# Diagnostics of Rare Diseases: Beyond the Exome

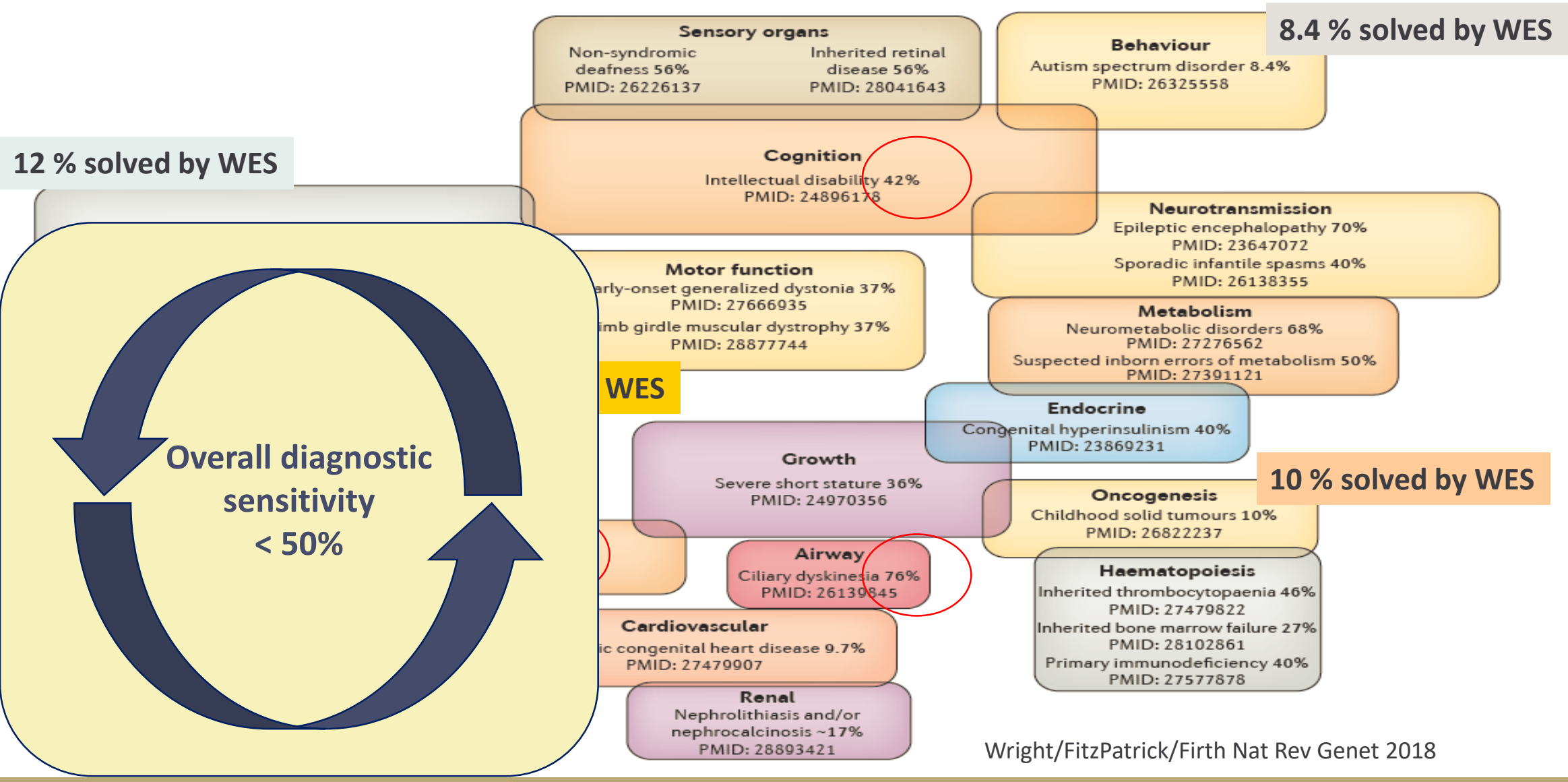
**Olaf Riess**

NCCT
NGS Competence Center Tübingen

Universitätsklinikum Tübingen

11.04.2021

# Diagnostic sensitivity of todays Whole Exome Sequencing in pediatric diseases




Wright/FitzPatrick/Firth Nat Rev Genet 2018

**UNSOLVED after WES:**
**50% of all patients with a rare disease will not have access to health care without having a clear diagnosis**

**300 Mio RD patients worldwide**
**150 Mio patients unsolved**

**30 Mio patients in Europe**
**15 Mio unsolved**

**3-4 Mio RD patients in Germany**
**1.5 Mio unsolved after WES**

# Use case: Rare diseases

## Diagnosis: A disease has to have a name to

- Communicate with doctors, relatives, authorities, health insurance
- Better define disease impact and progression and disease management
- Potentially apply disease-specific treatment
- Select disease information via internet

# Solving the Unsolved: Rare diseases

## Major limitations to increase diagnostic sensitivity:

- Bureaucracy, ethical hurdles,  data protection regulations
- Financing
- Limited knowledge
- Clinical processes in patient tissue/sample collection

- But presumably NOT technical limitations

# From Exome to Genome

# Limitations of Whole Exome Sequencing (WES)

Important: Type of enrichment system: SureSelectXT Human All Exon v6

Statistics of coverage: complete coding sequence +/-5bp intronic region

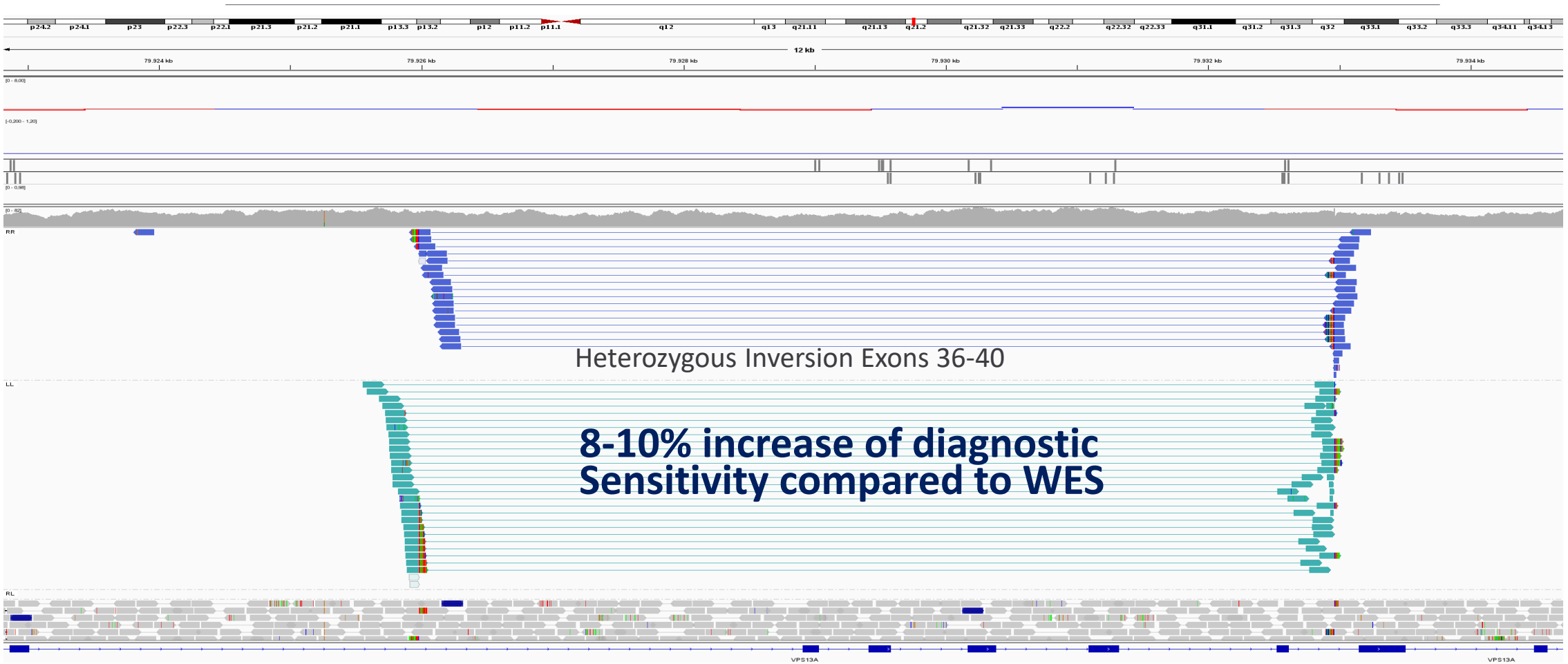
depth of sequencing (at least 20 fold)

coverage: 98.99%

## Limitations of WES:

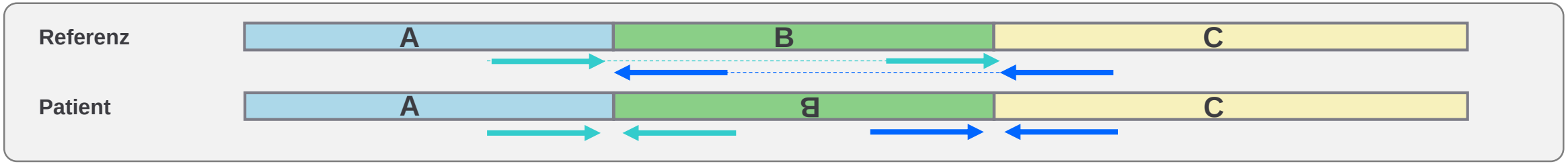
- Coverage
- Copy number
- Aberrant splicing
- Structural aberrations
- Regulatory regions

# WGS in unsolved chorea acanthocytosis: Structural variants in VPS13A



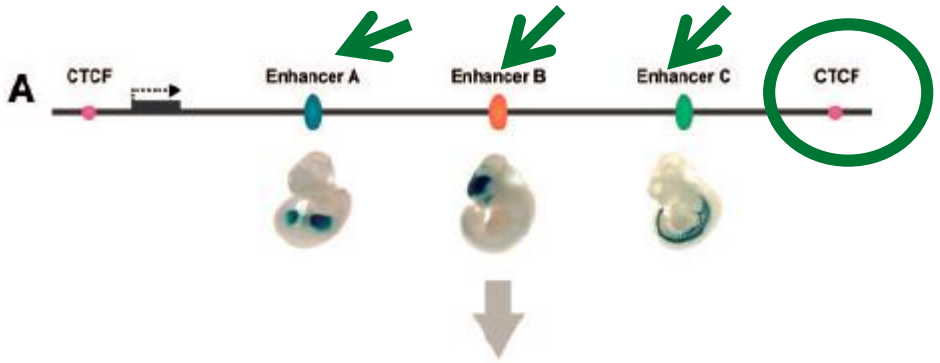

Heterozygous Inversion Exons 36-40

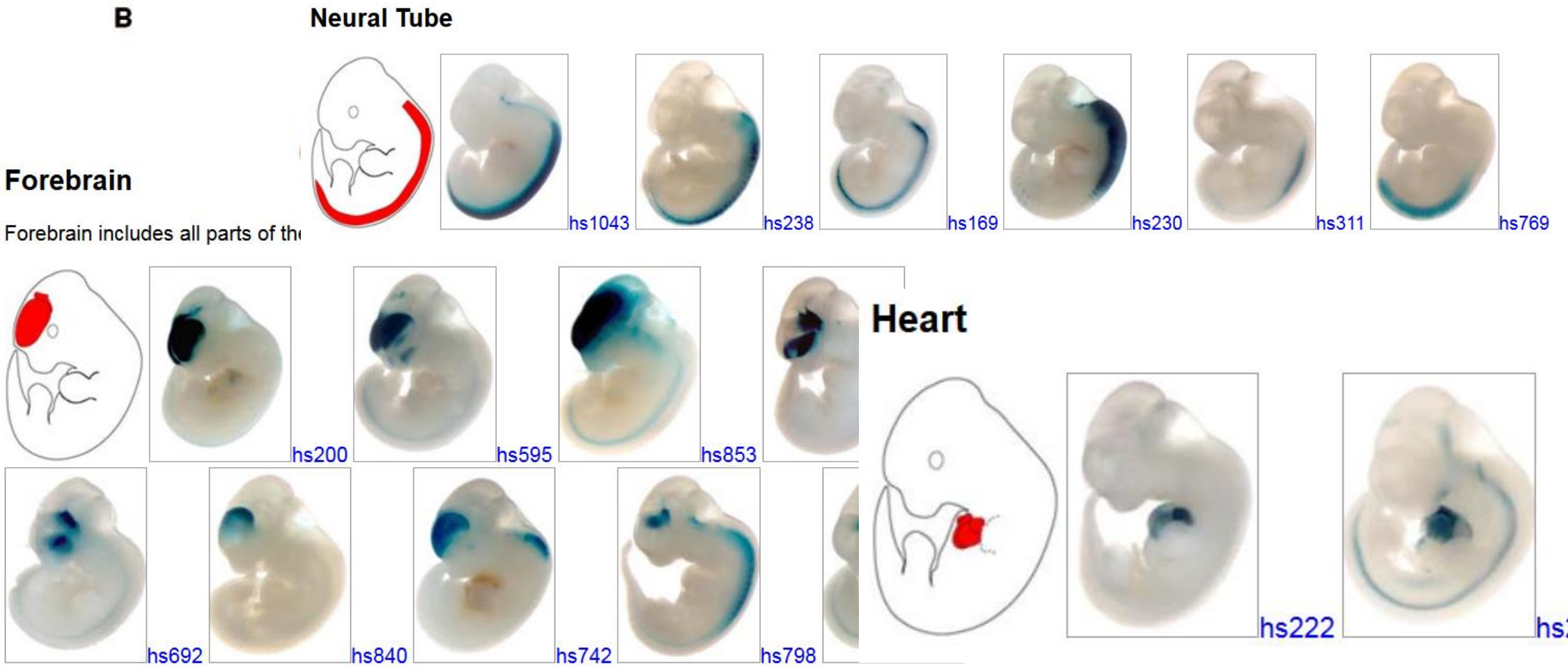
**8-10% increase of diagnostic Sensitivity compared to WES**



