



A Real-life Experience on the Challenge Called Genetics



2017 Paper AS05

Sigrid Fouchier, OCS Life Sciences, the Netherlands

Pharmacogenetics (PGt)



Pharmacogenetics (PGt) , study of variations in DNA sequence related to drug response.

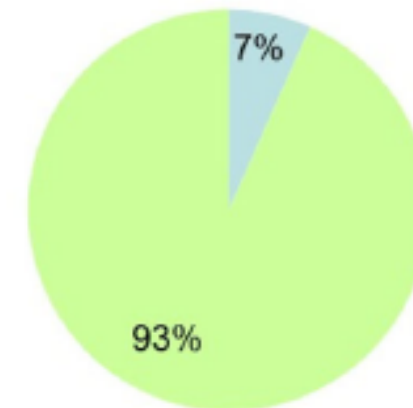
PGx has been widely recognized as the fundamental steps toward personalized medicine.

Clinical implementation of pharmacogenomics

Only about 7% of medications have actionable germline pharmacogenetics

Only 17 of ~ 19,000 human genes are considered clinically actionable for germline pharmacogenomics

By FDA approved medications (n=1200)



Relling *et al.* Nature. 2015;526: 343–350.



Protein-Coding Genes and Mendelian Phenotypes

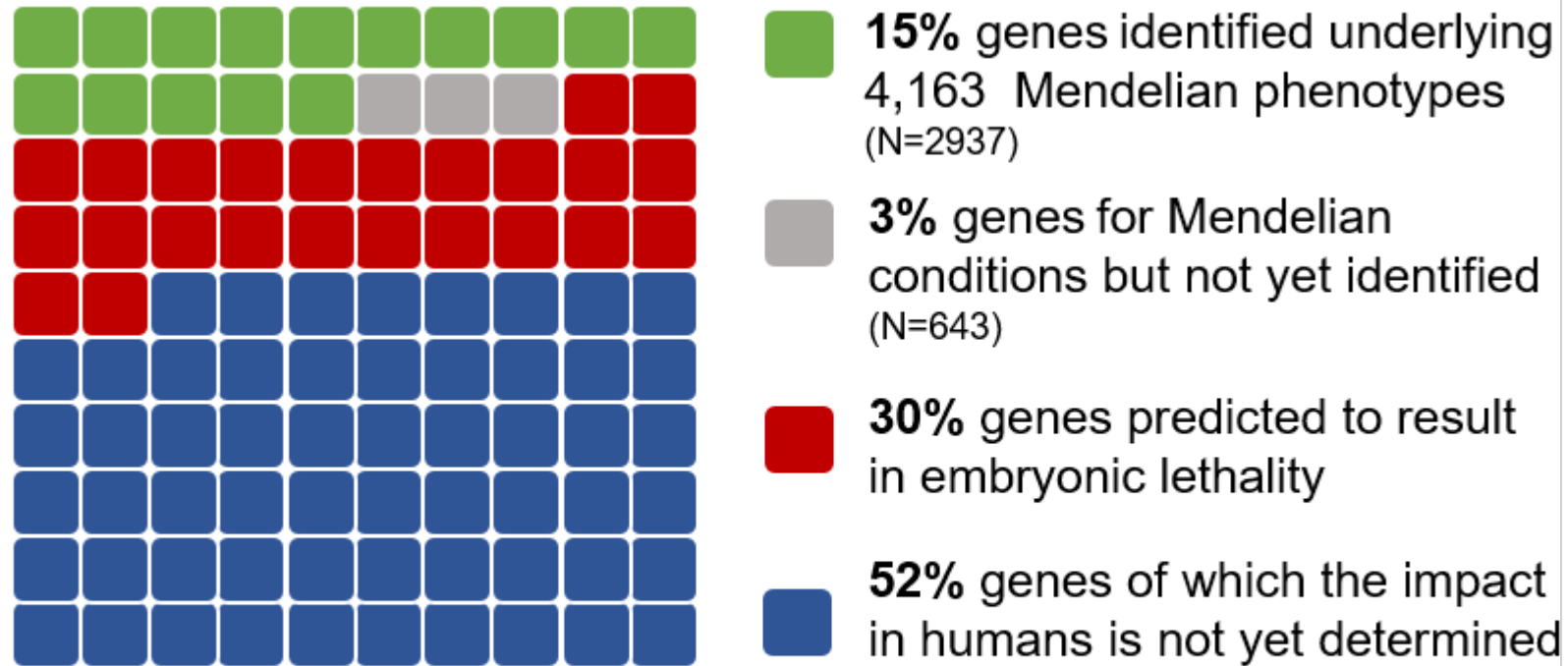
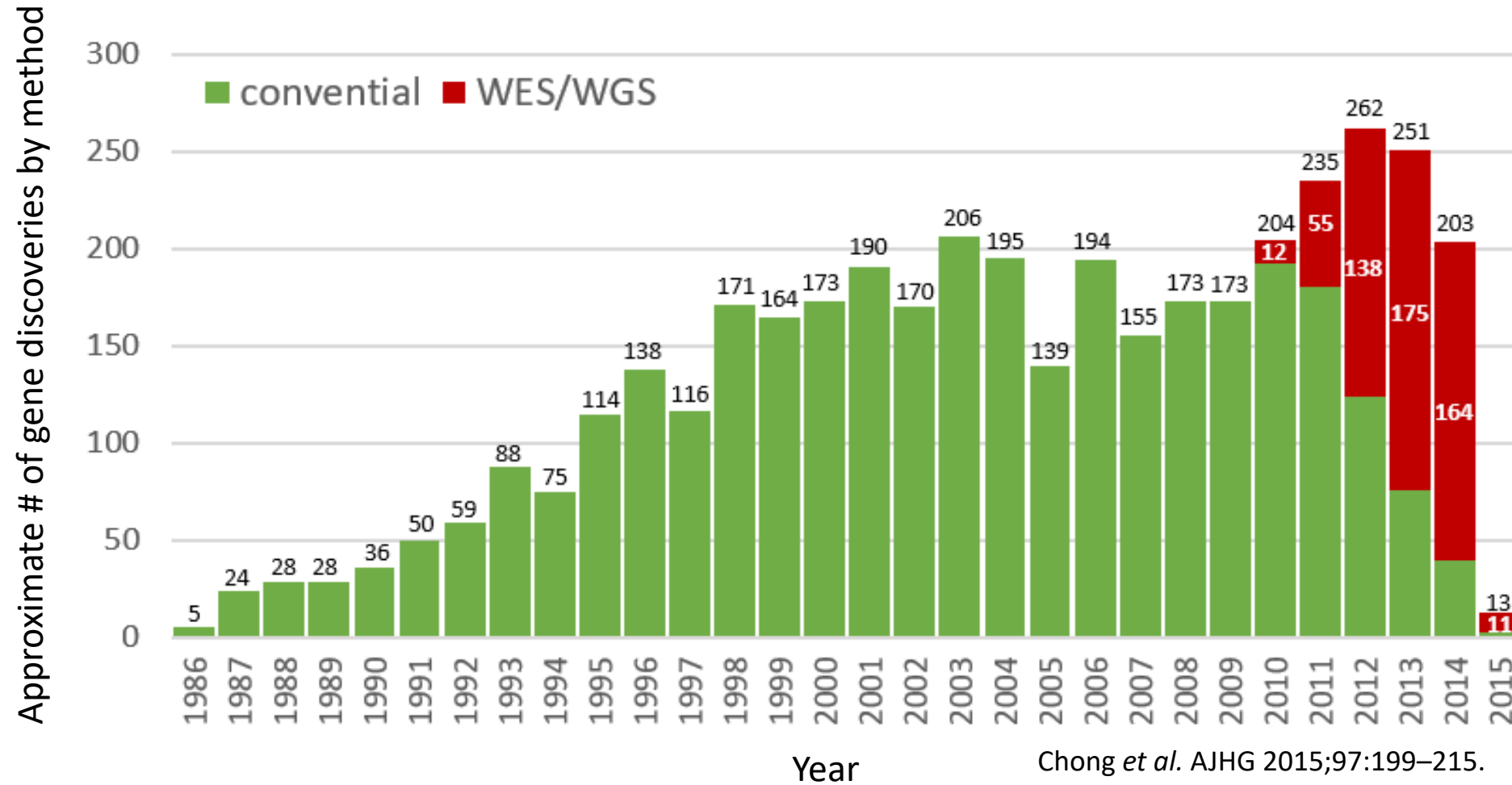


