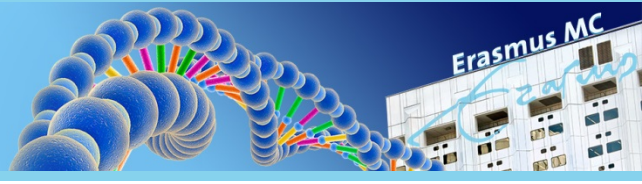


Erasmus MC
University Medical Center Rotterdam

Whole Exome Sequencing as a diagnostic test

Quality standards and reporting VEP results

International Post-graduate Course

Marjon van Slegtenhorst
laboratoriumspecialist klinische genetica
m.vanslegtenhorst@erasmusmc.nl

This course

- Gene variant database sharing
- HPO in genomic medicine
- Variant classification
- Protein prediction tools
- Viewing the diagnostic data
- [Phenotypic sharing (Face to Gene)]

How we use all these different tools in WES diagnostics:

Successes
Limitations and pitfalls

Exome sequencing

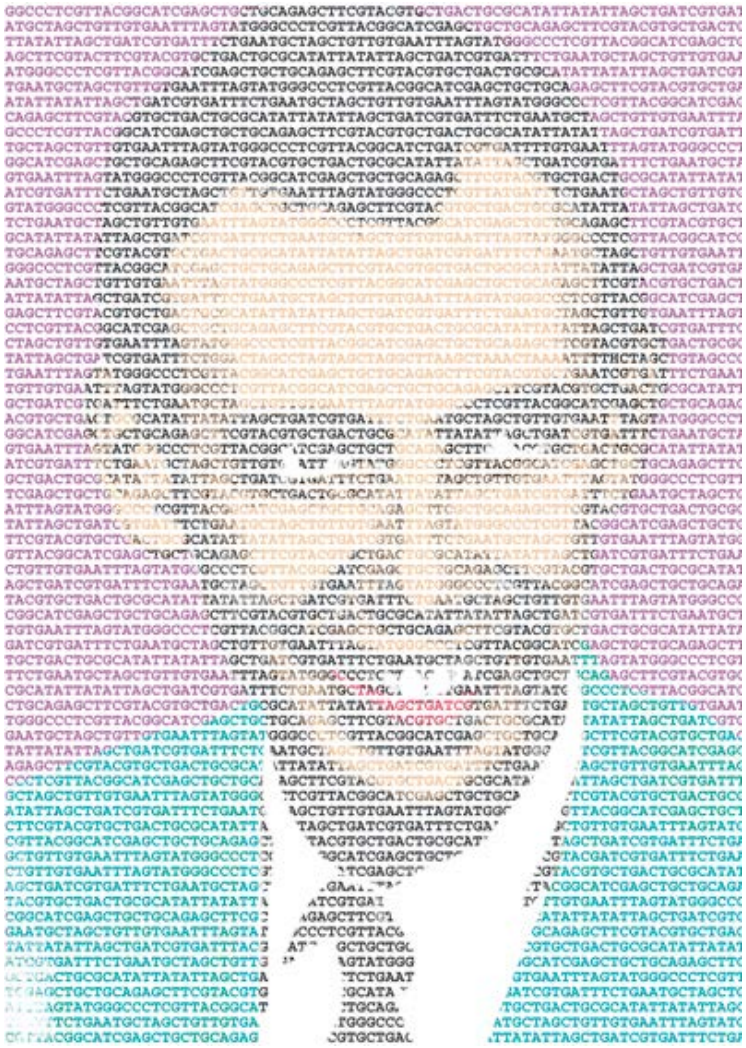
Human Genome Project

↓

Start in 1990

↓

10 years later; first exome sequenced



NGS in DNA diagnostics

Targeted Sequencing
-enrichment group of genes
-enrichment for all coding genes (WES)

(NIPT)

(Whole Genome Sequencing)

Why WES in DNA diagnostics?

One test fits all

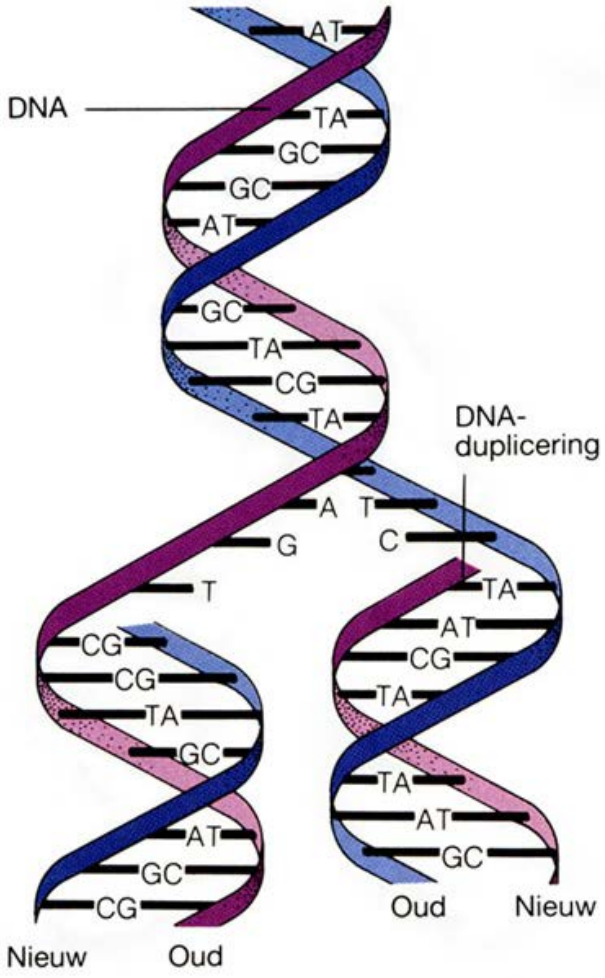
- Indications like ID are heterogeneous
- Diagnostic yield increases
- Flexible
- FULL WES analysis possible

Many children are in diagnotic routing, still without diagnosis

⬇

More information for parents about recurrent risk and prognosis
Organ involvement
Contact with other parents/ patients

