.
Mita Mancini
Chiara Dal Fiume

Biodiscovery, Inc.
Soheil Shams