

Next Generation Sequencing

Mapping, variant calling, identification de novo & inherited variants,
CNVs

Christian Gilissen PhD

christian.gilissen@radboudumc.nl

12-12-2022

Contents

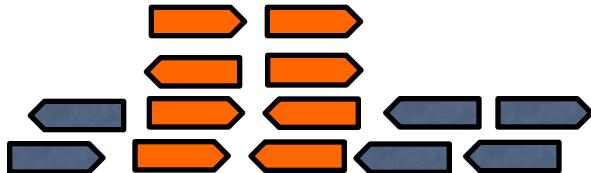
- CNV analysis in exomes
- De novo mutations
- Clinical exome sequencing
- Whole genome sequencing
- *Optional: Long read sequencing*

Copy Number Variants (CNVs) in WES



Rolph Pfundt

Structural variant calling for NGS



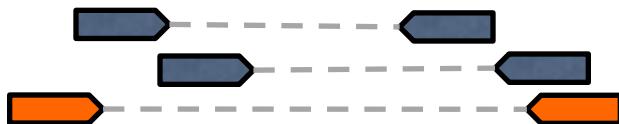
Read depth

CNVs > 500bp



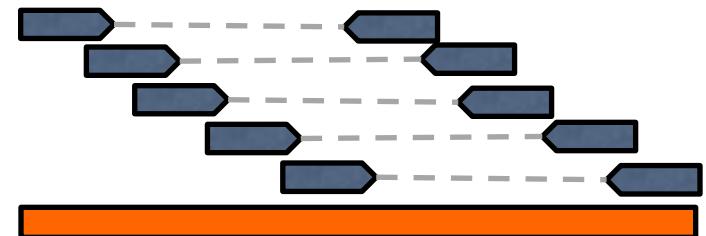
Split Read mapping

CNVs < 50bp



Discordant pairs

CNVs > 500bp



De novo assembly

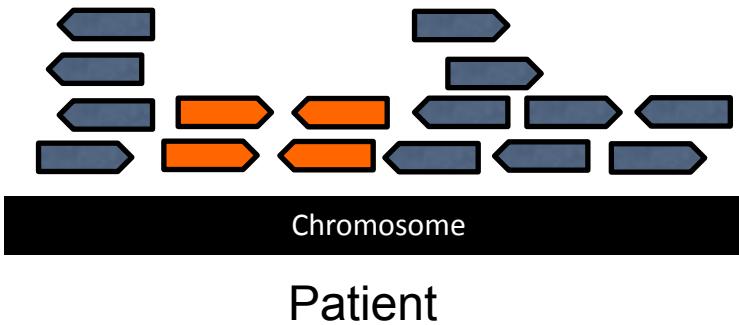
All CNVs

Exome sequencing

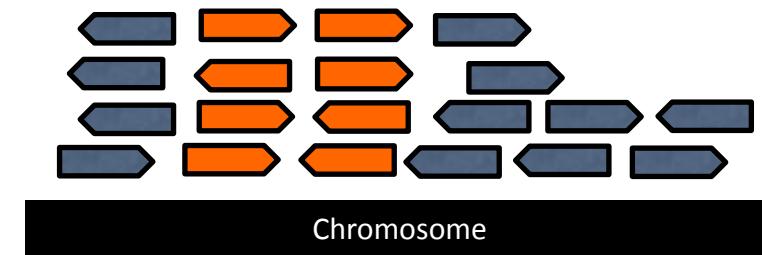
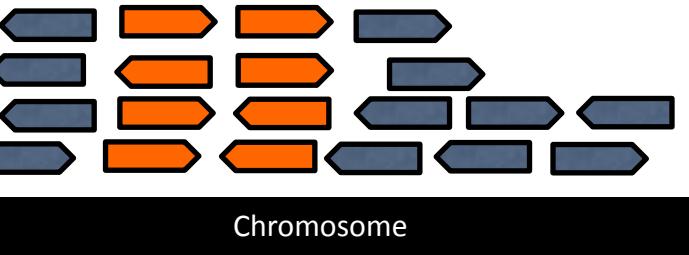
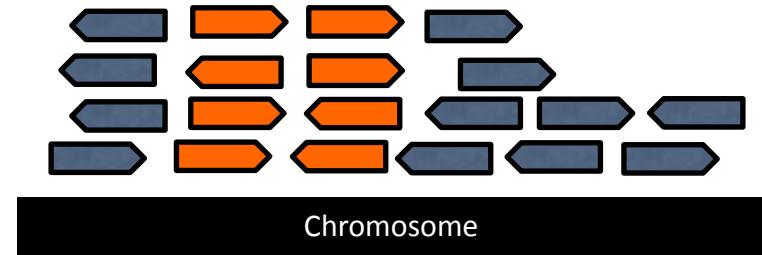
- **Best option: depth-of-coverage**
- Because:
 - Split reads relies on reads covering the break-points and these will likely be outside of the exome targets
 - Discordant reads, unlikely that your two pairs are outside of the CNV event, because you only sequence exons
- **Difficulties depth of coverage:**
 - Enrichment biases make calling on targeted assays difficult (correct for GC content)
 - Naturally duplicated regions in the genome
 - Mapping algorithms only consider a haploid genome
 - Imperfect reference genome
- Easy to get false positives!
- Multiple exons needed to identify a CNV

How to detect CNVs?

Detection by inference from
read-depth differences



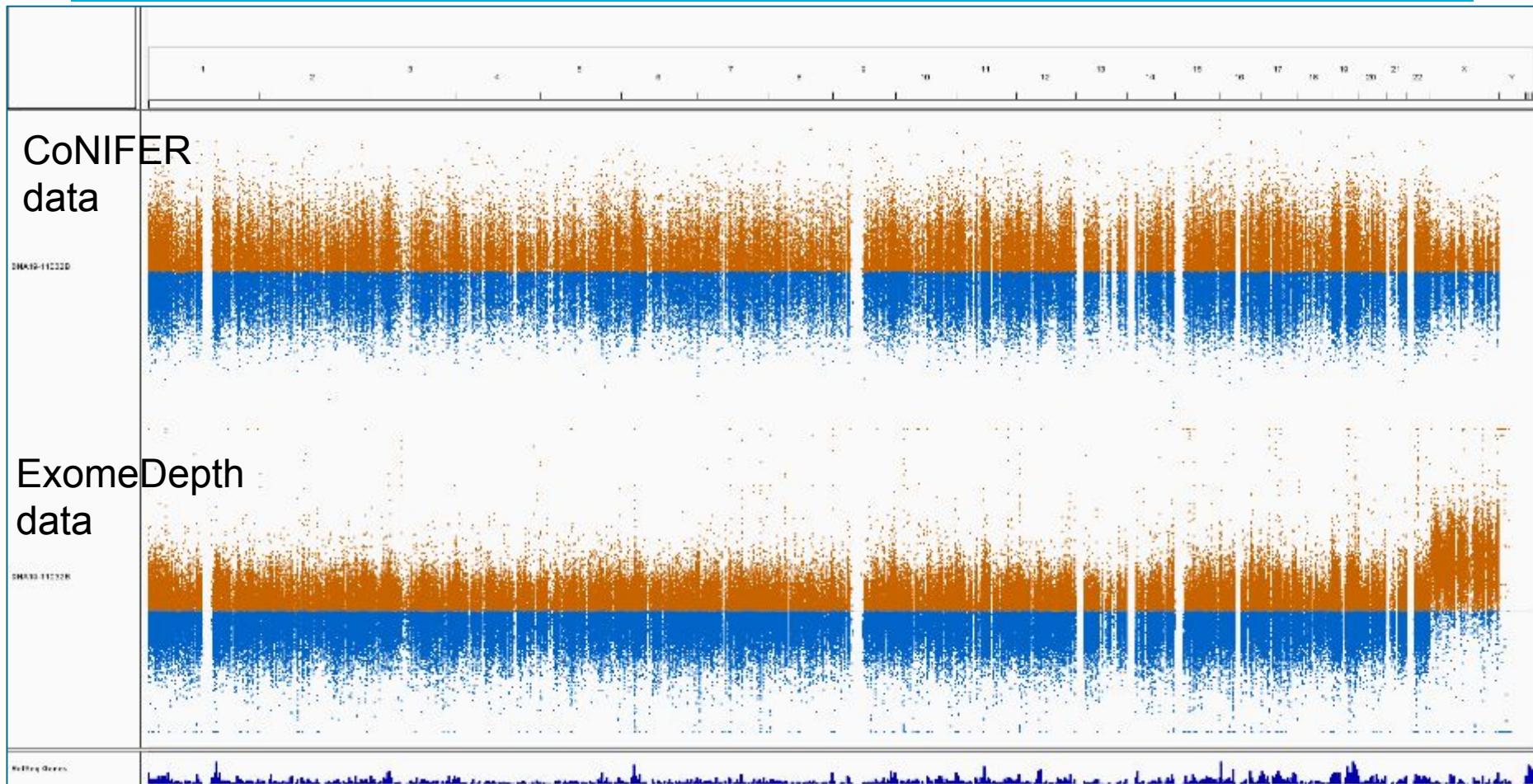
Healthy controls



CNV/SV tools

- CNV calling methods for exomes all rely on depth of coverage.
- They differ in their normalization methods, and calling method.
 - **Conifer: depth of coverage, detects rare CNVs**
 - Excavator: Hidden Markov model, detecting 5 CNV states
 - XHMM: PCA normalization, HMM calling
 - ExomeCNV: Event wise testing
 - Convex: direct wavelet transform and HMM calling
 - **ExomeDepth: single exons**

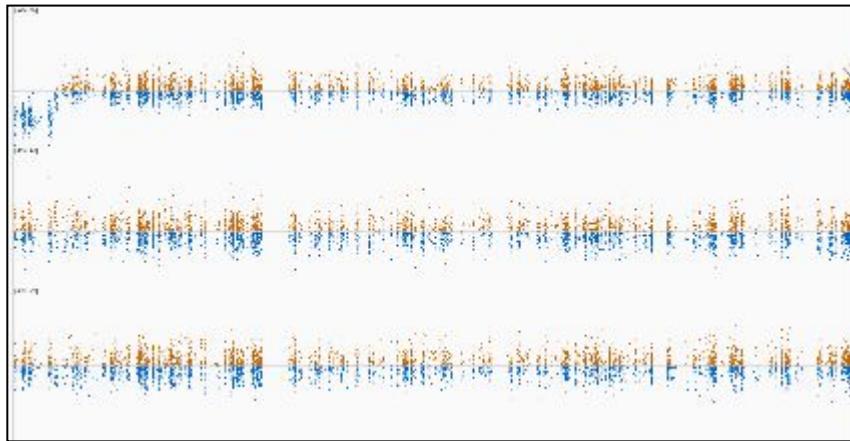
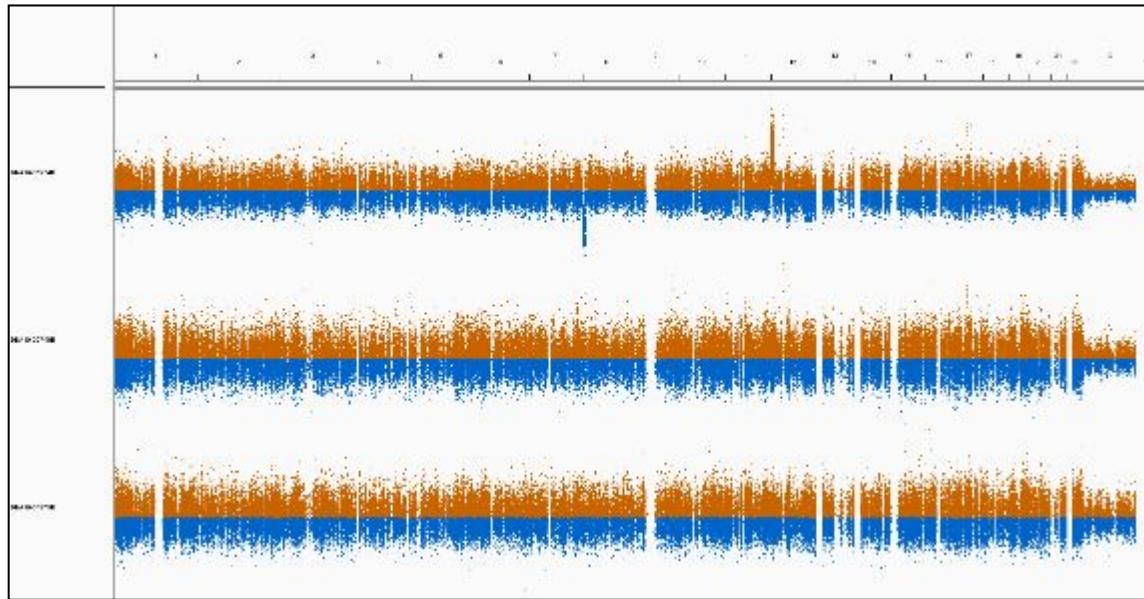
Triple-X; 47,XXX casus visible in ExomeDepth but not in CoNIFER



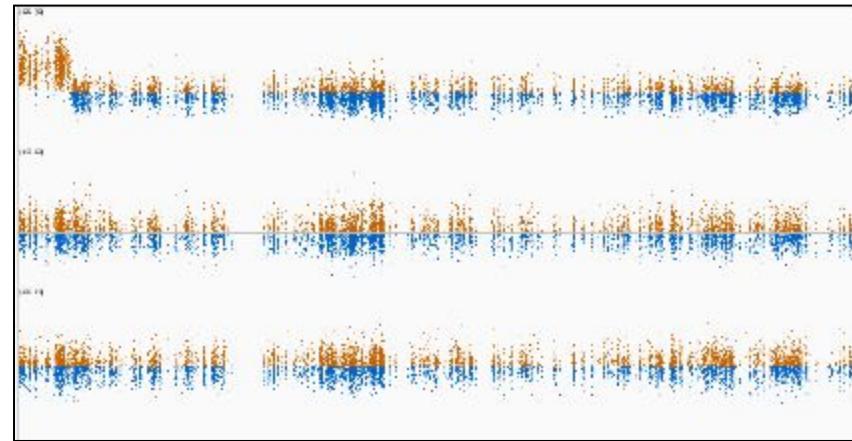
- Each dot represents the (normalized) coverage of a single exome target (i.e. an exon)
- Normalisation is done by comparing the coverage to other samples

Unbalanced translocation t(8;12)