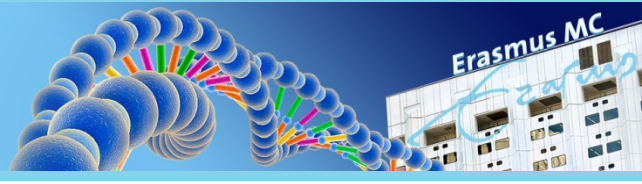
DNA as starting material



EDTA blood
Fibroblasts
Amniotic fluid
Chorionic villi
Sputum
Tumor

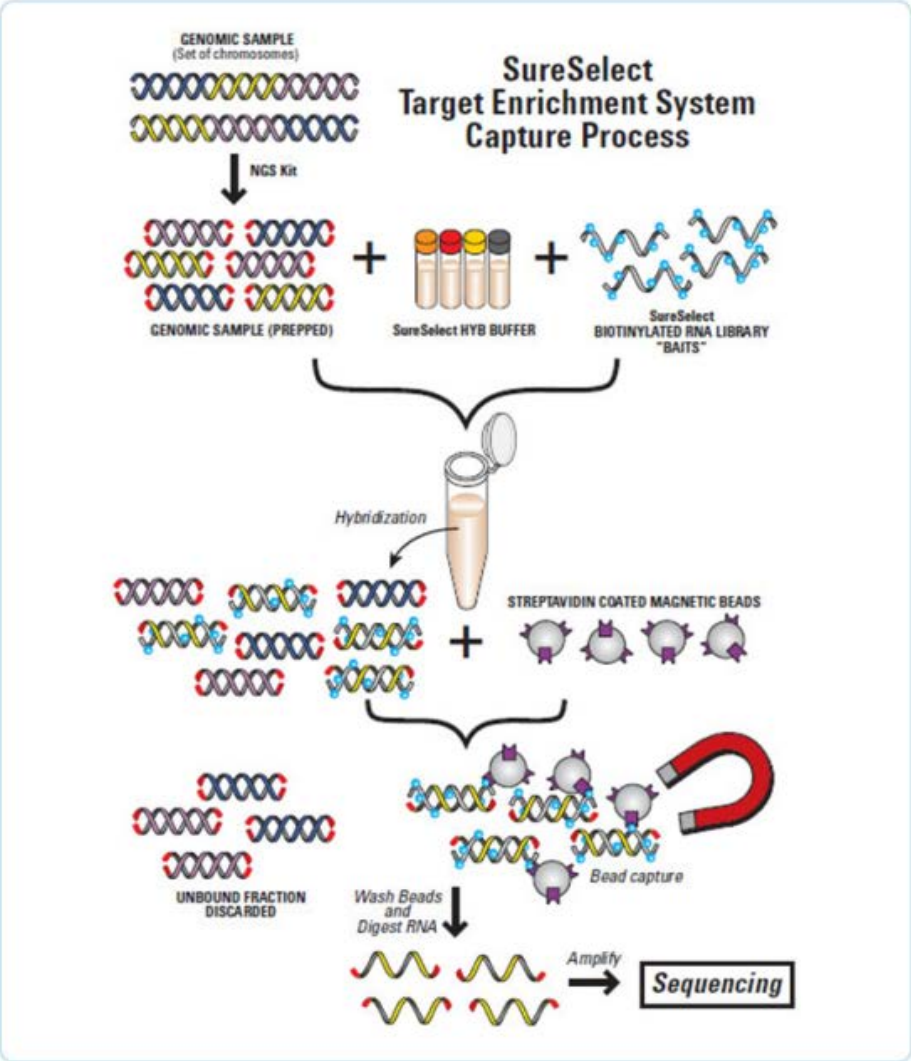
Important!
Quantity DNA
Quality DNA

Technical Information



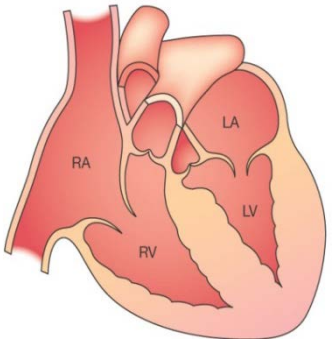
- Quality check; gender match for 96 well plate
- Agilent SureSelect **C**linical **R**esearch **E**xome V2 capture (paired-end sequenced /Illumina platform) (**outsourced**).
- Duplicate reads are excluded
- Mapping to the genome using the BWA-MEM algorithm (reference:http://bio-bwa.sourceforge.net/).
- Variant detection: Genome Analysis Toolkit HaplotypeCaller (reference:http://www.broadinstitute.org/gatk/).
- Filtering and annotation with Cartagenia software
- Classification with Alamut Visual
- Sample ID check standard for single patients (process not automated)

Sample prep and run on Novaseq



Single exome versus Trio exome

Exome for single patient:
HCM: Left ventricular hypertrophy; diagnostic yield 45%



▶ Indicatie

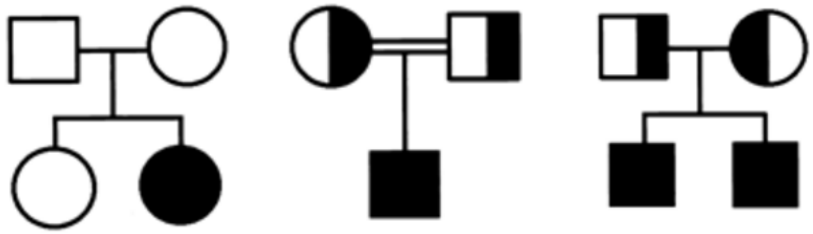
- ☐ Aangeboren afwijking (alleen trio analyse) (0161)
- ☐ Aneurysma (2454)
- ☐ Anusatresie (0057)
- ☐ Autisme (1486)
- ☐ Bewegingsstoornis (incl. voormalige panels ataxie, Parkinson en paroxismale dyskinesie) (5222)
- ☐ Ceroïdlipofuscinose (CLN) (3512)
- ☐ Ciliopathie, incl. Bardet Biedl syndroom (5599)
- ☐ Disorders of Sex Development (DSD) (1439)

  resultaten karyotypering:.....
- ☐ Doofheid (0800)
- ☐ Epilepsie (2011)
- ☐ Hernia diafragmatica (0271)
- ☐ Neurodegeneratie (incl. voormalige panels dementie/FTD/ALS, NBIA en Parkinson (1656)
- ☐ Oesophagusatrasie (0905)
- ☐ Oncopakket voor kinderen (4722)
- ☐ Oncopakket voor volwassenen (5407)
- ☐ Verstandelijke beperking (alleen trio analyse) (0311)
- ☐ Visusstoornis (2089)
- ☐ Ouder voor trio analyse

WES Trio diagnostics

Intellectual disability and congenital abnormalies

- *de novo* variants for intellectual disability (ID)
- consanguineous families for AR (homozygous variants)
- recessive diseases (compound heterozygous variants)