Figure 1. Relationship Human Protein-Coding Genes and Mendelian Phenotypes.

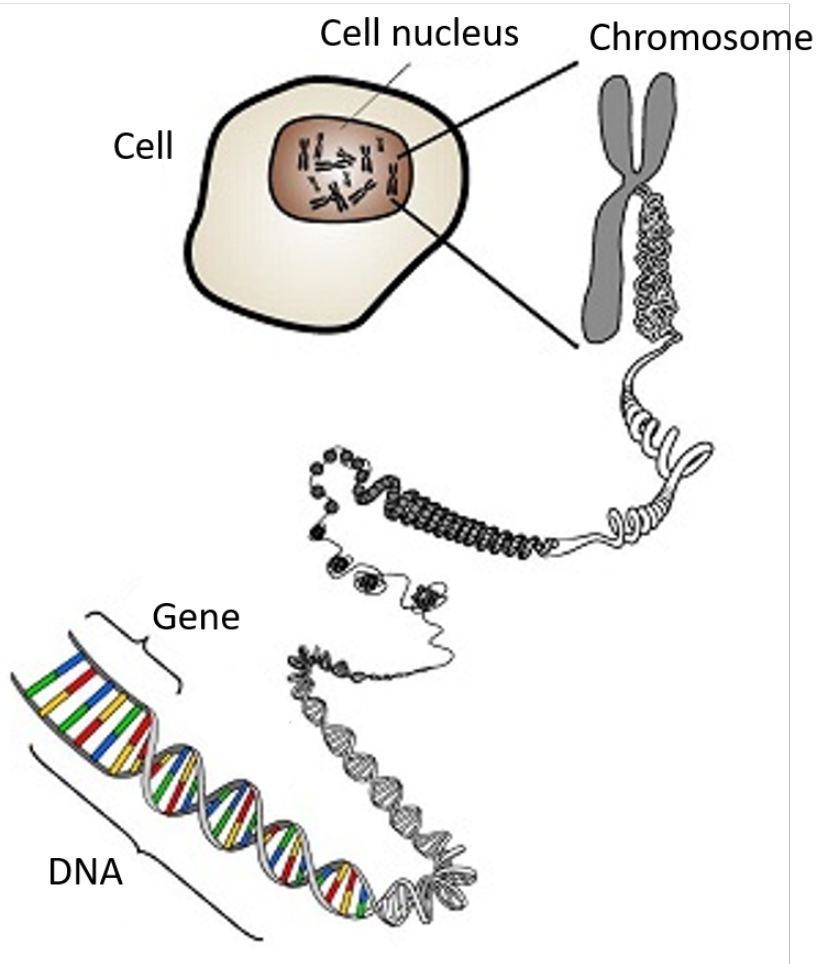
Chong *et al.* AJHG 2015;97:199–215.



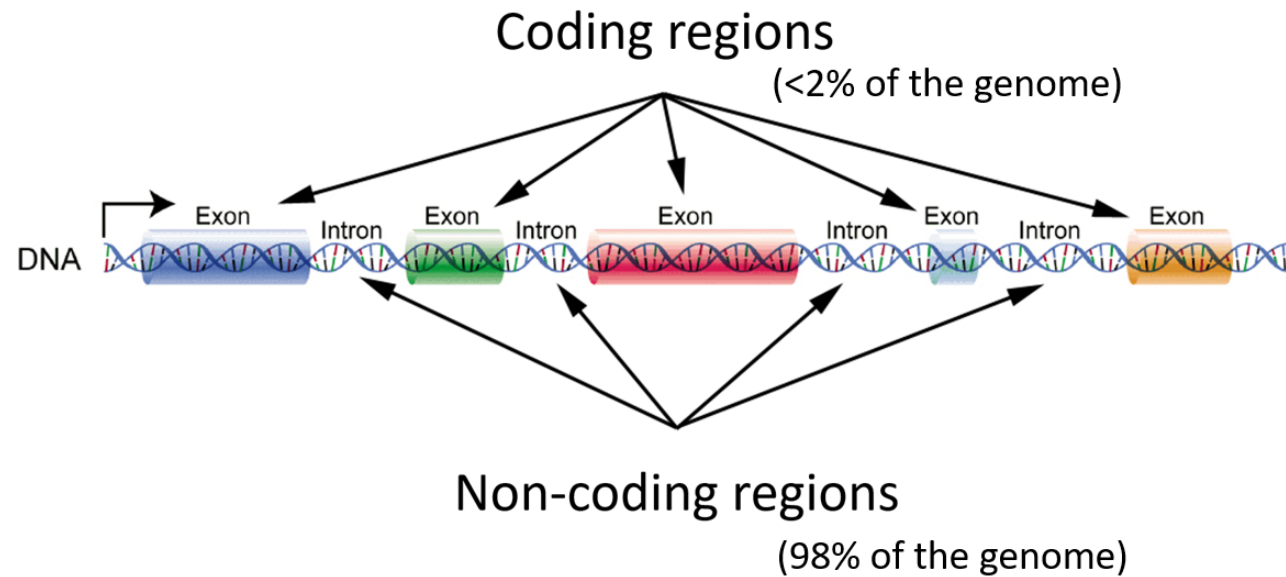
Protein-Coding Genes and Mendelian Phenotypes



