# Detección de posibles mosaicismos estructurales y puntuales

**Curso Medicina Genómica Personalizada**

Barcelona, 5-7 Noviembre 2018

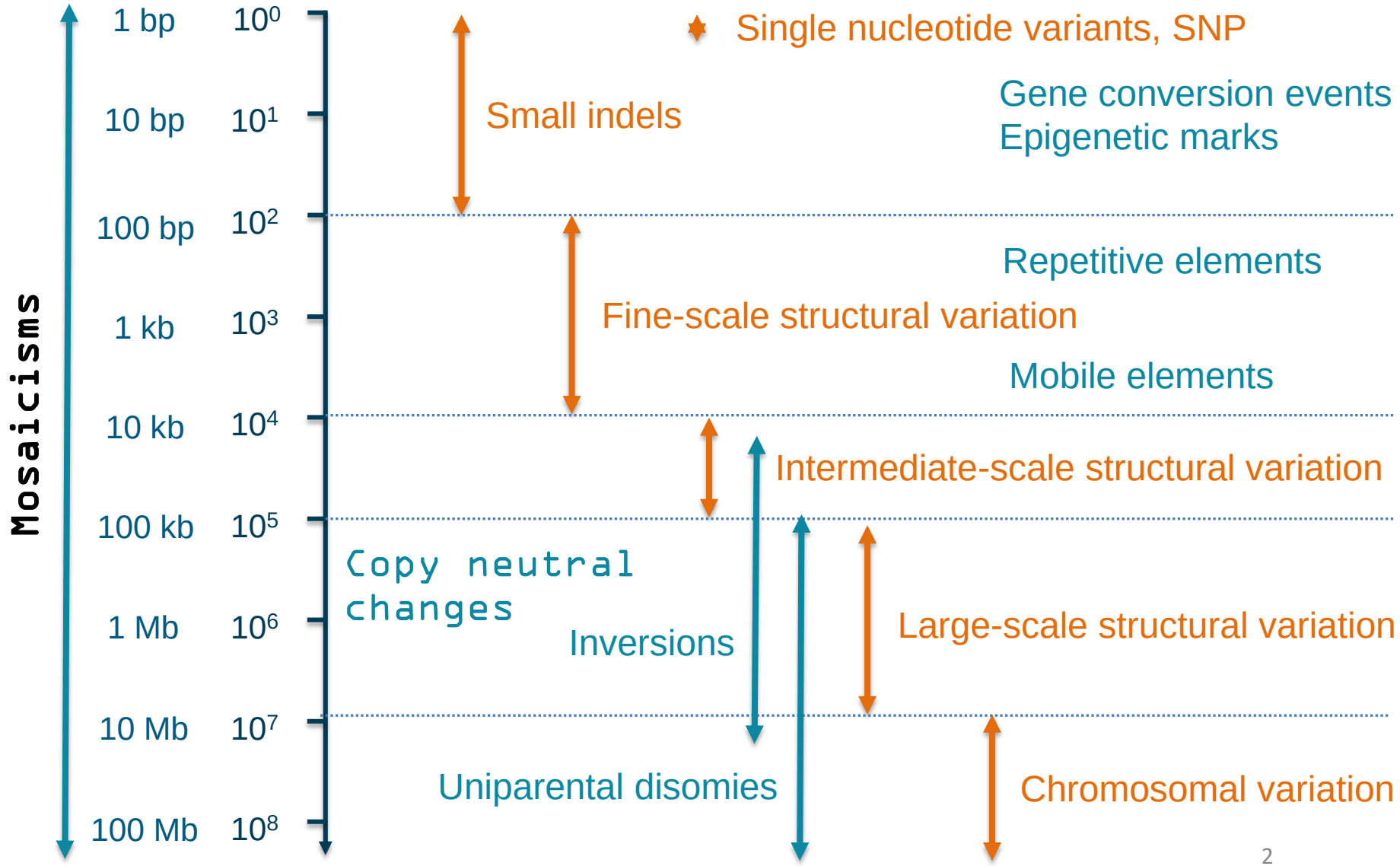

**Juan R Gonzalez**

Bioinformatics Research Group in Epidemiology (BRGE), ISGlobal

www.brge.isglobal.org

**email:** juanr.gonzalez@isglobal.org

# Spectrum of variation in the human genome

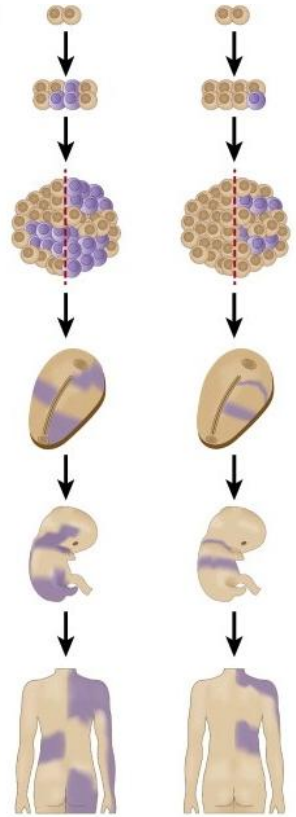


# Genetic Mosaicisms

**Genetic Mosaicism:** Presence of two population of cells with different genotype in one individual

# Genetic Mosaicisms

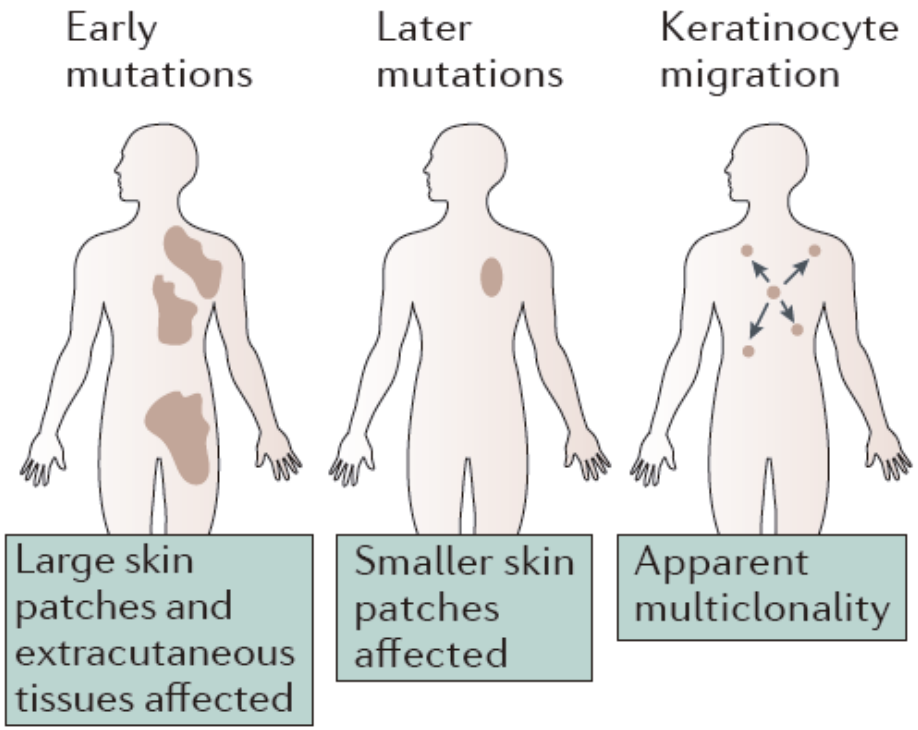
## Early somatic mosaicism



Early developmental event propagated to daughter cells

Adapted from Campbell IM 2015

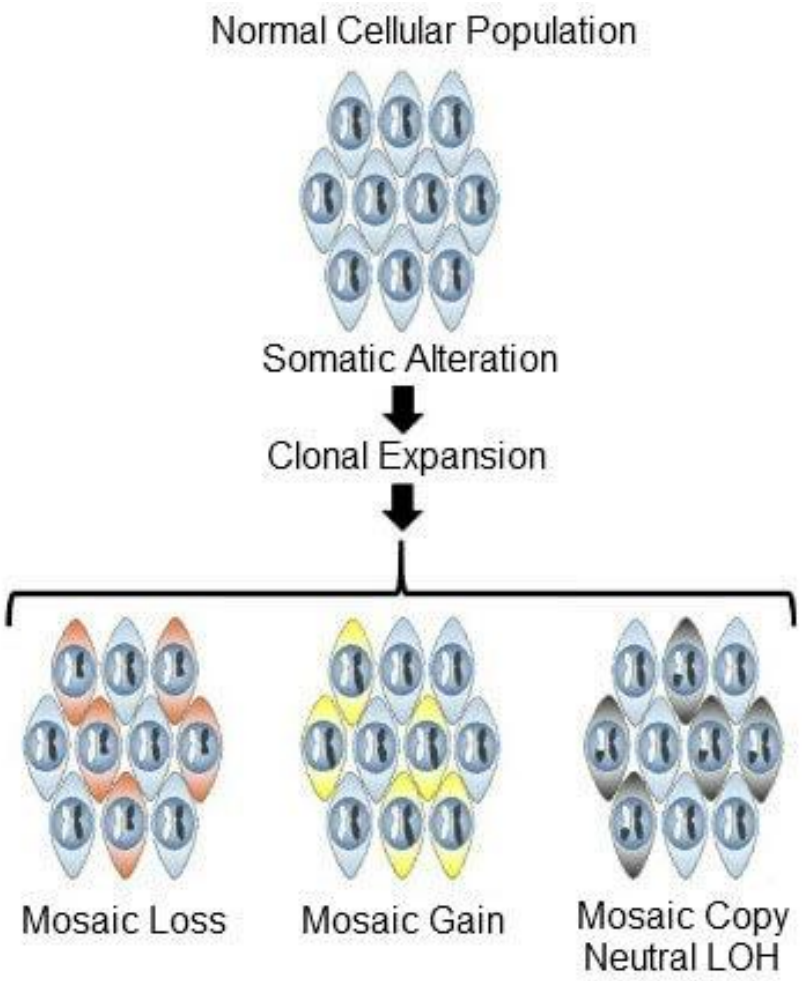
## Late somatic mosaicism




Early or late somatic event with clonal selection for variable reasons

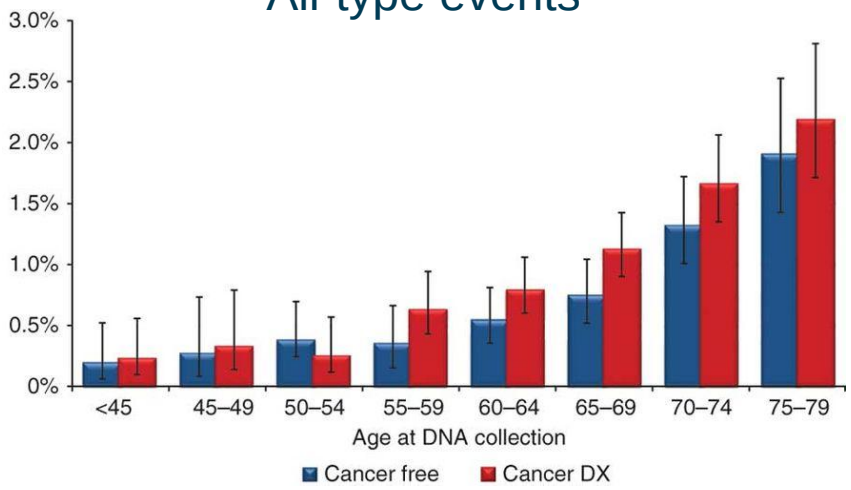
Adapted from Fernández, Luis C 2015

# Genetic Mosaicisms



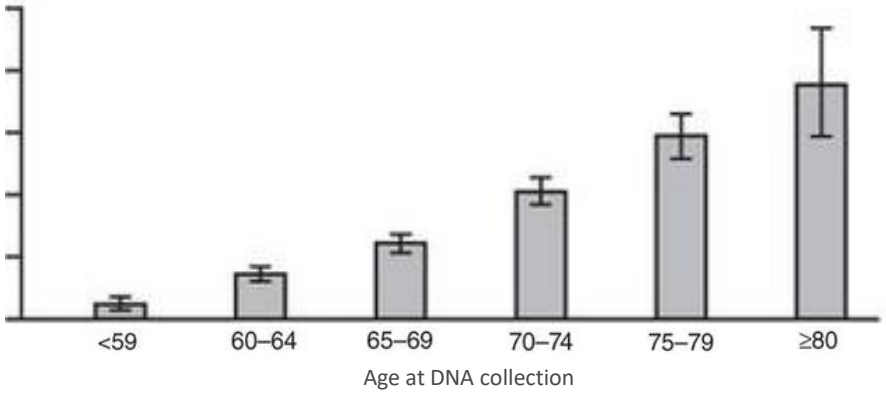
# Genetic Mosaicisms

## All-type events




Adapted from Jacobs, KB et al. Nat Genet, 2012

## Mosaic loss of ChrY in males




Adapted from Zhou, W et al. Nat Genet, 2016

## Associated with

- Late events:
  - Aging
  - Solid and hematological cancer
- Early events:
  - Developmental disorders
  - Brain and other malformations
  - Pregnancy loss

- Late events:
  - All-type mortality
  - Alzheimer, cardiovascular events
  - Hematological cancer
