The circle of NGS innovation

from research to diagnostics and back

RE(ACT) Congress 2018

Lisenka ELM Vissers, PhD
PI Translational Genomics,
Radboudumc, The Netherlands
Lisenka.Vissers@radboudumc.nl

Radboudumc

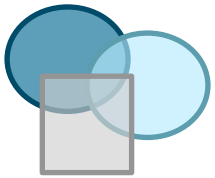
Next generation sequencing innovation

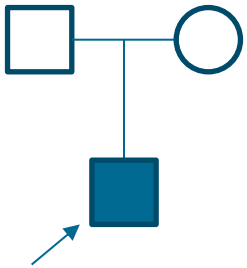
2010

2012-2013

2014

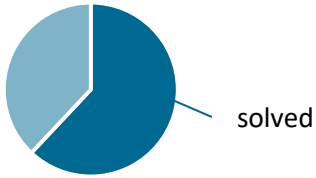
Momentum for
Rare Disease Gene Discovery





Diagnostic use of WES for RD

De Ligt et al. NEJM 2010 -- ID
Rauch et al. Lancet 2012 -- ID
Neveling et al. Hum Mut 2013 -- many
Yang et al. NEJM 2013



A de novo paradigm for ID

Vissers et al. Nat Genetics 2010

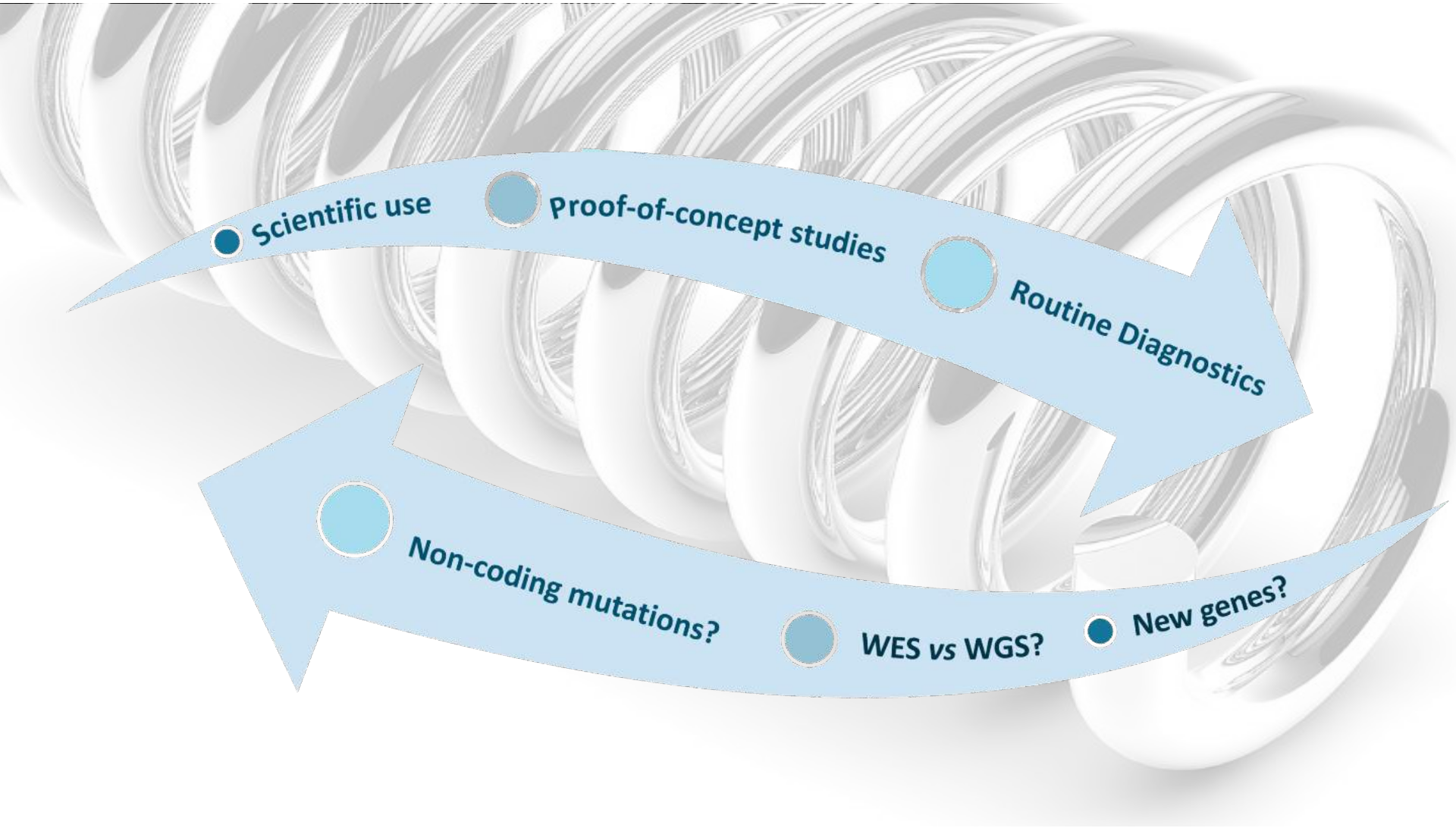
Gene discovery by exome sequencing for dominant disorders

Ng et al. Nat Genetics 2010
Hoischen et al. Nat Genetics 2010

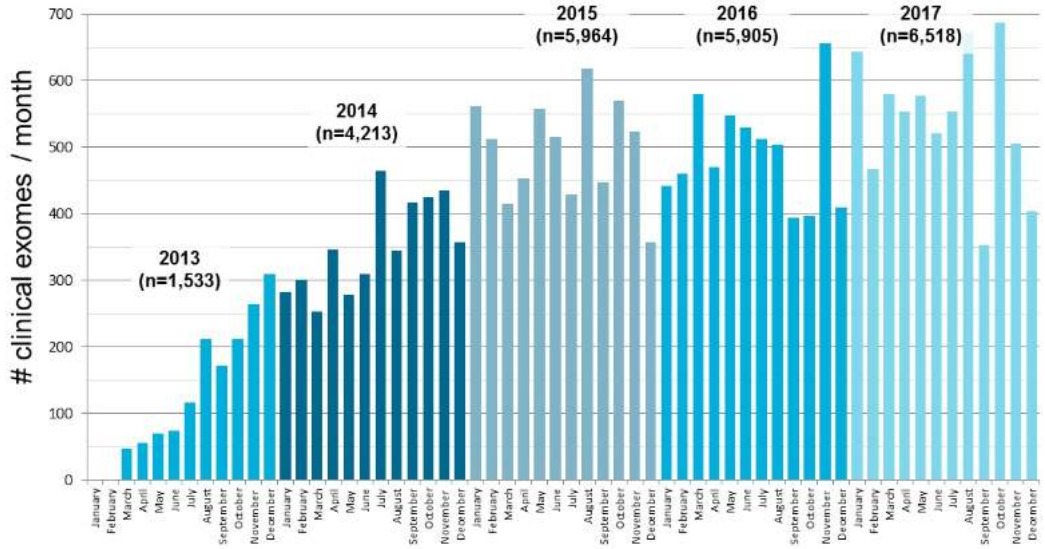
Diagnostic use of WGS for ID

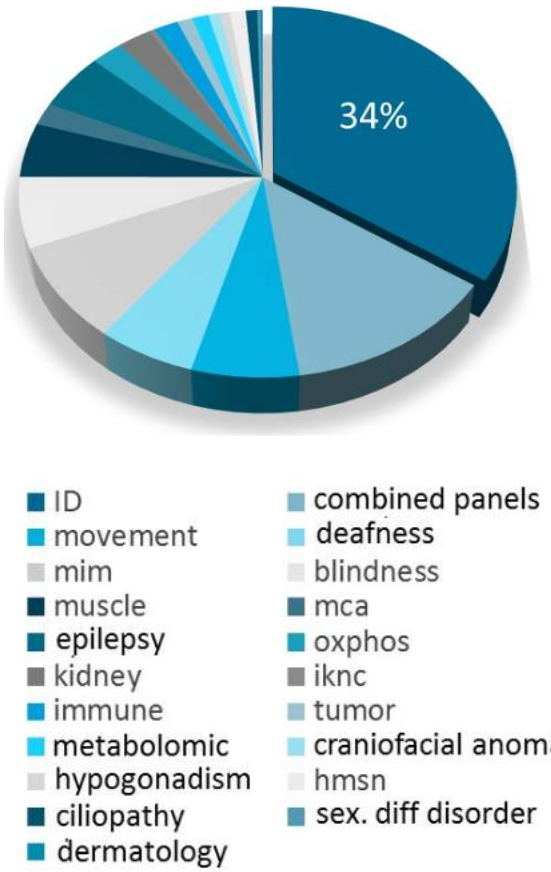
Gilissen et al. Nature 2014

From research to clinic and back



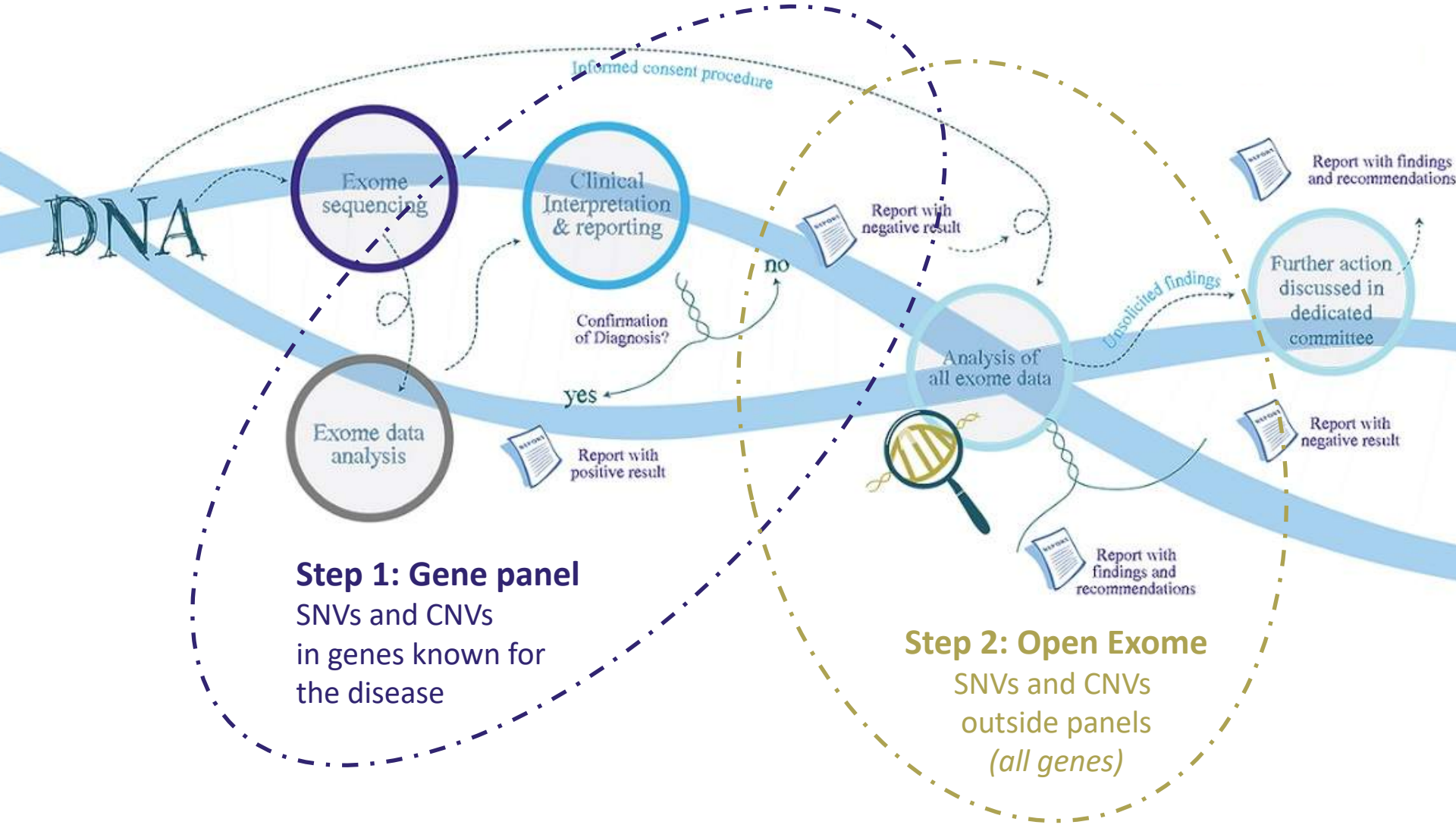
Clinical WES @Radboudumc since 2012





- 25 different gene panels
- 3,742 unique gene entries
- **>20,000 index cases tested**

Clinical WES analysis: a two step process



Why (do) we need exome sequencing (?)

A 5 year clinical utility study in pediatric neurology

Radboudumc

From last resource to first tier test?

150 consecutive patients consulted in clinical practice of pediatric neurology with a phenotype of presumed genetic origin

Conventional process

- MRI
- Blood withdrawal
- Gene test 1
- Lumbal punction
- Gene test 2
- Genomic array
- Gene test 3
- …
- Diagnosis?

Whole exome sequencing

- Blood withdrawal
- Exome sequencing
- Diagnosis?

Compare diagnostic yield of both trajectories
(Human Genetics/ Pediatric Neurology)

Patient Perspective

Interviews
assessing expectations and experience of diagnostic process so far
(20 patients)

Interviews
assessing experience of diagnostic process and obtained diagnosis
(IQ Health care)

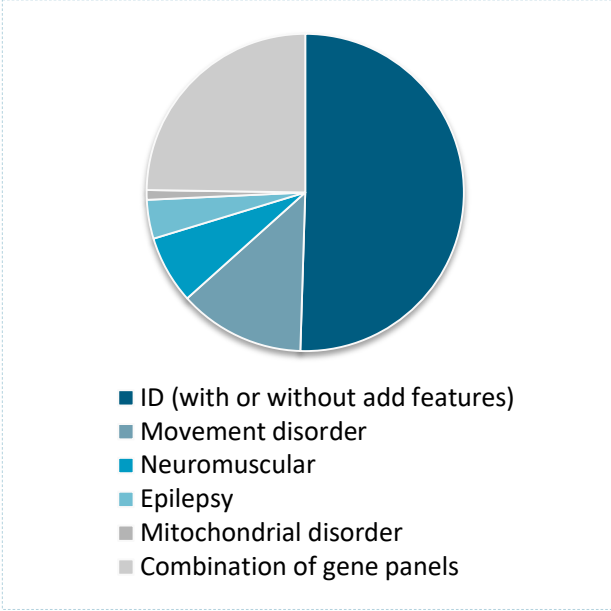
Cost effectiveness *(HTA)*

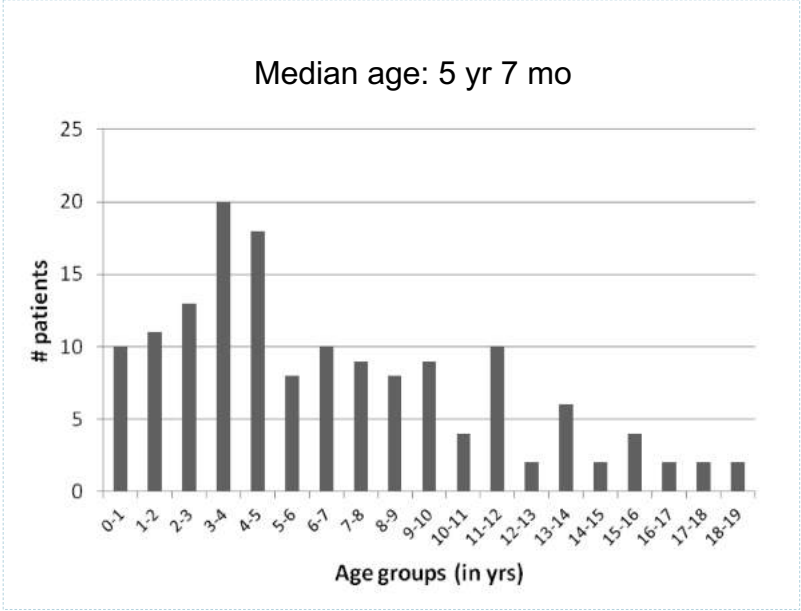
Project support: ZonMW Healthcare Efficiency Studies (40-41200-98-9131)