Genetics – The Basics

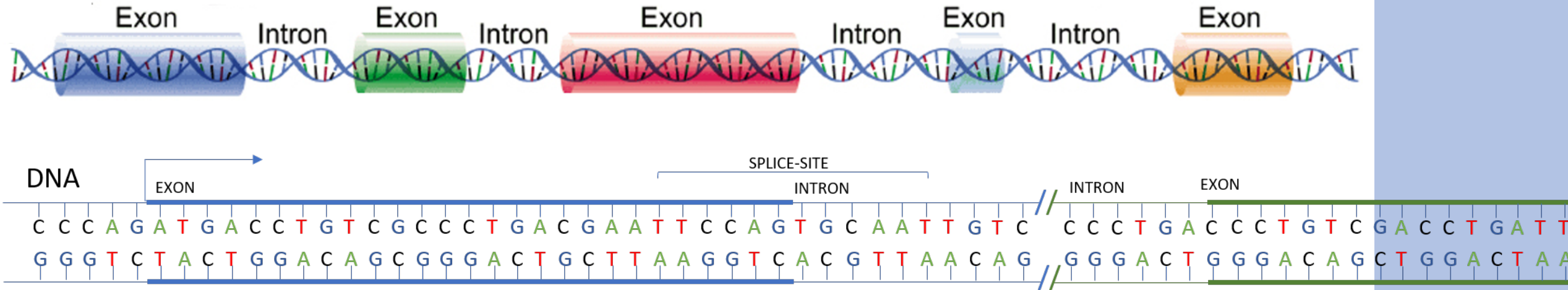


Whole-exome sequencing (WES) is an application of the next-generation technology that determines the variations in and around all coding regions of known genes (~19,000 genes, <2% of the human genome).





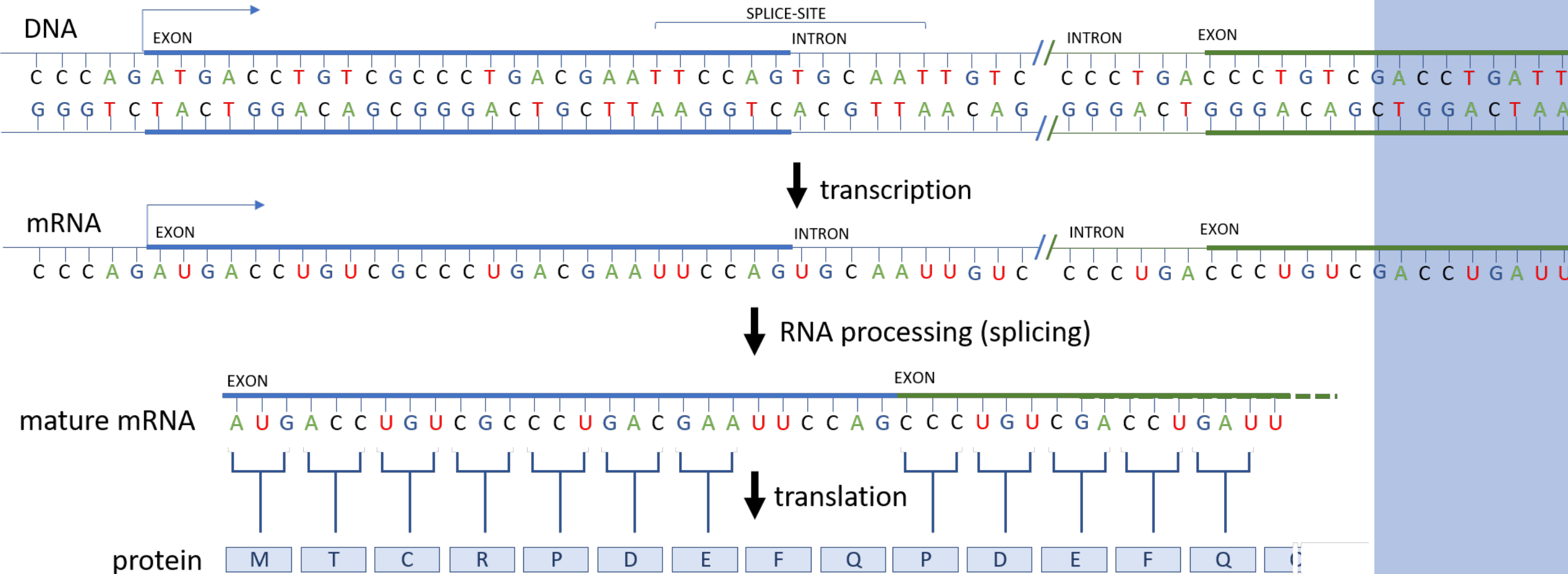
DNA – RNA - Protein



DNA is the blueprint of life containing the instructions needed for an organism to grow, develop, survive and reproduce, controlling protein synthesis.

Proteins do most of the work in cells, and are the basic unit of structure and function in the cells of organisms.

DNA – RNA - Protein



Genetic variants



A single nucleotide variation (SNV) is a variation in a single nucleotide.