- Early events:
  - Mosaic Turner syndrome and related effects

# Phenotypic Spectrum

| Mosaic Disorders | Tissue Affected | Chromosome |
|---|---|---|
| Chronic lymphocytic leukemia | Blood | 13q14 |
| Down's syndrome | Multiple | 21 |
| Keratinocytic epidermal nevus | Skin | 11p |
| Maffucci syndrome | Multiple | 2q,15q |
| McCune-Albright syndrome | Multiple | 20q |
| Nevus sebaceous | Skin | 11p,12p |
| Ollier disease | Connective | 2q,15q |
| Proteus syndrome | Multiple | 14q |
| Schimmelpenning syndrome | Multiple | 11p,12p |
| Turner's syndrome | Multiple | X |

# Structural mosaic events

By event type and location

| | Mosaic Chromosome Count | | | | | Mosaic Chromosome Frequency (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Event Location | Gain | Loss | CN LOH | Mixed | Total | Gain | Loss | CN LOH | Mixed | Total |
| Chromosome | 70 | 11 | 45 | 5 | 131 | 6.7 | 1.0 | 4.3 | 0.5 | 12.5 |
| Telomeric p | 16 | 24 | 144 | 1 | 185 | 1.5 | 2.3 | 13.7 | 0.1 | 17.6 |
| Telomeric q | 22 | 26 | 232 | 0 | 280 | 2.1 | 2.5 | 22.1 | 0.0 | 26.6 |
| Interstitial | 46 | 379 | 2 | 1 | 428 | 4.4 | 36.1 | 0.2 | 0.1 | 40.7 |
| Span centromere | 2 | 1 | 2 | 0 | 5 | 0.2 | 0.1 | 0.2 | 0.0 | 0.5 |
| Complex | 1 | 5 | 9 | 7 | 22 | 0.1 | 0.5 | 0.9 | 0.7 | 2.1 |
| Total | 157 | 446 | 434 | 14 | 1051 | 14.9 | 42.4 | 41.3 | 1.3 | 100.0 |

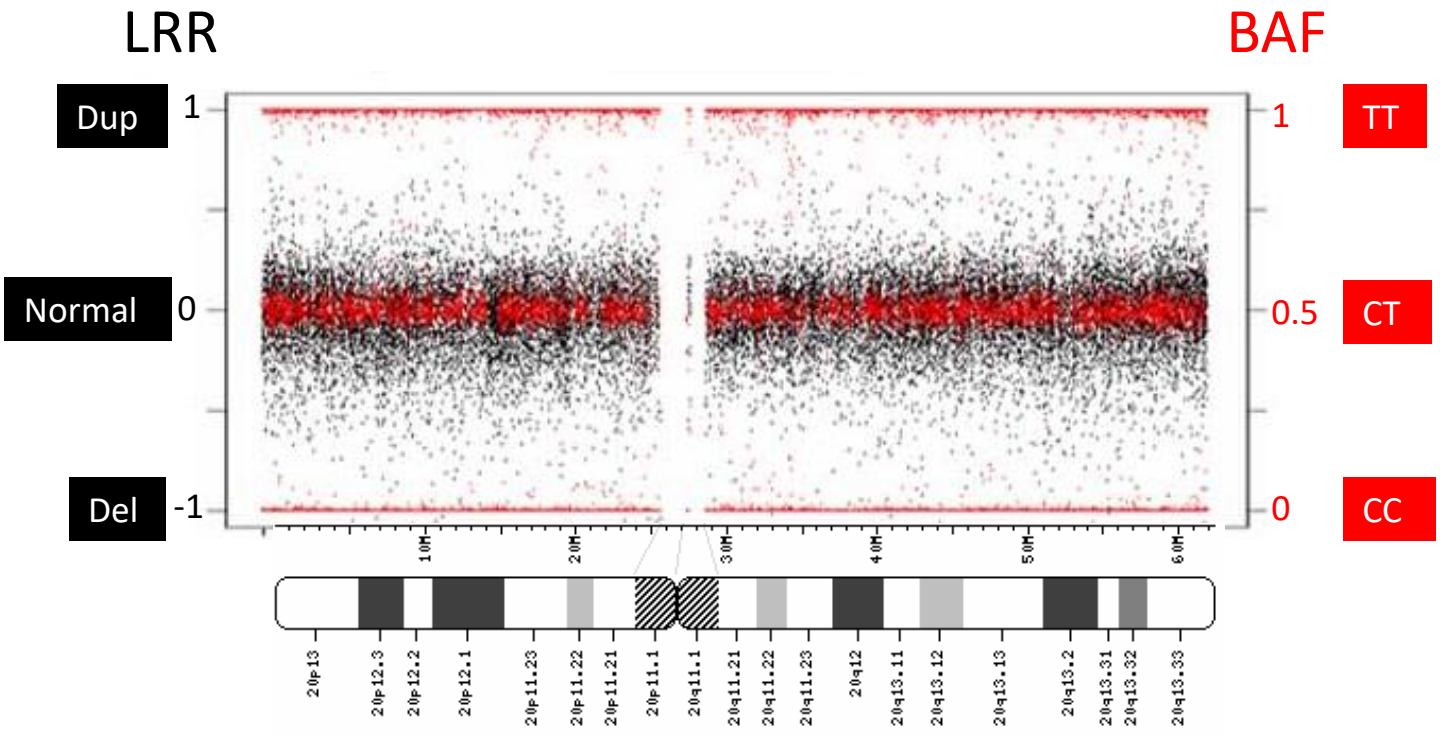
Abbreviation: CN LOH, copy-neutral loss of heterozygosity.