# Genome: Regulatory regions

# Regulatory regions: Enhancers



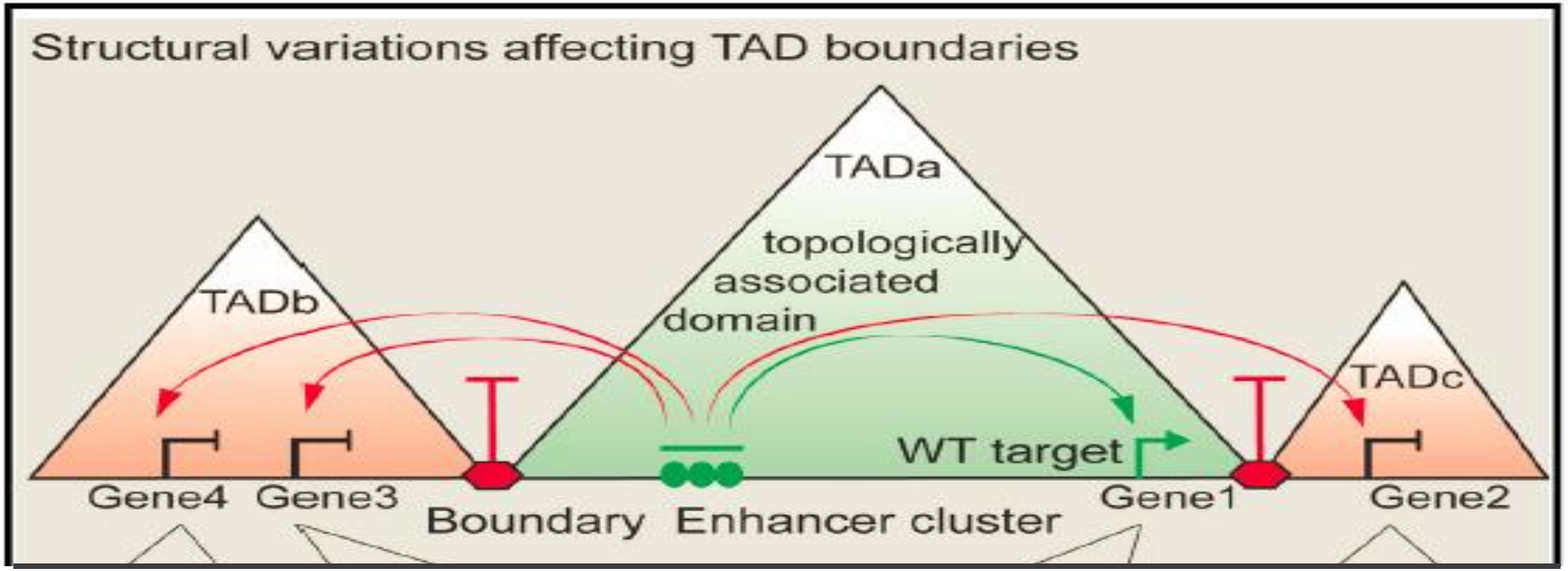

https://enhancer.lbl.gov/gallery_n.html

Cell



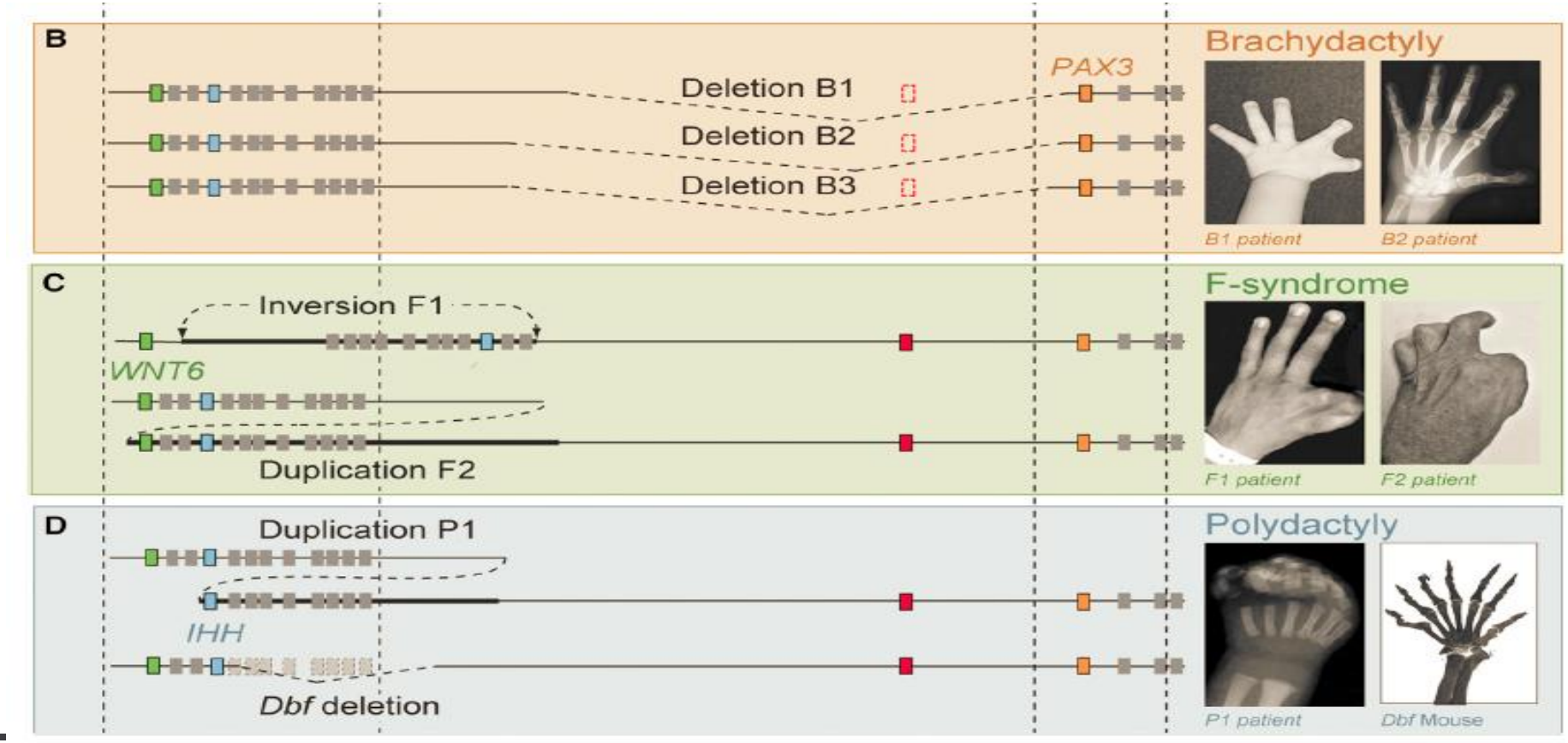
# Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions

Darío G. Lupiáñez,[1,2] Katerina Kraft,[1,2] Verena Heinrich,[2] Peter Krawitz,[1,2] Francesco Brancati,[3] Eva Klopocki,[4] Denise Horn,[2] Hülya Kayserili,[5] John M. Opitz,[6] Renata Laxova,[6] Fernando Santos-Simarro,[7,8] Brigitte Gilbert-Dussardier,[9] Lars Wittler,[10] Marina Borschiwer,[1] Stefan A. Haas,[11] Marco Osterwalder,[12] Martin Franke,[1,2] Bernd Timmermann,[13] Jochen Hecht,[1,14] Malte Spielmann,[1,2,14] Axel Visel,[12,15,16] and Stefan Mundlos[1,2,14,*]



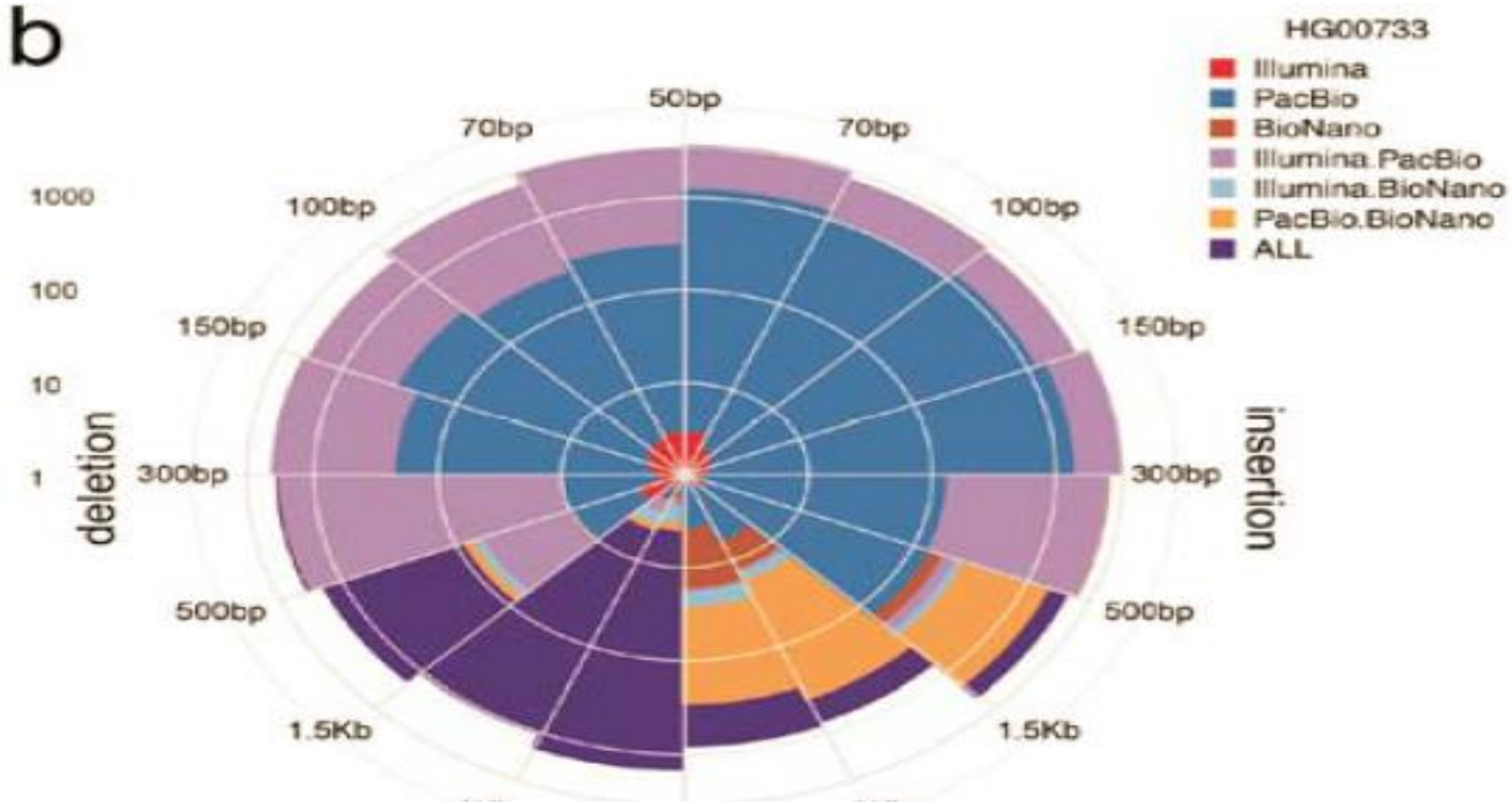
# Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions

Darío G. Lupiáñez,[1,2] Katerina Kraft,[1,2] Verena Heinrich,[2] Peter Krawitz,[1,2] Francesco Brancati,[3] Eva Klopocki,[4] Denise Horn,[2] Hülya Kayserili,[5] John M. Opitz,[6] Renata Laxova,[6] Fernando Santos-Simarro,[7,8] Brigitte Gilbert-Dussardier,[9] Lars Wittler,[10] Marina Borschiwer,[1] Stefan A. Haas,[11] Marco Osterwalder,[12] Martin Franke,[1,2] Bernd Timmermann,[13] Jochen Hecht,[1,14] Malte Spielmann,[1,2,14] Axel Visel,[12,15,16] and Stefan Mundlos[1,2,14,*]



