Exome and genome sequencing in clinical practice

2nd IRDiRC conference - Shenzhen

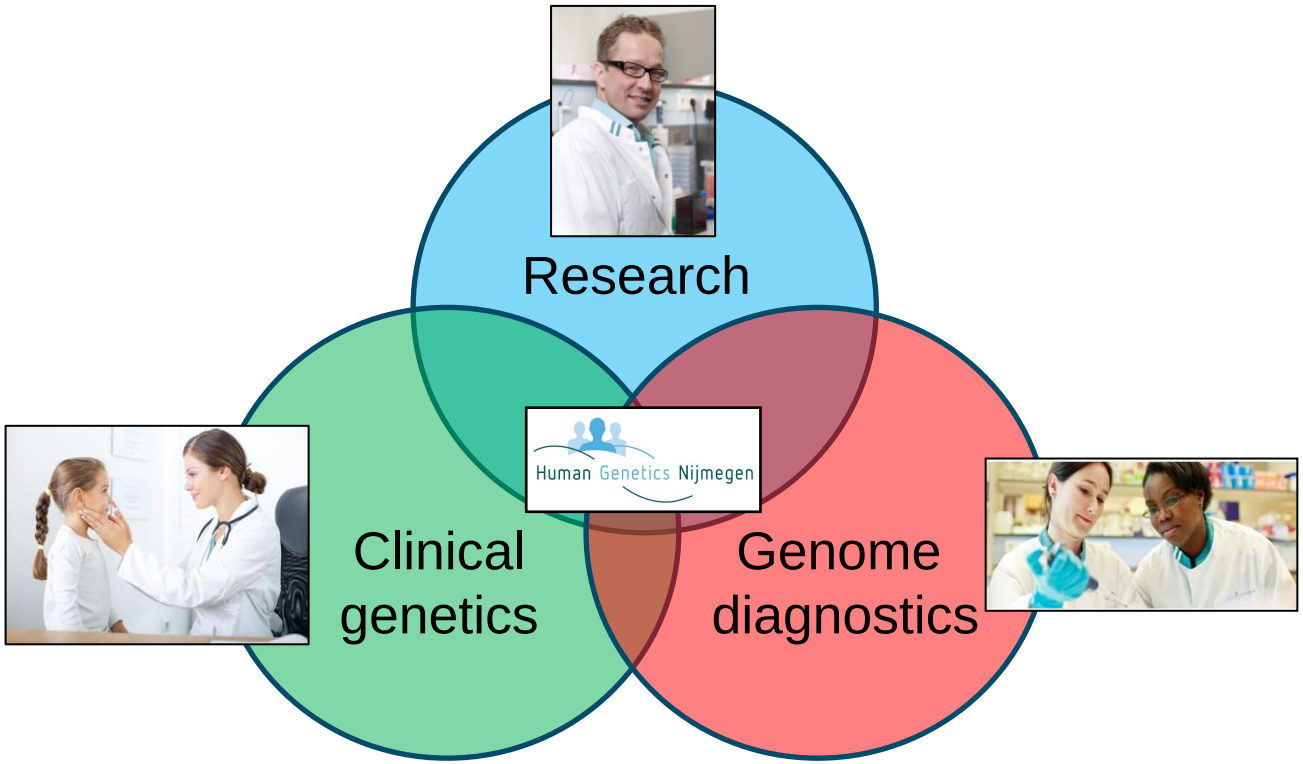
Christian Gilissen Ph.D.
christian.gilissen@radboudumc.nl
07-11-2014

Human Genetics Nijmegen

Radboudumc
university medical center

Human genetics Nijmegen

Human genetics Nijmegen



- Clinical genetics, diagnostics and research in one department
- Largest department of Human Genetics in the Netherlands

The challenge in diagnostics

- Single gene testing offered for more than 400 Mendelian diseases and more than 800 genes.

Good diagnostic yield for diseases with few causative genes (e.g. Noonan syndrome).

Poor diagnostic yield for genetically heterogeneous diseases. (e.g. Blindness, > 100 genes)

- **Solution**: Next generation sequencing?

Why exome sequencing?

Sanger	Targeted	Exome	Genome
• Very accurate • Cheap per exon • High turn-around	• Optimization possible • Low chance of incidental findings • “Easy” analysis • “Easy” interpretation	• No bias for genes • Standardized workflow • Re-use of performed exomes to interpret new ones • Simple to add new genes	• No bias in what you sequence • Little technical biases • Allows detection of SVs and SNVs in one experiment
• Low diagnostic yield for genetically heterogeneous diseases	• Design and re-design required • Different designs for different disorders • Sufficient patients required	• Sequencing bias • No non-coding regions • Incidental findings	• Data analysis bottleneck • Interpretation of non-coding variants • Expensive, time-consuming

Current exome sequencing workflow

Out-sourced: BGI

Enrichment
- Enrichment (Agilent v4)
- ~ 21,000 genes

Sequencing
- Sequencing at BGI Copenhagen
- Using Illumina 2x100bp, 75x median coverage

Primary Analysis
- Read mapping with BWA
- Variant calling with GATK

In-sourced: Nijmegen

Secondary Analysis
- Quality control
- Sample mix-up check
- Variant annotation

Inter-pretation
- Gene package visualization
- Standardized interpretation protocol
- Independent interpretation by 2 people

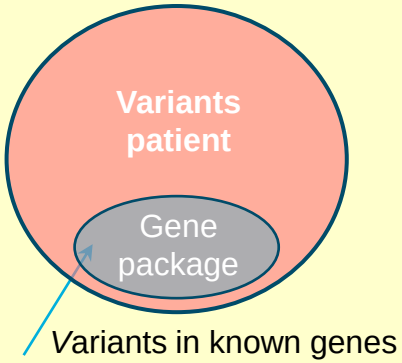
Report
- Validation by Sanger
- [Segregation analysis and functional confirmation]
- Report of results

±350 exome samples per month!

Diagnostic exome approaches

Gene package approach

Most genes known



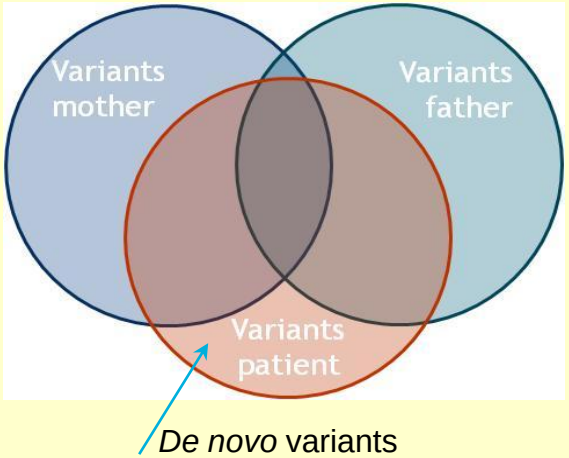
Pilot study: 50 exomes for 5 disorders

Neveling *et al*. Hum mut. 2013

- Sequence the exome of the patient
- Look only at known genes

Trio approach

Most genes unknown



Pilot study: 100 trios for intellectual disability

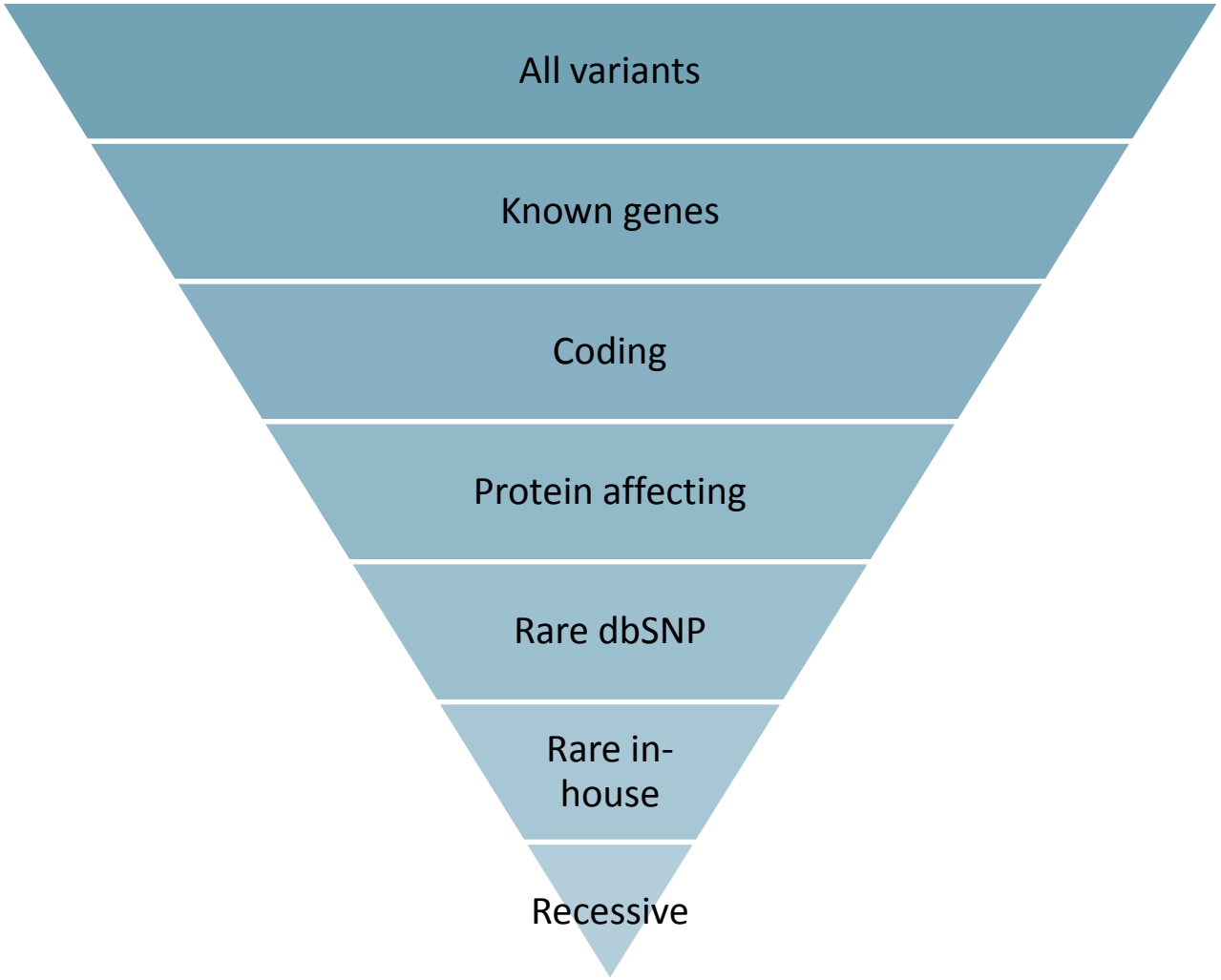
De Ligt *et al*. NEJM, 2012

- Sequence the exome of father, mother and patient.
- Look for *de novo* mutations

Challenges in clinical exome sequencing

When it all works out…

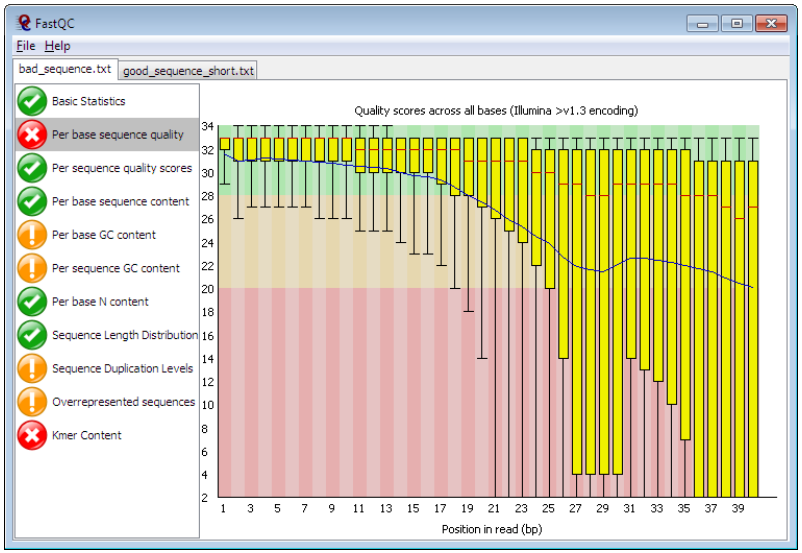
- Prioritization of variants found by exome sequencing:



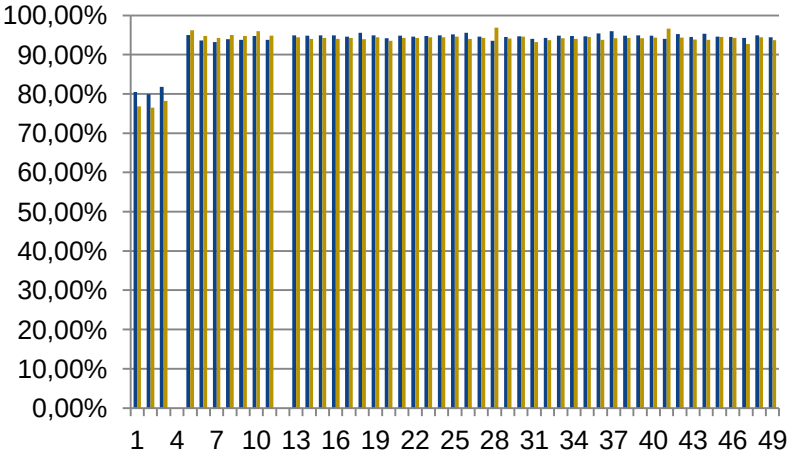
43,906
404
156
63
10
7
2

What can go wrong? Sample quality

- **Raw sequence and mapping statistics**
  - FastQC tool
  - Bedtools – coverage statistics
  - Per gene / exon target coverage