de novo (trio) homozygous Compound heterozygous

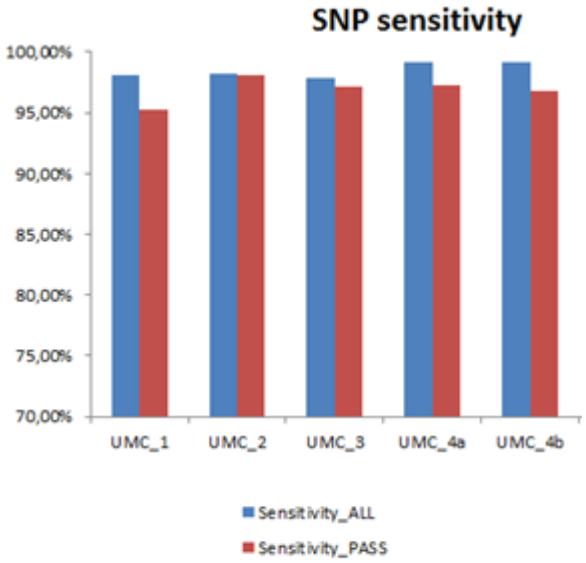
Include parents for filtering

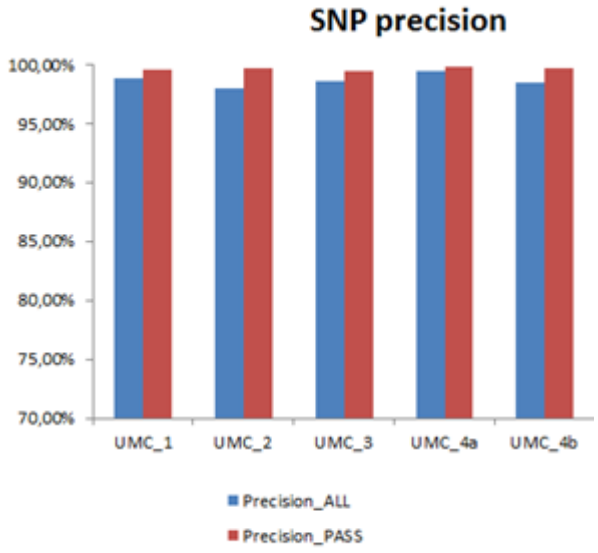
gene panels (>1000 genes) and for “FULL exome”

Quality issues; GIAB

How do you determine how much you sequence?

More is better, but also more is more expensive





Martin Elferink & Koen van Gassen, UMCU
22-02-2017

vcf files (list with all variants in sample)

#CHROM	POS	ID	REF	ALT
chr1	762273	rs3115849	G	A
chr1	792263	rs1044922	A	G
chr1	792480	rs2905036	C	T
chr1	808922	rs6594027	G	A
chr1	866319	rs9988021	G	A
chr1	877715	rs6605066	C	G
chr1	879676	rs6605067	G	A
chr1	879687	rs2839	T	C
chr1	880238	rs3748592	A	G
chr1	882033	rs2272756	G	A
chr1	883625	rs4970378	A	G

In total ~ >100 000 variants

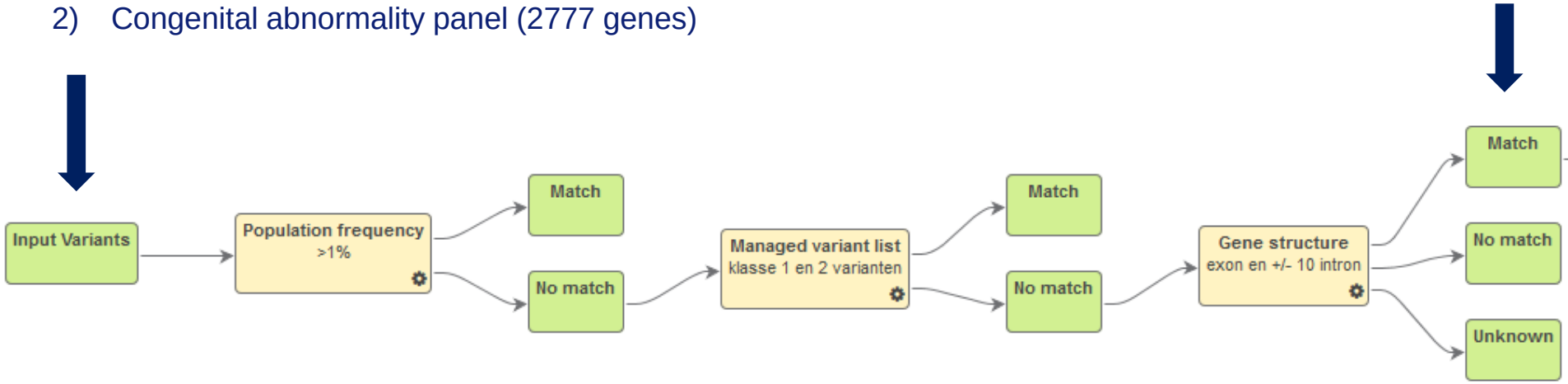
How do you find the variants of interest?

1. Filtering (Software versus Pipeline homemade)
2. Classification
3. Interpretation
4. Reporting

Cartagenia Trio Filtering

Inheritance filter:
1) De novo variants
2) Compound heterozygous variants
3) Homozygous/hemizygous variants
4) Imprinted genes

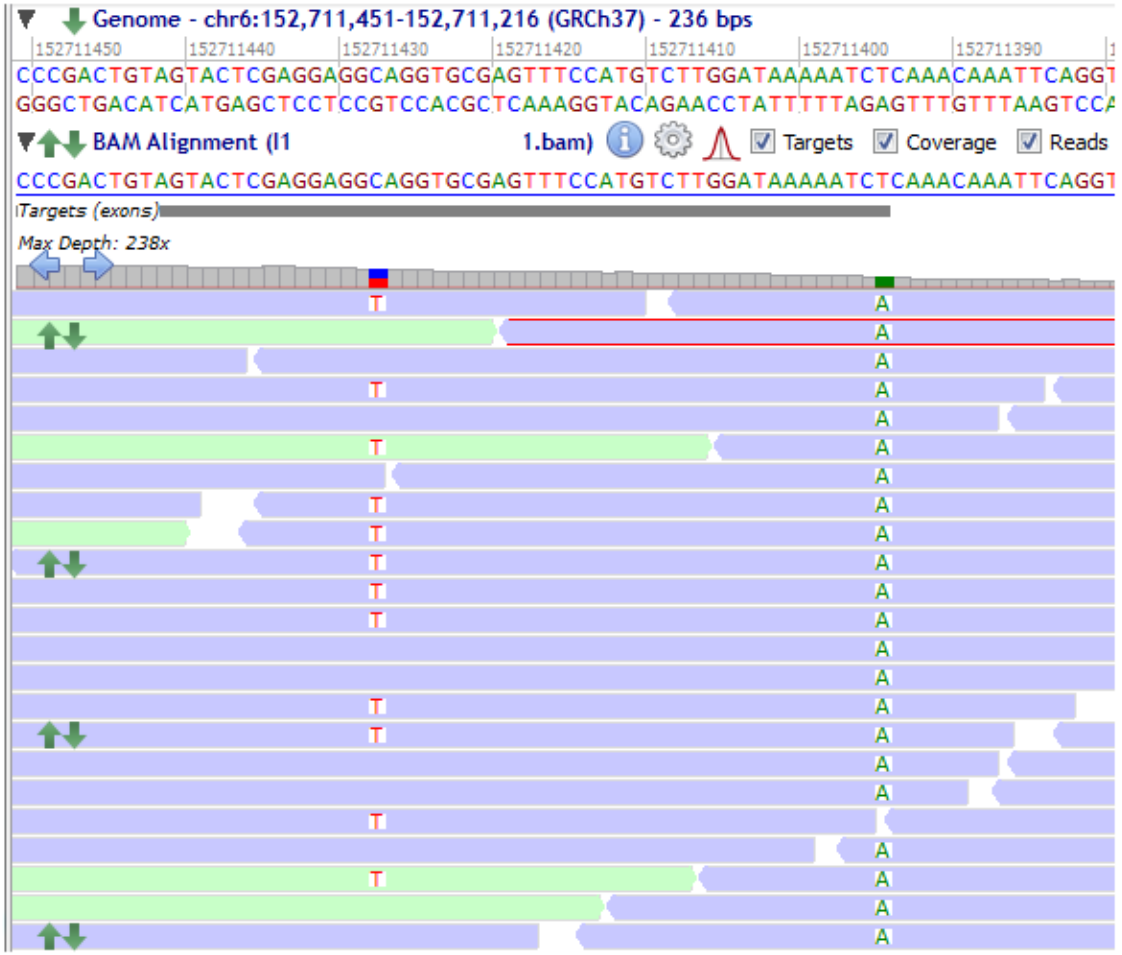
Variants loaded with genepanel filter
1) Intellectual disability panel (1162 genes)
2) Congenital abnormality panel (2777 genes)



