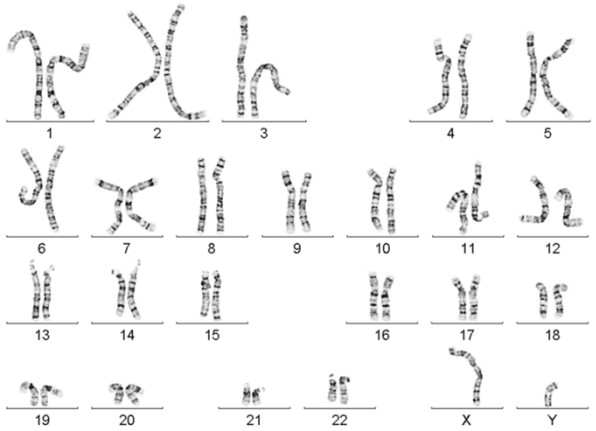
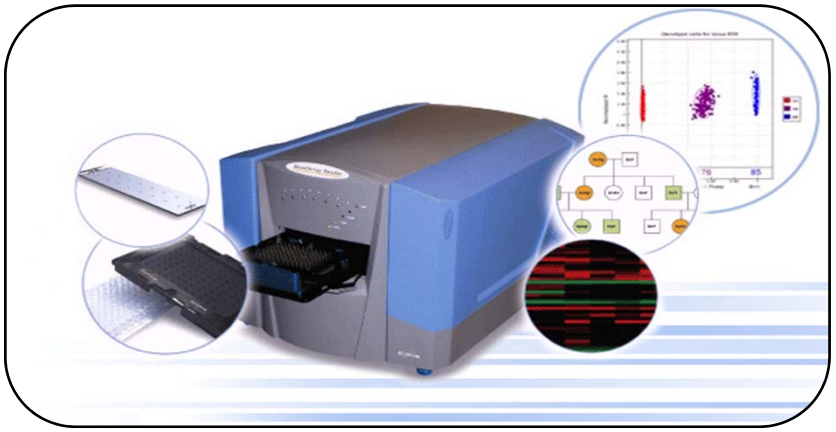




The Importance of Bioinformatics in the Detection of Cognitive Disorders

Topics in Bioinformatics

Frank Kooy

Universiteit Antwerpen

Université de Liège

Cognitive Genetics
Centre of Medical Genetics
University of Antwerp

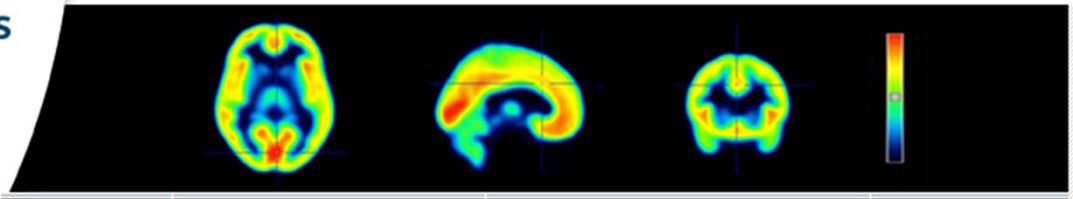
University of Antwerp

Search

Nederlands

Cognitive Genetics

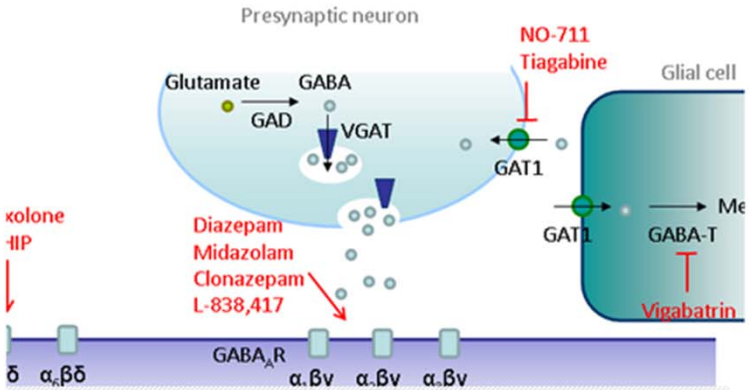
Centre of Medical Genetics
University of Antwerp



Research Interests | People | For Students | Voor de patiënt | Contact

You are here: UAntwerp > Research groups > Cognitive Genetics

Tweet Like



The GABA(A) Receptor as a Therapeutic Target for Neurodevelopmental disorders. Neuron 86:1119–1130, 2015

Research Interests

- Fragile X syndome
- Dynamic mutations
- Novel syndromes and disorders
- Gene interaction networks
- Bioinformatics

Voor de patiënt

- Patienten organisaties
- Behandeling van het fragiele X syndroom

Cognitive Genetics

The term cognitive genetics was introduced to bridge the concepts of genetics and cognitive processes in the first decade of the 21st century. This relatively novel branch of genetics studies the influence of genetic variation on cognition and central nervous system disorders.

Mission

Our mission is to identify genetic causes of cognitive disorders and to study the molecular defects in order to eventually develop rational therapy.

Embedding

Our research group Cognitive Genetics is part of the research cluster Medical Genetics of the Faculty of Pharmaceutical, Biomedical and Veterinary Sciences of the University of Antwerp. The research cluster, in combination with diagnostic and clinical units forms the Center of Medical Genetics. We are part of the Universty of Antwerp research excellence center GENOMED.

Databases

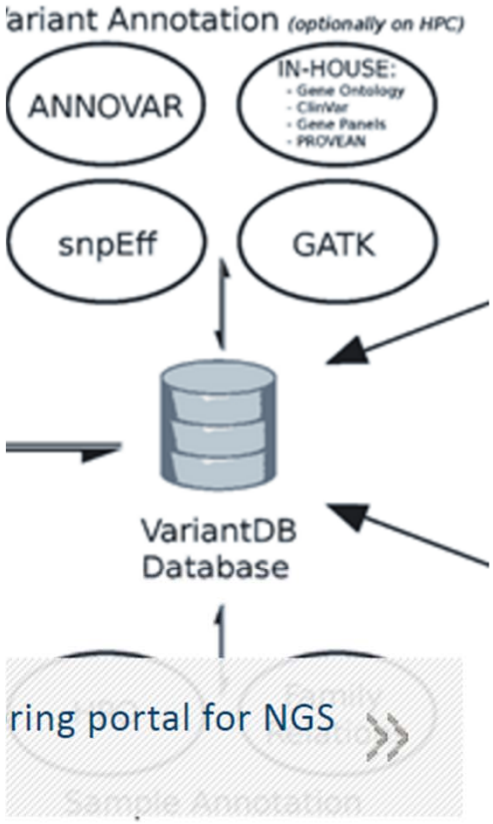
Maintained by the group:

VariantDB
NGS data analysis suite

Galaxy@BioMina
Customised microarray and NGS data analysis tools

CNV-WebStore
MicroArray data analysis and interpretation

ADNP gene
Phenotype information



Cognitive Genetics

The term cognitive genetics was introduced to bridge the concepts of genetics and cognitive processes in the first decade of the 21st century. This relatively novel branch of genetics studies the influence of genetic variation on cognition and central nervous system disorders.

Mission

Our mission is to identify genetic causes of cognitive disorders and to study the molecular defects in order to eventually develop rational therapy.

Databases

Maintained by the group:

VariantDB
NGS data analysis suite

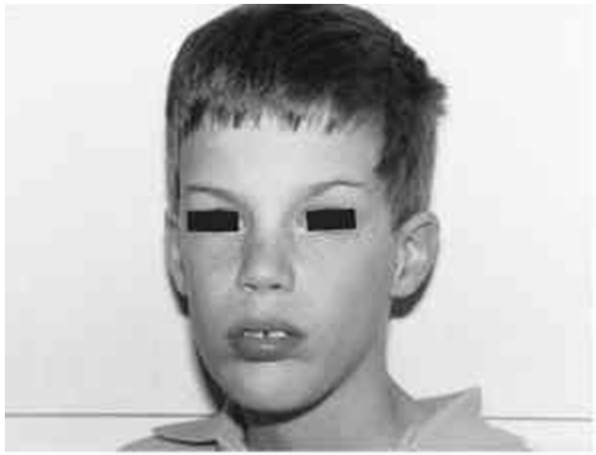
Galaxy@BioMina
Customised microarray and NGS data analysis tools

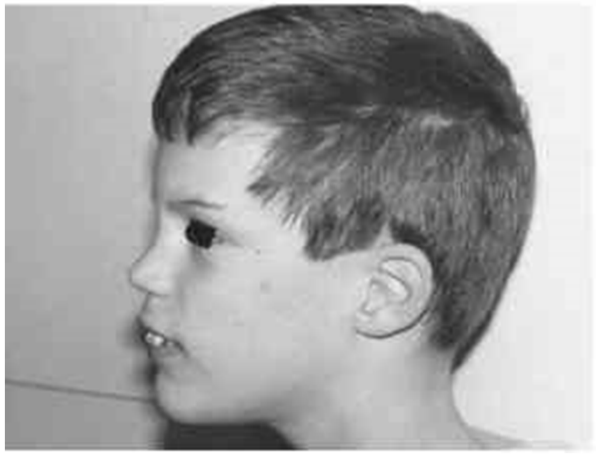
CNV-WebStore

Intellectual disability

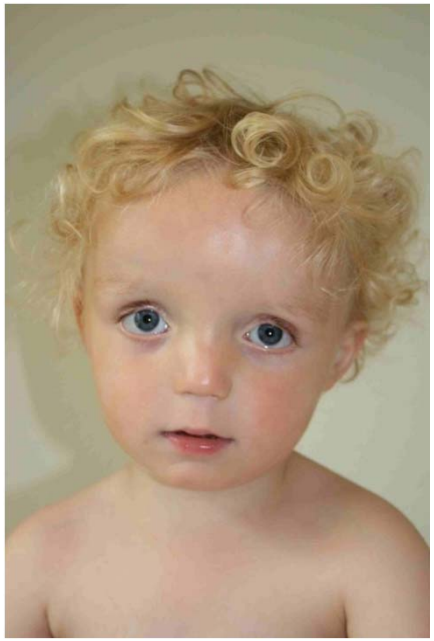
Formerly “mental retardation”

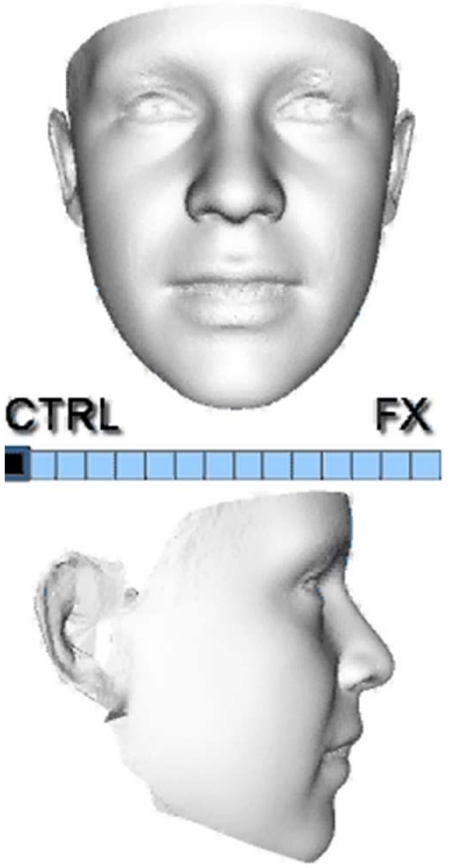
IQ below 70





- Diagnosed in childhood
- In combination with adaptive limitations

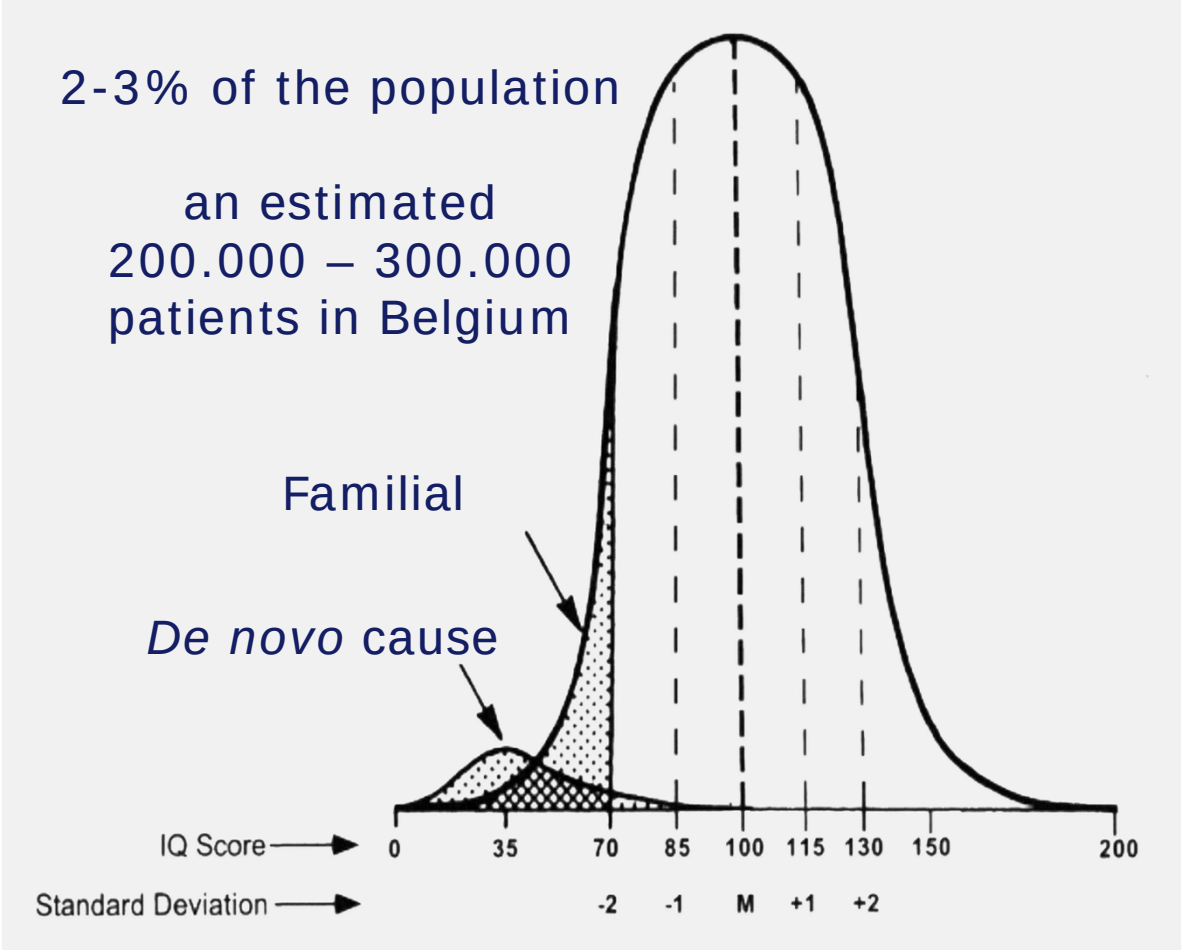




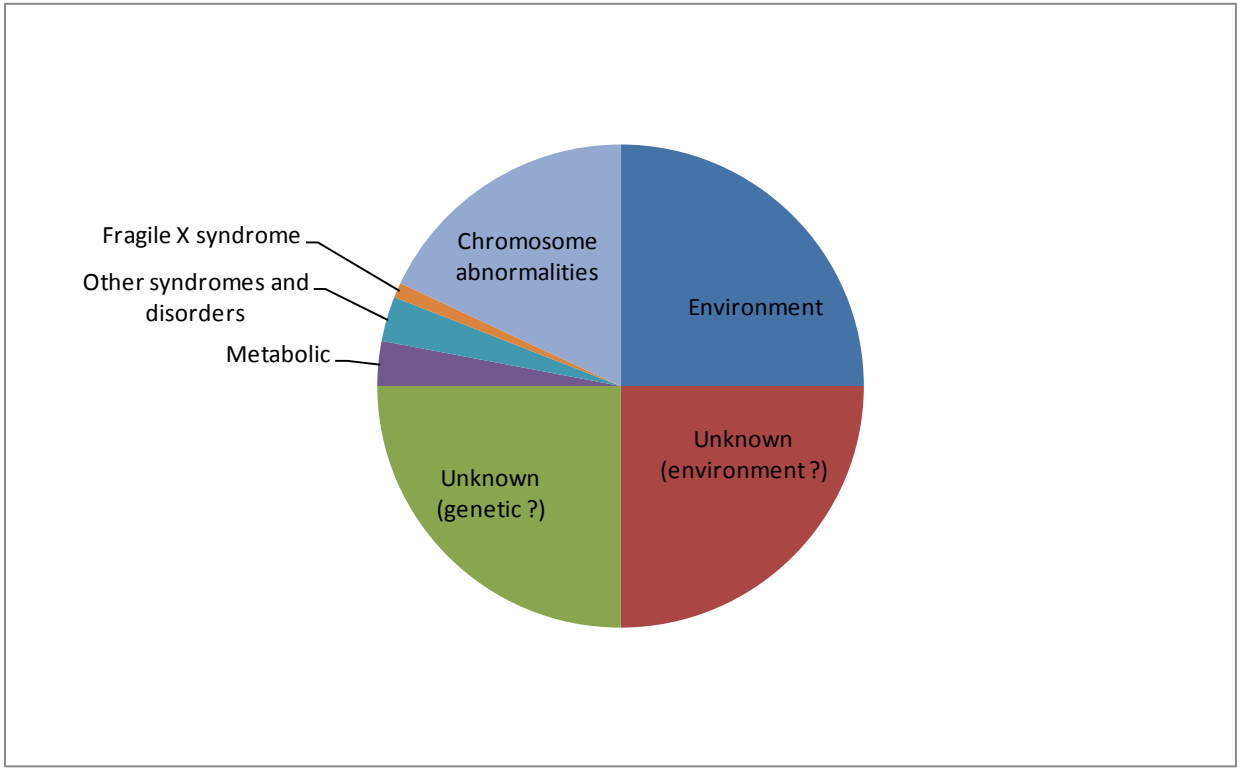
Frequency of Intellectual disability

2-3% of the population

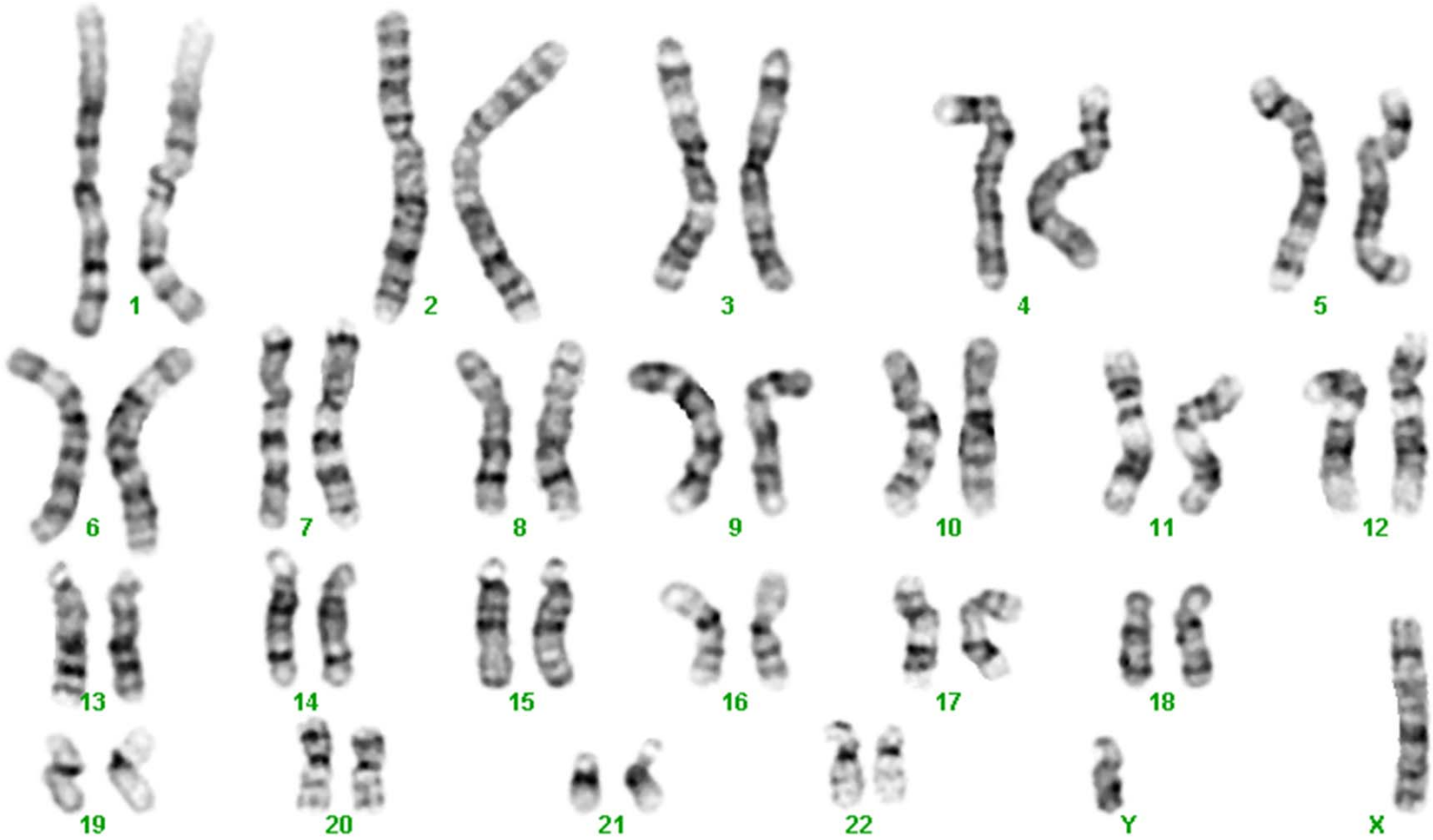
an estimated
200.000 – 300.000
patients in Belgium