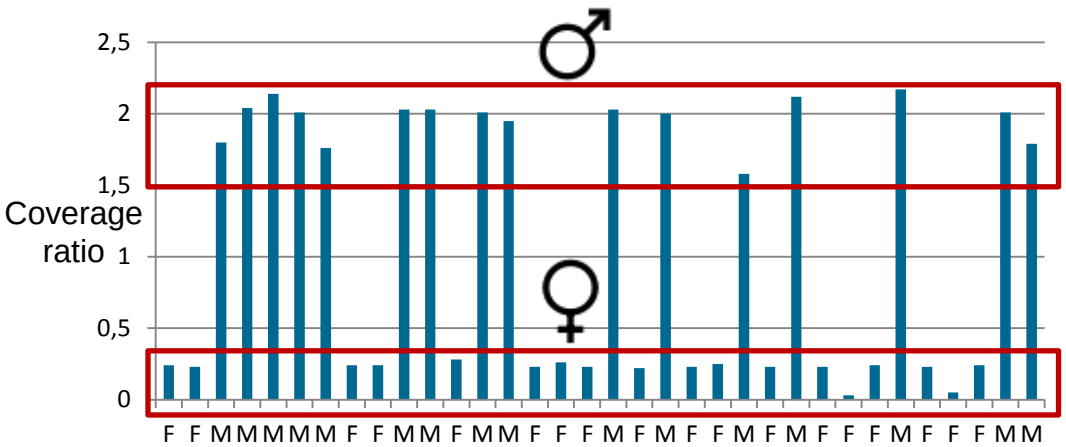


- **Variant statistics**:
  - Overlap dbSNP
  - Number of truncating mutations
  - Tr/Ti ratio



What can go wrong? Sample mix-up

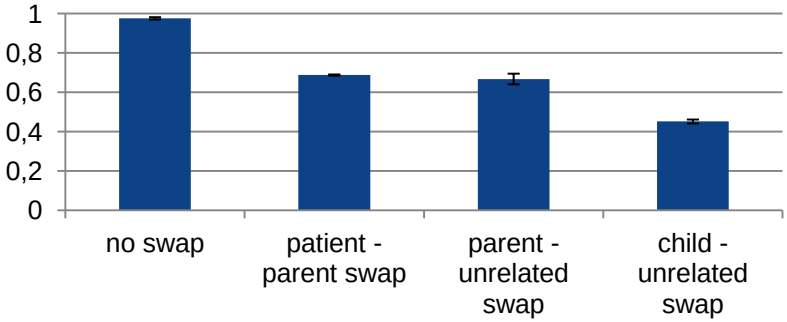
- **Gender check**:
  - Calculate chrY/chrX target coverage ratio



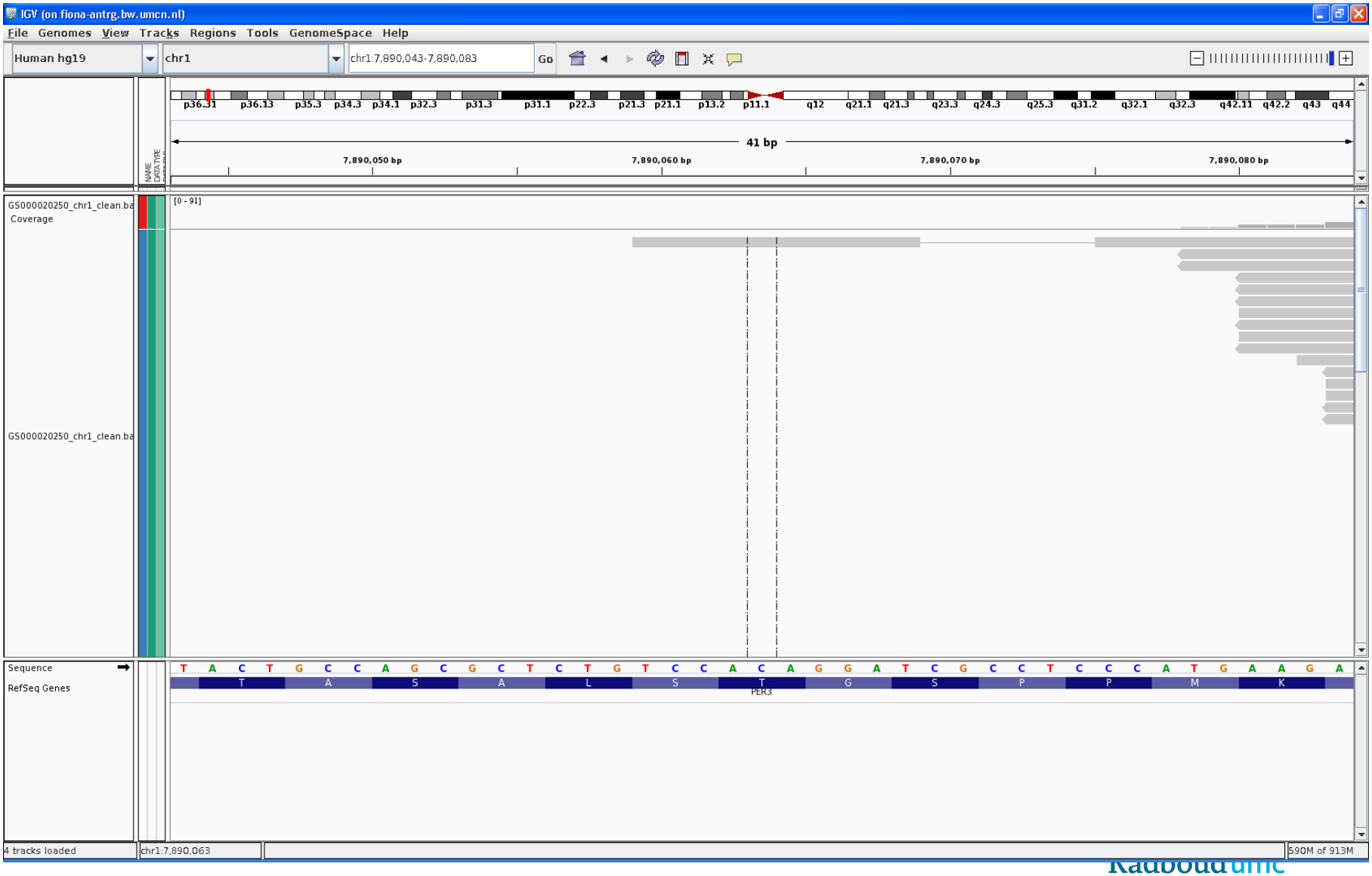
Gender according to patient database (M=male, F = Female)

- **SNP Test**:
  - 12 common SNPs tested separately by Sanger sequencing

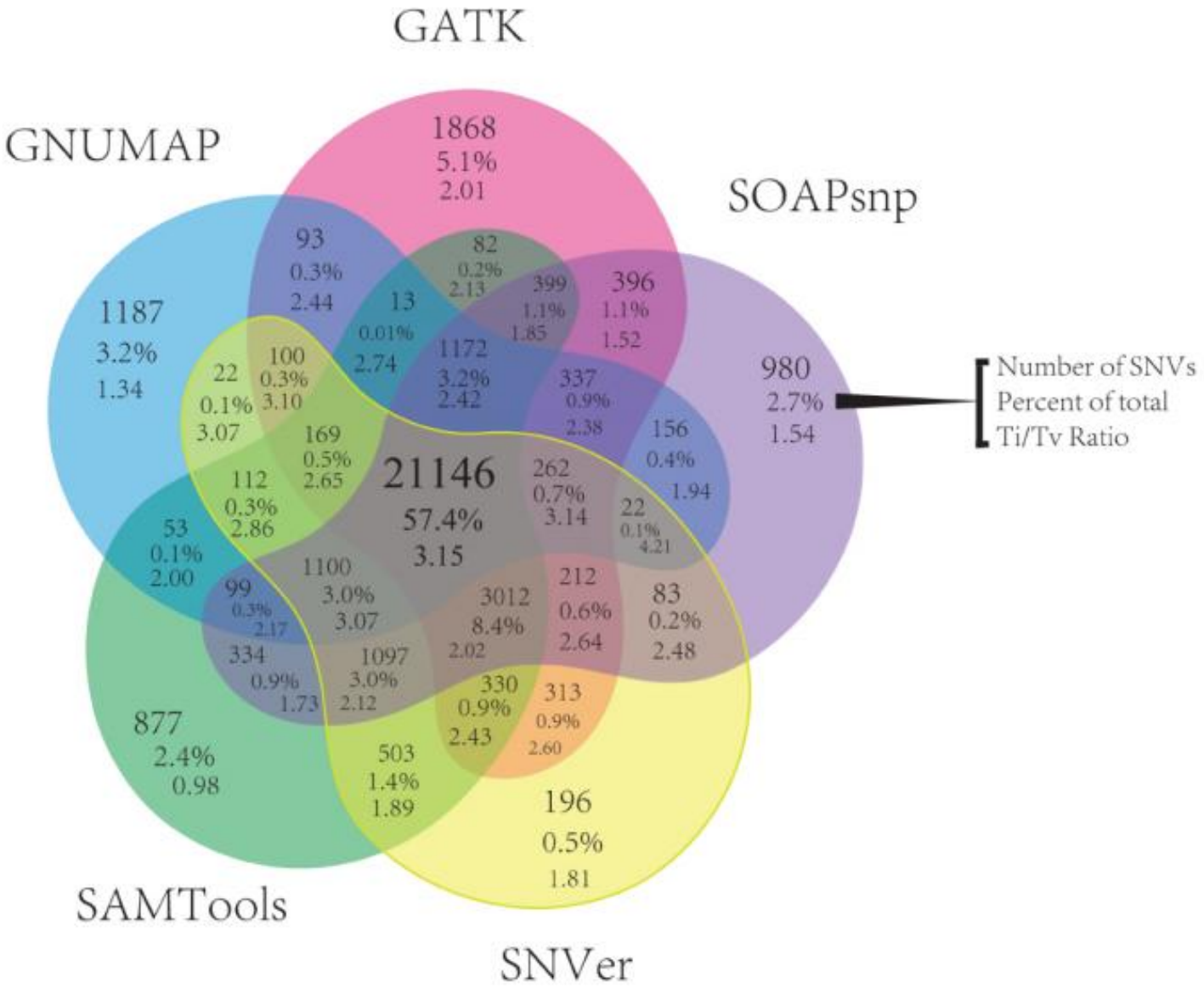
- **Trio check**:
  - Compare high quality variant calls between child and parents



What can go wrong? No sequence coverage



What can go wrong: uncalled variants



O'Rawe et al. Genome med. 2013

Radboudumc

What can go wrong: incorrect interpretation

Variant Pos	rs ID	Alleles	EA Allele #	AA Allele #	All Allele #	Avg. Sample Read Depth	Genes	mRNA Accession #	GVS Function	Amino Acid
20:31021163	rs145699348	A/G	A=1/G=8599	A=0/G=4406	A=1/G=13005	81	ASXL1	NM_015338.5	missense	ILE,VAL
20:31021190	unknown	T/C	T=1/C=8599	T=0/C=4406	T=1/C=13005	90	ASXL1	NM_015338.5	missense	CYS,ARG
20:31021211	unknown	T/C	T=0/C=8600	T=2/C=4404	T=2/C=13004	92	ASXL1	NM_015338.5	stop-gained	stop,ARG
20:31021232	rs148964601	T/C	T=1/C=8599	T=0/C=4406	T=1/C=13005	90	ASXL1	NM_015338.5	missense	CYS,ARG
20:31021233	rs143719307	A/G	A=0/G=8600	A=1/G=4405	A=1/G=13005	90	ASXL1	NM_015338.5	missense	HIS,ARG
20:31021250	unknown	T/C	T=1/C=8599	T=0/C=4406	T=1/C=13005	87	ASXL1	NM_015338.5	stop-gained	stop,ARG
20:31021324	unknown	C/T	C=0/T=8600	C=1/T=4405	C=1/T=13005	98	ASXL1	NM_015338.5	coding-synonymous	none
20:31021332	unknown	G/C	G=1/C=8599	G=0/C=4406	G=1/C=13005	97	ASXL1	NM_015338.5	stop-gained	stop,SER
20:31021337	unknown	A/G	A=1/G=8599	A=0/G=4406	A=1/G=13005	97	ASXL1	NM_015338.5	mis[illegible]e	ILE,VAL
20:31021384	unknown	A/G	A=2/G=8598	A=0/G=4406	A=2/G=13004	102	ASXL1	NM_015338.5	coding[illegible]ous	none
20:31021389	unknown	A/G	A=1/G=8599	A=1/G=4405	A=2/G=13004	104	ASXL1	NM_015338.5	[illegible]	ASN,SER
20:31021430	rs141346625	C/G	C=2/G=8598	C=17/G=4389	C=19/G=12987	107	ASXL1	NM_015338.5	[illegible]	GLN,GLU
20:31021466	rs142172134	G/C	G=0/C=8600	G=10/C=4396	G=10/C=12996	106	ASXL1	NM_015338.5	[illegible]	GLY,ARG
20:31021475	rs145913172	C/G	C=0/G=8600	C=2/G=4404	C=2/G=13004	109	ASXL1	NM_015338.5	[illegible]	PRO,ALA
20:31021521	rs138971201	A/T	A=0/T=8600	A=14/T=4392	A=14/T=12992	131	ASXL1	NM_015338.5	[illegible]	ASN,ILE
20:31021544	unknown	A/G	A=0/G=8600	A=1/G=4405	A=1/G=13005	140	ASXL1	NM_015338.5	[illegible]	MET,VAL

Variation Color Code:
splice or nonsense or frameshift
missense
coding-synonymous
coding
utr
codingComplex

Bohring-Opitz syndrome is often fatal in early childhood.

http://evs.gs.washington.edu/EVS/

Better diagnostics by whole exome sequencing?