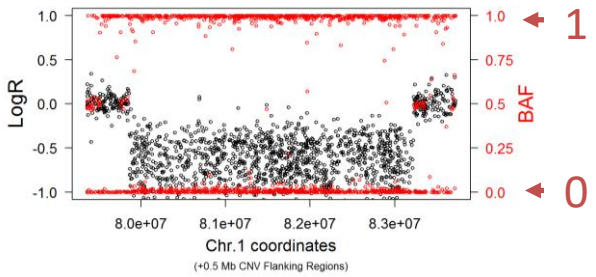
Adapted from Laurie et al. Nat Genet, 2012 and Jacobs et al. Nat Genet, 2012

# Mosaic detection

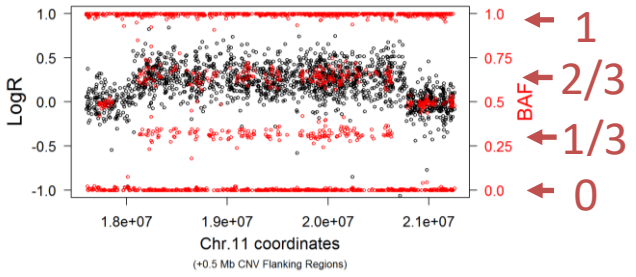
- Cytogenetics and fluorescent in situ hybridization
- Sanger sequencing and pyrosequencing
- Single cell sequencing
- Personalized assays (targeted approach)
- **Arrays (aCGH or SNPs)**
- **Massively-parallel sequencing (WES or WGS)**