Whole genome view



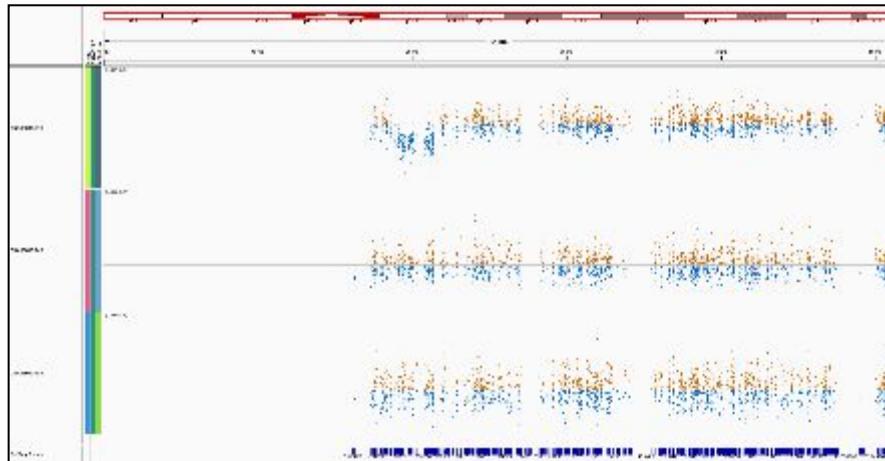
Chromosome 8



Chromosome 12

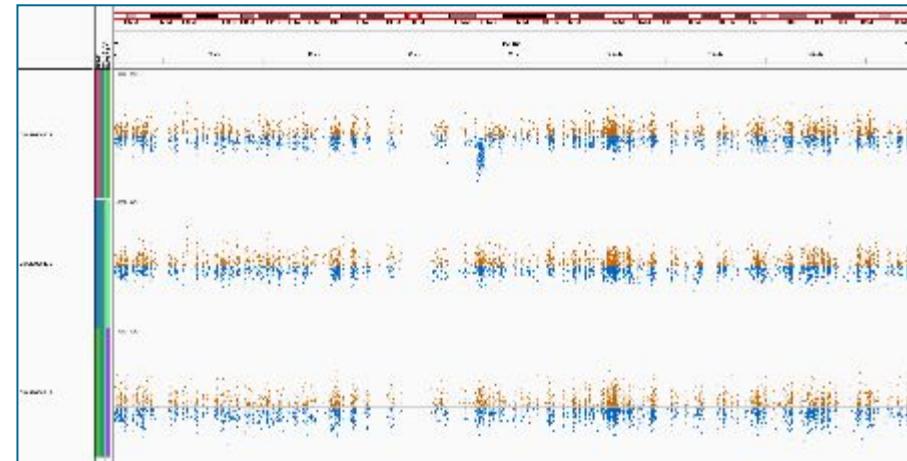
Large CNV's

3 Mb 22q11 deletion



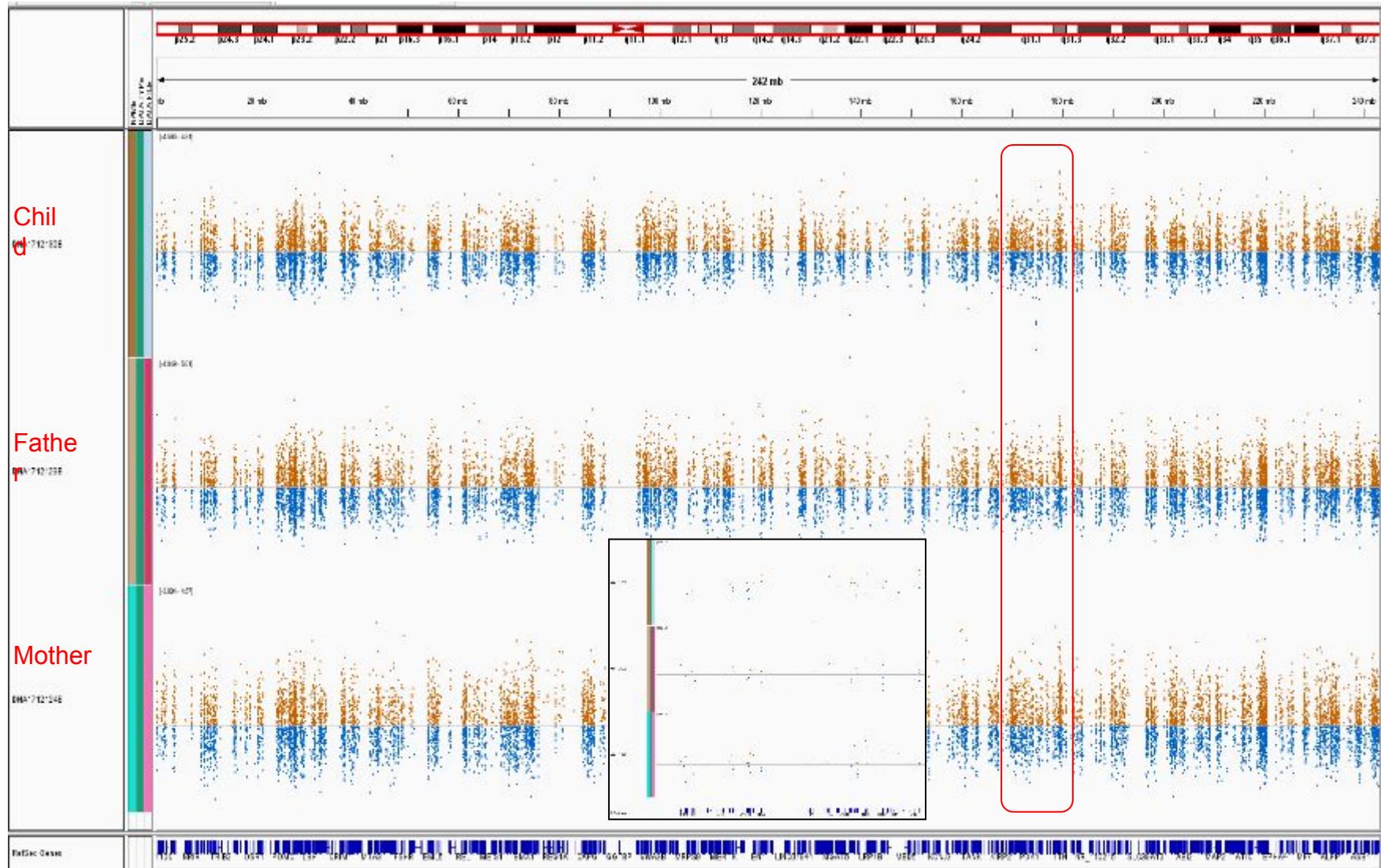
DiGeorge syndrome

~1.5 Mb 7q11.23 deletion



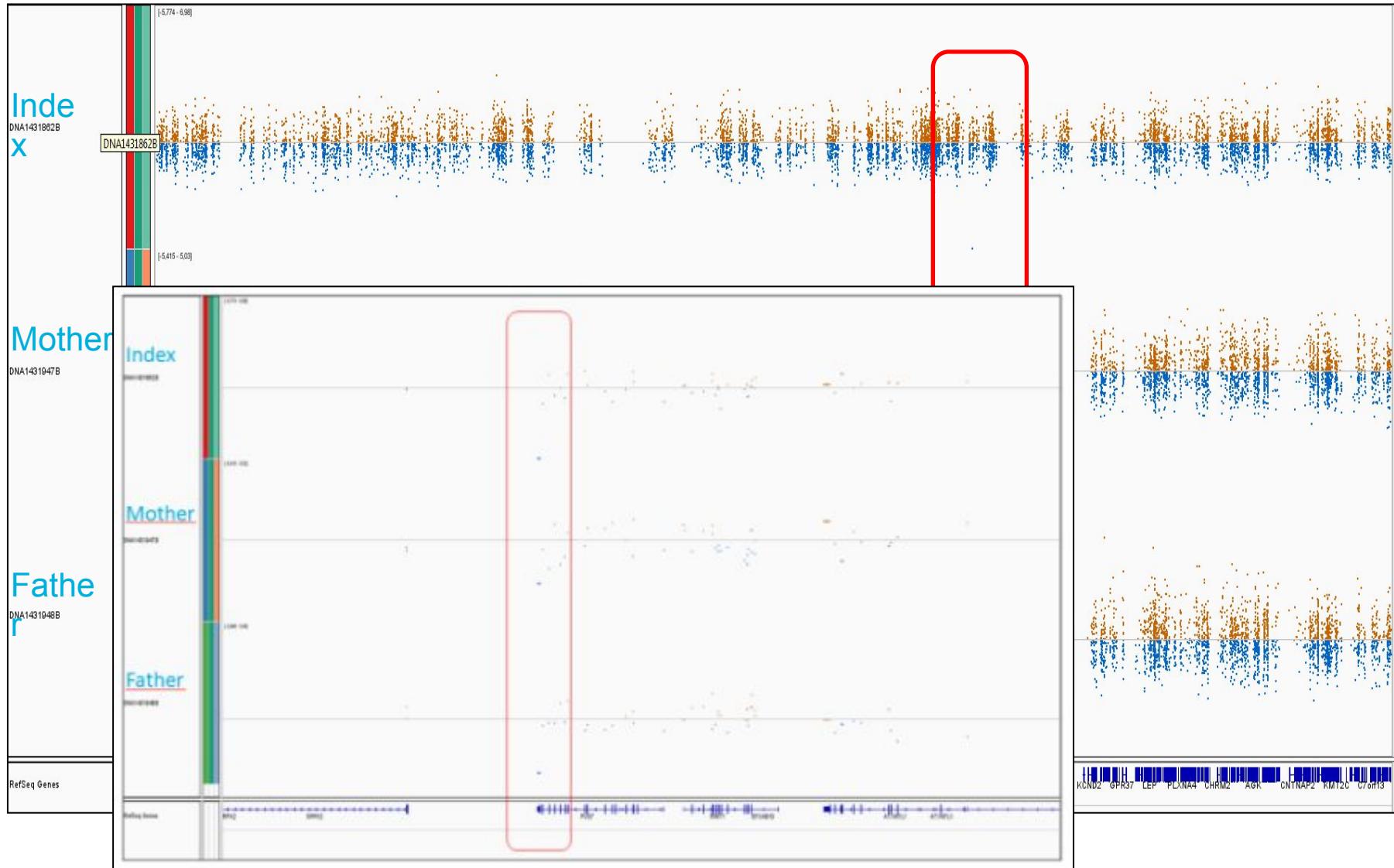
Williams-Beuren syndrome

Small CNVs (>3 exons)



Profile of CNV from WES algorithm of chromosome 2

Single exon deletions



Single exon CNVs are visible in the data
but not called by CoNIFER

Heterodisomy 15

(F) - ED, CNV, ROH & SNP visualization

ED homozygous deletions - N/A
Conifer CNV

chr13

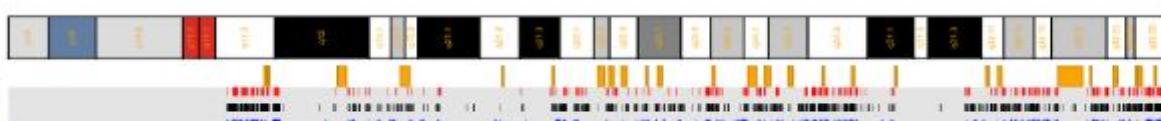
ROH calls
MV
Inheritance PV MV
PV



ED homozygous deletions - N/A
Conifer CNV

chr14

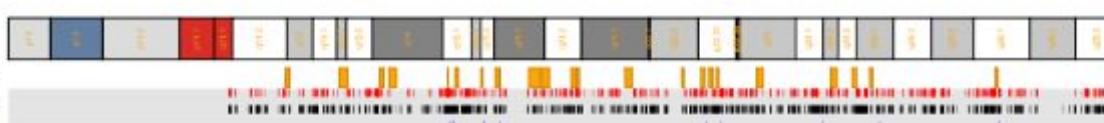
ROH calls
MV
Inheritance PV MV
PV



ED homozygous deletions - N/A
Conifer CNV

chr15

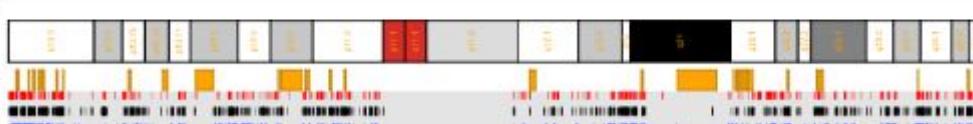
ROH calls
MV
Inheritance PV MV
PV



ED homozygous deletions - N/A
Conifer CNV

chr16

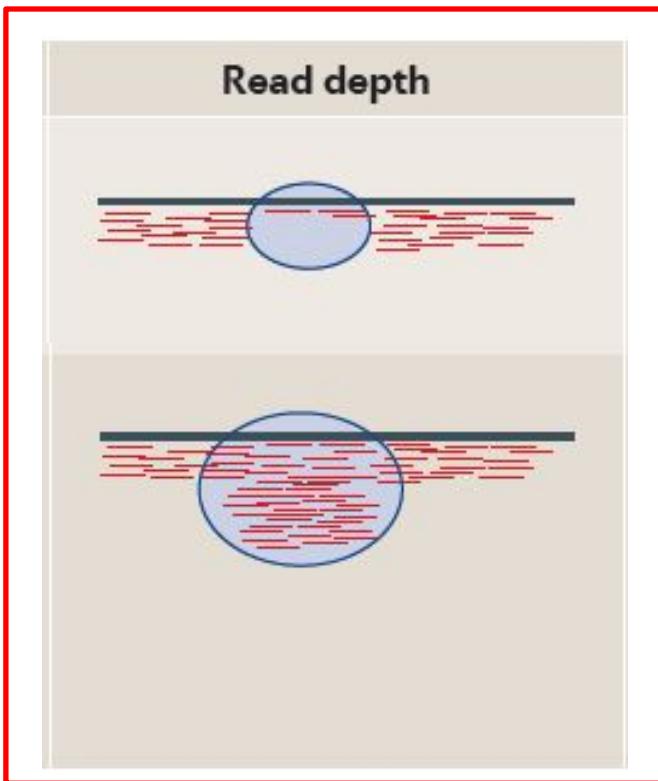
ROH calls
MV
Inheritance PV MV
PV



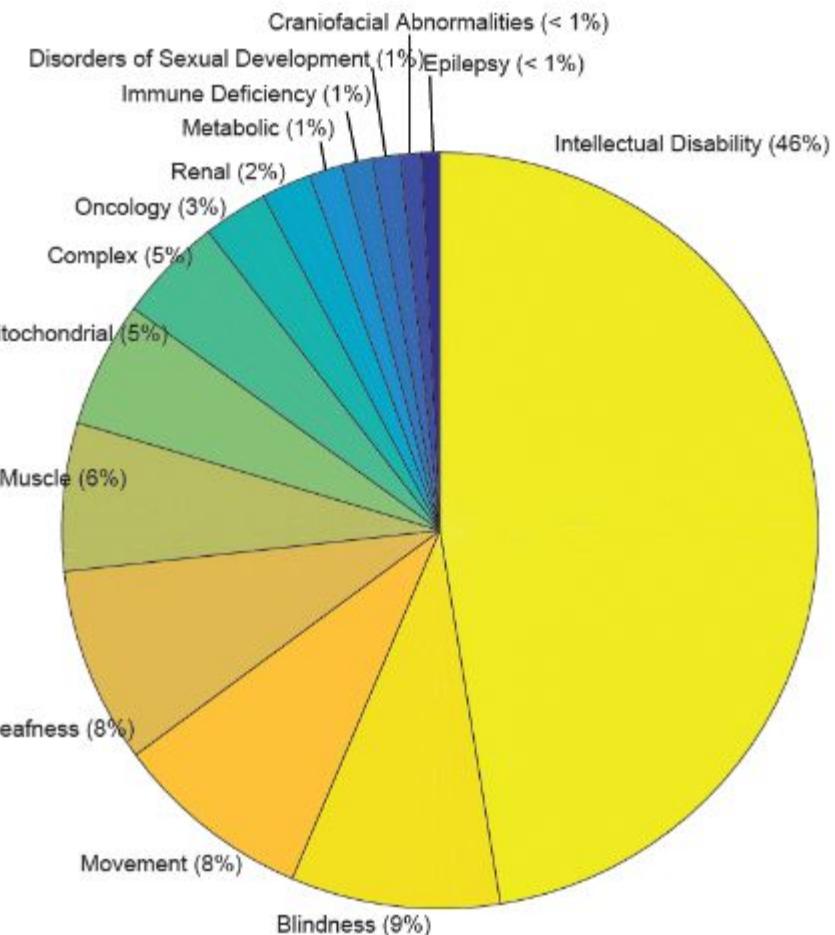
How well does it work overall?

conifer

COPY NUMBER INFERENCE FROM EXOME READS



Copy number variation detection and genotyping from exome sequence data. Krumm et al. Genome Res. 2012; 22:1525-1532



In total 2,645 clinical exomes

Results

	Number of patients	CNVs overlapping		Diagnostic Yield	
		Autosomal Dominant Genes	Autosomal Recessive Genes with 2nd hit	nr CNVs	% of cohort
Epilepsy	21	2	0	2	9.5%
Craniofacial anomalies	31	0	0	0	0.0%
Disorders of sexual development	38	0	0	0	0.0%
Immune Deficiency	24	0	0	0	0.0%
Metabolic disorders	34	0	0	0	0.0%
Hereditary Cancer ¹	74	0	0	0	0.0%
Renal disorders	56	1	1	2	3.6%
Complex phenotypes ²	183	8	1	9	4.9%
Mitochondrial disorders	142	0	1	1	0.7%
Muscle disorders	171	1	0	1	0.6%
Deafness	223	0	10	10	4.5%
Movement disorders	217	2	0	2	0.9%
Blindness	237	2	2	4	1.7%
Intellectual Disability	1194	22	0	22	1.8%

On average a 2% increase in diagnostic yield

Recycle your exomes!