# Genome: Towards long read sequencing

# Combination of short and long read NGS
# Required for sensitive detection of deletions and insertions




NCCT
NGS Competence Center Tübingen

NGS-CN
Next Generation Sequencing Network

**Multi-platform discovery of haplotype-resolved structural variation in human genomes**

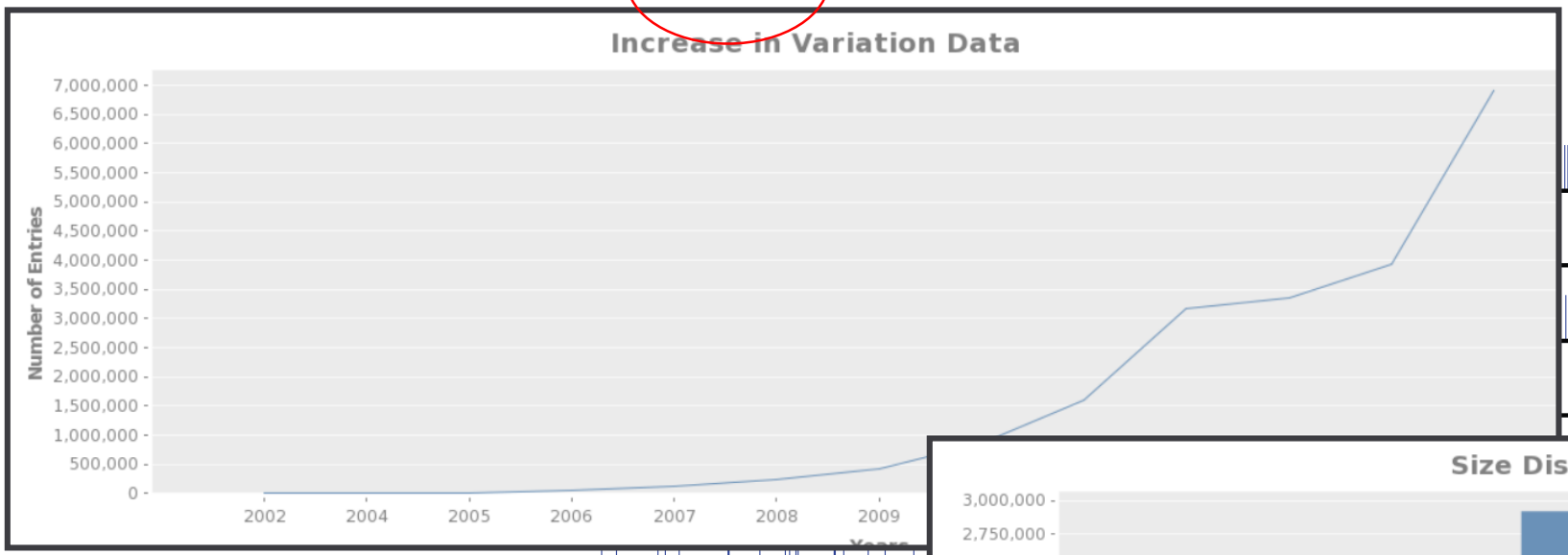
Mark J.P. Chaisson[1,2*], Ashley D. Sanders[3*], Xuefang Zhao[4,5*], Ankit Malhotra[6†], David

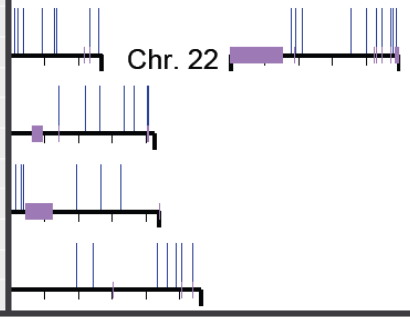
# Characterization of missing human genome sequences and copy-number polymorphic insertions

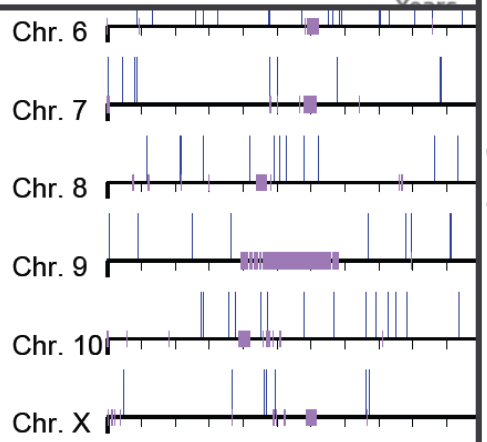
Jeffrey M Kidd[1], Nick Sampas[2], Francesca Antonacci[1], Tina Graves[3], Robert Fulton[3], Hillary S Hayden[1], Can Alkan[1], Maika Malig[1], Mario Ventura[4], Giuliana Giannuzzi[4], Joelle Kallicki[3], Paige Anderson[2], Anya Tsalenko[2], N Alice Yamada[2], Peter Tsang[2], Rajinder Kaul[1], Richard K Wilson[3], Laurakay Bruhn[2] & Evan E Eichler[1,5]

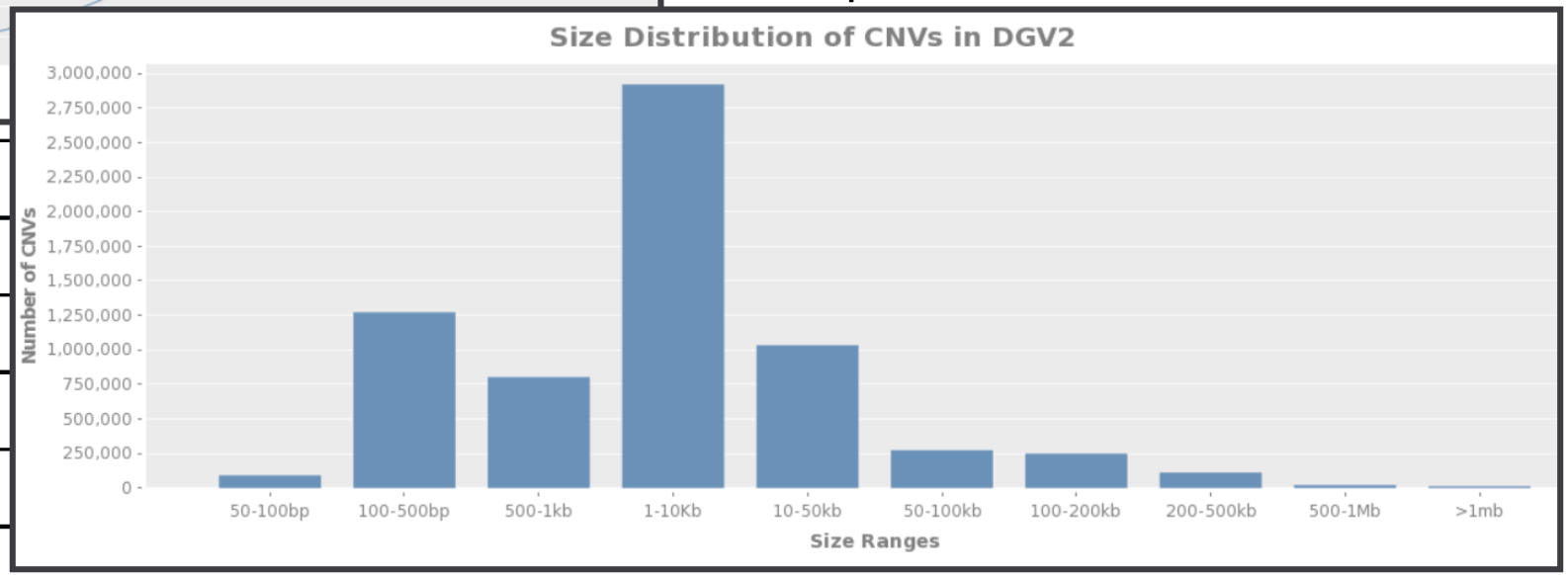
NATURE METHODS | VOL.7 NO.5 | MAY 2010 |

**2,363** new insertion sequences corresponding to **720** genomic loci.



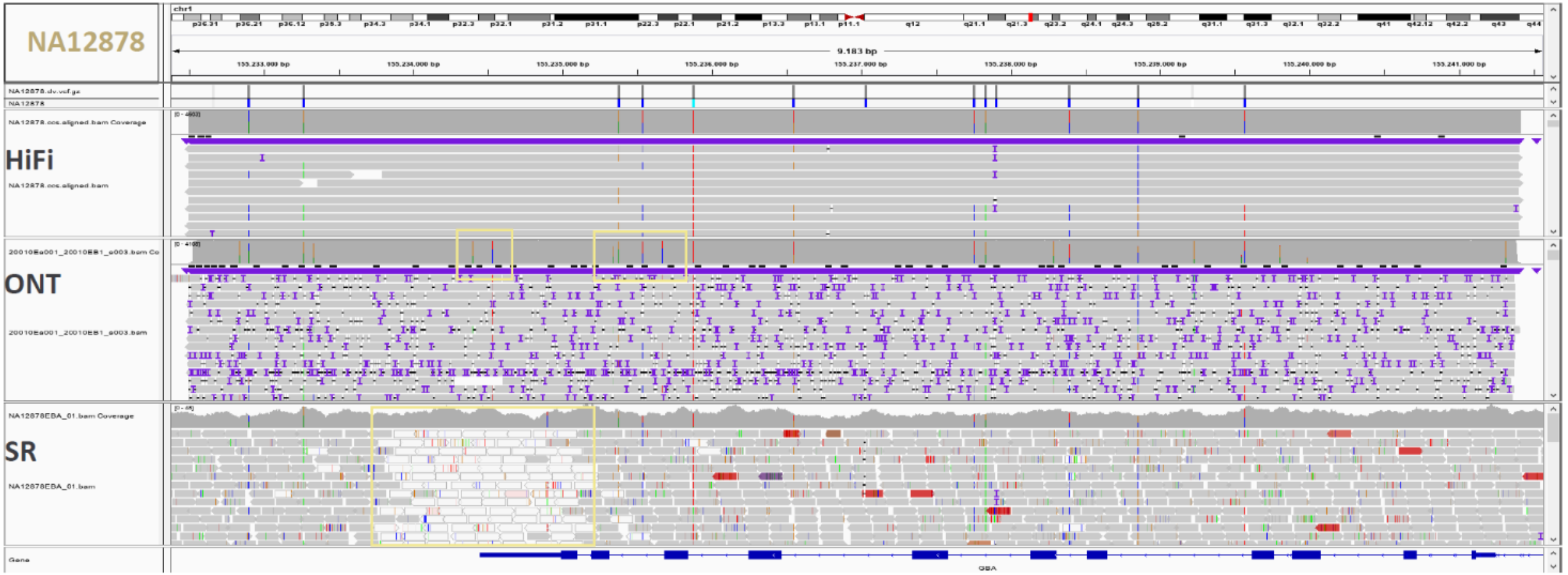



# Highly Homologous Genes neglected in Short-Read WGS

HiFi reads to distinguish functional genes (GBA) from its pseudogenes (ψGBA) and allocate variants