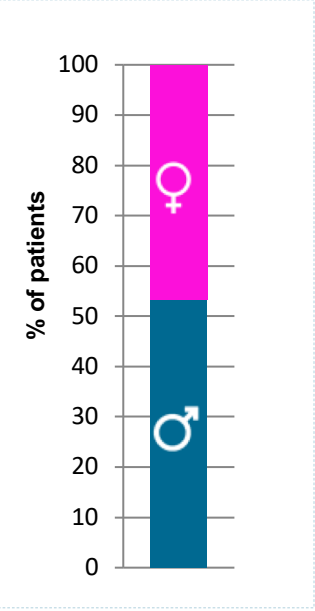
Radboudumc

Pediatric neurology cohort description

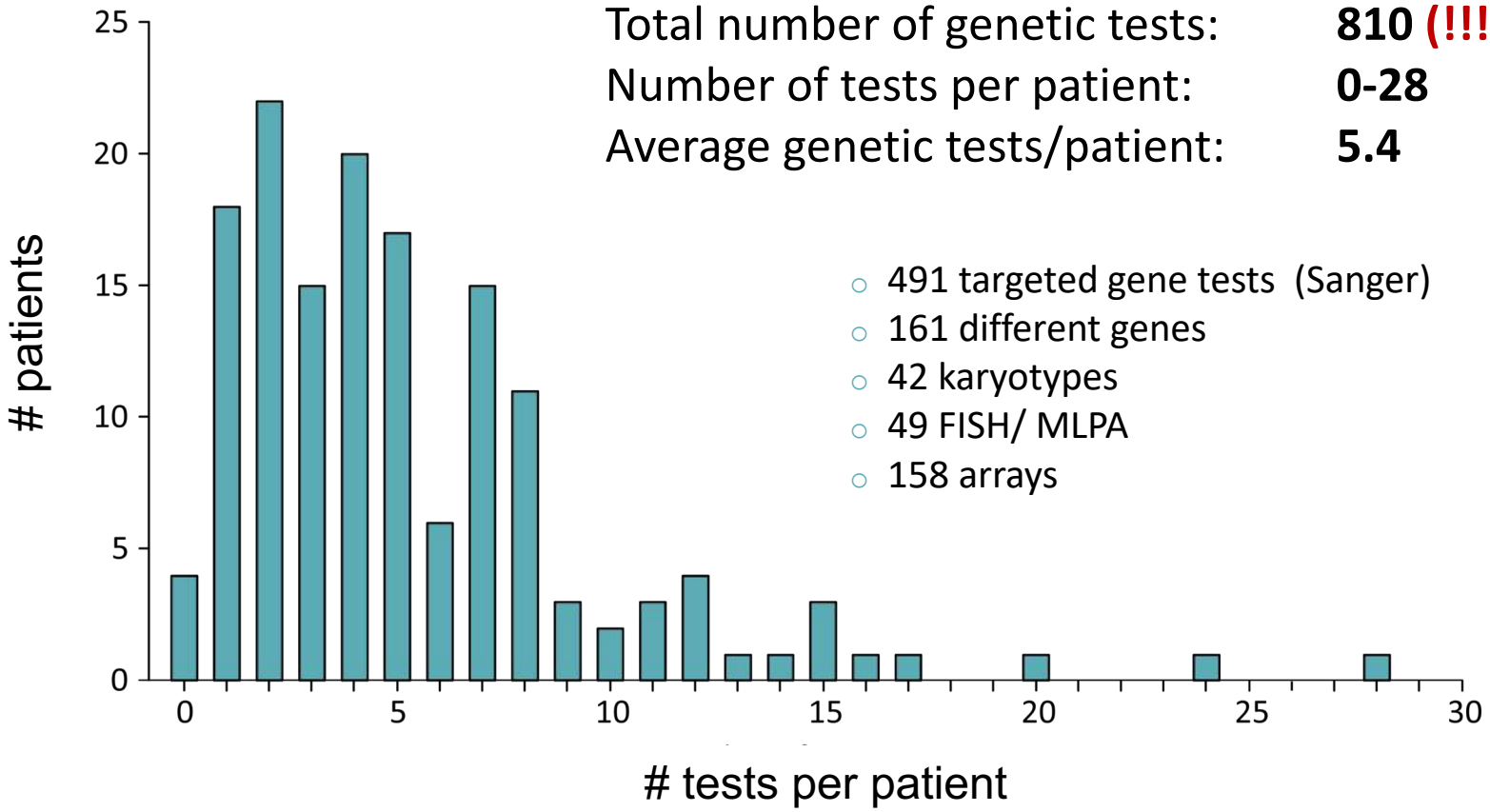
- 150 patients recruitment between Nov 2011 and Jan 2015
- Representation of ‘every day practice’ in pediatric neurology clinic
- Mixture of new patient referrals (n=66) and patients ‘somewhere’ in the diagnostic tract (n=84)







Number of genetic tests per patient



Total number of genetic tests: **810 (!!!)**
Number of tests per patient: **0-28**
Average genetic tests/patient: **5.4**

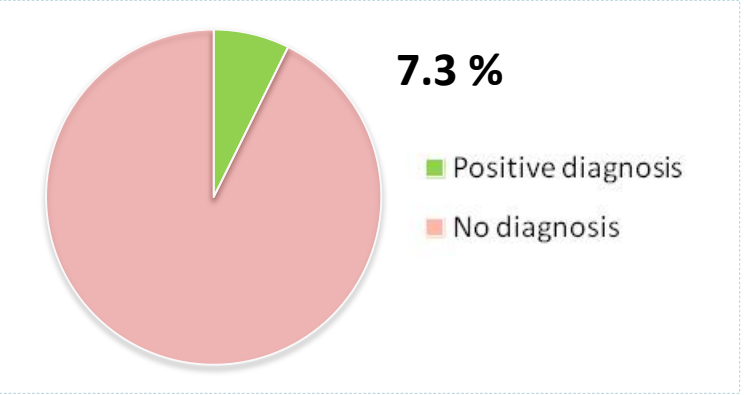
- 491 targeted gene tests (Sanger)
- 161 different genes
- 42 karyotypes
- 49 FISH/ MLPA
- 158 arrays

Diagnostic outcome in standard testing

Standard GENETIC testing cohort (n=150)

Total number of tests: 810
Number of tests per patient: 0-28
Average tests/patient: 5.4

Genetic cause identified: 11



	Number of genetic tests needed to obtain diagnosis
Patient 1	5 (*NSD1*)
Patient 2	5 (*PMP22*)
Patient 3	2 (*FLNA*)
Patient 4	7 (*SCN8A*)
Patient 5	3 (CNV)
Patient 6	7 (CNV)

	Number of genetic tests needed to obtain diagnosis
Patient 7	5 (CNV)
Patient 8	3 (CNV)
Patient 9	3 (*NKX2-1*)
Patient 10	6 (*FMR1*)
Patient 11	5 (CNV)

- Mean: 4.6 genetic tests needed to obtain diagnosis
- No correlation between obtaining a diagnosis and the number of genetic tests (p=0.57)

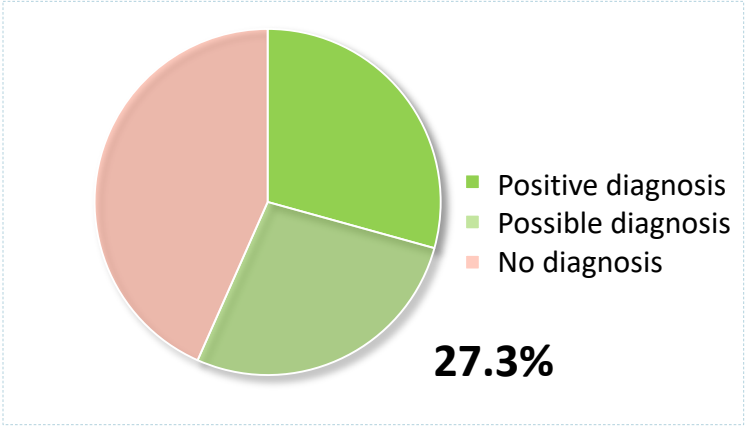
Diagnostic outcome WES

WES testing in cohort (n=150)

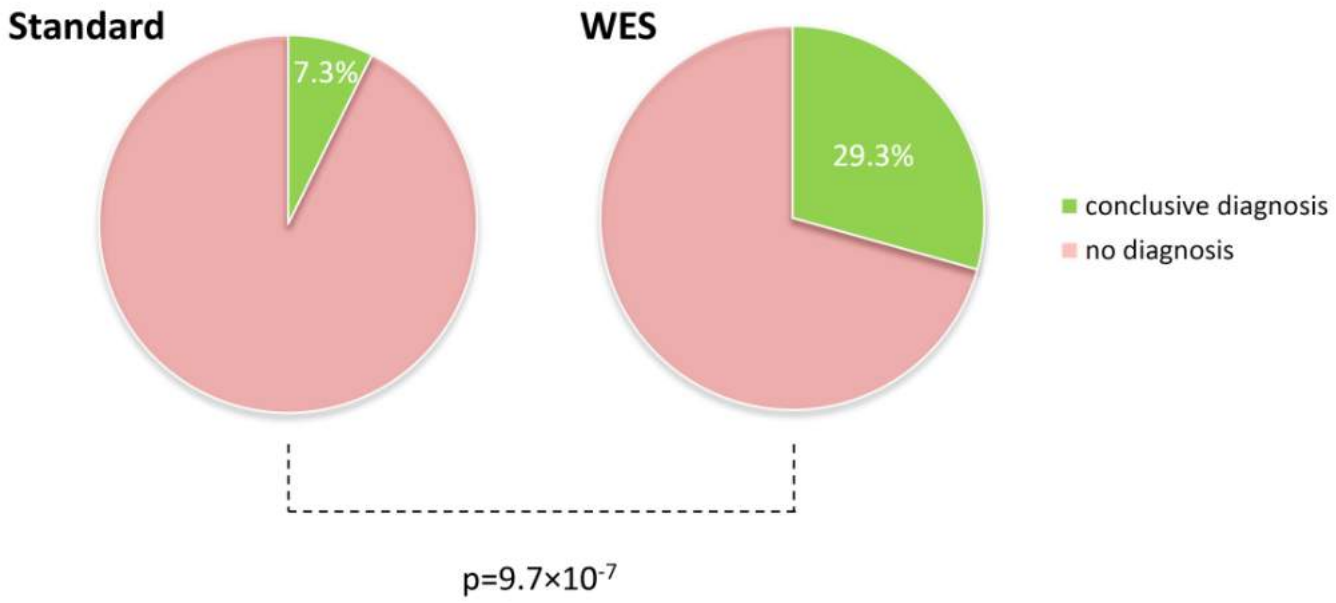
Total number of tests: 150 (x3)
Number of tests per patient: 1
Average tests/patient: 1

Added value of WES?

Total number of patients
with a ***possible*** **diagnosis**: **41**



WES provides significantly more diagnosis



At cohort level:
WES better for patient care

At patient level:
3 diagnoses in standard care were *not identified* by WES!

Vissers, van Nimwegen, Schieving et al. Genetics in Medicine 2017

There is more to the coding sequence than meets the eye

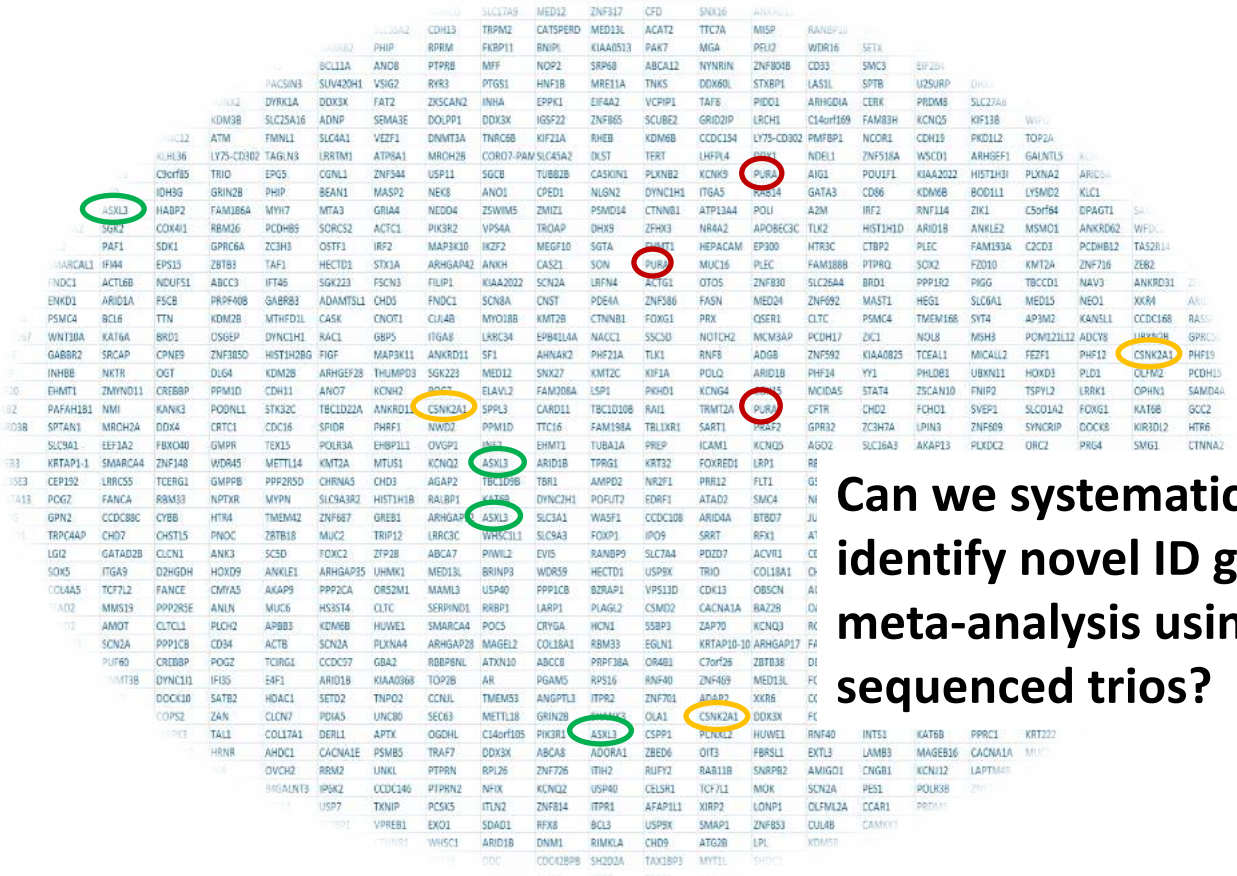
Radboudumc

Diagnostic exome sequencing for ID

Step1: Gene-Panel Strategy: SNVs and CNVs in ±1,000 established ID genes

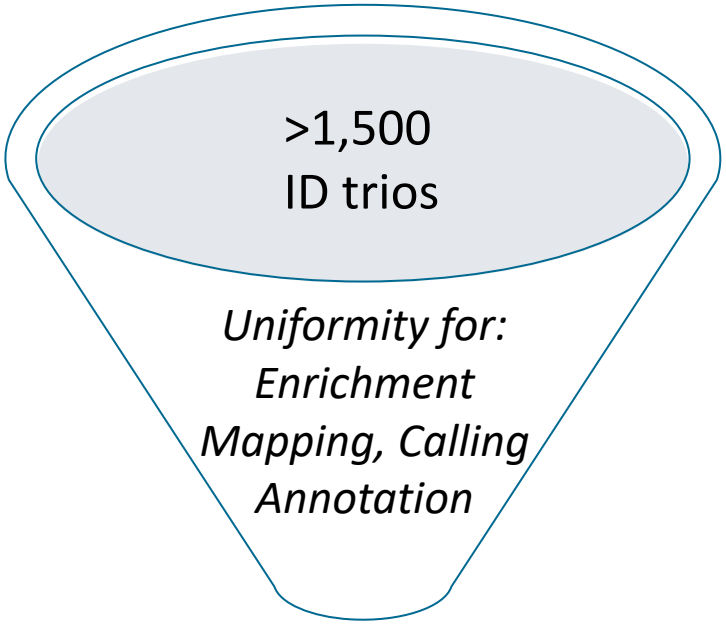
Step2: Exome-wide *de novo* mutation analysis