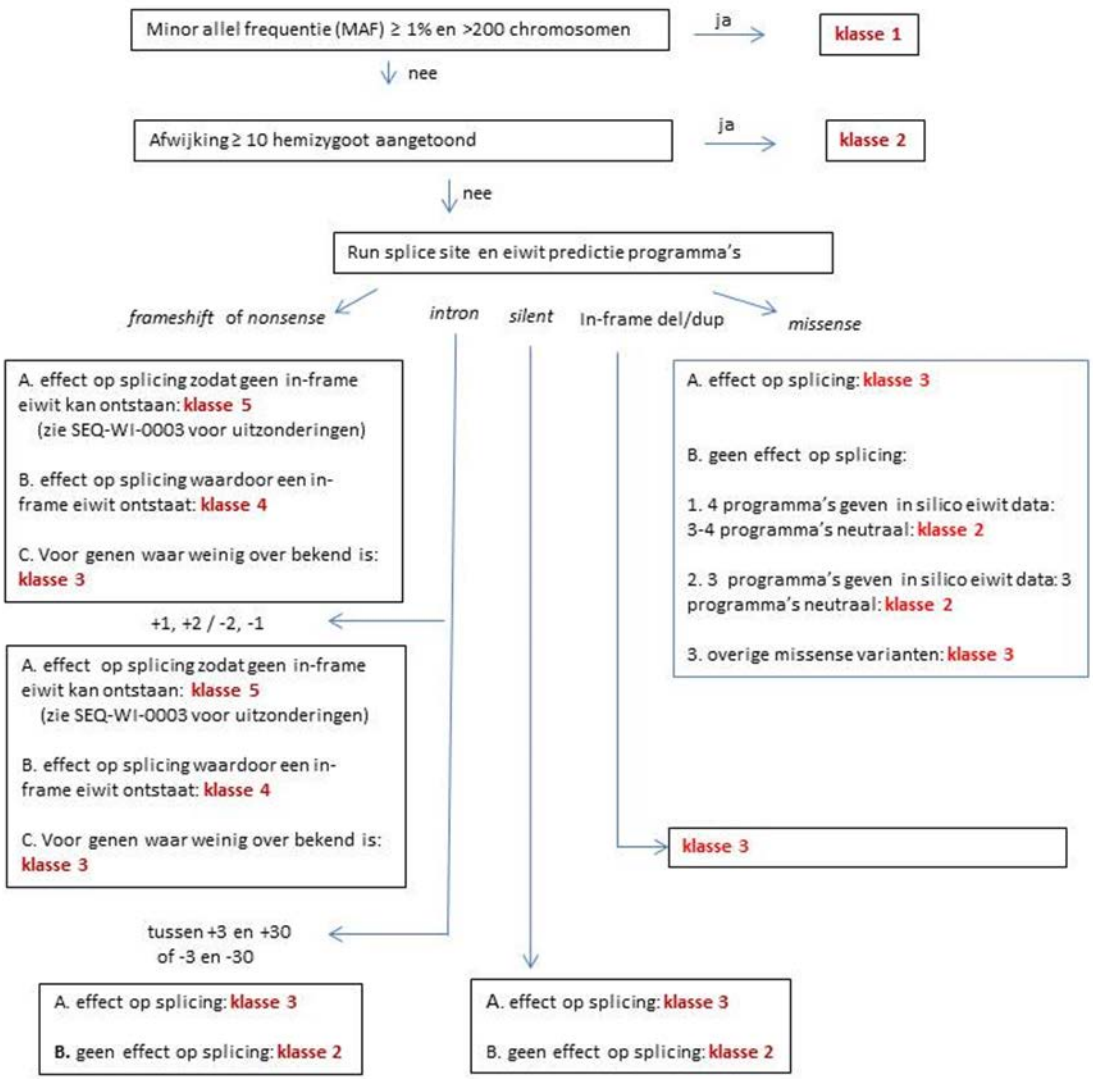
Variant detection/ coverage

Important:

- -quality of reads
- -quality of base
- -coverage of variant



Classification scheme Sanger genes



Follow-up variants

After filtering, variants are classified

Classification of sequence variants (known genes)

- Class 1; benign variants (frequent in control populations)
- Class 2; silent variants and intronic variants with no effect splicing
- Class 3; “rest group”
- Class 4; Likely pathogenic, HGMD link (inspecion!); more often found in patients
- Class 5; pathogenic (frameshift, nonsense, splicesite (+/- 1 and 2)), functional data

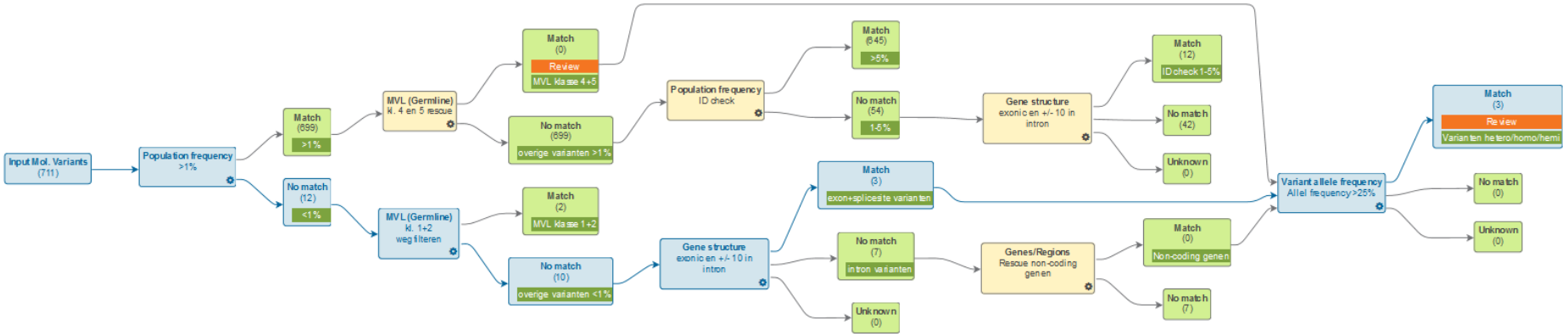
What to report?
What to confirm with Sanger sequencing?