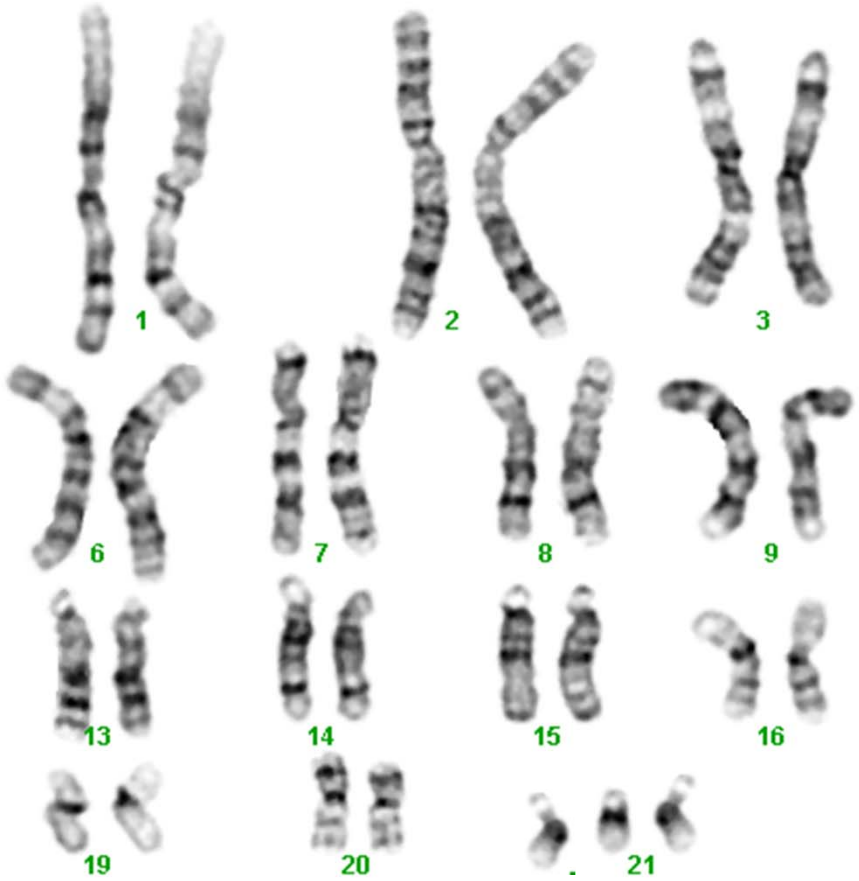
Causes of Intellectual disability a reflection

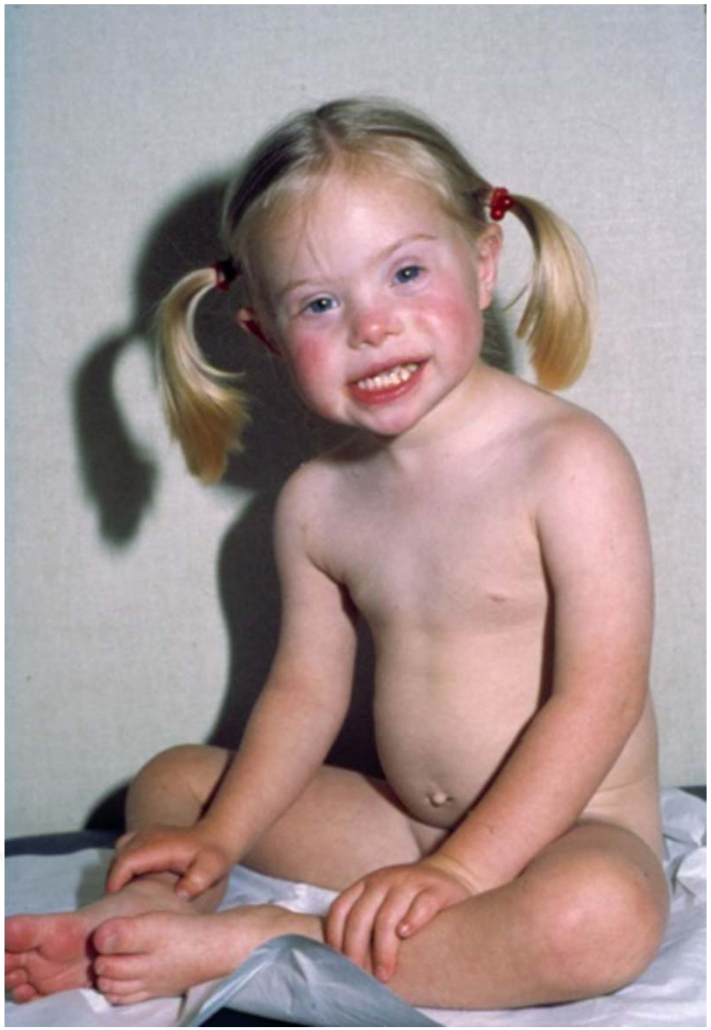


Our genome: Karyotype

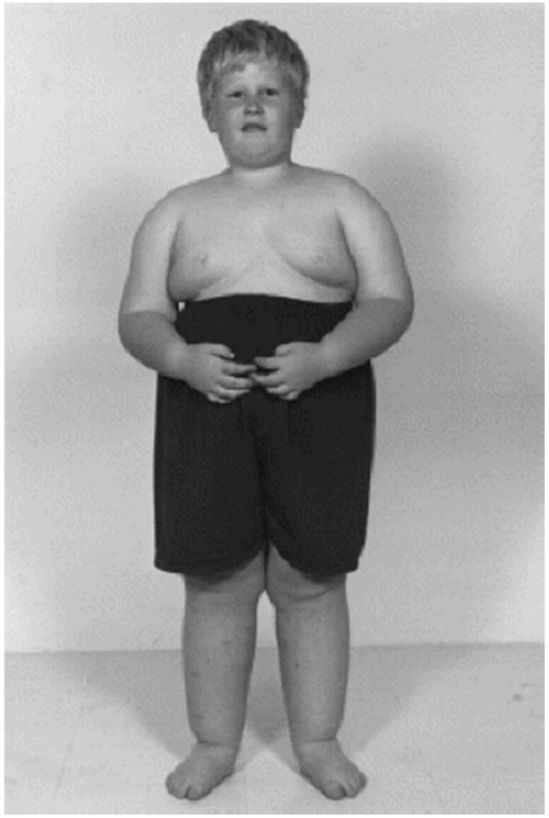


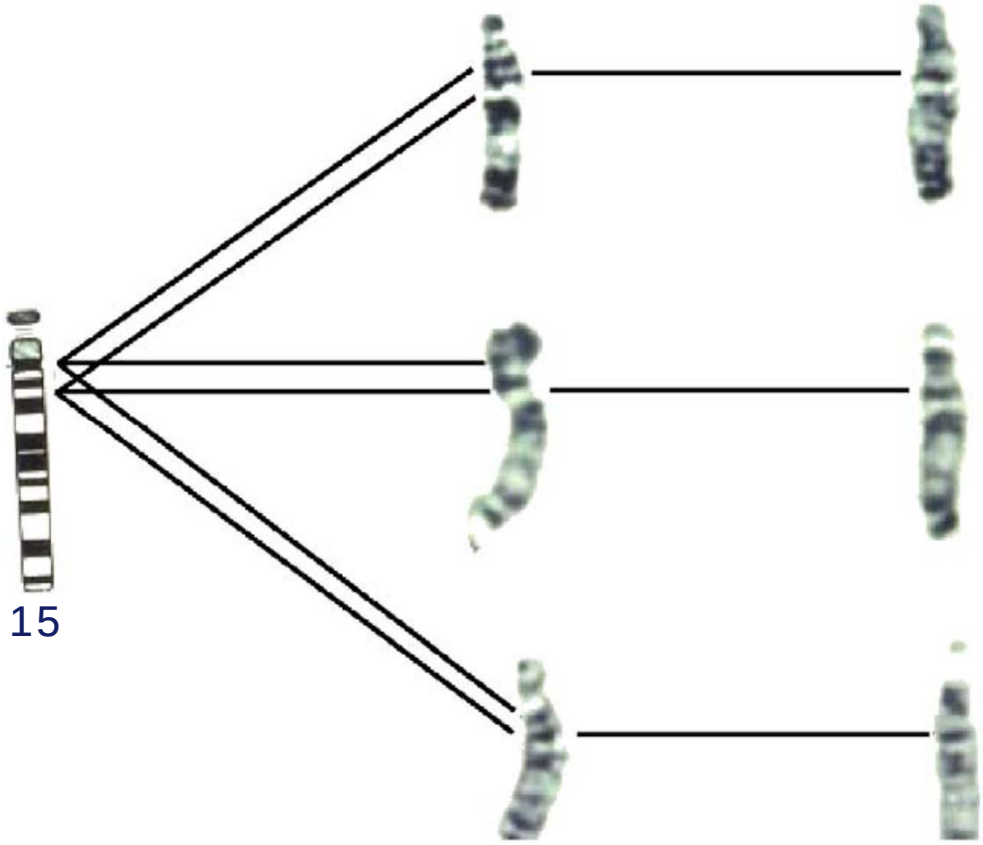
Down syndrome - Trisomy 21





Prader-Willi syndrome

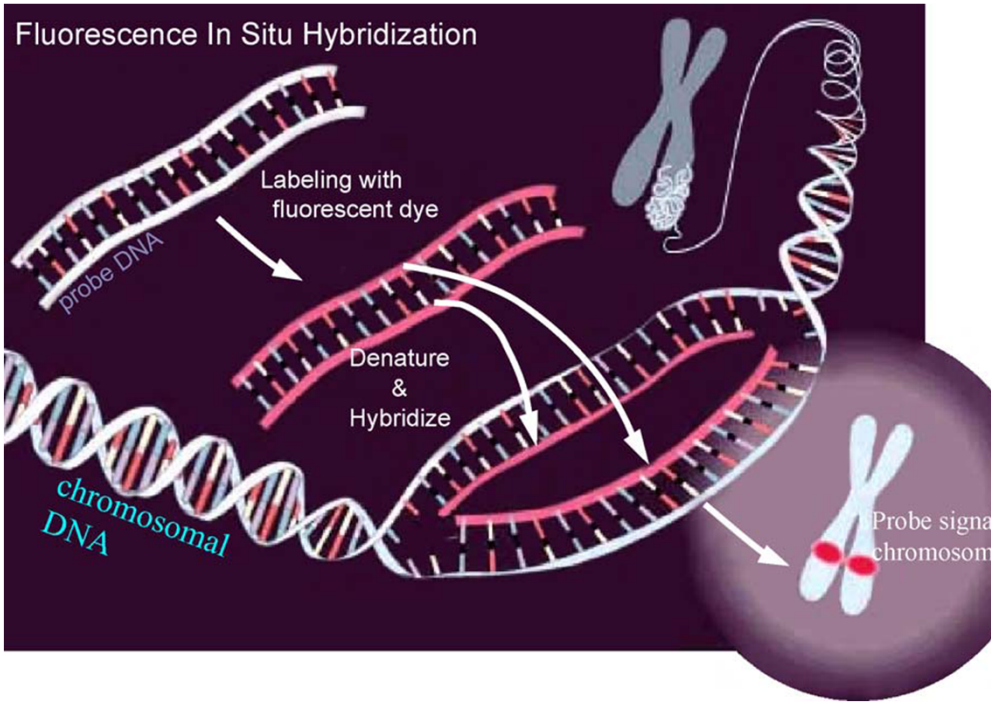


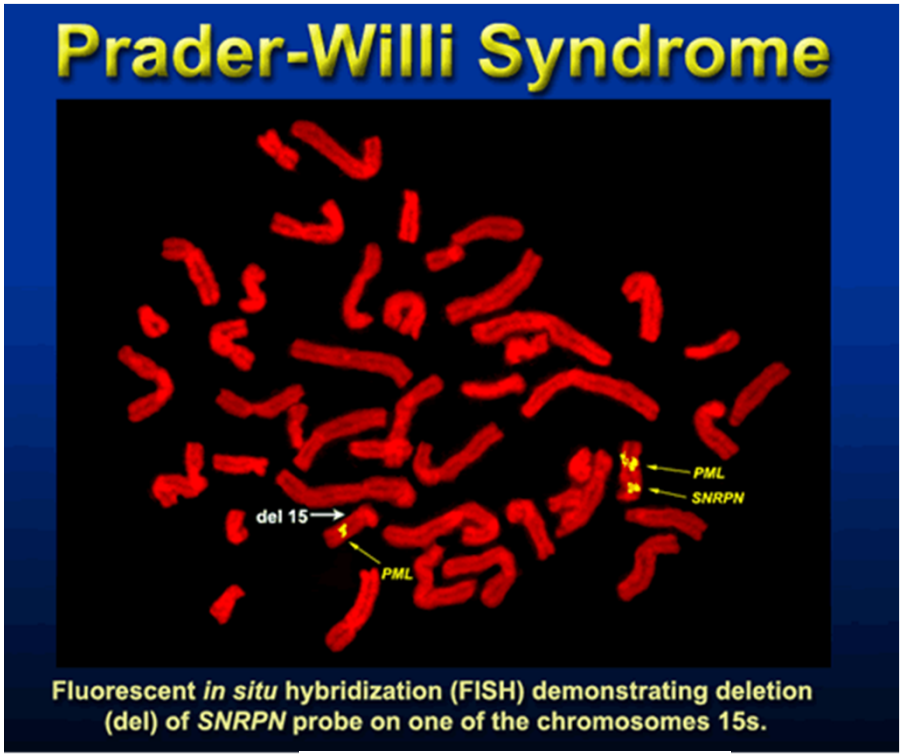


No Deletion Deletion 15q11-13

F3

FISH


Fluorescence In Situ Hybridization
Labeling with
fluorescent dye
probe DNA
Denature
&
Hybridize
chromosomal
DNA
Probe signa
chromosom


Prader-Willi Syndrome
del 15
PML
PML
SNRPN
Fluorescent *in situ* hybridization (FISH) demonstrating deletion
(del) of *SNRPN* probe on one of the chromosomes 15s.

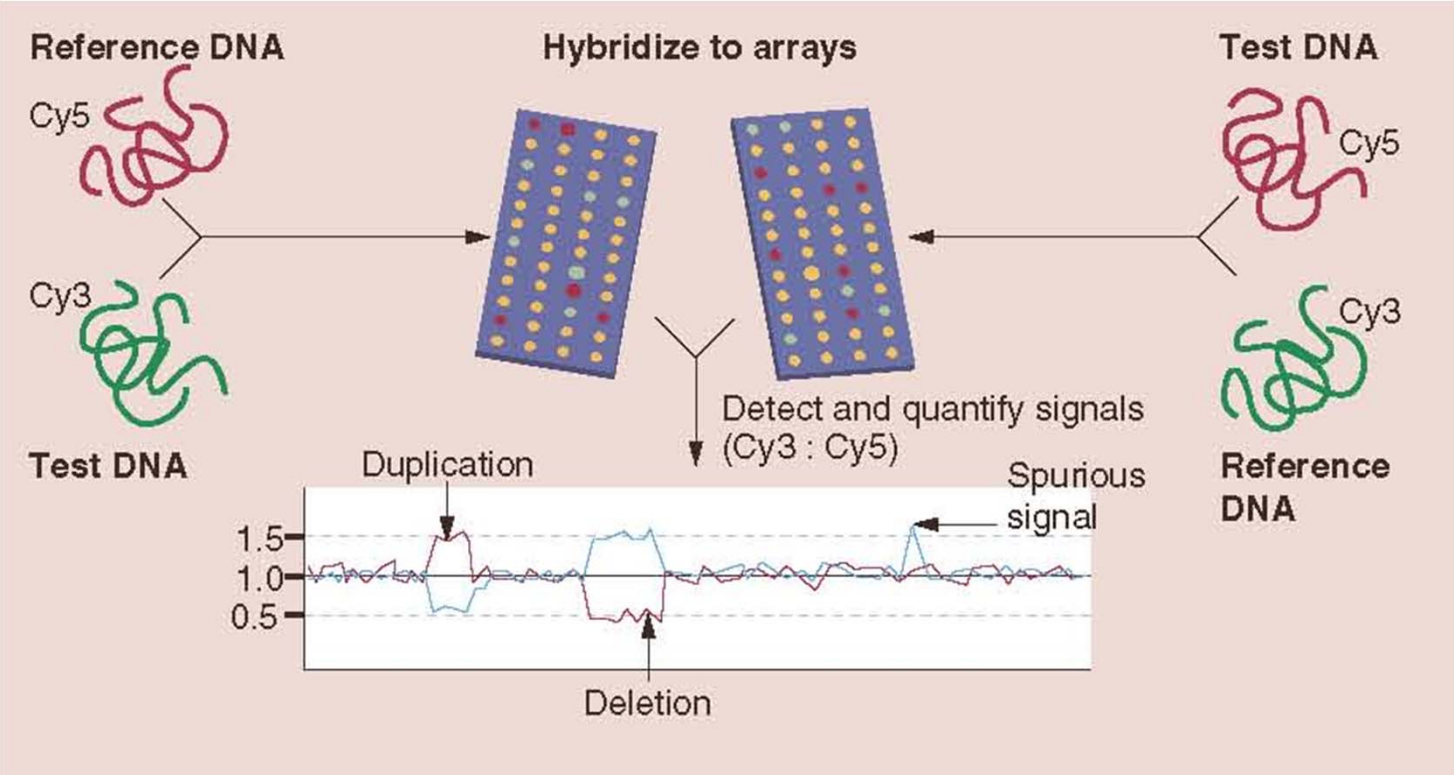
Dia 10

F3 Hier echt recentere plaatjes. Ook prader willi

Frank; 29-10-2014

F4

Array-CGH

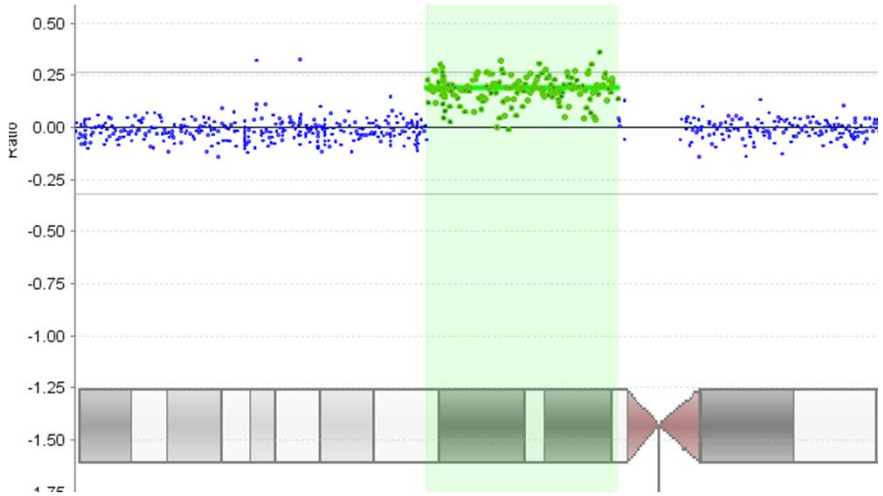

Reference DNA
Cy5
Cy3
Test DNA
Hybridize to arrays
Test DNA
Cy5
Cy3
Reference DNA
Detect and quantify signals
(Cy3 : Cy5)
Duplication
Spurious signal
1.5
1.0
0.5
Deletion

Dia 11

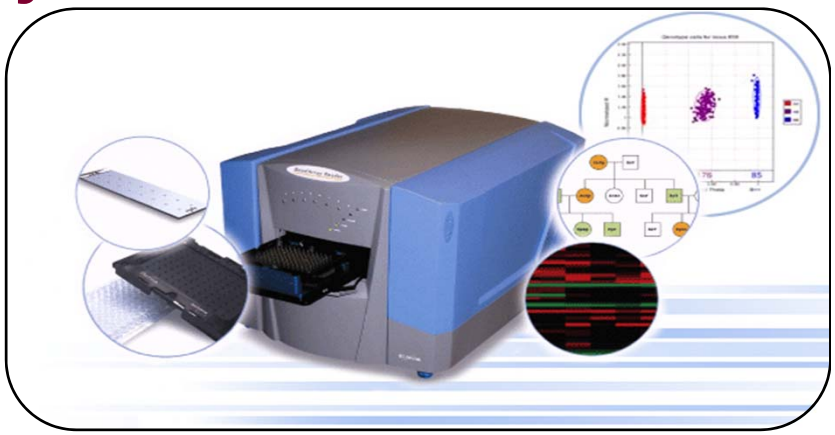
F4 Nieuwe dia

Frank; 29-10-2014

Array-CGH

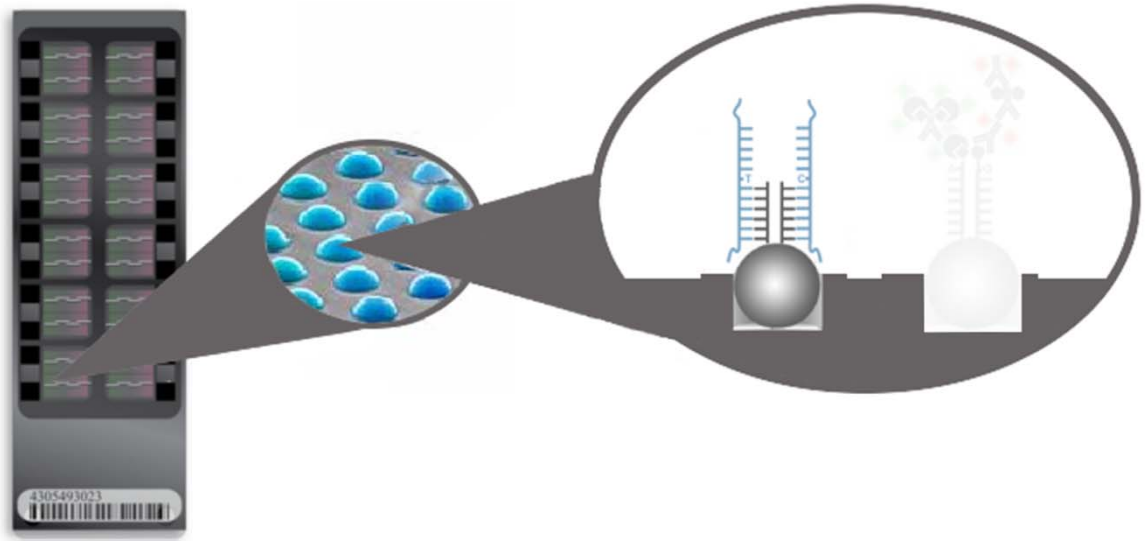


SNP Array

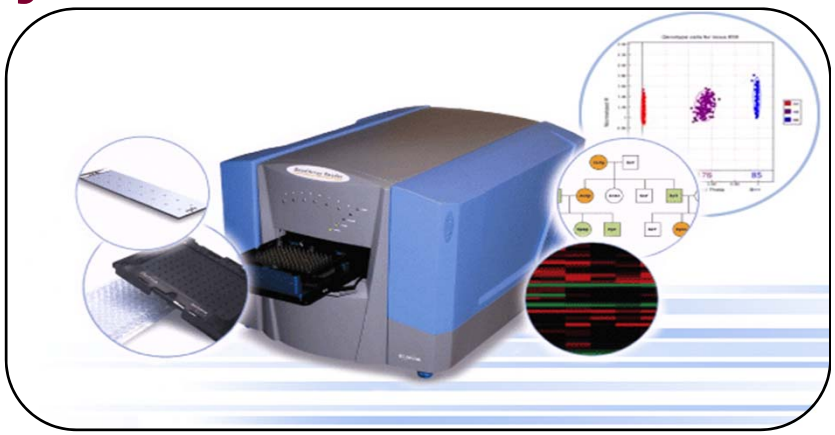


:

- Two information channels: intensity and genotype

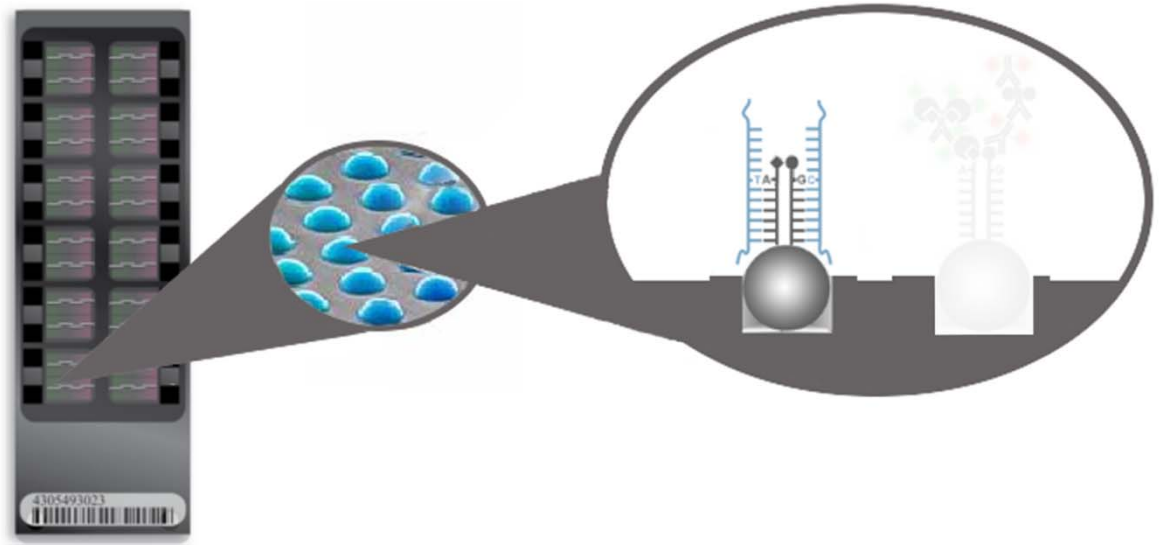


SNP Array

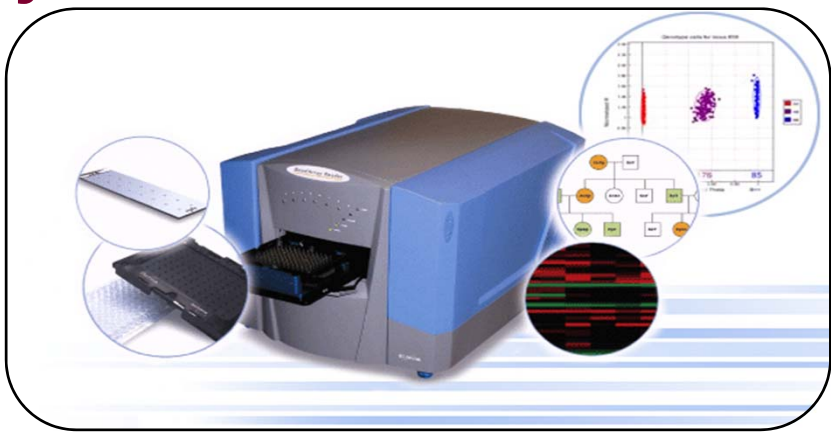


:

- Two information channels: intensity and genotype

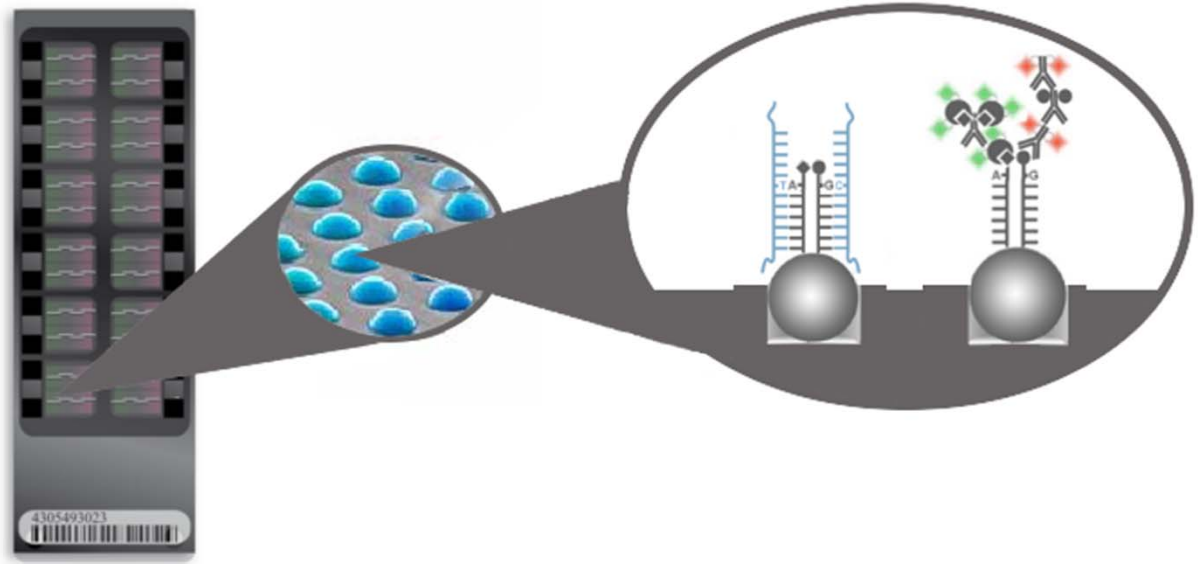


SNP Array

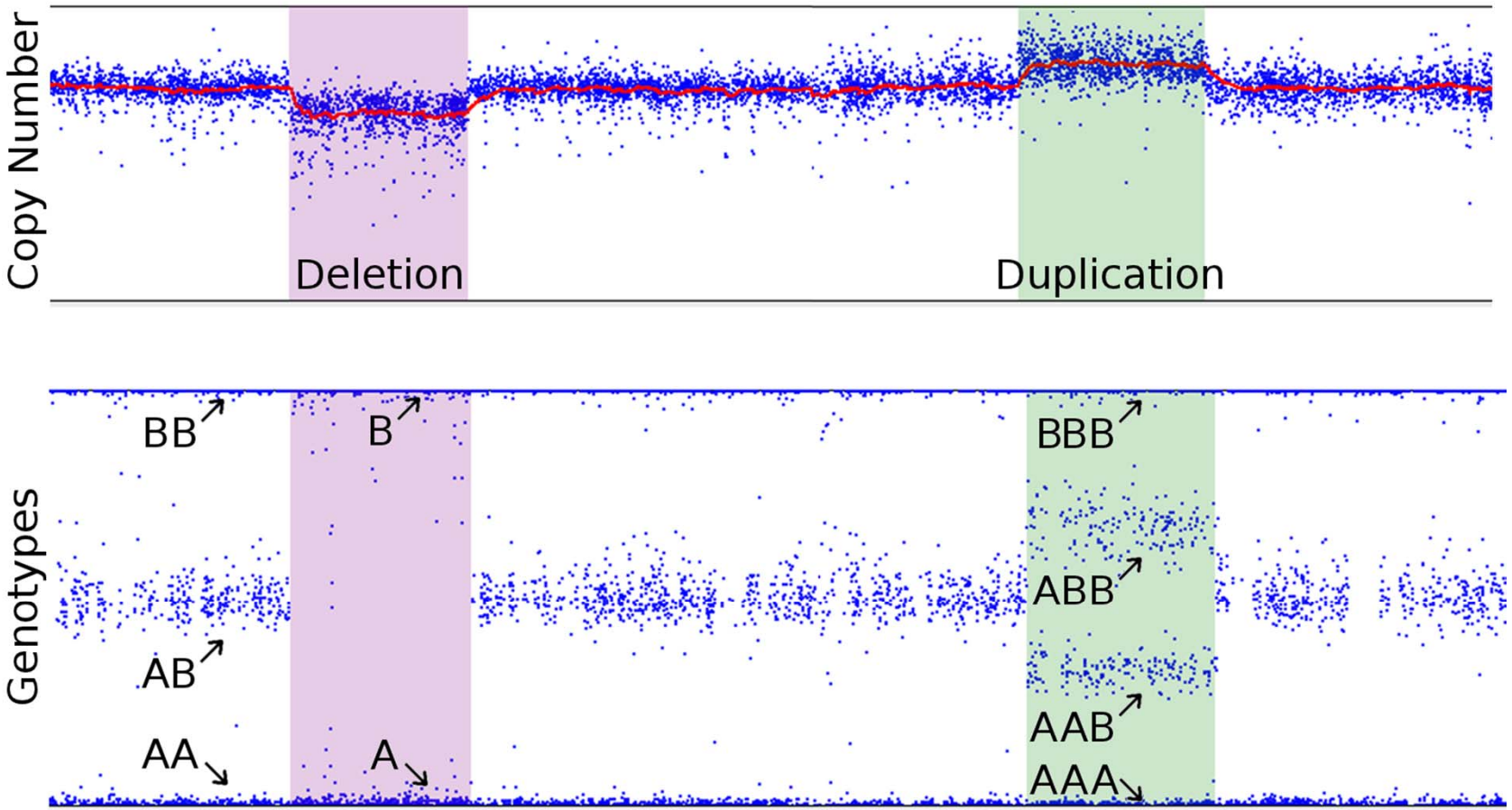


:

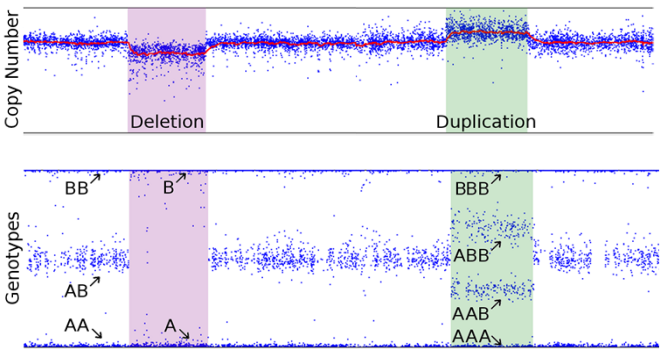
- Two information channels: intensity and genotype

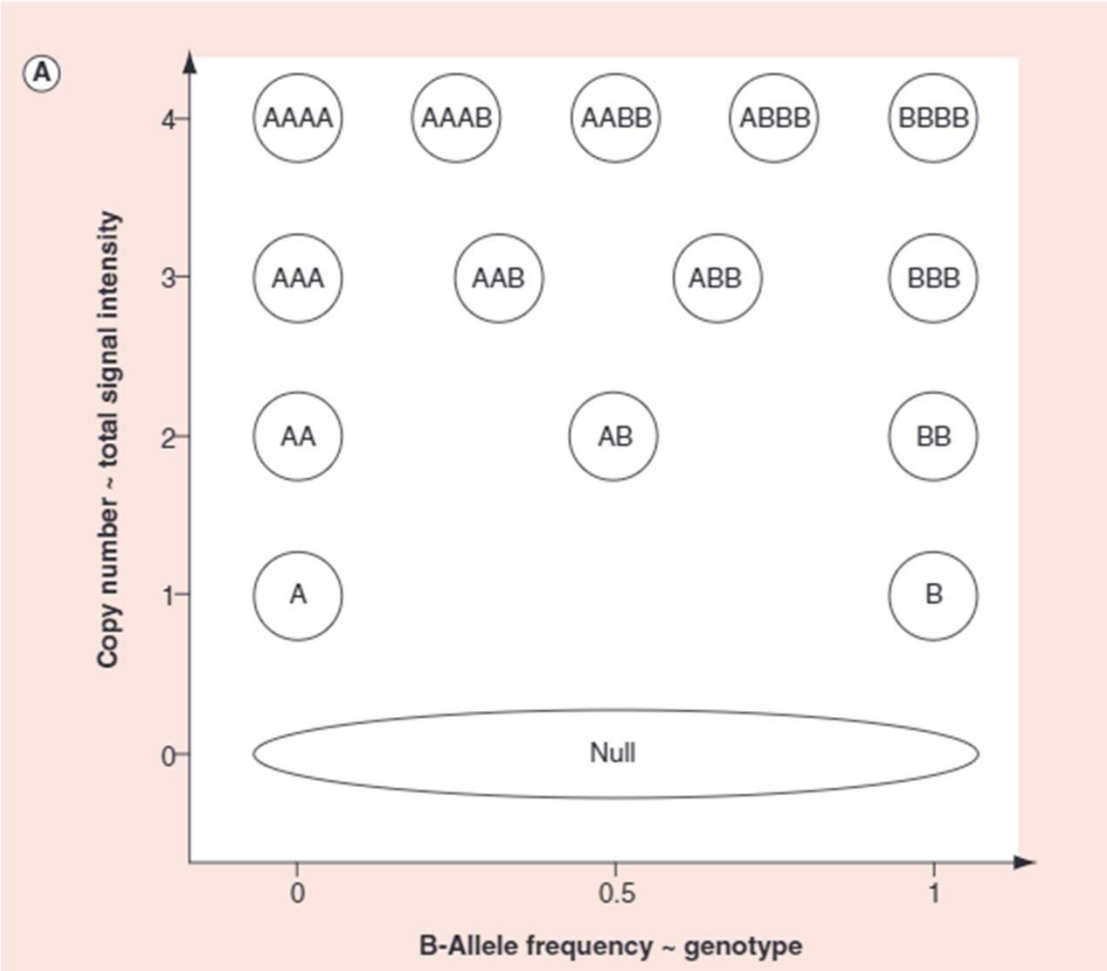


SNP Array


Copy Number
Deletion
Duplication
Genotypes
BB
B
BBB
ABB
AB
AAB
AA
A
AAA

SNP Array


Copy Number
Deletion
Duplication
Genotypes
BB
B
BBB
ABB
AB
AAB
AA
A
AAA


A
AAAA
AAAB
AABB
ABBB
BBBB
AAA
AAB
ABB
BBB
AA
AB
BB
A
B
Null
4
3
2
1
0
0
0.5
1
Copy number ~ total signal intensity
B-Allele frequency ~ genotype

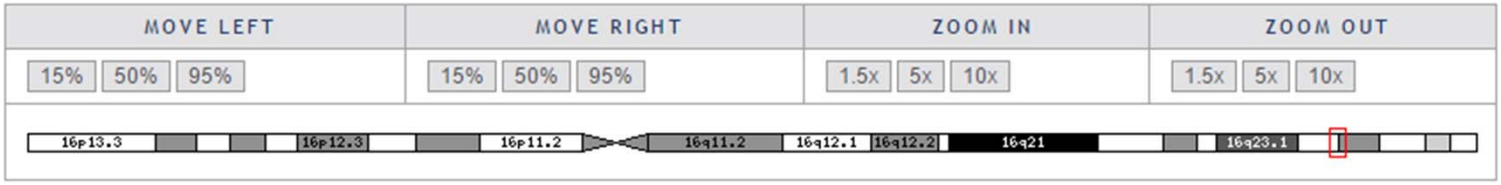
Different algorithms for Deletion / duplication testing

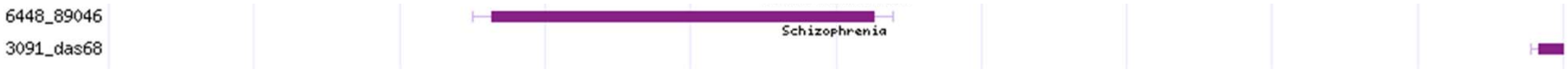
Table 1 Proportion of CNVs called by Pinto et al. detected by the different methods used

	Pinto *et al.*					
	Deletions *(n = 378)*		Duplications *(n = 130)*		Overall *(n = 508)*	
Algorithm	**%**	**P-Val**	**%**	**P-Val**	**%**	**P-Val**
2/3 Vote	81.5		69.6		74.4	
QuantiSNP	79.0	0.19	66.0	0.26	71.3	0.13
PennCNV	80.0	0.3	75.2	0.16	77.2	0.15
VanillaICE	83.0	0.31	55.1	0.0073	66.3	0.0023

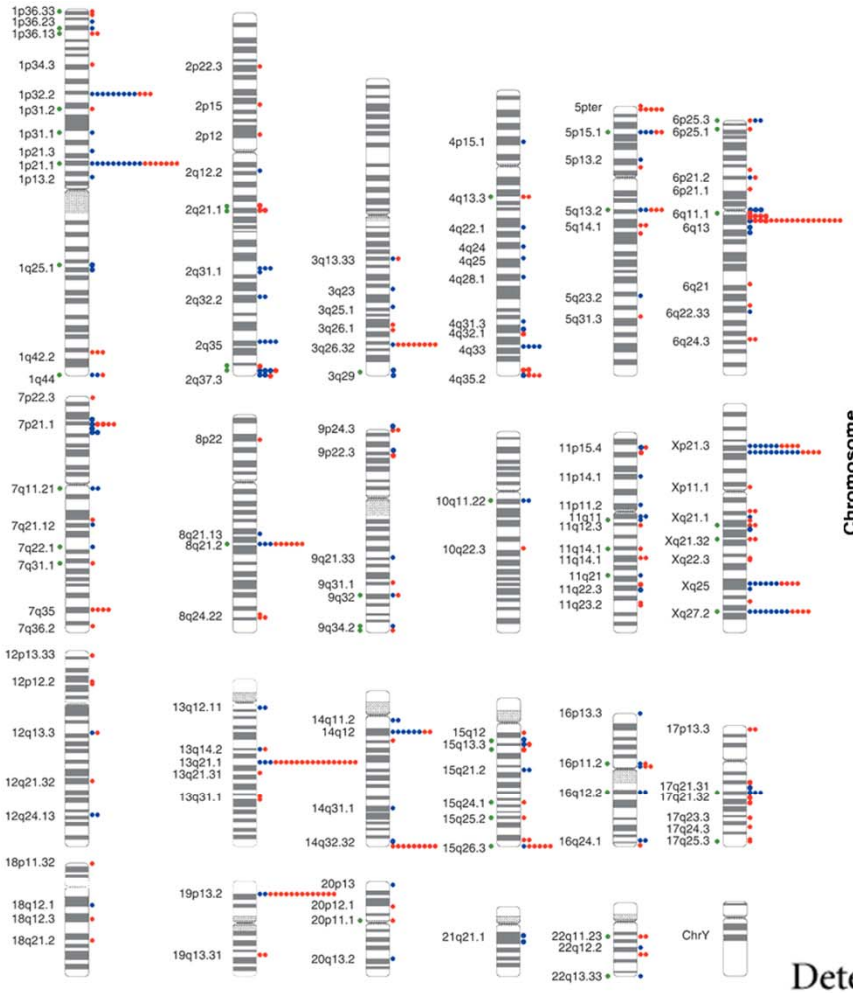
Results are averaged over 181 HapMap samples. P-values are calculated using a one-tailed z-test for difference between proportions, comparing the majority vote against each separate method.

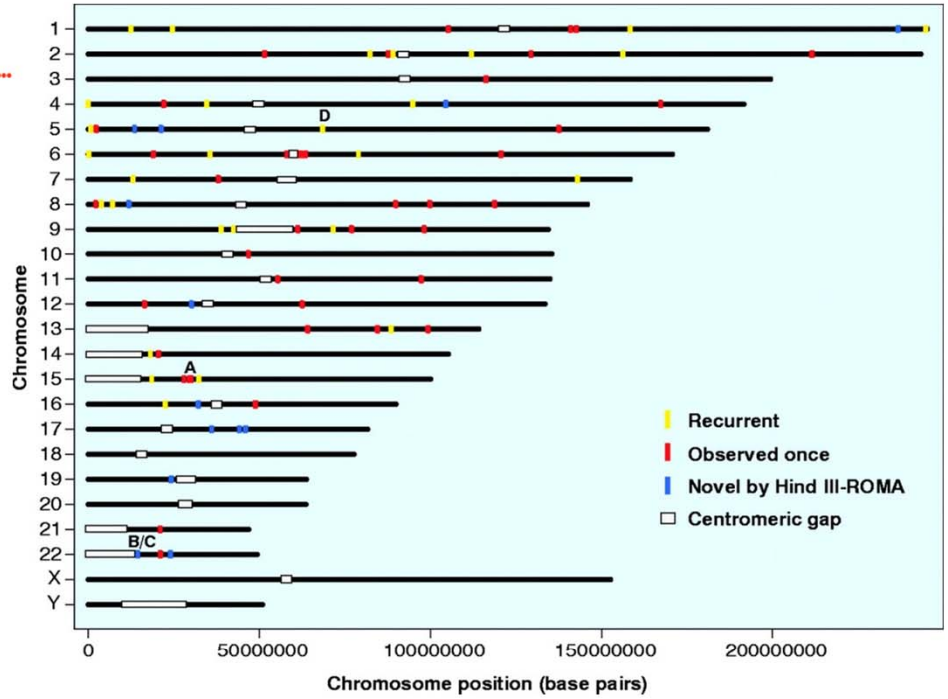
But what if we find a deletion / duplication?