Single Nucleotide Polymorphism (SNP)

- common (found in at least 1% of the population)
- non-functional
- located in the non-coding regions

Mutation

- rare
- functional, disease-causing
- located in functional domain of the genome (i.e. exon, splice-site, promoter region)

Types of mutations: silent, missense, nonsense, frame-shift, indels, splice.

Mutations



Type of mutation	Genotype	Effect	Phenotype
Consensus	EXON SPLICE-SITE INTRON C C A G A T G A C C T G T C G C C C T G A C G A A T T C C A G T G C A M T C R P D E F Q		
Silent	C C A G A T G A C C T G T C G C C C C G A C G A A T T C C A G T G C A M T C R P D E F Q	Base pair change NO amino acid change (synonymous)	none
Missense	C C A G A T G A C C T G T C G C C T T G A C G A A T T C C A G T G C A M T C R L D E F Q	Base pair change AND amino acid change (non-synonymous)	moderate
Nonsense	C C A G A T G A C C T G A C G C C C T G A C G A A T T C C A G T G C A M T X	Base pair change, resulting in a stop codon	severe
Frame-shift (Indel)	C C A G A T G A C C T G T C G C C T G A C G A A T T C C A G T G C A M T C R L T N S S	Deletion of one base pair (indel), resulting in frame-shift	severe
Splice	C C A G A T G A C C T G T C G C C C T G A C G A A T T C C A G T A C A M T C R P D E F Q Y	Base pair change in splice-site, resulting in translation of intronic sequence	severe

Coding variants by WES

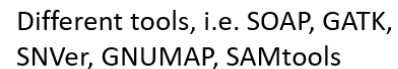


Variant type	Total variants	SD	Non-novel variants	SD	Novel variants	SD
synonymous	10,645	(± 286)	10,536	(± 288)	109	(± 16)
missense	9,511	(± 244)	9,319	(± 233)	192	(± 21)
nonsense	93	(± 6)	89	(± 6)	5	(± 2)
splice	34	(± 4)	32	(± 3)	2	(± 1)
total	20,283	(± 523)	19,976	(± 505)	307	(± 33)

Mean number of coding variants in one European individual Bamshad *et al.* Nature 2011;12:745–755.

Non-novel variants: reported in publicly available SNV databases
(dbSNP, the 1000 Genomes Project, ExAC Database, GoNL)

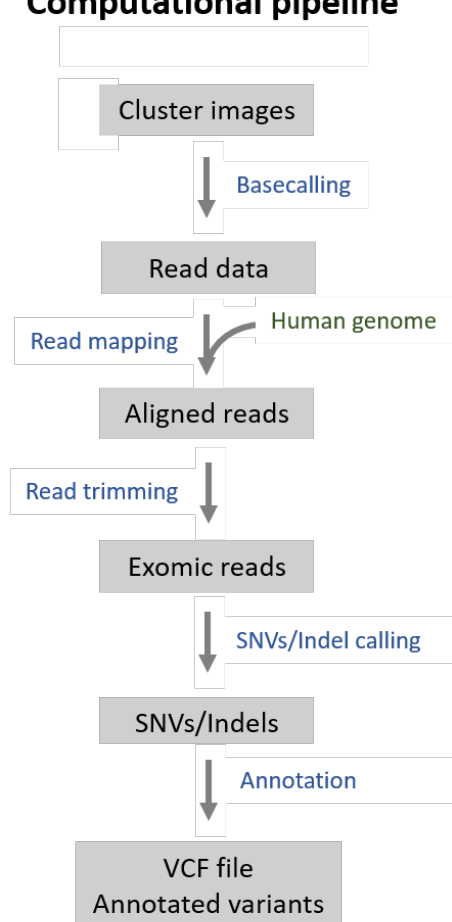
Novel variants: private and rare variations in one individual.



Typical WES analysis pipeline



Computational pipeline



Variant type	Total variants	SD
synonymous	10,645	(± 286)
missense	9,511	(± 244)
nonsense	93	(± 6)
splice	34	(± 4)
total	20,283	(± 523)

correctly called variants

Workflow for exclusion and prioritization of exome sequencing variants

Different tools, i.e. SOAP, GATK, SNVer, GNUMAP, SAMtools



Typical WES analysis pipeline

Workflow for exclusion and prioritization of exome sequencing variants

Exclusion

1. Selecting variants fitting the disease model (heterozygous, homozygous)
2. The disease-causing variant should be absent or with low frequency (<1%) present in publicly available databases

Prioritization

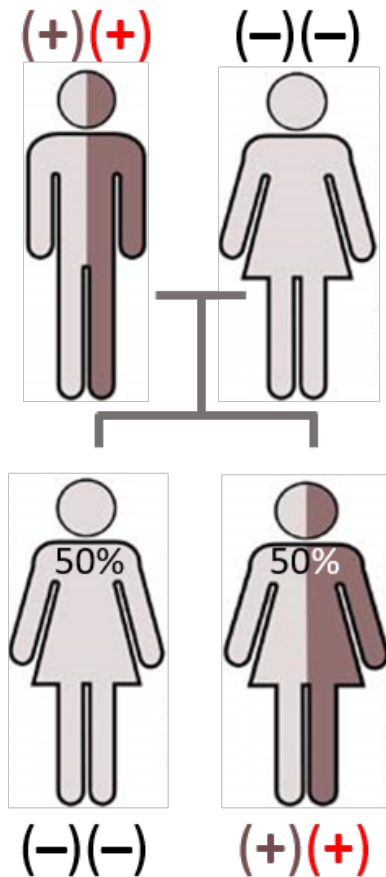
1. Focus on variants likely to effect protein function (missense, nonsense, frame-shift (indels) and splice mutations) using different type of prediction models (SIFT, Polyphen-2, splicing prediction tool)
2. Confirm the variant is not false-positive by traditional methods (i.e. Sanger sequencing)
3. Co-segregation analysis in families (genotype-phenotype correlation)
4. Sequencing or screening for the absent of the genetic variant in ethnically matched healthy controls
5. Sequencing the whole gene of interest in a cohort of disease-related individuals to identify the same or additional variants in the gene of interest

