Human Genetics Nijmegen

UMC St Radboud

"Genomic Resequencing"
27-29 May 2013, Erasmus MC Rotterdam, NL

# Diagnostics in sporadic disorders:
## *Intellectual Disability*

**Lisenka Vissers, PhD**

Dept. of Human Genetics
Radboud University Nijmegen Medical Centre
Nijmegen, The Netherlands

## Intellectual disability

- Affects ~1.5-2% of the Western population
- Clinically and genetically heterogeneous
- Genetic diagnosis remains elusive in 55-60%
- Associated with reproductive lethality

**Paradox in evolutionary theory**

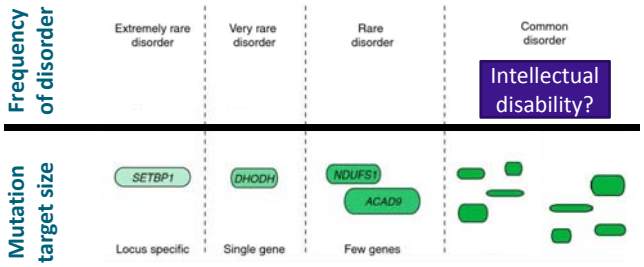
**Why are** (and remain) **reproductively lethal disorders, like intellectual disability, so frequent in our population?**

# The human per generation mutation rate, mutational target size and disease frequency

**… estimation of per generation mutation rate**

$7.6 \times 10^{-9}$ to $2.2 \times 10^{-8}$
= 50-100 genetic errors ***(or de novo mutations)*** per genome per generation

**… if *de novo* mutations are important for disease, then the mutational target size determines the frequency of disease**

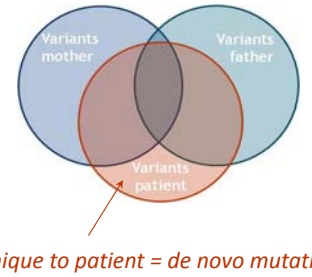



*Gilissen et al. Genome Biol 2011*



# A *de novo* paradigm for intellectual disability

- Each patient has a different mutation in a different gene
- Exome sequencing is the way to find these mutations




*Unique to patient = de novo mutation*

- Selected cohort of 10 patients
- Conventional tests normal
- Nine *de novo* mutations in nine genes
  - *Conclusive diagnosis: 3 patients*
    *(RAB39B, SYNGAP1, JARID1C)*
  - *Possible diagnosis: 4 patients*
    *(YY1, DYNC1H1, DEAF1, CIC)*
  - *Several de novo "background" mutations*

- May have implications for diagnostic strategy for patients with ID

*Vissers et al. Nat Genet 2010*

## Diagnostic evaluation of exome sequencing for ID

- Family-based exome sequencing of 100 patient with severe ID and their unaffected parents
  - Moderate to severe ID
  - No family history of ID
- Patients have reached the end stage of conventional diagnostic strategies
  - Targeted gene tests negative
  - Genomic array profile negative
- Evaluate variants identified in the clinical context of the patient assuming various inheritance models



## How representative is the cohort?

| | *% of total series (n=5,621)* | *% this study (n=100)* |
|---|---|---|
| **Level of ID** | | |
| **IQ <50** | **62** | **100** |
| *IQ <30* | *34* | *62* |
| *IQ 30-50* | *28* | *38* |
| **IQ 50-70** | **38** | **0** |
| **Clinical features** | | |
| Short stature | 23 | 24 |
| Microcephaly | 18 | 30 |
| Macrocephaly | 6 | 4 |
| Epilepsy | 21 | 52 |
| Abnormal brain imaging | 18 | 30 |
| Cardiac malformation | 11 | 2 |
| Urogenital abnormalities | 17 | 13 |