2004: CNV in the control population





Initial maps showed > 10% of the genome to be copy number variant

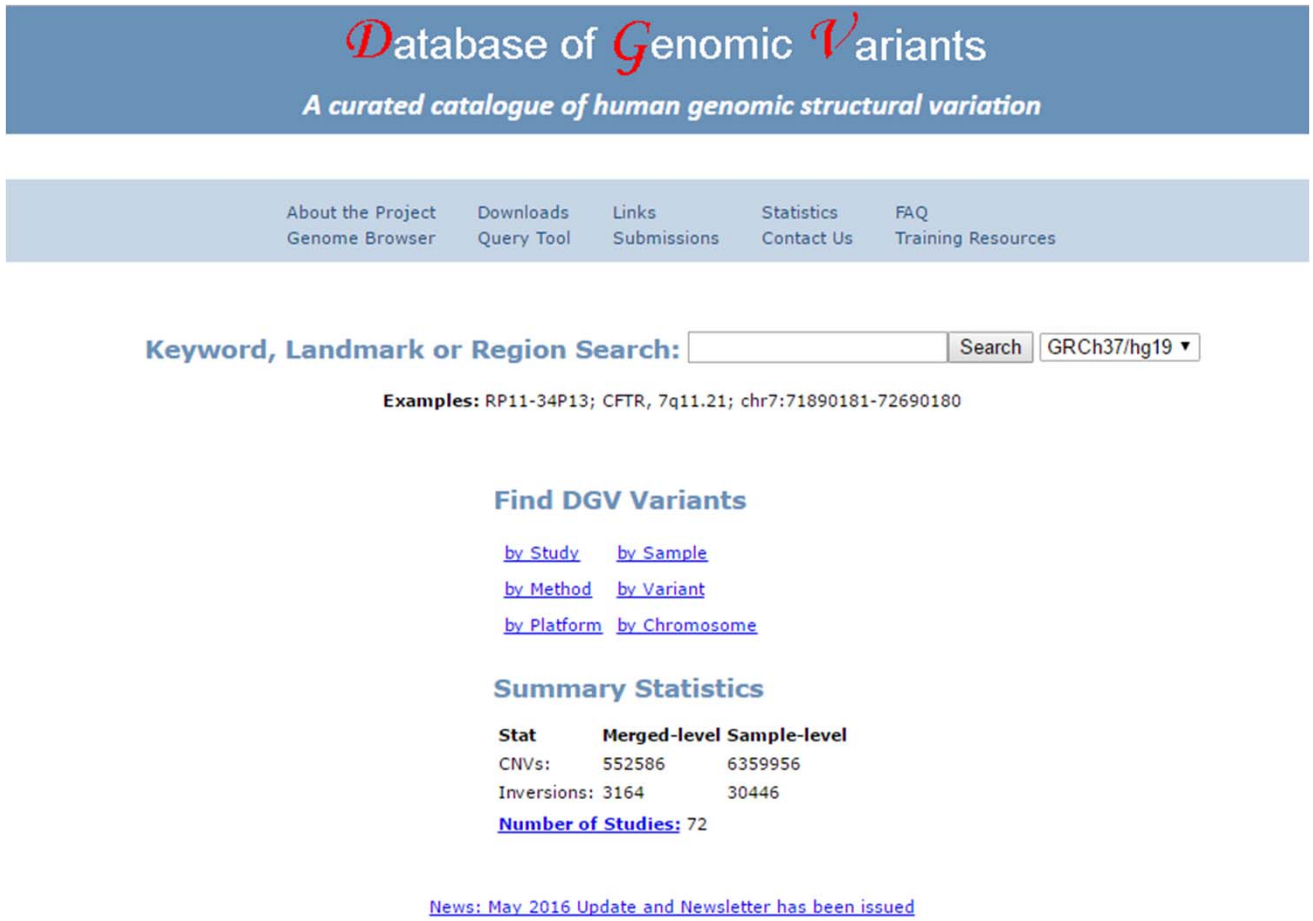
Detection of large-scale variation in the human genome

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]

Large-Scale Copy Number Polymorphism in the Human Genome

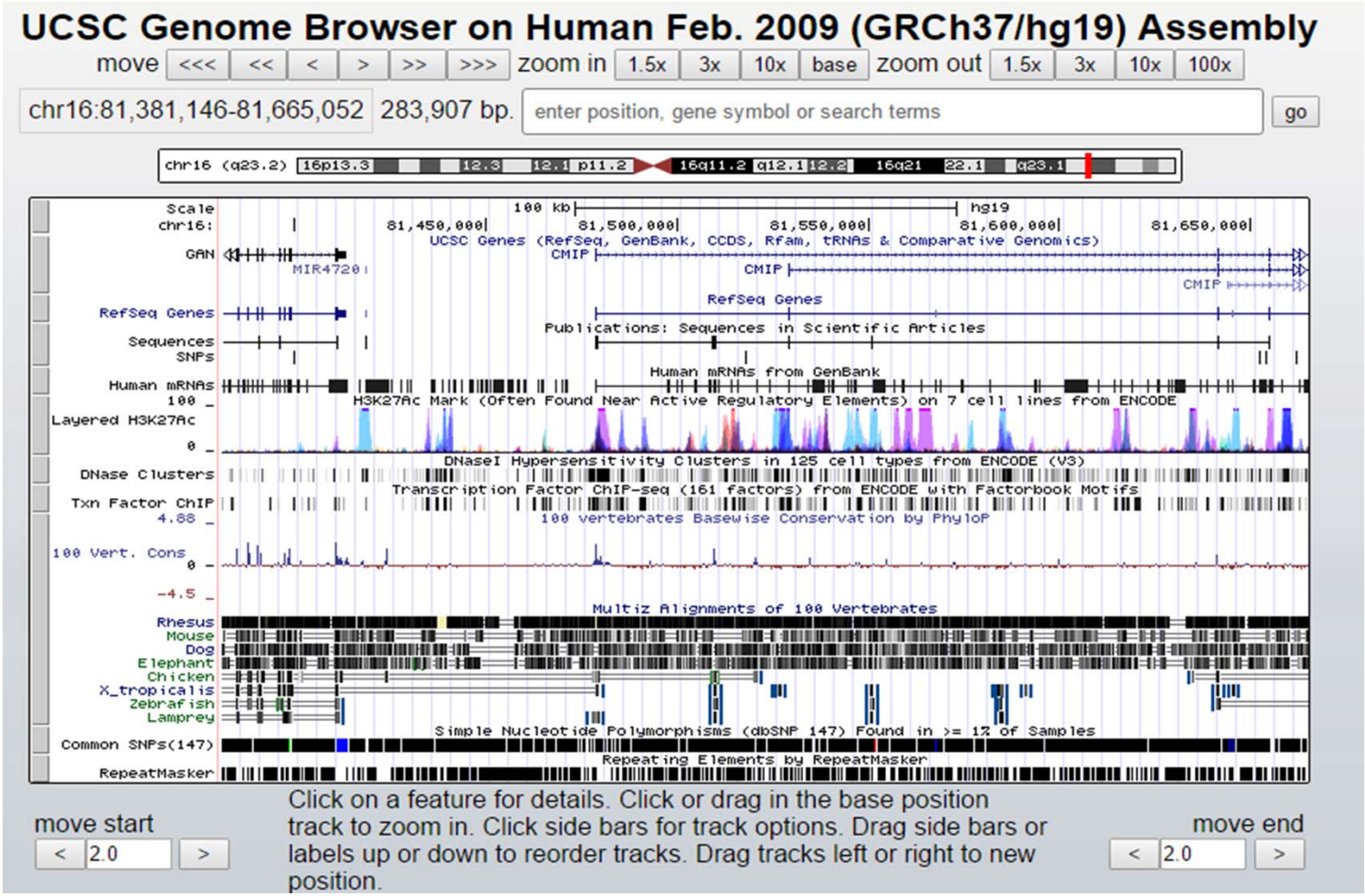
Jonathan Sebat,[1] B. Lakshmi,[1] Jennifer Troge,[1] Joan Alexander,[1] Janet Young,[2] Pär Lundin,[3] Susanne Månér,[3] Hillary Massa,[2] Megan Walker,[2] Maoyen Chi,[1] Nicholas Navin,[1] Robert Lucito,[1] John Healy,[1] James Hicks,[1] Kenny Ye,[4] Andrew Reiner,[1] T. Conrad Gilliam,[5] Barbara Trask,[2]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

CNV are stored in database of genomic variants DGV



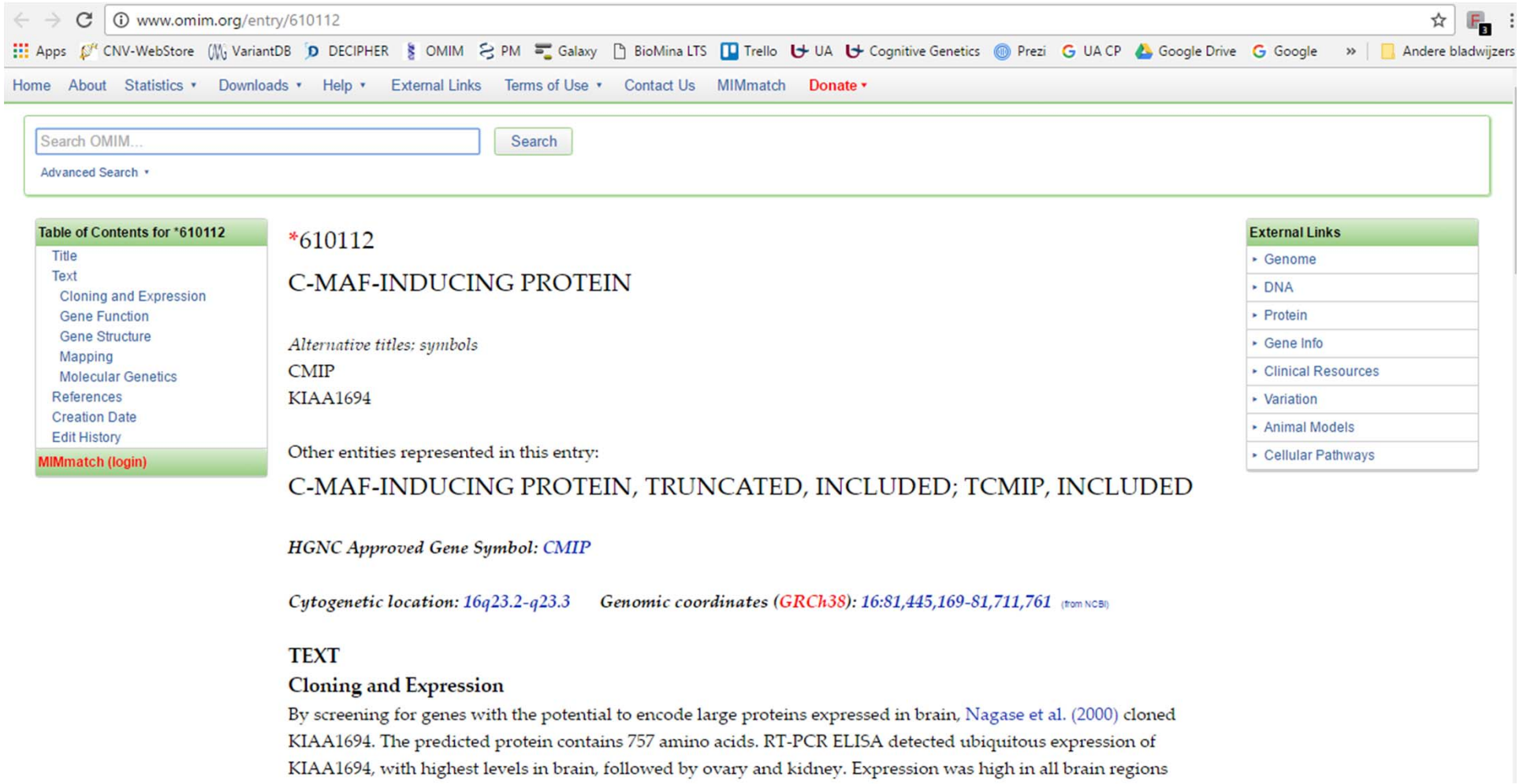
Are there genes in the region?

Genome is annotated in UCSC



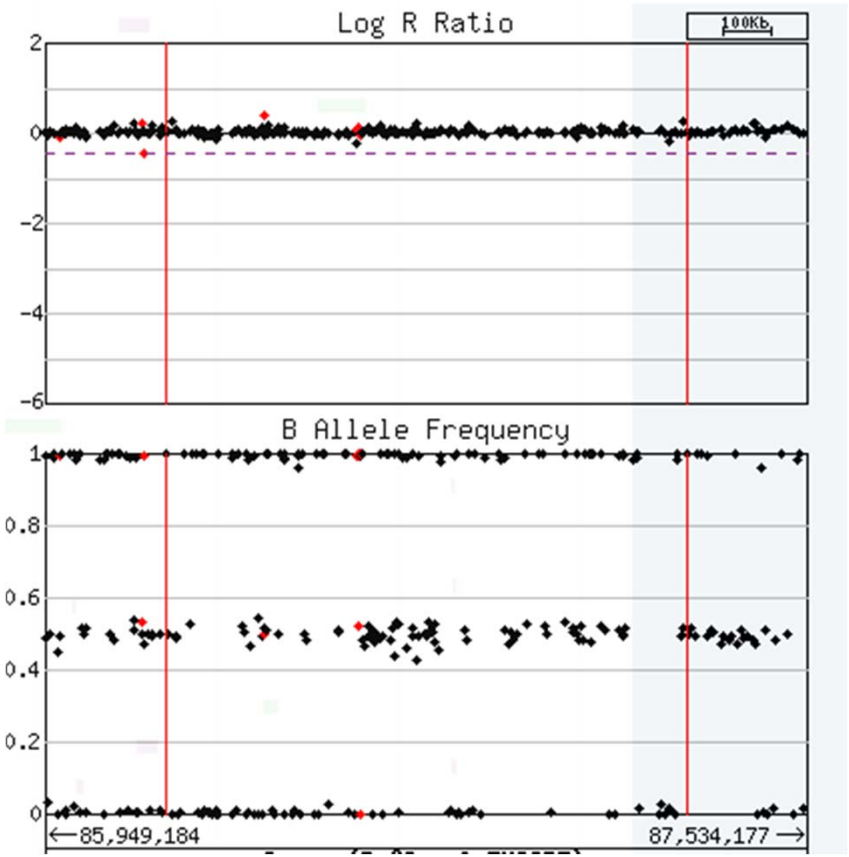
Are there relevant genes in the region?

Disease genes are annotated in the OMIM

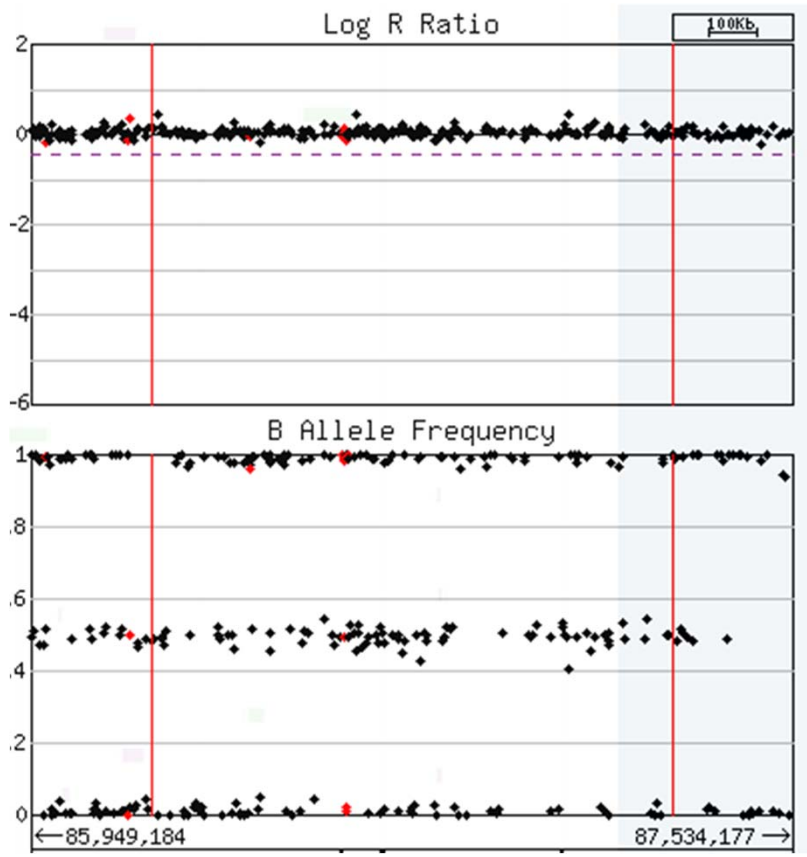


Do the parents have the same deletion?

Father

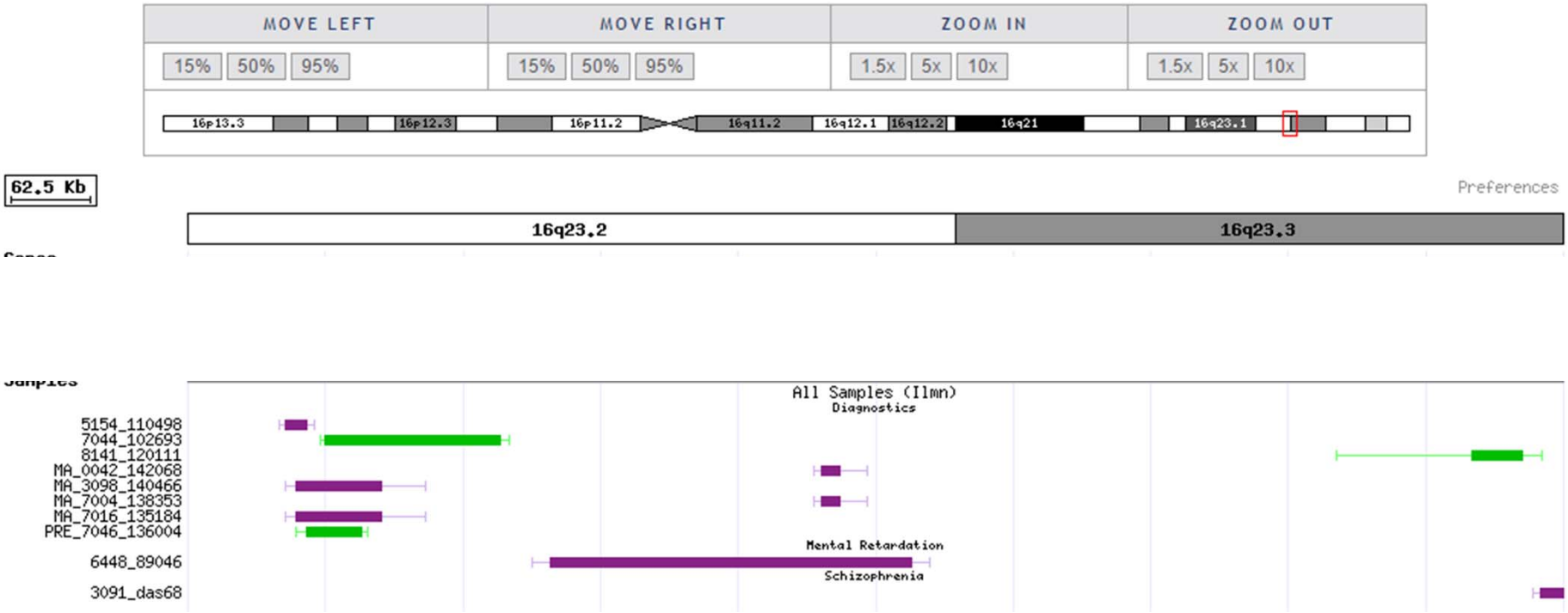


Mother



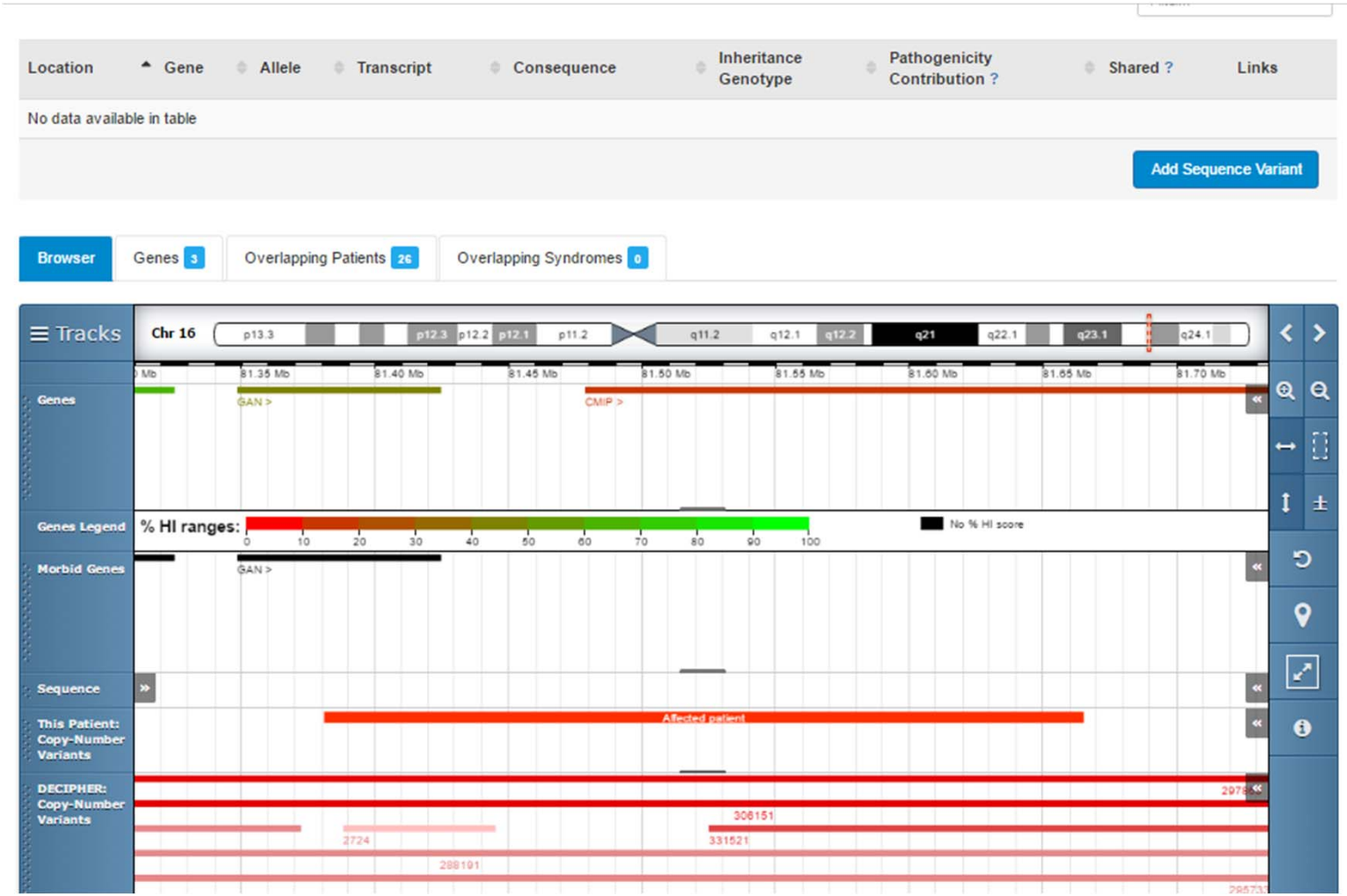
Did we find the same deletion before?