-

Pilot for the gene panel approach

Pilot study – gene package approach

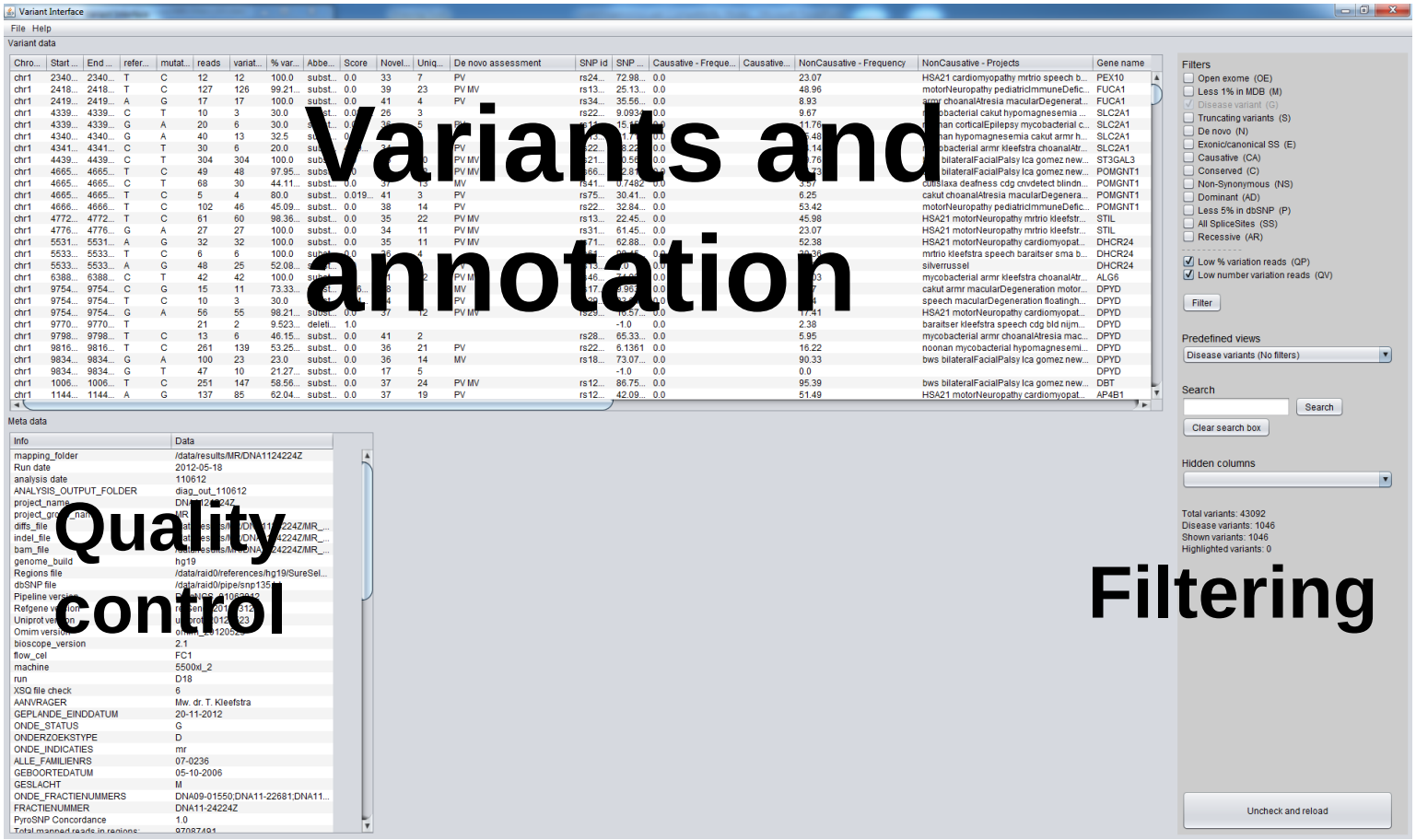
- **250 exomes**: 50 exomes for 5 genetically heterogeneous diseases

- Gene package design:
  - Only known genes are allowed, no candidate disease genes
  - Gene lists must be up-to-date and is updated every ~3 months
  - Created by team of experts from clinic, diagnostic and research division

	Number of genes (Sept. 2011)
Blindness	144
Deafness	98
Early onset colorectal cancer	115
Mitochondrial disorders	207
Movement disorders	152

Neveling *et al*. Hum mut. 2013

How to do 400 samples per month?

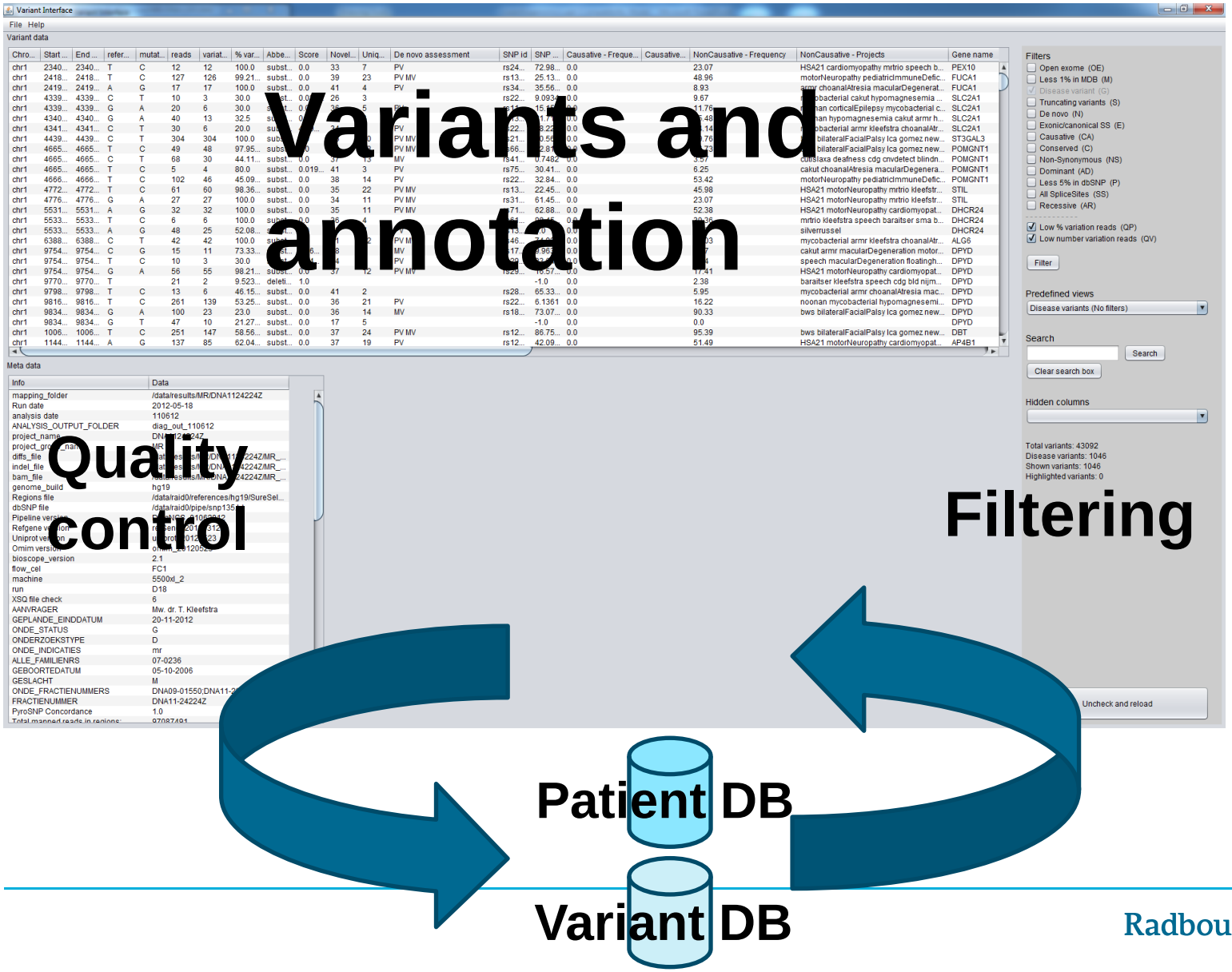


Variants and annotation

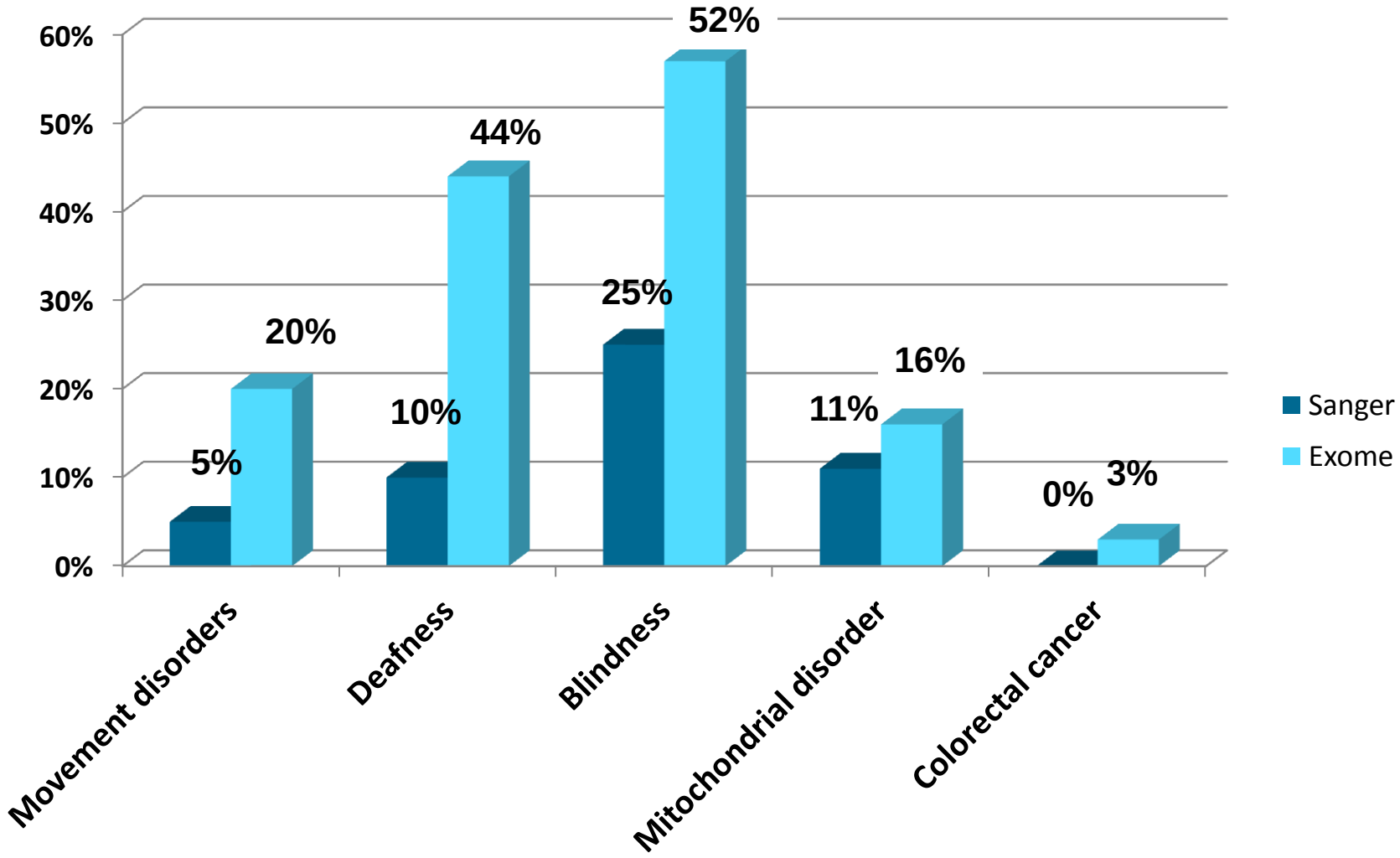
Quality control

Filtering

How to do 400 samples per month?