## Summary of clinical and genetic work-up of the 100 patients

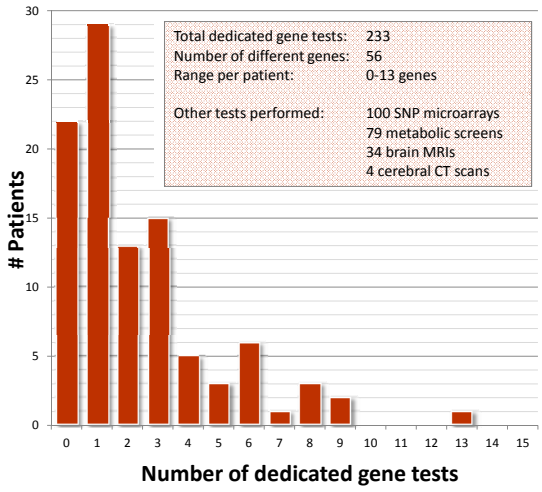
| Level of ID | Number of patients |
|---|---|
| IQ <30 | 62 |
| IQ 30-50 | 38 |
| IQ 50-70 | 0 |

| 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 |

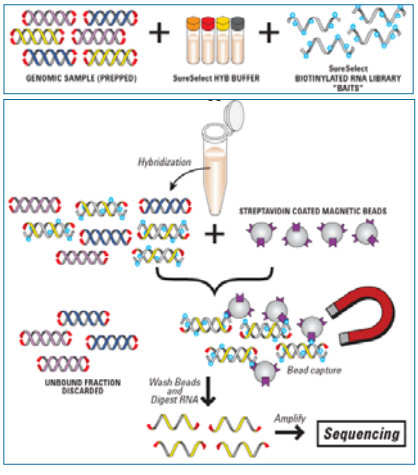
| Number of major congenital anomalies | |
|---|---|
| 0 | 62 |
| 1 | 31 |
| 2 | 7 |
| 3 | 0 |





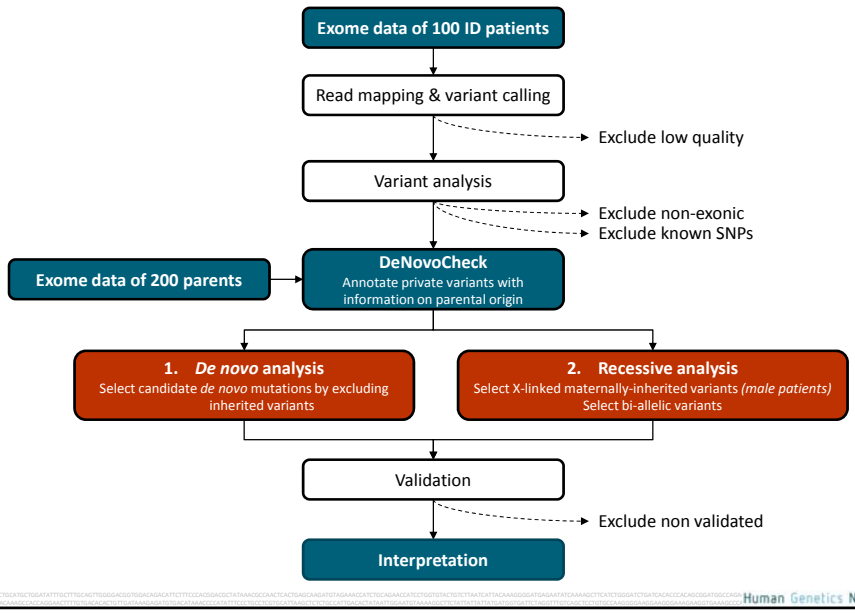

## Capturing genomic variation in all coding regions

*In liquid exome enrichment*
Agilent SureSelect v2, 50 Mb
enriching ~21,000 genes



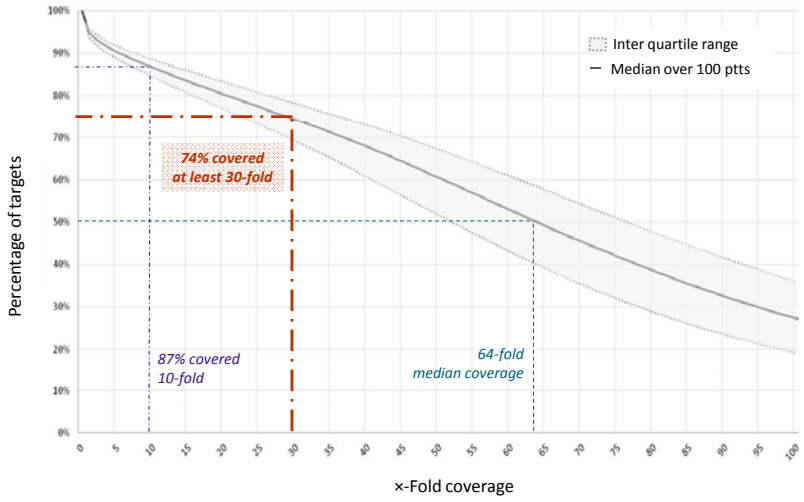

*Massive parallel sequencing*
Life Technologies SOLiD 4
~125-150M 50 bp reads/ sample