--> Targeted Application with Amplicons: HiFi reads to screen a large cohort



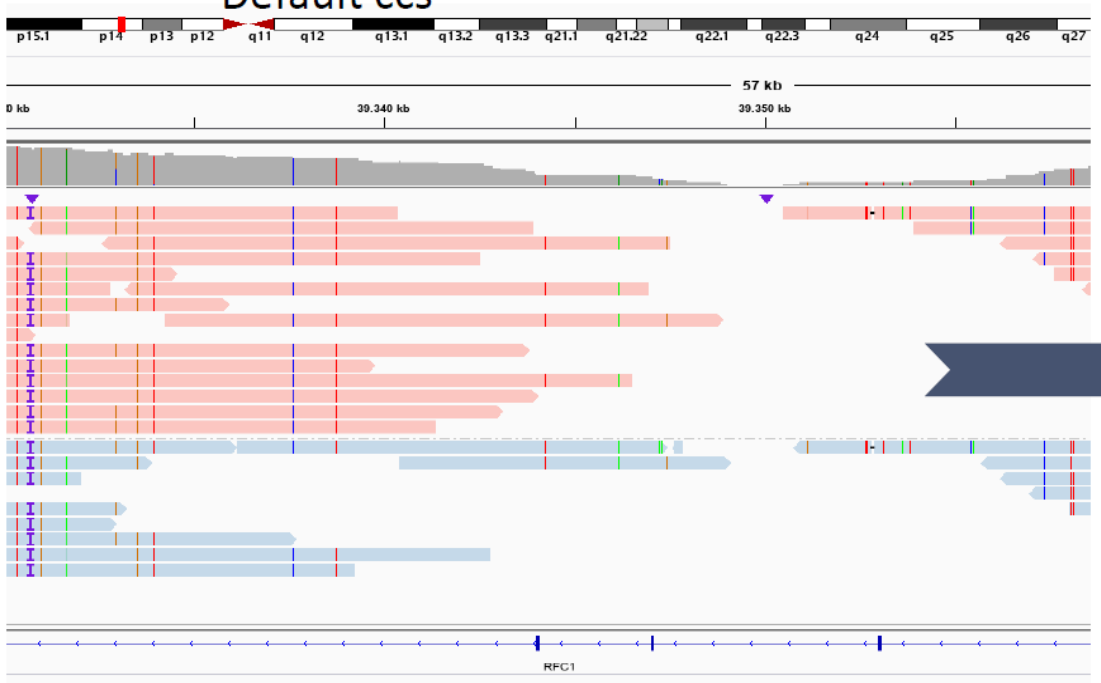

Bad mapping quality

# WGS – Variant Calling: HiFi data from NCCT establishment

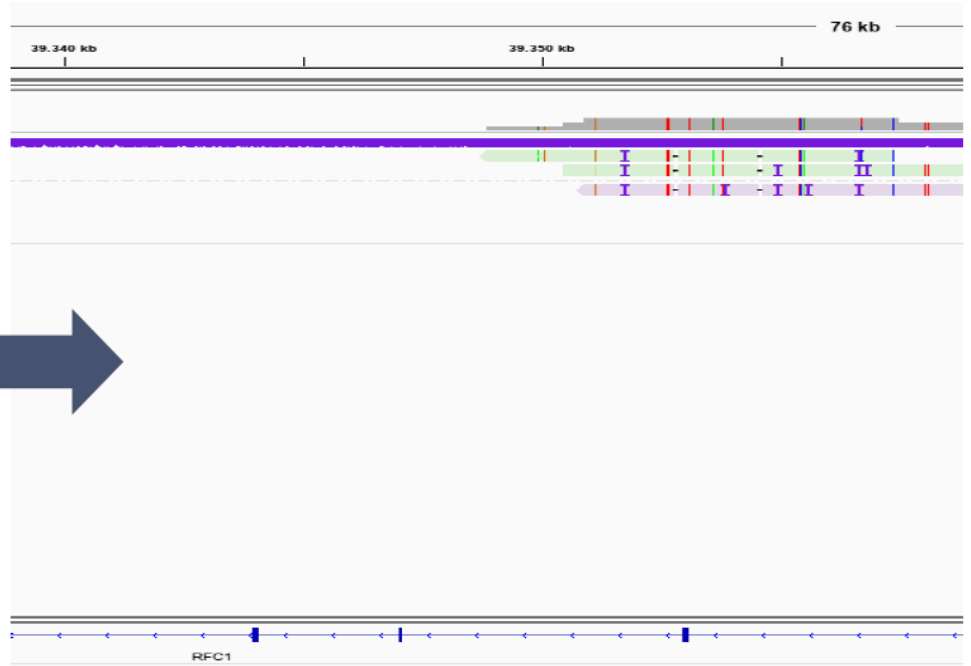
## RFC1 Repeat Expansion

**I19D045a01**

Default ccs




Local ccs without heuristics




**No reads mapping to RFC1**

Problem: Not enough reads in close distance to RFC1 region.

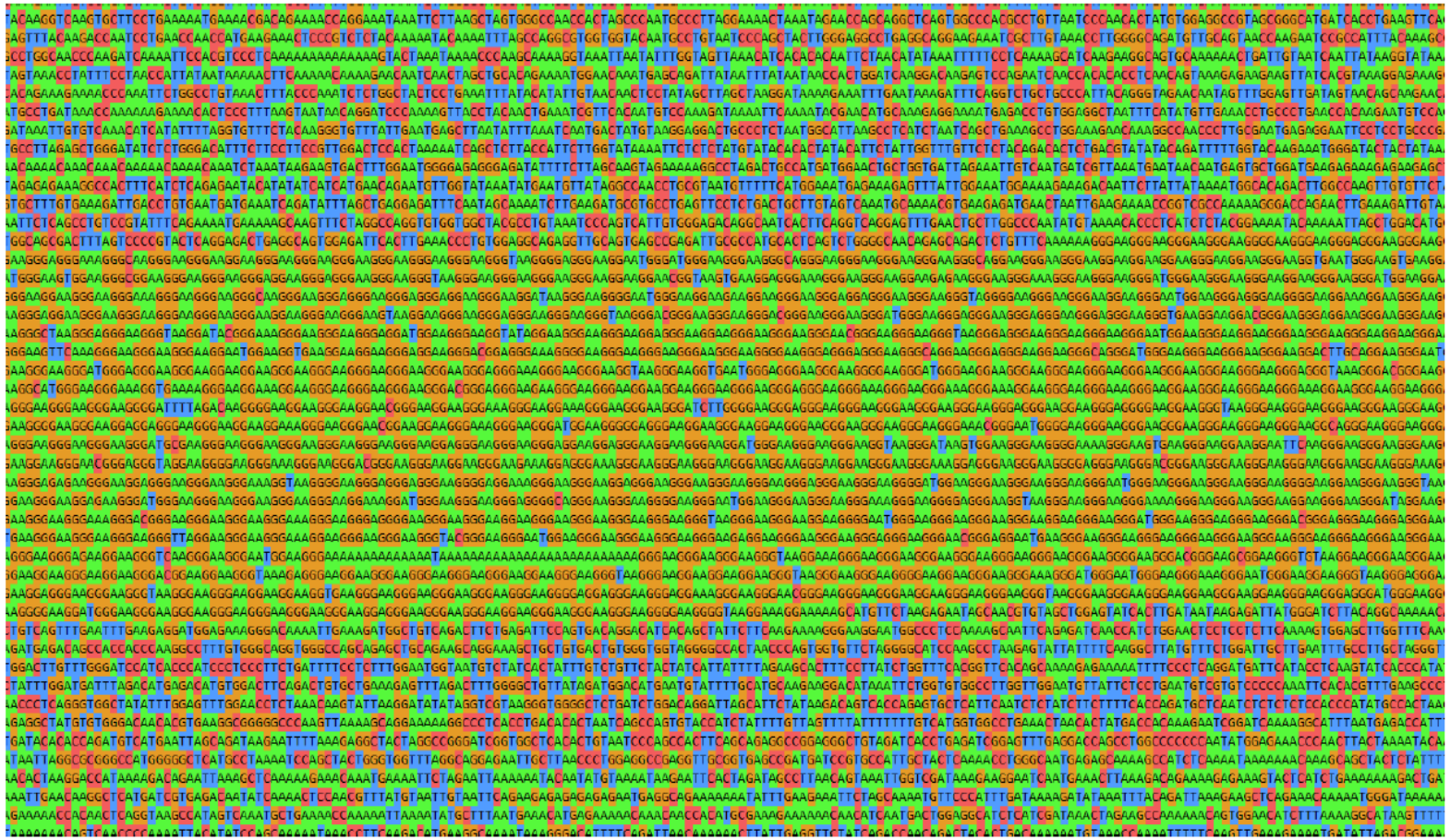
Currently all steps are repeated with larger window around RE site.

# WGS – Variant Calling: HiFi data from NCCT establishment

## RFC1 Repeat Expansion

**I19D045a01**

**Local assembly confirms repeat expansion** undetected by mapping



Length:
3753 bases

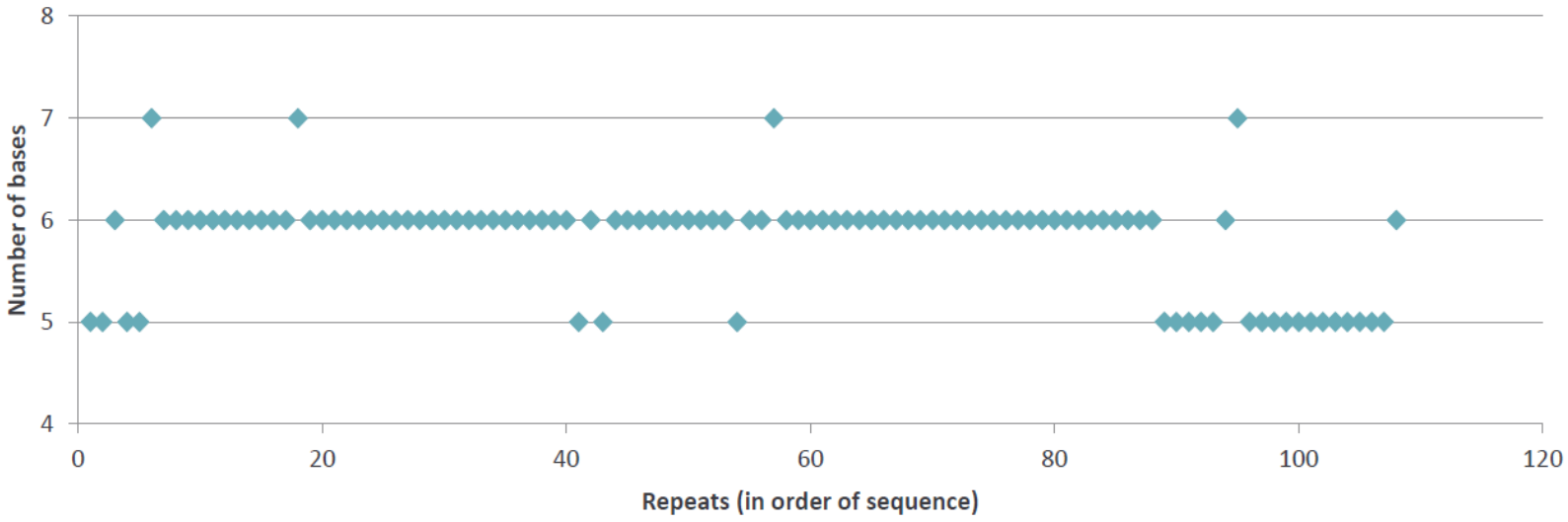
# WGS – Variant Calling: HiFi data from NCCT establishment

## RFC1 Repeat Expansion

**I19D045a02**

### HiFi Read (mapped) Sequence of the 571-bp Insertion and flanking bases

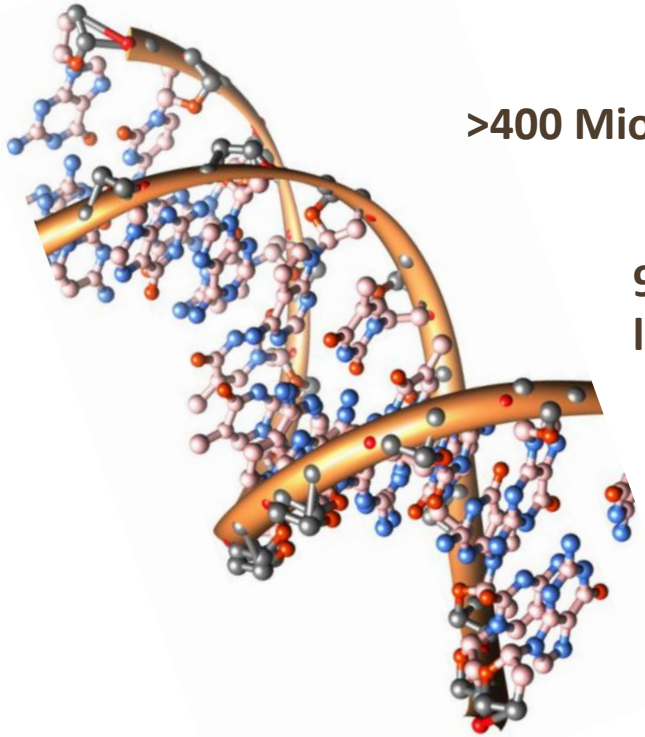
AAAGGAAAGGAAAGGGAAAGGAAGGGAAAGGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAG
GGAAAGGGAAAGGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAG
GGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGAAAGGGAAAGGGAAAGGGAAAGGGGA
AAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAGGGAAAGGGAAAGGGAAAAGGGAAAGGGAAAGGGAAAGGGAAA
AGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAA
GGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAGGGAAAAGGAAAAGGAAAAGGAAAGGAAAGGAAAGGAAAGGAA
AGGAAAAGGAAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAAGAAAAGAAAAAA



☐ Yes/Ja ☐ No/Nein

# Complexity of analysing genomes in the clinical context

## TOPMed study (Trans-Omics for Precision Medicine)



53,831 Genomes analysed

>400 Mio variants

97% of the variants
less than 1%

53% present
in only 1 individual

-> 4,6 Mio protein-coding variants
Total 230,000 putative loss of function variants in 18.493 genes
-> 104,000 frameshift variants
-> 97,000 putative splice and truncation variants

2.5 unique pLOF per individual

Additional read out needed !

https://www.shutterstock.com/de/image-illustration/organic-chemistry-model-dna-molecule-illustration-93597241?id=93597241