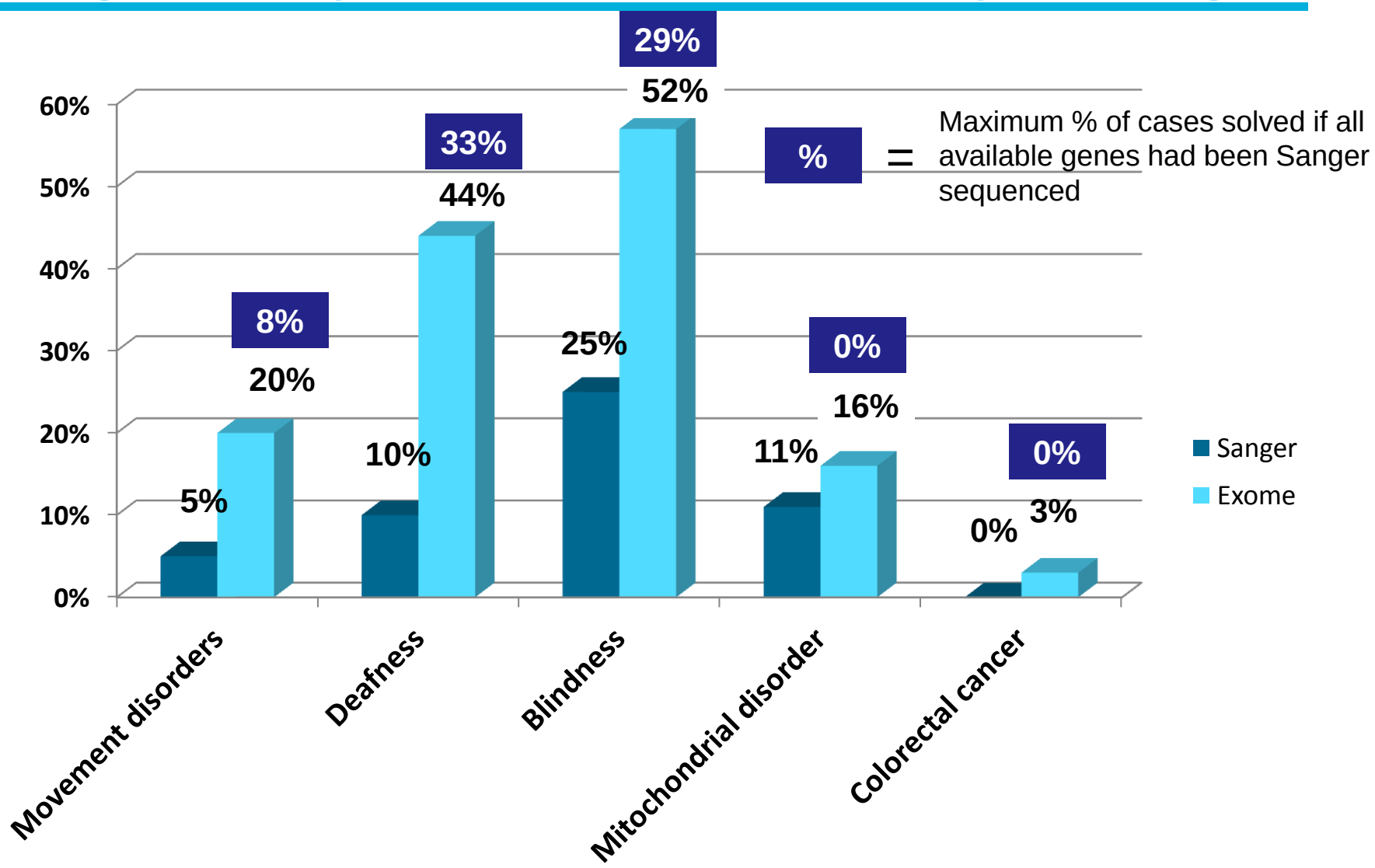
Diagnostic yield from exome sequencing



Neveling *et al*. Hum mut. 2013

Radboudumc

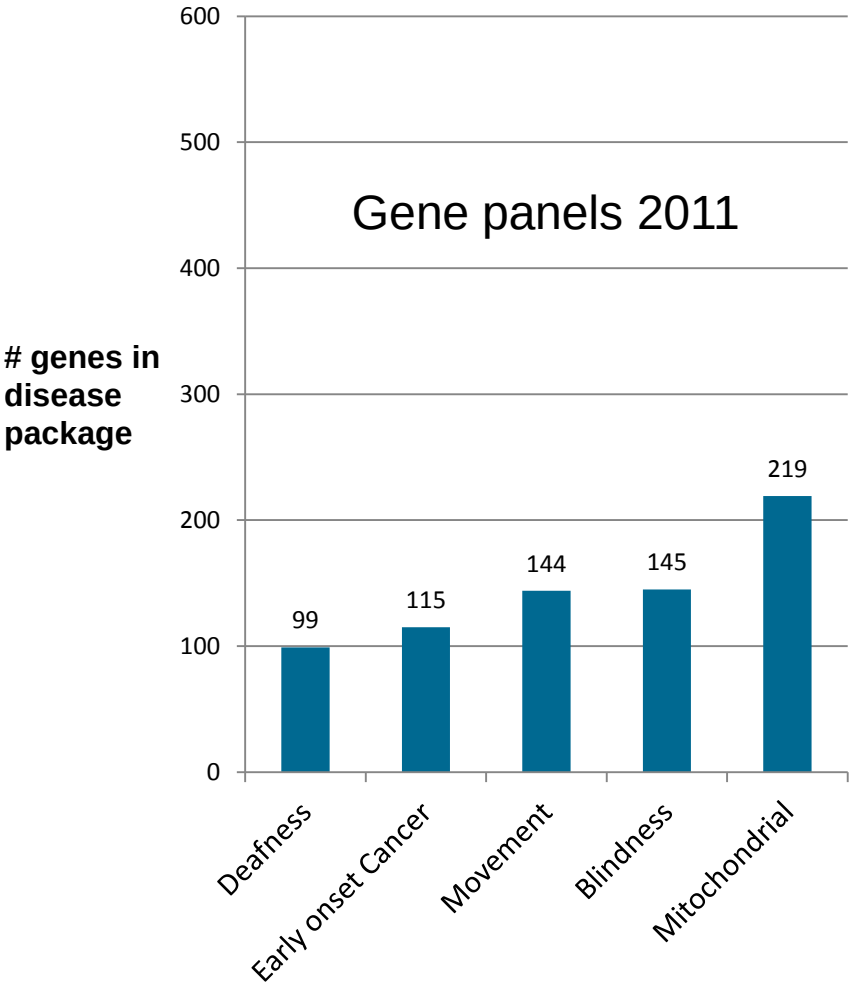
Diagnostic yield from exome sequencing



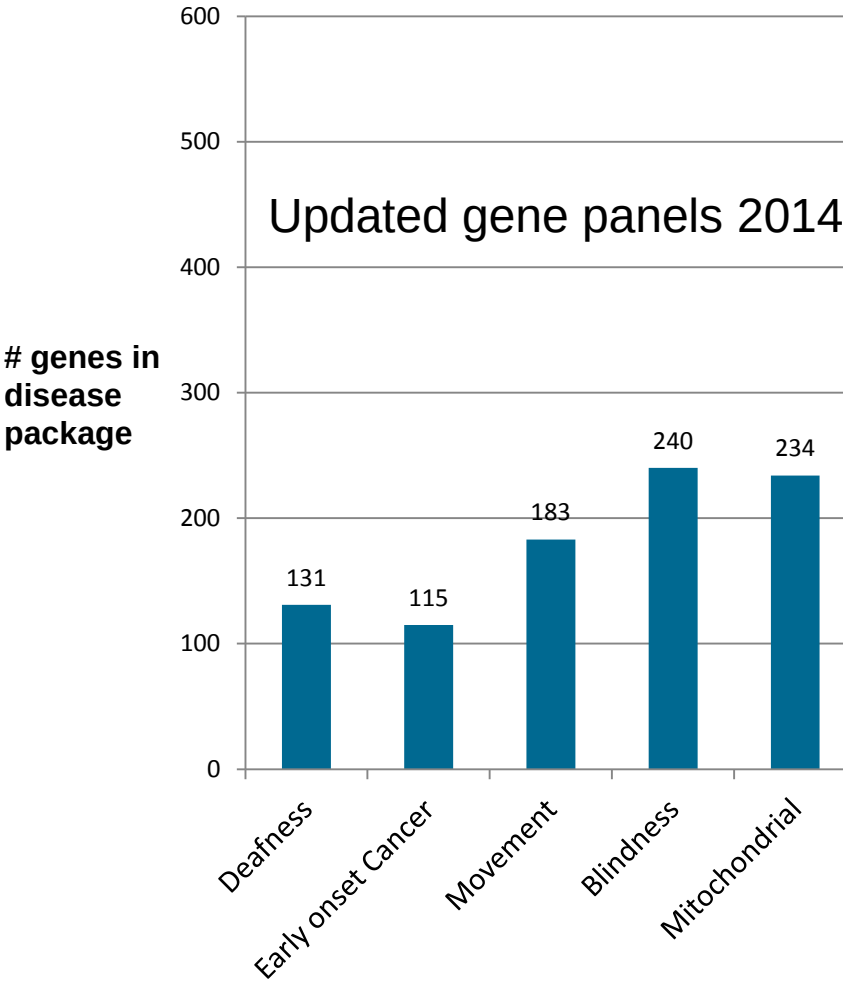
Neveling *et al*. Hum mut. 2013

Radboudumc

Current exome panels

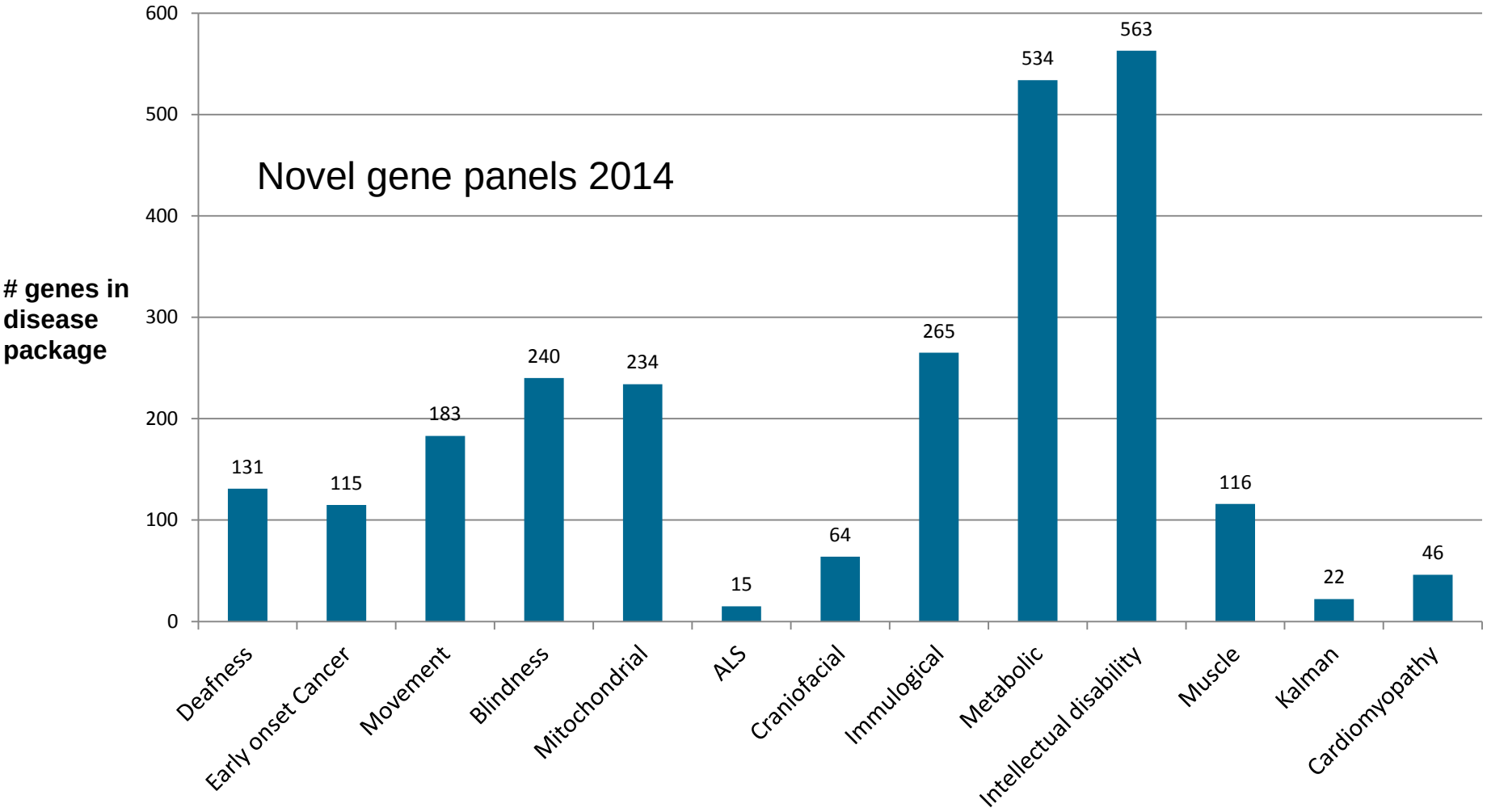


Current exome panels



Exome sequencing can be cost-efficient compared to Sanger when sequencing 3 genes or more..

Current exome panels

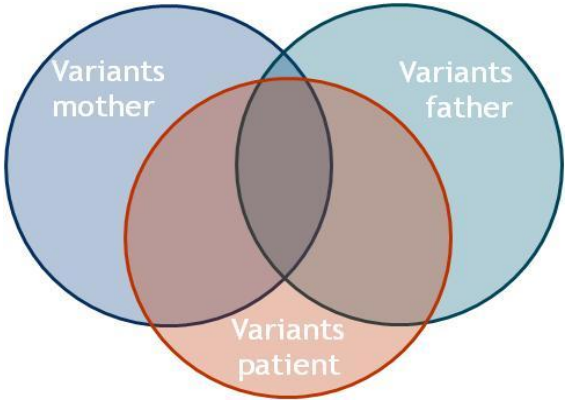


Better diagnostics by whole exome sequencing?

-

Pilot for the trio approach

Pilot study – *de novo* approach



- **100 patients + 200 parents!**
  - Severe intellectual disability (IQ<50)
  - No etiological or syndromic diagnosis
  - Negative family history

- Patients have reached the end stage of conventional strategies
  - Targeted gene tests negative
  - Genomic array profile negative

Yield in 100 ID patients

Positive diagnosis	June 2012	June 2013
All mutations	**16**	**29**
***De novo* mutations**	**13**	**28**
Autosomal dominant	10	23
X-linked	2	4
Autosomal recessive	1	1
Inherited mutations	**3**	**1**
X-linked	3	1
Autosomal recessive	0	0
Candidates	19	11

Yield of ~30% in patients with severe ID

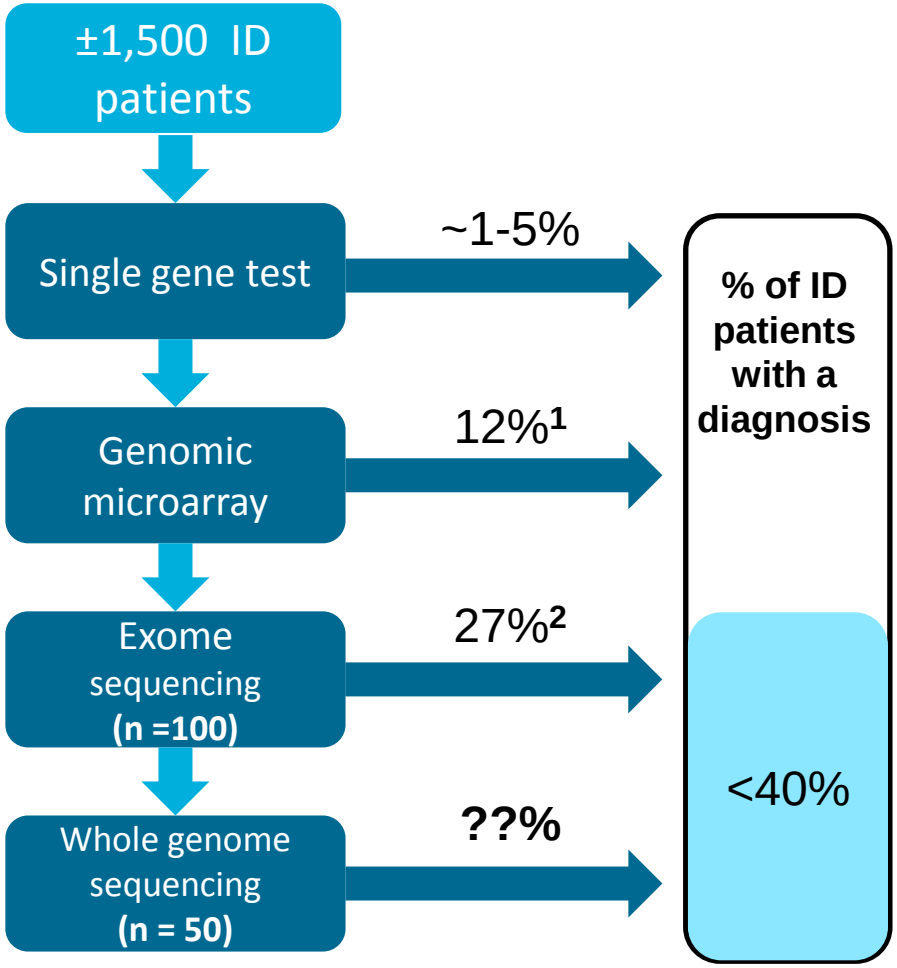
De Ligt *et al.* NEJM, 2012

Better diagnostics by whole genome sequencing?