# Diagnostic trio exome sequencing workflow






# Coverage distribution




*Incomplete coverage but may be sufficient to detect majority of de novo mutations*

# Prioritization and distribution of *de novo* variants

| | Average over 100 patients | SD |
|---|---|---|
| High confidence variant calls | *24,324* | *2,403* |
| After exclusion of non-genic & intronic variants | *11,615* | *890* |
| After exclusion of known variants (dbSNPv132 + inhouse db) | *156* | *37* |
| After exclusion inherited variants (candidate *de novo* mutations) | *7 (range 2-20)* | *4* |

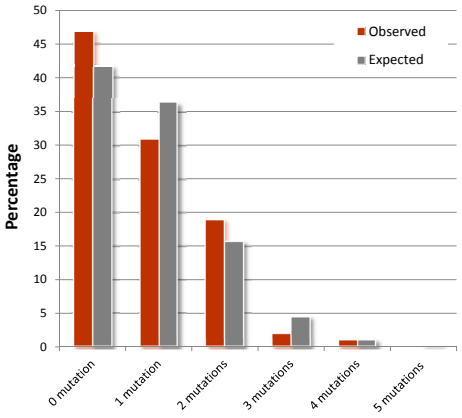
*Systematic validation by Sanger sequencing*

| | *# variants* |
|---|---|
| Not confirmed in patient | *530* |
| Confirmed in patient | *160* |
| ***De novo* occurrence** | ***79*** |

Human Genetics Nijmegen

# Number and type of *de novo* mutations detected

| | # mutations |
|---|---|
| ***Synonymous*** | ***16*** |
| ***Nonsynonymous*** | ***63*** |
| *Splice site* | 2 |
| *Nonsense* | 4 |
| *Missense* | 48 |
| *Frameshift* | 9 |
| ***total number of de novo mutations*** | ***79*** |




Patients with severe ID **do not have more *de novo* mutations**, but may have "more severe mutations"

Human Genetics Nijmegen

# Excess of predicted pathogenic mutations in ID

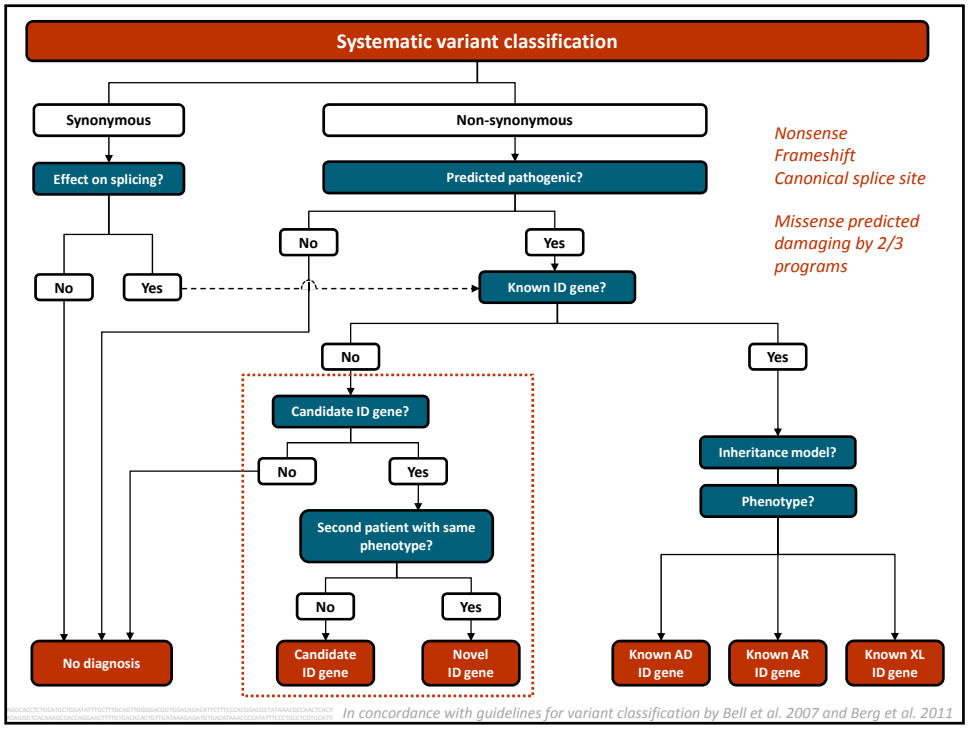
*Table S5: Single nucleotide mutation types and comparison to "random de novo mutations"*