# Copy number variants (CNVs)



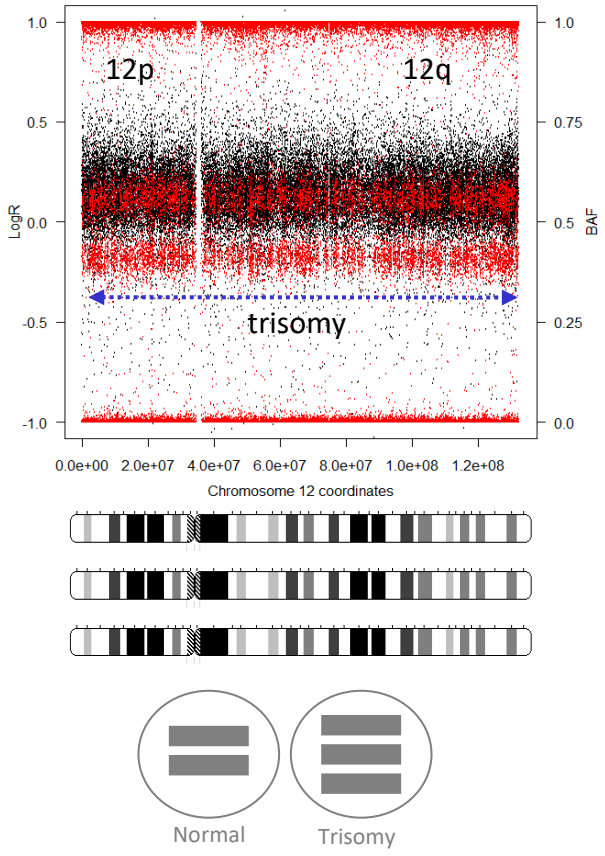

CNV Deletion

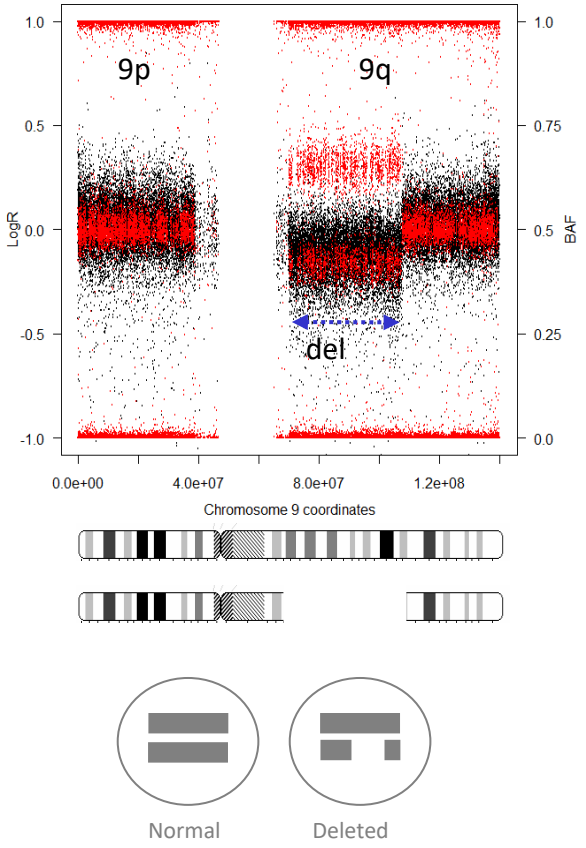


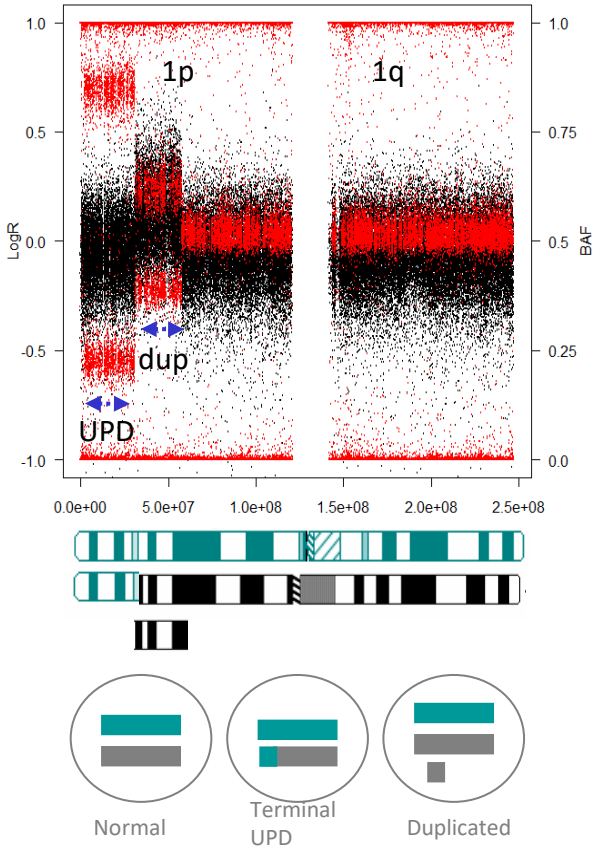

CNV Duplication

# Structural mosaic events

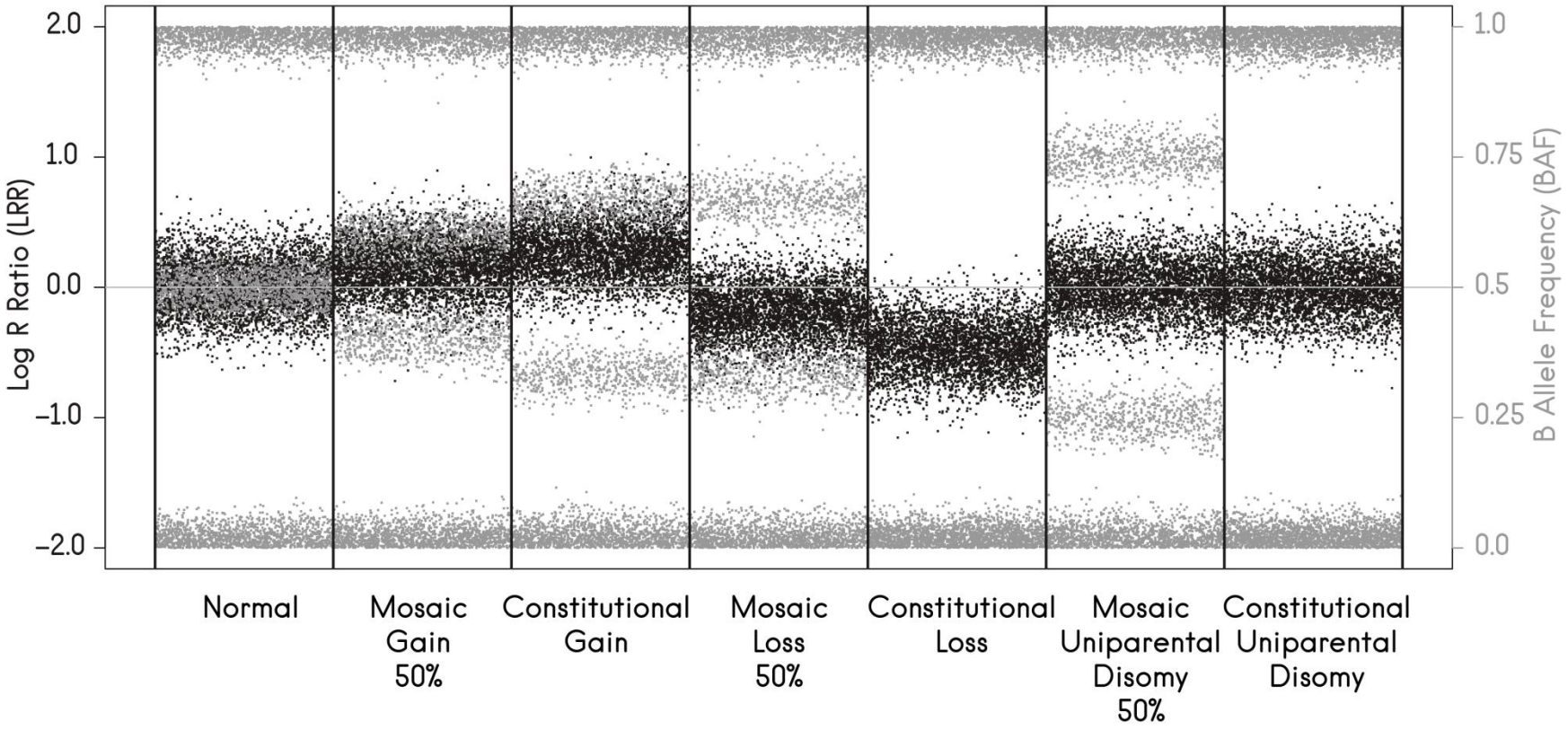
## Type of Mosaic rearrangements



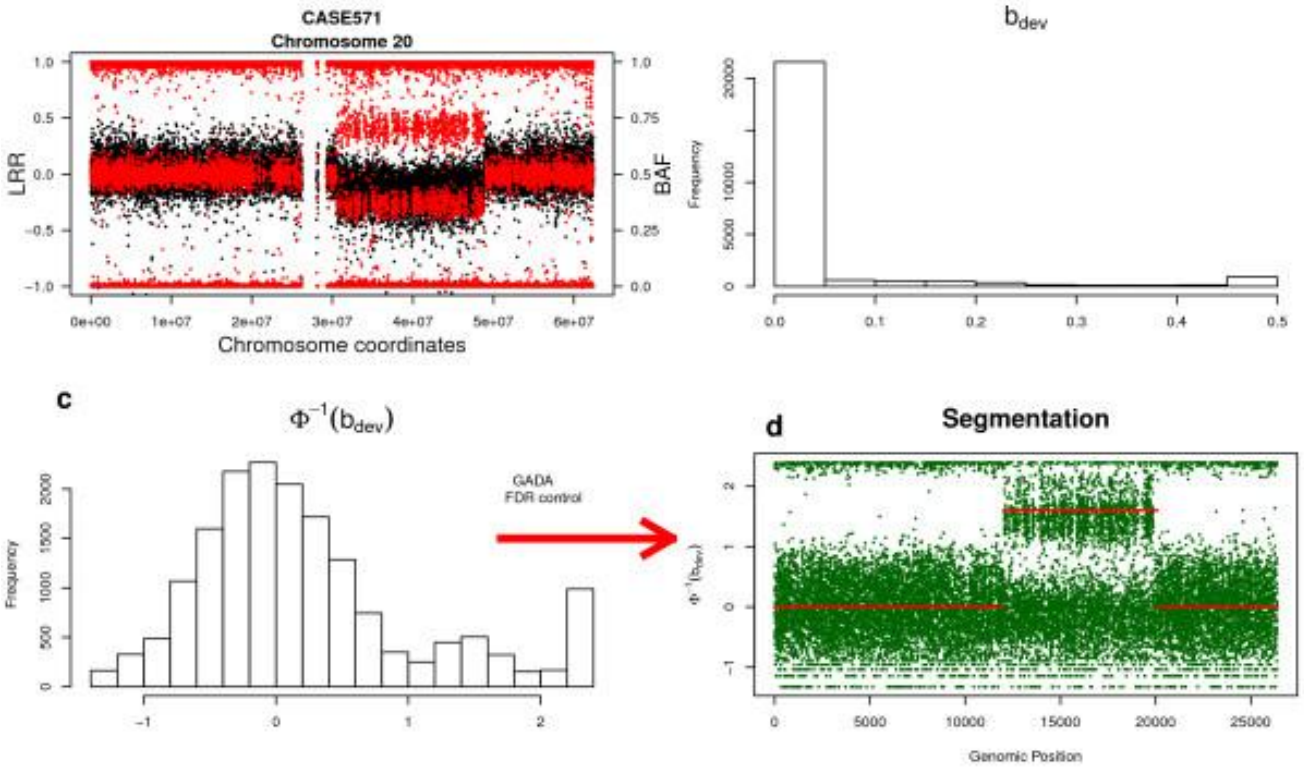



# Structural mosaic events



# Structural mosaic events

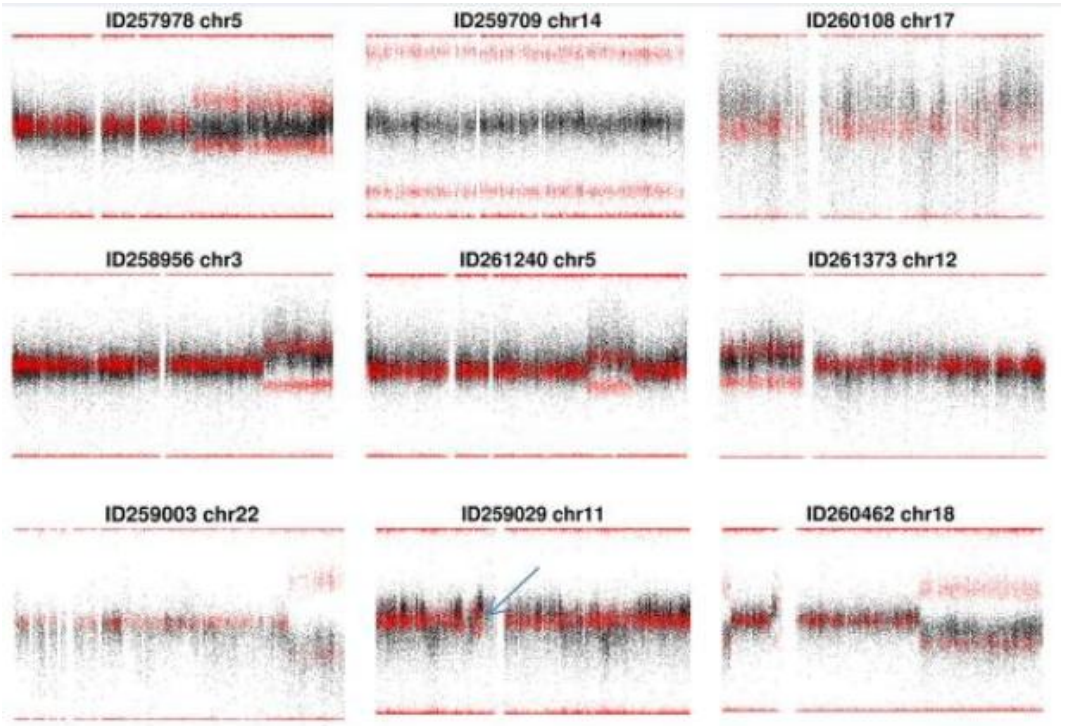
**MAD:** https://github.com/isglobal-brge/MAD




González JR et al, BMC Bioinf, 2011

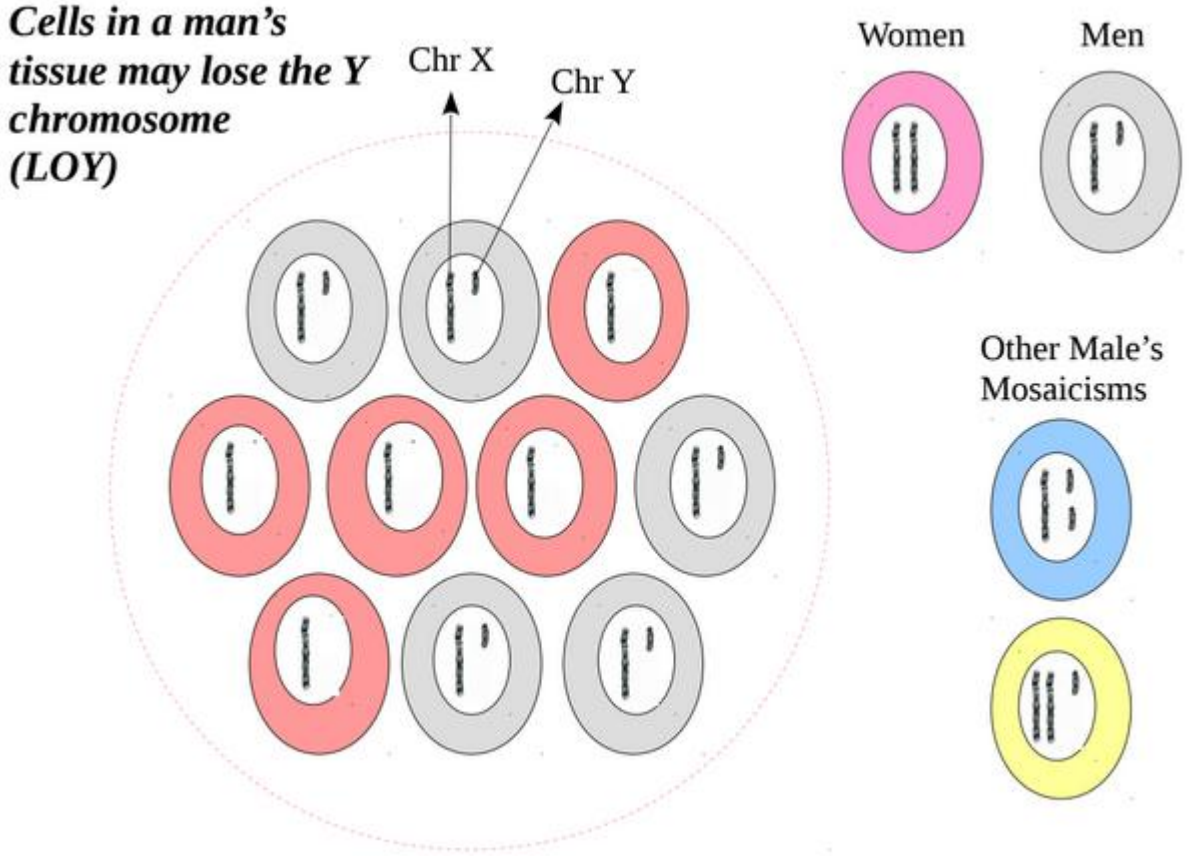
# Structural mosaic events

- Case–control analysis: 1303 cases with **developmental disorders** and 5094 controls (OR = 39.4, *P*-value 1.073e − 6)
- A meta-analysis that included frequency estimates 7000 children with congenital diseases yielded an even stronger statistical enrichment (*P*-value 1.784e − 11)