-

pilot study

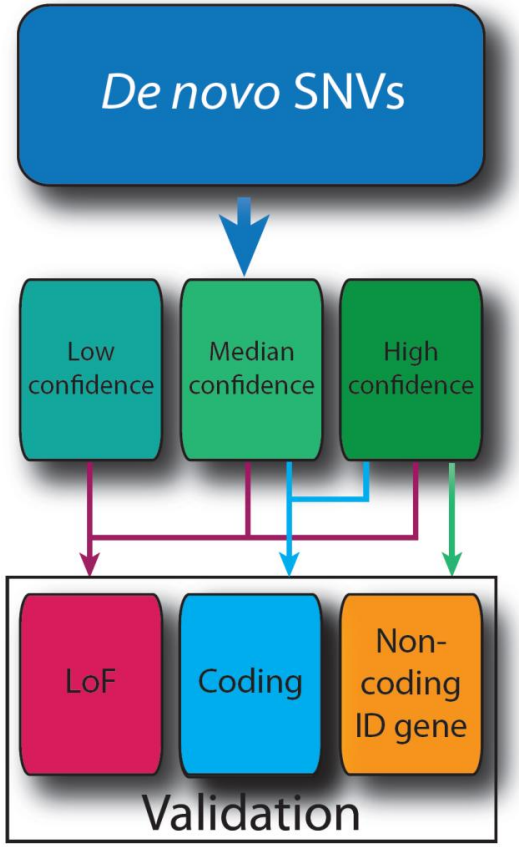
Diagnosis in patients with severe intellectual disability (ID)



- Whole genome sequencing
- 50 trios at 80x coverage
- A *de novo* approach

[1] Vulto-van Silfhout, A. T. *et al* Hum mut 2013
[2] de Ligt, J. *et al.* NEJM 2012

Can we identify *de novo* mutations?

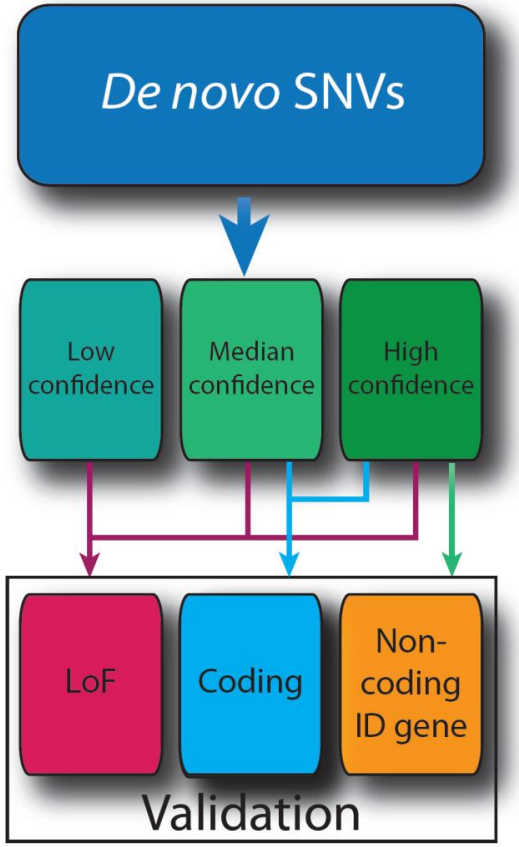


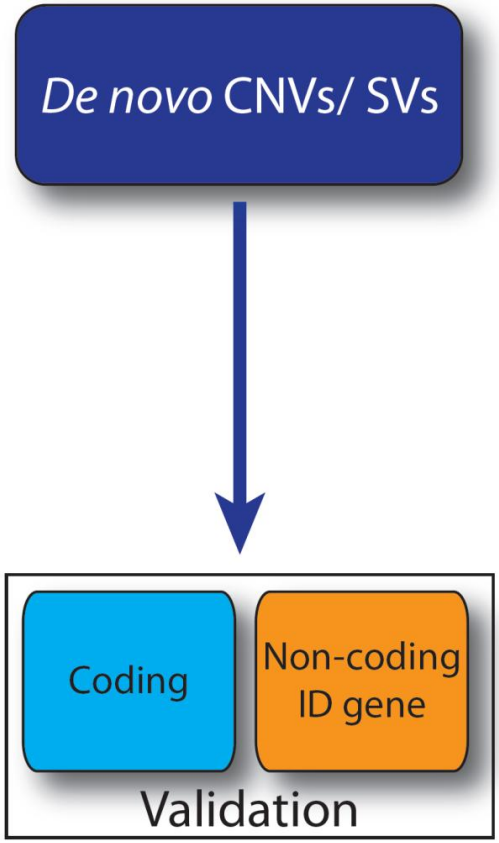
- Validation rate: **38%**

(80% for high confidence!)

- Coding *de novo* SNVs: **84**
- Comparison to WES SNVs: **12%**

Can we identify *de novo* mutations?





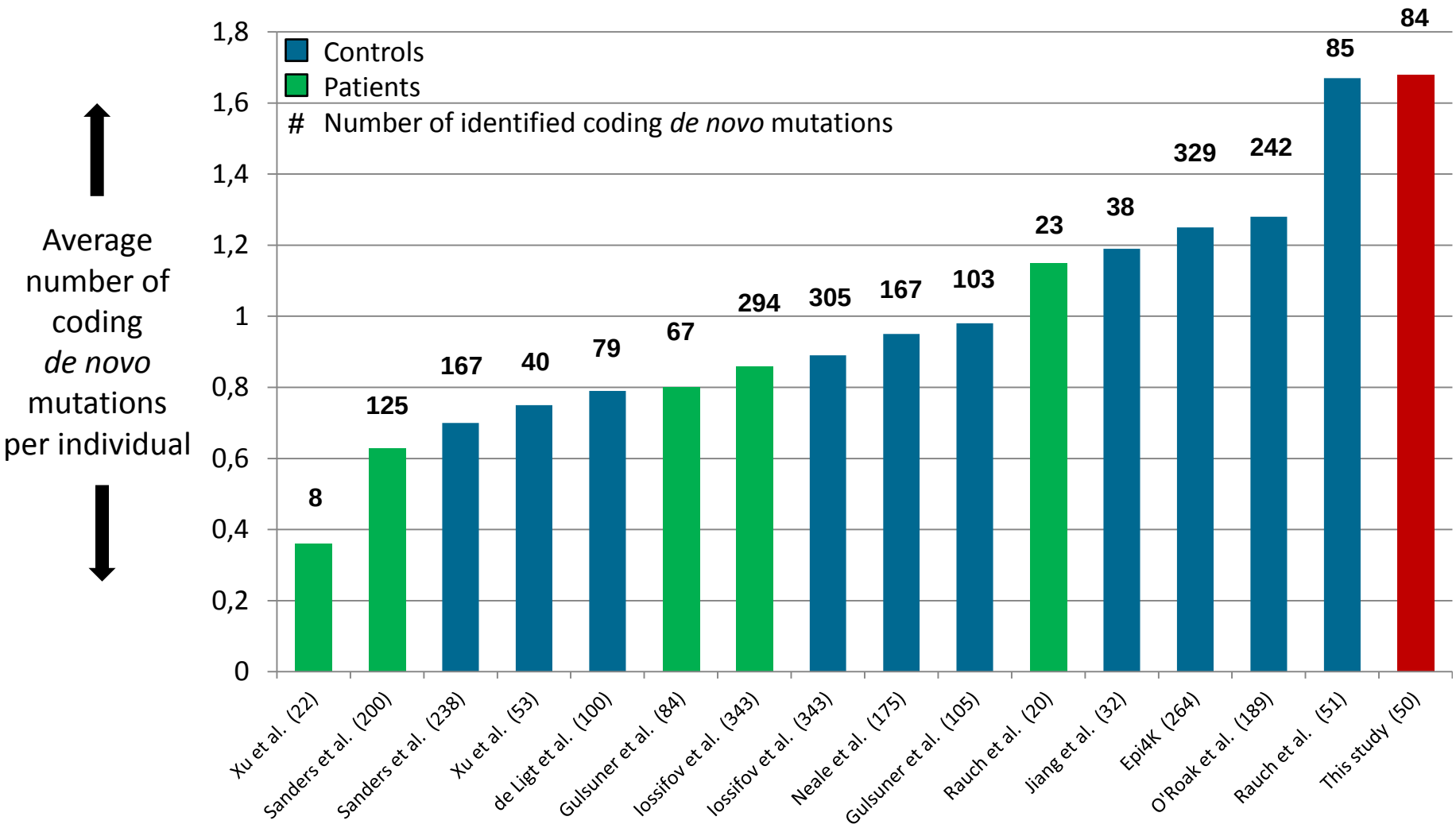
- Validation rate: **38%**

(80% for high confidence!)

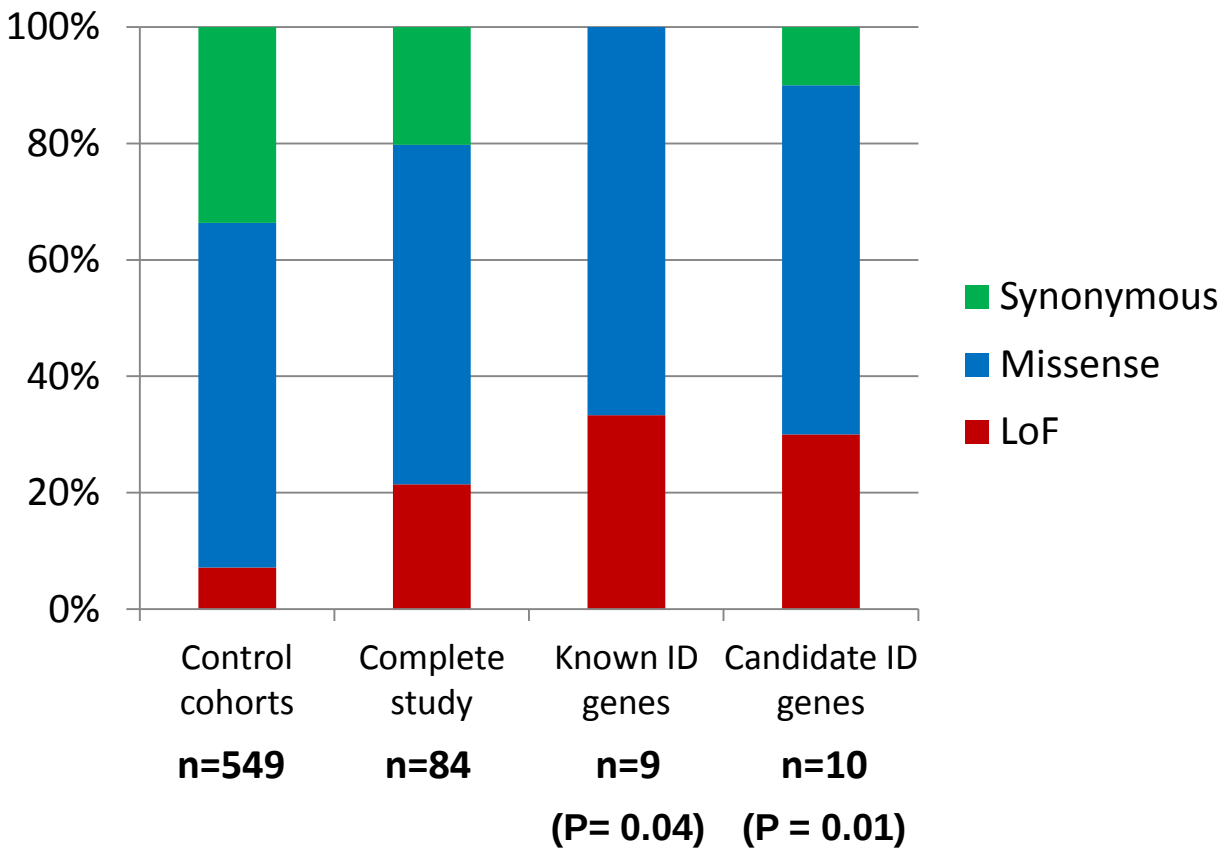
- Coding *de novo* SNVs: **84**
- Comparison to WES SNVs: **12%**

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

Did we find more coding *de novo* mutations than other studies?



Are these *de novo* SNVs related to ID?