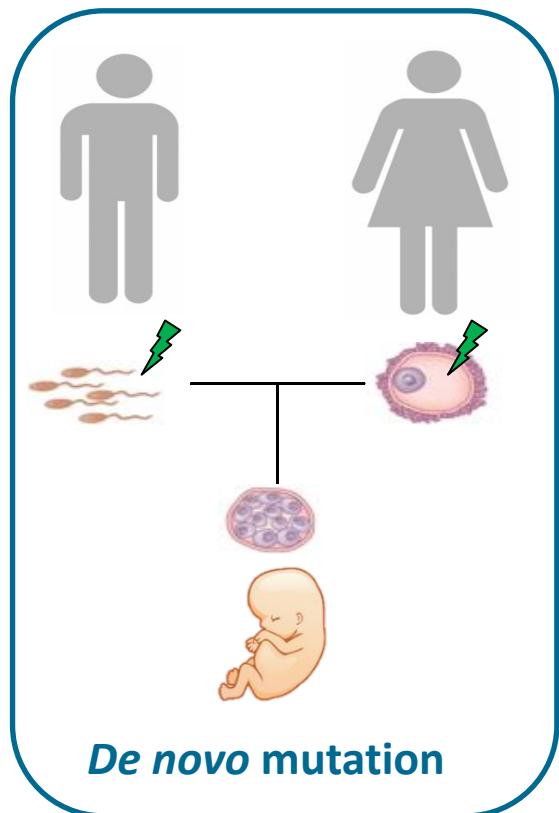
Data from exomes can be reused to identify other types of variants:

- Copy number variation calling (CNV) – *Pfundt et al. Genet med. 2017*
- Regions of homozygosity (ROH) – *Magi et al. Bioinformatics 2014*
- Mobile Element Insertions (MEI) – *Torene et al. Genet Med. 2019, Wijngaard et al. in prep.*
- Nucleotide repeat expansions – *Dolzhenko et al. Bioinformatics. 2019, van der Sanden et al. Genet. Med. 2021*
- Uniparental disomy (UPD) – *Yauy et al. Genet. Med. 2019*
- Mitochondrial DNA (mtDNA) – *Griffin et al. Genet med. 2014*
- Identify mosaic mutations – *Acuna-Hidalgo et al. AJHG 2015*
- Variation in homologous regions – *Steyaert et al. in prep.*

De novo mutations

De novo mutations

- Sporadic DNA change in the sperm or oocyte of the parents that give rise to a new (de novo) mutations in the offspring

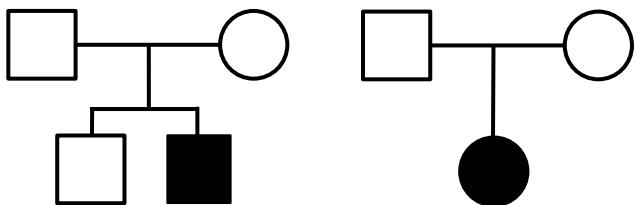


- On average, **~74** *de novo* single nucleotide mutations genome-wide (ranging from 30-100 DNM)*
- Of which 0-5 (avg. ~1.5) in the coding regions
- Changes are mostly single base substitutions but can also be indels and larger events

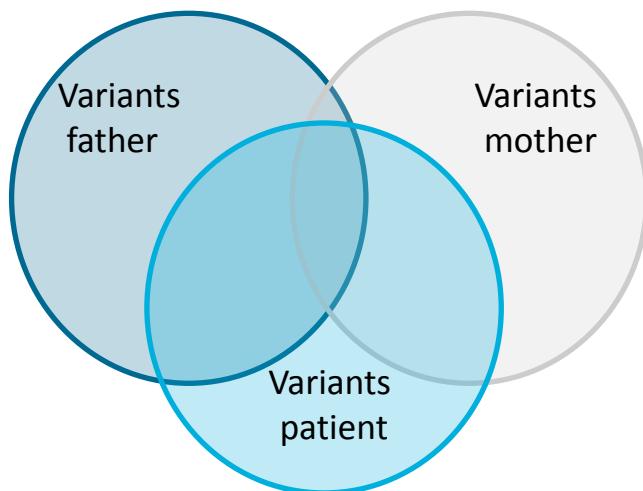
*Vissers, Gilissen & Veltman, Nat Rev Gen. 2016 • *Goldmann, Veltman & Gilissen, Trends Genet. 2019

~75-80% occurs on the paternal allele

De novo mutations?



- 100 patient-parent trios
- Severe intellectual disability ($\text{IQ} < 50$)
- No etiological or syndromic diagnosis
- Negative family history
- Normal karyotype & CNV profile



Yield in 100 ID patients

Positive diagnosis

June 2013

All mutations **29**

De novo mutations **28**

Autosomal dominant **23**

X-linked **4**

Autosomal recessive **1**

Inherited mutations **1**

X-linked **1**

Autosomal recessive **0**

Candidates **11**

Yield of ~30% in patients with severe ID

Current advancement in neurodevelopmental disorders

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability

Joep de Ligt, M.Sc., Marjolein H. Willemsen, M.D., Bregje W.M. van Bon, M.D., Ph.D., Tjitske Kleefstra, M.D., Ph.D., Helger G. Yntema, Ph.D., Thessa Kroes, B.Sc., Anneke T. Vulto-van Silfhout, M.D., David A. Koelen, M.D., Ph.D., Petra de Vries, B.Sc., Christian Gilissen, Ph.D., Marisol del Rosario, B.Sc., Alexander Hoischen, Ph.D., Hans Scheffer, Ph.D., Bert B.A. de Vries, M.D., Ph.D., Han G. Brunner, M.D., Ph.D., Joris A. Veltman, Ph.D., and Lisenka E.L.M. Vissers, Ph.D.

Articles

Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study

Anita Rausch¹, Dagmar Wieszorek², Elisabeth Graf¹, Thomas Wieland^{1*}, Sabine Endege, Thomas Schwarzmayr, Beate Albrecht, Deborah Bartholdi, Jasmin Reysel, Nadiajlo Di Donato, Andree Dufke, Kirsten Cremers, Maja Hempel, Denise Horn, Juliane Hoyer, Pascal Jostet, Albrecht Ropke, Ute Moog, Angelika Rieß, Christian T. Thiel, Andreu Tzschach, Antje Wiesener, Eva Wohlfleber, Christiane Zwerter, Arif B. Eki, Alexander M. Zirk, Andreas Rump, Christa Meisinger, Hans-Joachim Graefel, Heinrich Sticht, Annette Schenck, Hartmut Engels, Gudrun Rappold, Evelyn Schreiber, Peter Wiescker, Olaf Rieß, Thomas Mettinger, André Reis, Tim M. Strom¹

LETTER

doi:10.1038/nature12439

De novo mutations in epileptic encephalopathies

Epi4K Consortium* & Epilepsy Phenome/Genome Project*

Neuron
Article

Cell
PRESS

De Novo Gene Disruptions in Children on the Autistic Spectrum

Ivan Iossifov^{1,6}, Michael Ronemus^{1,6}, Dan Levy¹, Zihua Wang¹, Inessa Hakker¹, Julie Rosenbaum¹, Boris Yamrom¹, Yoon-ha Lee¹, Giuseppe Narzisi¹, Anthony Leotta¹, Jude Kendall¹, Ewa Grabowska¹, Beicong Ma¹, Steven Marks¹, Linda Devi¹, Asya Stepansky¹, Jennifer Troge¹, Peter Andrews¹, Mitchell Becker¹, Kith Pradhan¹, Elena Gliban¹, Melissa Kramer¹, Jennifer Parla¹, Ryan Demeter², Lucinda L. Fulton², Robert S. Fulton², Vincent J. Magrini², Kenny Ye³, Jennifer C. Damell⁴, Elaine R. Mardis², Richard K. Wilson², Michael C. Schatz¹, W. Richard McCombie¹, and Michael Wigler^{1*}.

LETTER

doi:10.1038/nature11011

Patterns and rates of exonic *de novo* mutations in autism spectrum disorders

Benjamin M. Neale^{1,2}, Yan Kou^{3,4}, Li Liu⁵, Avi McCayam³, Caitlin E. Samocha^{1,2}, Antiko Sabo⁶, Chiao-Feng Lin⁷, Christine Stevens², Li-San Wang⁸, Vladimir Makarov^{9,10}, Paz Polak^{11,12}, Semiratul Yoon^{1,8,13}, Jennifer Maguire¹⁴, Emily L. Crawford¹⁵, Nicholas G. Campbell¹⁰, Elizabeth M. Vives¹⁶, Venkateswaran Sankararaman¹⁷, Daniel J. MacArthur¹⁸, Daniel C. Geschwind¹⁹, Michael J. Bamshad²⁰, Khalid A. Fakhry^{1,2}, Joseph Glessner²¹, Halon Hakonarson^{1,3,22}, Michael J. Italia²³, Jonathan K. Kline²⁴, Juan Carlos Rodriguez²⁵, Richard Kim²⁶, Jennifer K. Kline²⁶, Teresa Lee²⁷, Jeremy Leipzig²⁸, Alexander P. Pukac²⁹, Michael M. Mangano³⁰, Linda E. Mardis³¹, Jane W. Newell³², Michael J. Bamshad³³, Isik Pe'er²⁴, George Porter²⁴, Amy E. Roberts²⁴, Ravi Sachidanandam²⁷, Stephen J. Sanders^{22,34}, Howard S. Seiden²⁴, Matthew W. State²⁴, Saileshkumar Subramanian²⁴, Irina R. Tikhonova^{1,6,7}, Wei Wang^{32,33}, Dorothy Warburton^{4,20}, Peter S. White^{14,15}, Ismee A. Williams⁴, Hongyu Zhao^{22,27}, Jonathan G. Seidman², Martina Buccineider²⁸, Wendy K. Chung^{1,29}, Bruce D. Gelb^{22,24,30}, Elizabeth Goldmuntz^{14,31}, Christine E. Seidman^{3,5,32} & Richard P. Lifton^{1,2,6,7,33}.

LETTER

De novo mutations in histone-modifying genes in congenital heart disease

Samir Zaidi^{1,2*}, Murim Choi^{1,2*}, Hiroko Wakimoto³, Lijiang Ma⁴, Jianming Jiang^{1,5}, John D. Overton^{1,6,7}, Angela Romano-Adesman⁸, Robert D. Björnsson^{7,9}, Roger E. Breitbart¹⁰, Kerry K. Brown¹¹, Nicholas J. Carrasco^{1,3,4}, Yee Him Cheung¹², John DeAngelis¹², Steve DePalma¹², Khalid A. Fakhry^{1,2}, Joseph Glessner²¹, Halon Hakonarson^{1,3,22}, Michael J. Italia²³, Jonathan K. Kline²⁴, Juan Carlos Rodriguez²⁵, Richard Kim²⁶, Jennifer K. Kline²⁶, Teresa Lee²⁷, Jeremy Leipzig²⁸, Alexander P. Pukac²⁹, Michael M. Mangano³⁰, Linda E. Mardis³¹, Jane W. Newell³², Michael J. Bamshad³³, Isik Pe'er²⁴, George Porter²⁴, Amy E. Roberts²⁴, Ravi Sachidanandam²⁷, Stephen J. Sanders^{22,34}, Howard S. Seiden²⁴, Matthew W. State²⁴, Saileshkumar Subramanian²⁴, Irina R. Tikhonova^{1,6,7}, Wei Wang^{32,33}, Dorothy Warburton^{4,20}, Peter S. White^{14,15}, Ismee A. Williams⁴, Hongyu Zhao^{22,27}, Jonathan G. Seidman², Martina Buccineider²⁸, Wendy K. Chung^{1,29}, Bruce D. Gelb^{22,24,30}, Elizabeth Goldmuntz^{14,31}, Christine E. Seidman^{3,5,32} & Richard P. Lifton^{1,2,6,7,33}.

LETTERS

LETTER

doi:10.1038/nature10989

Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations

Brian J. O'Roak¹, Laura Vives¹, Santhosh Girirajan¹, Emre Karakoc¹, Niklas Krumm¹, Bradley P. Coe¹, Roie Levy¹, Arthur Ko¹, Choih Lee¹, Joshua D. Smith¹, Emily H. Turner¹, Ian B. Stanaway¹, Benjamin Vernot¹, Maika Malig¹, Carl Baker¹, Beau Reilly², Joshua M. Akey¹, Elhanan Borenstein^{1,3,4}, Mark J. Rieder¹, Deborah A. Nickerson¹, Raphael Bernier¹, Jay Shendure¹ & Evan E. Eichler^{1,5}.

LETTER

doi:10.1038/nature10945

De novo mutations revealed by whole-exome sequencing are strongly associated with autism

Stephan J. Sanders¹, Michael T. Murtha¹, Abha R. Gupta^{2,8}, John D. Murdoch^{1,9}, Melanie J. Raubeson^{1,8}, A. Jeremy Willsey^{1,8}, A. Gullhan Eranc-Sencicek^{1,8}, Neelroop N. Parkash^{1,9}, Jason L. Stein^{1,9}, Michael F. Walker¹, Gordon T. Ober¹, Nicole A. Teran¹, Youenne Song¹, Paul El-Fishawy¹, Ryan C. Murtha¹, Murim Choi¹, John D. Overton¹, Robert D. Björnsson¹, Nicholas J. Carrasco^{1,3,4}, Kyle A. Meyer¹, Kaya Bilgavur¹, Shirkan M. Mane¹, Nenad Sestan¹, Richard P. Lifton¹, Murat Gunel¹, Kathryn Roeder¹, Daniel H. Geschwind¹, Bernie Devlin¹⁰ & Matthew W. State¹.