| Type of *de novo* mutation (SNVs)* | *De novo* mutation in 100 ID patients (%) | Modeling of random *de novo* mutations (%) | p-value# |
|---|---|---|---|
| Missense | 69.7 | 66.1 | 0.65 |
| Nonsense | 6.1 | 3.3 | 0.45 |
| Synonymous | 24.2 | 30.6 | 0.42 |
| ***PolyPhen-2*[15] *missense classification*** | | | |
| Benign | 17.4 | 35.8 | 0.03 |
| Possibly damaging | 17.4 | 18.9 | 1.00 |
| Probably damaging | 65.2 | 45.2 | 0.03 |

**: Single nucleotide variants included only (indels/frameshift mutations excluded). ; #: p-values were determined using absolute counts and a two-tailed Fishers exact test. Modeling data was derived from* Neale et al. 2012[13].

➡ *Fraction of these missense mutations may have phenotypic consequences (relevant to ID)*

Human Genetics Nijmegen

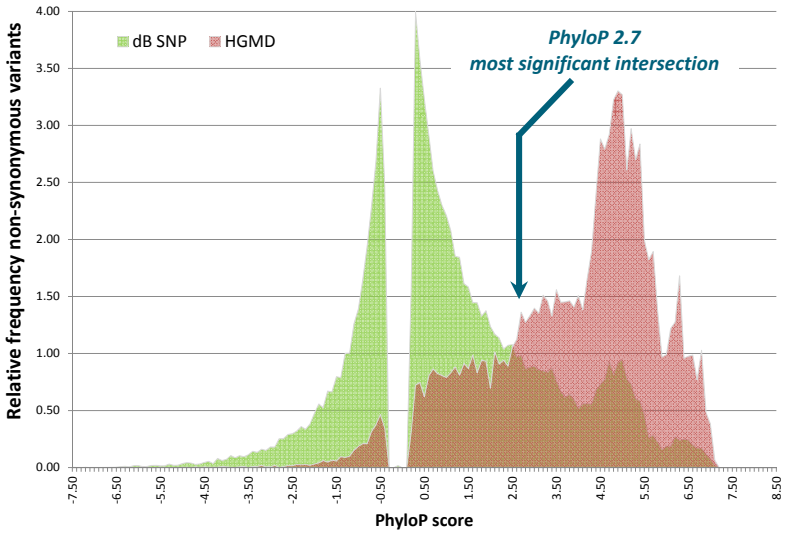


## Definition of a candidate ID gene

- Not previously associated with ID
- Assesment of seven parameters:
  1. Literature studies show function in essential development process for ID *(e.g. brain development)*
  2. Disruptive mutation *(e.g. nonsense, frameshift and canonical splice site)*
  3. Pathogenicity predication by 3 programs (*majority vote by PolyPhen2, Condel and SIFT)*
  4. Evolutionary conserved base *(PhyloP> 3.0)*
  5. Overrepresentation of GO terms associated with known ID genes
  6. Overrepresentaion of Mouse Phenotypes associated with known ID genes
  7. Brain expressed at 'significant' levels for 'significant' period of time *(Human brain transcriptome database, score >6 for at least 3 periods)*



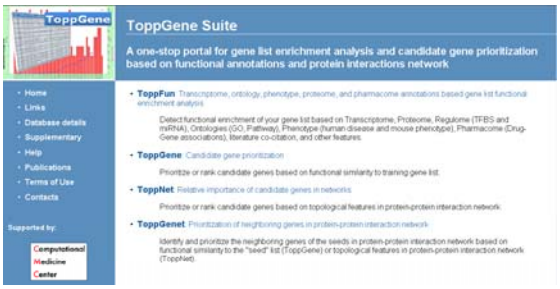
## Distinguishing benign *(dbSNP)* from causal *(HGMD)* variants