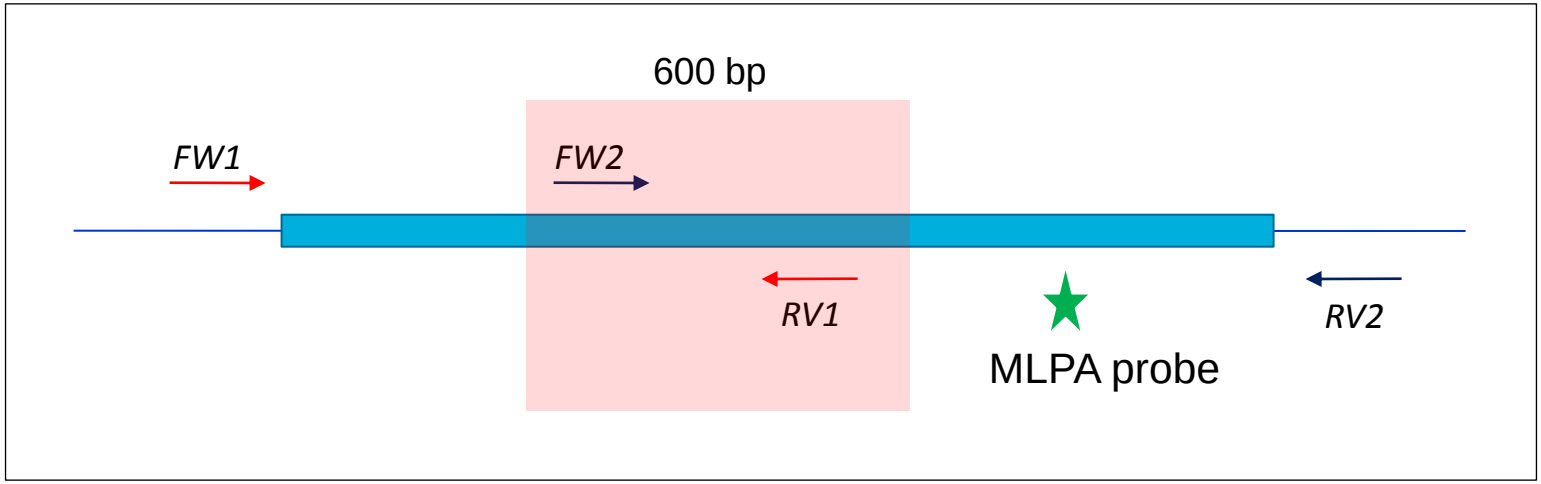


- Significantly more *de novo* mutations in known/candidate ID genes
- Significantly more loss-of-function mutations ($P=4.8^{-06}$, P=0.02)

De novo structural variants, example 1

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

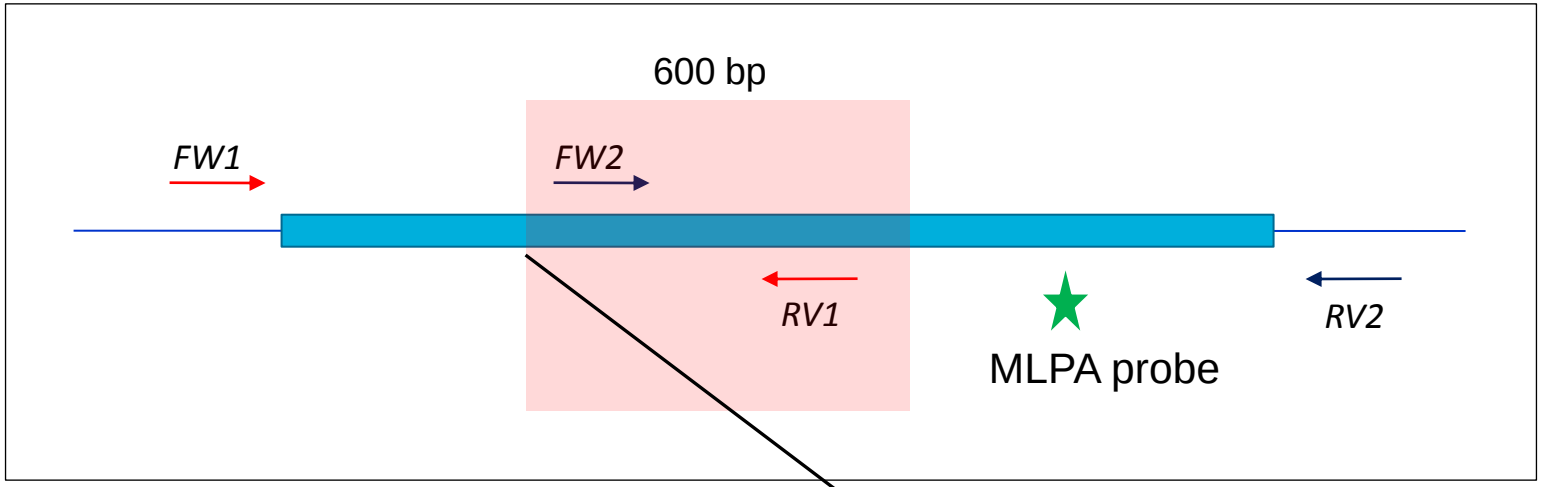
MECP2, exon 4 (~1000 bp)

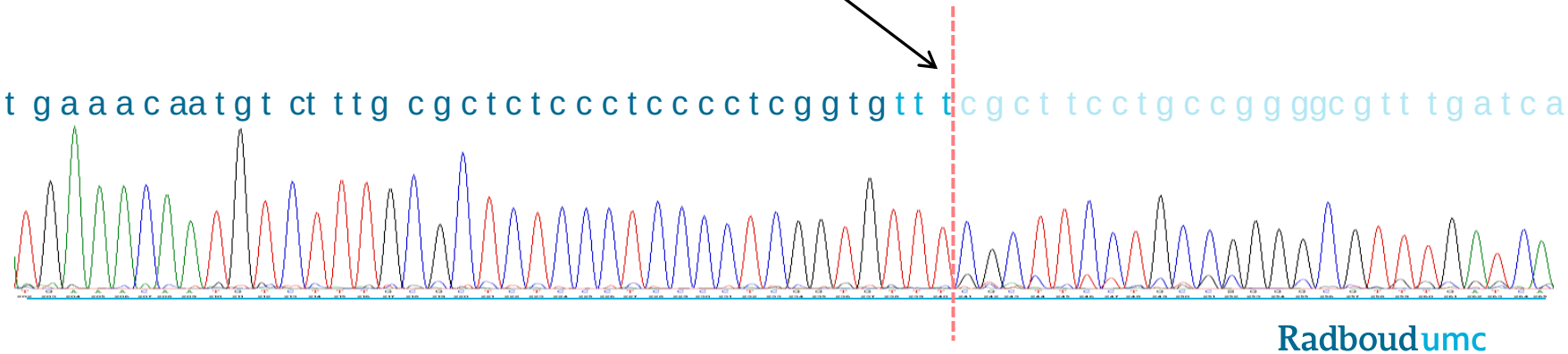


De novo structural variants, example 1

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

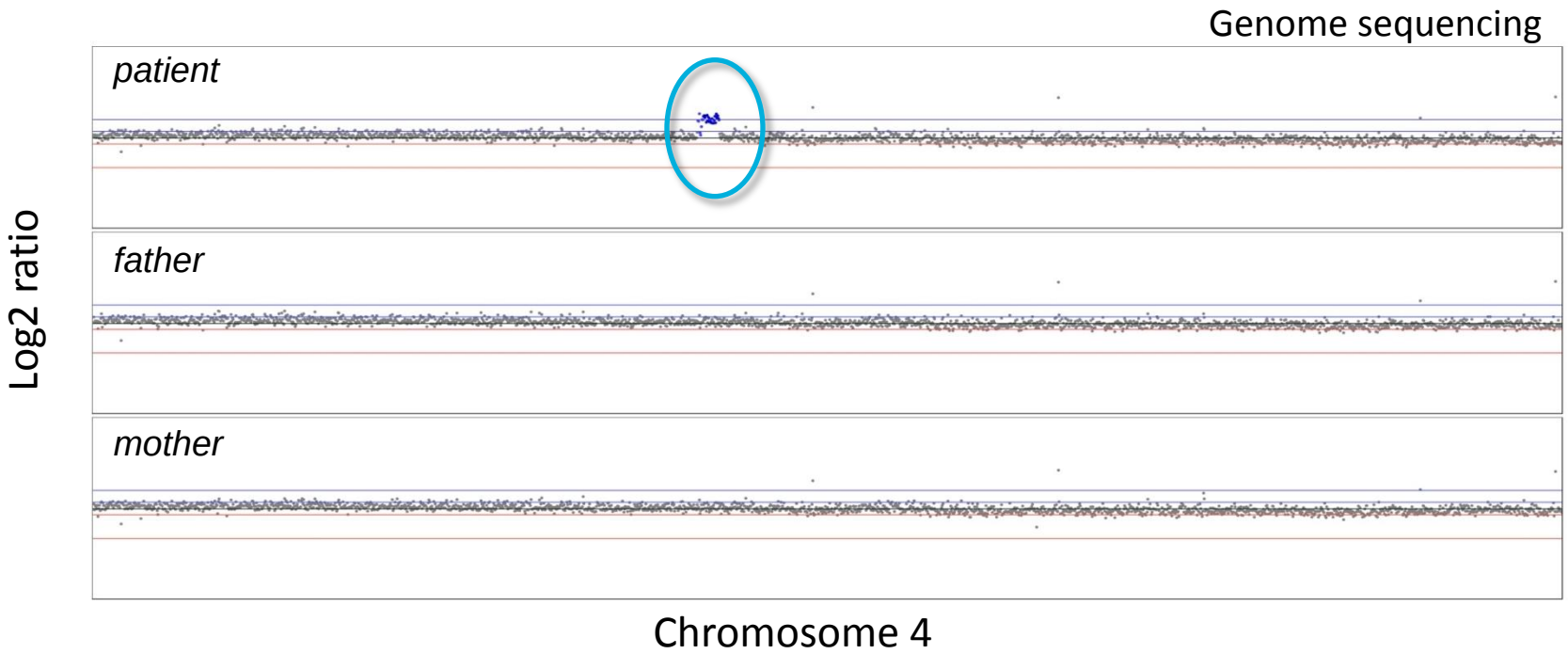
MECP2, exon 4 (~1000 bp)