Build and search in house database

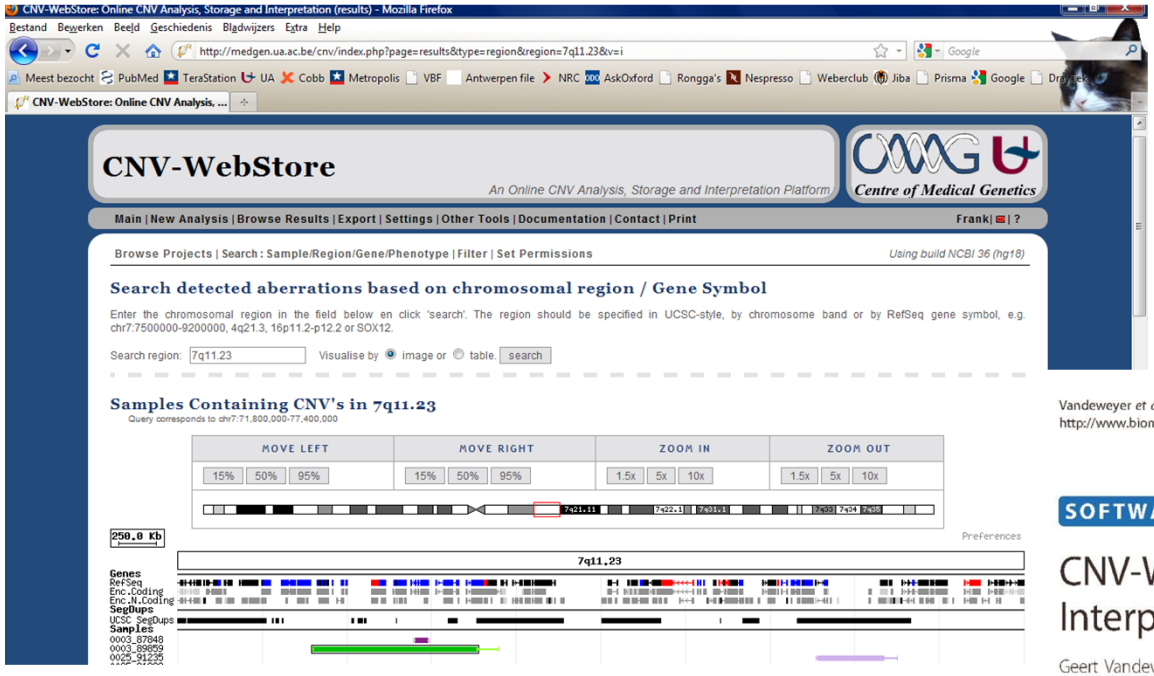


Did others find the same deletion before?

Search DECIPHER



CNV-WebStore a graphical interface



Vandeweyer *et al. BMC Bioinformatics* 2011, **12**:4
http://www.biomedcentral.com/1471-2105/12/4

BMC Bioinformatics

SOFTWARE | Highly accessed | Open Access

CNV-WebStore: Online CNV Analysis, Storage and Interpretation

Geert Vandeweyer[1], Edwin Reyniers[1,2], Wim Wuyts[2], Liesbeth Rooms[1], R Frank Kooy[1*]

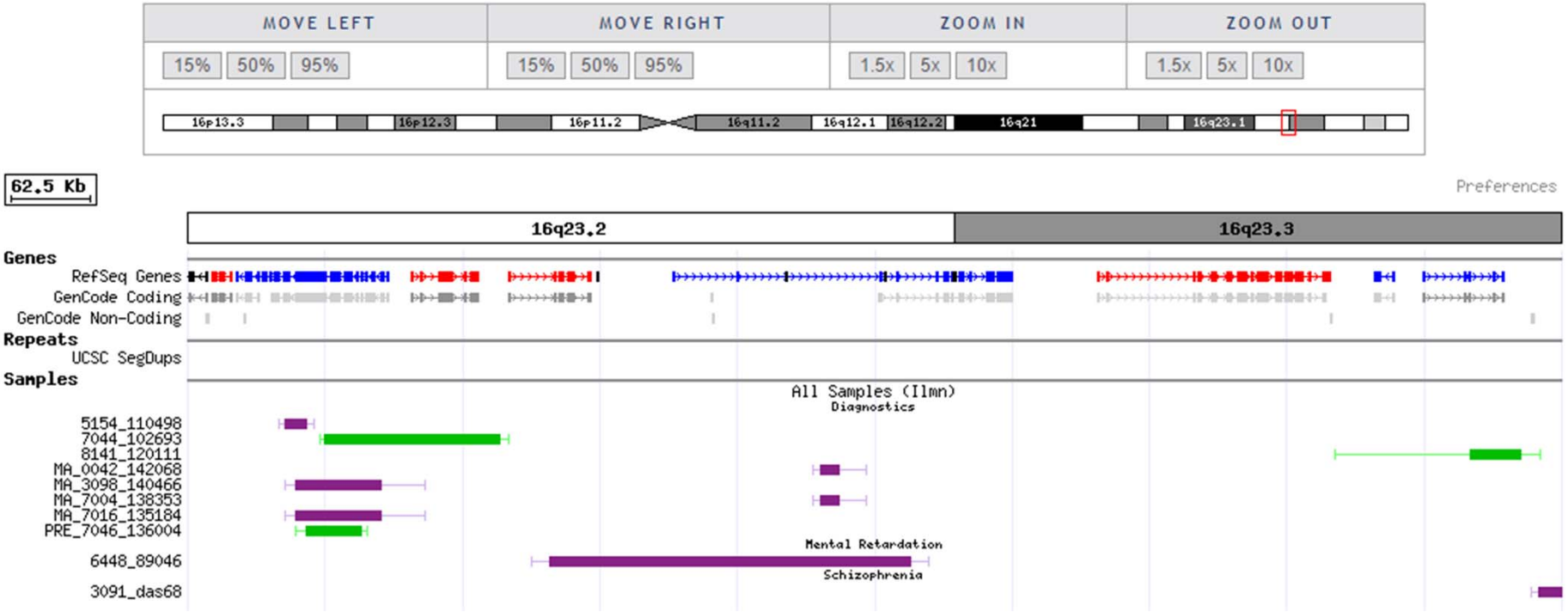
Abstract

Background: Microarray technology allows the analysis of genomic aberrations at an ever increasing resolution, making functional interpretation of these vast amounts of data the main bottleneck in routine implementation of high resolution array platforms, and emphasising the need for a centralised and easy to use CNV data management and interpretation system.

Results: We present CNV-WebStore, an online platform to streamline the processing and downstream interpretation of microarray data in a clinical context, tailored towards but not limited to the Illumina BeadArray platform. Provided analysis tools include CNV analsyis, parent of origin and uniparental disomy detection. Interpretation tools include data visualisation, gene prioritisation, automated PubMed searching, linking data to several genome browsers and annotation of CNVs based on several public databases. Finally a module is provided for uniform reporting of results.

Conclusion: CNV-WebStore is able to present copy number data in an intuitive way to both lab technicians and clinicians, making it a useful tool in daily clinical practice.

Analysis using CNV-WebStore



CNV-WebStore

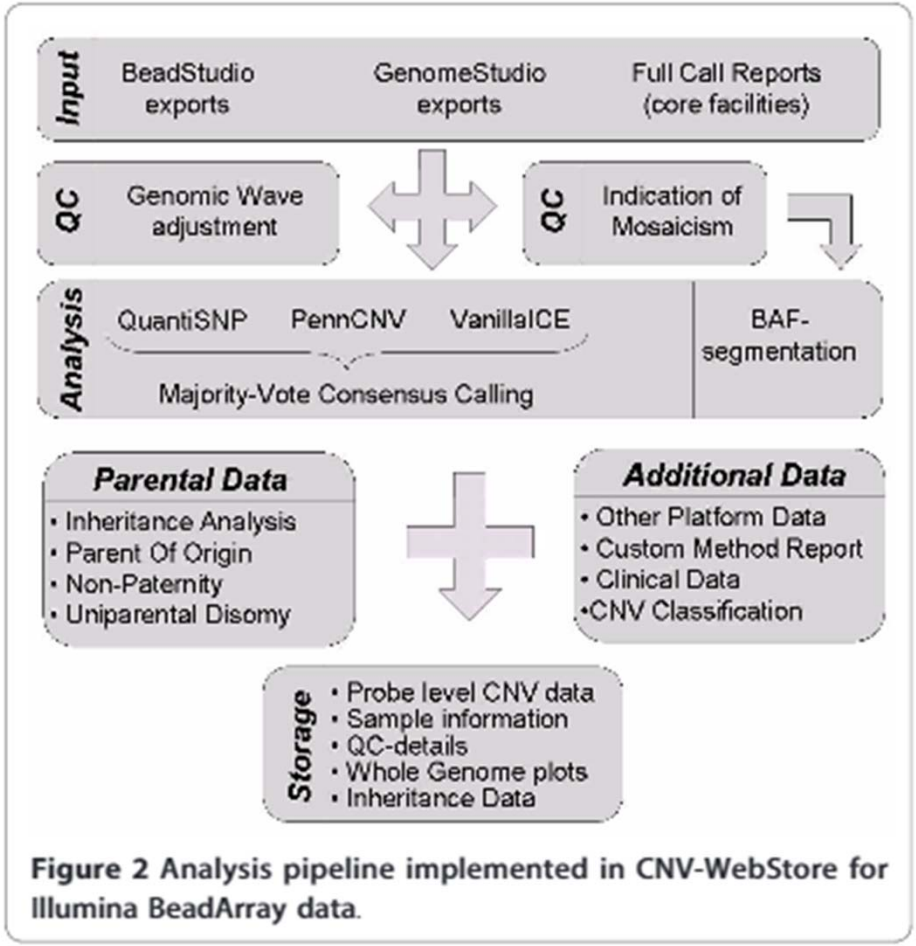
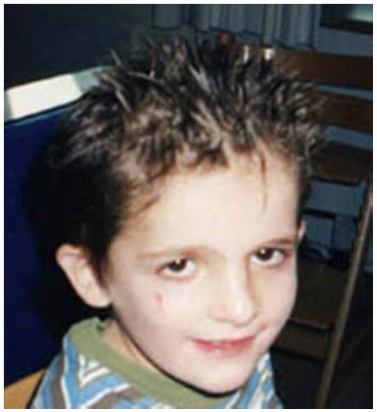
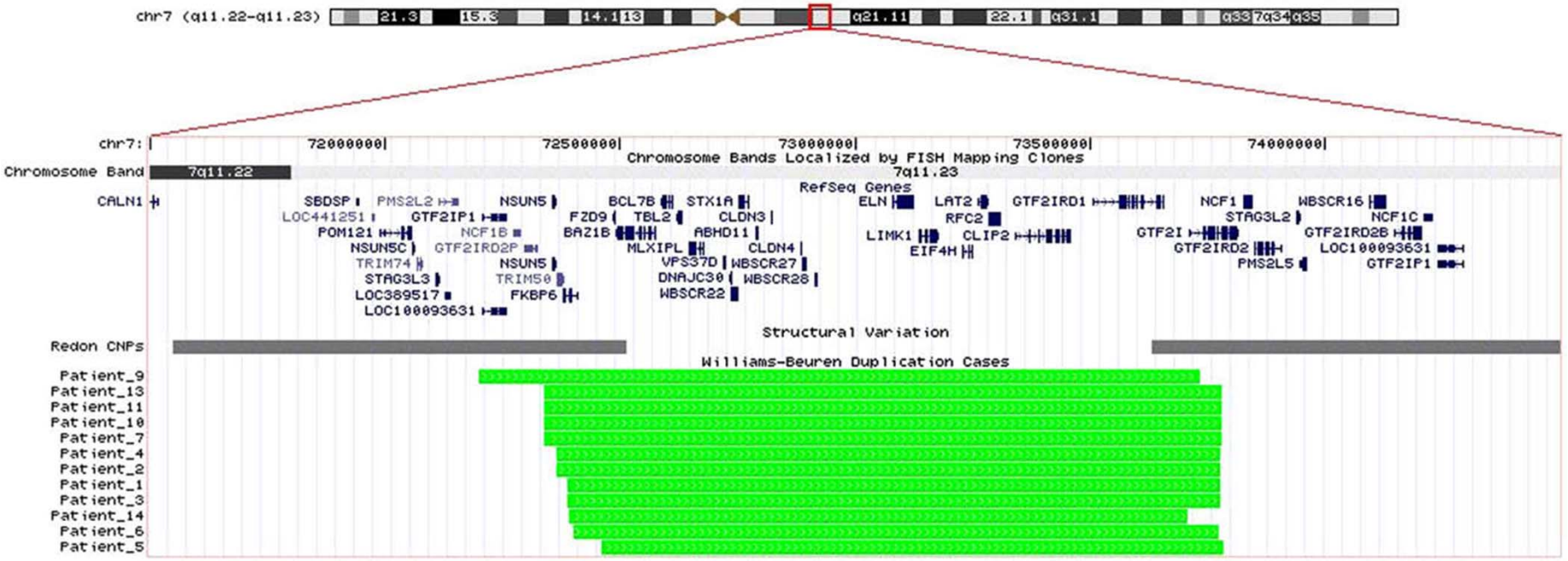


Figure 2 Analysis pipeline implemented in CNV-WebStore for Illumina BeadArray data.

Duplications of the Williams-Beuren region at 7q11.23





Universiteit Antwerpen

European Journal of Medical Genetics 52 (2009) 94–100

Contents lists available at ScienceDirect

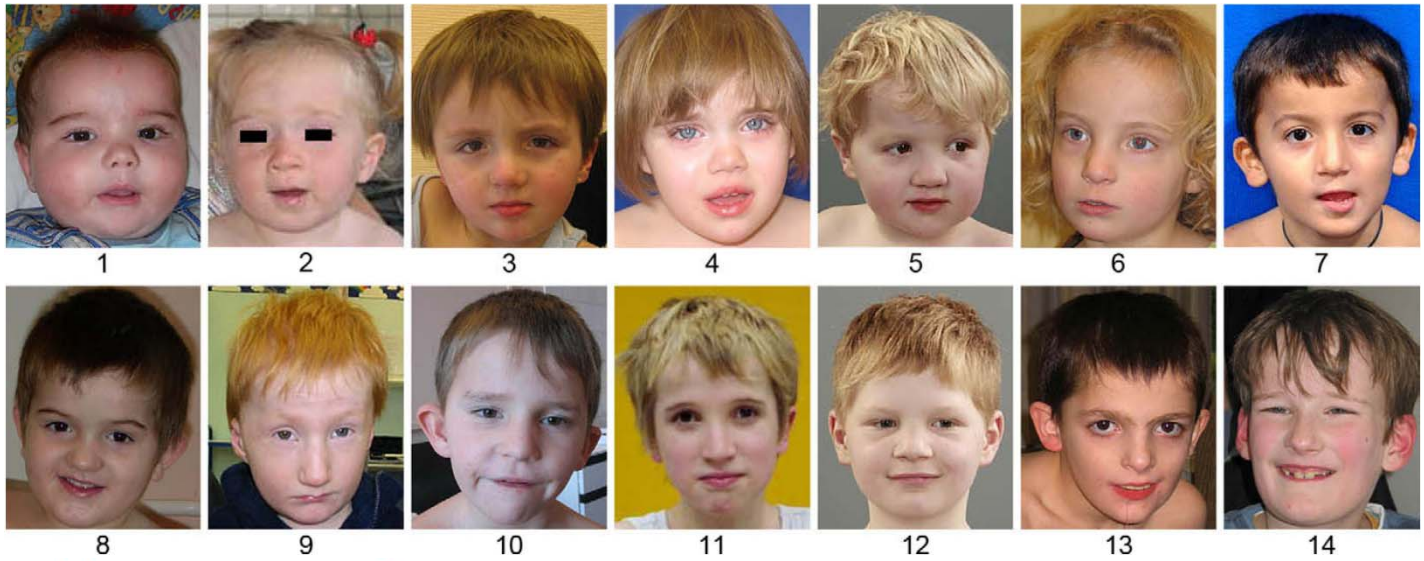
European Journal of Medical Genetics

journal homepage: http://www.elsevier.com/locate/ejmg

Original article

Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome

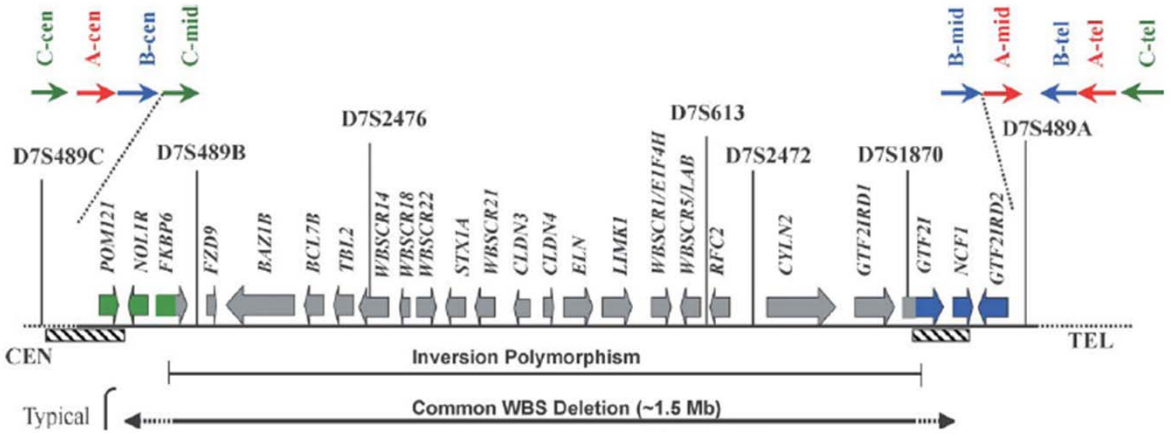
Nathalie Van der Aa [a,*], Liesbeth Rooms [a], Geert Vandeweyer [a], Jenneke van den Ende [a], Edwin Reyniers [a], Marco Fichera [b], Corrado Romano [b], Barbara Delle Chiaie [c], Geert Mortier [c], Björn Menten [c], Anne Destrée [d], Isabelle Maystadt [d], Katrin Männik [e], Ants Kurg [e], Tiia Reimand [f], Dom McMullan [g], Christine Oley [g], Louise Brueton [g], Ernie M.H.F. Bongers [h], Bregje W.M. van Bon [h], Rolph Pfund [h], Sebastien Jacquemont [i], Alessandra Ferrarini [i], Danielle Martinet [i], Connie Schrander-Stumpel [j], Alexander P.A. Stegmann [j], Suzanna G.M. Frints [j], Bert B.A. de Vries [h], Berten Ceulemans [k], R. Frank Kooy [a]



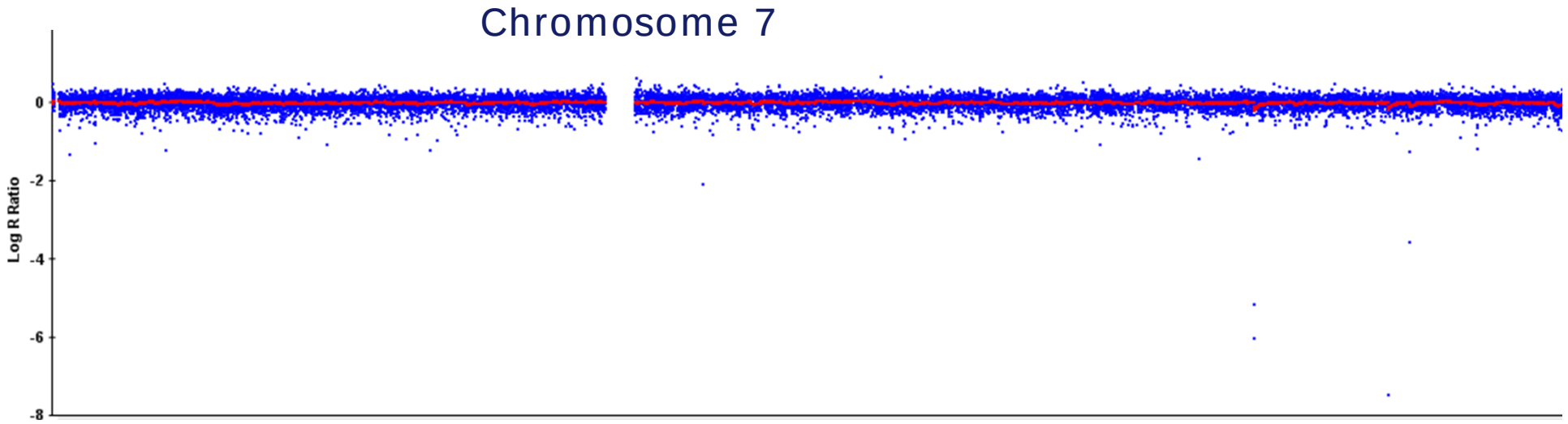
Williams syndrome: a common deletion in the 7q11.23 region

Deletion of 1.55 Mb

contiguous gene syndrome



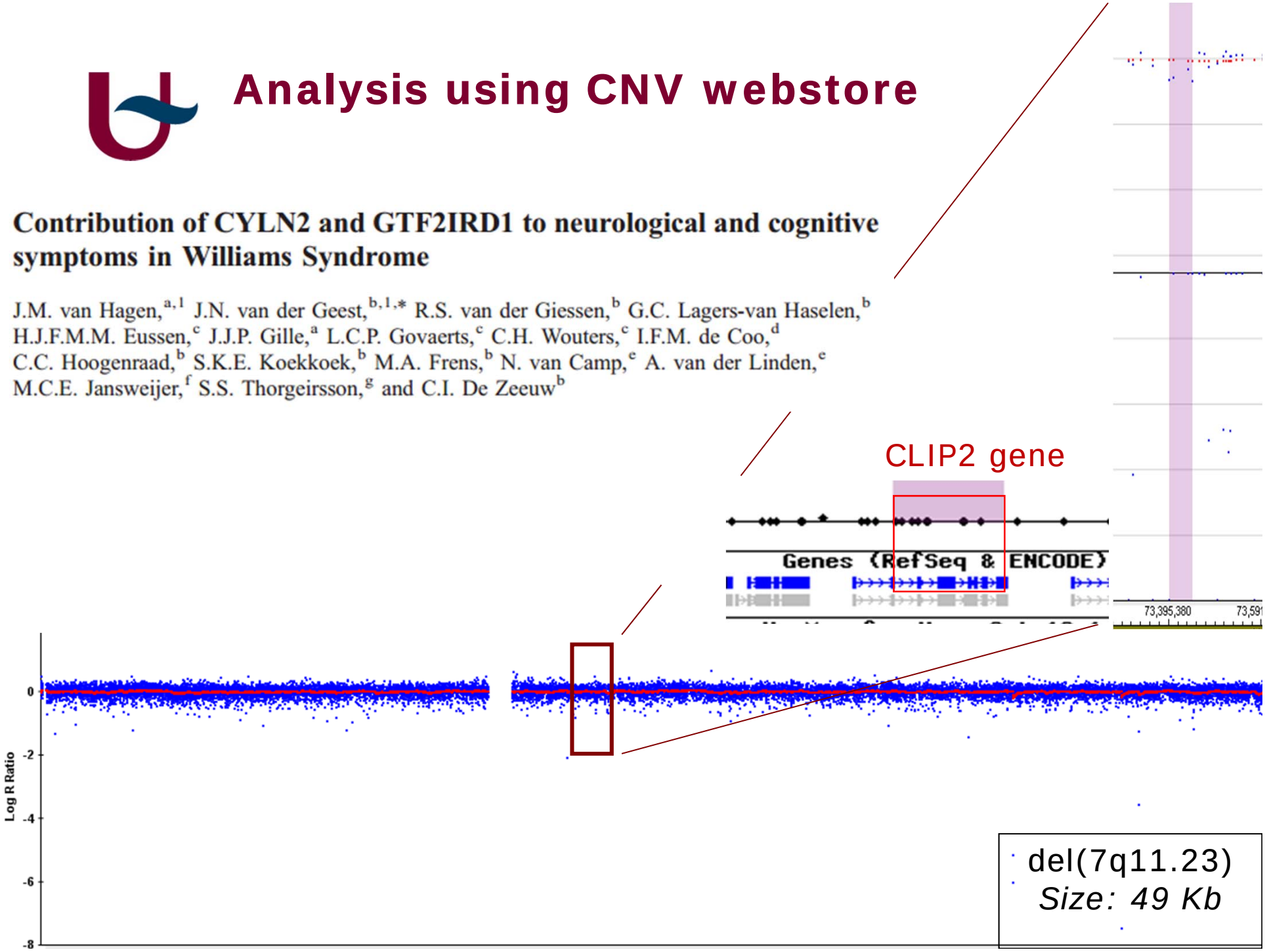
Analysis using CNV webstore



Analysis using CNV webstore

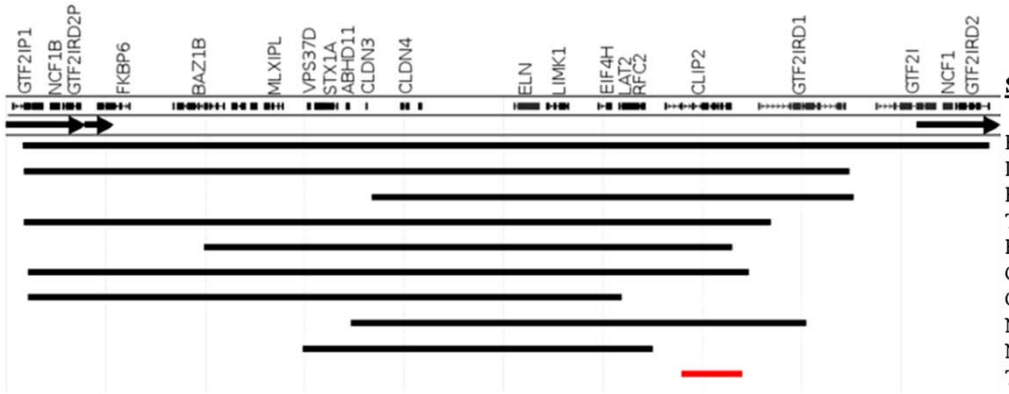
Contribution of CYLN2 and GTF2IRD1 to neurological and cognitive symptoms in Williams Syndrome