Nature Vol.590, 2021

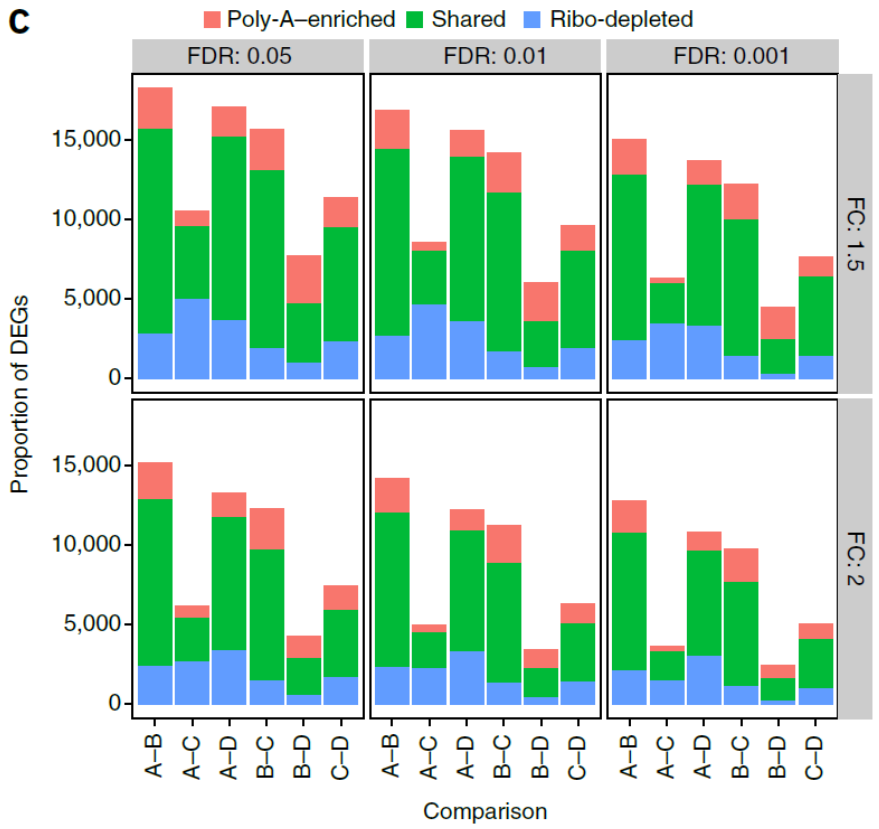
# Genome: Complemented with RNAseq

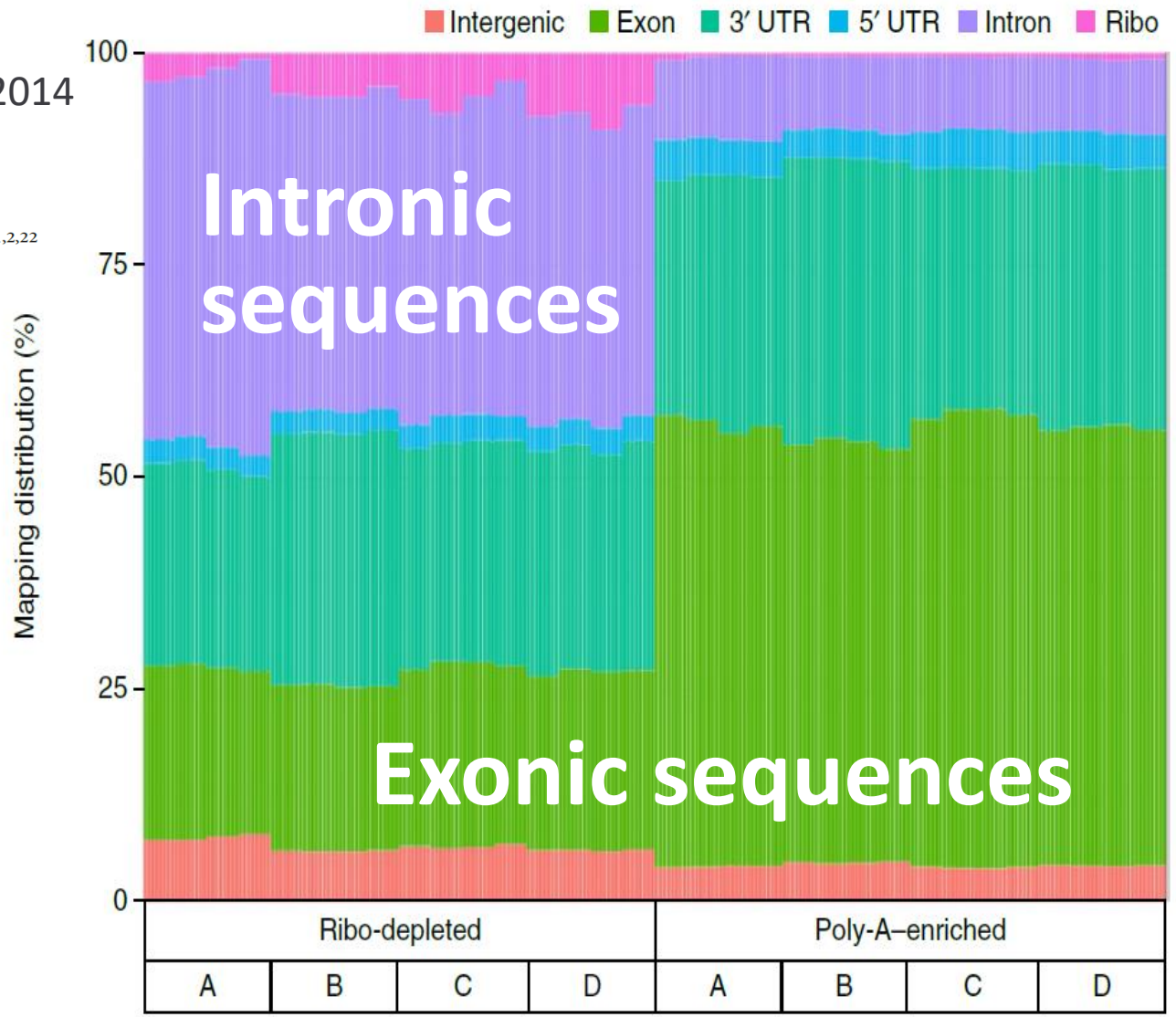
# RNAseq in research and diagnostics – Technical „details“

Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study

Nature Biotechn 2014

Sheng Li[1,2,24], Scott W Tighe[3,24], Charles M Nicolet[4], Deborah Grove[5], Shawn Levy[6], William Farmerie[7], Agnes Viale[8], Chris Wright[9], Peter A Schweitzer[10], Yuan Gao[11], Dewey Kim[11], Joe Boland[12], Belynda Hicks[12], Ryan Kim[13,23], Sagar Chhangawala[1,2], Nadereh Jafari[14], Nalini Raghavachari[15], Jorge Gandara[1,2], Natàlia Garcia-Reyero[16], Cynthia Hendrickson[6], David Roberson[12], Jeffrey A Rosenfeld[17], Todd Smith[18], Jason G Underwood[19], May Wang[20], Paul Zumbo[1,2], Don A Baldwin[21], George S Grills[10] & Christopher E Mason[1,2,22]



# RNAseq in diagnostics

94 patients / 1600 controls
343 outliers per sample
7.5% diagnostic rate

Additional 16.7% improved candidate gene resolution

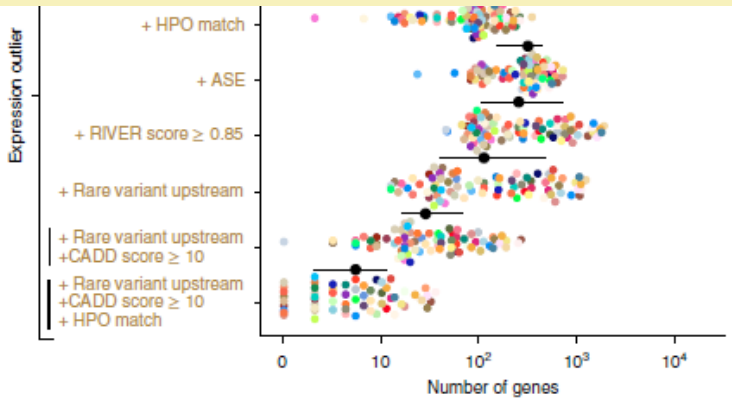
nature medicine
LETTERS
https://doi.org/10.1038/s41591-019-0457-8

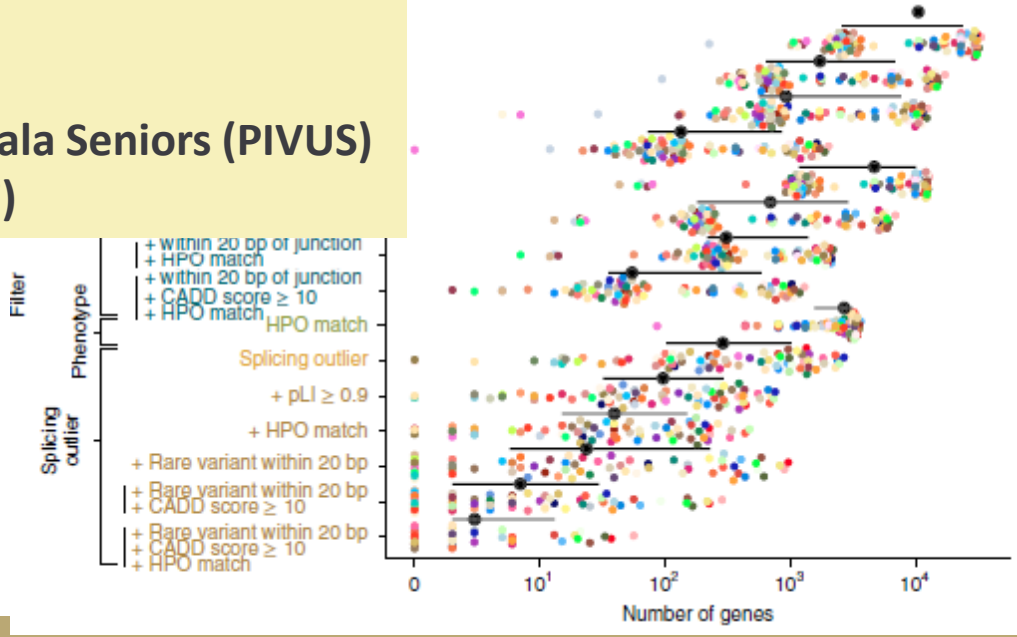
## Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts

Laure Frésard[1*], Craig Smail[2], Nicole M. Ferraro[2], Nicole A. Teran[3], Xin Li[1], Kevin S. Smith[1], Devon Bonner[4], Kristin D. Kernohan[5], Shruti Marwaha[4,6], Zachary Zappala[3], Brunilda Balliu[1], Joe R. Davis[3], Boxiang Liu[7], Cameron J. Prybol[3], Jennefer N. Kohler[4], Diane B. Zastrow[4], Chloe M. Reuter[4], Dianna G. Fisk[8], Megan E. Grove[8], Jean M. Davidson[4], Taila Hartley[9], Ruchi Joshi[8], Benjamin J. Strober[10], Sowmithri Utiramerur[8], Undiagnosed Diseases Network[11], Care4Rare Canada Consortium[11], Lars Lind[12], Erik Ingelsson[6,13], Alexis Battle[10,14], Gill Bejerano[15,16,17,18], Jonathan A. Bernstein[16], Euan A. Ashley[3,4,13], Kym M. Boycott[9], Jason D. Merker[1,8,19], Matthew T. Wheeler[4,6] and Stephen B. Montgomery[1,3*]


**RNAseq control data:**
909: Depression Genes and Network (DGN)
65: Prospective Investigation of the Vasculature in Uppsala Seniors (PIVUS)
620: Genotype-Tissue Expression Consortium (GTEx vs 7)