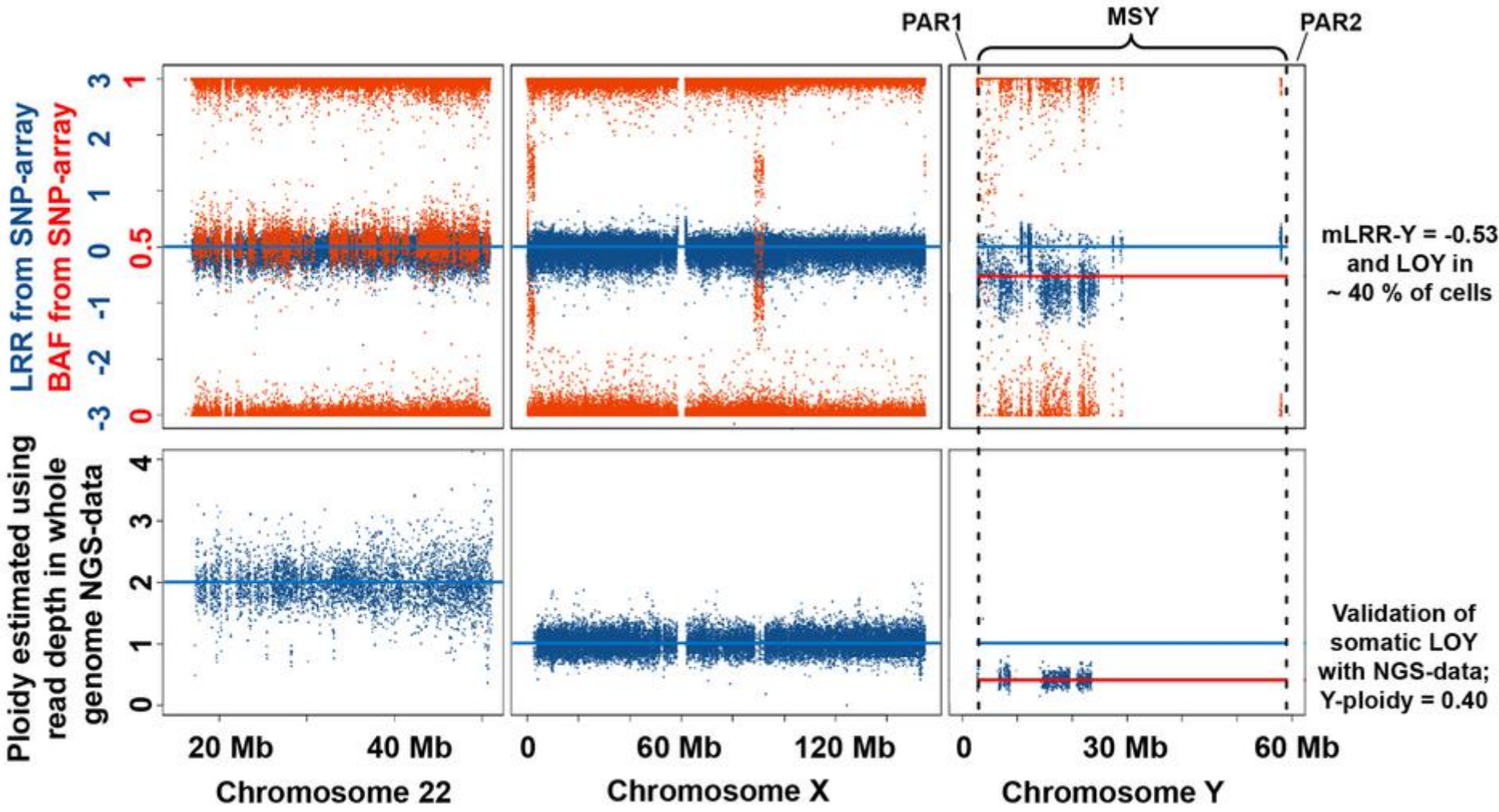



King DA et al, Hum Mol Genet, 2015

# Loss of Chromosome Y



# Loss of Chromosome Y



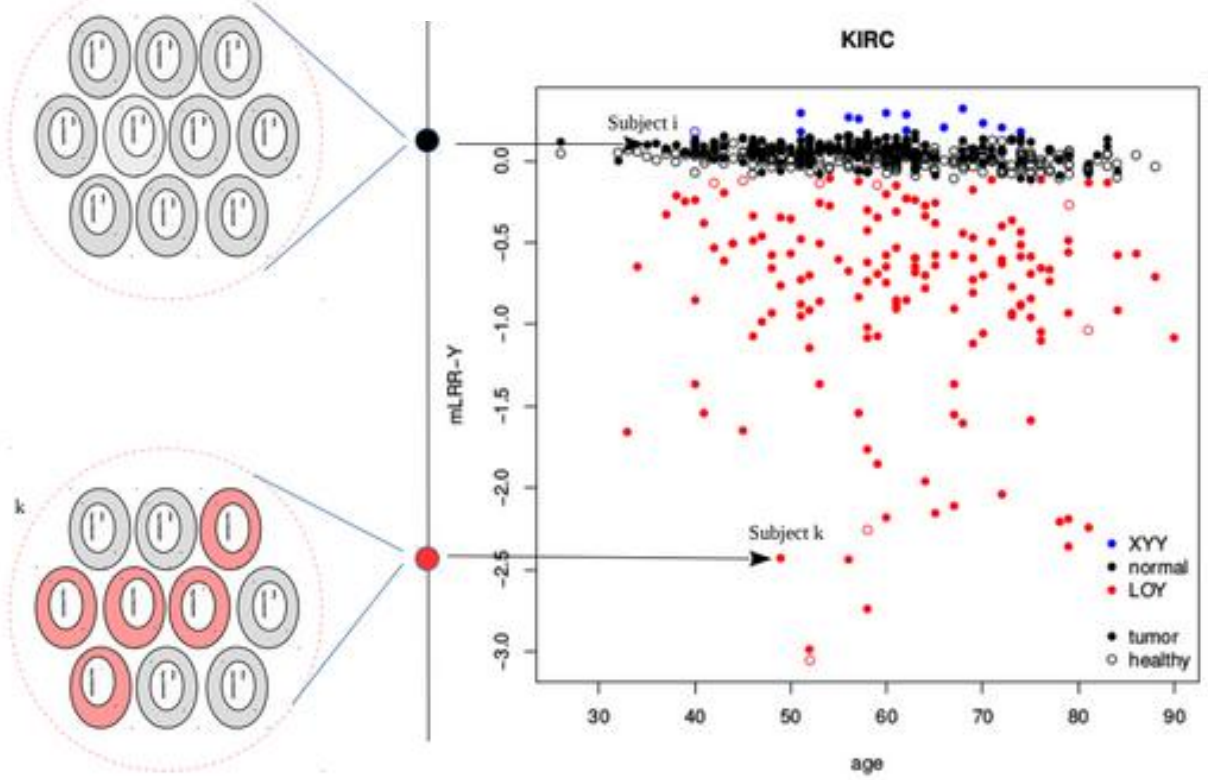
# Loss of Chromosome Y

**MADloy:** https://github.com/isglobal-brge/MADloy

Relative amount of chr Y
DNA in the tissue

Subject i

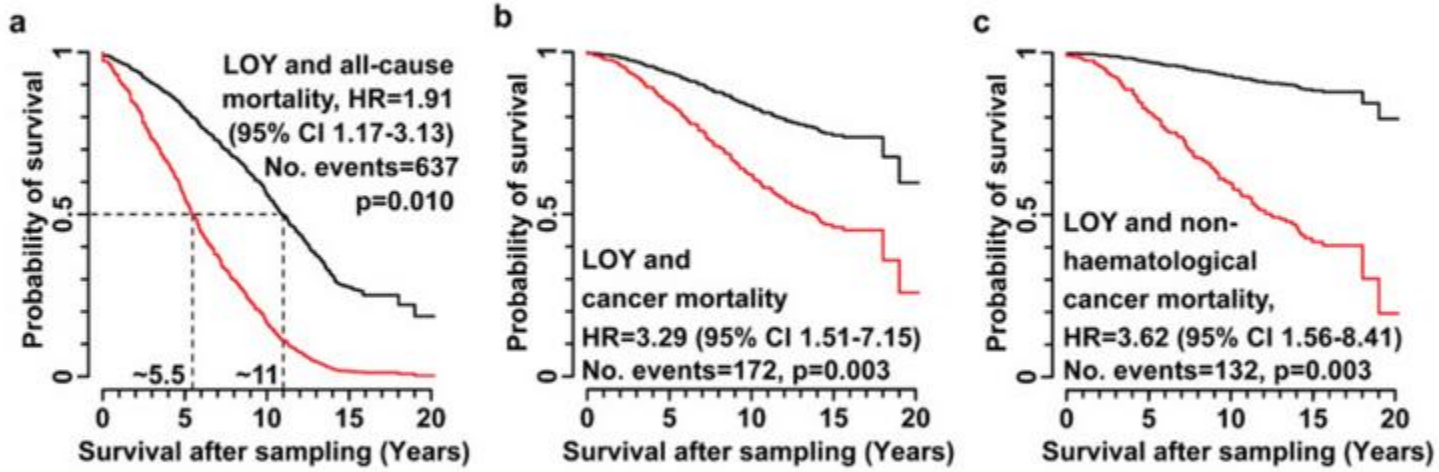
Subject k




González JR et al, bioRxiv, 2018

# Loss of Chromosome Y

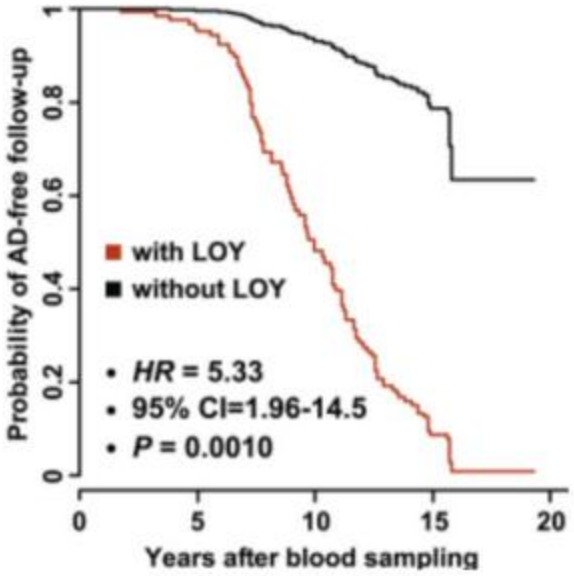
- Associated to all cause and cancer mortalities