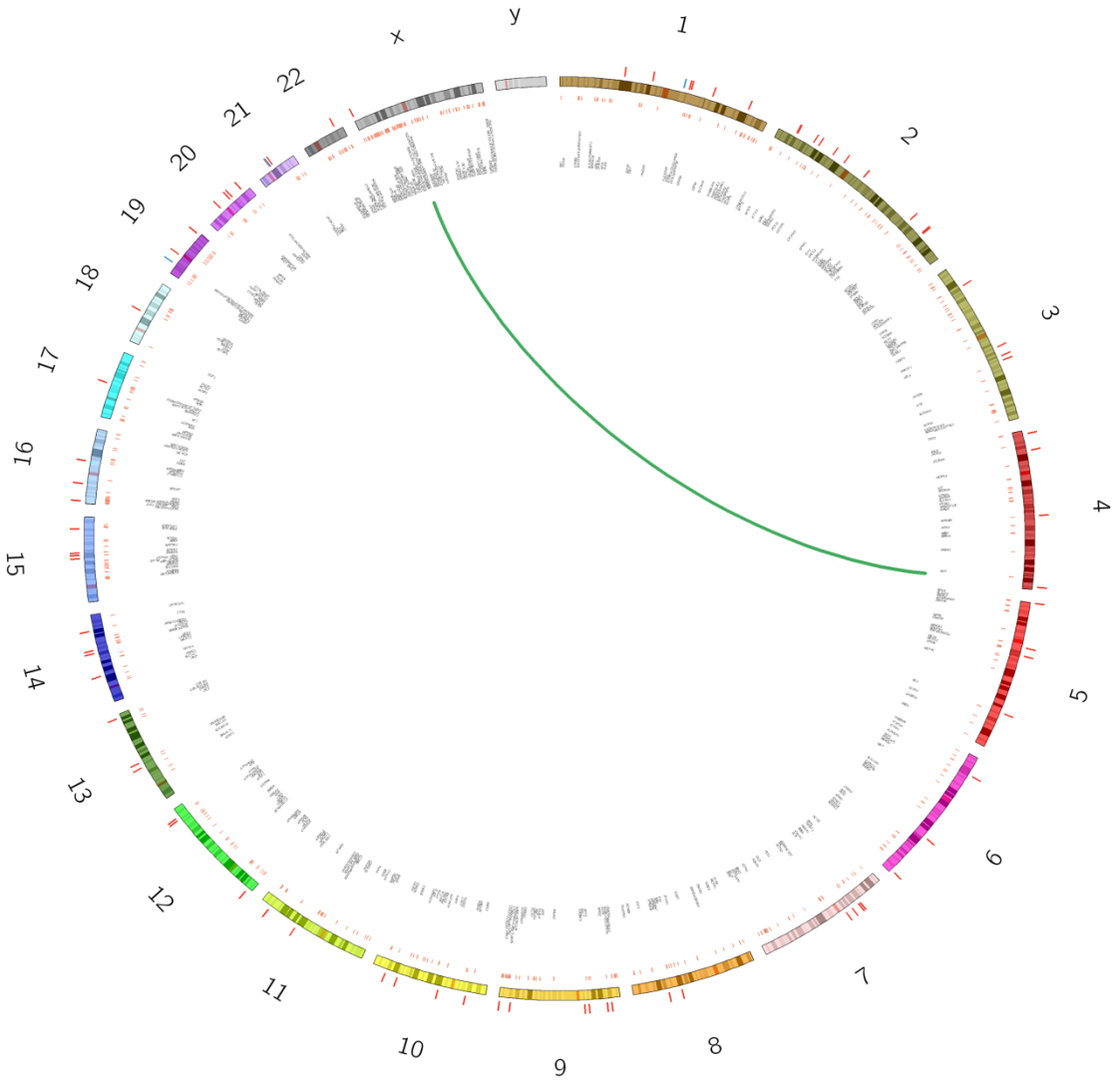
Radboudumc

Example 2: A *de novo* duplication on chromosome 4

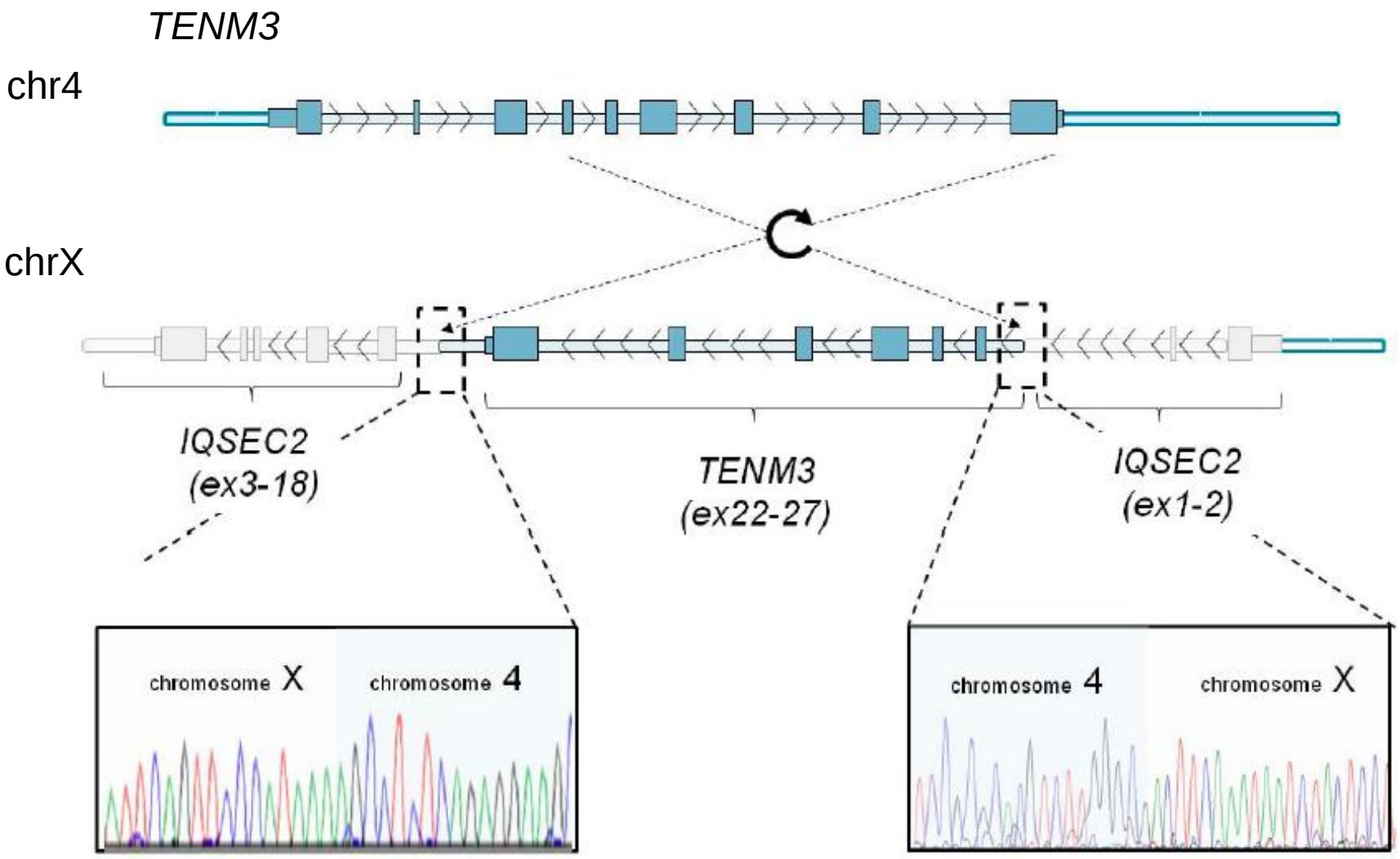


- ~60 kb *de novo* duplication on chromosome 4
- Affecting the last 6 exons of the *TENM3* gene
- *TENM3* is associated with coloboma, and microphthalmia

Duplication from chr4 to chrX



Duplication from chr4 to chrX



Only possible with genome sequencing!

Radboudumc

Diagnostic yield genome sequencing

Highly likely diagnosis[1]

SNV	SV
TBR1 (2x)	***SHANK3***
WDR45	***VPS13B****
SMC1A	***MECP2***
SPTAN1	***IQSEC2***
RAI1	***STAG1***
MED13L	***SMC1A***
SATB2	16p11.2 microdel. syndrome
PPP2R5D	Multiple genes
KCNA1	
SCN2A	
POGZ	
KANSL2	

21/50 cases diagnosed: **42%**

* Recessive variants. Known genes in bold.

[1] de Ligt *et al*. NEJM 2012.

Radboudumc

Diagnostic yield genome sequencing

Highly likely diagnosis[1]		Possible diagnosis
SNV	**SV**	**SNV**
TBR1 (2x)	***SHANK3***	*NGFR*
WDR45	***VPS13B****	*GFPT2*
SMC1A	***MECP2***	*WWP2*
SPTAN1	***IQSEC2***	*ASUN*
RAI1	***STAG1***	*BRD3*
MED13L	***SMC1A***	*MAST1*
SATB2	16p11.2 microdel. syndrome	*APPL2*
PPP2R5D	Multiple genes	*NACC1*
KCNA1		
SCN2A		
POGZ		
KANSL2		

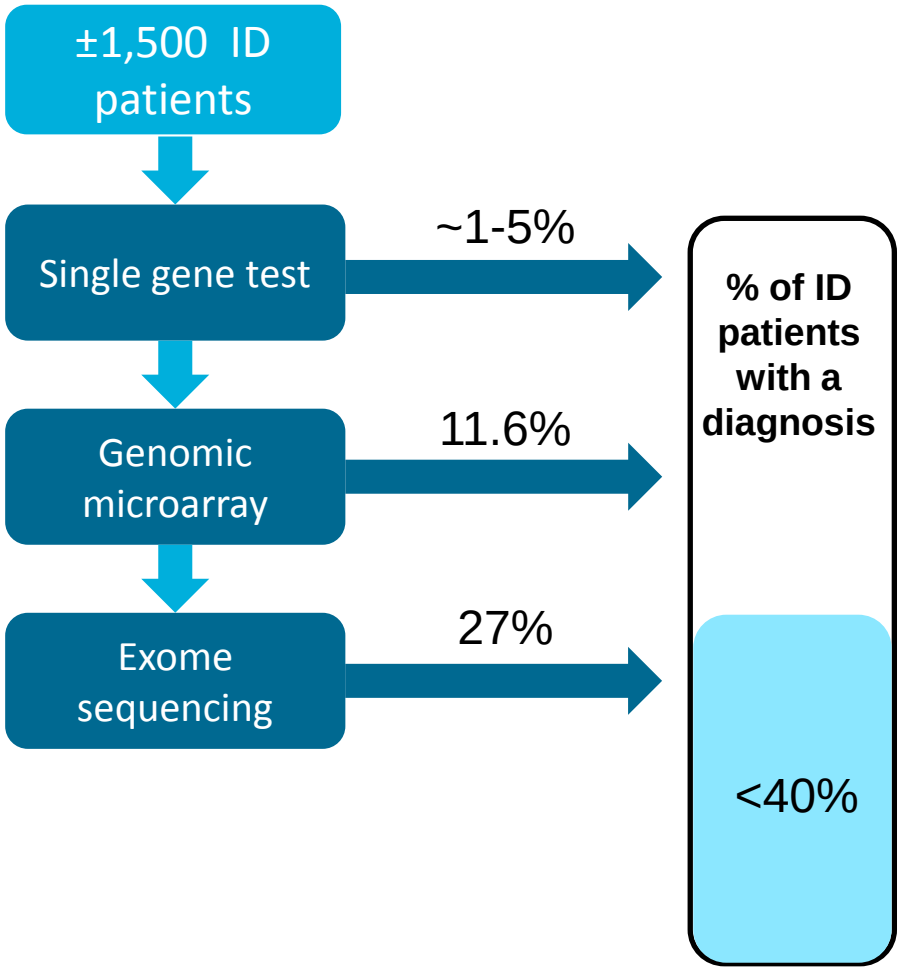
21/50 cases diagnosed: **42%**

* Recessive variants. Known genes in bold.

[1] de Ligt *et al*. NEJM 2012.

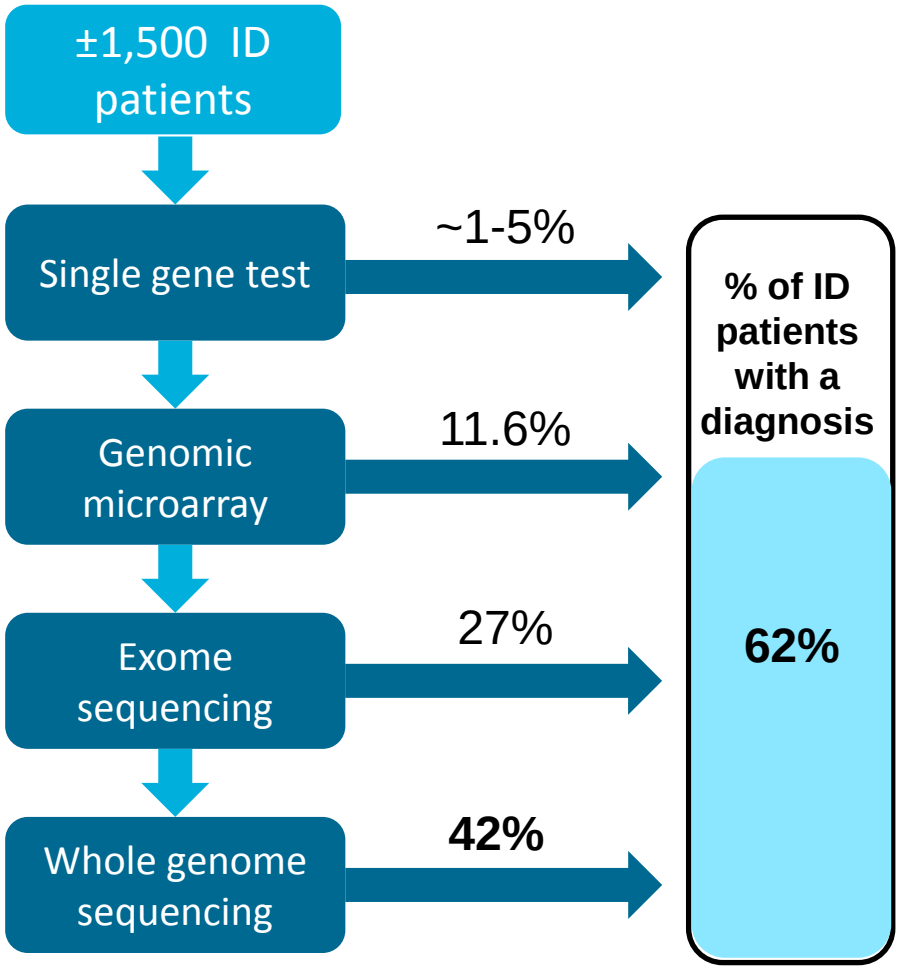
Radboudumc

Diagnostic yield



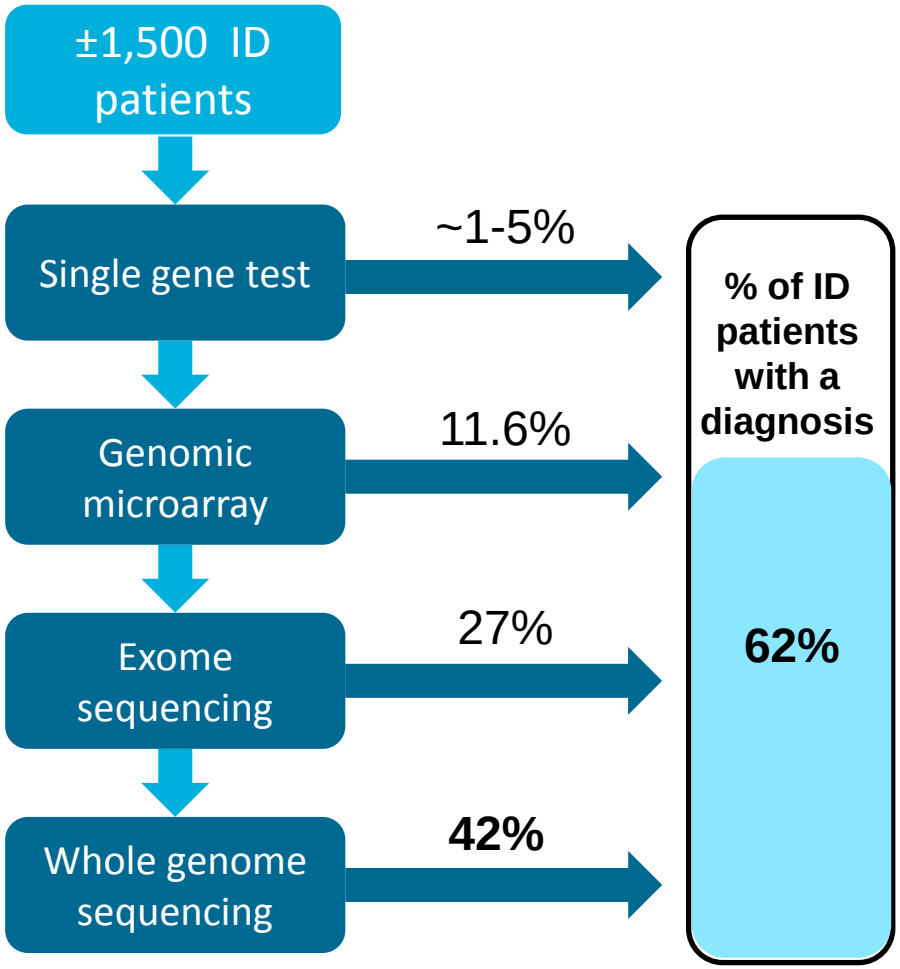
Gilissen *et al*. Nature 2014

Diagnostic yield

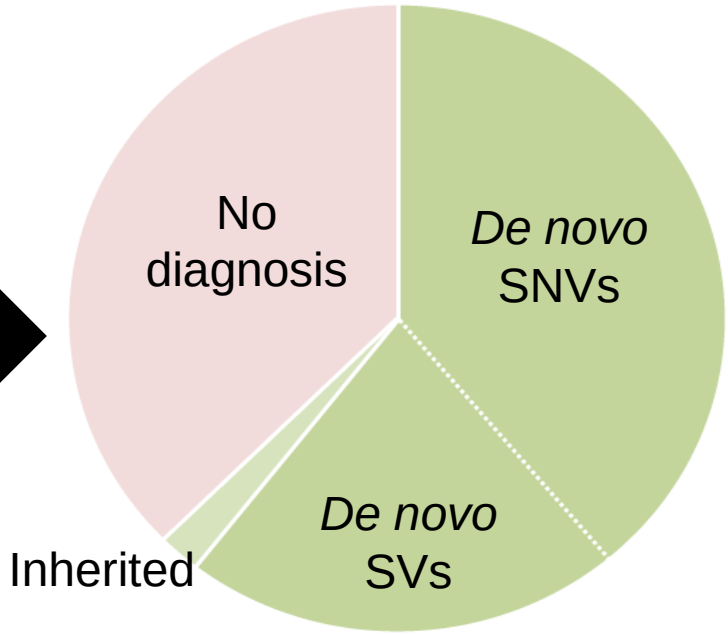


Gilissen *et al*. Nature 2014

Diagnostic yield



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



Gilissen *et al*. Nature 2014

Conclusions

1. Exome sequencing results in a higher diagnostic yield for genetically heterogeneous diseases than Sanger-based approaches