Wallis et al (ACGS & VKGL; 2013) Practice guidelines for the evaluation of pathogenicity and the reporting of sequence variants in Clinical Molecular Genetics

ACMG guidelines for variant classification in near future as the standard?

Erasmus MC
University Medical Center Rotterdam

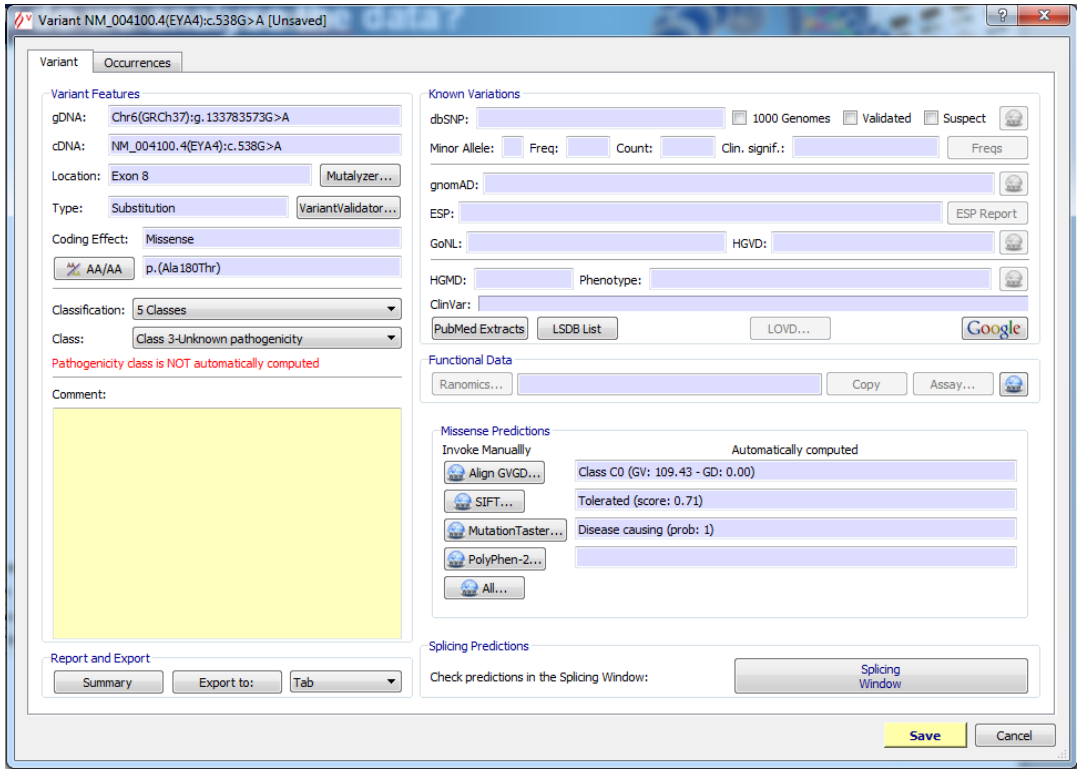
Cardiomyopathy single patient reporting

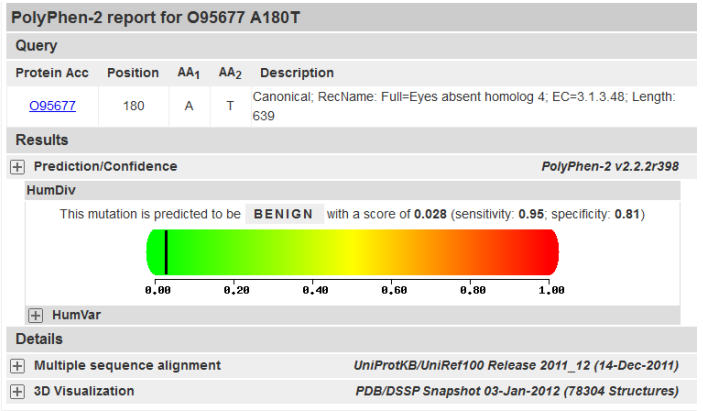


Gene	Position	Ref	Patient	Read Depth	Type	Transcript	cDNA	Location	Exon	Effect	Protein
TTN	2:179,454,135	G	G A	40	snp	NM_001267550.1	c.62317C>T	exonic	304	synonymous	p.Leu20773=
EYA4	6:133,783,573	G	G A	79	snp	NM_004100.4	c.538G>A	exonic	8	nonsynonymous	p.Ala180Thr
MYBPC3	11:47,359,280	.	. C	160	insertion	NM_000256.3	c.2373dupG	exonic	24	frameshift	p.Trp792Valfs*41

Erasmus MC
University Medical Center Rotterdam

How do we analyse the data?





Gene	Position	Classification	External Databases	MVL Eras	MVL Gene	MVL niet	MVL verk	MVL Other	MVL VKGL	MVL Simil.
TTN	2:179,454,135	Likely benign								
EYA4	6:133,783,573	VOUS								
MYBPC3	11:47,359,280	Pathogenic								

Erasmus MC
University Medical Center Rotterdam

Variant classification

It's not good, should be deleted

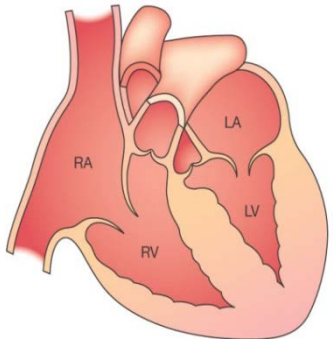
HGMD Professional

Gene	Position	Classification	External Databases	MVL Eras	MVL Gene	MVL niet	MVL verk	MVL Other	MVL VKGL	MVL Simil.
TTN	2:179,454,135	Likely benign								
EYA4	6:133,783,573	VOUS								
MYBPC3	11:47,359,280	Pathogenic								

HGMD accession	Reported disease/phenotype	Variant class	Gene symbol	Insertion
CI983160	Cardiomyopathy, hypertrophic	DM	MYBPC3	CACAGTA^CAGgTGGGAGCCGC