*Forsberg, 2014*

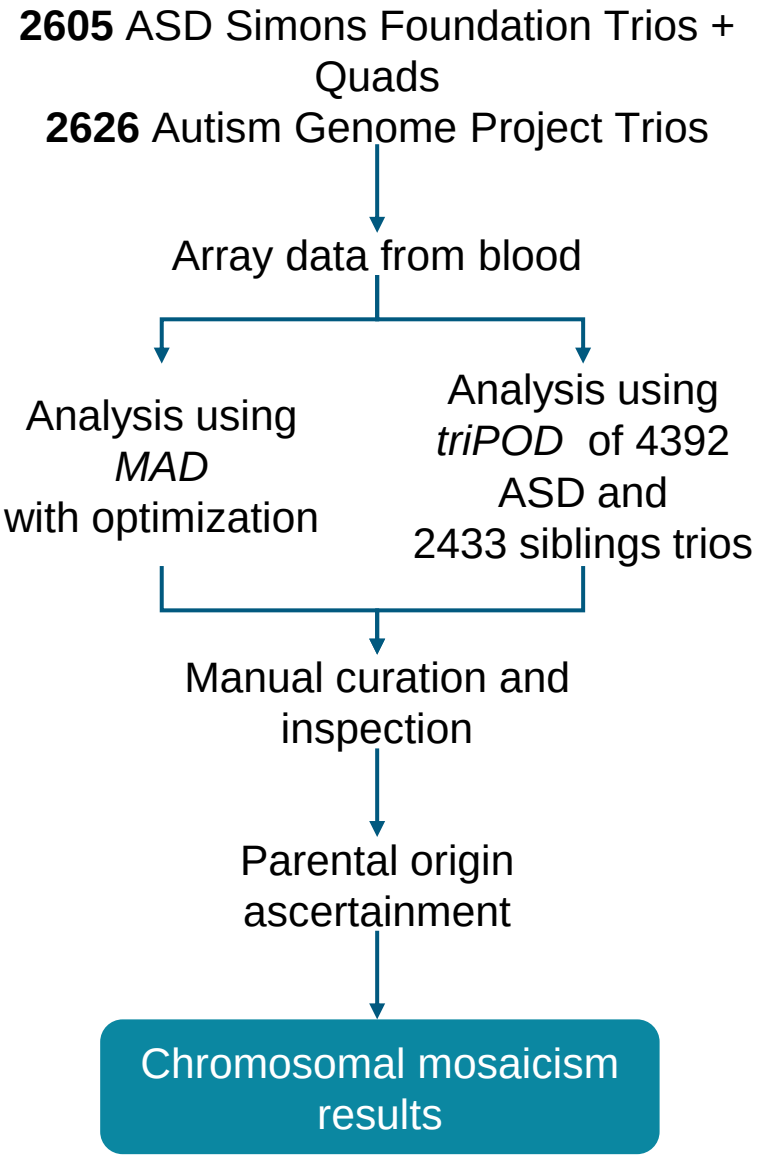
# Loss of Chromosome Y

- Associated with higher risk of developing Alzheimer's disease

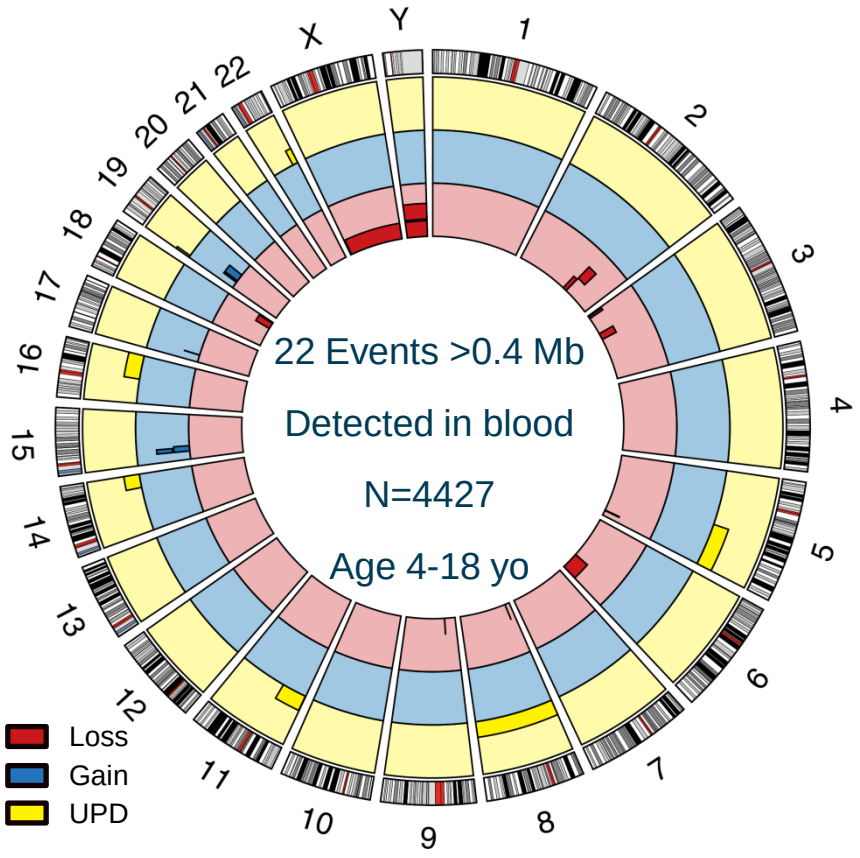


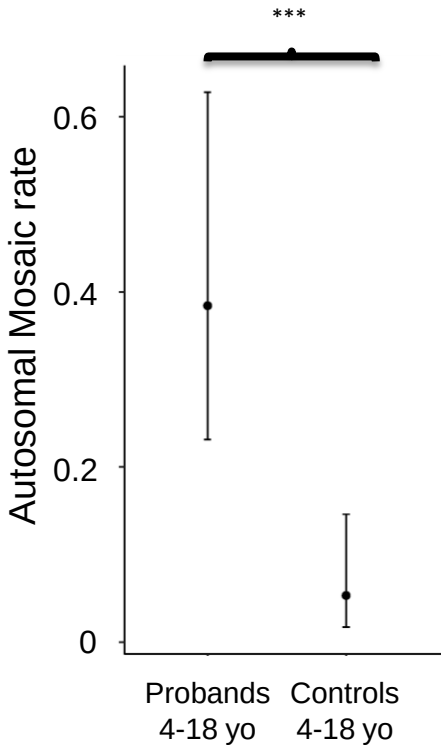

*Dumanski, 2016*

# Autism spectrum disorder



# Autism spectrum disorder




0.43% rate (autosomes)

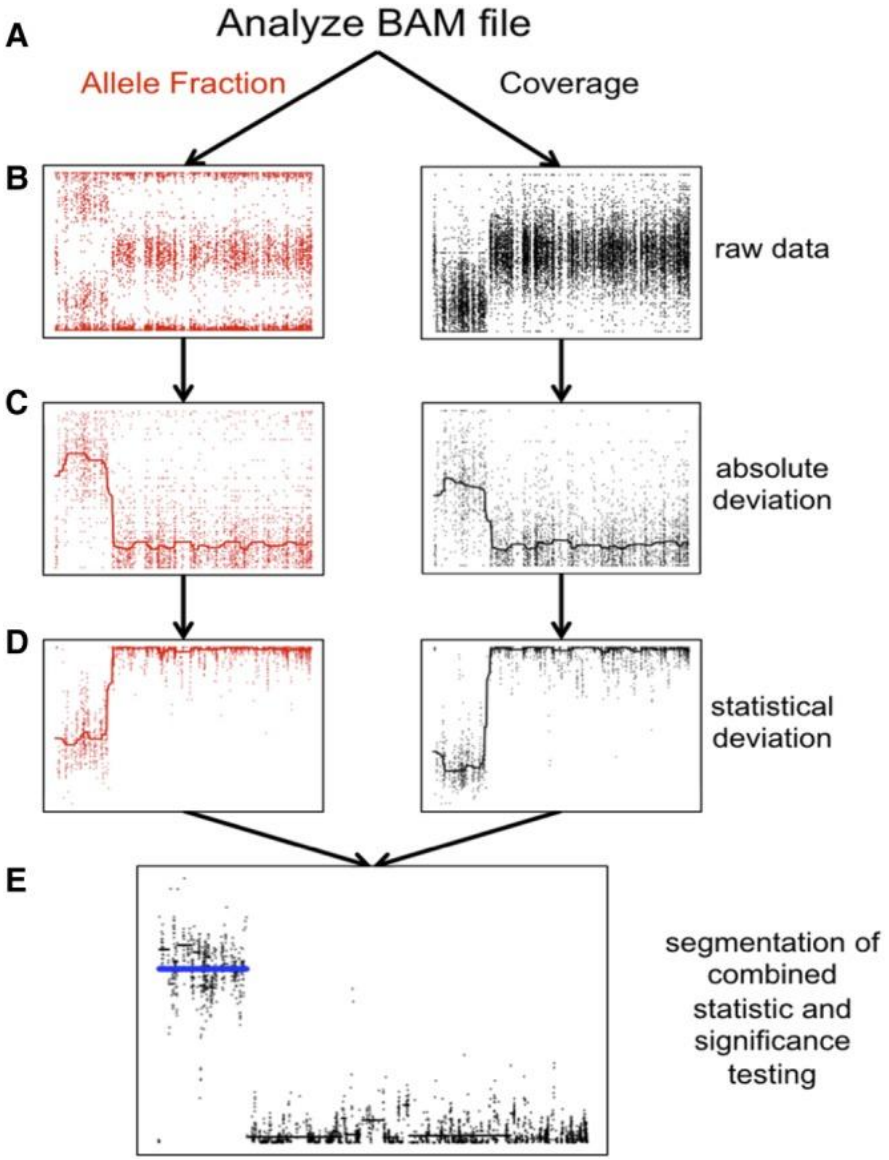
0.07% rate (gonosomes)

79.16% events not previously detected

- Significantly higher mosaicism rate in ASD patients than controls in the same age bin (OR = 8.11, $p=9.798 \cdot 10^{-6}$).