2. Based on clinical diagnostic criteria we can provide a genetic diagnosis for the majority of severe ID cases by using WGS

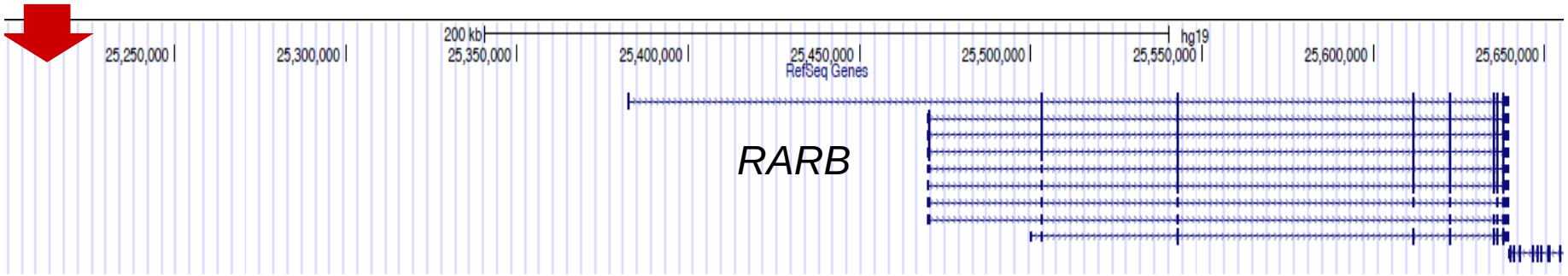
1. *De novo* coding mutations are the major cause of severe ID

Discussion

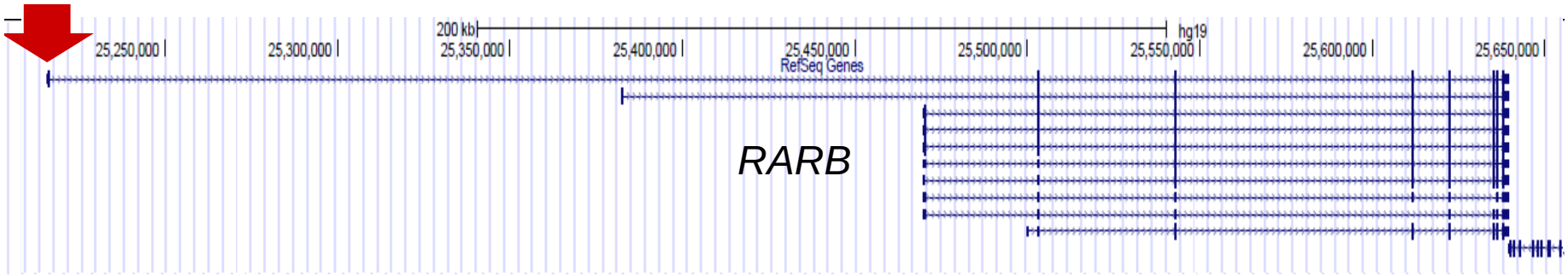
Other advantages of genome sequencing:

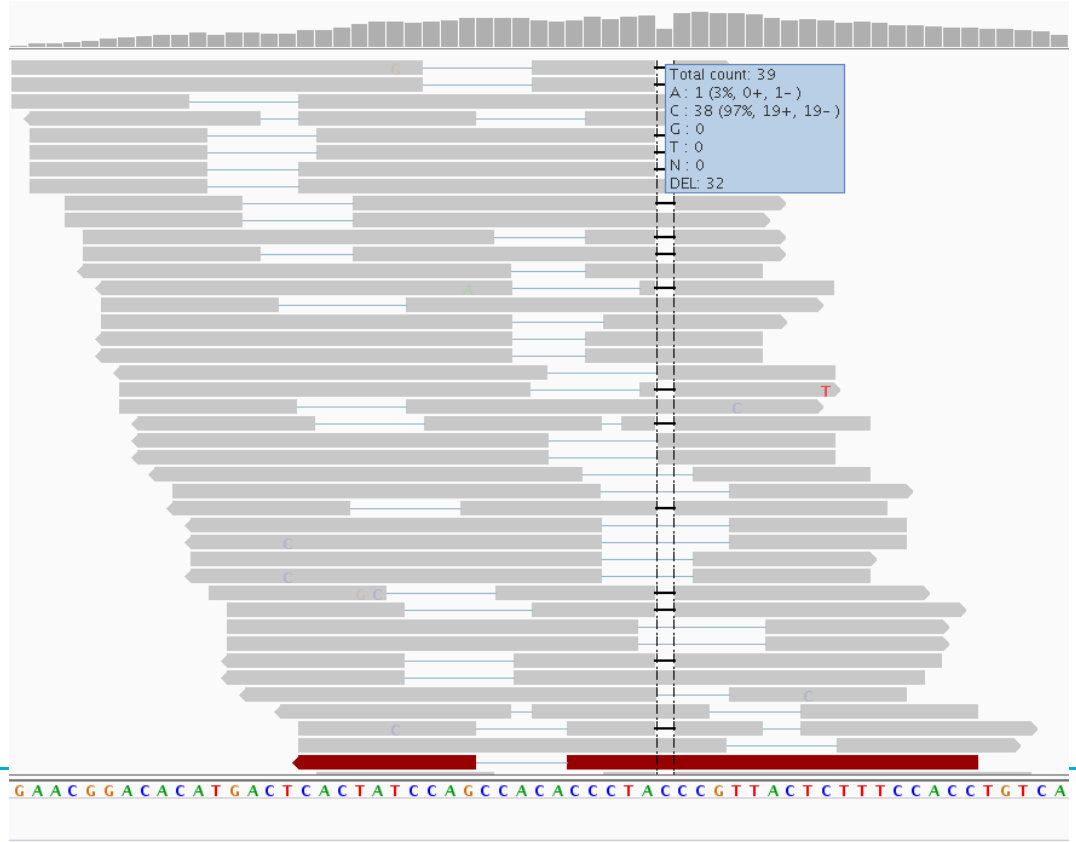
- Comprehensive set of variants: novel genes/exons annotations are still being added
- Non-coding variants: *de novo* mutations introducing novel splice sites
- Mosaic variants: variants occurring in sub-population of cells
- Phasing: identifying whether two variants are on the same allele

One more coding *de novo* variant?

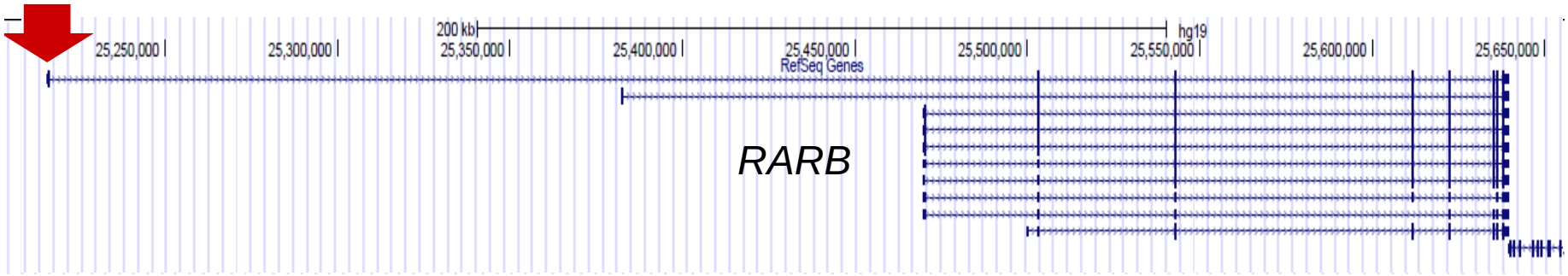


One more coding *de novo* variant?





One more coding *de novo* variant?



- Mutations in the original *RARB* transcript are a known cause for Microphthalmia
- Validation confirmed: coding *de novo* mutation #85
- Expression analysis: this specific exon is expressed in brain / fetal brain.

Acknowledgments

Lisenka Vissers

Kornelia Neveling

Alex Hoischen

Jayne Hehir-Kwa

Joris Veltman

Marcel Nelen

Hans Scheffer

Han Brunner

Genome Diagnostics

Helger Yntema
Erik-Jan Kamsteeg
Lies Hoefsloot
Willy Nillesen
Marjolijn Ligtenberg
Arjen Mensenkamp
Dorien Lugtenberg
Rolph Pfundt

Genome Research

Djie Thung
Maartje van de Vorst
Rick de Reuver
Marisol del Rosario
Nienke Wieskamp
Petra de Vries
Michael Kwint
Irene Janssen
Marloes Steehouwer

Clinical genetics

Marjolein Willemsen
Tjitske Kleefstra
Ernie Bongers
David Koolen
Anneke Vulto-van Silfthout
Wendy van Zelst-Stams

erc European Research Council

TECHGENE

Complete genomics

NWO Netherlands Organisation for Scientific Research

SEUVADIS

Diagnostic exome sequencing in 100 ID trios

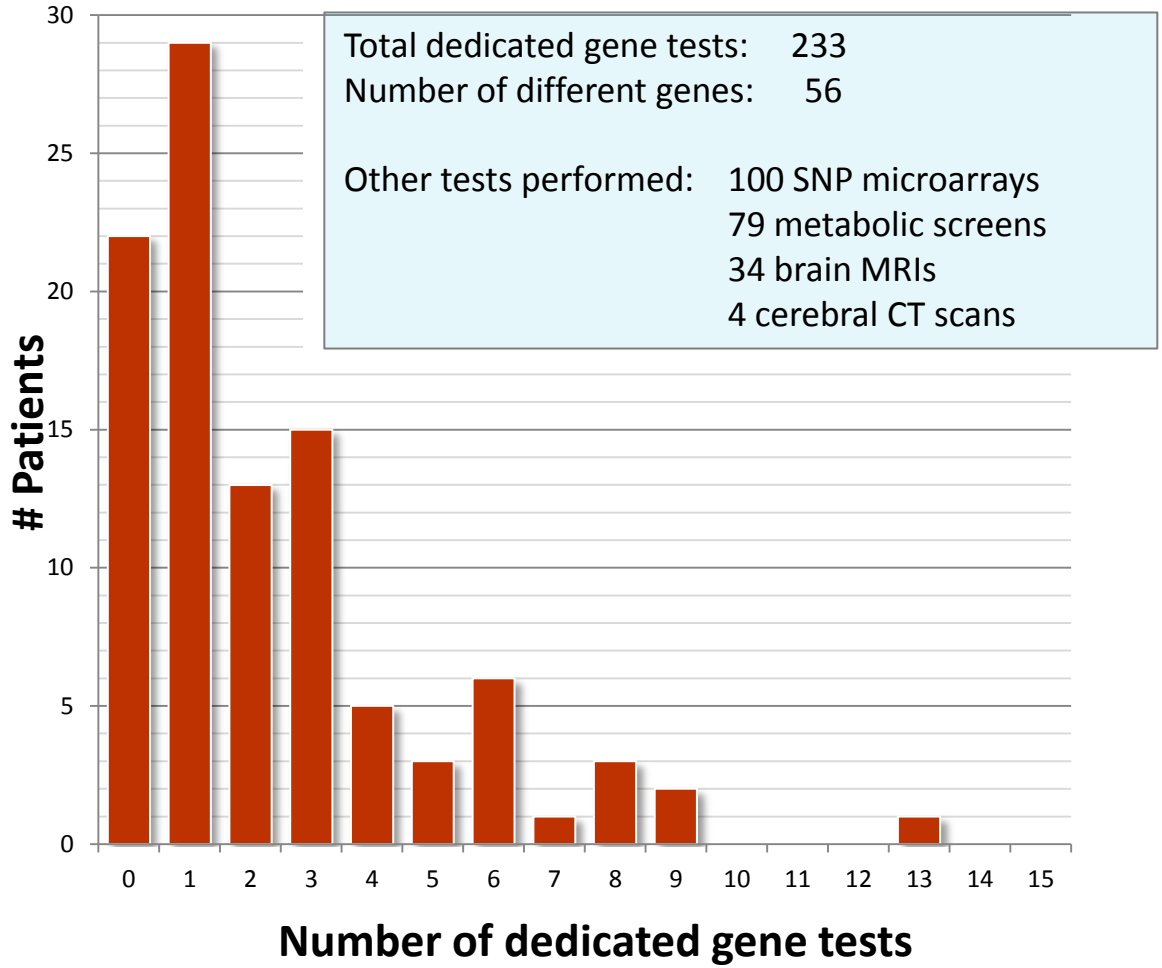
Level of ID	Number of patients
IQ <30	62
IQ 30-50	38
IQ 50-70	0

Gender	
Male	47
Female	53

Age groups	
<10 yrs	37
10-20 yrs	41
>20 yrs	22

Sibship size	
1	12
2	47
3	36
4	1
≥5	2
unknown	2

Number of major congenital anomalies	
0	62
1	31
2	7
3	0



Radboudumc

De Ligt *et al.* NEJM, 2012