Literature citation	Citation type	Support	
1. Niimura (1998) *N Engl J Med* **338:** 1248 PubMed: 9562578 Mutations in the gene for cardiac myosin-binding protein C and late-onset familial hypertrophic cardiomyopathy.	Primary literature report	✓	aka c.2373_2374insG/c.2373dupG.
2. Marston (2009) *Circ Res* **105:** 219 PubMed: 19574547 Evidence from human myectomy samples that MYBPC3 mutations cause hypertrophic cardiomyopathy through haploinsufficiency.	Additional literature report		None
3. van Dijk (2009) *Circulation* **119:** 1473 PubMed: 19273718 Cardiac Myosin-Binding Protein C Mutations and Hypertrophic Cardiomyopathy. Haploinsufficiency, Deranged Phosphorylation, and Cardiomyocyte Dysfunction.	Functional characterisation		None
4. Christiaans (2010) *Neth Heart J* **18:** 248 PubMed: 20505798 Founder mutations in hypertrophic cardiomyopathy patients in the Netherlands.	Additional literature report	✓	Founder mutation in Dutch.
5. Yiu (2012) *PLoS One* **7:** e36115 PubMed: 22574137 Myocardial structural alteration and systolic dysfunction in preclinical hypertrophic cardiomyopathy mutation carriers.	Additional literature report		None
6. Birket (2015) *Cell Rep* **13:** 733 PubMed: 26489474 Contractile Defect Caused by Mutation in MYBPC3 Revealed under Conditions Optimized for Human PSC-Cardiomyocyte Function.	Functional characterisation	✓	None
7. Murphy (2016) *J Cardiovasc Transl Res* **9:** 153 PubMed: 26914223 Evaluation of the Mayo Clinic Phenotype-Based Genotype Predictor Score in Patients with Clinically Diagnosed Hypertrophic Cardiomyopathy.	Additional literature report	✓	Descr. in Supplemental Table 2 (online).
8. Wijnker (2016) *J Mol Cell Cardiol* **97:** 82 PubMed: 27108529 Comparison of the effects of a truncating and a missense MYBPC3 mutation on contractile parameters of engineered heart tissue.	Functional characterisation	✓	None
9. Baudhuin (2017) *Circ Cardiovasc Genet* **10:** e001844 PubMed: 29237689 Technical Advances for the Clinical Genomic Evaluation of Sudden Cardiac Death: Verification of Next-Generation Sequencing Panels for Hereditary Cardiovascular Conditions Using Formalin-Fixed Paraffin-Embedded Tissues and Dried Blood Spots.	Additional case report	✓	None
10. Burns (2017) *Circ Cardiovasc Genet* **10:** e001666 PubMed: 28790153 Multiple Gene Variants in Hypertrophic Cardiomyopathy in the Era of Next-Generation Sequencing.	Additional literature report	✓	None
11. Miller (2017) *Circ Cardiovasc Genet* **10:** e001735 PubMed: 29212898 Genetic Testing in Pediatric Left Ventricular Noncompaction.	Additional phenotype	✓	Noncompaction, left ventricular
12. van Velzen (2017) *Circ Cardiovasc Genet* **10:** e001660 PubMed: 28794111 Clinical Characteristics and Long-Term Outcome of Hypertrophic Cardiomyopathy in Individuals With a MYBPC3 (Myosin-Binding Protein C) Founder Mutation.	Additional literature report	✓	None
13. Viswanathan (2017) *PLoS One* **12:** e0187948 PubMed: 29121657 Hypertrophic cardiomyopathy clinical phenotype is independent of gene mutation and mutation dosage.	Additional literature report	✓	Potentially recessive. See Table S4 and S5.

What to confirm/ report in this case?



Single patient;

Class 4 or 5 that confirms the diagnosis; MYBPC3 pathogenic variant
All other class 3 variants! EYA4 variant (involvement in DCM)

Conclusion in letter: Diagnosis HCM has been confirmed.
Further testing for the pathogenic variant in the familiy of tested person is possible

What to confirm with Sanger?

You do not confirm all the variants from the report!

Single patient (more general);

1) Class 4 or 5: that confirms the diagnosis

→ in most cases there will be presymptomatic testing of family members

2) Class 3: good candidate

→ when you expect testing of affected family members or parents in case of suspicion the novo

3) Class 1 or 2: Fragment that is working in the lab

→ ID check of a known variant (filterstep 1-5% in controls)

You often have to design the Sanger test

What to confirm with Sanger?

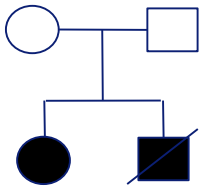
1) Class 4 and 5 variants which confirm the diagnosis
2) Class 3 variants:
    - Only the novo variants when the coverage in the parents is <30 reads
    - Autosomal recessive disease you have internal controles (parents)

What to look out for:

False positives!

AF between 25% and 35%: often T-stretches, CG rich regions, deletions/duplications

Germline mosaicims in parents (always check BAM files in the novo calls!)



Gene	Position	Ref	Patient	Read Depth	Type	Transcript	cDNA	Location	Exon	Effect	Protein
NAA10	X:153,197,863	G	G A	76	snp	NM_003491.3	c.247C>T	exonic	5	nonsynonymous	p.Arg83Cys

Trio reporting

The novo variants (class 3 or higher for AR and AD)
Compound heterozygous variants (class 3 or higher)
Homozygous variants (class 3 or higher)
X-linked variants (class 3 or higher)

In all cases a is clinical geneticist is involved for informed consent; secundary findings

The phenotype on application form is often not complete.

Discussions in Multidisciplinary Expertise Group (MEG) to decide what can fit

Helpful tools with classification with examples:

- Gene variant database sharing in Netherlands
- Genematcher
- HPO in genomic medicine

Sharing information VKGL

VKGL: Dutch database of the diagnositic laboratories

Participant	Classification	
vkgl (Consensus list)	*Variant not present*	
vkgl (VKGL-AMC)	VOUS	
vkgl (VKGL-ERASMUS)	*Variant not present*	
vkgl (VKGL-LUMC)	*Variant not present*	
vkgl (VKGL-NKI)	*Variant not present*	
vkgl (VKGL-RADBOUD)	*Variant not present*	
vkgl (VKGL-UMCG)	*Variant not present*	
vkgl (VKGL-UMCU)	*Variant not present*	
vkgl (VKGL-VUMC)	*Variant not present*	

DPYSL5; 2 children share the same de novo variant

Severe brain abnormalities in these two children: are we convinced this is it?

Important for solving exomes:
Datasharing (worldwide); databases and genematcher
Multidisciplinairy discussions
Software tools; face to gene

Erasmus MC
University Medical Center Rotterdam

Multidisciplinary expertise group (MEG)

Child with short stature, abnormal teeth, epilepsy, no speech, macrocephaly, dysmorphic face, cataract at age 5, abnormal toes.

Trio ID panel : negative

⬇

FULL exome

Power of Genematcher

INTS1 homozygous nonsense variant (Integrator Complex Subunit 1)

c.5351C>A, p.Ser1784* (NM_001080453)

Follow up after variant in gene with no linked fenotype in OMIM

Discuss in MEG/ submit to Genematcher:
very striking features is helpful
Fenotype recognized by other clinical geneticist and hit genematcher
Now 3 extra patients found with same nonsense variant worldwide

3 patients from the Netherlands: information meeting organized for parents

Oegema et al, PLoS Genet. 2017 May 25;13(5):e1006809

Google your genes/ variants!