Expression outliers




Splicing outliers

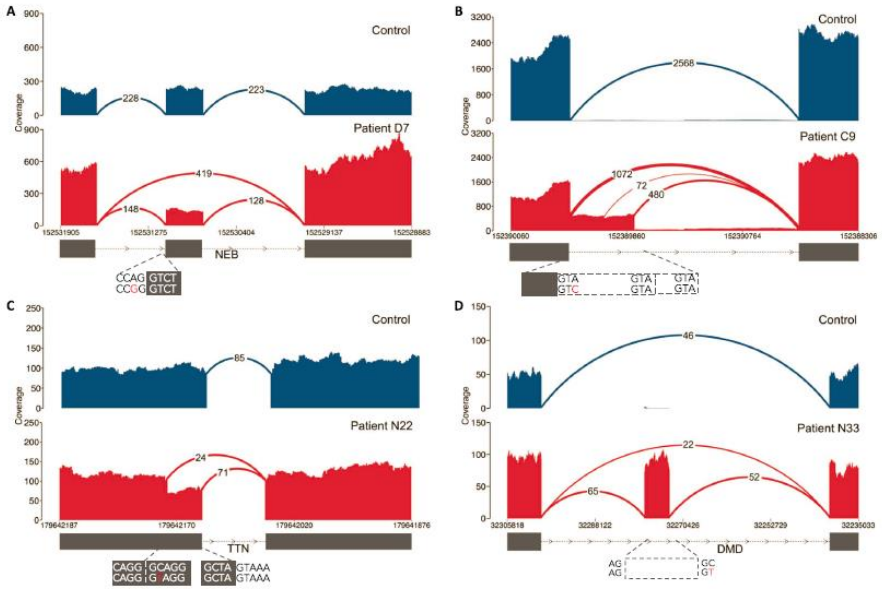
# RNAseq in diagnostics

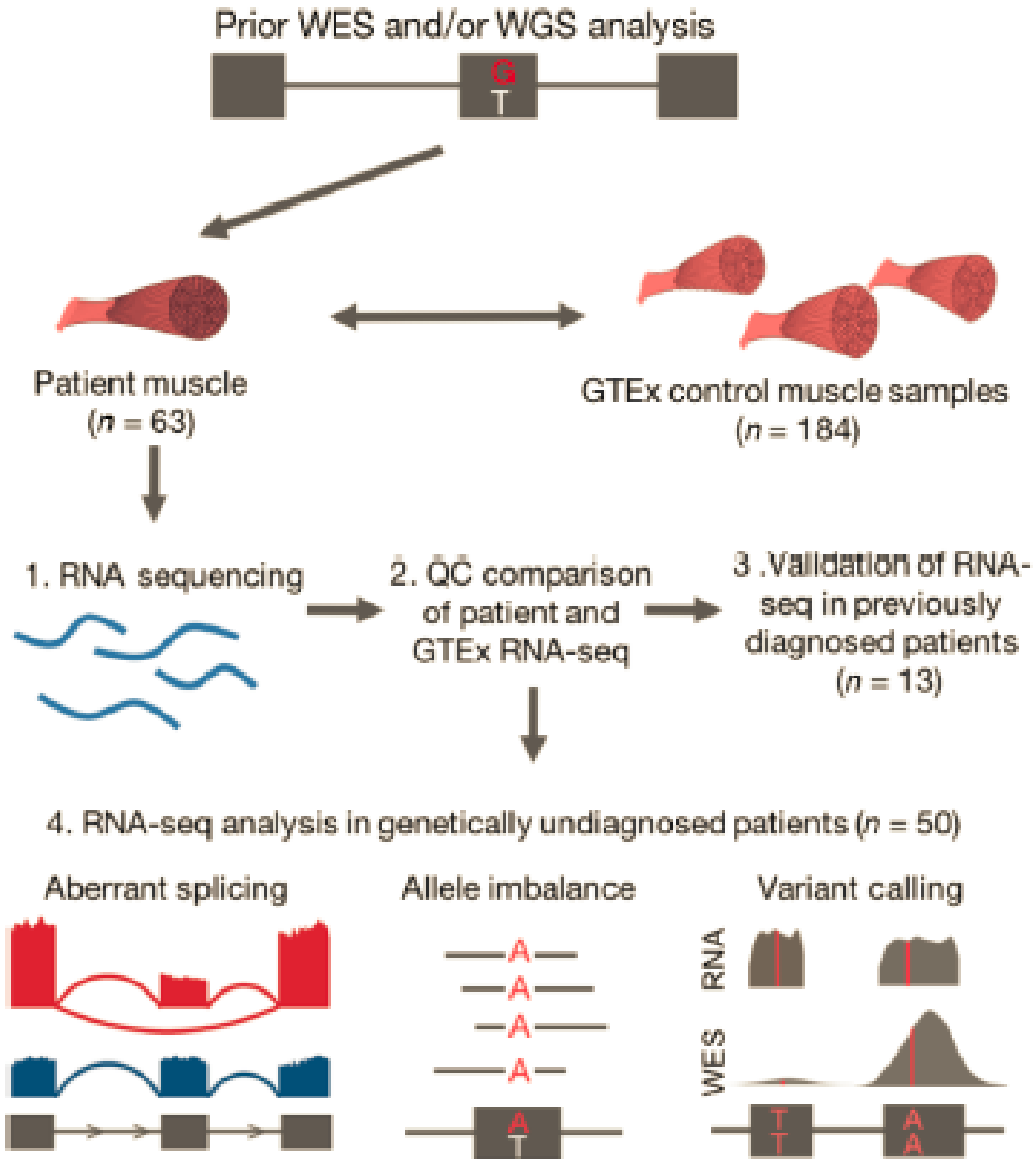
SCIENCE TRANSLATIONAL MEDICINE | RESEARCH ARTICLE

GENETIC DIAGNOSIS

## Improving genetic diagnosis in Mendelian disease with transcriptome sequencing

Beryl B. Cummings,[1,2,3] Jamie L. Marshall,[1,2] Taru Tukiainen,[1,2] Monkol Lek,[1,2,4,5] Sandra Donkervoort,[6] A. Reghan Foley,[6] Veronique Bolduc,[6] Leigh B. Waddell,[4,5] Sarah A. Sandaradura,[4,5] Gina L. O'Grady,[4,5] Elicia Estrella,[7] Hemakumar M. Reddy,[8] Fengmei Zhao,[1,2] Ben Weisburd,[1,2] Konrad J. Karczewski,[1,2] Anne H. O'Donnell-Luria,[1,2] Daniel Birnbaum,[1,2] Anna Sarkozy,[9] Ying Hu,[6] Hernan Gonorazky,[10] Kristl Claeys,[11] Himanshu Joshi,[5] Adam Bournazos,[4,5] Emily C. Oates,[4,5] Roula Ghaoui,[4,5] Mark R. Davis,[12] Nigel G. Laing,[12,13] Ana Topf,[14] Genotype-Tissue Expression Consortium, Peter B. Kang,[7,8] Alan H. Beggs,[7] Kathryn N. North,[15] Volker Straub,[14] James J. Dowling,[10] Francesco Muntoni,[9] Nigel F. Clarke,[4,5]* Sandra T. Cooper,[4,5] Carsten G. Bönnemann,[6] Daniel G. MacArthur[1,2†]



# RNAseq in diagnostics

RNASeq revealed clear indication of loss of the second allele

**Ataxia patient**

WES defined heterozygous frameshift

**Mutation of one allele in SIL1**

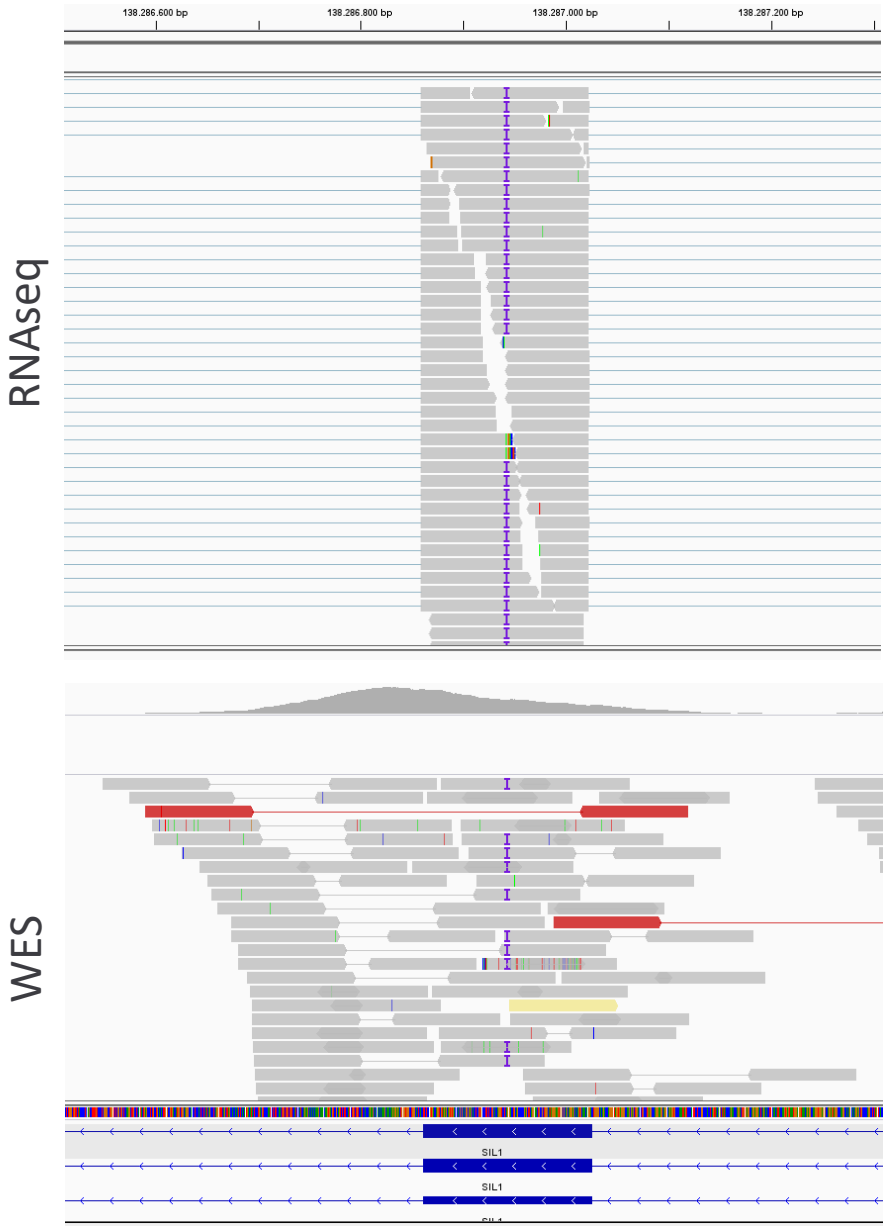
If both alleles affected, it would cause

**Marinesco-Sjörgen-Syndrome**

No indication of deletion of the second allele

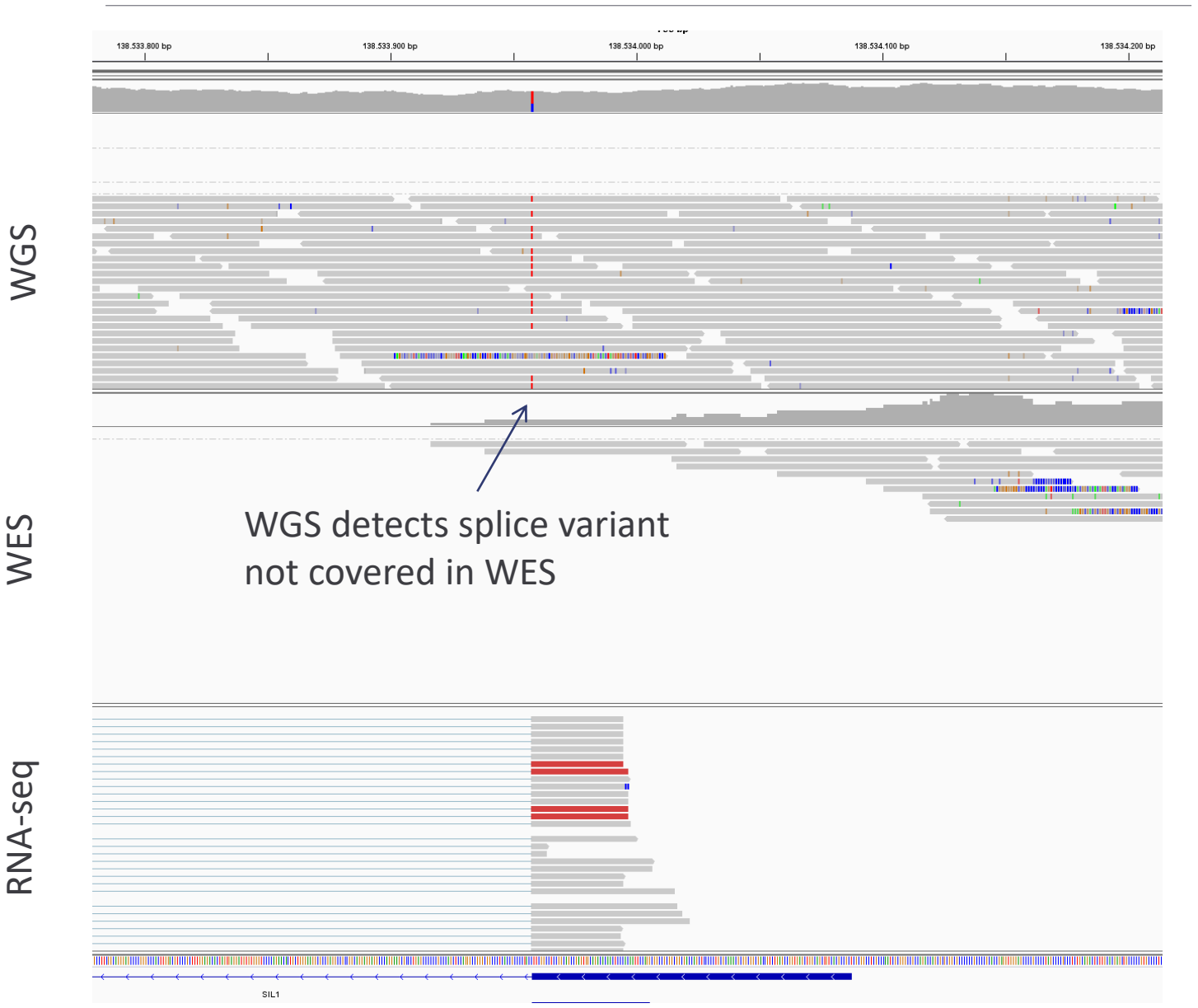
**Phenotype of the patient would fit MSS**