The Real Life-Experience

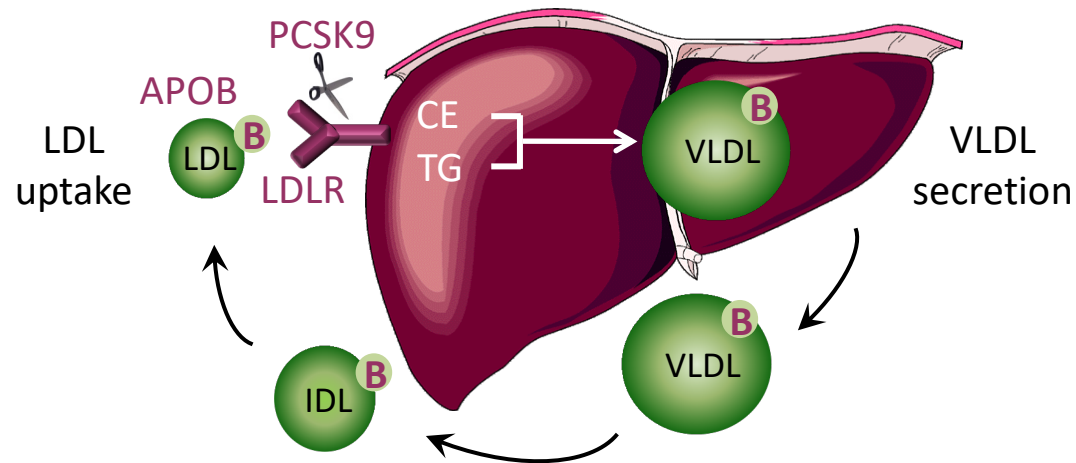


Autosomal Dominant Hypercholesterolemia

Prevalence 1:200



- Increased total- and LDL-cholesterol (> 95th percentile)
- Increased risk for premature cardiovascular disease (CVD)
- Since 1991 ~20,000 clinical diagnosed ADH



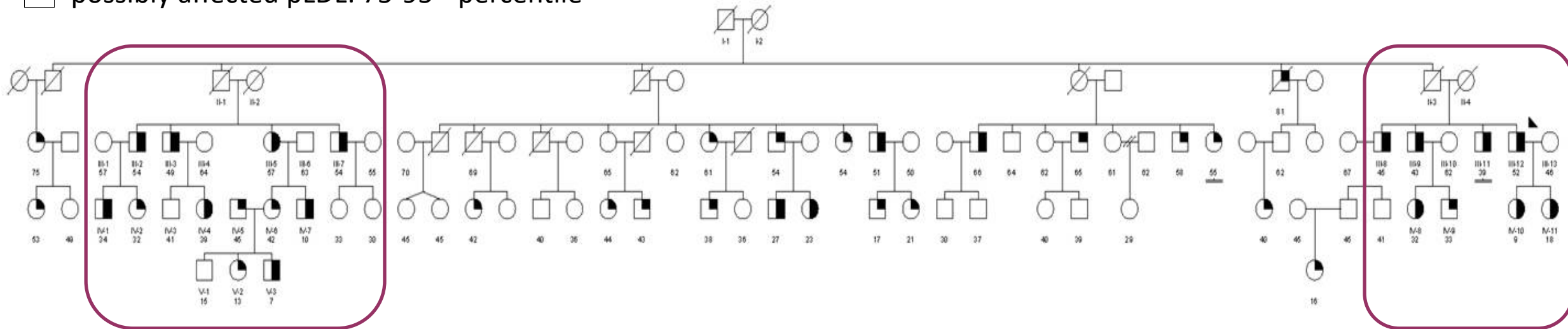
- ~30% molecular explained by 3 genes (LDLR, APOB and PCSK9)
- 70% 'FH4'; polygenetic or **novel** monogenetic genes