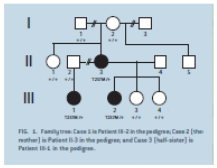
Program #: 3367

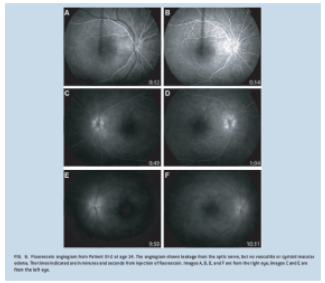
Whole Exome Sequencing (WES) Identifies a Mutation in ALPK1 Responsible for a Novel, Autosomal Dominant Disorder of Vision Loss, Splenomegaly, and Pancytopenia

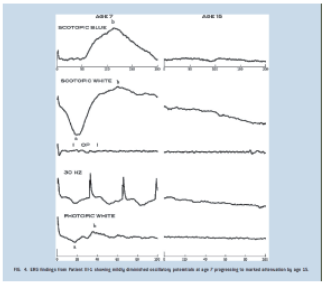
Lloyd B. Williams, Chad D. Huff, Denise J. Morgan, Rosann Robinson, Margaux A. Morrison, Krista Kinard, George Rodgers, Kathleen B. Digre, Margaret M. DeAngelis

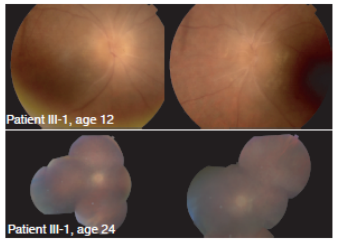
T237M Mutation in ALPK1 is identified as the likely causative mutation in Autosomal Dominant Digre-Williams Syndrome

Phenotype







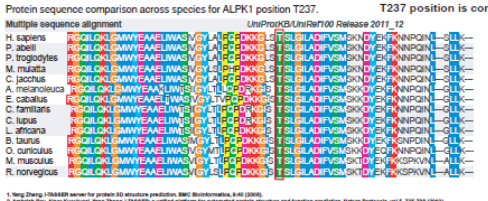


Genotype

Candidate genes identified using VAAST

Gene name	p-value	chromosome	position (hg19)	Allele change	Protein change
PRAMEF11	6.10E-06	chr1	12887174	C->T	R->H
ANKRD20A4	7.32E-06	chr9	69423637	G->A	E->K
MRPL4	8.55E-06	chr19	10367459	C->T	R->W
FAM90A10	9.77E-06	chr8	7629232	G->T	A->S
GOLGA6L10	9.77E-06	chr15	82635194	T->C	E->G
FAM90A20	1.28E-05	chr8	7155458	C->G	A->G
EEF1A1	1.65E-05	chr6	74228474	C->T	R->H
MSI1	1.65E-05	chr12	120800875	C->T	V->M
VDAC2	1.71E-05	chr10	76980685	G->T	A->S
PIM1	2.08E-05	chr6	37138779	A->T	K->M
USP11	2.26E-05	chrX	47104817	G->A	V->M
TAS2R31	2.62E-05	chr12	11183427	T->G	S->R
ABCF1_DUP_06	2.87E-05	chr6	30539278	C->T	P->L
STAU2	3.30E-05	chr8	74529664	T->C	I->V
PRAMEF4	3.48E-05	chr1	12939510	C->T	S->N
FRG1	3.97E-05	chr4	190876196	G->A	A->T
TAS2R19	4.52E-05	chr12	11174390	G->A	L->F
ZNF527	4.64E-05	chr19	37879853	C->T	P->L
NBPF10	4.64E-05	chr1	145368518	C->T	S->L
CTBP2	4.82E-05	chr10	126678163	G->A	A->V
PRSS1	5.43E-05	chr7	142460764	G->A	V->I
ZNF846	6.96E-05	chr19	9869202	T->A	N->I
CLDN25	7.94E-05	chr11	113651170	T->C	F->S
SIGLEC10	9.03E-05	chr19	51917709	G->A	T->M
SIRPA	0.000136	chr20	1895965	C->A	N->K
THNSL1	0.000197	chr10	25314307	G->T	E->*
TBRG1	0.000222	chr11	124495568	T->C	Y->H
PDIA2	0.000263	chr16	334579	C->T	P->L
AKAP6L	0.000302	chr19	15514392	C->T	A->T
CHST15	0.000337	chr10	125780759	G->C	P->A
STARD8	0.000596	chrX	67937592	A->C	E->A
ALPK1	**0.000840**	**chr4**	**113348736**	**C->T**	**T->M**
PTPN13	0.001565	chr4	87691019	C->G	S->R
FAM90A13_DUP_020	0.002901	chr8	7575210	G->C	M->I
DMRTA2	0.003176	chr1	50886700	G->C	A->G
DUX4L4	0.003529	chr4	191003471	C->T	S->L
CDKN2A	0.003686	chr9	21971185	C->A	R->L
USP17	0.003960	chr4	9217197	T->C	S->P
FDX1	0.004235	chr11	110300857	G->A	G->R
NBPF16_DUP_01	0.004645	chr1	148754897	A->G	Y->C
MUC17	0.008888	chr7	100678029	G->C	R->T
MLL3	0.009841	chr7	151945225	T->C	E->G

Protein sequence comparison across species for ALPK1 position T237.



Methods / Results

WES on 4 individuals - 2 affected II-3 III-2, and 2 unaffected II-2 and III-3
Illumina Truseq Enrichment Kit
Quantitative PCR enrichment
Sequencing with Illumina HiSeq 2000 101 cycle paired end sequencing

Mapped and aligned with Picard Tools (http://picard.sourceforge.net)
SNPs and indels identified with Genome analysis toolkit (GATK)
Manual inspection and curation was done with Intergrative Genomics Viewer (http://www.broadinstitute.org/igv)

VAAST - compares to HGMD, dbSNP, ESP, and1000 Genomes panels to rule out common variants.

Non-synonymous, synonymous | Exonic

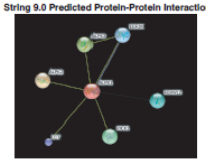
SIFT, PolyPhen 2 QFRP
Alternative Splicing
Splice site prediction

Swiss model
Functional assays, search for available animal models, qRT-PCR, RNA-Seq

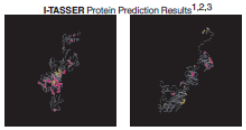
In silico bioinformatics: sequence conservation in species, gene expression in rat/mar/spleen, gene card (www.genecards.org)

Secondary Structure Predictions, Solvent Accessibility, Fold Prediction, Sequence Information (PredictProtein, PsiPred, I-TASSER)

String 9.0 Predicted Protein-Protein Interactions[4]



I-TASSER Protein Prediction Results[1,2,3]



Conclusions:

WES identifies ALPK1 mutation in Digre-Williams Syndrome
Disease causing mutation is inherited in AD pattern
SNP is chr4:113348736 C→T, located in exon 7 of 16 in *ALPK1*
ALPK1 T237M mutation cosegregates with disease
3 of 3 affected, 0 of 6 unaffected.
T237M mutation is predicted to be damaging - SIFT, PolyPhen2
T237 position is conserved across species