Patient-parent trios are interpreted ‘in isolation’

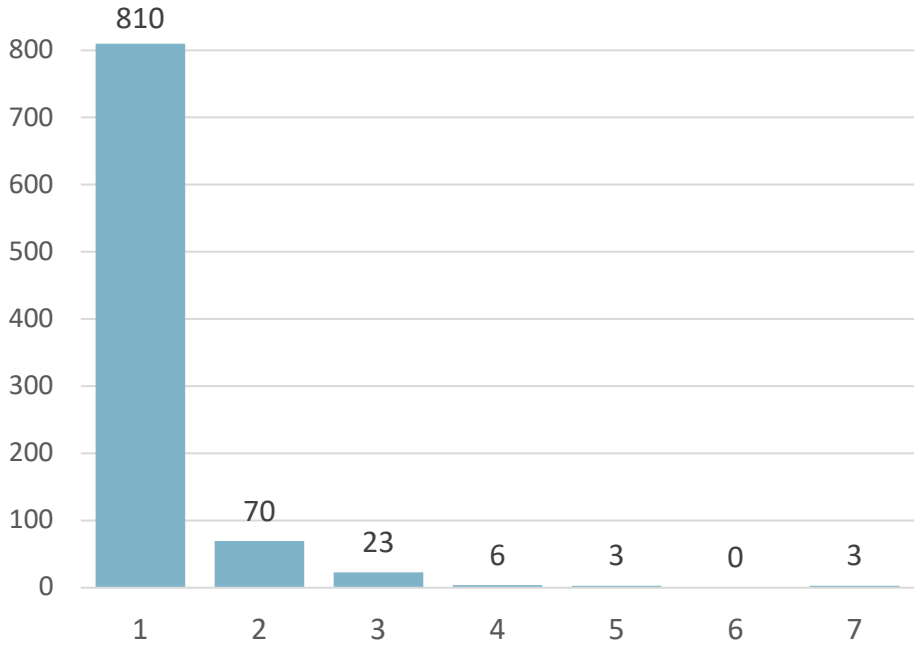


Can we systematically identify novel ID genes by meta-analysis using all sequenced trios?

ID trios for meta-analysis



820 ID trios
for analysis



- 1,079 *de novo* mutations (1.3 per patient)
- 915 different genes
- **105 with more than one DNM**

How to identify novel ID genes given the per generation mutation rate?

- Use **gene specific mutation rates** (GSMR)[1] to determine the excesses in DNMs per gene for 820 ID trios

- **Categorize mutations** in:

Loss-of-function (LoF):

stop-gain
frameshift
canonical splice sites,
start-loss
nonsense

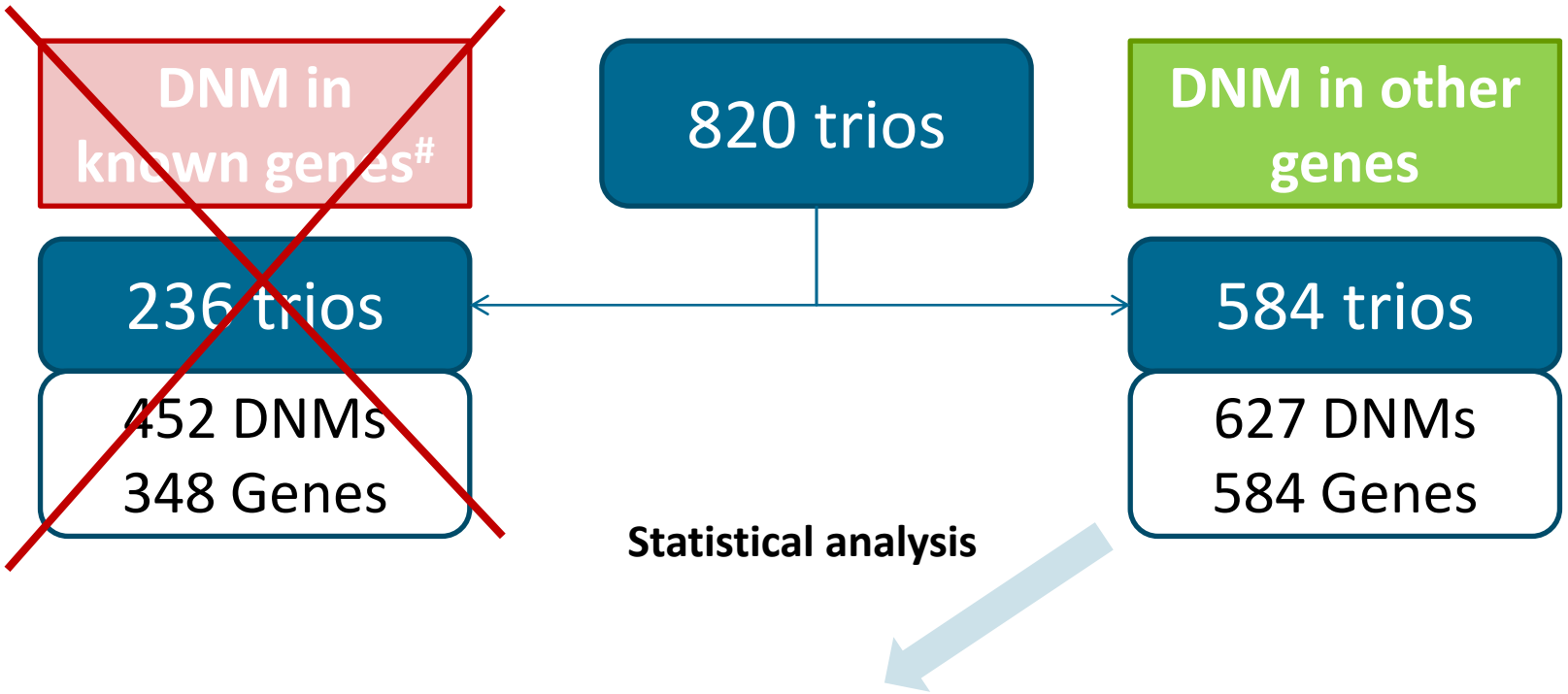
Functional:

missense
in-frame indels
LoF

1. Samocha et al. Nat Gen. (2014)

Theory put into practice

1. Remove all individuals with a DNM in a known gene



2. Identify genes with more Lof and/or functional DNMs than expected based on the gene specific mutation rate$

1,537 known ID genes (DD2GP + RUMC Diagnostics); $ Samocha et al. Nat Gen. (2014)

4 novel candidate ID genes in the RUMC cohort

Gene	p-value* LoF	Description
DLG4	**1.13×10^{-04}** *(n=3)*	Post-synaptic density gene
PPM1D	**0.047** *(n=2)*	Ser/Thr phosphatase mediating regulation of p38-p53 signaling
SOX5	**0.016** *(n=2)*	Member of Transcription factors that regulate embryonic development.
TCF20	**6.22×10^{-06}** *(n=4)*	Transcriptional activator of matrix metalloproteinase 3 and (co)activator of various other transcriptional activators.

Increase power!

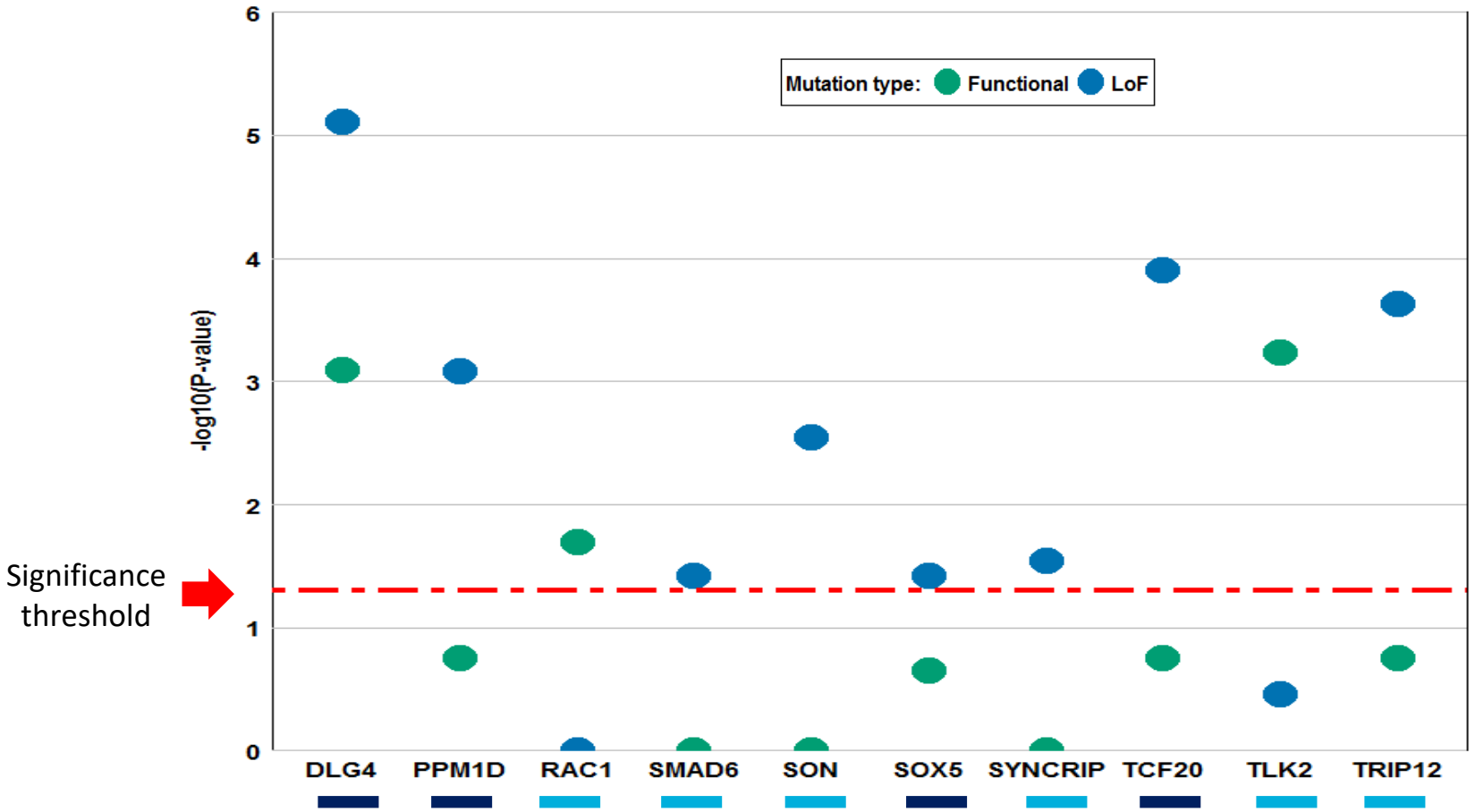
2,104 ID trios*
2,637 DNMs
(2,073 genes)

Repeat analysis!

*Rauch *et al*. Lancet (2012), de Ligt *et al*. NEJM (2012), Gilissen *et al*. Nature (2014), DDD study. Nature (2015)

Radboudumc

Meta-analysis of 2,104 trios identifies 10 novel ID genes



- Reproduce the 4 novel candidate ID genes in RUMC cohort
- 6 additional novel candidate ID genes

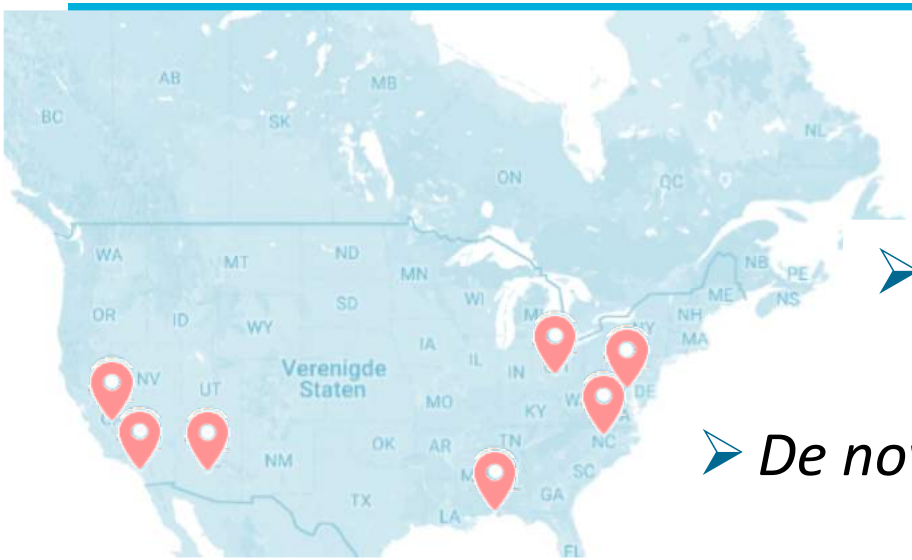
Lelieveld & Reijnders et al. Nature Neuroscience 2016

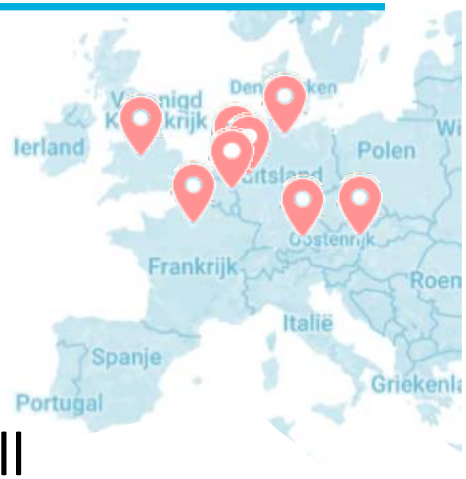
Radboudumc

Statistics do not prove disease causality,
patients, their phenotypes
and functional follow-up of the mutations do!

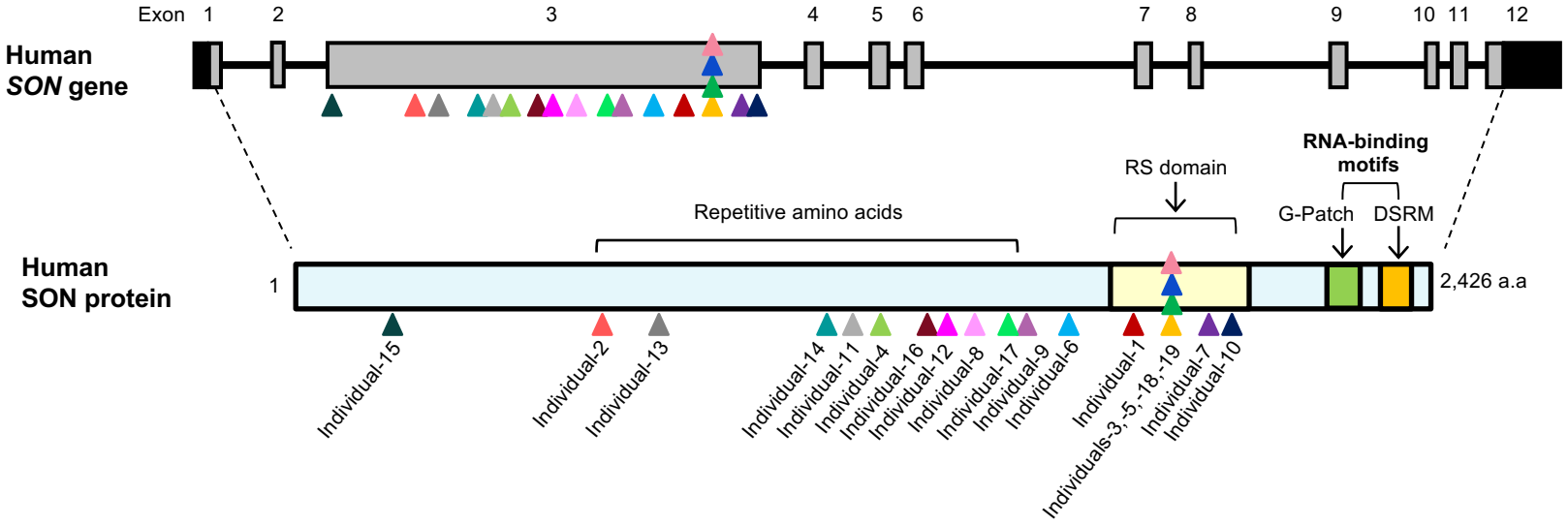
(Right?)