The Real Life-Experience

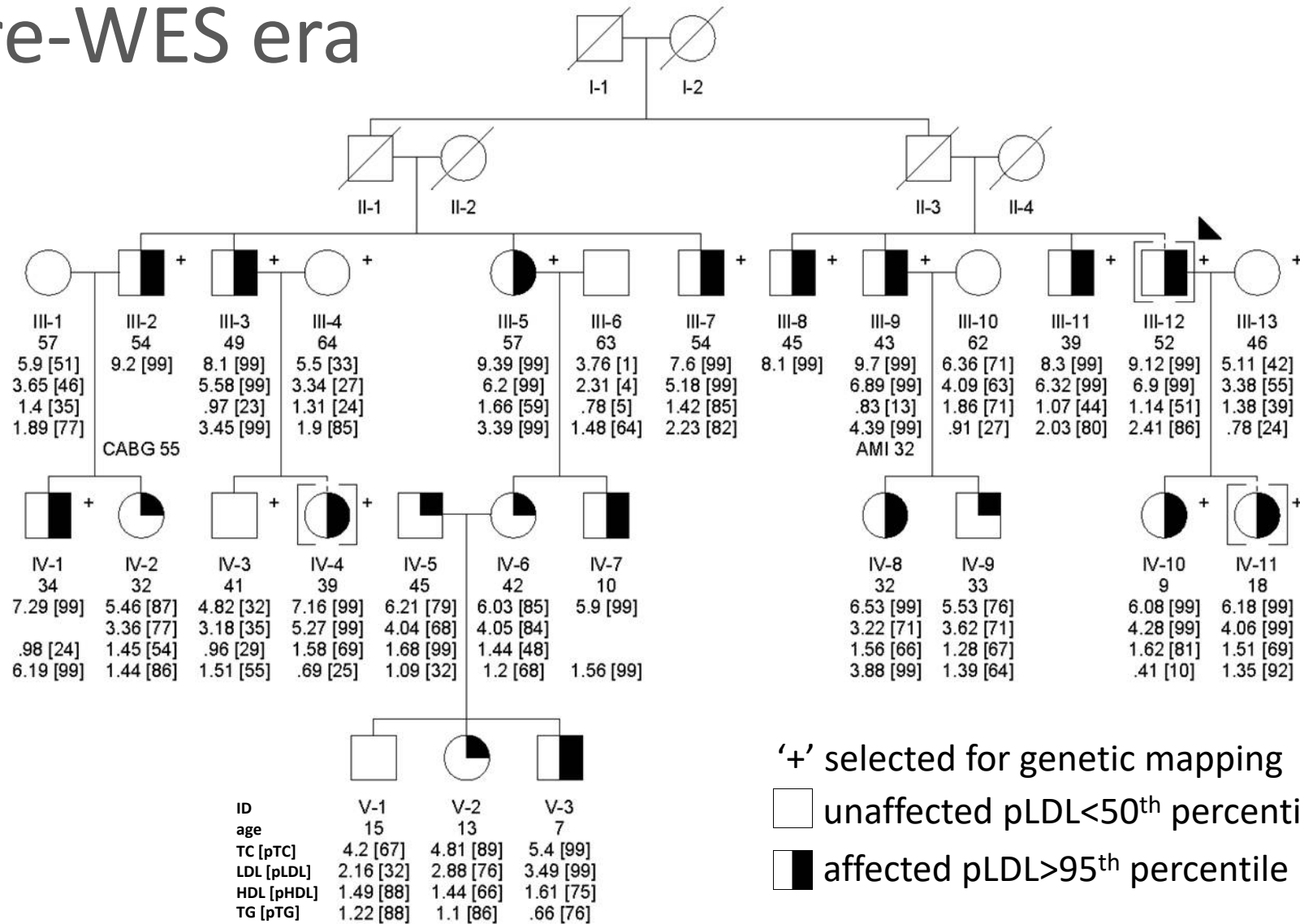


Identification of novel ADH genes is largely challenged by the phenotypic and genetic heterogeneity of this disease.