# Structural mosaic events

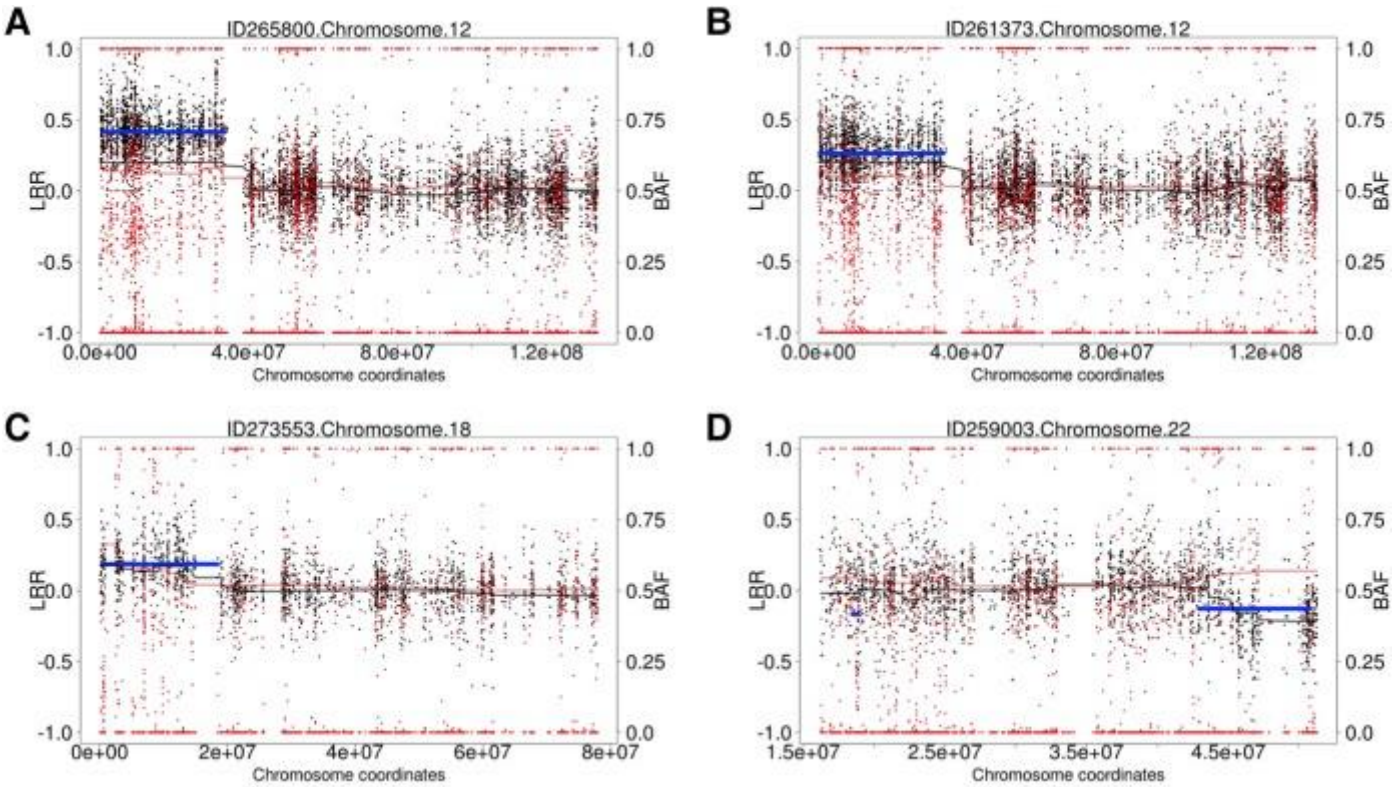
MrMosaic




King, et al. Genome Res. 2017

# Structural mosaic events

4911 patients with undiagnosed developmental disorders, and 11 events among nine patients were detected.




King, et al. Genome Res. 2017

# Structural mosaic events

| | Exome detections | | | | | | | | SNP validation | |
|---|---|---|---|---|---|---|---|---|---|---|
| Decipher ID | Chromosome | Type | Start (GRCh37) | End (GRCh37) | $B_{dev}$ | $\log_2$ ratio | Tissue | Clonality | Clonality saliva | Clonality blood |
| 265800 | 12 | Gain | 988,894 | 33,535,510 | 0.201 | 0.140 | Saliva | 1.34 | 0.68[a] | Absent |
| 261373 | 12 | Gain | 283,642 | 33,535,289 | 0.131 | 0.262 | Saliva | 0.72 | 0.45[a] | Absent |
| 273553 | 18 | Gain | 670,541 | 18,534,702 | 0.186 | 0.185 | Saliva | 1.18 | 0.6[a] | Absent |
| 259003 | 22 | Loss | 42,912,136 | 50,717,129 | 0.131 | −0.129 | Blood | 0.42 | 0.54 | 0.34 |
| 274013 | 10 | Loss | 121,717,932 | 134,916,366 | 0.159 | −0.324 | Saliva | 0.48 | 0.44 | Absent |
| 274600 | 18 | Loss | 48,458,662 | 76,870,586 | 0.190 | −0.434 | Saliva | 0.55 | 0.49 | Absent |
| 260462 | 18 | Loss | 662,103 | 2,740,714 | 0.171 | −0.339 | Saliva | 0.51 | 0.46 | Absent |
| 260462[b] | 18 | Gain | 12,702,610 | 15,323,214 | 0.118 | 0.263 | Saliva | 0.41 | 0.5 | Absent |
| 260462 | 18 | Loss | 48,466,843 | 74,962,645 | 0.153 | −0.345 | Saliva | 0.47 | 0.45 | Absent |
| 257978 | 5 | LOH | 146,077,526 | 179,731,635 | 0.167 | −0.002 | Blood | 0.33 | 0.24 | 0.26 |
| 274396 | 11 | LOH | 66,834,252 | 134,126,612 | 0.255 | −0.0047 | Saliva | 0.51 | 0.28 | 0.17 |

| Decipher ID | Phenotypes |
|---|---|
| 257978 | Intellectual disability profound, seizures, somnolence, thoracolumbar scoliosis, gastroesophageal reflux, abnormality of neuronal migration |
| 259003 | Generalized hypotonia, global developmental delay |
| 260462 | Microcephaly, muscular hypotonia, short philtrum, upslanted palpebral fissure |
| 261373 | Moderate global developmental delay |
| 265800 | Global developmental delay, meningocele, delayed closure of the anterior fontanelle, macroglossia, sparse scalp hair, ligamentous laxity, delayed speech and language development, coarse facial features |
| 273553 | Global developmental delay, joint laxity, hypermetropia, strabismus |
| 274013 | Severe expressive language delay, global developmental delay, abnormal facial shape, brachydactyly syndrome, thick hair, coarse facial features, abnormality of facial musculature, joint stiffness |
| 274396 | Congenital hypothyroidism, congenital microcephaly, moderately short stature, mild global developmental delay, premature anterior fontanel closure, fine hair, sparse scalp hair, long palpebral fissure, wide mouth, short broad hands, excessive wrinkling of palmar skin, excessive skin wrinkling on dorsum of hands and fingers, strabismus, generalized hypopigmentation of hair, progressive hyperpigmentation, mixed hypo- and hyperpigmentation of the skin, axillary and groin hyperpigmentation and hypopigmentation |
| 274600 | Microcephaly, progressive microcephaly, severe global developmental delay, abnormal posturing, brachycephaly, epicanthus, muscular hypotonia, narrow palate, hypotelorism, broad distal phalanx of finger |

King, et al. Genome Res. 2017

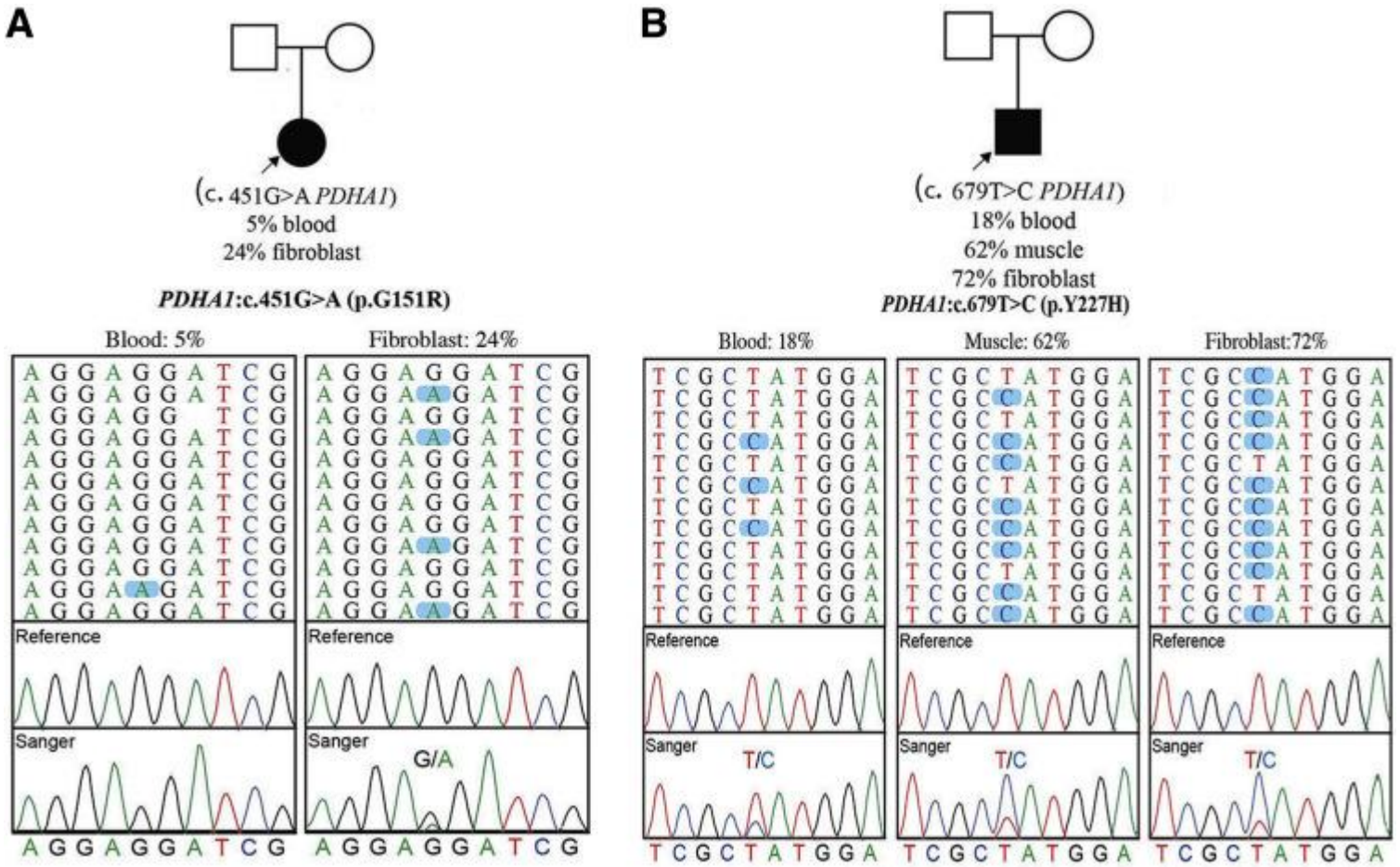
# Mosaic point mutations

- Acuna-Hidalgo, et al. AJHG, 2015 found that ~6.5% of presumed germline ***de novo*** **mutations** were present as **mosaic mutations** in the blood of the offspring and were therefore likely to have occurred post-zygotically.