Genotype-first approach for involvement of *SON* in ID



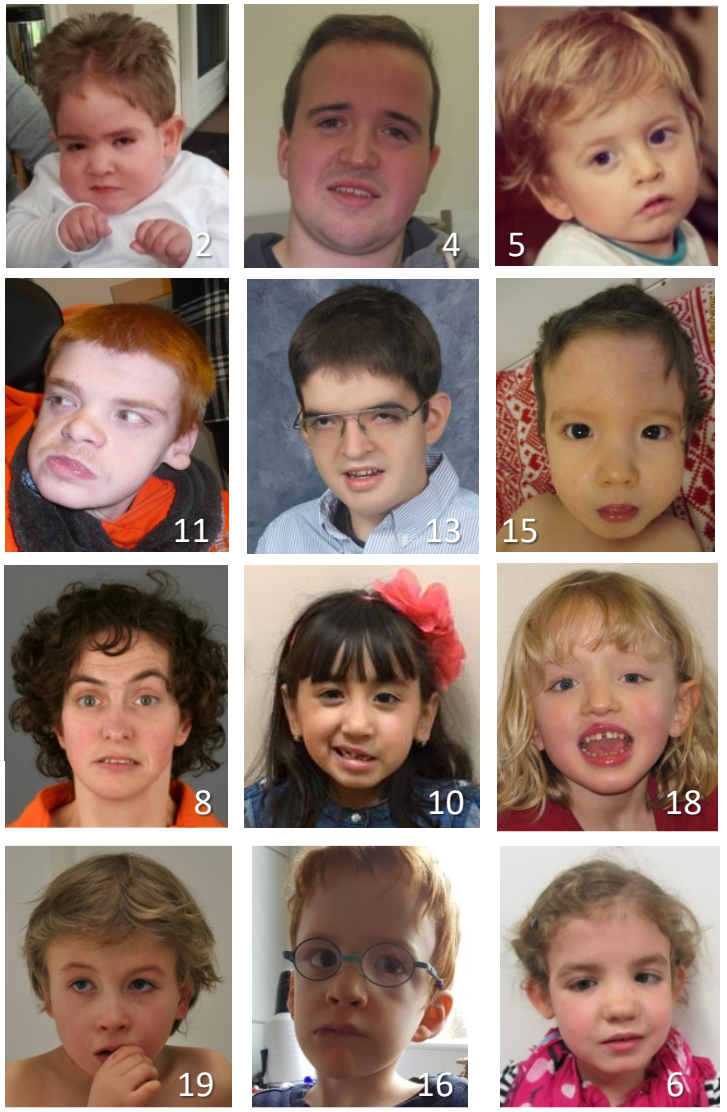


- 20 patients
- Loss-of-function mutations only
- *De novo* occurrence in all

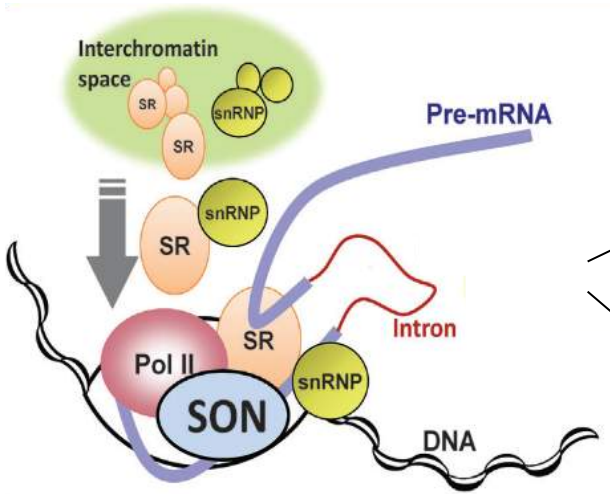


Phenotype associated with *SON* haploinsufficiency

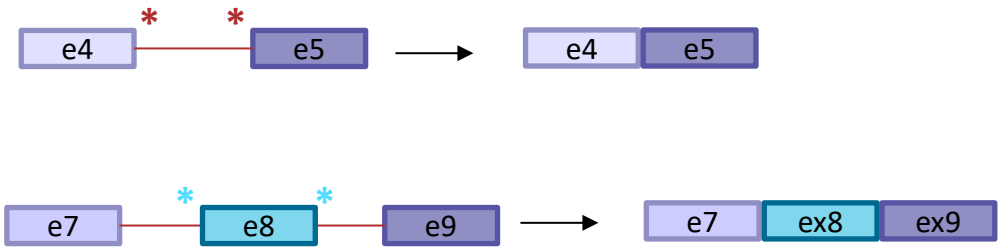
	Percentage	Number of affected individuals
Intellectual disability	**100%**	**20/20**
Brain malformation	**89%**	**17/19**
Neurological features	**85%**	**17/20**
Seizures	55%	11/20
Hypotonia	75%	15/20
Musculoskeletal abnormalities	**85%**	**17/20**
Hypermobility	40%	8/20
Scoliosis/Kyphosis	20%	4/20
Hemivertebrae	10%	2/20
Contractures	10%	2/20
Eye/Vision abnormality	**75%**	**15/20**
Strabismus	55%	11/20
(Suspicion) CVI	20%	4/20
Hypermetropia	30%	6/20
Heart defect	**25%**	**5/20**
Gastrointestinal malformation	**15%**	**3/20**
Urogenital malformation	**30%**	**6/20**
Horseshoe kidney	10%	2/20
Facial dysmorphism	**100%**	**20/20**
Short stature	**50%**	**10/20**
Craniosynostosis	**15%**	**3/20**

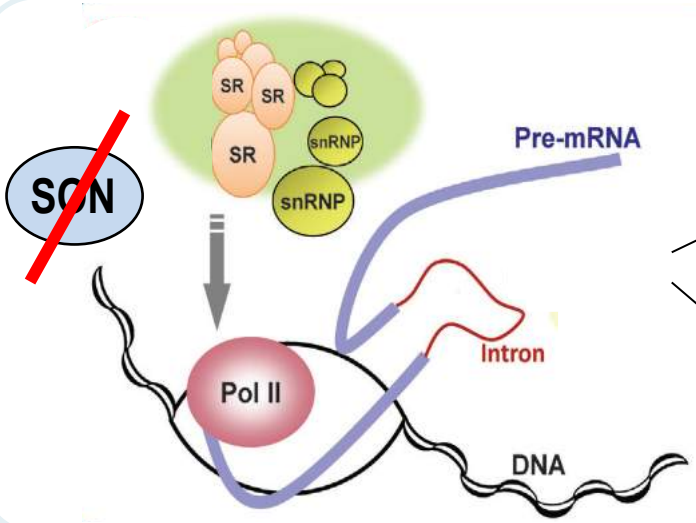


SON facilitates spliceosome recruitment to the RNA Pol II complex and enhances splicing

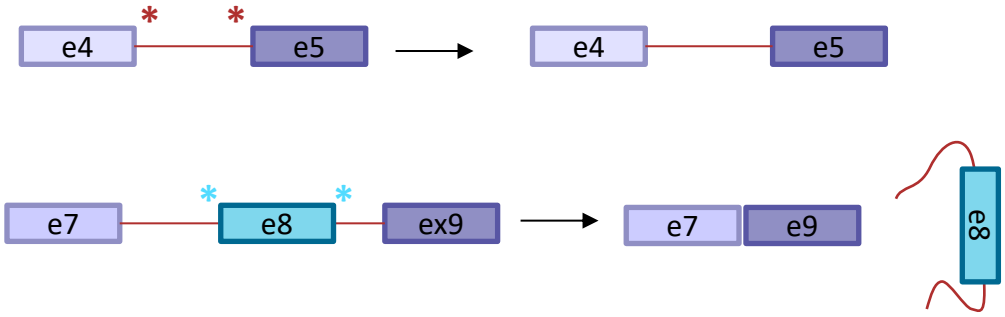


Recognition of weak and alternative splice sites





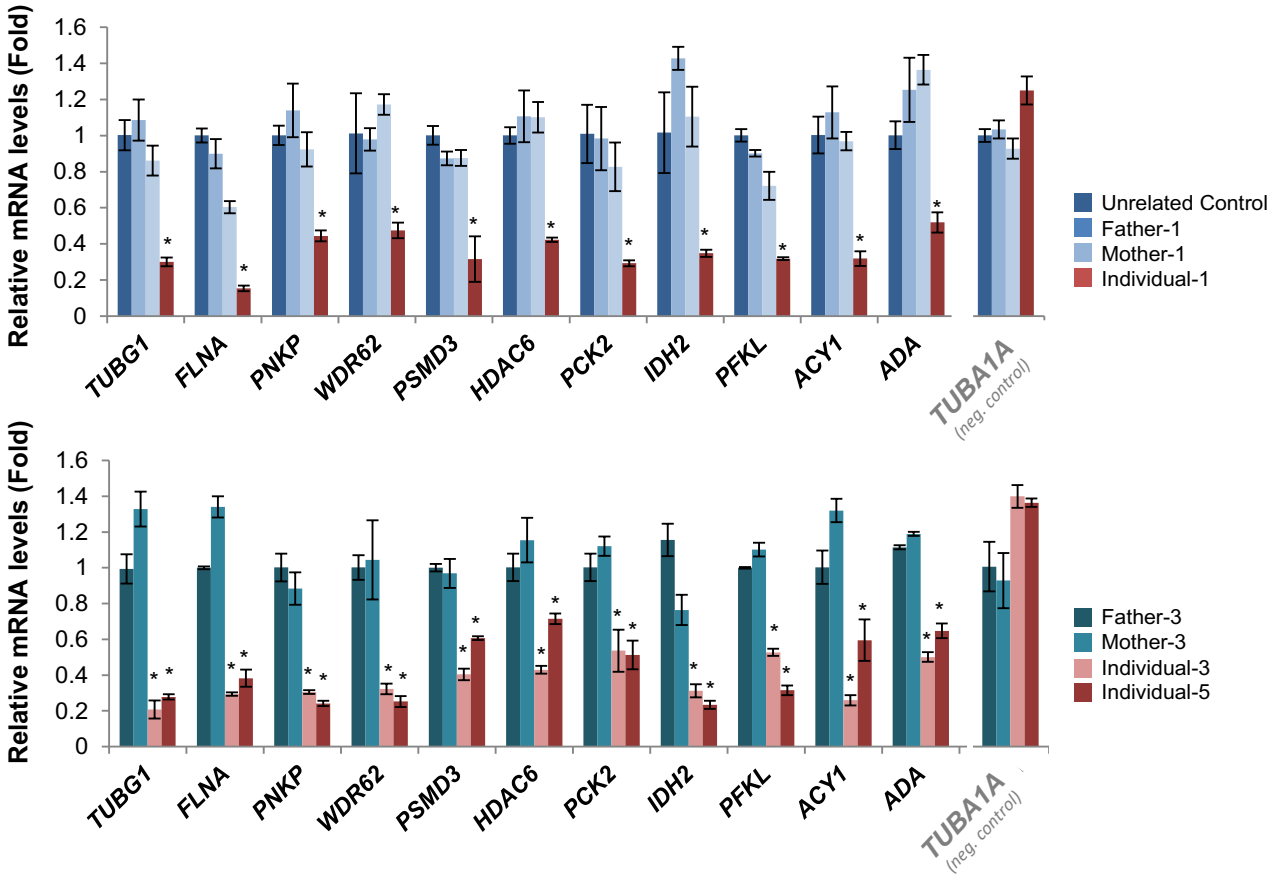
Intron retention and skipping of (alternative) exons



Adapted from Hickey et al. J Cell Biochem 2014

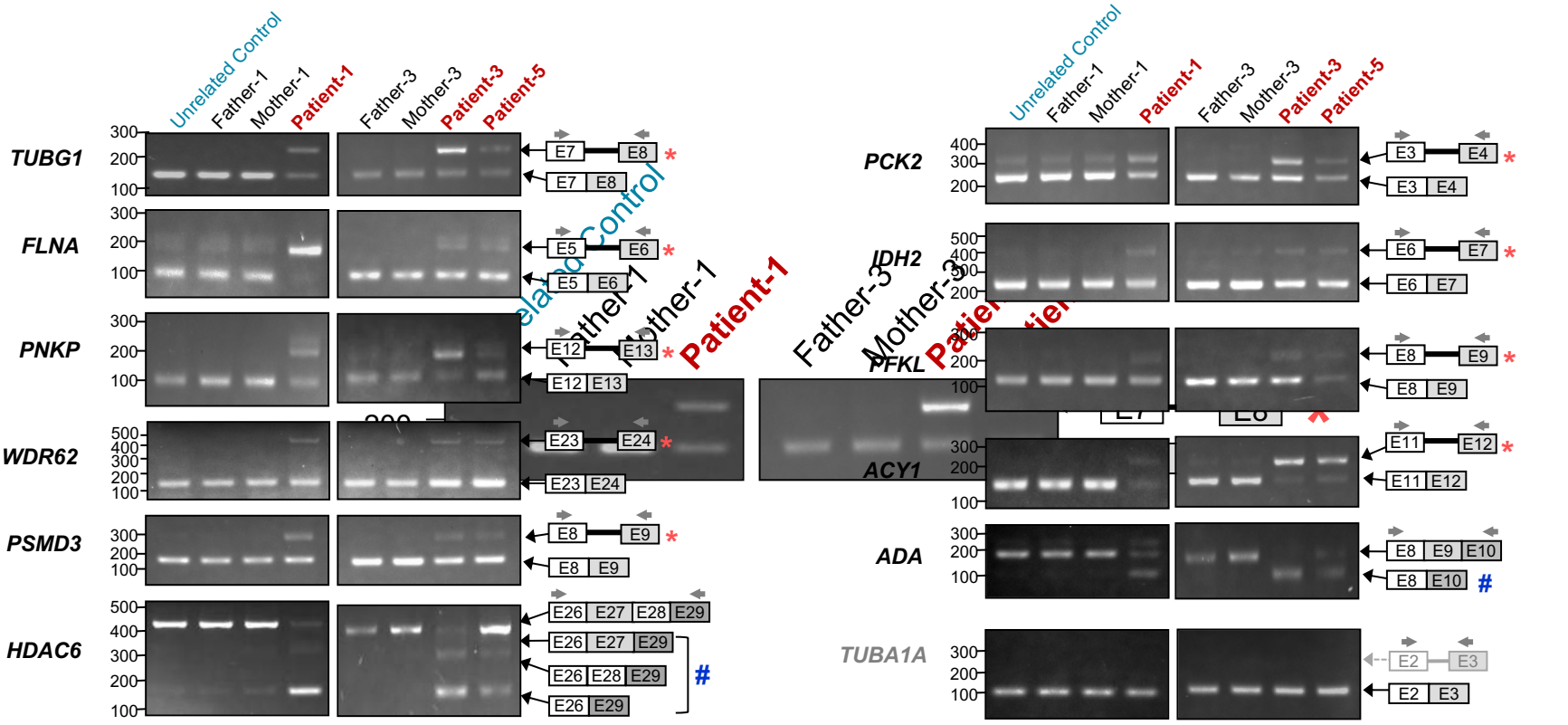
Radboudumc

SON haploinsufficiency downregulates genes essential for neurodevelopment in patients