*Vissers et al. Nat Genet 2010*

## Enrichment for Mouse Phenotypes and GO terms

Upload list of 78 known ID genes in ToppGene Suite to analyze for enrichment of “specific terms”




**54 GO terms**
**27 Mouse Phenotypes**

Examples: “abnormal learning/memory/conditioning”
“abnormal synaptic transmission”

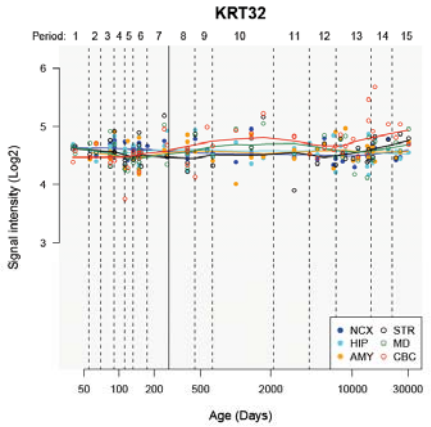
*Positive score for “Enrichment”:*
*Mutated gene also presents with (one of) the GO terms or Mouse Phenotypes*



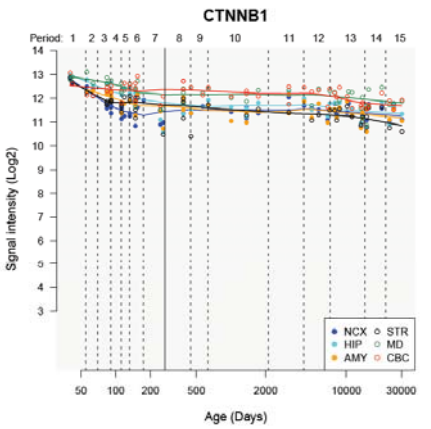
## Human brain transcriptome database

*Positive score for “Brain Expressed”:*
*>6 for at least 3 periods*

***Not expressed***




***Brain expressed***




*http://hbatlas.org/*

## Definition of a candidate ID gene

- Not previously associated with ID
- Assesment of seven parameters:
  1. Literature studies show function in essential development process for ID *(e.g. brain development)*

     **Required**
  2. Disruptive mutation *(e.g. nonsense, frameshift and canonical splice site)*
  3. Pathogenicity predication by 3 programs (*majority vote by PolyPhen2, Condel and SIFT)*

     **Either** disruptive **or** predicted pathogenic
  4. Evolutionary conserved base *(PhyloP> 3.0)*
  5. Overrepresentation of GO terms associated with known ID genes
  6. Overrepresentaion of Mouse Phenotypes associated with known ID genes
  7. Brain expressed at 'significant' levels for 'significant' period of time

     *(Human brain transcriptome database, score >6 for at least 3 periods)*

     **"Positive"** for **at least 2 of these parameters**

Human Genetics Nijmegen

## 35 genes affected by *de novo* mutations with a link to ID

| Type of mutation | Known ID genes | Candidate ID genes | | |
|---|---|---|---|---|
| Missense | GRIN2A[¥] | DYNC1H1 | GRIA1 | PROX2 |
| | GRIN2B | TANC2 | CAMKIIG | |
| | TCF4 | TNPO2 | KIF5C | |
| | TUSC3[§] | PPP2R5D | COL4A3BP | |
| | ARFGEF2[§] | ASH1L | RAPGEF1 | |
| | | LRP1 | PHACTR1 | |
| | | MIB1 | PSMA7 | |
| | | TRIO[¥] | EEF1A2 | |
| Nonsense | SCN2A | GATAD2B | PHIP | WAC |
| Frameshift | LRP2[*] | CTNNB1 | ZMYM6 | MTF1 |
| | PDHA1 | | | |
| | TUBA1A | | | |
| | SLC6A8 | | | |
| Splice site | SYNGAP1 | | MYT1L | |