De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia

Bin Xu¹, Iuliana Ionita-Laza², J Louw Roos^{3,4}, Braden Boone⁵, Scarlet Woodrick^{1,6}, Yan Sun¹, Shawn Levy⁵, Joseph A. Gogos^{6,7} & Maria Karayiorgou¹

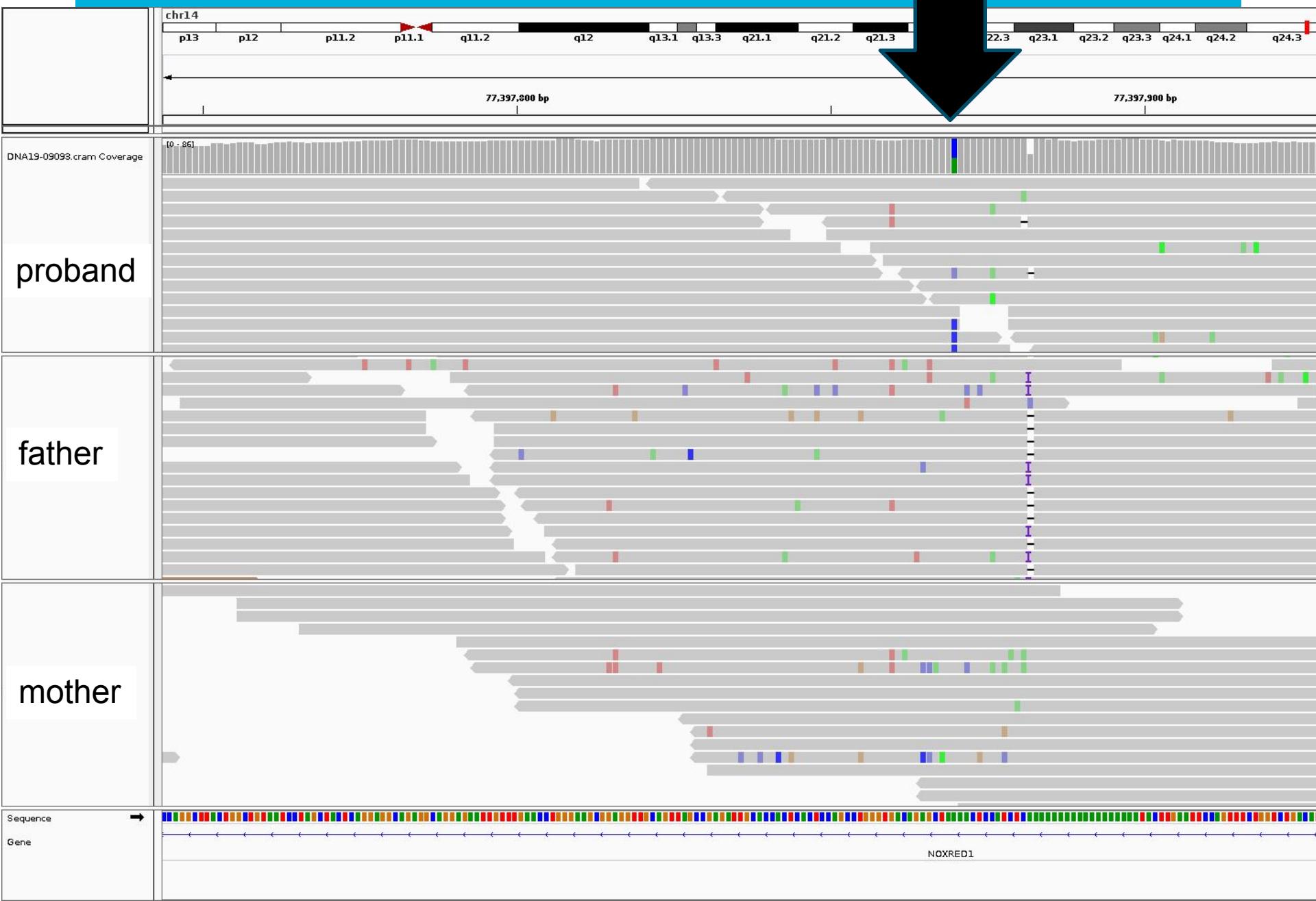
Cell

Spatial and Temporal Mapping of De Novo Mutations in Schizophrenia to a Fetal Prefrontal Cortical Network

Suleyman Gubum^{1,2}, Tom Welsh^{1,3}, Amanda C. Watts^{1,2}, Ming K. Lee¹, Anne M. Thornton¹, Silvia Cassadei¹, Caitlin Rippey¹, Hashem Shahin^{1,3}, Consortium on the Genetics of Schizophrenia (COGS)^{7,11}, PAARTNERS Study Group^{6,11}, Vishwanaj L. Nitinagonkar⁴, Rodney C.P. Go⁵, Robert M. Savage⁴, Neal R. Swerdlow⁷, Raquel E. Gur⁵, David L. Braff⁷, Mary-Claire King^{1,4} & Jon M. McClellan².

Radboudumc

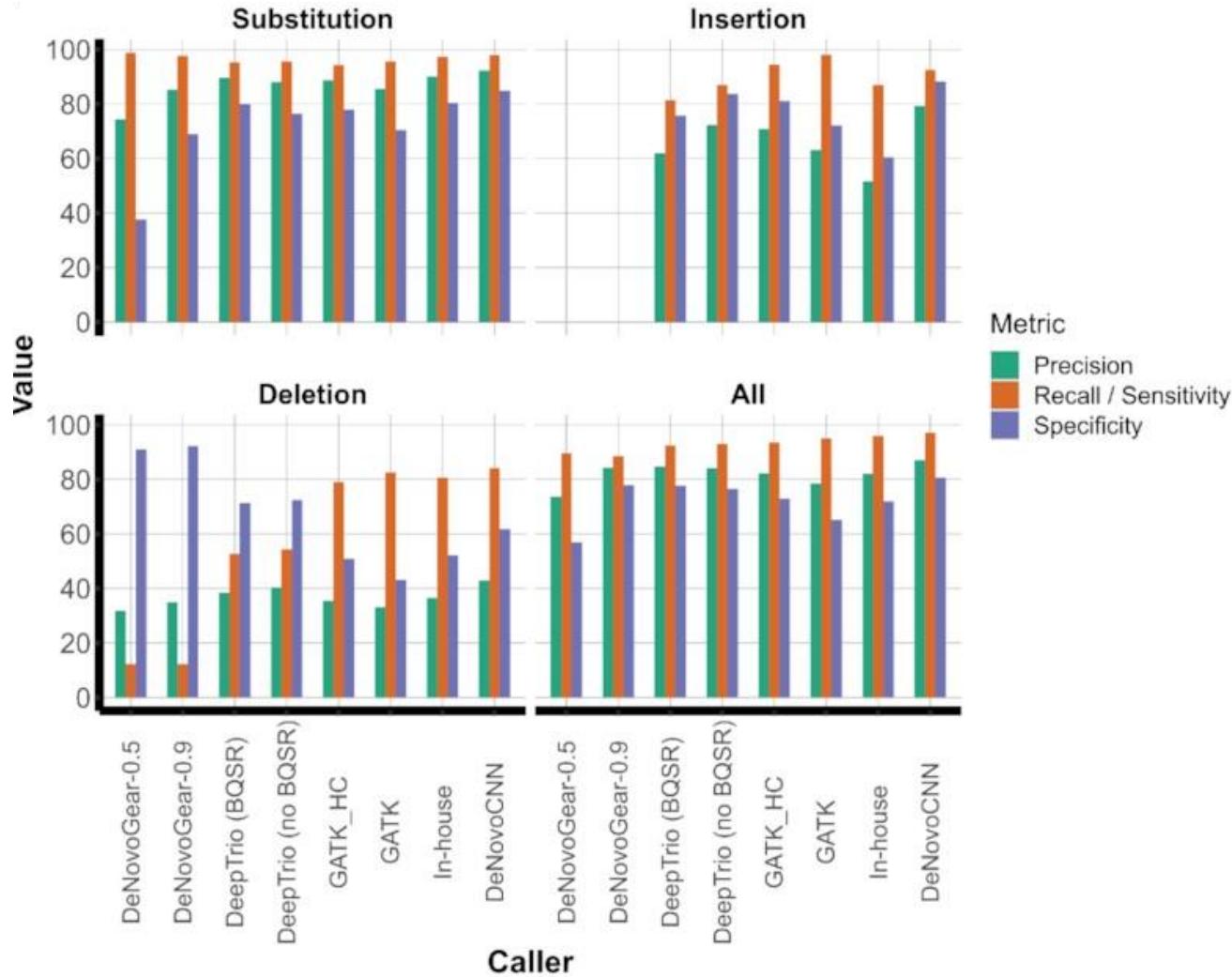
Detection of DNMs



Detection of DNMs

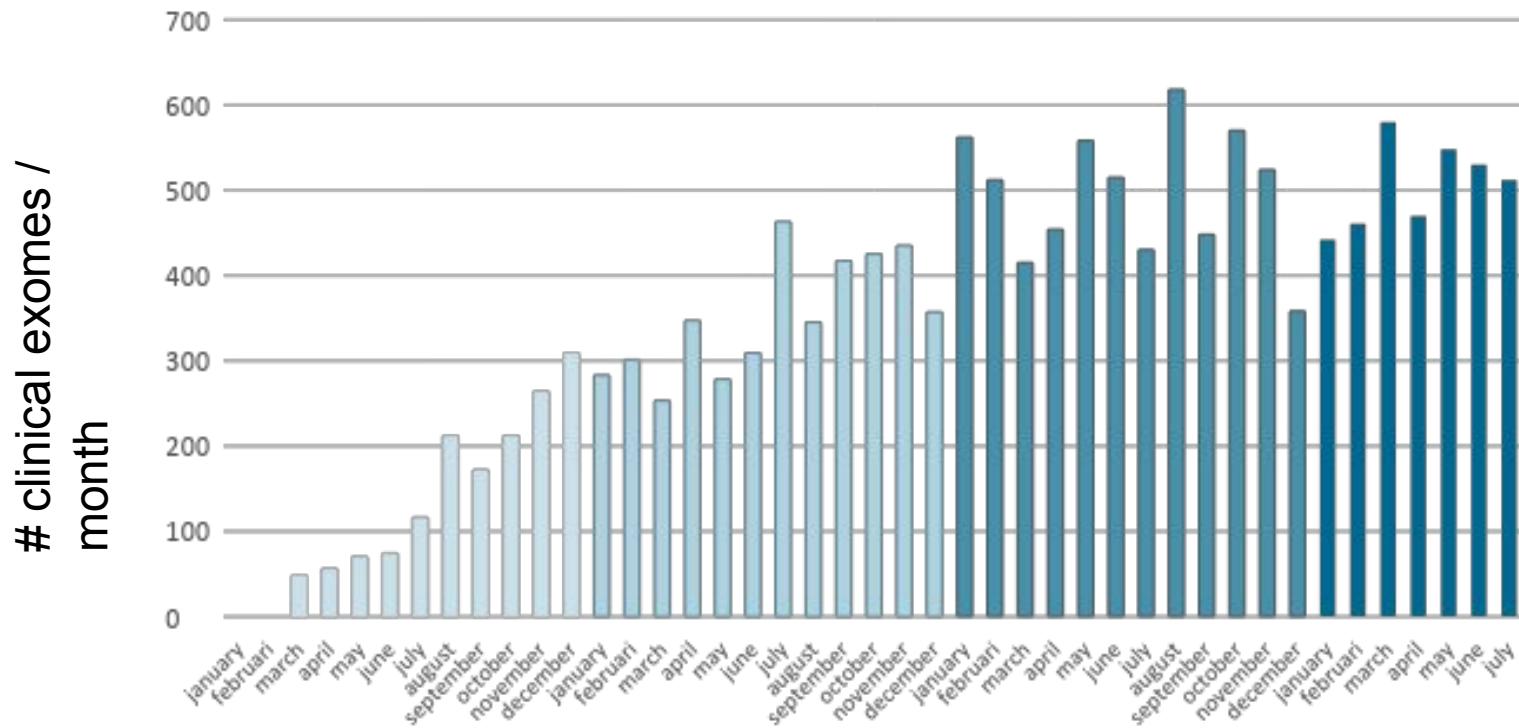
- Compare genotypes of patient and parents, but:
- Results in hundreds (exome) or thousands (genome) candidates due to:
 - False positive variants in the proband
(sequence error, mapping artifact, indels)
 - Lack of sequence coverage in the parents
 - Mosaicism
- **Solutions:**
 - Joint genotyping
 - Excluding difficult regions
 - Excluding common variants
 - Quality filtering of variants
 - Detailed investigation of potential de novo sites

Deep learning - DeNovoCNN



Exome sequencing
for
patient diagnostics

Routine clinical use WES since Sept 2012



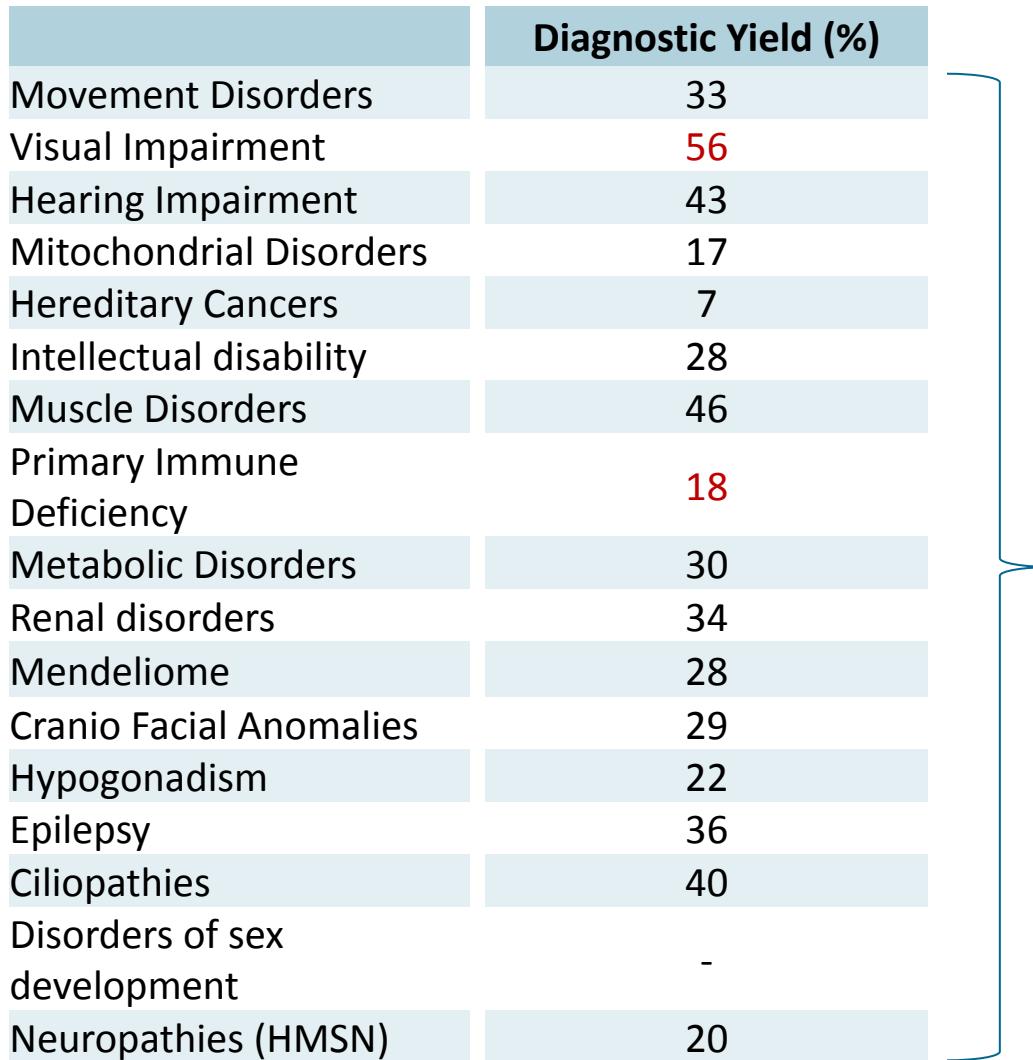
- Currently about 800 samples / month
- >60,000 individuals sequenced

Evolution of genes panels

The number of genes in a gene panel

	2017
Movement Disorders	255
Visual impairment	375
Deafness Disorders	142
Metabolic Disorders	548

Diagnostic yield per panel



Average diagnostic yield of
WES using panel strategy
~30%

Additional yield from WES

Much more is possible than detecting only SNVs:

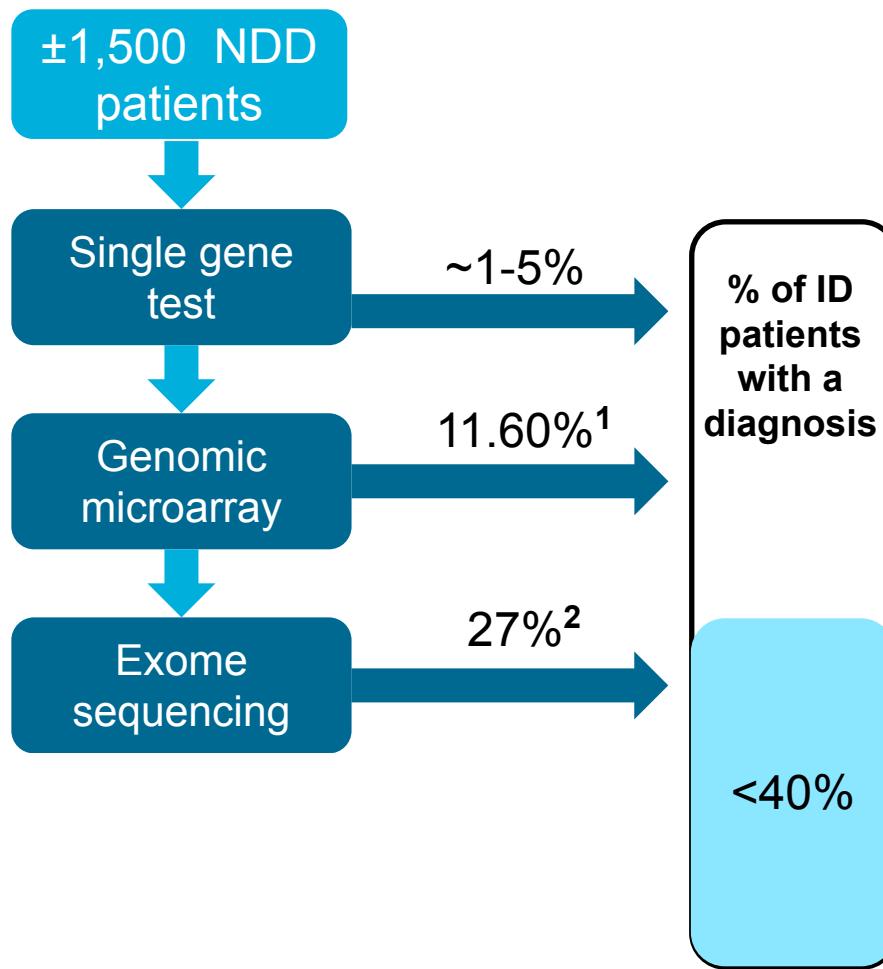
- **Copy number variation** calling (CNV) – *Pfundt et al. Genet med. 2017*
- Regions of **homozygosity** (ROH) – *Magi et al. Bioinformatics 2014*
- **Mobile Element Insertions** (MEI) – *Torene et al. Genet Med. 2019, Wijngaard et al. submitted.*
- Nucleotide **repeat expansions** – *Dolzhenko et al. Bioinformatics. 2019, van der Sanden et al. Genet. Med. 2021*
- **Uniparental disomy** (UPD) – *Yauy et al. Genet. Med. 2019*
- **Mitochondrial DNA** (mtDNA) – *Griffin et al. Genet med. 2014*
- Identify **mosaic** mutations – *Acuna-Hidalgo et al. AJHG 2015*
- Variation in **homologous regions** – *Steyaert et al. submitted.*

Take home message

- Exome sequencing provides a better diagnostic test in terms of diagnostic yield, compared to traditional methods
- Exome sequencing is a generic method that is easy to scale
- There is a lot of additional information to get out of an exome

Whole genome sequencing

Diagnosis in patients with severe NDD

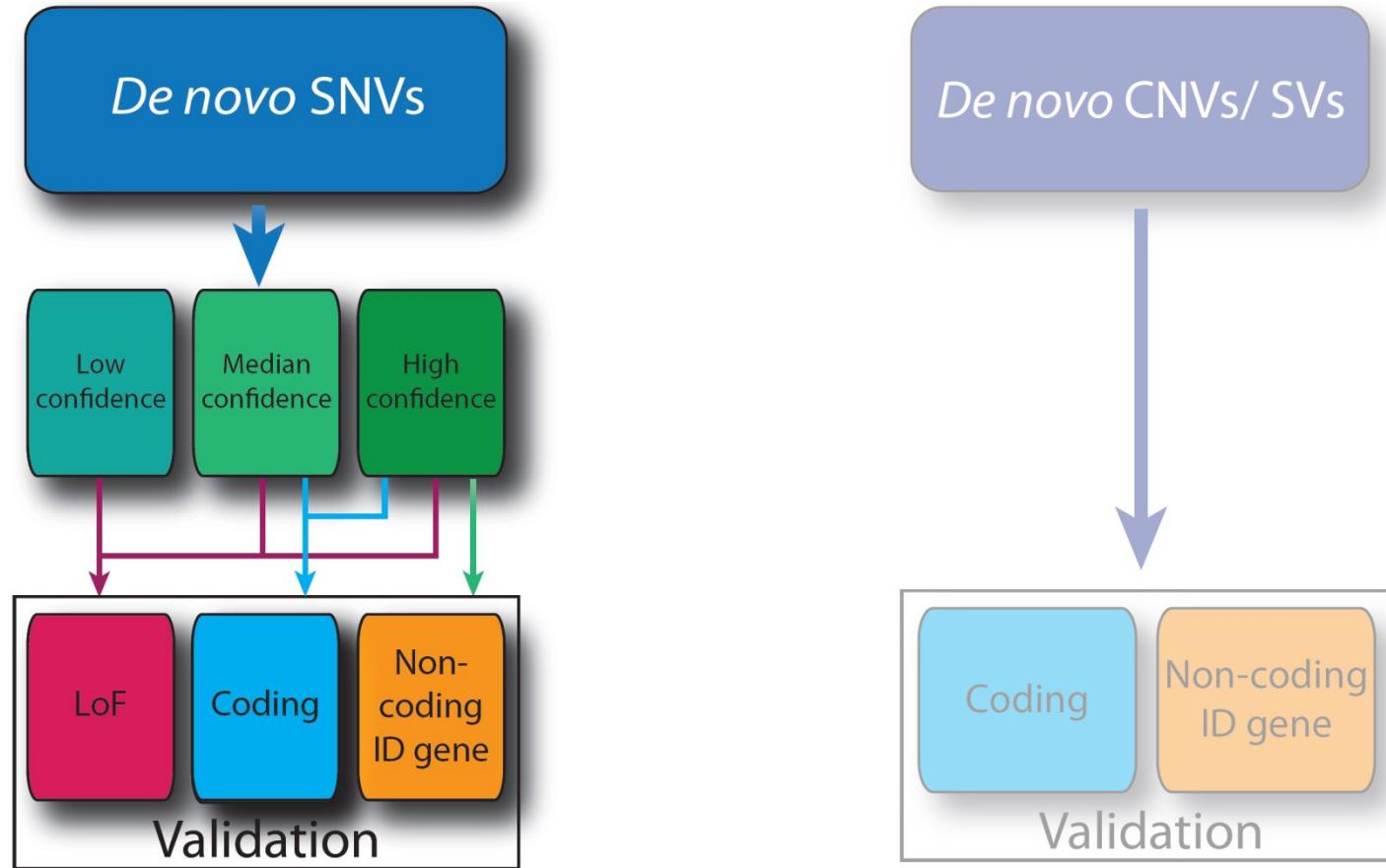


- Whole genome sequencing
- 50 trios at 80x coverage
- Focus on *de novo* mutations

¹Vulto-van Silfhout, A. T. et al Hum mut 2013

²de Ligt, J. et al. NEJM 2012

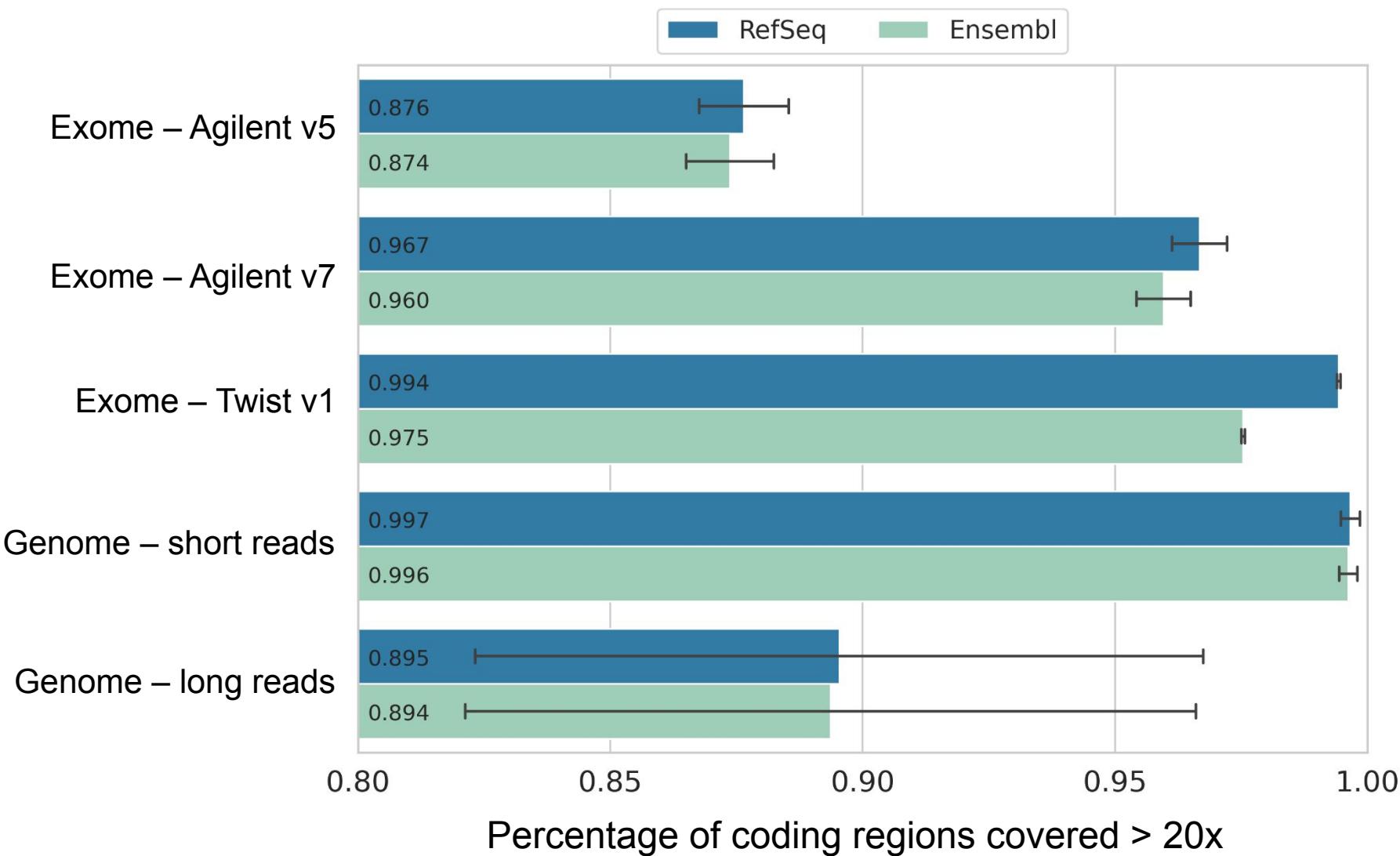
Can we identify *de novo* mutations?



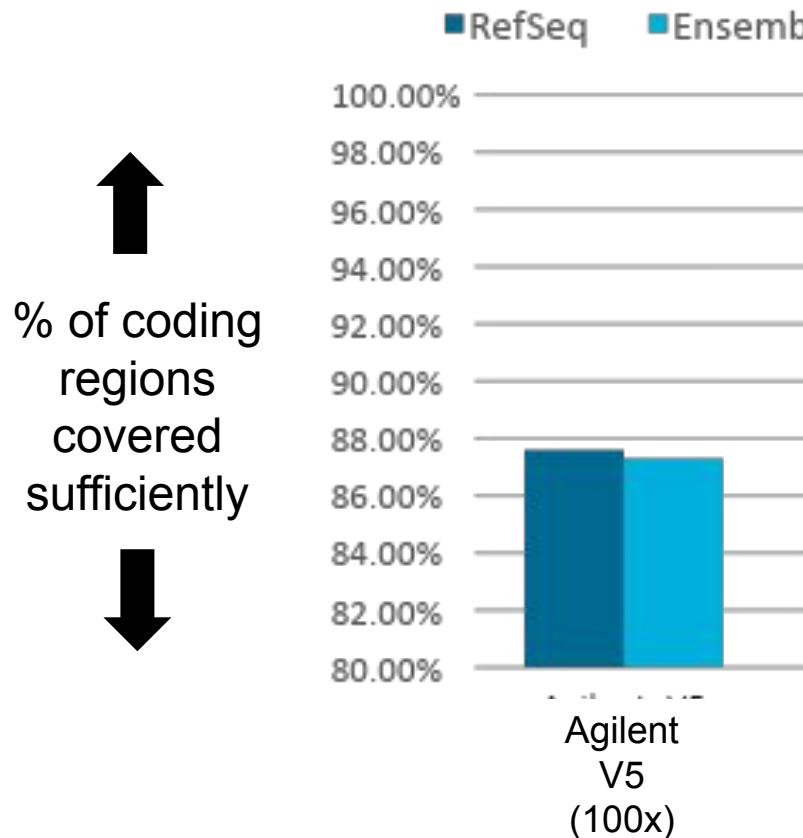
- Validation rate: **38%**
(80% for high confidence!)
- Coding *de novo* SNVs: **84**

- Validation rate: **82%**
- Coding *de novo* SVs: **9**

1. Coverage of protein-coding regions



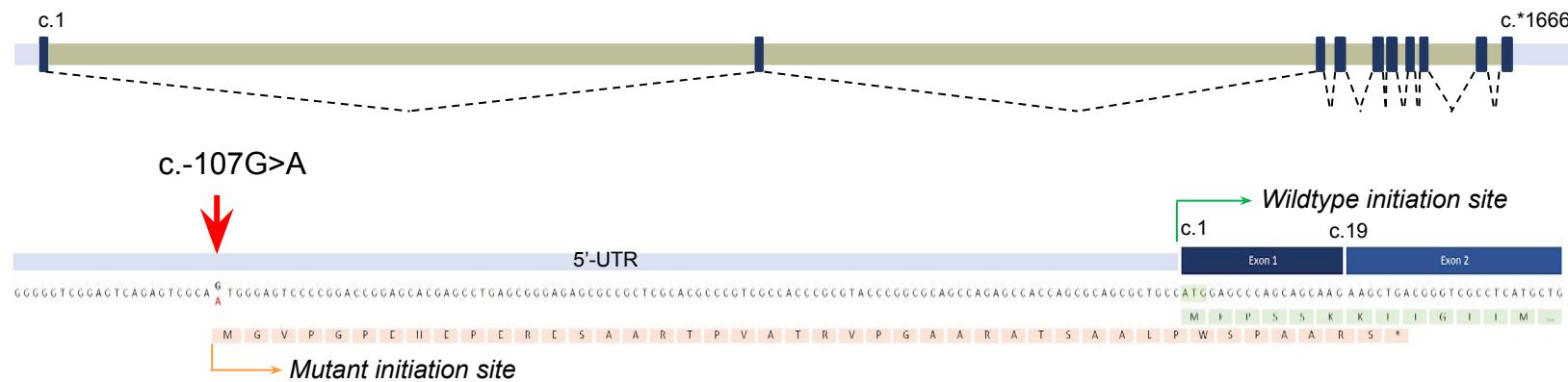
1. Coverage of protein-coding regions



- Short read WGS covers close to 99% of coding regions
- Twist (core) exome does equally well, if a target is present

2. Non-coding pathogenic mutations

Patient where biochemical test indicated a mutation in *SLC2A1*, but no mutation could be identified by Sanger or WES □ WGS