Variant 1 (frameshift) DX174053_01



# RNAseq in diagnostics

## WGS finds Second Hit



# The Genome is not „naked“: The Epigenome

# Don‘t forget the epigenome in diagnostics !

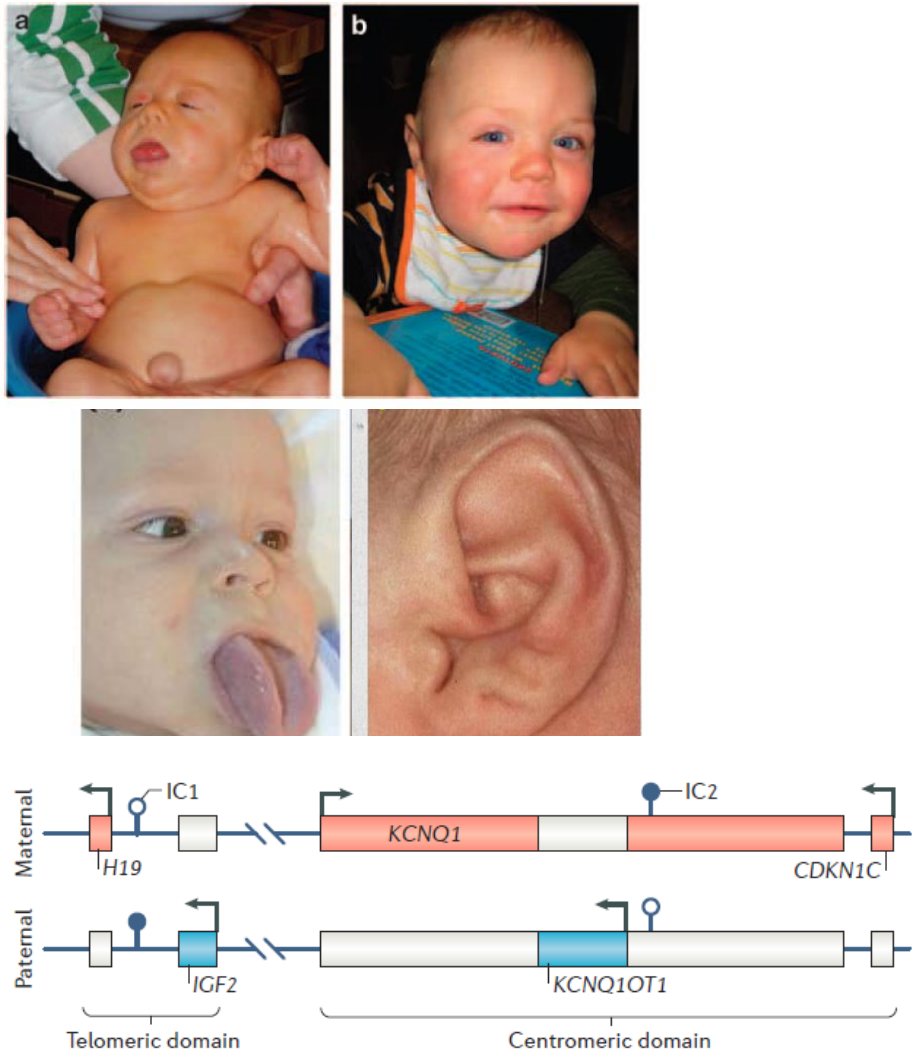



Figure 3 | **The Beckwith–Wiedemann syndrome locus at chromosome 11p15.5.**

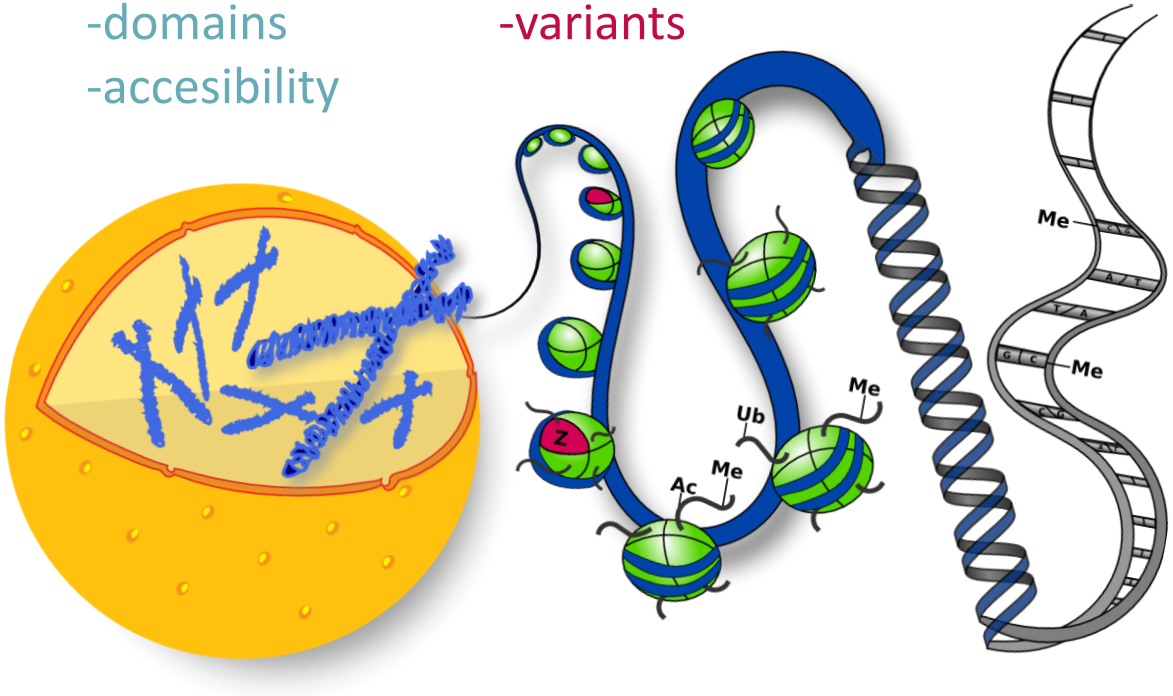
Beckwith-Wiedemann-Syndrome (BWS)
Transient Neonatal Diabetes mellitus
Birk-Barel intellectual disability syndrome
Silver-Russel syndrome
Temple syndrome
Kagami-Ogata syndrome
Central Precocious Puberty
Prader-Willi syndrome
Angelman syndrome
Schaaf-Yang syndrome
Mulchandani-Bhoi-Conlin syndrome
Pseudo-hypoparathyoridism

# Epigenome Profiling using NGS

**Chromatin**
-structure
-domains
-accesibility

**Histone**
-modifications
-variants

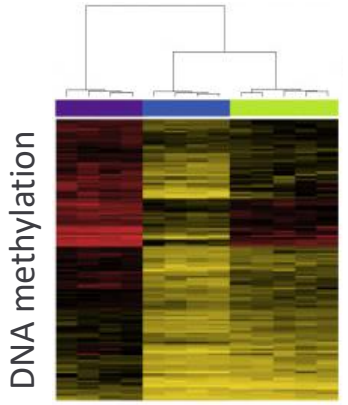
**DNA**
-modifications




| | | |
|---|---|---|
| 3C, 4C, 5C, Hi-C, ChIA-PET, DNAse-seq, FAIRE-seq, ATAC-seq | ChIP-seq, Mnase-seq | WGBS, RRBS, MeDIP-seq, Methyl-seq |

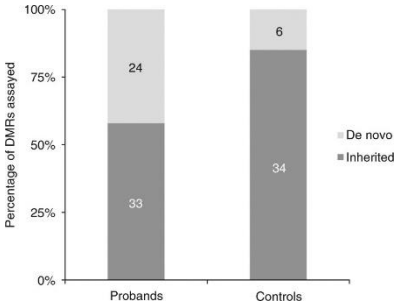
## Rare Diseases

### *Episignatures*




Erfan Aref-Eshghi *et al* 2020

### *Epivariants*




Mafalda Barbosa *et al* 2018

Please cite this article in press as: Aref-Eshghi et al., Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected..., The American Journal of Human Genetics (2019), https://doi.org/10.1016/j.ajhg.2019.03.008

**Methylation-sensitive arrays for detection of epigenomic disease specific profiles**

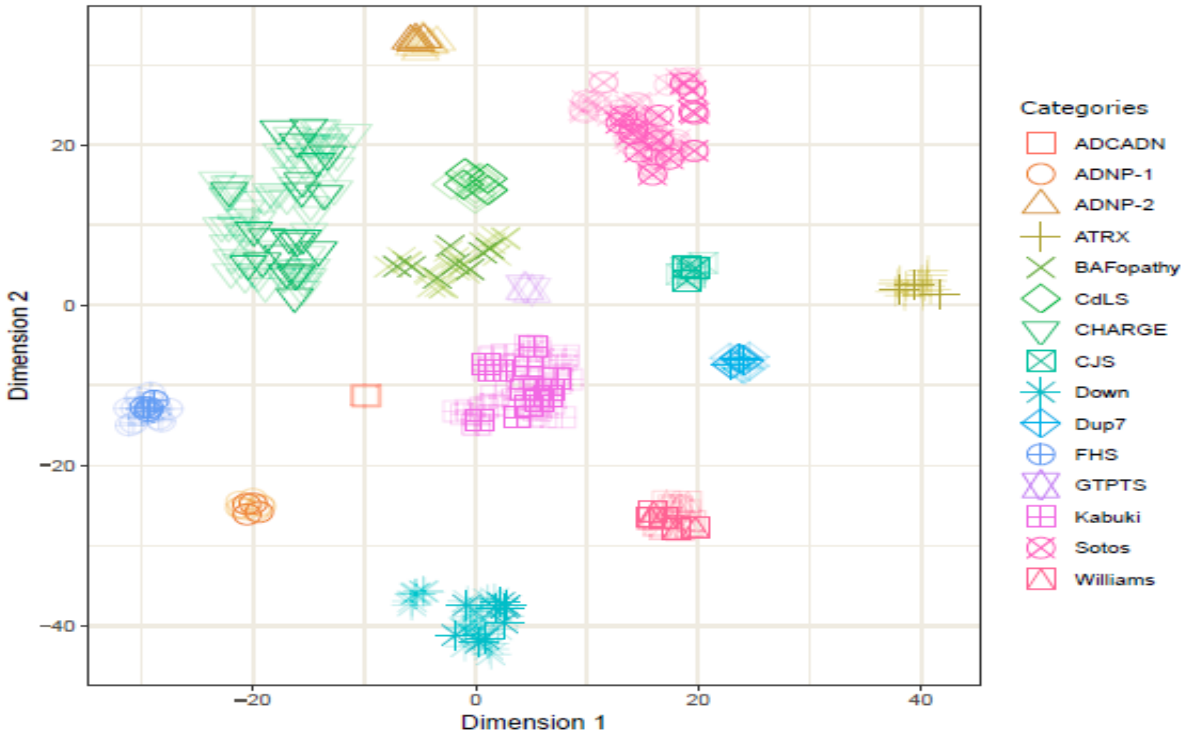
# Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions

Erfan Aref-Eshghi,[1,2] Eric G. Bend,[3] Samantha Colaiacovo,[4] Michelle Caudle,[4] Rana Chakrabarti,[4] Melanie Napier,[4] Lauren Brick,[5] Lauren Brady,[5] Deanna Alexis Carere,[2] Michael A. Levy,[1,2] Jennifer Kerkhof,[2] Alan Stuart,[2] Maha Saleh,[4] Arthur L. Beaudet,[6] Chumei Li,[5] Maryia Kozenko,[5] Natalya Karp,[4] Chitra Prasad,[4] Victoria Mok Siu,[4] Mark A. Tarnopolsky,[5] Peter J. Ainsworth,[1,2] Hanxin Lin,[1,2] David I. Rodenhiser,[7] Ian D. Krantz,[8] Matthew A. Deardorff,[8] Charles E. Schwartz,[3] and Bekim Sadikovic[1,2,*]