- affected pLDL > 95th percentile
- ◼ possibly affected pLDL: 75-95th percentile



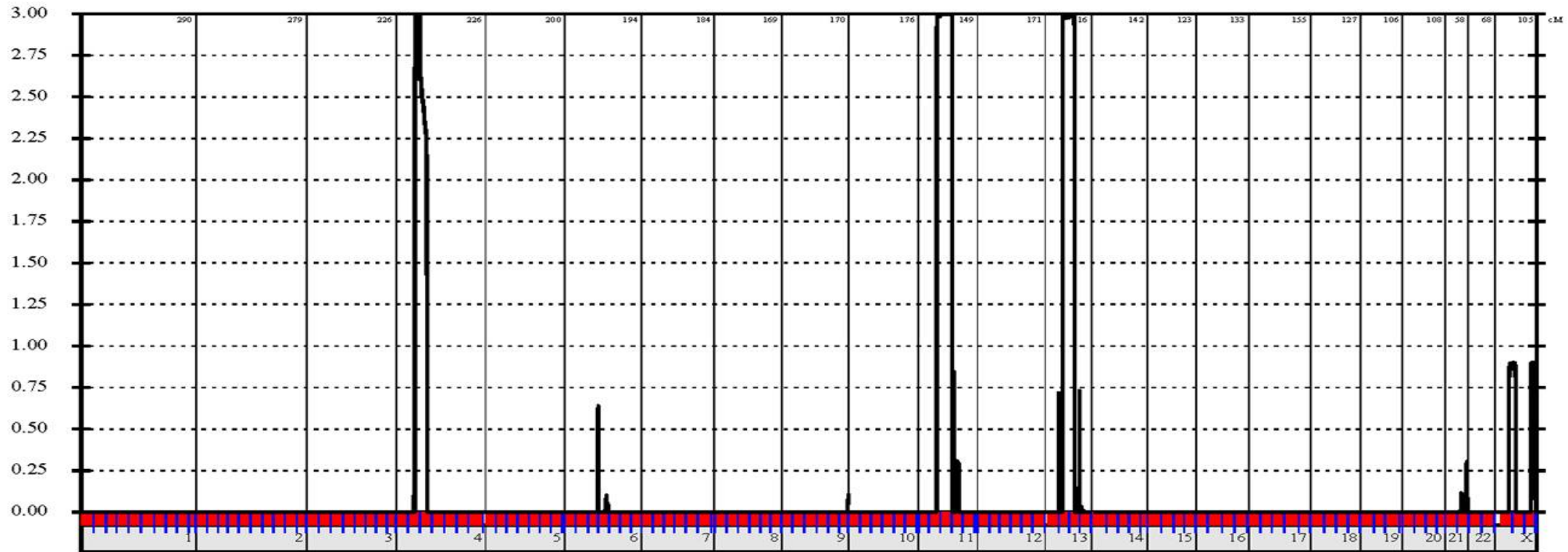
The pre-WES era



Linkage analysis

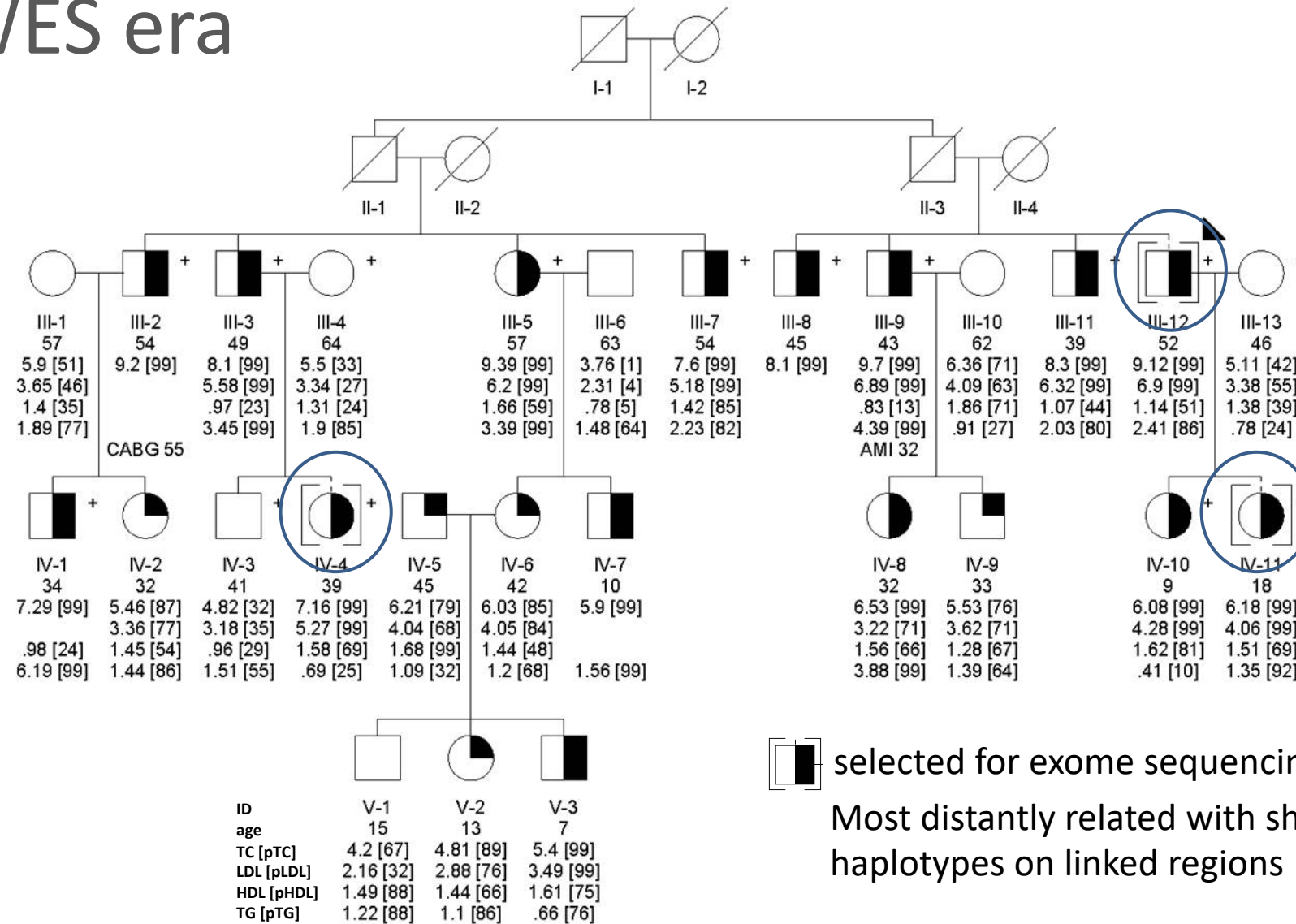


Parametric LOD scores



Multi-point linkage analysis revealed highly suggestive LOD scores of 3.0 ($\theta = 0.0$) at chromosomes 4p15.1-q13.3, 11p13-q14.1, and 13q14.13-q32.1

Into WES era





Workflow for exclusion and prioritization of exome sequencing variants

Exclusion

	# candidate variants	# candidate genes
1. disease model (heterozygous)	5,953	2,822
2. publicly databases (<1%)	394	315

Prioritization

- Variants likely to effect protein function

coding variant	149	131
exclusion of silent mutations	64	58
- Confirm the variant is not false-positive by traditional methods (i.e. Sanger sequencing)
- Co-segregation analysis in families (genotype-phenotype correlation)
- Sequencing or screening for the absent of the genetic variant in ethnically matched healthy controls
- Sequencing the whole gene of interest in a cohort of disease-related individuals to identify the same or additional variants in the gene of interest



Workflow for exclusion and prioritization of exome sequencing variants

Exclusion

	# candidate variants	# candidate genes
1. disease model (heterozygous)	5,953	2,822
2. publicly databases (<1%)	394	315
3. In-house database with phenotype	[167]	[143]
4. linkage analysis (Chr 4, 11, 13)	[33]	[27]

Prioritization

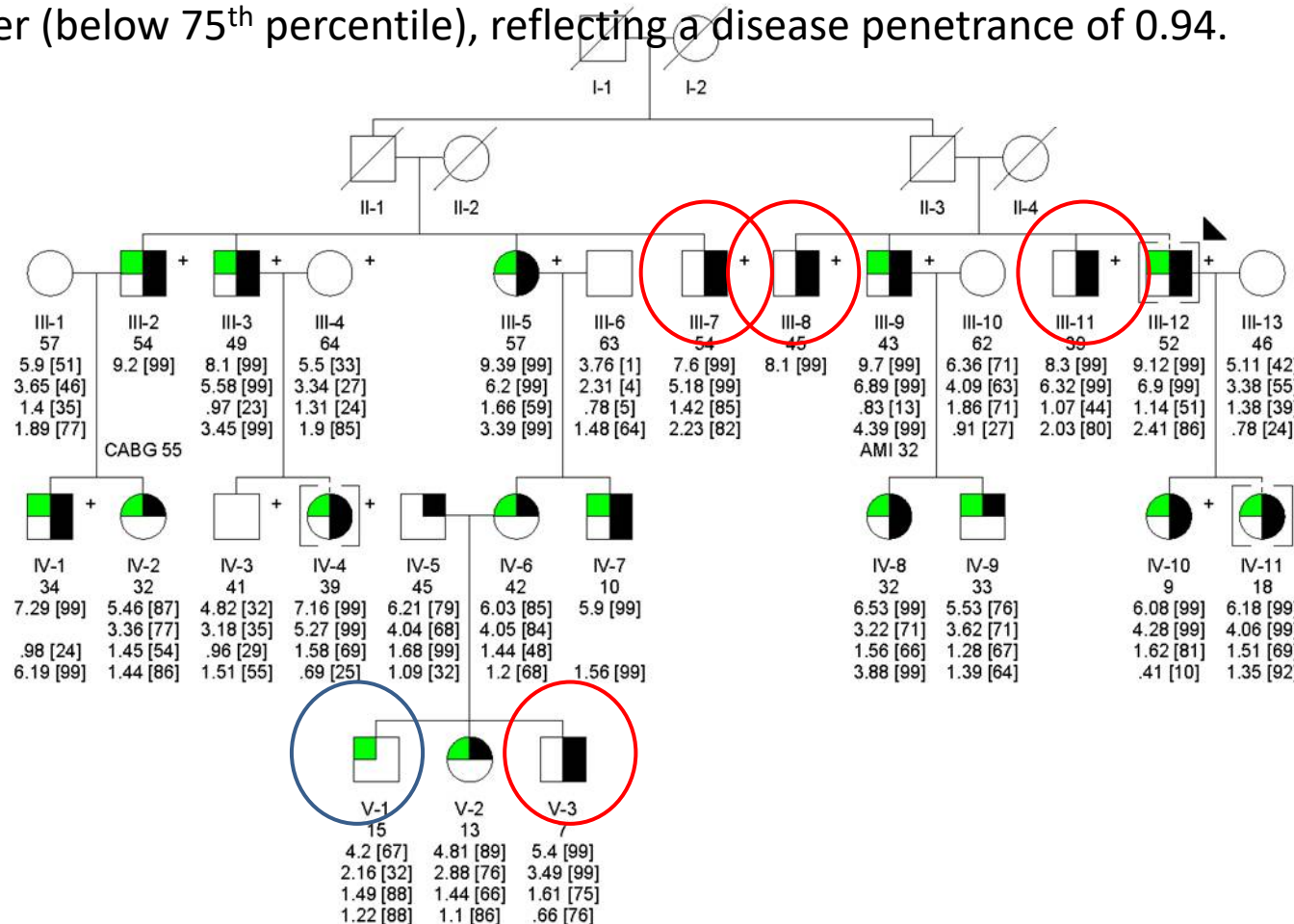
- Variants likely to effect protein function

coding variant	149	[13]	131	[8]
exclusion of silent mutations	64	[3]	58	[3]
- Confirm the variant is not false-positive by traditional methods (i.e. Sanger sequencing)
- Co-segregation analysis in families (genotype-phenotype correlation)
- Sequencing or screening for the absent of the genetic variant in ethnically matched healthy controls
- Sequencing the whole gene of interest in a cohort of disease-related individuals to identify the same or additional variants in the gene of interest