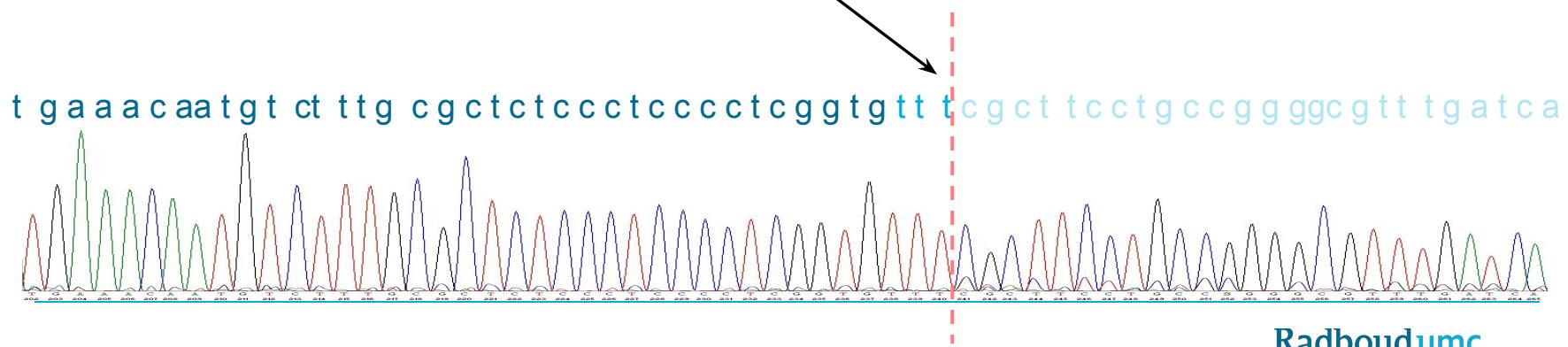
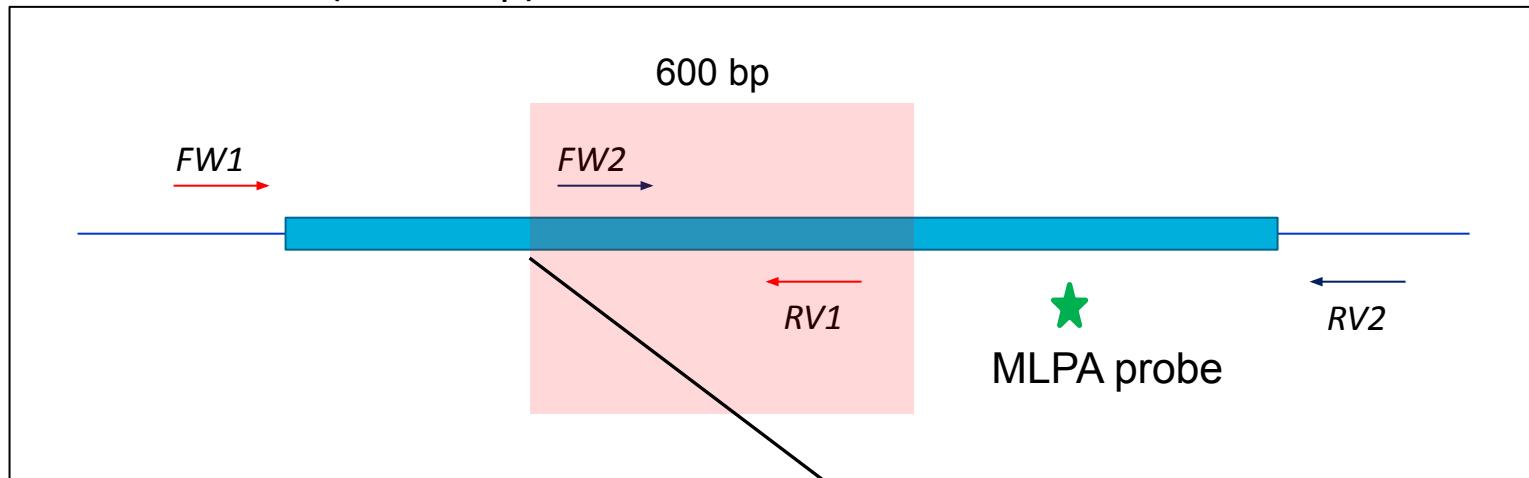


5' UTR *de novo* mutation introducing a new start site for *SLC2A1*, but causing a pre-mature stop codon targeted by nonsense-mediated-decay

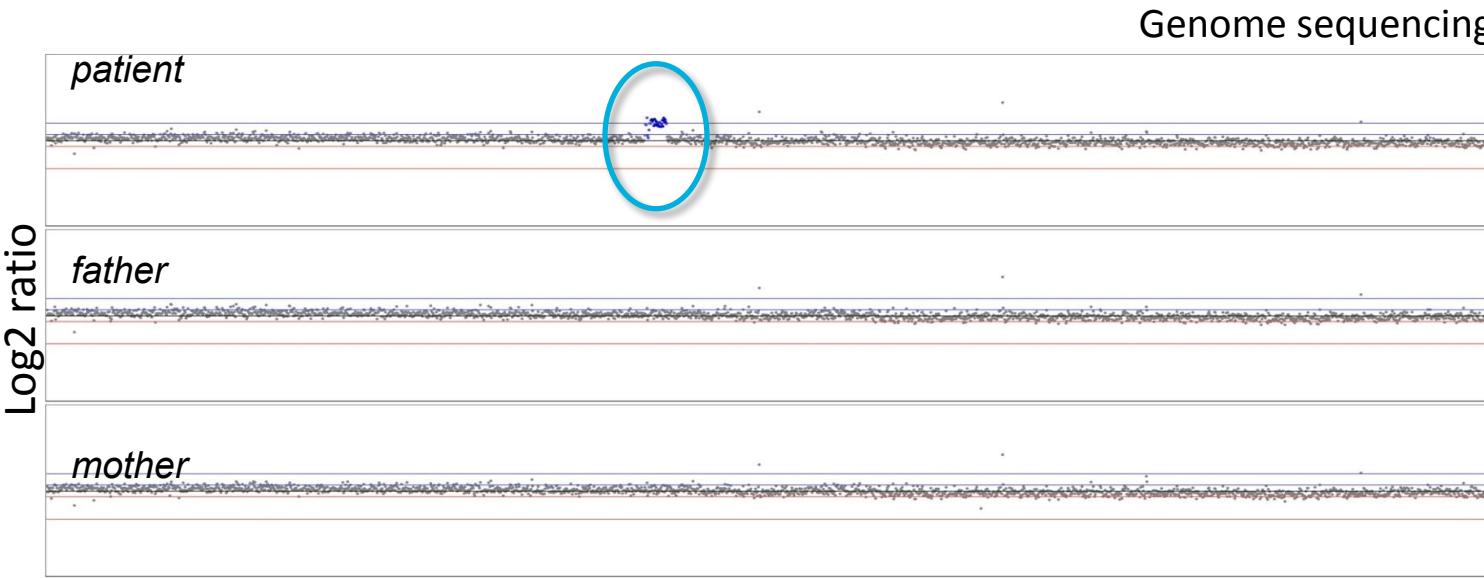
3. Structural variants, example 1

- A patient with the clinical suspicion of Rett syndrome
- *MECP2* gene tested by Sanger sequencing but no mutations identified

MECP2, exon 4 (~1000 bp)



3. Structural variants, example 2



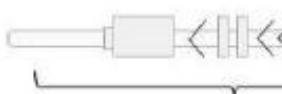
- Chromosome 4
- ~60 kb duplication on chromosome 4
 - Affecting the last 6 exons of the *TENM3* gene
 - *TENM3* is associated with coloboma, and microphthalmia

Duplication from chr4 to chrX

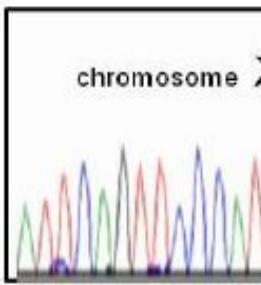
TENM3



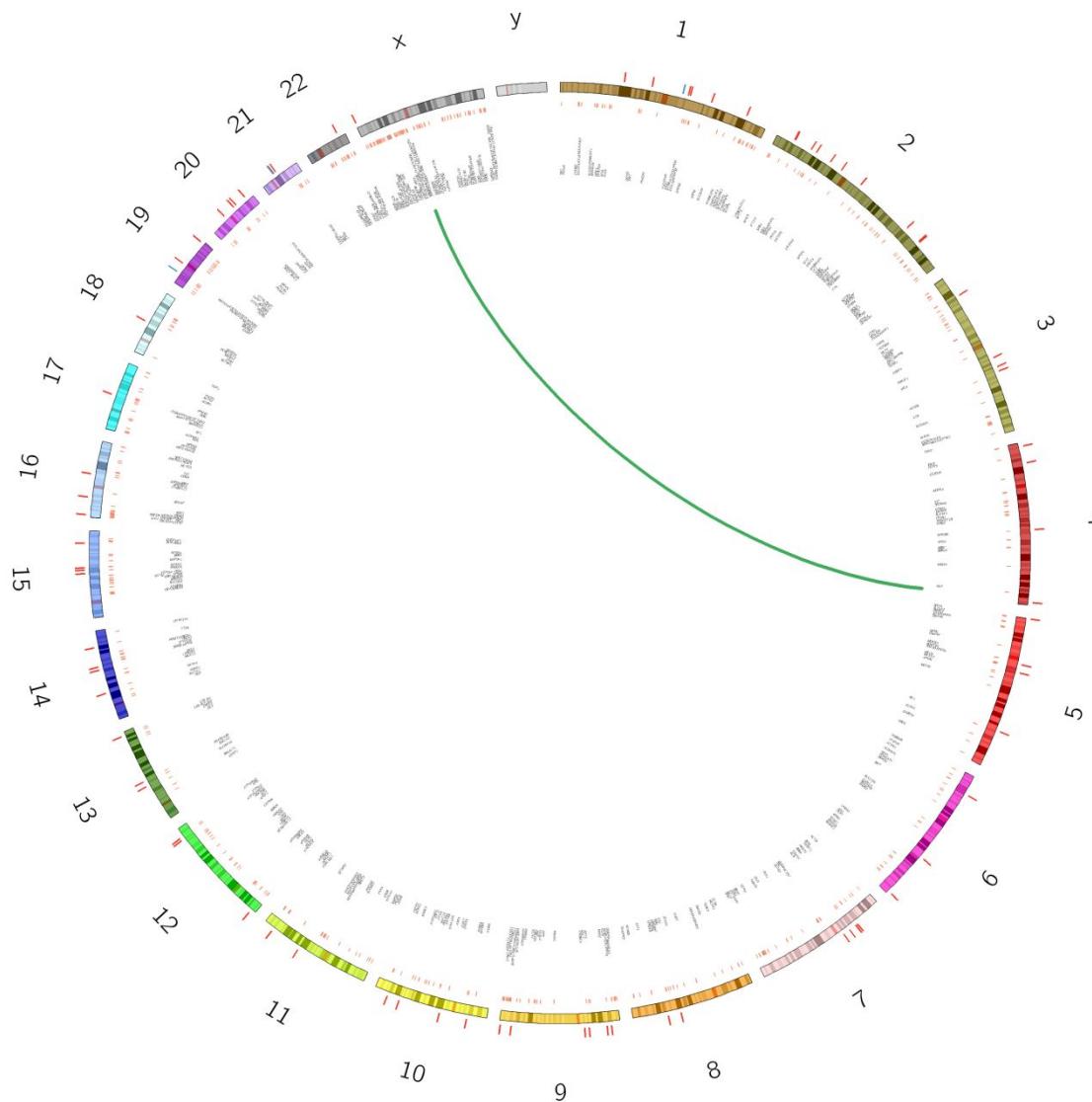
*IQSEC
(ex3-1)*



chromosome

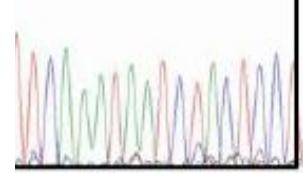


- First time



*E2
(1-2)*

chromosome X



event.

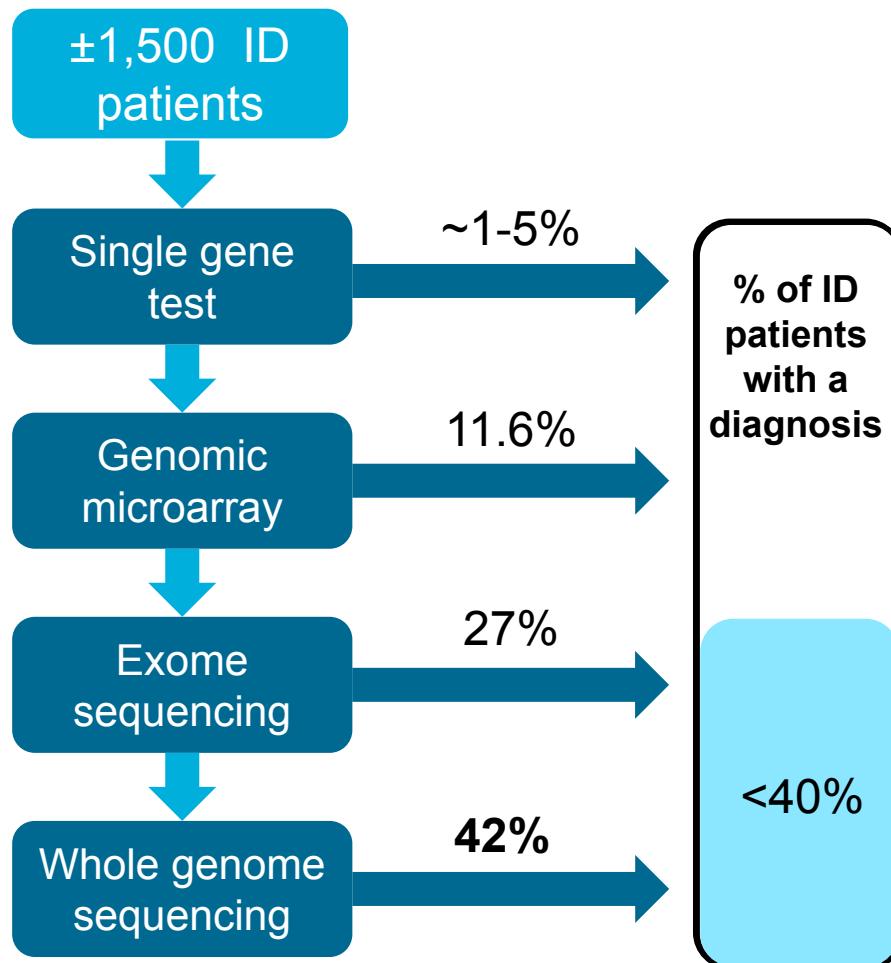
Diagnostic yield genome sequencing

Highly likely diagnosis

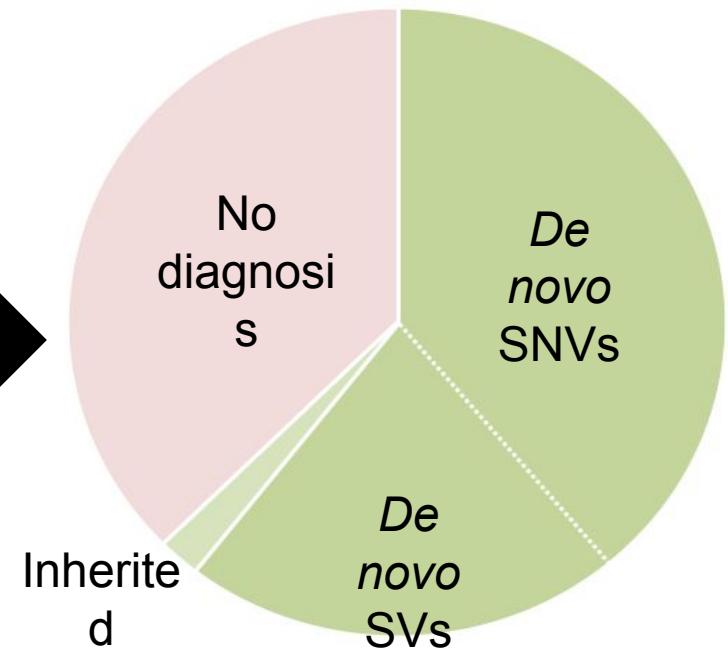
SNV	SV
<i>TBR1</i> (2x)	<i>SHANK3</i>
<i>WDR45</i>	<i>VPS13B</i> *
<i>SMC1A</i>	<i>MECP2</i>
<i>SPTAN1</i>	<i>IQSEC2</i>
<i>RAI1</i>	<i>STAG1</i>
<i>MED13L</i>	<i>SMC1A</i>
<i>SATB2</i>	16p11.2 microdel. syndrome
<i>PPP2R5D</i>	Multiple genes
<i>KCNA1</i>	
<i>SCN2A</i>	
<i>POGZ</i>	
<i>KANSL2</i>	21/50 cases diagnosed: 42%

* Recessive variants. Known genes in bold.