Gene selection for RNA expression analysis based on downregulation of these genes in *SON* knockdown in cell line systems, and known to cause disease by themselves

Dosage effects of essential developmental genes result from erroneous SON-mediated RNA splicing



SON haploinsufficiency causes a novel ID syndrome by erroneous splicing

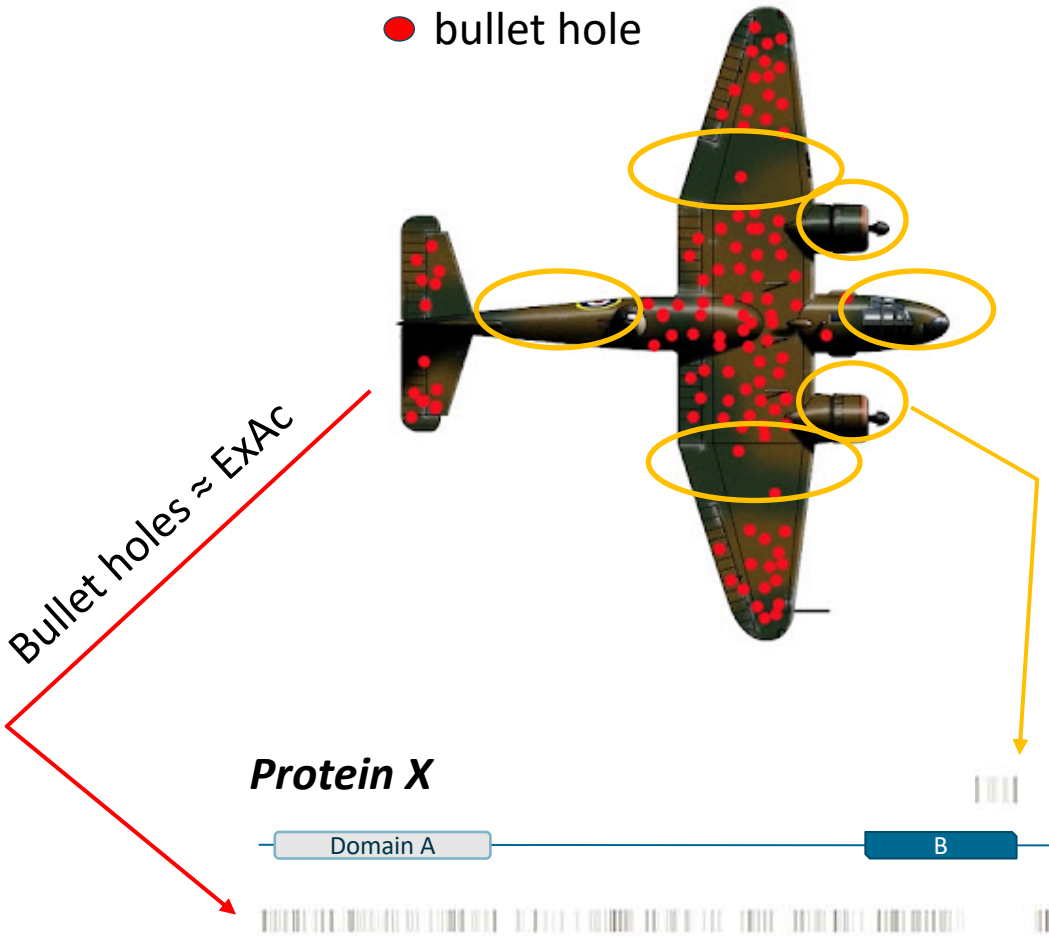
Kim, Shinde, Reijnders et al. Am J Hum Genet 2016

Radboudumc

Let’s go back to statistics, mutation location and the legend of Abraham Wald

Radboudumc

Abraham Wald's theory of survivorship bias

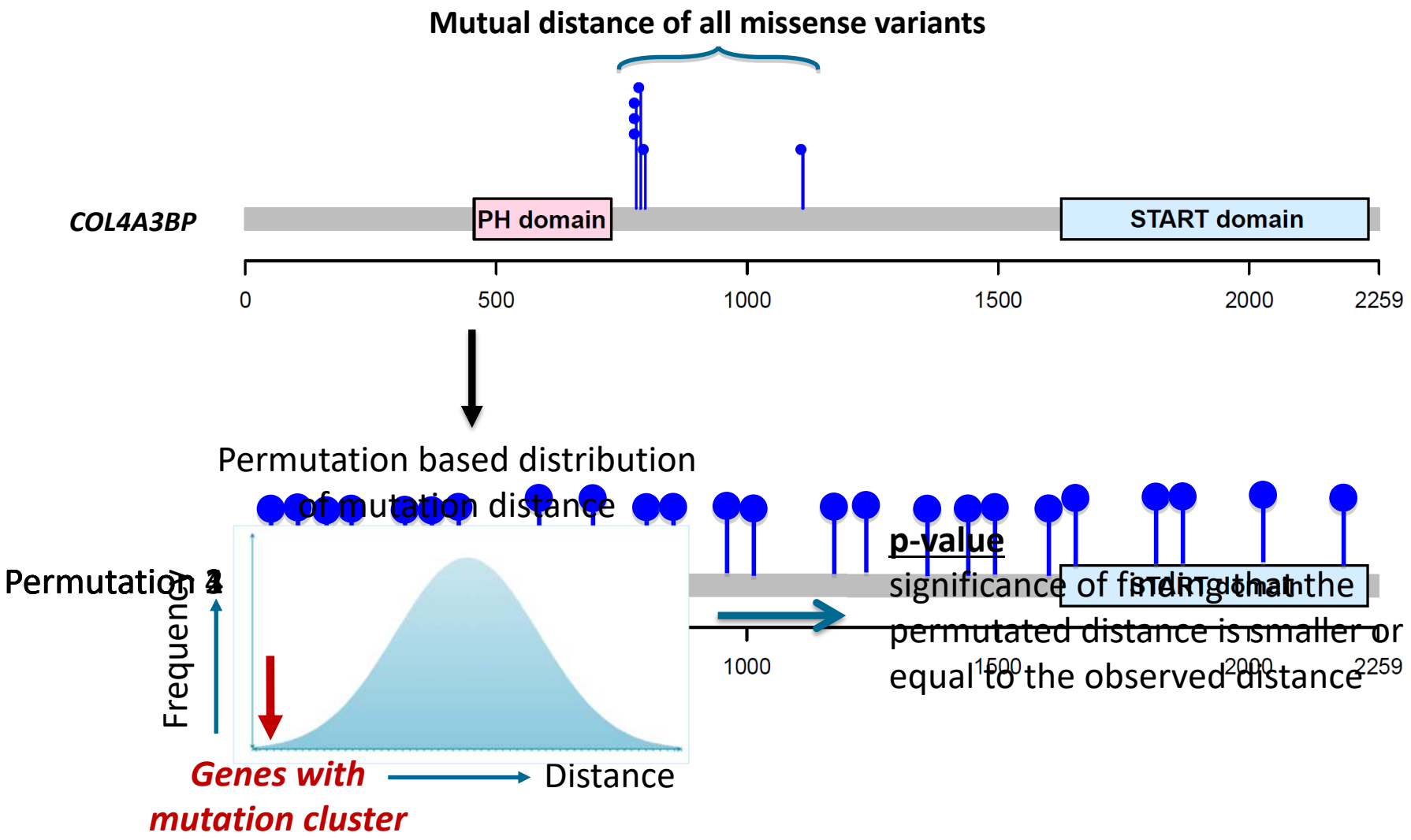


Absence of missense variation in ExAc define the critical regions

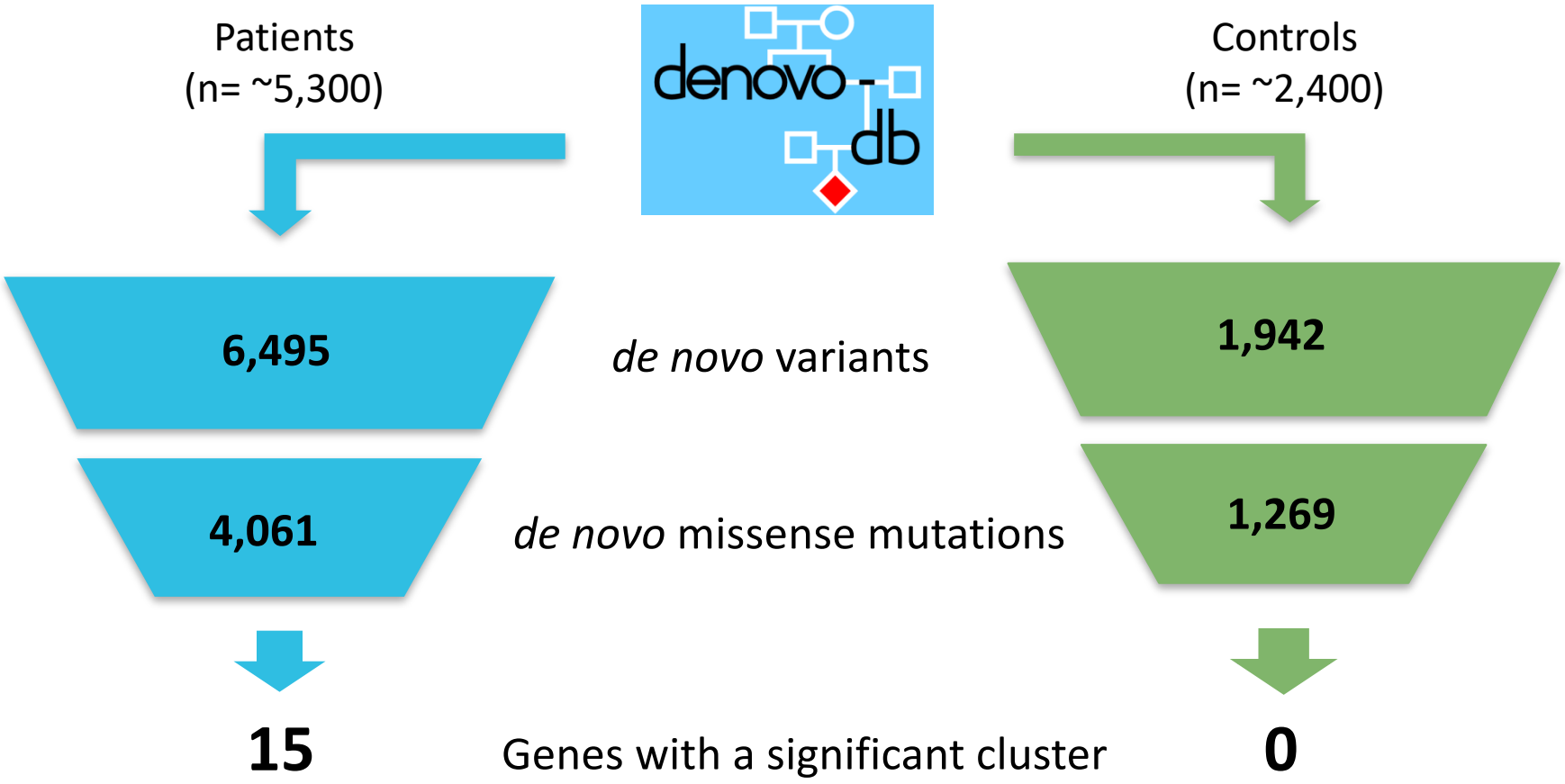
Any *de novo* mutation in a patient is highly interesting

Clustering suggestive for mechanisms other than haploinsufficiency

How to identify clustering?



15 genes with clusters of missense mutations



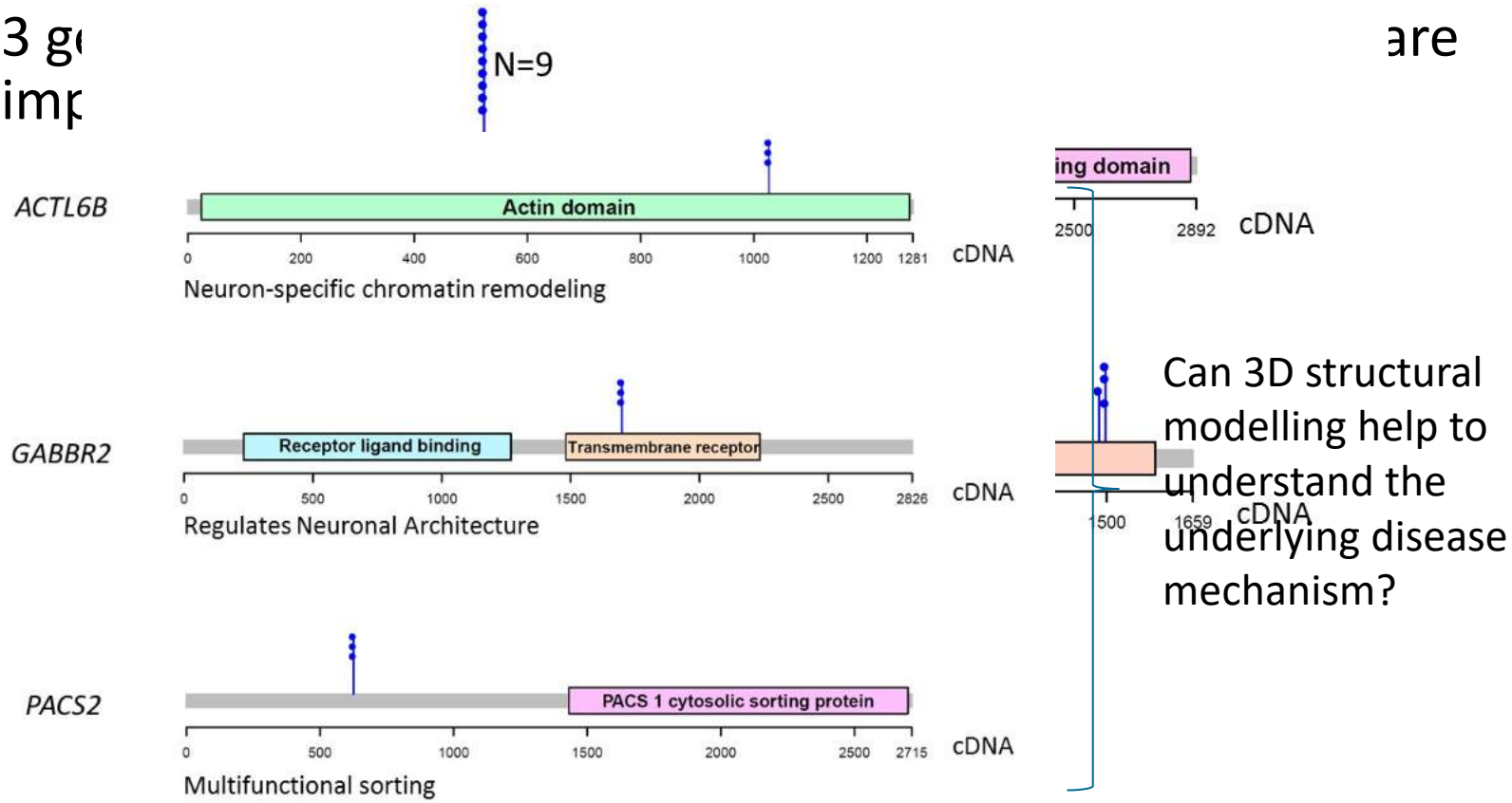
Denovo-dB: Turner *et al.* Nucleic Acids Research (2016)

15 genes show clustering of missense mutations

- 12 genes are well-known ID genes

  For all, the disease mechanism is reported to be gain-of-function or dominant-negative

- 3 g… …are imp…

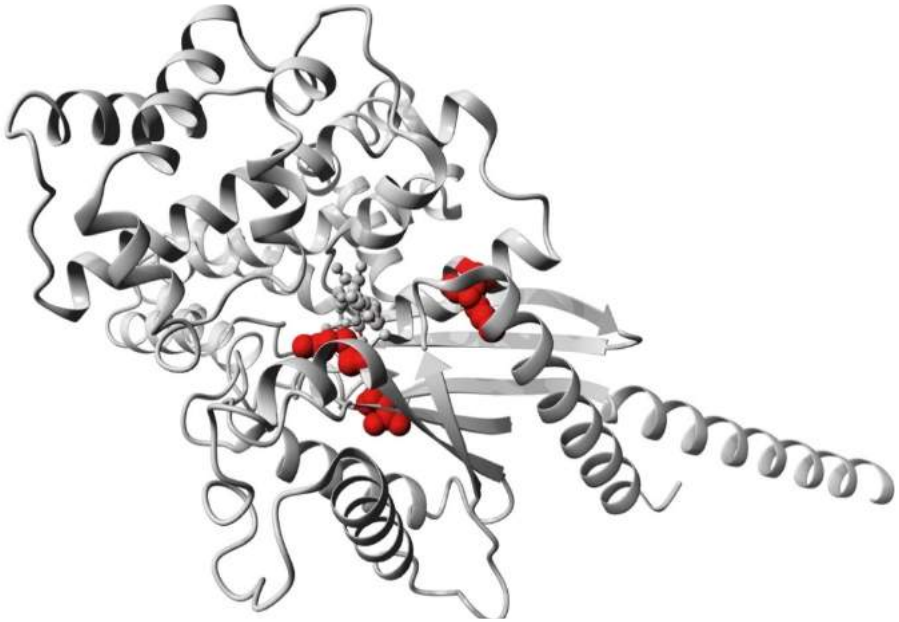


Can 3D structural modelling help to understand the underlying disease mechanism?