J.M. van Hagen,[a,1] J.N. van der Geest,[b,1,*] R.S. van der Giessen,[b] G.C. Lagers-van Haselen,[b] H.J.F.M.M. Eussen,[c] J.J.P. Gille,[a] L.C.P. Govaerts,[c] C.H. Wouters,[c] I.F.M. de Coo,[d] C.C. Hoogenraad,[b] S.K.E. Koekkoek,[b] M.A. Frens,[b] N. van Camp,[e] A. van der Linden,[e] M.C.E. Jansweijer,[f] S.S. Thorgeirsson,[g] and C.I. De Zeeuw[b]



del(7q11.23)
Size: 49 Kb

Williams-Beuren syndrome: a common deletion in the 7q11.23 region



Study	Affected Genes	VIQ/PIQ/FSIQ
Howlin *et al.* 2010	GTF2IP1-GTF2IRD2	63/55/55
Dai *et al.* 2009	GTF2IP1-GTF2IRD1	88/71/78
K3321, Morris *et al.* 2003	CLDN4-GTF2IRD1	./.90
Tassabehji *et al.* 2005	GTF2IP1-GTF2IRD1	./..
Ferrero *et al.* 2010	BAZ1B-CLIP2	117/91/105
Case 1, Hirota *et al.* 2003	FKBP6-CLIP2	64/72/64
Case 2, Hirota *et al.* 2003	FKBP6-LIMK1	65/56/55
NW1, Antonell *et al.* 2010	CLDN3-GTF2IRD1	77/76/75
NW2, Antonell *et al.* 2010	VPS37D-RFC2*	79/90/82
This Study	CLIP2	106/107/106

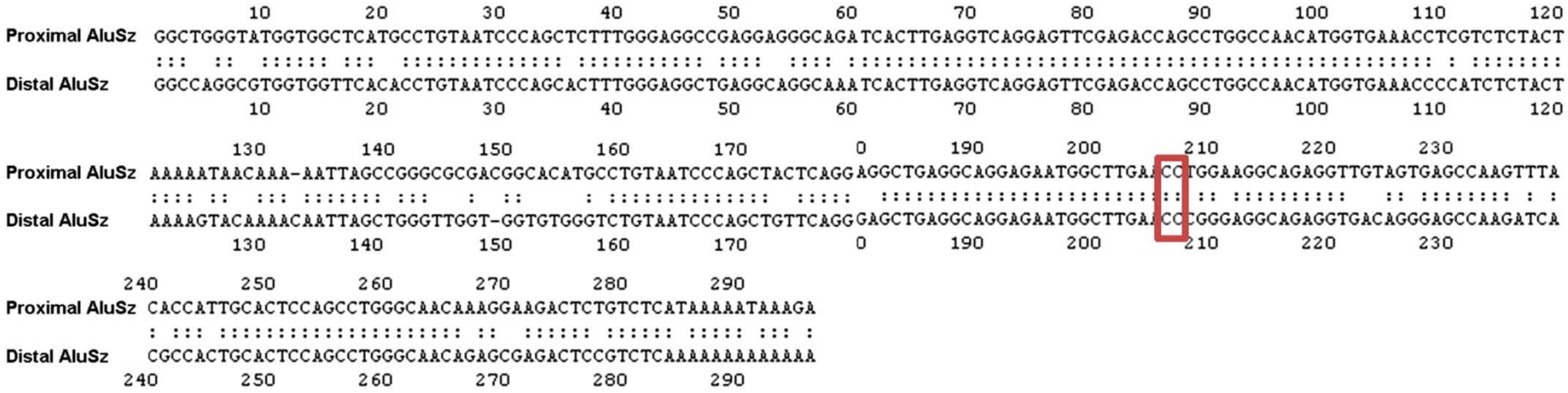
- ELN gene: SVAS
- BAZ1B : hypercalcemia
- LIMK1- CYLN2- GTF2
  - → Linked to aspects of cognitive delay
- GTF2IRD1
  - → Linked to aspects of craniofacial pathology

The American Journal of Human Genetics *90*, 1071–1078, June 8, 2012

REPORT

The Contribution of *CLIP2* Haploinsufficiency to the Clinical Manifestations of the Williams-Beuren Syndrome

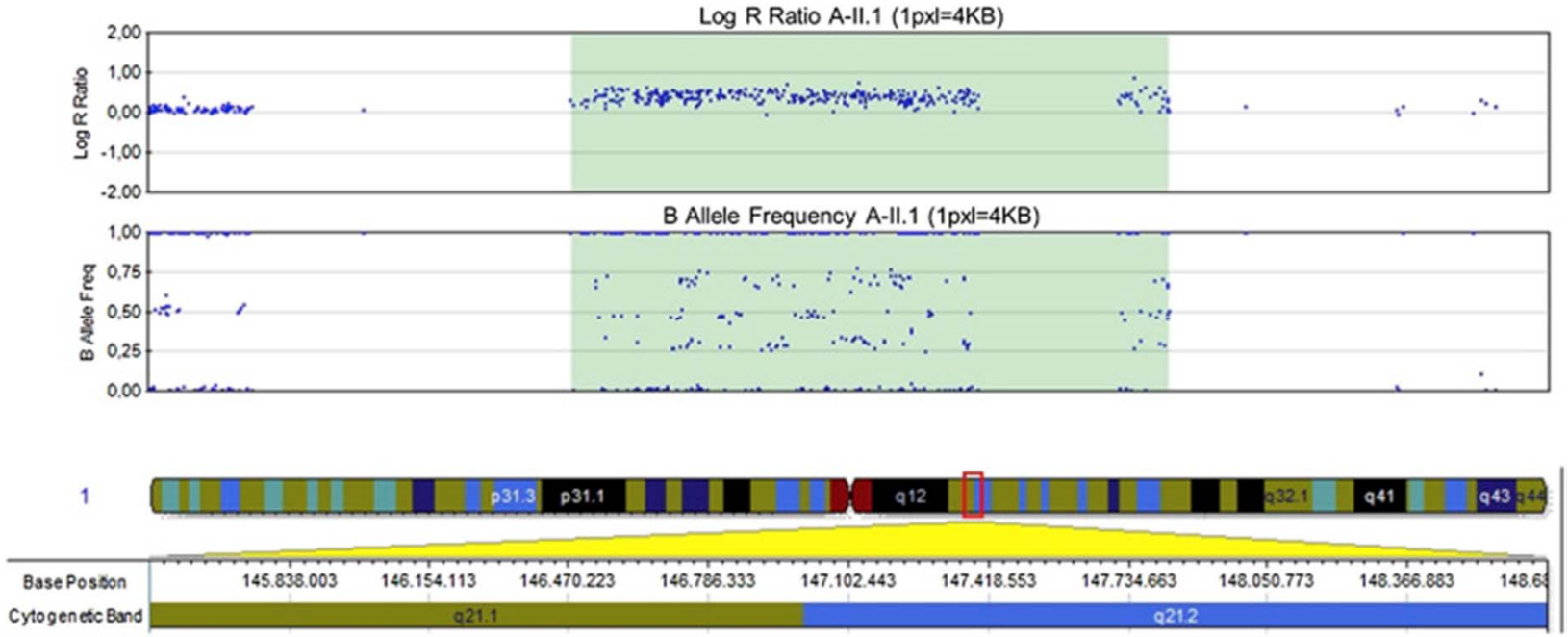
Geert Vandeweyer,[1] Nathalie Van der Aa,[1] Edwin Reyniers,[1] and R. Frank Kooy[1,*]



Five patients with a chromosome 1q21.1 triplication show macrocephaly, increased weight and facial similarities

Anke Van Dijck [a, b, *], Ilse M. van der Werf [a], Edwin Reyniers [a, b], Stefaan Scheers [a, b], Meron Azage [c], Kiana Siefkas [d], Nathalie Van der Aa [a], Amy Lacroix [e], Jill Rosenfeld [f], Bob Argiropoulos [g, h], Kellie Davis [g], A.Micheil Innes [g, h], Heather C. Mefford [d, e], Geert Mortier [a, b], Marije Meuwissen [b], R.Frank Kooy [a]

European Journal of Medical Genetics 58 (2015) 503–508



CNV is found in 10% of all patients with Intellectual disability

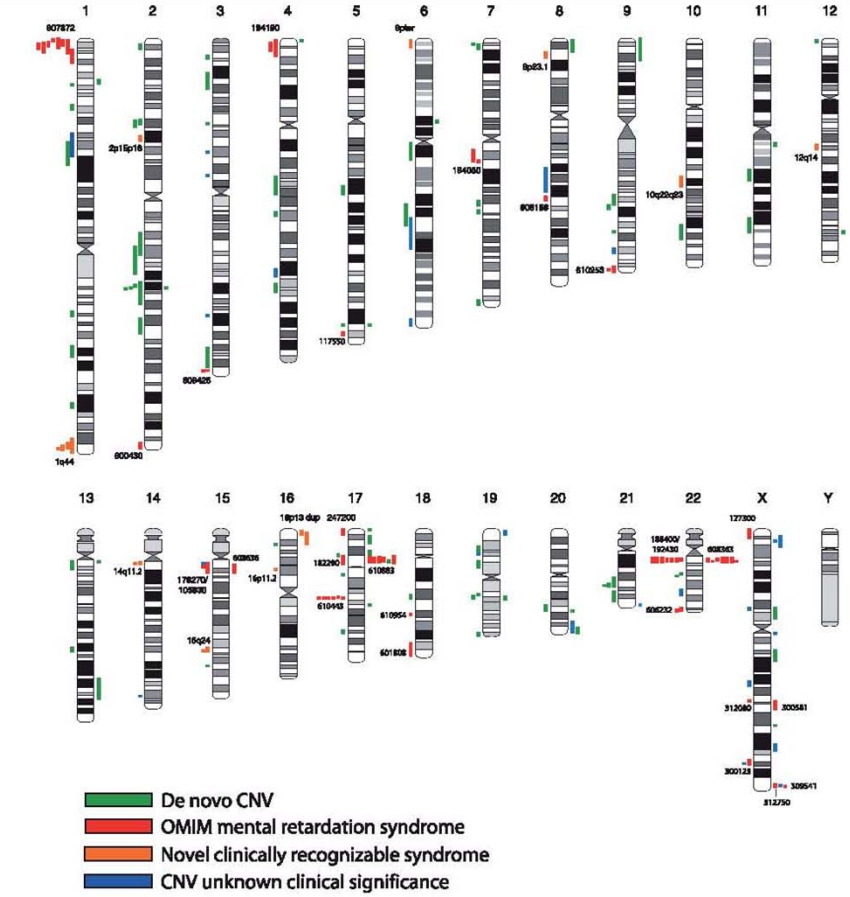
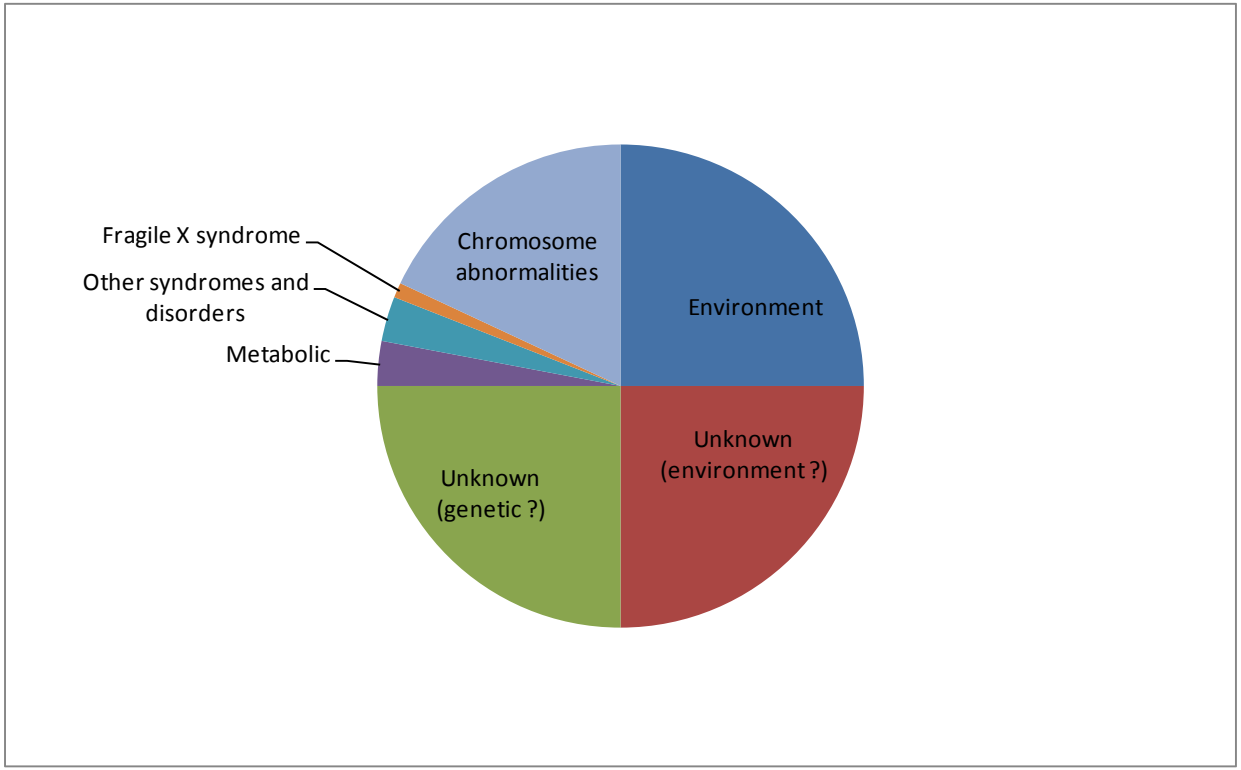


Figure 4. Overview of all CNVs reported in genome-wide microarrays studies in mental retardation. CNVs are represented by colored bars. Copy number losses and copy number gains are depicted, respectively, on the left-hand side and the right-hand side of the chromosomes. 1) Red bars, CNVs associated with a well-known OMIM syndrome; 2) orange bars, novel recurrent CNVs; 3) green bars, de novo CNVs not known to the previous categories; and 4) blue bars, CNVs of unknown clinical significance.

Causes of Intellectual disability a reflection



Draft sequence of the Human Genome
Human Genome Project (1990 – 2001)

Public consortium

Celera
Private company

February 2001

Next generation sequencing

Roche

454 SEQUENCING

illumina

HiSeq 2500

Oxford NANOPORE Technologies

PromethION: small benchtop system for high throughput real-time biological analyses and allowing large sample numbers

Ion Proton™ System

life technologies™

PACBIO®

NGS Data analysis

What kind of data are we working with?

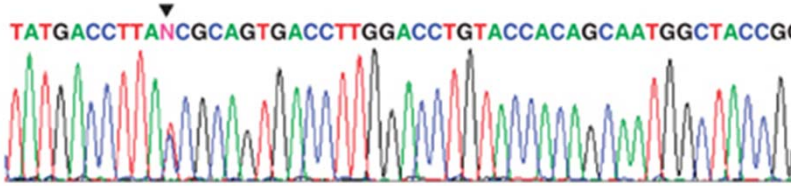
From sequence to variant: Analysis flow

From variant to knowledge: Interpretation flow

NGS Data analysis

What kind of data are we working with?

- Sanger Sequencing:
  - 1 amplicon / reaction
  - 1 sequence / amplicon (or 2)
  - Visual inspection for overlapping peaks

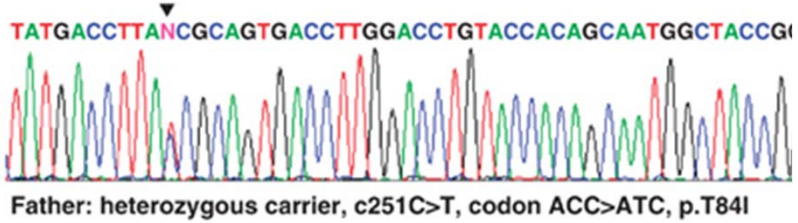


Father: heterozygous carrier, c251C>T, codon ACC>ATC, p.T84I

NGS Data analysis

What kind of data are we working with?

- Sanger Sequencing:
  - 1 amplicon / reaction
  - 1 sequence / amplicon (or 2)
  - Visual inspection for overlapping peaks

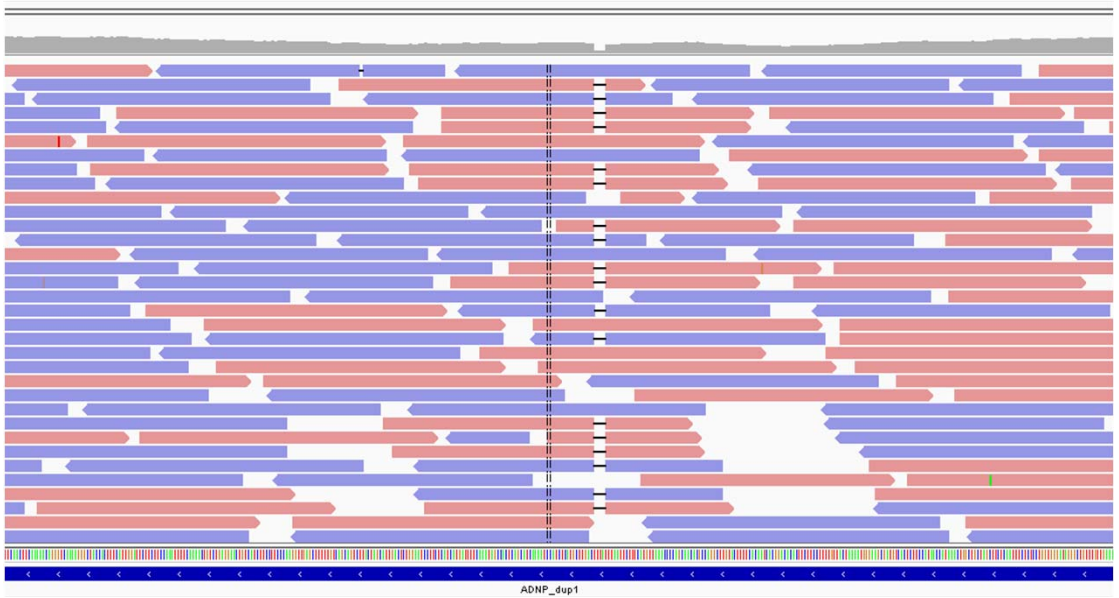


Father: heterozygous carrier, c251C>T, codon ACC>ATC, p.T84I

- Next-Generation Sequencing:
  - Massive Parallel sequencing
  - *exome* panel: > 200.000 targets (all exons in the genome)
  - Multiple amplicons / target
  - *optimal design:* > 40 unique fragments covering *every* nucleotide in targets.

NGS Data analysis

What kind of data are we working with?



=> Amount of data : > 8.000.000 sequences / sample
=> Impossible to manually inspect all reads for mutations.

NGS Data analysis

What kind of data are we working with?

- Data format : FASTQ
- FASTA :

>Sequence_Name
AACTACTAGATACTGATAGTATATCTCTCTTAATCGA
GCTCTAGATCGATCTATACCGAT

- Add Quality (fasta-Q => FASTQ)

@Read_Name
AACTACTAGATACTGATAGTATATCTCTCTTAATCGA
+
BCEECEEFFECGECGECFGFF@?<<=??<>53@##

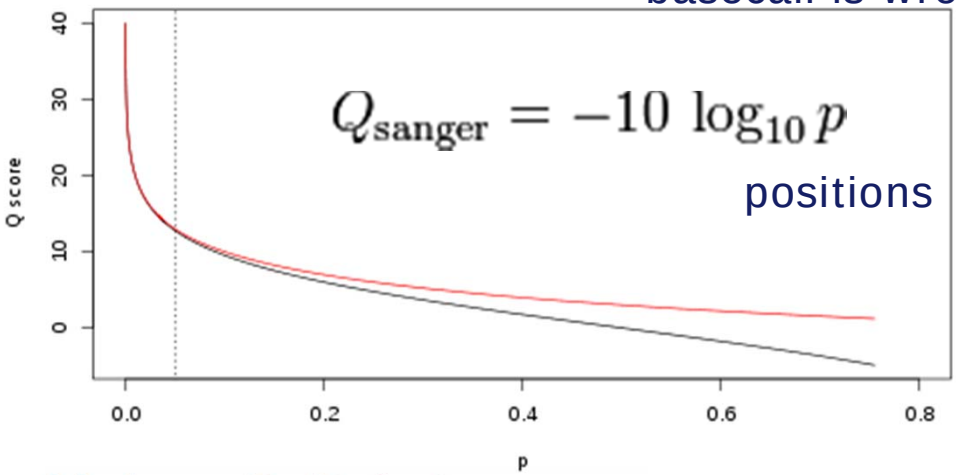
NGS Data analysis

What kind of data are we working with?