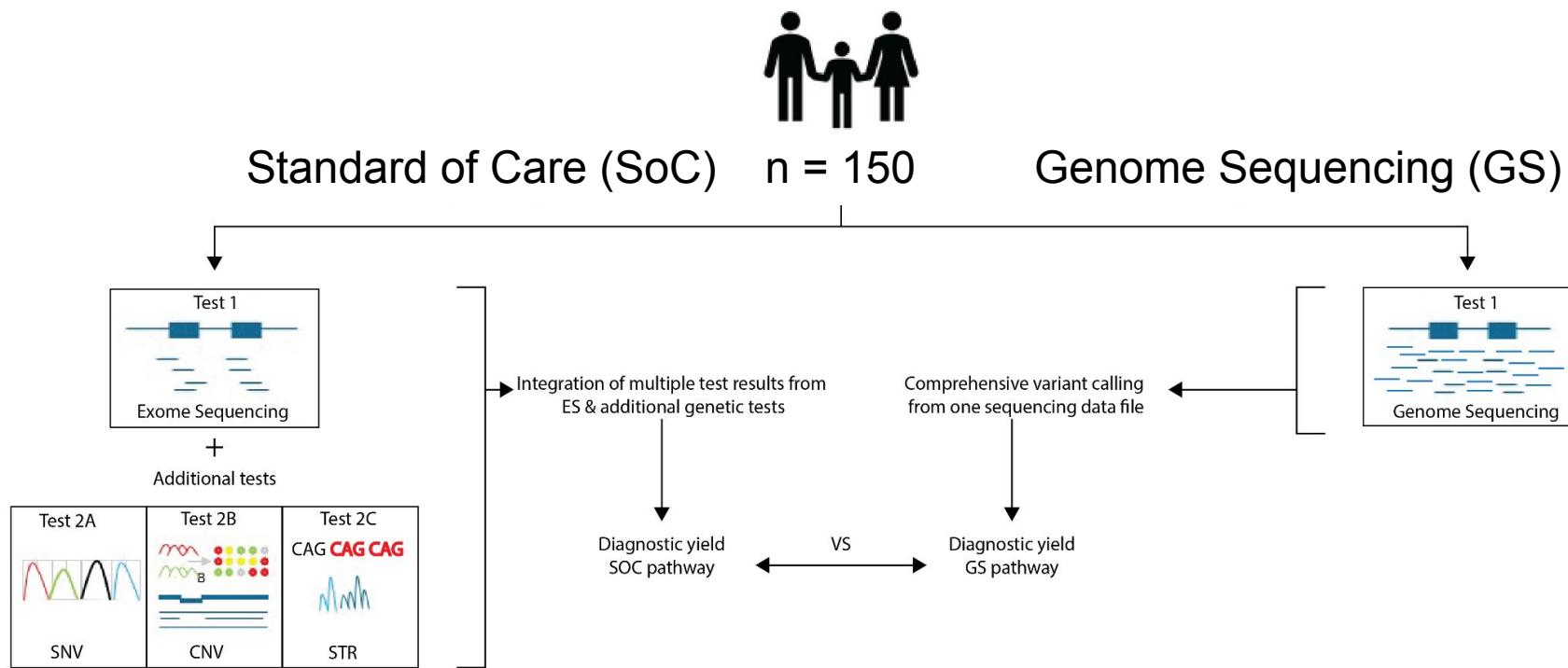
Diagnostic yield



- **Majority is *de novo*!**

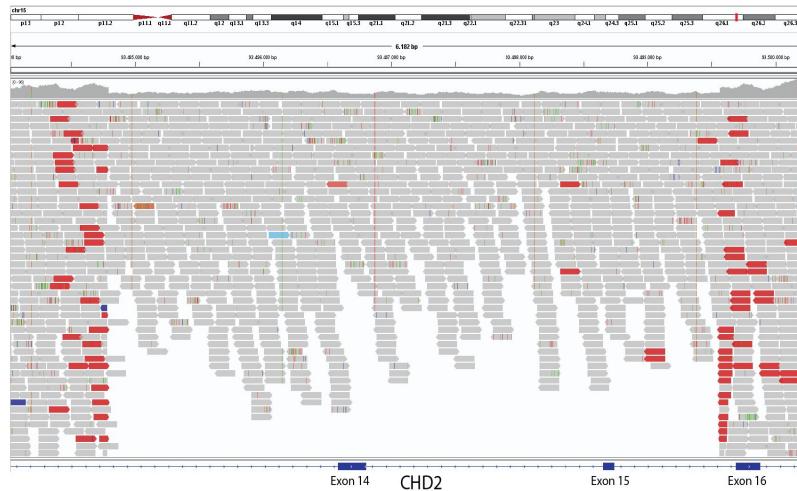


WGS to replace all other tests?

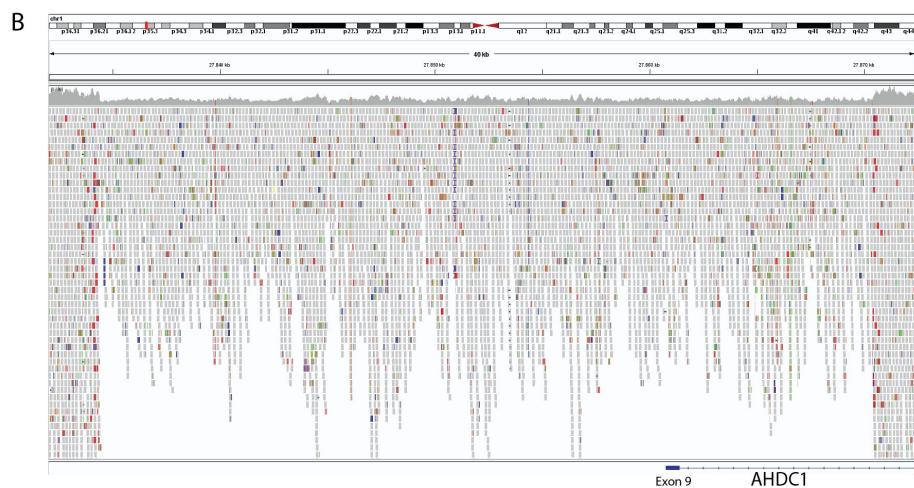


Comparing genome sequencing

		Genome sequencing			
Standard of Care	DIAGNOSIS	YES	POSSIBLE	NO	TOTAL
	YES	43	0	0	43
	POSSIBLE	0	31	0	31
	NO	2	4	70	76
	TOTAL	45	35	70	150



5kb deletion in *CHD2* at 15q26.1



36kb deletion in *AHDC1* at 1p36.11

Conclusions

1. WGS gives you better coverage in the coding regions than exomes
2. WGS allows the comprehensive detection of *de novo* SNVs and CNVs/SVs in a single experiment
3. Based on clinical diagnostic criteria we can provide a genetic diagnosis for the majority of severe ID cases
4. Many challenges remain:
 - cost, storage and compute
 - interpretation of (non-coding) variation

Long read sequencing

How are neurodevelopmental disorders diagnosed?

Diagnostic yield from a clinical exome is approximately **30-40%**.

What about the remaining ~60 - 70%?

1. Environmental factors (foetal alcohol syndrome)
2. Interpretation of variation in known NDD genes
3. Novel (undiscovered) NDD genes
4. **'Tricky variation': variants in repetitive regions and complex structural rearrangements**
5. Interpretation of non-coding variation
6. Genetic factors (penetrance, expressivity, epistasis)

Short versus long reads

Short read sequencing	Long read sequencing
Read lengths of $\pm 150\text{bp}$	Read lengths of $>15\text{ kb}$
\$800 for a 30x genome	\$6000 for a 30x genome

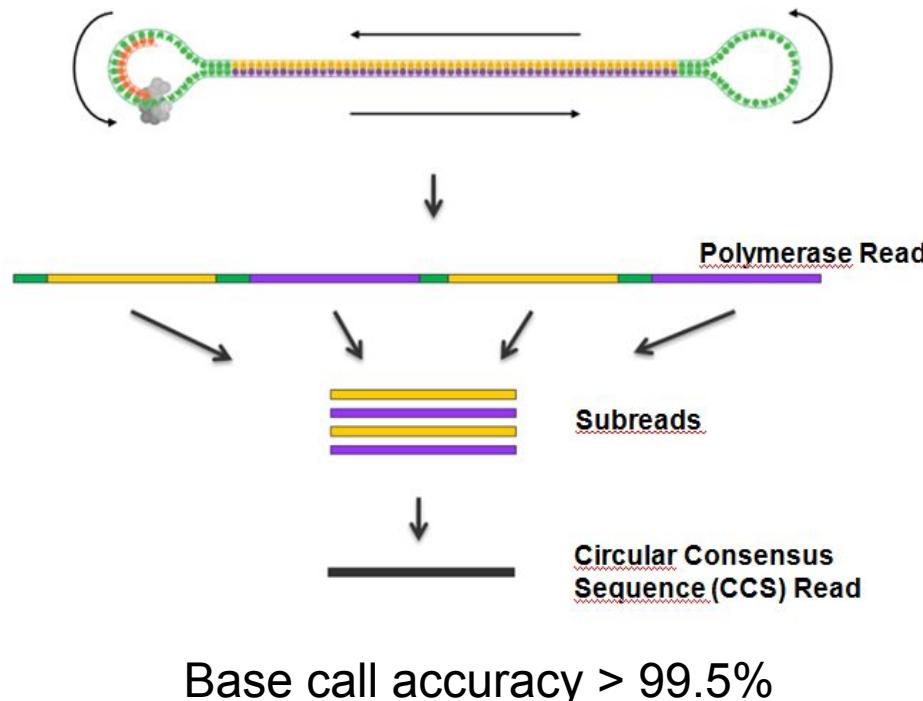
PacBio Single Molecule Real Time sequencing



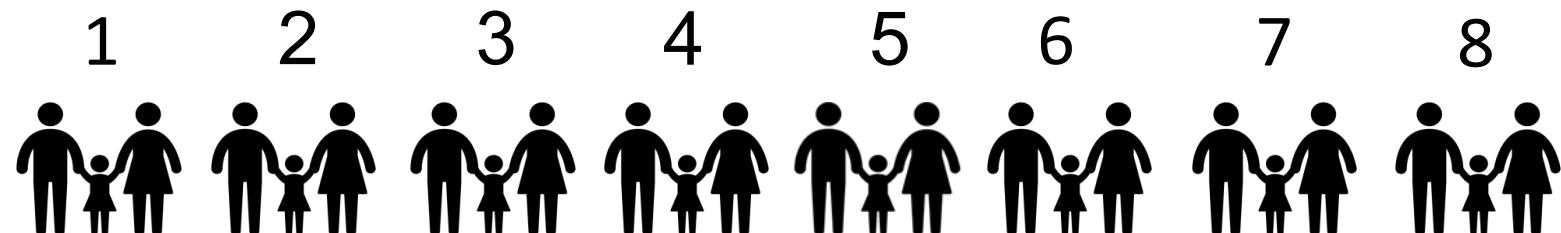
- Sequel IIe: 50-fold throughput
- Allow for accurate HiFi sequencing
- One high quality human genome
with 30x coverage ~6k€

HiFi sequencing

- High Fidelity (HiFi) reads are generated by sequencing the circularized input DNA repeatedly and generating a consensus sequence of all passes
- Their higher accuracy allows for more precise detection of single nucleotide variants



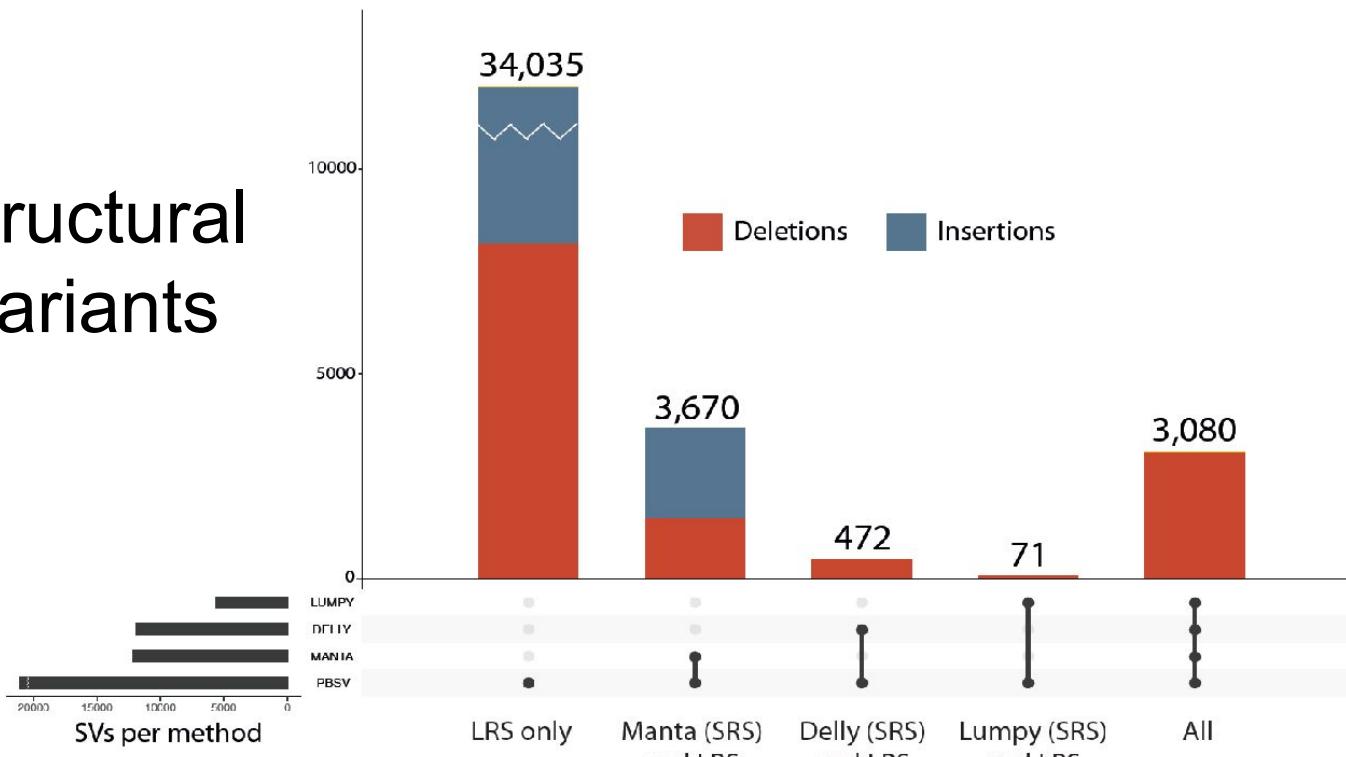
Experimental Design: HiFi Trios



- We sequenced 8 trios unresolved ID with Pacbio HiFi reads to 10-15x coverage
- We focused on discovering *de novo* variation
- We have Illumina WGS data for all trios for comparison