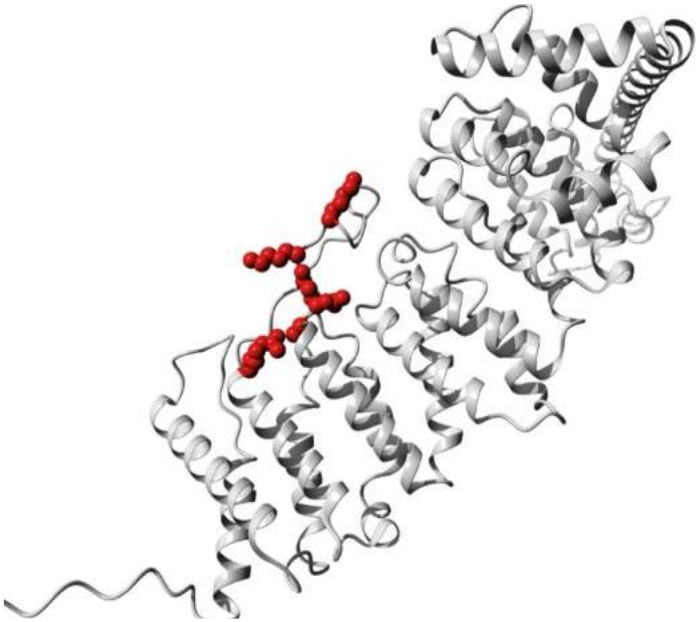
3D modeling: mutational mechanisms

- *De novo* missense variants in non-haploinsufficiency genes are located at the surface of protein structures
- They disrupt protein-protein interactions (complex formation), rather than protein folding (stability)

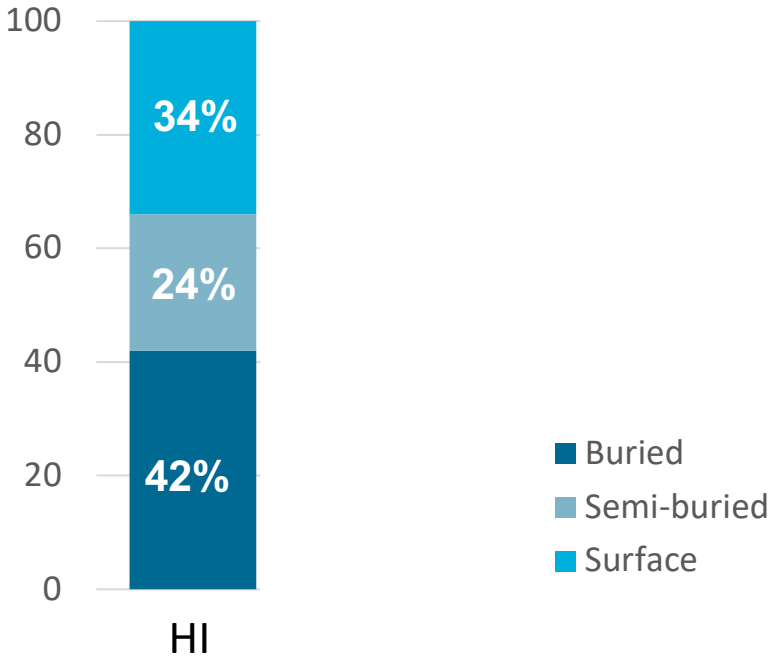
GNA1
(Haploinsufficient mechanism)



PPP2R5D
(Non-haploinsufficient mechanism)



3D protein modeling: mutational mechanism



- HI genes: 25 genes, 75 *de novo* missense mutations
- Non-HI genes: 10 genes (incl. *ACTL6B* and *GABBR2*), 48 *de novo* missense mutations

Genes with clustering mutations show mutation clustering at the protein surface and suggest non-HI mechanisms as underlying cause

Why NGS innovation matters in diagnostic care

- Novel genes for ID and/or NDD (and beyond)
- Insight into disease mechanisms and underlying biology

n = 2,419

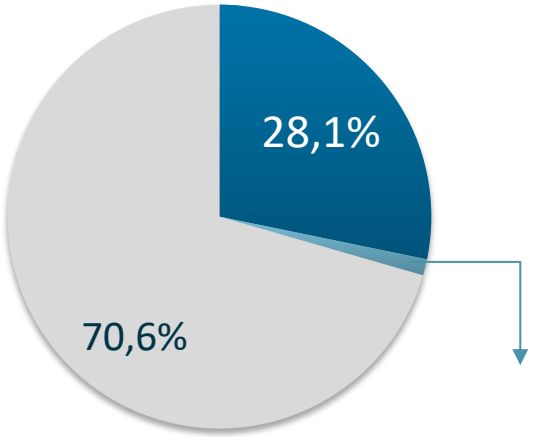
ID trios @Radboudumc
(with consent for diagnostic innovation)

n = 713

Patients with a *de novo* mutation any of the known ~1,000 ID genes

n = 30

Patients with a *de novo* mutation in one of the 13* novel genes



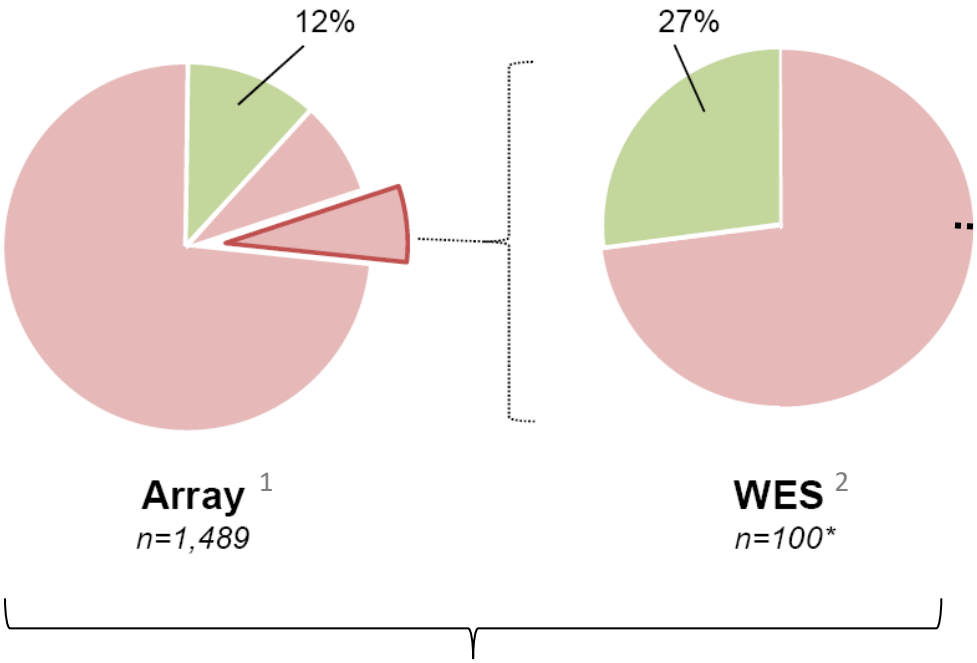
>200 patients per year in Europe alone!

**Lelieveld & Reijnders et al. Nat Neurosc 2016; Lelieveld & Wiel et al AJHG 2017*

Exome vs. Genome sequencing in a clinical setting

Radboudumc

Diagnostic yield of current technologies



50 'WES negative' patients for WGS

Complete genomics

- Cumulative 50 patient-parent trios in collaborative project
- Sequence depth: 80-fold

Sequence 50 patient-parent trios in collaborative project
~25-5% of patients with severe ID

1. Vulto-van Silfhout et al. Human Mutation 2013
2. De Ligt et al. New England Journal of Medicine 2012

WGS with focus on clinically relevant mutations

1. **Did WES/ arrays miss mutations?**

   → Re-assess the coding sequence for SNVs and CNVs

2. **What is the added diagnostic value of WGS over WES?**

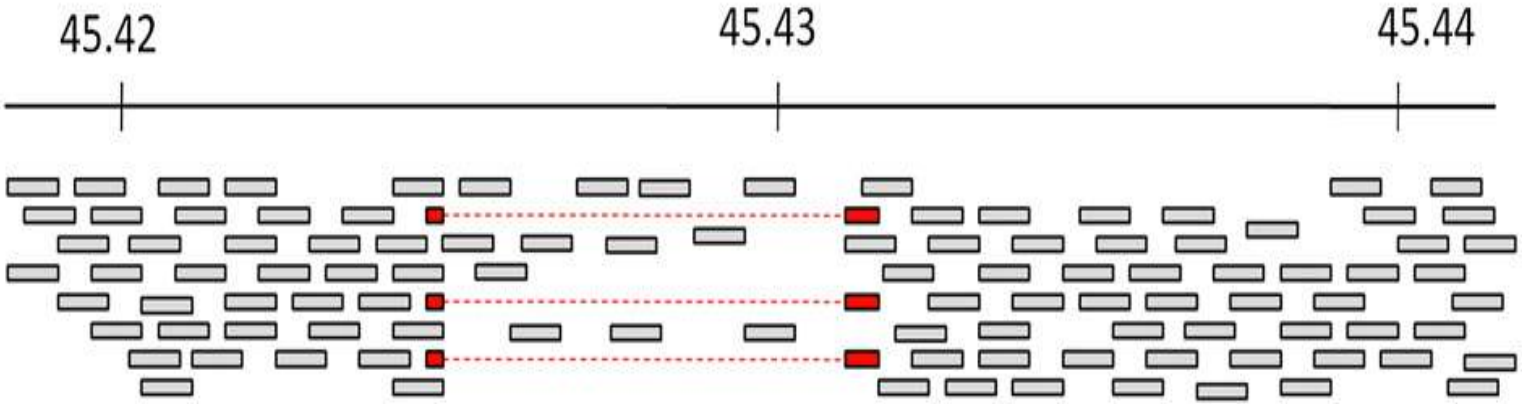
   → Non-coding SNVs, and SVs affecting known ID genes
   - UTR
   - Introns
   - Promoters

26% of cohort has a clinically relevant *de novo* SNV

	Genes previously implicated in ID	Novel Candidate ID gene
Nonsense mutation	*RAI1* *SCN2A* *TBR1* *POGZ*	*ASUN* *APPL2* *SON*
Missense mutation	*KCNA1* *TBR1* *SPTAN1* *MED13L* *PPP2R5D*	*NGFR* *GFPT2* *BRD3* *NACC1* *MAST1*
Insertion/deletion	*WDR45* *SMC1A* *SATB2*	
Synonymous w/ effect on splicing	*KANSL2*	*WWP2*

Structural variation dectection in WGS

- Read depth approach
- Discordant reads pairs/ split reads method



16% of patients has a clinically relevant SV

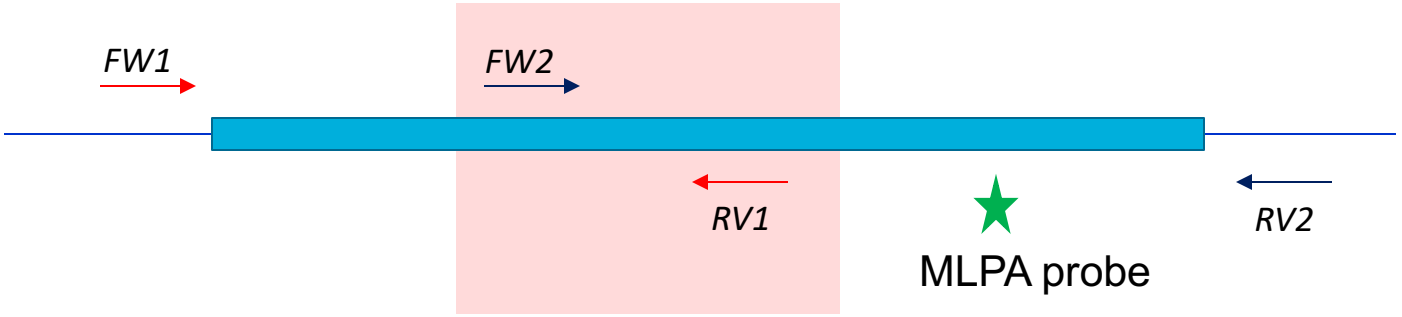
De novo intra-exonic deletion *MECP2* exon4