*De novo* mutations in two independent patients

Recessive ID gene, no 2nd mutation identified

Recessive ID gene, 2nd paternally- inherited, predicted pathogenic mutation

Human Genetics Nijmegen

## Finding additional *de novo* mutations to 'prove causality'

- Through routine diagnostic exome sequencing

*Pilot study*

Table 2 Overview of all *de novo* variants identified by exome sequencing in ten individuals with unexplained mental retardation

| Gene | Trio | Sex[a] | NM number | cDNA level change | Protein level change | PhyloP score | Grantham score | Probability of being observed in dbSNP[b] | Probability of being observed in HGMD[b] | Gene function |
|---|---|---|---|---|---|---|---|---|---|---|
| ***De novo* mutations** | | | | | | | | | | |
| *DYNC1H1* | 1 | M | NM_001376 | c.11465A>C | p.His3822Pro | 5.5 | 77 | 0.20 | 0.80 | Retrograde axonal transporter; interacts with *PAFAH1B1* (mutation of which causes lissencephaly, a neurodevelopmental disorder) |

*Diagnostic series of 100 patients*

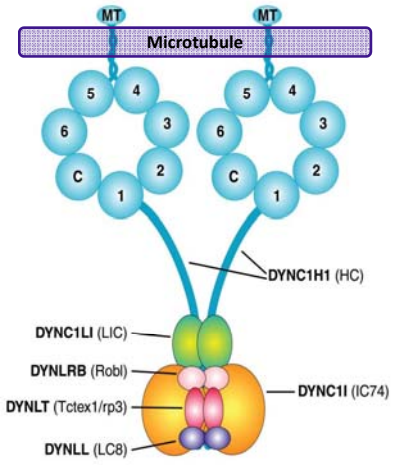
| Gen | *DYNC1H1* |
|---|---|
| **Ref Seq ID** | *NM_001376.4* |
| **cDNA** | c.4552G>A |
| **Protein level** | p.(Glu1518Lys) |
| **PhyloP score** | 6.03 |
| **Grantham score** | 56 |

- Two *de novo* missense mutations could be random chance
- **Function/Phenotype *DYNC1H1*?**
- **Do both patients look alike?**



## *DYNC1H1: dynein cytoplasmic 1 heavy chain 1*

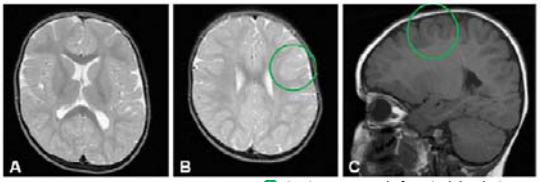
- 78 coding exons, 4646 amino acids
- Encodes the heavy chain of the cytoplasmic dynein heavy chain 1 motor complex
- Key role in retrograde axonal transport in neurons
- Interacts with *LIS1*
- 3 mouse models for DYNC1H1 show varies defects including
  - Gait abnormalities
  - Reduced muscle strength
  - Reduced reflexes
  - Neuronal migration defects




Adapted from: *Pfister et al. PLoS Genetics 2006*

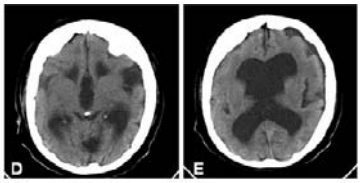
## Mutations in DYNC1H1 cause ID with neuronal migration defects

**Patient 1**




Region suspected of cortical dysplasia

**Patient 2**




| | Patient 1 (age 5 years) | Patient 2 (age 51 years) |
|---|---|---|
| ***ID/DD/LD*** | Severe ID | Severe ID |
| ***Neuronal migration defect*** | Mild | Severe |
| ***Delayed motor milestones*** | + | + |
| ***Epilepsy*** | - | + |
| ***Gait abnormalities*** | + | Cannot walk |
| ***Reduced reflexes*** | + | ? |
| ***Hypotonia and/or lordosis*** | + | Hypertonia |
| ***Clubfeet/ feet surgery*** | - | + |