- Data format : FASTQ

@Read_Name
AACTACTAGATACTGATAGTATATCTCTCTTAATCGA
+
BCEECEEFFECGECGECFGFF@?<<=??<>53@##

=> *Phred Score* : correlates with the chance on error (probability that basecall is wrong)

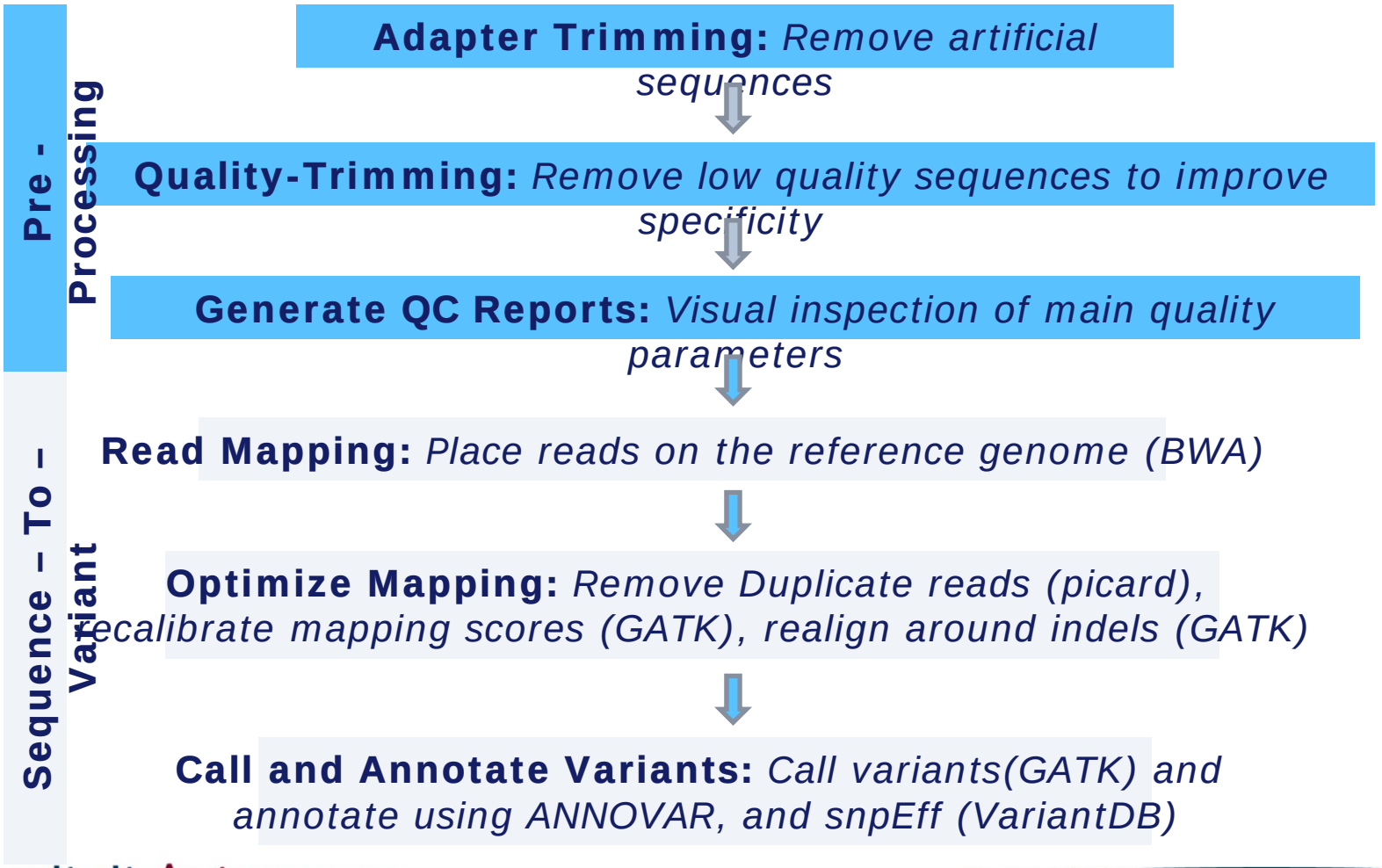


=> *Example:*
Q30 => 1 error / 1000

Q10 => 1 error / 10 positions

NGS Data analysis

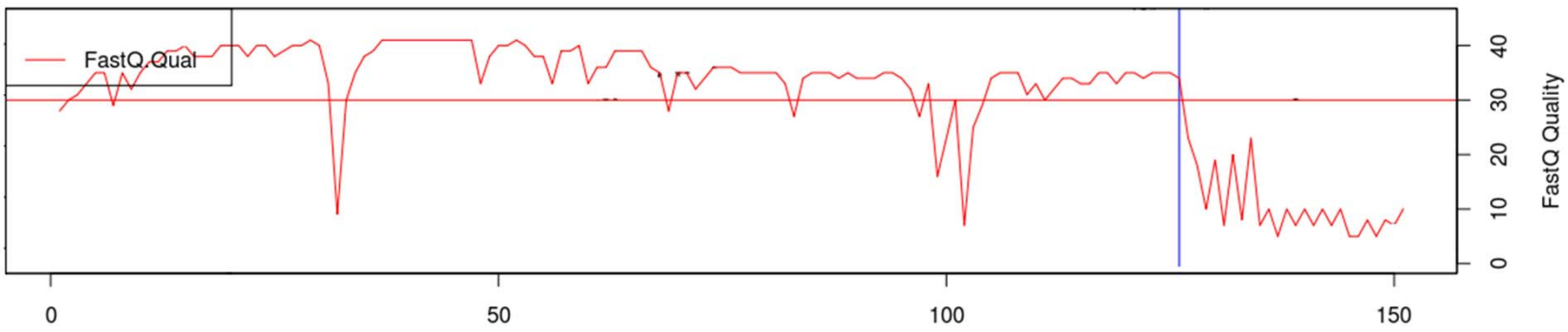
From sequence to variant: Analysis flow



NGS Data analysis

From sequence to variant: Analysis flow

Pre - Processing | **Quality-Trimming**



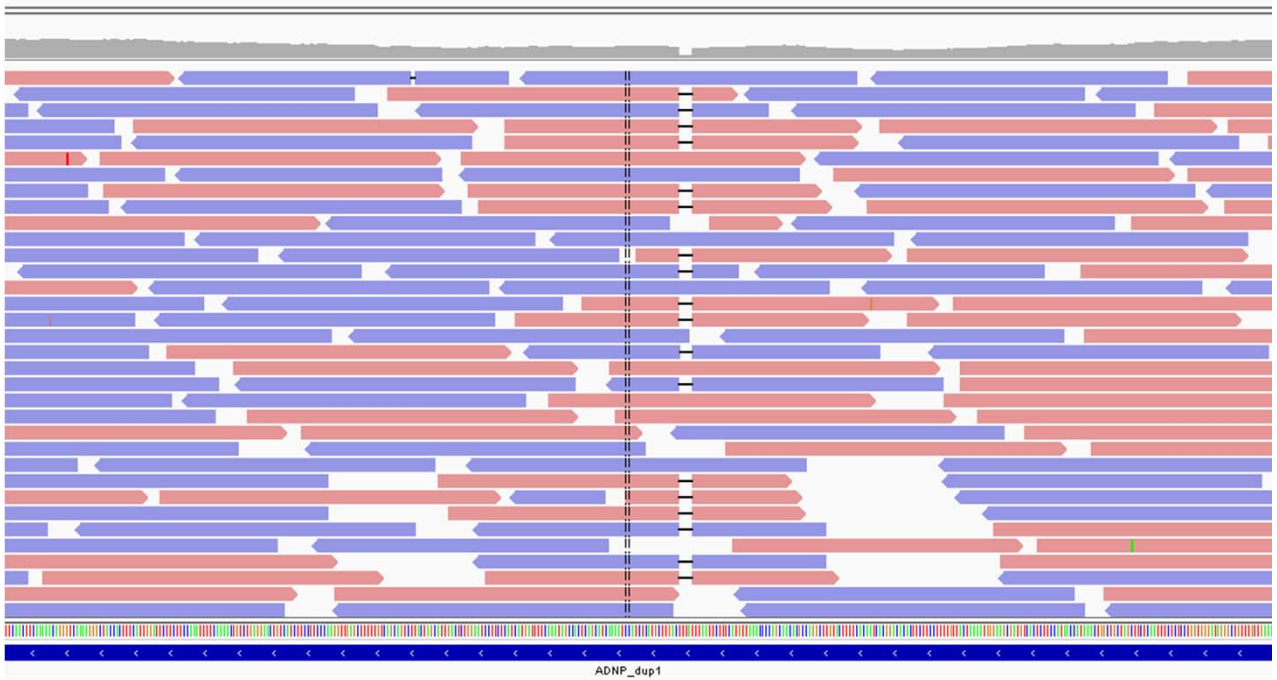
Low quality leads to high error rates (cfr Phred Score)
=> We want a limit of 1 error in 1000 positions
=> Due to chemical degradation, 3' ends are lower quality
=> Trim everything on 3' end with quality < 30

NGS Data analysis

From sequence to variant: Analysis flow

Sequence – To – Variant

Read Mapping



NGS Data analysis

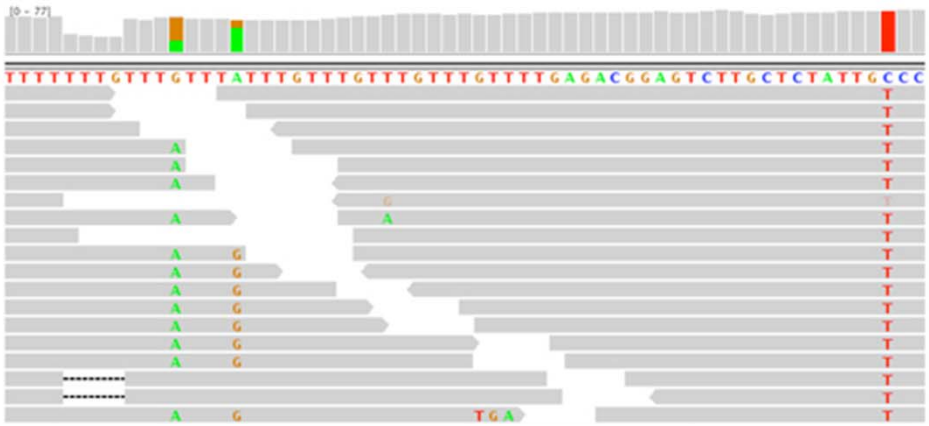
From sequence to variant: Analysis flow

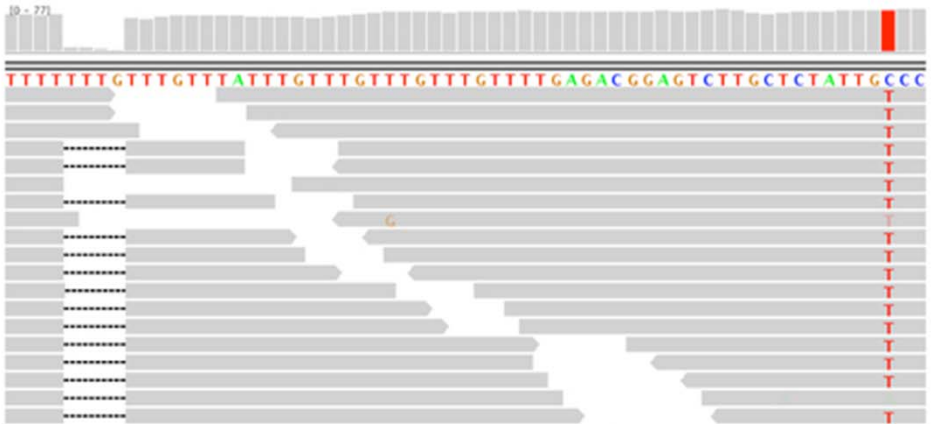
Sequence – To – Variant

Optimize Mapping

- Remove Duplicate reads (picard)
=> Reduce computational time
=> Reduce amplification bias

- Realign around indels (GATK)



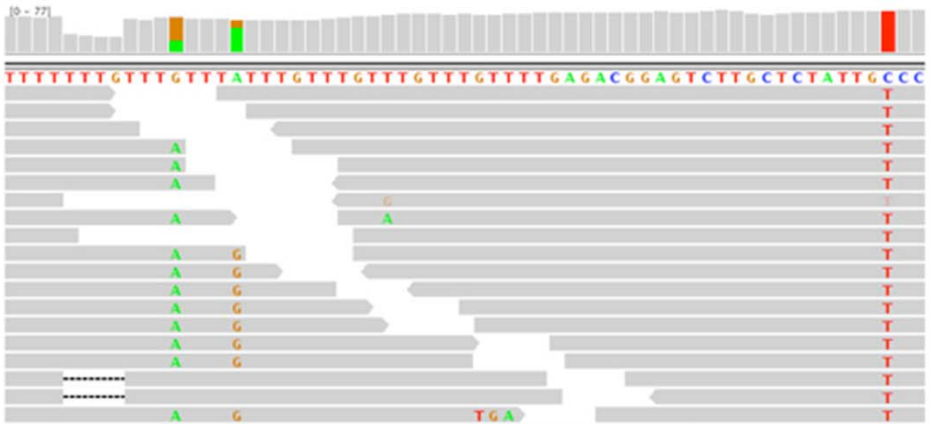


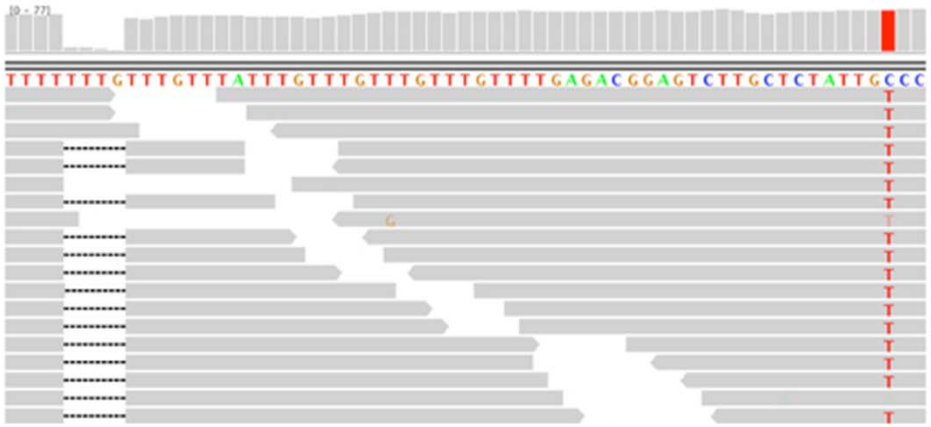
NGS Data analysis

From sequence to variant: Analysis flow

Sequence – To – Variant **Call And Annotate Variants**

- *Call Variants (GATK)*
- *Search for positions with statistically significant evidence for a non-reference nucleotide*
- *Take into account: base-quality, position in read, strand bias, …*

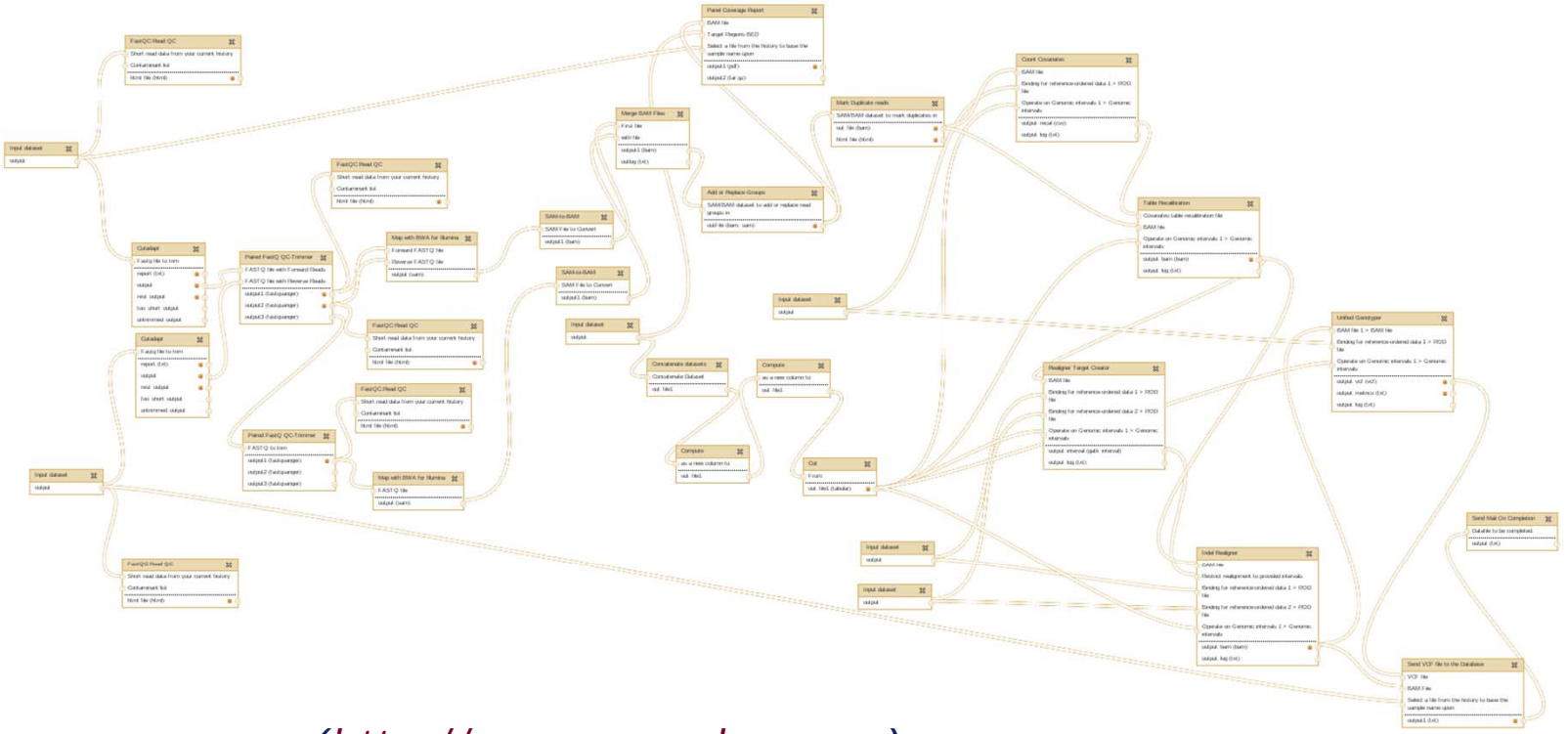




NGS Data analysis

Galaxy

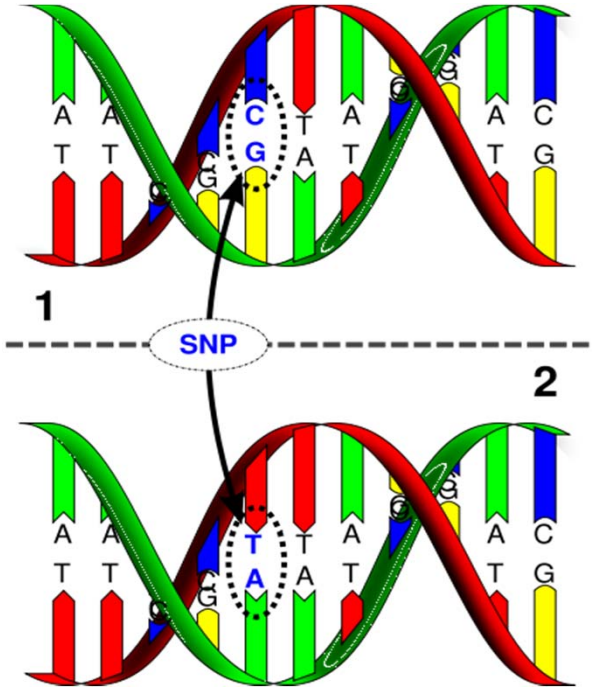
- A website offering an easy way to run complete pipelines
- No programming skills needed, very usefull for dynamic analysis



(http://www.usegalaxy.org)

Genetic Variation

- Single Nucleotide Polymorphism
  - Fixed position in the genome that *might* differ in sequence between individuals
  - 1 SNP / 700bp
  - Long thought to be major source of variation



NGS Data analysis

From sequence to variant: Analysis flow

Sequence – To – Variant **Call And Annotate Variants**

Annotate Variants (ANNOVAR, snpEff, …)

- Add as much information to the variant as possible to ease interpretation

- Effect on Gene transcription (RefSeq, Ensembl, UCSC)

- Quality parameters (GATK)

- Occurence in control populations (dbSNP, ESP, HapMap, 1KG, …)

- Known pathogenic variations (dbSNP, OMIM, …)

- Effect on gene function (PolyPhen, MutationTaster, Sift, …)

- …

NGS Data analysis

From variant to knowledge:
Interpretation flow

Effect on gene function (~ from high to low)

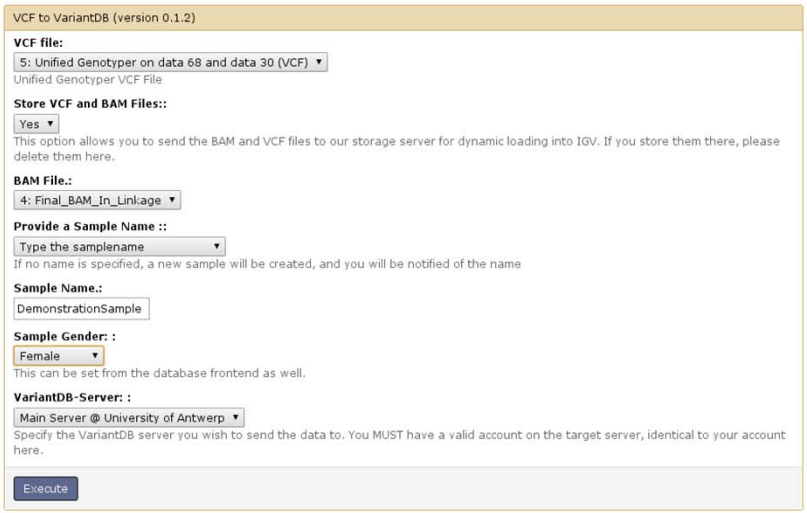
- Variant causes gain/loss of stop/start coding?
- Variant causes aberrant splicing of the transcript?
- Variant replaces a highly conserved nucleotide/amino acid ?
- Variant replaces an aminoacid, and is not reported in control populations ?
- Variant can modify binding of regulatory elements?
- ...