UNIVERSITY OF UTAH HEALTH CARE
Moran Eye Center

Corresponding Author:
Margaret M. Deangelis.
margaret.deangelis@utah.edu

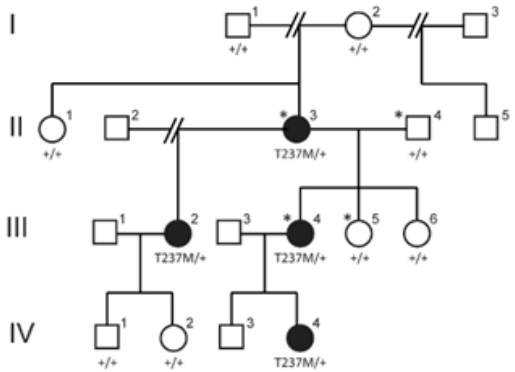
1. Yang Zhang. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9:40 (2008).
2. Ambrish Roy, Alper Kucukural, Yang Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, vol 5, 725-738 (2010).
3. Ambrish Roy, Jianyi Yang & Yang Zhang. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Research, vol 40, W471-W477 (2012).
4. Franceschini A., Szklarczyk D., Frankild S., et al., STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013 Jan;41.

Poster ARVO congres 2013

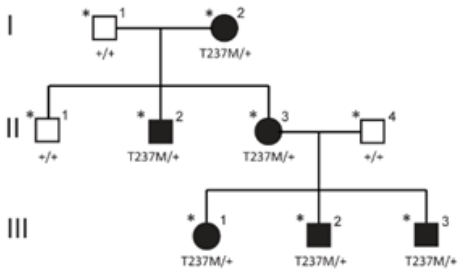
Erasmus MC
University Medical Center Rotterdam

Contact with group in VS

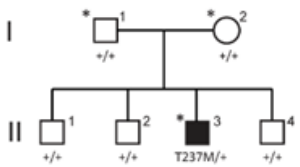
Family 1 (Utah cohort)



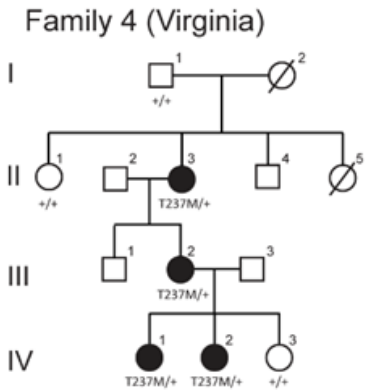
Family 2 (Australian cohort)



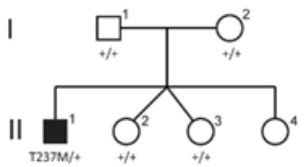
Family 3 (Netherlands)

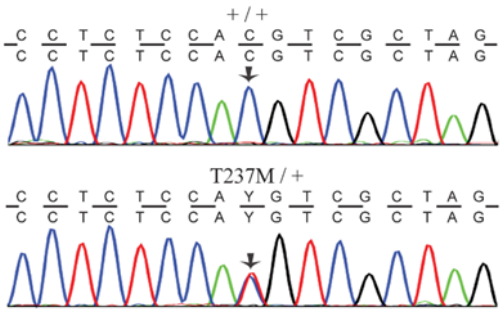


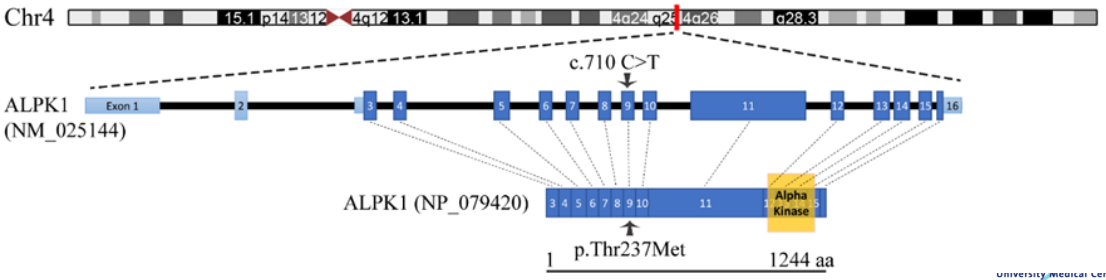
Family 4 (Virginia)



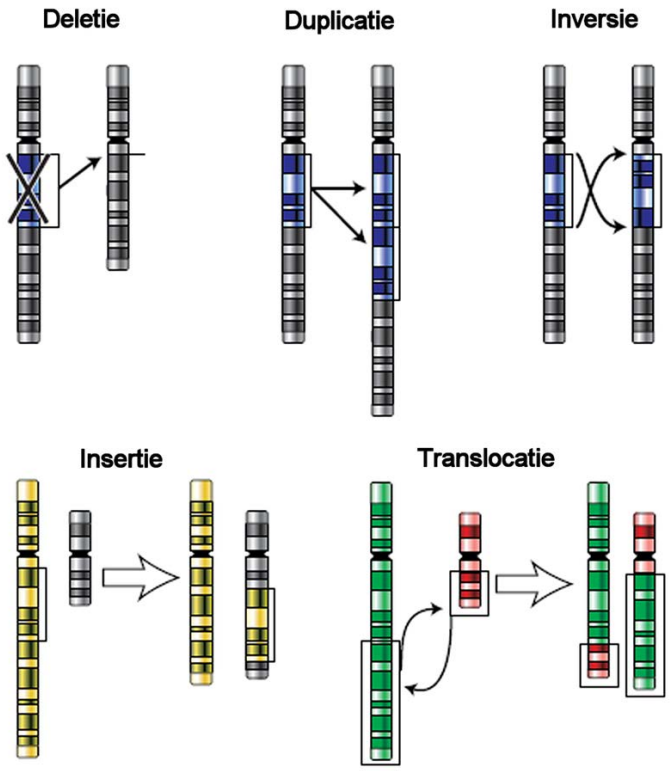
Family 5 (Delaware cohort)







Limitations of WES



Also deeper intronic variants
Variants in promotor region

Whole Genome Sequencing?
The holy grail?

Conclusions/ Discussion

Important for WES diagnostics:

Datasharing, good databases
Multidisciplinairy meetings
Bioinfomaticians
HPO terms in future (not used yet on routine basis)

Discussion: What to report is a main discussion point!

Examples incidental findings:
1) Finding XXY when you do the quality control
2) You find a pathogenic BRCA1 variant in child and parent
3) DPYP pathogenic variants: "Dihydropyrimidine dehydrogenase deficiency", risk factor for 5-FU hypersensitivity
4) Carrier of AR disease: CFTR

Questions?

Acknowlegdements

Clinical Genetics EURMC

NGS unit

Joan Kromosoeto

Grazia Mancini

Alice Brooks

Iris Hollink

Walter de Valk

Frank Sleutels

Martina Wilke

Rick van Minkelen

Ans van den Ouweland

Hennie Bruggenwirth

Lies Hoefsloot