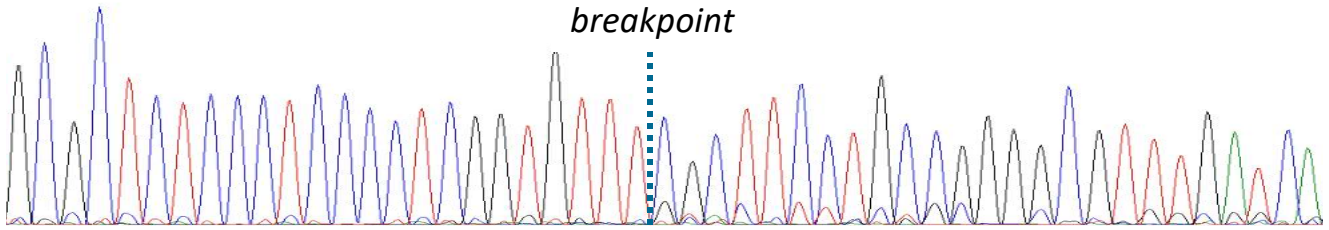
Girl, with signs of Rett syndrome

MECP2 tested by Sanger sequencing and no mutations identified

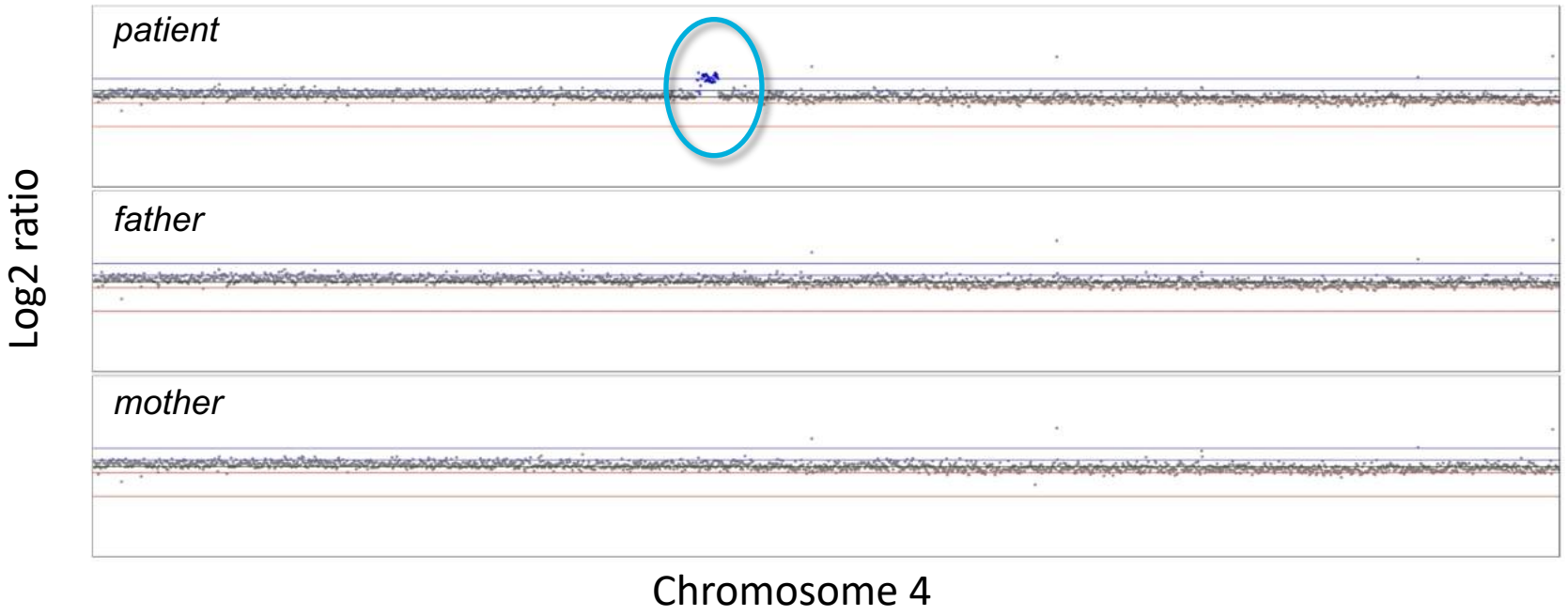
Exon 4: ~1kb in size, amplified by two PCRs

one MLPA probe for CNV detection



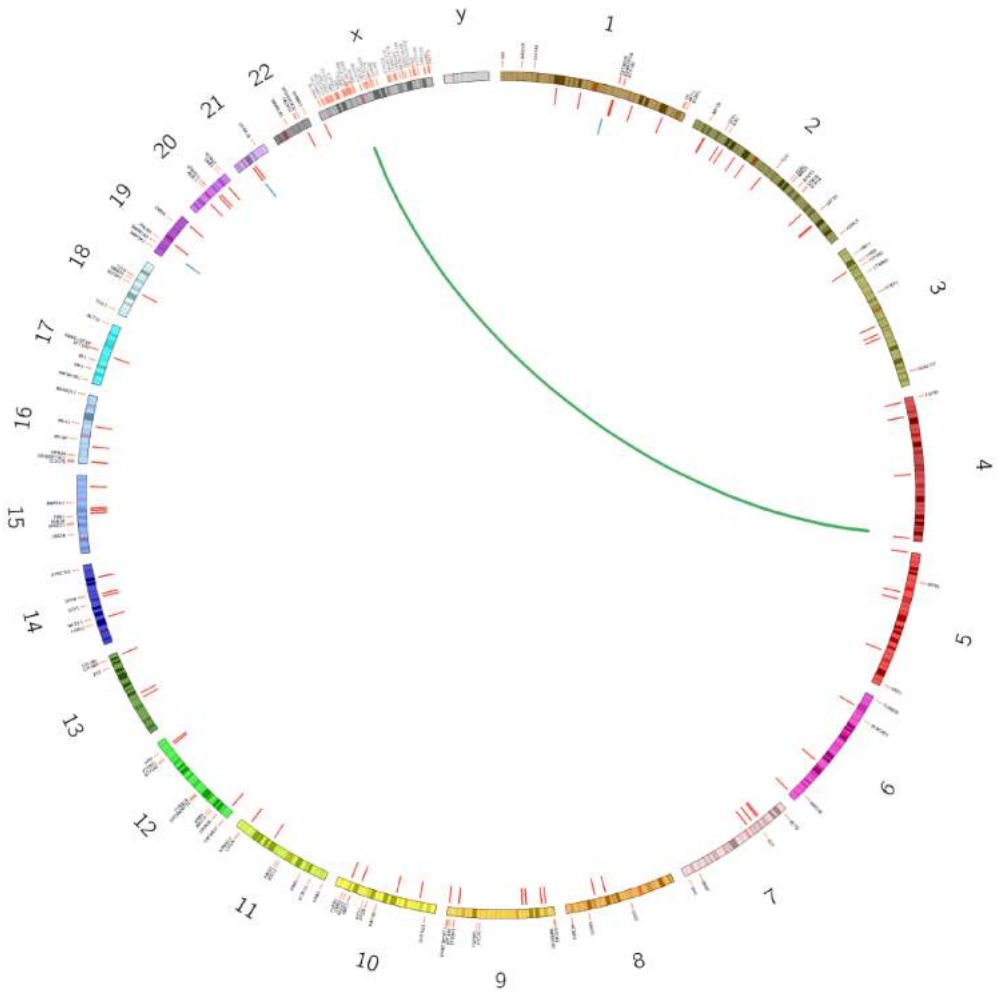


A *de novo* duplication on chromosome 4….

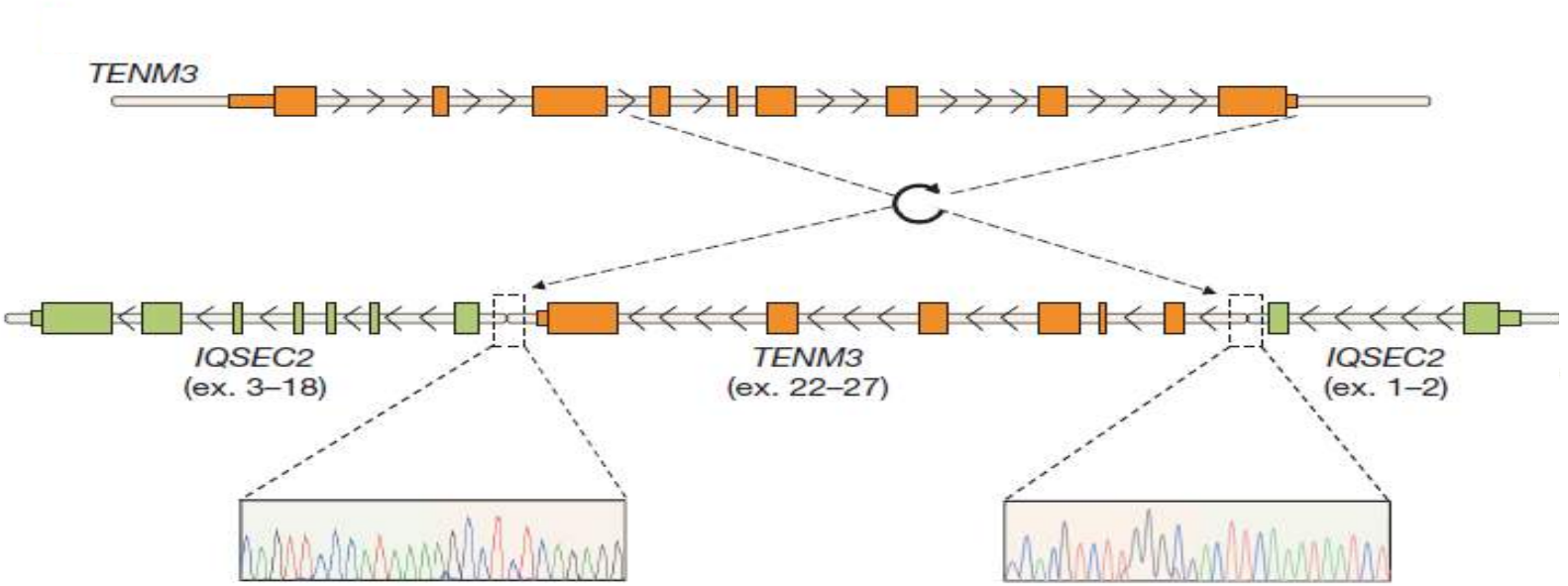


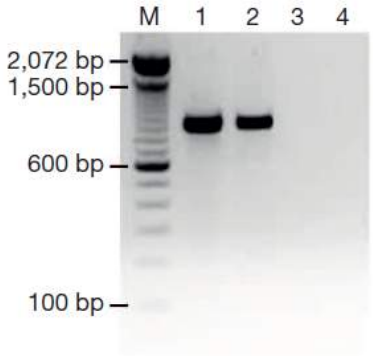
Chr4(GRCh37):g.183693000-183756000
Estimated size ~60 kb
Contains one gene (in part): *TENM3*

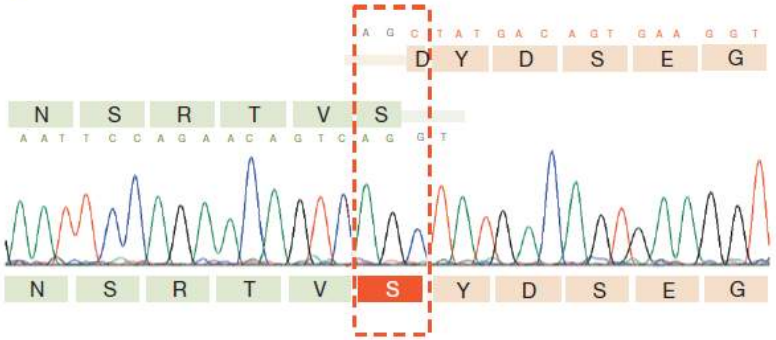
Duplication inserted into chromosome X…



….creating a stable fusion gene disrupting *IQSEC2*







So what about non-coding variants in ID...?

Location*	# variants
Non-coding	**43**
Promoter	1
Intron	38
Splice site	1
3'UTR	1
5'UTR	3

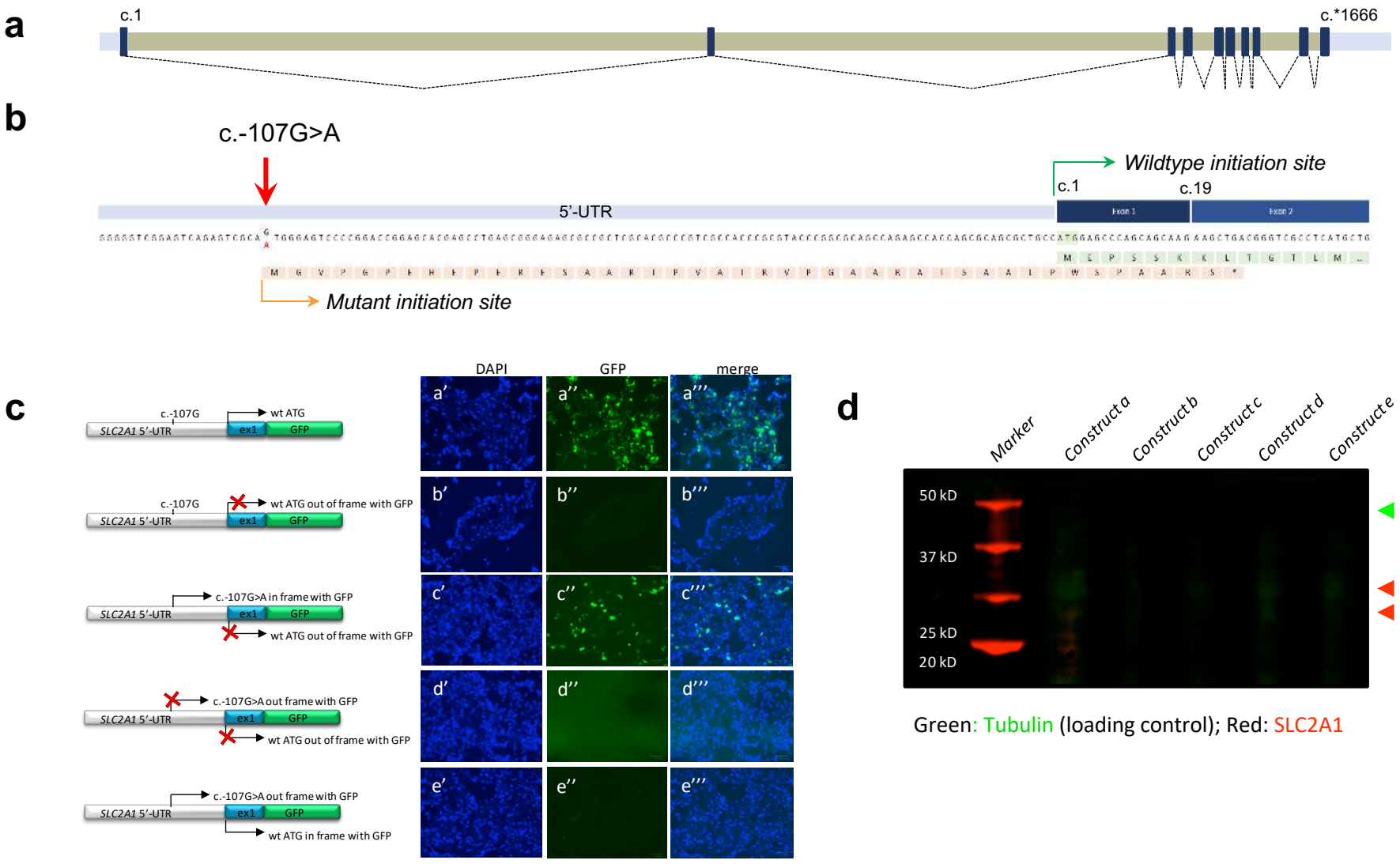
**of known ID genes (n=528)*

- Predicted effect on splicing? NO
- Predicted effect for miRNA binding? NO
- Predicted effect promotor? Weak, and gene does not fit patients phenotype

- Role? Significance? Currently unknown....

BUT…….

