Leiden University
Medical Center

Exome sequencing in Osteoarthritis

A compelling example to identify causal genes

Ingrid Meulenbelt

Molecular Epidemiology

LEIDEN, THE NETHERLANDS

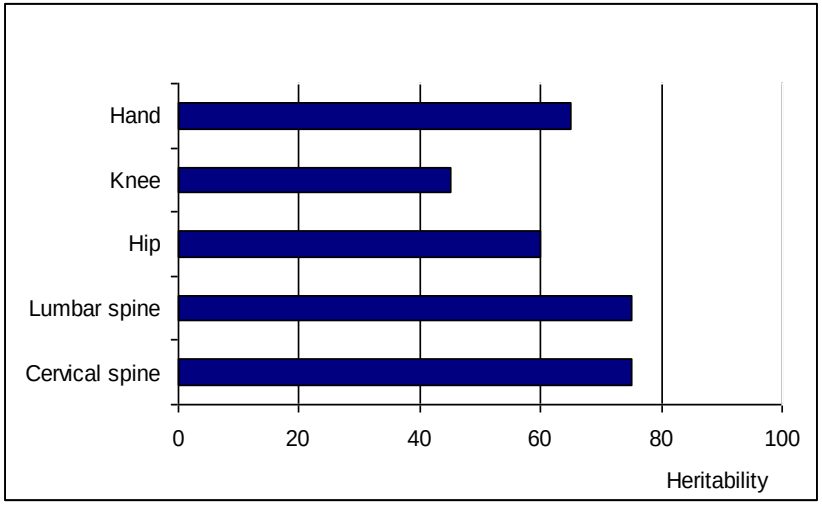
Osteoarthritis (OA)

- Degenerative disease of joint tissues
- Prevalent
- No effective treatment

Major cause of disability among elderly affecting mobility and hampering daily life activities

Risk factors of osteoarthritis

- Age
- Obesity
- Bone mineral density
- Genetics



A genetic component allows you to identify genes that are involved in the disease process

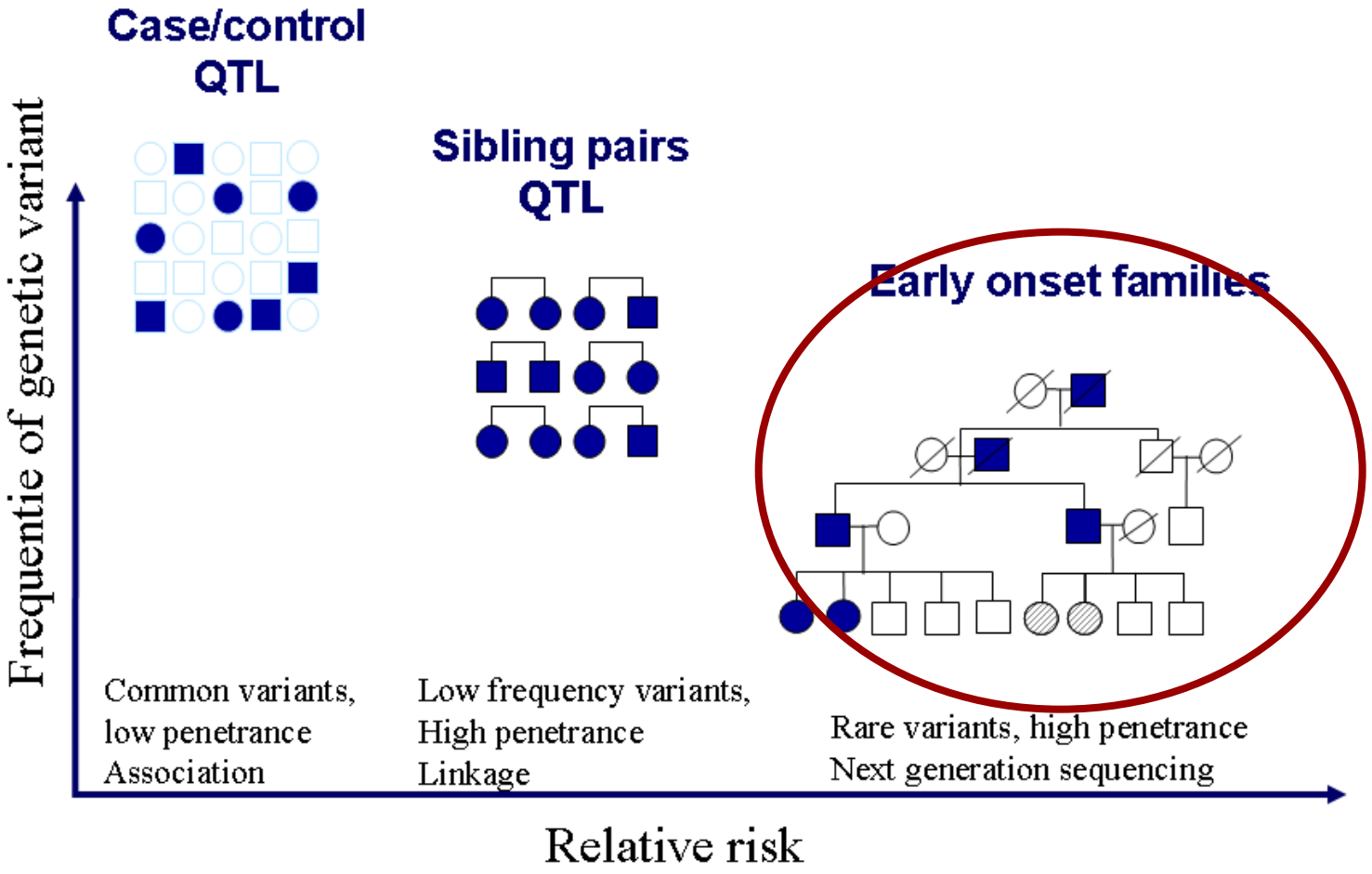
Balance between heritability and environment