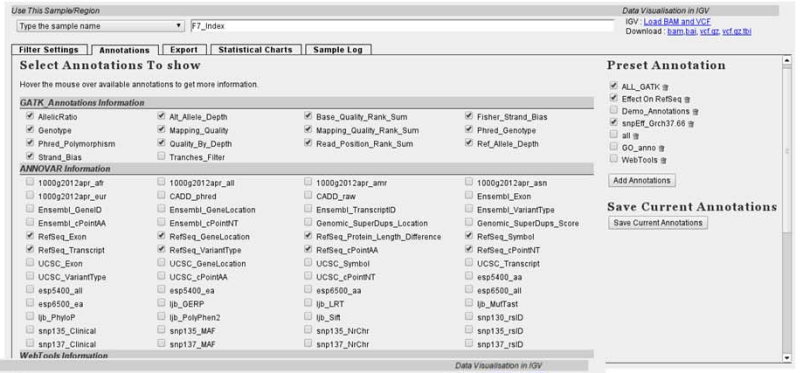
⇒ Extended annotation is critical

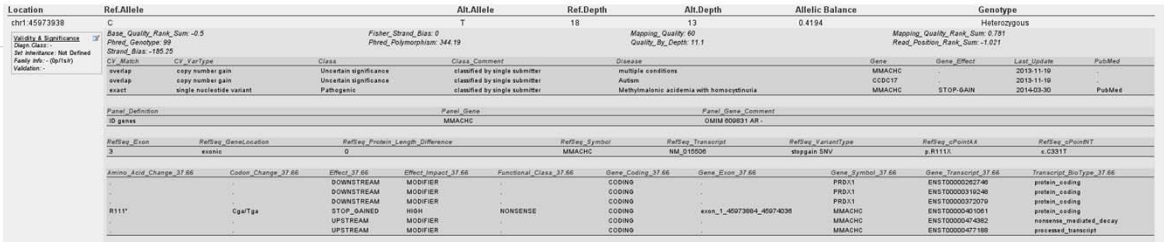
⇒ Manual inspection of > 20.000 variants/sample is impossible.
⇒ Again: automation is needed

VariantDB

Web-based interface





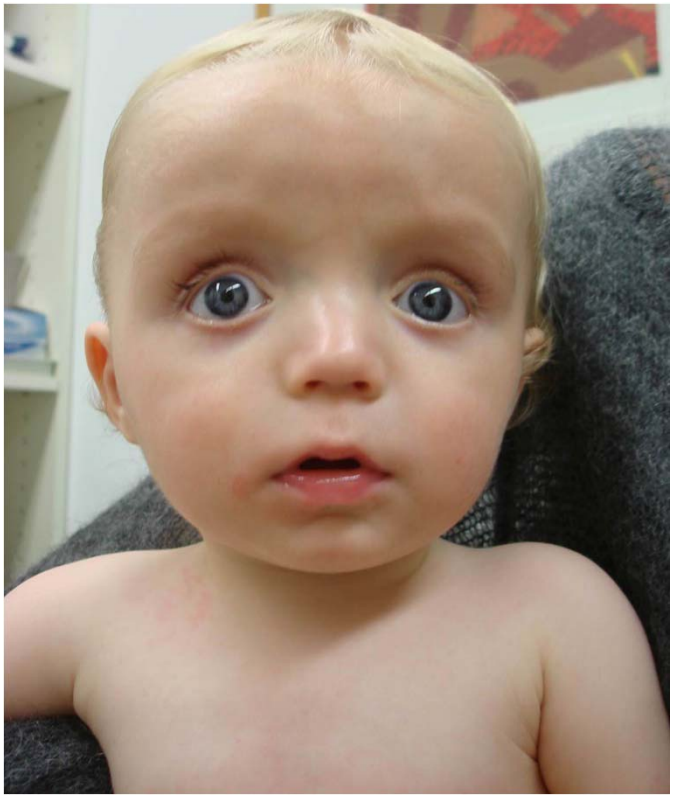


Universiteit Antwerpen

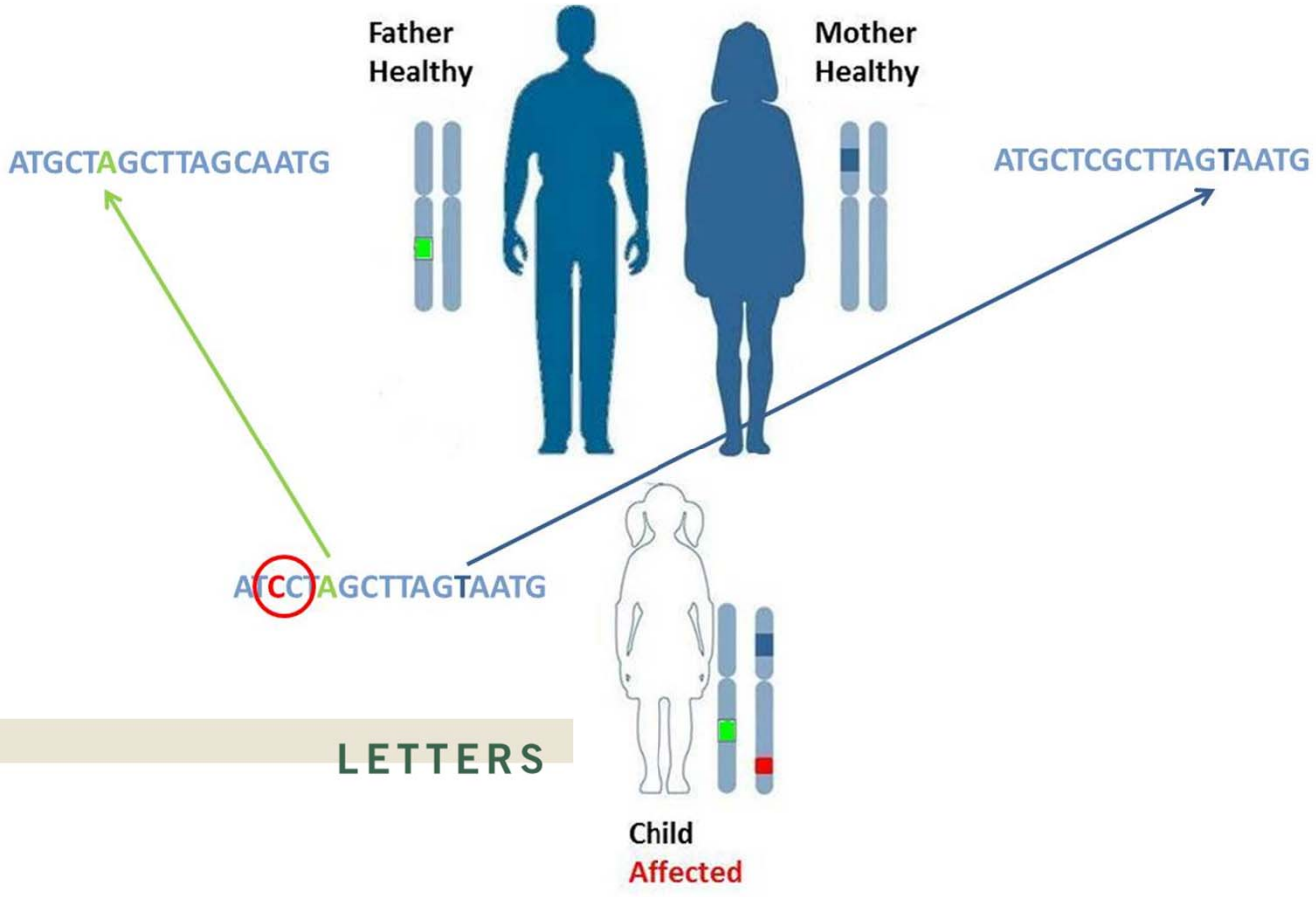
Vandeweyer *et al.*, Genome Medicine 2014

Young boy

- Global Developmental Delay
- Failure to Thrive
- Autism Spectrum Disorder
- Cardiac: Atrial Septal Defect
- White Matter Lesions
- Facial Dysmorphism



Next Generation Sequencing
Trio approach



LETTERS

nature genetics

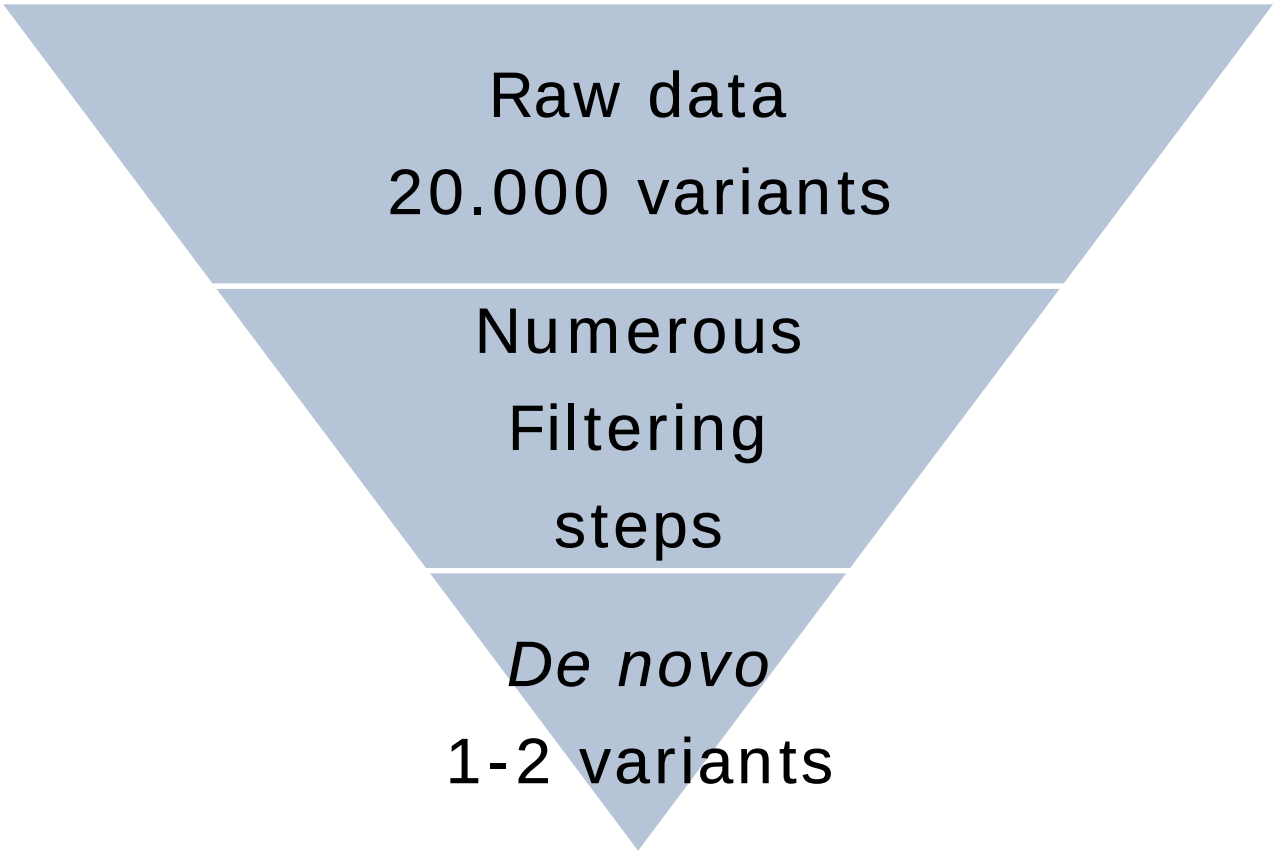
A *de novo* paradigm for mental retardation

Lisenka E L M Vissers[1,2], Joep de Ligt[1,2], Christian Gilissen[1], Irene Janssen[1], Marloes Steehouwer[1], Petra de Vries[1], Bart van Lier[1], Peer Arts[1], Nienke Wieskamp[1], Marisol del Rosario[1], Bregje W M van Bon[1], Alexander Hoischen[1], Bert B A de Vries[1], Han G Brunner[1,3] & Joris A Veltman[1,3]

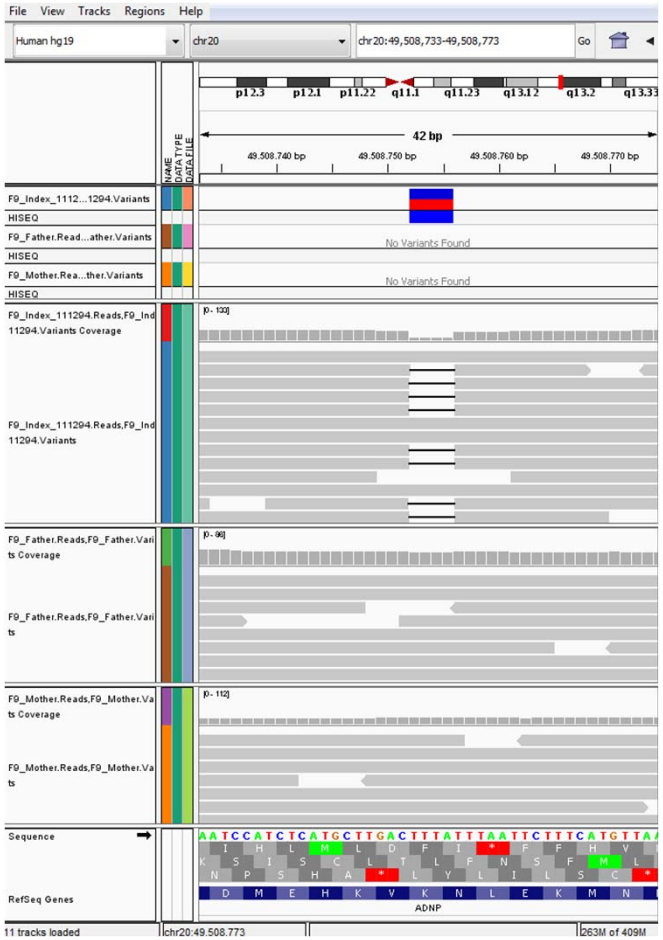
Nature Genetics **42**, 1109–1112 (2010)

Trio approach

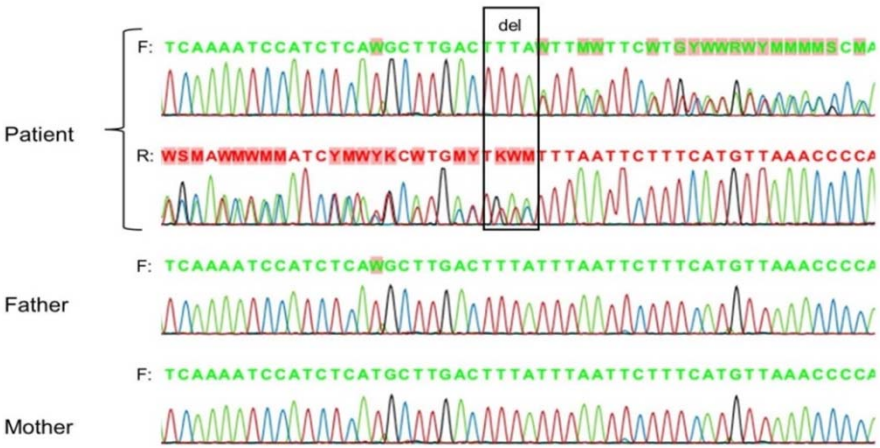
Variant calling



NGS analysis



NGS analysis:
a single *de novo* mutation



- 4bp del in *ADNP*
- Causes frameshift introducing stop codon
- Mutation not control databases, e.g.,ESP, 1000g, dbSNP

ADNP gene

Activity-Dependent Neuroprotective Protein

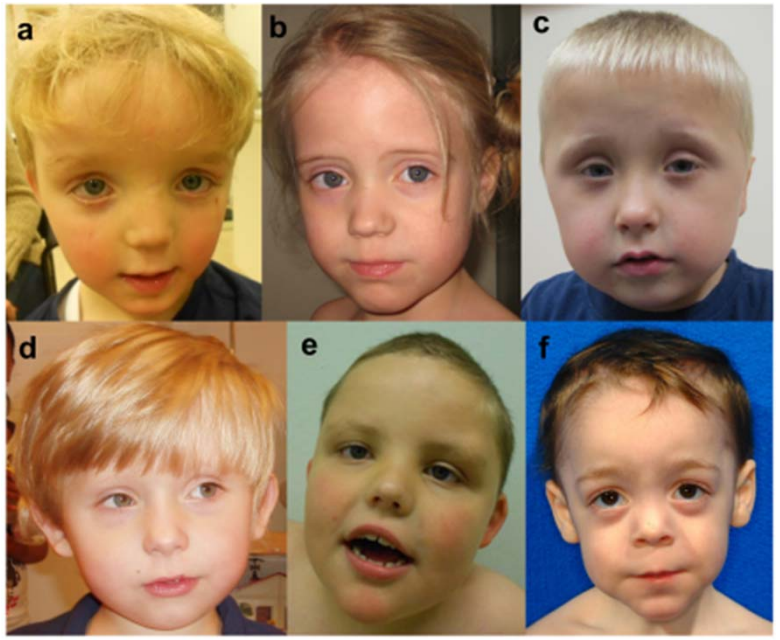
- Expressed in brain
- Zinc fingers/homeobox domain: potential transcription factor
- Involved in neurogenesis
- Involved in heart development
- Homozygous KO mice are embryonically lethal
- Heterozygous KO mice have cognitive & behavioral problems

Ten patients with truncating mutations in *ADNP*

Patient	Patient ID	Origin	Screening method	Cohort composition	Cohort size	mutation in genomic DNA (chr20)	mutation in cDNA (NM_015339.2)	Protein	Mutation Type	Inheritance
1	111294	Antwerp	WES	Moderate to severe ID and/or autism + dysmorphic features	10	g.49508752_49508755 delTTTA	c.2496_2499delTAAA	p.Asp832Lysfs*80	Frameshift	*de novo*
2	11-08612	Nijmegen	WES	Non-syndromic severe ID	100	g.49510040G>T	c.1211C>A	p.Ser404*	Nonsense	*de novo*
3	12130.p1	Seattle	WES[2,16]	ASD from the Simon Simplex Collection	189	g.49510028_49510029 delTT	c.1222_1223delAA	p.Lys408Valfs*31	Frameshift	*de novo*
4	1050237	Westmead	WES	Non-syndromic severe ID	95	g.49509086_49509098 delATTTGCTCGTAAG	c.2153_2165delCTTAC GAGCAAAT	p.Thr718Glyfs*12	Frameshift	*de novo*
5	3061-08D	Stockholm	WES	Moderate to severe ID and/or autism + dysmorphic features	45	g.49509094G>C	c.2157C>G	p.Tyr719*	Nonsense	*de novo*
6	122793	Antwerp	HRM	Autism	148	g.49508757_49508760 delTTAA	c.2491_2494delTTAA	p.Lys831Ilefs*81	Frameshift	*de novo*
7	07-06960	Nijmegen	MIPS	ID and/or autism	2743	g.49508443delG	c.2808delC	P.Tyr936*	Frameshift	*de novo*
8	2376	Troina	MIPS	ID and/or autism	Idem as patient 7	g.49508757_49508760 delTTAA	c.2491_2494delTTAA	p.Lys831Ilefs*81	Frameshift	*de novo*
9	2533	Troina	MIPS	ID and/or autism	Idem as patient 7	g.49509321G>A	c.1930C>T	p.644Arg*	Nonsense	parents not available
10	13545.p1	Seattle	MIPS[16]	ASD from the Simon Simplex Collection	2446	g.49509094_49509095 insT	c.2156_2157insA	p.Tyr719*	Frameshift	*de novo*

Dysmorphic features

Phenotype	Frequency
Prominent forehead	5/8
High hairline	7/8
Eversion/notch eyelid	3/7
Hypertelorism	1/8
Broad nasal bridge	6/8
Short nose	2/8
Thin upper lip	6/7



Clinical characteri

Phenotype	Frequency
Autism Spectrum Disorder (ASD)	**10/10**
Intellectual Disability (ID)	**10/10**
Developmental delay (motor)	**9/10**
Developmental delay (speech)	8/9
ADHD	2/9
Hypotonia	7/9
Growth retardation / Short stature	5/8
Feeding problems	5/8
Recurrent infections	5/8
Congenital heart defect	3/8
Hyperlaxity	6/8
Obesity	4/7
Hypermetropia	6/6
Seizures	2/7
Behavior	5/7
Insensitivity to pain	2/5
MRI brain abnormality	5/9
Hand abnormalities	6/8
Constipation	2/6

Statistics

- The frequency of truncating *de novo* mutations in *ADNP* is significantly higher (p: 0.001852, odds ratio 13.24668, one-sided Fisher's exact test) in patients compared to the ESP cohort and additional controls from the Simons Siblings.
- The probability of detecting 8 or more *de novo* truncating events in *ADNP* within our cohort by chance was estimated to be p = $2.65e^{-18}$ (binomial test) under a *de novo* rate of 1.2 non-synonymous coding variants per individual according to a probabilistic model of a locus specific enrichment for truncating variation [O'Roak et al., 2012].

De novo mutations in *ADNP* cause new autism syndrome

Autism spectrum disorder
Intellectual Disability
Facial dysmorphology
Congenital malformations

may explain etiology of
0.17% of ASD patients
(95% confidence interval:
0.083–0.32%)

LETTERS

nature genetics

A SWI/SNF-related autism syndrome caused by *de novo* mutations in *ADNP*

Céline Helsmoortel[1], Anneke T Vulto-van Silfhout[2,16], Bradley P Coe[3,4,16], Geert Vandeweyer[1,5], Liesbeth Rooms[1], Jenneke van den Ende[6], Janneke H M Schuurs-Hoeijmakers[2], Carlo L Marcelis[2], Marjolein H Willemsen[2], Lisenka E L M Vissers[2], Helger G Yntema[2], Madhura Bakshi[7], Meredith Wilson[8], Kali T Witherspoon[3,4], Helena Malmgren[9], Ann Nordgren[9], Göran Annerén[10], Marco Fichera[11,12], Paolo Bosco[13], Corrado Romano[14], Bert B A de Vries[2,15], Tjitske Kleefstra[2,15], R Frank Kooy[1], Evan E Eichler[3,4] & Nathalie Van der Aa[1,6]

Helsmoortel *et al.*, Nat Genet 2014

Acknowledgements

Céline Helsmoortel
Ilse van der Werf
Anke Van Dijck
Geert Vandeweyer
Nathalie Van der Aa

Liesbeth Rooms
Jenneke van den Ende