De novo 5´UTR *SLC2A1* causes GLUT1 deficiency



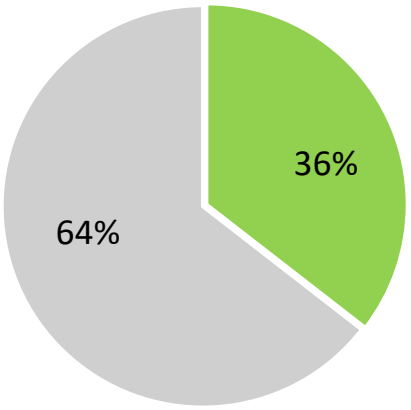
Green: Tubulin (loading control); Red: SLC2A1

Willemsen et al. EJHG 2017

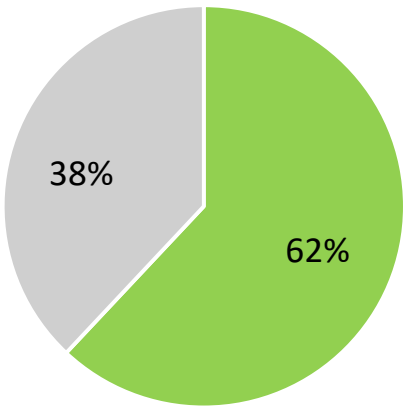
What if... WGS would be introduced in clinic

WGS provided a conclusive genetic diagnosis in 21/50 cases

Cumulative estimate of cases solved
standard diagnostics



Cumulative estimate of cases solved
with WGS



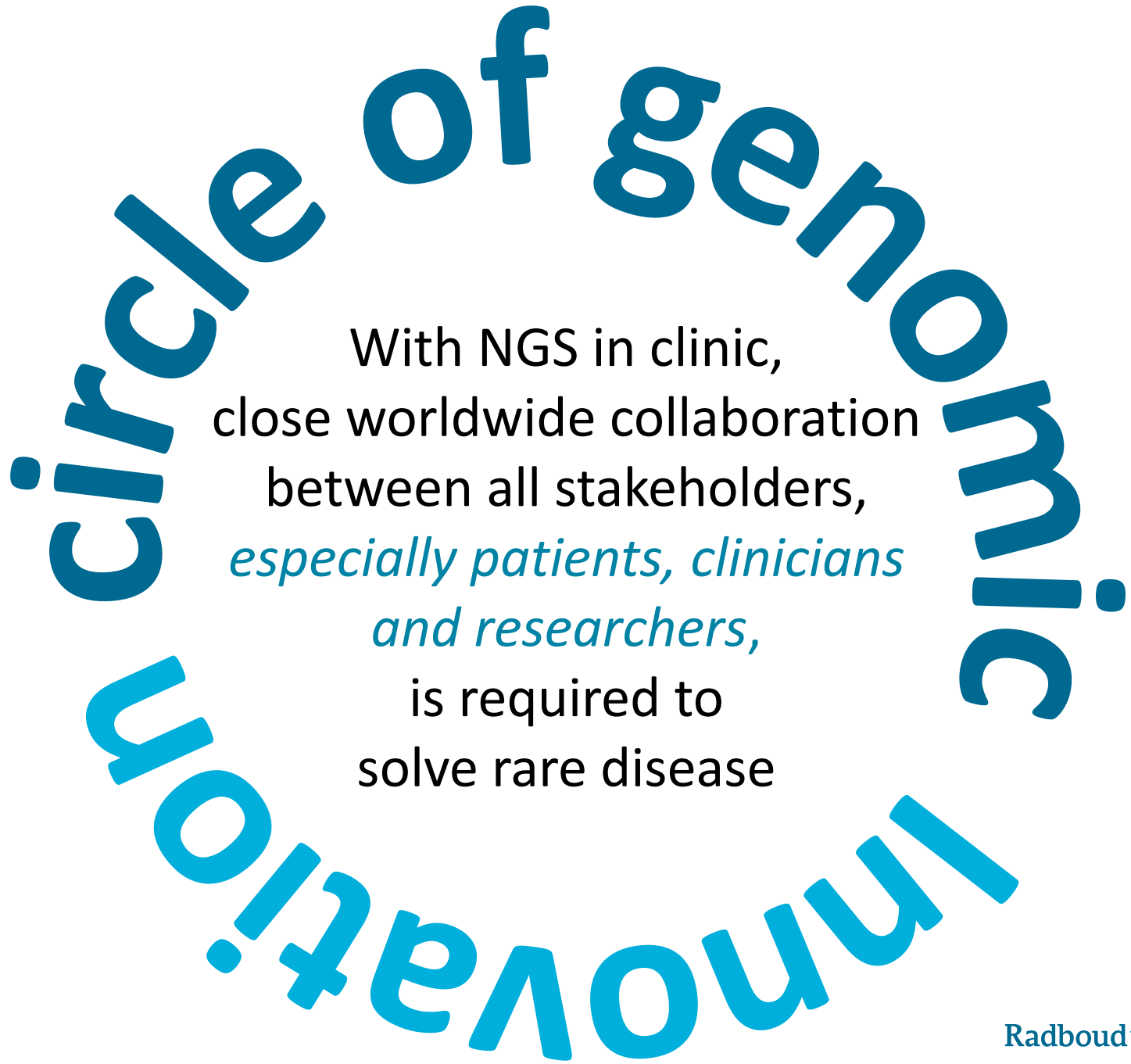
WGS has the potential to provide a conclusive molecular diagnosis for 62% of patients with severe ID

Gilissen & Hehir-Kwa et al. 2014 Nature

Radboudumc

Take home messages

1. Exomes are a (cost-)effective strategy in clinic and outperforms routine diagnostic by number of diagnoses
2. Statistical large-scale analyses on existing WES data allows identification of novel disease genes
3. WGS provides a better representation of the exome and more detailed insight in structural variation
4. Most disease causing mutations are located in coding sequence, and the importance of non-coding mutations is largely to be explored


circle of genomic innovation
With NGS in clinic,
close worldwide collaboration
between all stakeholders,
especially patients, clinicians
and researchers,
is required to
solve rare disease

Radboudumc

Acknowledgements

Genome Research

Christian Gilissen
Stefan Lelieveld
Jayne Hehir-Kwa
Laurens Wiel
Alex Hoischen
Djie Tjwan Thung
Maartje van de Vorst
Petra de Vries
Irene Janssen
Michael Kwint
Irene Janssen
Sinje Geuer
Joris Veltman

Genome Diagnostics

Erik-Jan Kamsteeg
Rolph Pfundt
Marcel Nelen
Servi Stevens
Alexander Stegmann
Dorien Lugtenberg
Helger Yntema

Clinical Genetics

Margot Reijnders
Sandra Jansen
Marjolein Willemsen
Bregje van Bon
Tjitske Kleefstra
Bert de Vries
Maaike Vreeburg
Margje Sinnema
Connie Stumpel
Han Brunner

Pediatric Neurology Project

Michèl Willemsen
Kirsten van Nimwegen
Janneke Grutters
Jolanda Schieving

Complete Genomics

Rick Tearle
Robert Klein
Richard Leach

European Research Council

TECHGENE

SEUVADIS CONSORTIUM

XAnEUploidy

Human Genetics Nijmegen

NWO

Acknowledgments *SON* study

Jung-Hyun Kim

Eun-Young Erin Ahn

Deepali Shinde

Margot Reijnders

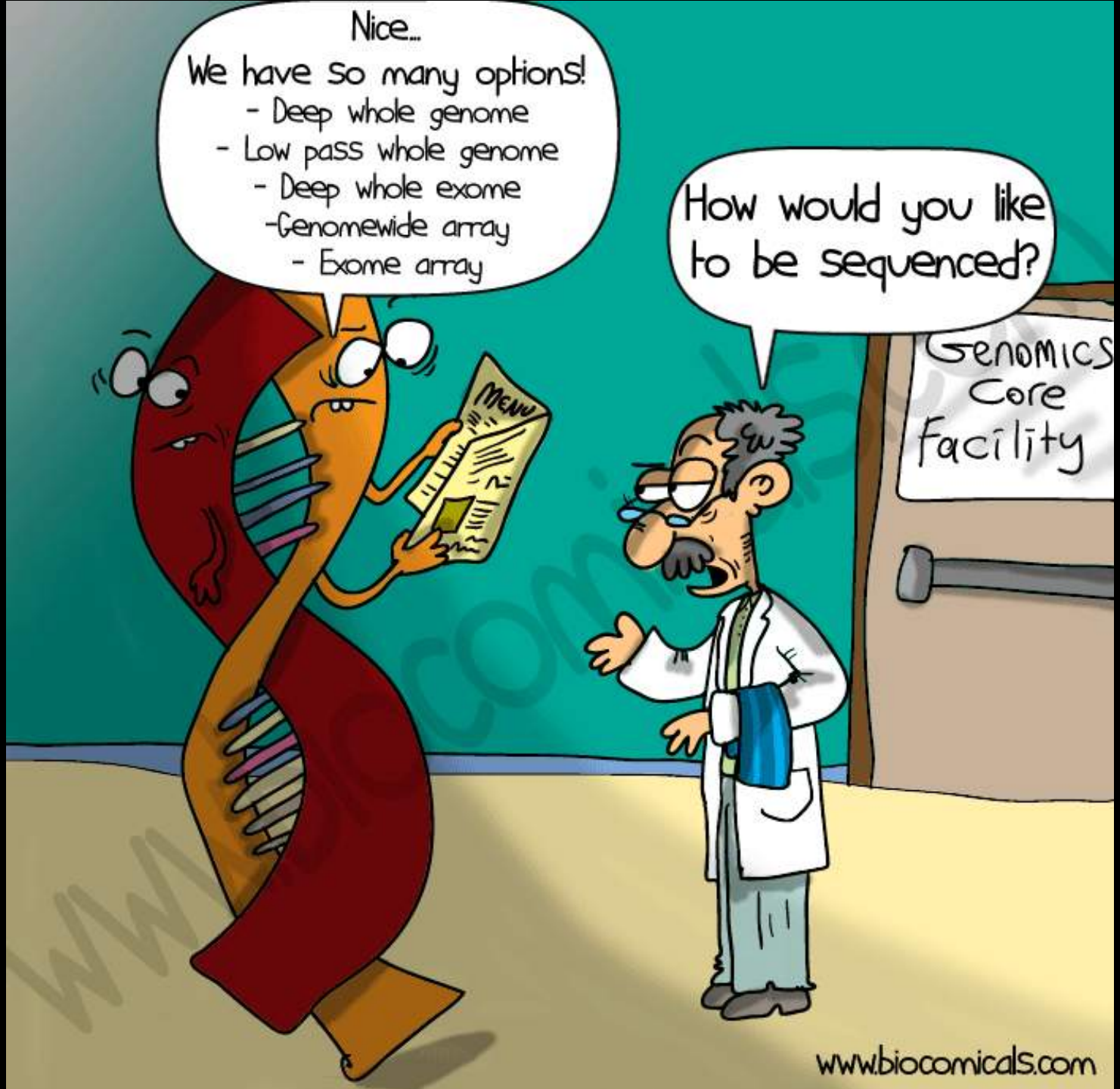
Natalie Hauser
Rebecca Belmonte
Gregory Wilson
Danielle Bosch
Paula Bubulya
Vandana Shashi
Slave Petrovski
Joshua Stone
Eun Young Park
Joris Veltman
Margje Sinnema
Connie Stumpel
Jos Draaisma

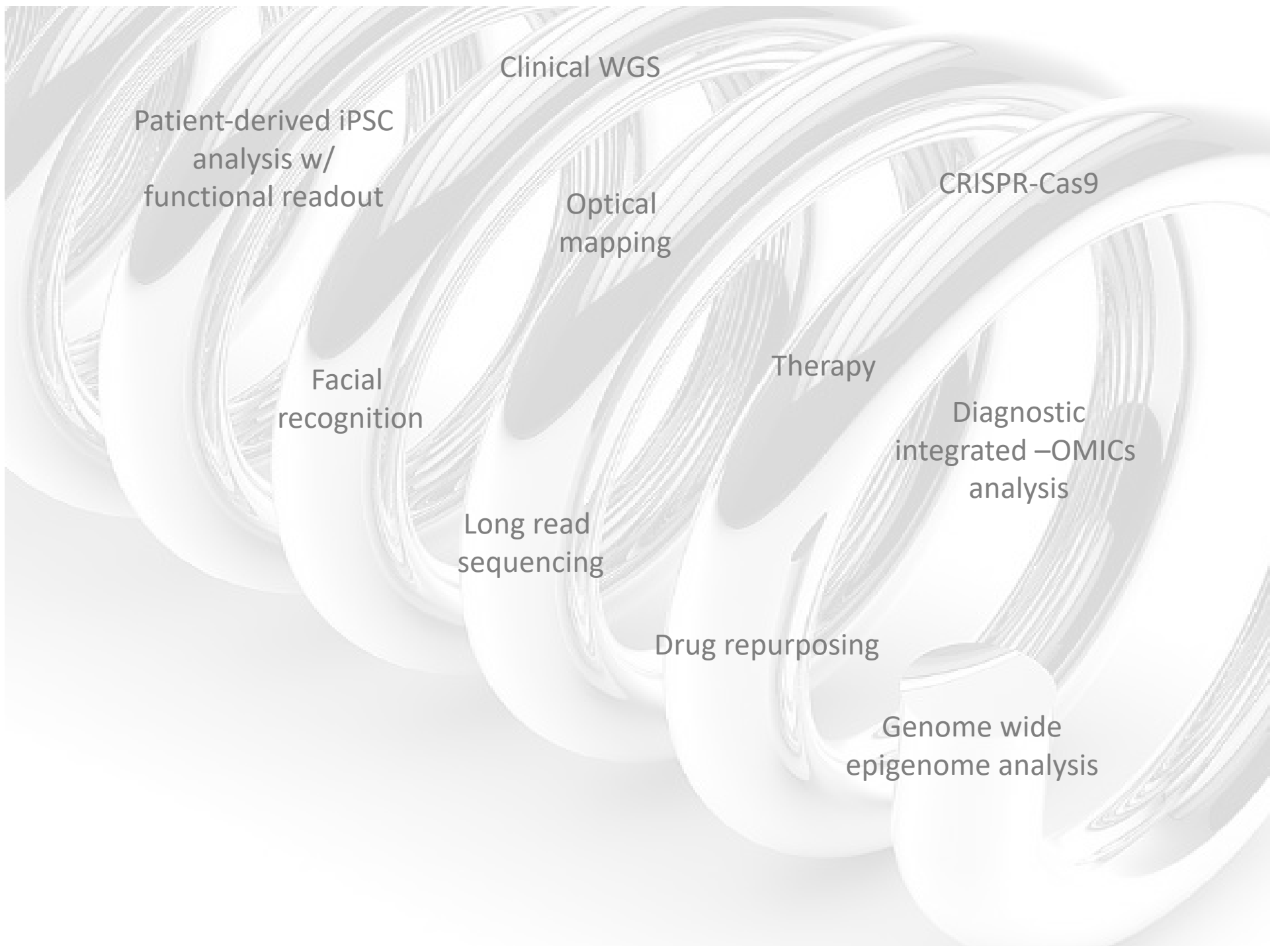
Joost Nicolai
UW CMG
Helger Yntema
Kristin Lindstrom
Bert de Vries
Tamison Jewett
Stephanie Santoro
Julie Vogt
The DDD study
Kristine Bachman
Andrea Seeley
Alyson Krokosky
Clesson Turner

Luis Rohena
Maja Hempel
Fanny Kortum
Davor Lessel
Axel Neu
Tim Strom
Dagmar Wieczorek
Nuria Bramswig
Franco Laccone
Jana Behunova
Helga Rehder
Christopher Gordon
Marlene Rio

Serge Romana
Sha Tang
Dima El-Khechen
Megan Cho
Kirsty McWalter
Ganka Douglas
Berivan Baskin
Amber Begtrup
Tara Funari
Kelly Schoch
Alexander Stegmann
Servi Stevens
Dong-Er Zhang

David Traver
Xu Yao
Daniel MacArthur
Han Brunner
Grazia Mancini
Richard Myers
Laurie Owen
Ssang-Taek Lim
David Stachura


Nice...
We have so many options!
- Deep whole genome
- Low pass whole genome
- Deep whole exome
-Genomewide array
- Exome array
MENU
How would you like to be sequenced?
Genomics Core Facility
www.biocomicals.com


Clinical WGS
Patient-derived iPSC analysis w/ functional readout
Optical mapping
CRISPR-Cas9
Therapy
Facial recognition
Diagnostic integrated –OMICs analysis
Long read sequencing
Drug repurposing
Genome wide epigenome analysis

Are the same diagnosis obtained?

Hypothesis:
WES will detect the vast majority, if not all, mutations identified in routine diagnostic genetic testing

Genetic path		WES			
		Yes	**Possible**	**No**	**Total**
Standard	**Yes**	8	-	3	**11**
	No	36	41	62	**139**
	Total	**44**	**41**	**65**	**150**

Standard care did better

WES did better

Standard care did better?

- In 3 patients WES did not identify the genetic cause of disease, whereas the standard of care did.

Disease	Cause	Why missed by WES
FraX	CGG triplet repeat expansion	*Triplet repeat expansions cannot be detected by WES*
PMR	Mosaic 27Mb duplication on chr 7 (20-30% cells)	*Mosaic CNV detection in WES still being optimized. In retrospect, CNV is visible in de data*
Benign hereditary chorea	NKX2-1 9bp duplication	*Larger indels are difficult to detect using WES and depend heavily on sequence read length*