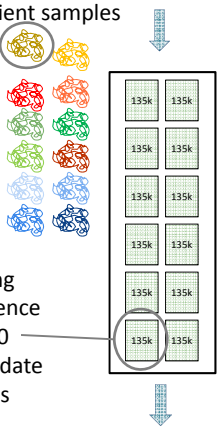
*Willemsen et al. JMG 2012*

Human Genetics Nijmegen

## Targeted re-sequencing of candidate ID genes

Equimolar pool of 10 patient samples

Hybridize individual pools onto 12-plex array

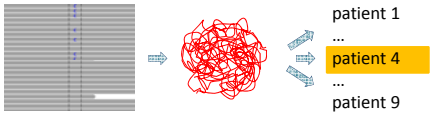



Coding sequence of ~20 candidate genes

Elute as a 'single sample' representing 120 novel patients

Sequence on SOLID 5500 XL

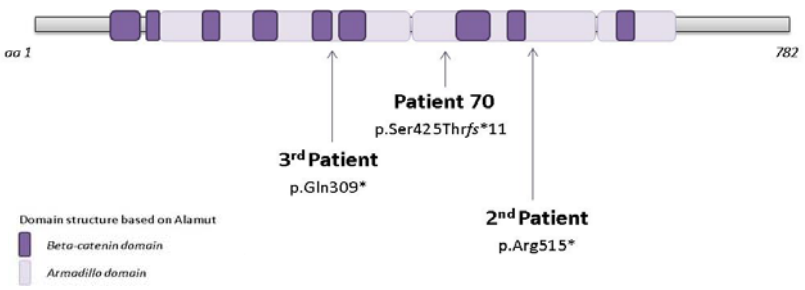
Find mutation, determine the sample pool, and repeat amplicon in individual patients of the pool




***Cohort size: 765 patients with ID***

Human Genetics Nijmegen

# Additional *CTNNB1* loss of function mutations




**Phenotypic comparison:**

- severe developmental delay with delayed motor milestones
-microcephaly
- spasticity due to which the ability to walk was severely impaired



# Defining novel ID genes

| Type of mutation | Known ID Genes | Novel ID Genes | Candidate ID Genes | |
|---|---|---|---|---|
| **Missense** | GRIN2A (2x) | **DYNC1H1** | | PROX2 |
| | GRIN2B | | CAMKIIG | TANC2 |
| | TCF4 | | **KIF5C** | TNPO2 |
| | | | COL4A3BP | PPP2R5D |
| | | | RAPGEF1 | **ASH1L** |
| | | | PHACTR1 | LRP1 |
| | | | PSMA7 | MIB1 |
| | | | EEF1A2 | TRIO (2x) |
| | | | GRIA1 | |
| **Nonsense** | SCN2A | **GATAD2B** | | WAC |
| | | | PHIP | |
| **Frameshift** | LRP2 | **CTNNB1** | | ZMYM6 |
| | PDHA1 | | MTF1 | |
| | TUBA1A | | | |
| | SLC6A8 | | | |
| **Splice site** | SYNGAP1 | | MYT1L | |

+ 3 Maternally inherited X-linked mutations
+ 0 Bi-allic inherited autosomal recessive mutations

* = recessive ID gene

## Diagnostic yield in 100 ID patients

| Positive diagnosis | No. of patients |
|---|---|
| **All mutations** | **16** |
| ***De novo* mutations** | **13** |
| Autosomal dominant | 10 |
| X-linked | 2 |
| Autosomal recessive | 1 |
| **Inherited mutations** | **3** |
| X-linked | 3 |
| Autosomal recessive | 0 |
| Candidates | 19 |



## Conclusions

### *De novo* mutations are a common cause of severe ID

- Exonic point mutations and indels explain 16-35%
- Most mutations of *de novo* origin

### From research to diagnostics

- Exome sequencing can be used as a generic diagnostic test and may include CNV analysis
- New ID genes/syndromes can be defined
- Practical workflow implemented with few incidental findings

*De Ligt, Willemsen, van Bon, Kleefstra et al. NEJM 2012*

# Acknowledgements


***Selection of patients***
**Marjolein Willemsen**
**Bregje van Bon**
**Tjitske Kleefstra**
David Koolen
Anneke Vulto-van Silfthout
Bert de Vries

***Exome sequencing & Validation***
Thessa Kroes
Petra de Vries
Gaby van de Ven-Schrobbers
Michael Kwint
Irene Janssen
Marloes Steehouwer
Alex Hoischen

***Bioinformaticians***
**Joep de Ligt**
Jayne Hehir-Kwa
Christian Gilissen
Marisol del Rosario
Nienke Wieskamp
Rick de Reuver

***Diagnostic interpretation***
**Helger Yntema**
Dorien Lugtenberg
Rolph Pfundt
Marcel Nelen
Hans Scheffer

***Supervision***
**Joris Veltman**
**Han Brunner**