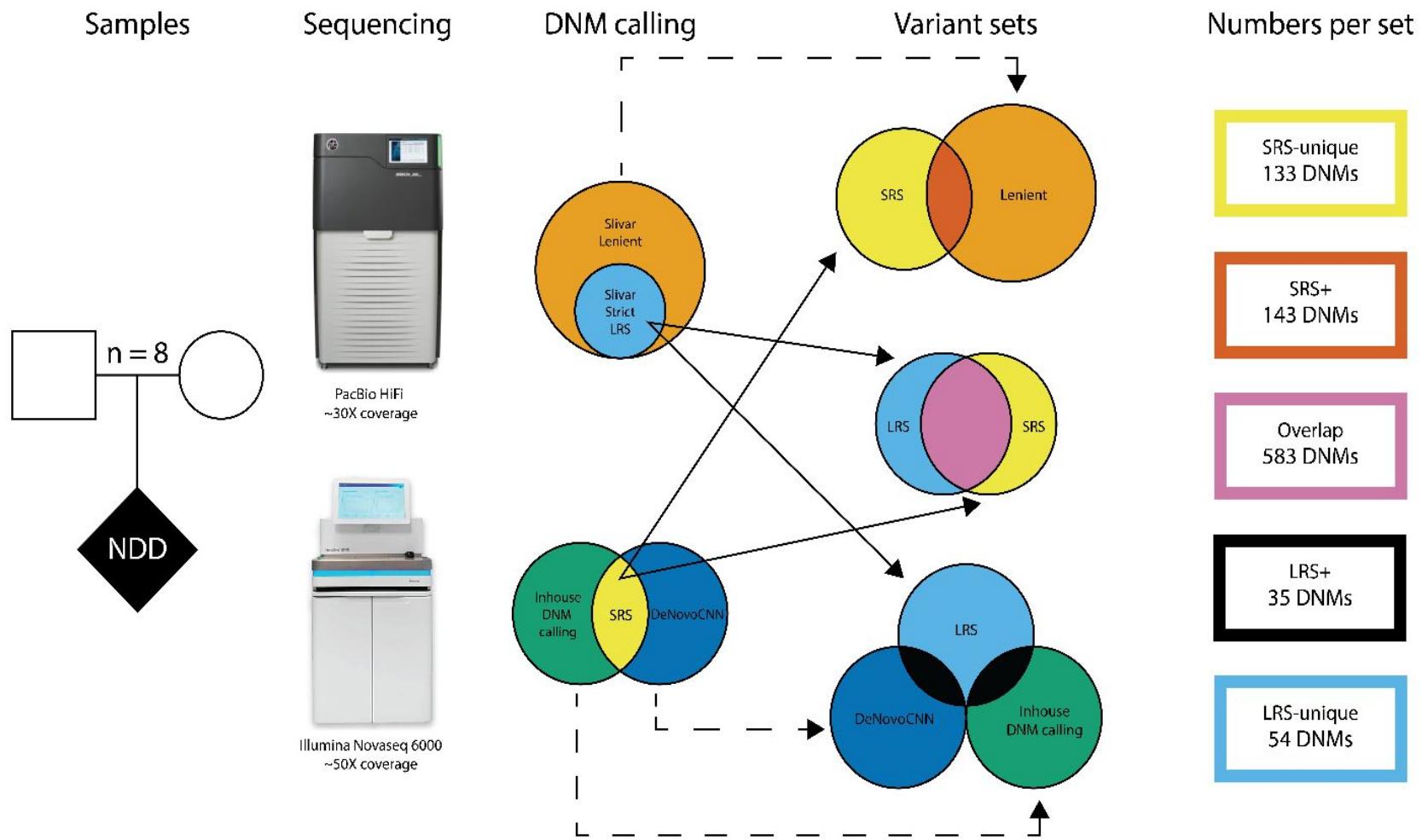
Comparison of LRS and SRS

Structural variants

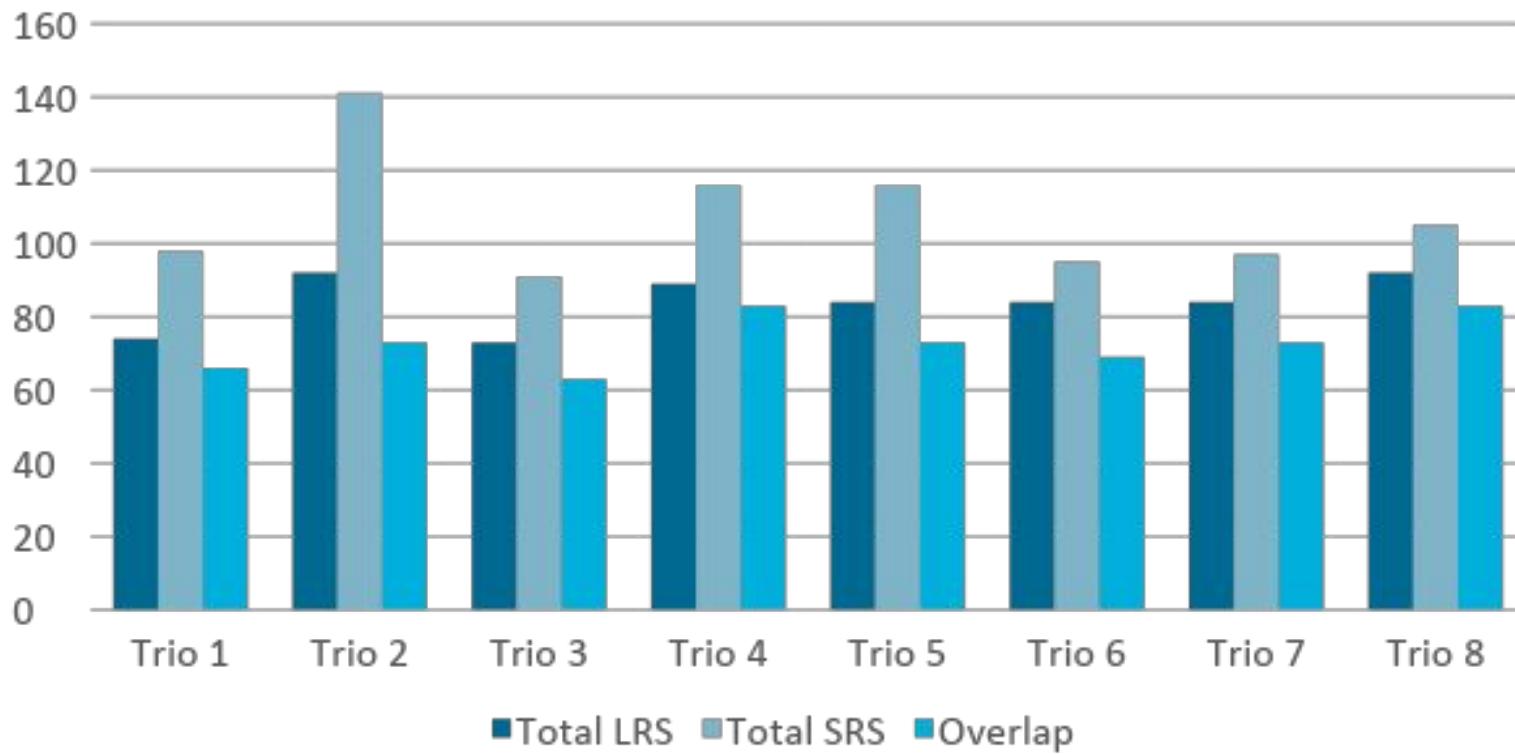


- **LRS will become the de-facto standard technique for future genetics research and diagnostics.**

HiFi Sequencing for DNM detection



Comparison of de novo mutations

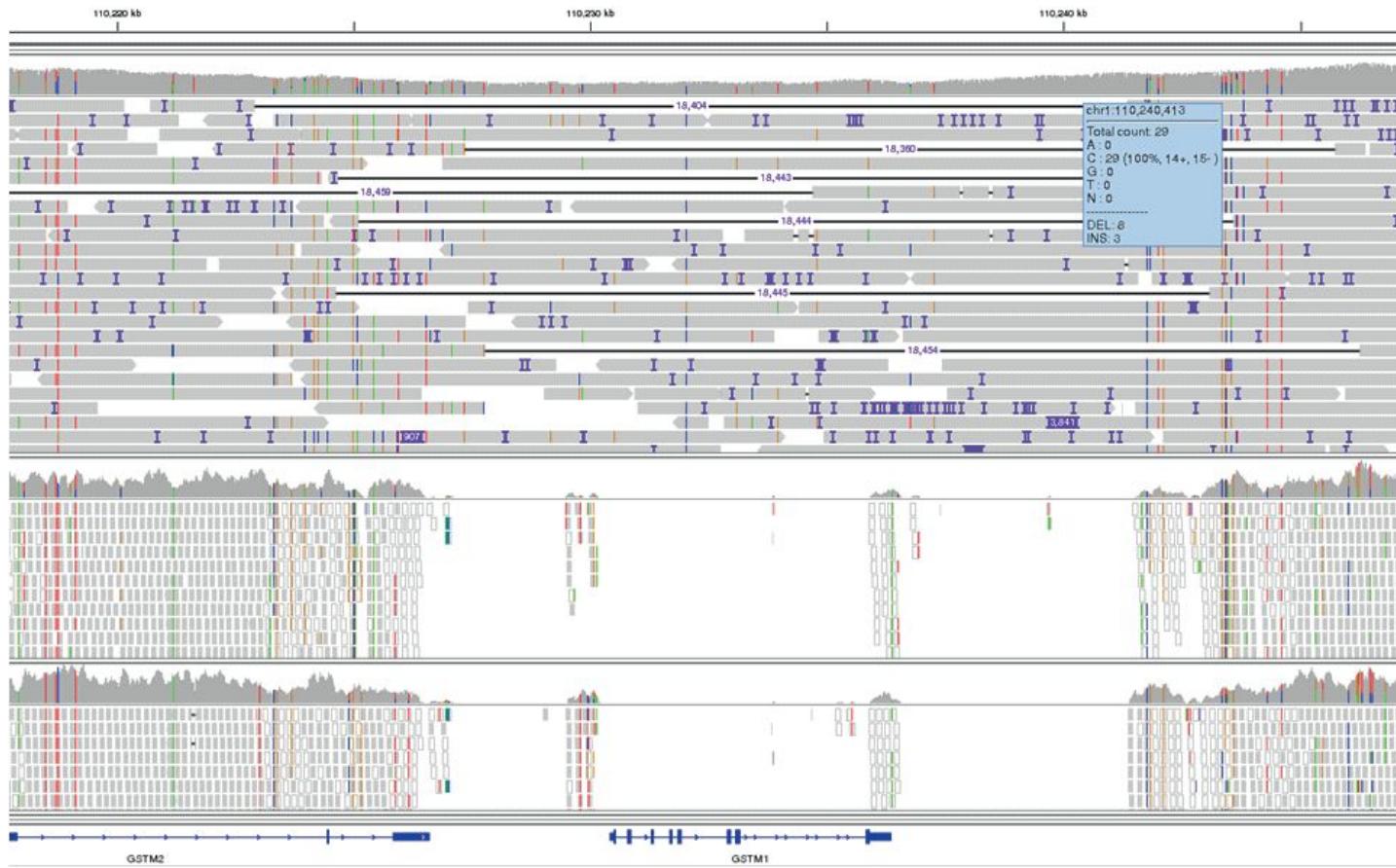


- Overlap between DNM^s is $\pm 84\%$ assuming SRS as the standard
- Validation of unique SRS variants by Sanger showed we missed at least **1-2 DNM^s per trio with LRS** (15 in total)
- Validation of unique LRS variants showed we missed **at least 1 DNM per trio with SRS** (11 in total)

Coverage again

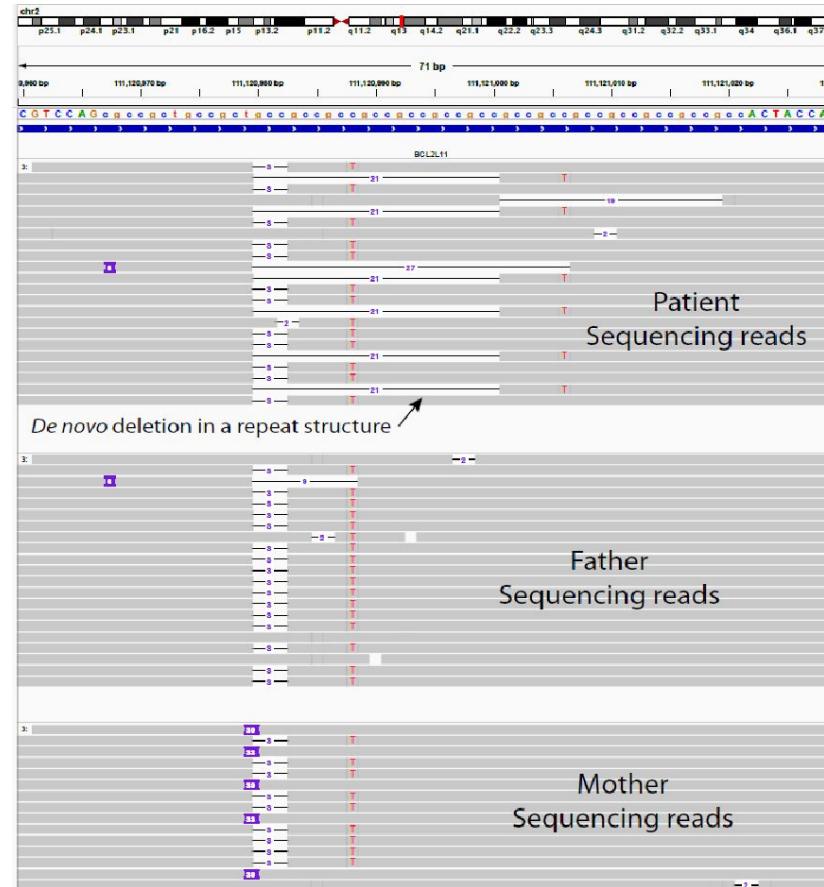
Long-read
sequencing
on PacBio
Sequel

Illumina
NovaSeq
(40x)



De novo structural variation

Long-read sequencing data identifying an *18 bp maternal deletion in repetitive coding regions of the gene BCL2L11.*



Reads with 3 bp deletion

Reads with 18 bp deletion

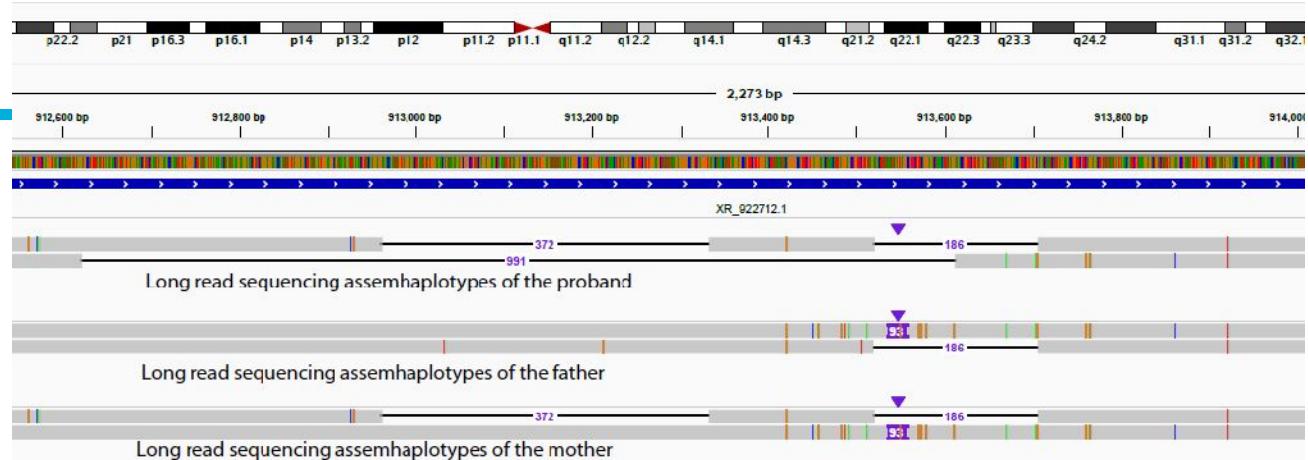
Only reads with a 3 bp deletion

Reads with a 3 bp deletion

Reads with a 33 bp insertion

Haplotype

*991 bp de novo
deletion on the
paternal allele*



Summary

- Long read sequencing detects $\pm 30,000$ structural variants per sample that are missed with standard short read genome sequencing
- With HiFi reads we gain accuracy sufficient to perform high quality de novo mutation detection
- Long read sequencing has the potential to increase the yield even further by being able to detect all variation in a single experiments.