# Mosaic point mutations

- Postzygotic mosaic mutations (PPM) analysis of ASD (Simon Simplex Collection)
- 470 PMMs detected in children
- Increase up to 22% SNVs mosaic detected
- The authors estimate that PPMs may contribute to 3-4% of simplex ASD case.

# Mosaic point mutations



# Mosaic point mutations

nature methods

BRIEF COMMUNICATION

https://doi.org/10.1038/s41592-018-0051-x

**Strelka2: fast and accurate calling of germline and somatic variants**

Kim et al. Nat Meth, 2018

https://github.com/Illumina/strelka

# Mosaic point mutations

- 251 samples belonging to URDCat
- Nimblegen_SeqCapEZExome_v3_64Mb
- Analysis using adhoc pipeline

# Mosaic point mutations

| Filter Description | # Variants (in total) |
|---|---|
| **Initial** | 1.200.0000 |
| **Software Developer** | 150.000 |
| **QC filters from Strelka** | 44.000 |
| **GATK MQ** | 41.000 |
| **SnpCluster & repetitive regions** | 37.352 |
| **Unique variants (1-2 individuals)** | 717 |
| **Depth coverage >= 30** | 578 |
| **Coverage ALT > 3** | 576 |
| **Coverage Forward & Coverage Reverse >= 1 (avoid strand bias)** | 565 |

# Mosaic point mutations

| Filter Description | # Variants |
|---|---|
| **DPForward & DPReverse >= 1** | 565 |
| **AAF >= 0.05 & <= 0.15 \| >= 0.85 & <= 0.95** | **7** |
| **Population (ExAC, gnomAD) <0.002** | 3 |
| **Segmental duplication** | 3 |
| **Known clinical significance (CLINSIG)** | 3 |
| **pLI >0.9** | 2 |

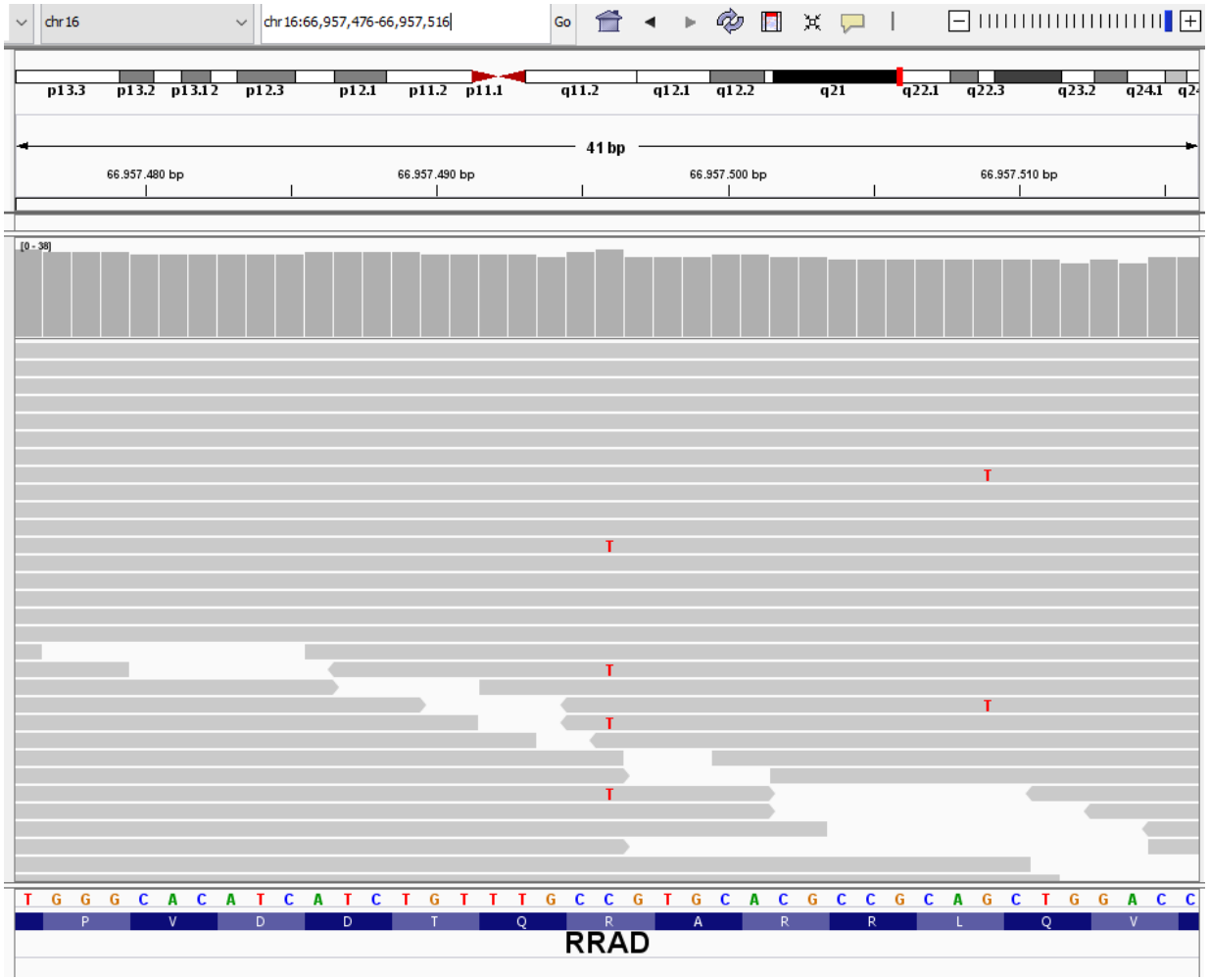
# Mosaic point mutations

False positive !!!

| FAMILIA | Relación | PHENOTIPS ID | CENTRE | RDCat ID | Mutación | PATOLOGIA |
|---|---|---|---|---|---|---|
| FAM0000238 | **Probando** | **P0000400** | VHIR | EPR762118 | Germinal | Progresiva |
| FAM0000238 | **Padre** | **P0001037** | VHIR | EPR943482 | Germinal | No afecto |
| FAM0000238 | **Madre** | **P0001038** | VHIR | EPR798030 | | No afecto |
| FAM0000307 | **Probando** | **P0000474** | IDIBELL | EPR236242 | De novo ? | Progresiva |
| FAM0000307 | **Padre** | **P0000806** | IDIBELL | EPR292176 | | No afecto |
| FAM0000331 | **Probando** | **P0000354** | IDIBELL | EPR211571 | De novo ? | Progresiva |
| FAM0000331 | **Madre** | **P0000814** | IDIBELL | EPR000148 | De novo ? | Afecto |
| FAM0000472 | **Probando** | **P0000786** | IDIBELL | EPR606291 | De novo | Neuromuscular |
| FAM0000472 | **Padre** | **P0000863** | IDIBELL | EPR255030 | | Afecto |
| FAM0000472 | **Madre** | **P0000863** | IDIBELL | EPR308956 | | Afecto |
| FAM0000664 | **Probando** | **P0001175** | VHIR | EPR228697 | De novo | no afecto |
| FAM0000664 | **Padre** | **P0001174** | VHIR | EPR105166 | | Epilepsia y TPNE |
| FAM0000664 | **Madre** | **P0001176** | VHIR | EPR802333 | | no afecto |
| FAM0000676 | **Probando** | **P0001212** | VHIR | EPR290453 | De novo | no afecto |
| FAM0000676 | **Padre** | **P0001210** | VHIR | EPR416046 | | Epilepsia y TPNE |
| FAM0000676 | **Madre** | **P0001211** | VHIR | EPR846345 | | no afecto |
| FAM0000683 | **Probando** | **P0001232** | VHIR | EPR296368 | De novo | no afecto |
| FAM0000683 | **Padre** | **P0001230** | **VHIR** | EPR334036 | | Metabólica Hereditaria |
| FAM0000683 | **Madre** | **P0001231** | VHIR | EPR909792 | | No afecto |

# Mosaic point mutations

| | Chrom | Pos | REF | ALT | Coverage | REF cov | ALT cov | AAF |
|---|---|---|---|---|---|---|---|---|
| EPR606291 | 16 | 66957496 | C | T | 33 | 27 | 5 | 0.15 |



# Software

- **MAD:** https://github.com/isglobal-brge/MAD/blob/master/vignettes/MAD.html
- **MADloy:** https://github.com/isglobal-brge/MADloy/
- **mrMosaic:** https://github.com/asifrim/mrmosaic
- **Strelka2:** https://github.com/Illumina/strelka
- **MADseq:** https://bioconductor.org/packages/release/bioc/html/MADSEQ.html
- URDCat pipeline: juanr.gonzalez@isglobal.org

# Detección de posibles mosaicismos estructurales y puntuales

**Curso Medicina Genómica Personalizada**

Barcelona, 5-7 Noviembre 2018


**Juan R Gonzalez**

Bioinformatics Research Group in Epidemiology (BRGE), ISGlobal

www.brge.isglobal.org

**email:** juanr.gonzalez@isglobal.org