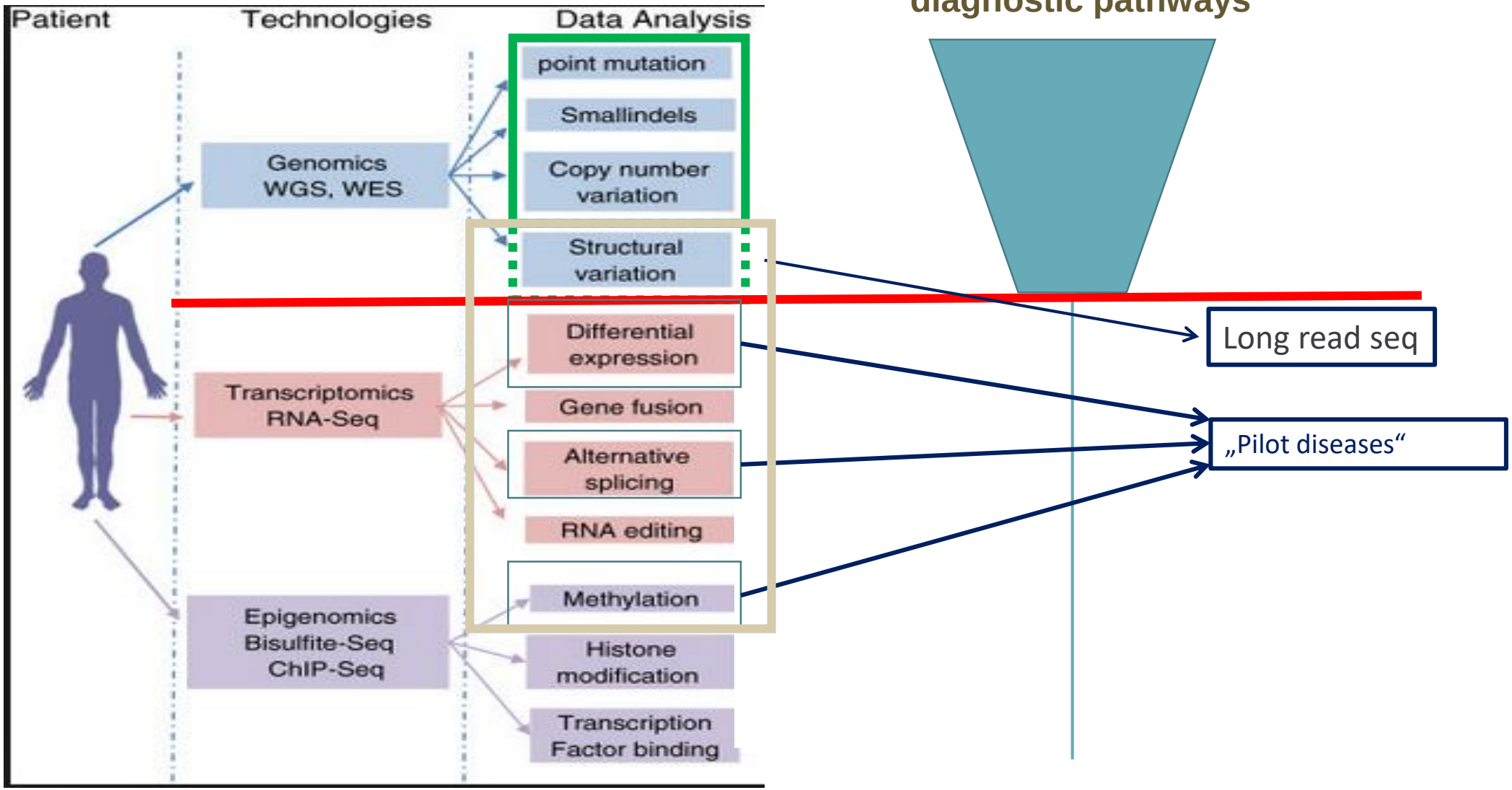
**MethBank 450K Array > 5000 controls**

**MethSeq control data:**

**Epigenie data base:**



# Technical hurdles in diagnostics

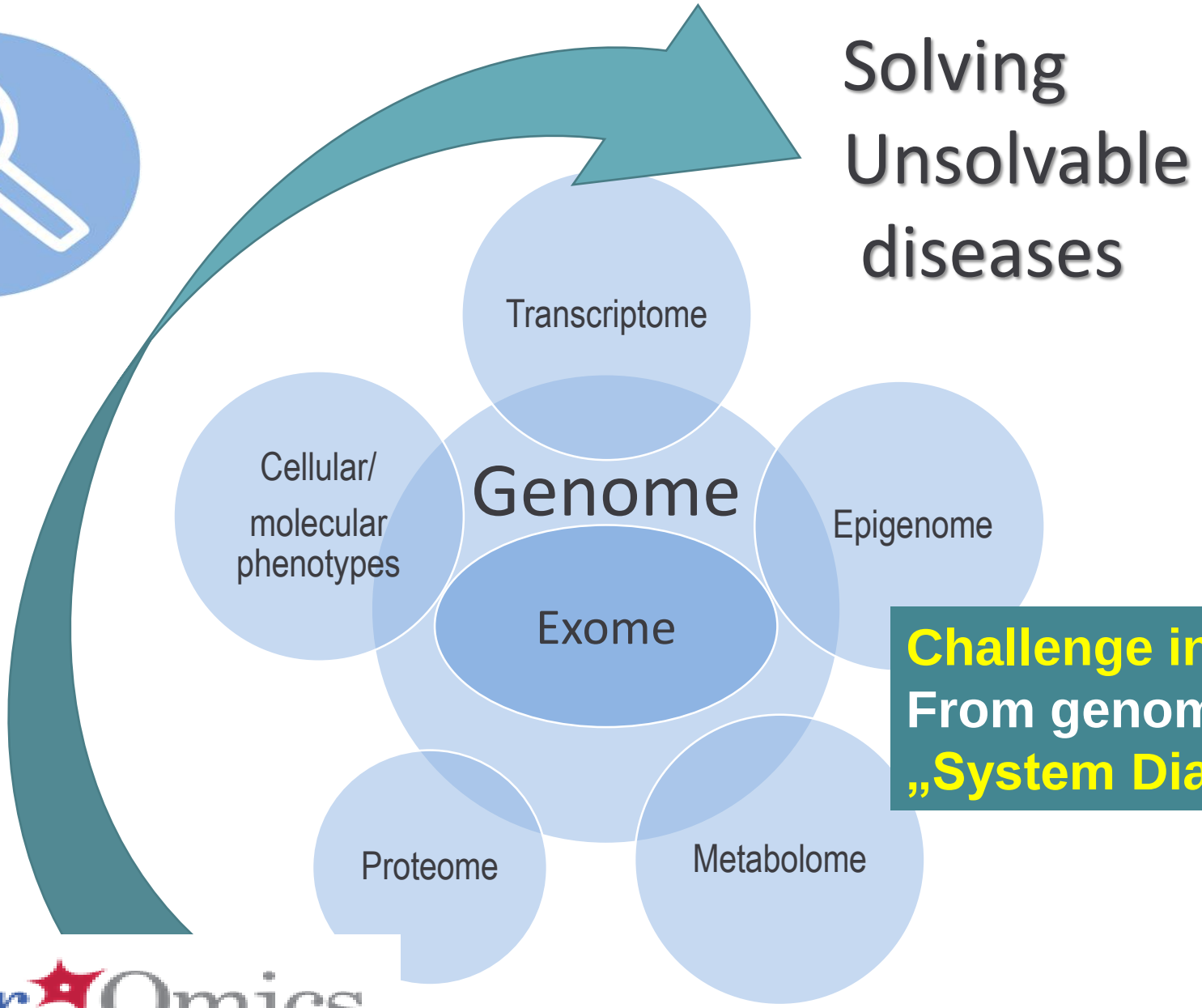



Modified from Shyr and Liu 2013

SolveRD

# Technological hurdles in diagnostics

Solving Unsolvable diseases




Solve-RD - Solving the unsolved Rare Diseases

European Commission | Horizon 2020 European Union funding for Research & Innovation

SolveRD

Solving the unsolved Rare Diseases

**Challenge in Diagnostic Transition:**
**From genome analysis towards „System Diagnostics“**

NeuroOmics

# „in numbers“

- Re-analysis of **19.000** exomes of unsolved cases
- **800 ultra-rare** RD patients presenting new phenotypes that will undergo WES/WGS
- **WGS for 2.000 cases** to achieve a more complete coding sequence
- **Novel omics approaches** (transcriptome, epigenome, proteome, metabolome, deep WES, deep molecular phenotyping) for more than **2.000 cases**
- **Long-read genomes for 500 cases** with smartly chosen phenotypes such as anticipated repeat expansion disorders (SBMA; DM1 and DM2, disorders with phenotype anticipation)

- **Multi-Omics approaches for 120 „unsolvable syndromes“**

**50 projects – 20.000€ each**

**Innovative and internationally interoperable small funding service**

# RDMM Europe

## Specialised Research Groups

**Validating gene discoveries** ⟺ **Topic & time-restricted collaborations**

**Modeling diseases** | **Therapeutic targets** | **Disease models**

Gene discovery group

Research group

**Proposal application**

**Catalyst Research grants**

Coordination and Subcontracting Office

**Clinical & Research Advisory Board: Grant evaluation**

ERN

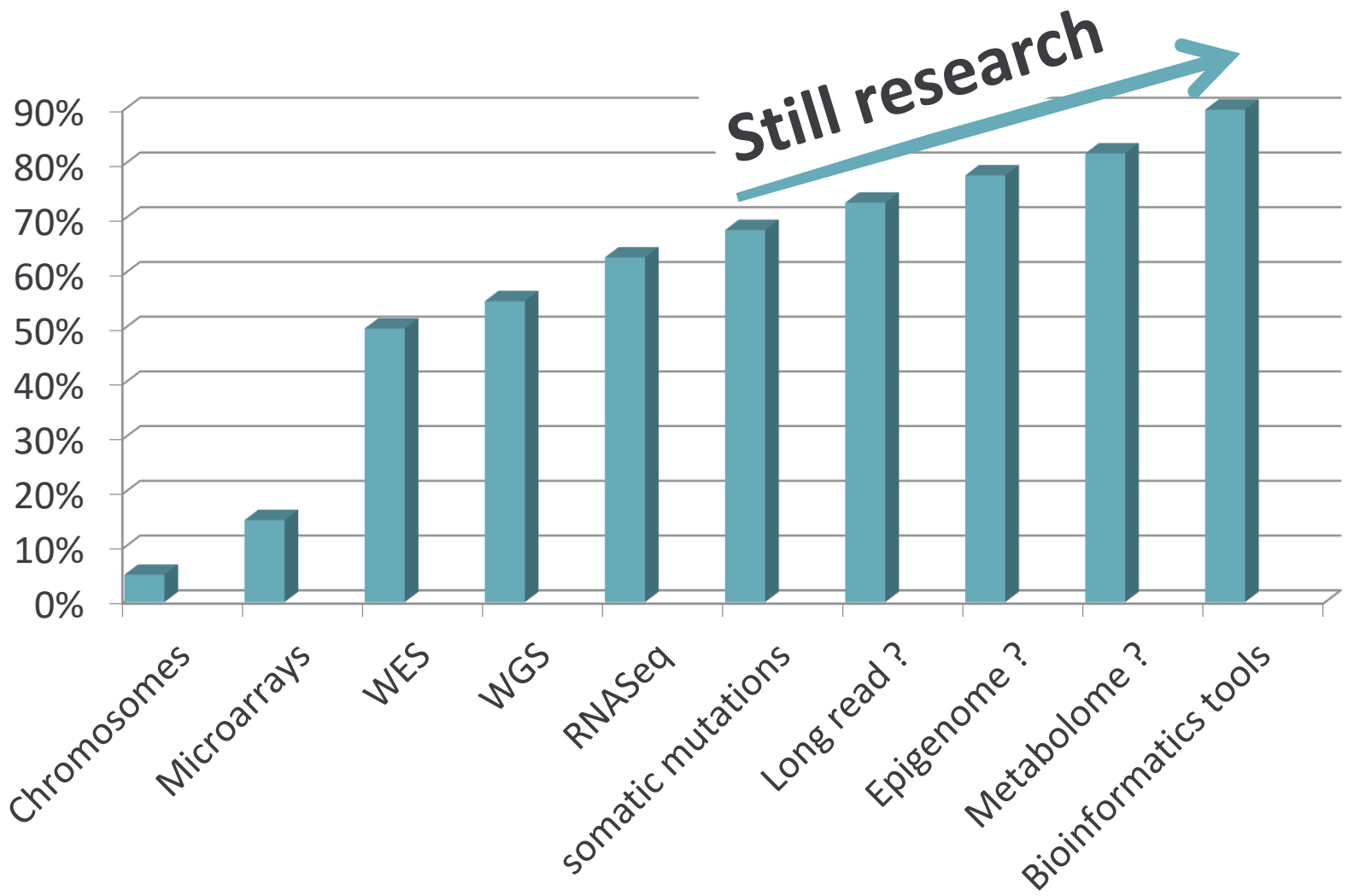
ERN

ERN….

REGISTER Technologies & Research groups

Canadian Network
(Australien Functional Network)
(Japanese Functional Network)

# Diagnostic yield of genetic technologies in rare diseases




**Cave: 80% of all RD have a clear genetic cause**

# Conclusions

1. Patient cohort of „Undiagnosed diseases“ should be subgrouped to define different diagnostic follow-up strategies
2. Development and availability of novel Omics technologies will contribute to deciphering unsolved and unsolvable diseases
3. Next steps in Omics diagnostics (system diagnostics) will be a challenge and require further strong support of research consortia
4. Implementing Machine learning and AI algorithms will greatly improve diagnostic work up
5. Whereas WES/WGS are more „global“ approaches of entire patient cohorts, next steps in terms of technology, methods, and analytics will be more patient individualized and require time
6. Special functional attention should be on solving VUS

SolveRD

ZSE TÜBINGEN
Zentrum für Seltene Erkrankungen


# Olaf Riess

olaf.riess@med.uni-tuebingen.de



I declare to receive an explorative grant from Illumina for implementation of WGS into clinical care.


Institut für Medizinische Genetik und Angewandte Genomik

NCCT
NGS Competence Center Tübingen

Deutschland Land der Ideen
Ausgewählter Ort 2011

www.ncct.life