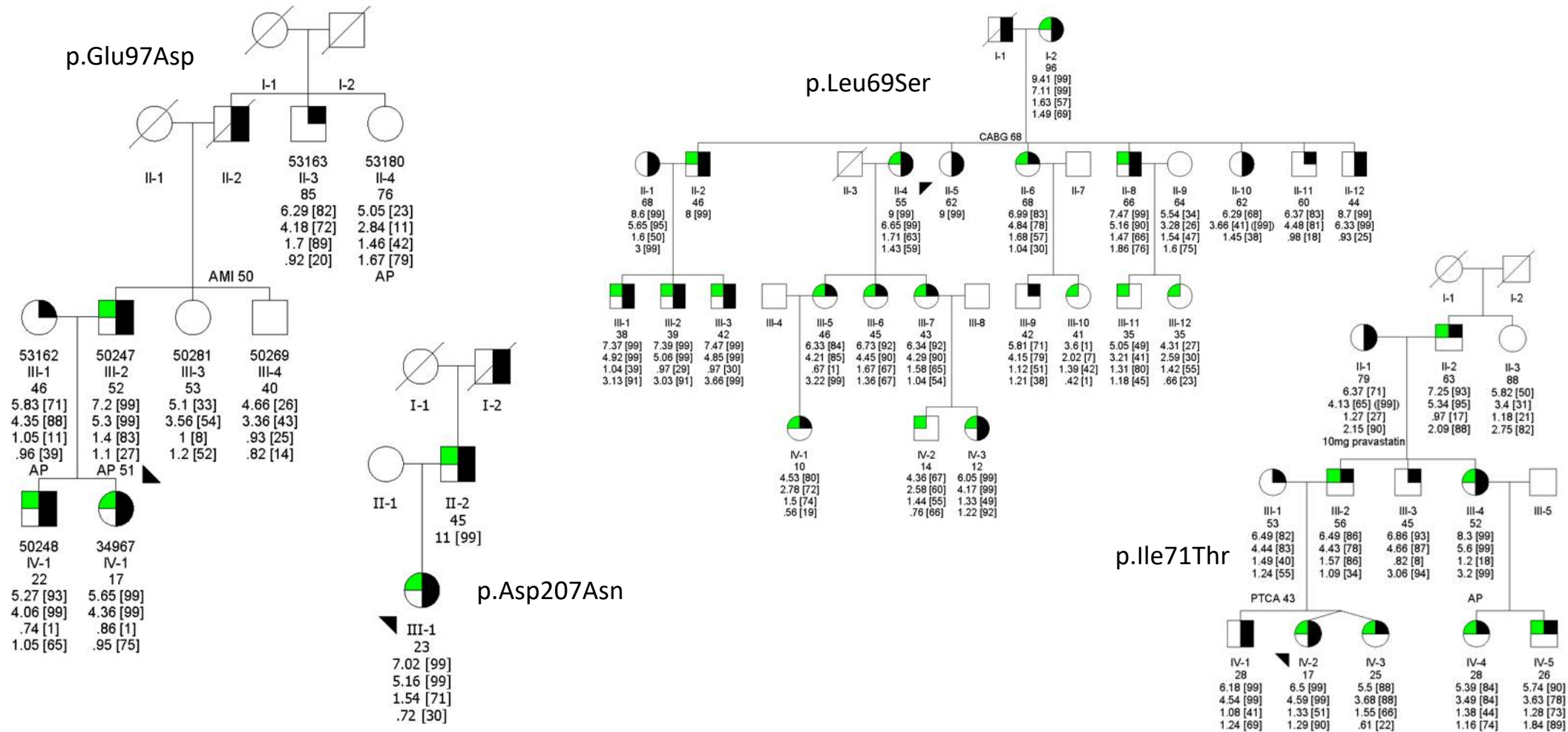


Co-segregation *STAP1*: p.Glu97Asp

11 affected carriers (total cholesterol (TC) or LDL-c levels above 95th percentile)
4 possibly affected carriers (between 75th and 95th percentile) and
1 unaffected carrier (below 75th percentile), reflecting a disease penetrance of 0.94.



Additional *STAP1* families





A novel ADH gene discovered?

- *STAP1* carriers higher cholesterol than non-carriers ($p=0.001$)
- Comparable phenotype *APOB* mutations carriers

Circulation
Research

JOURNAL OF THE AMERICAN HEART ASSOCIATION

Brief UltraRapid Communication

Mutations in *STAP1* Are Associated With Autosomal Dominant Hypercholesterolemia

Sigrid W. Fouchier, Geesje M. Dallinga-Thie, Joost C.M. Meijers, Noam Zelcer,
John J.P. Kastelein, Joep C. Defesche, G. Kees Hovingh

- *STAP1* expression absent in liver
- Function of *STAP1* protein unknown
- Fundamental research needed
- *STAP1* a convincing discovery?

Summary



WES/WGS alone not sufficient enough in novel- gene finding strategies

- Severe cases / Recessive model

Validity of discovery

- Hampered by genetic and phenotypic heterogeneity
- Fundamental research/knowledge essential

WES

- Covers <2% of the human genome (limited by gaps in coverage)
- ~1.5% rare, private variants (~300 out of ~20,000)
- Low concordance of multiple variant-calling pipelines

WGS

- Much higher rate of rare, private variants
- Publicly SNV databases very limited (no phenotype)

Pharmacogenetics (PGt) , a fundamental step toward **personalized medicine?**