complex genetic

100% environment **100 % genetic**

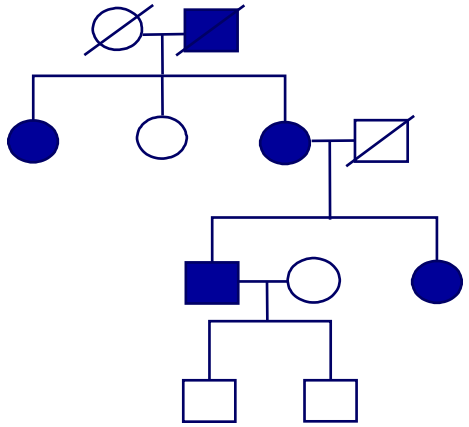
Determines genetic study design

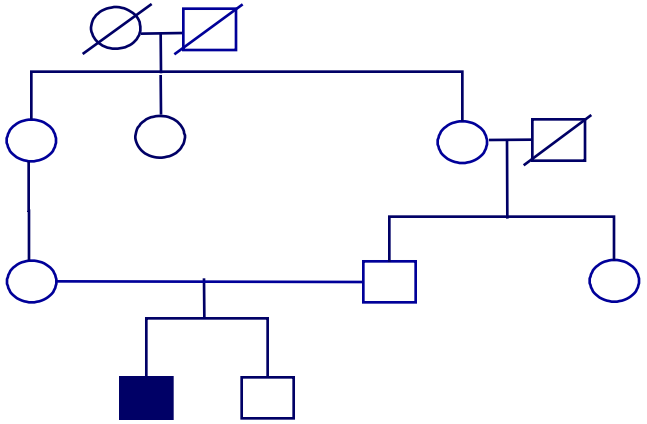
Depending on study design



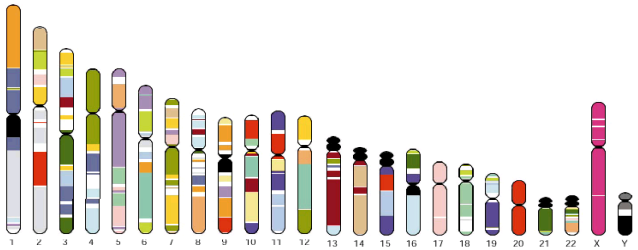
Genetic studies to early onset extended families

- Early onset distinct (rare) phenotype
- Proven powerfull to identify disease genes
- Identification of underlying disease pathways
- Generalizability to common disease forms





Identification of disease genes



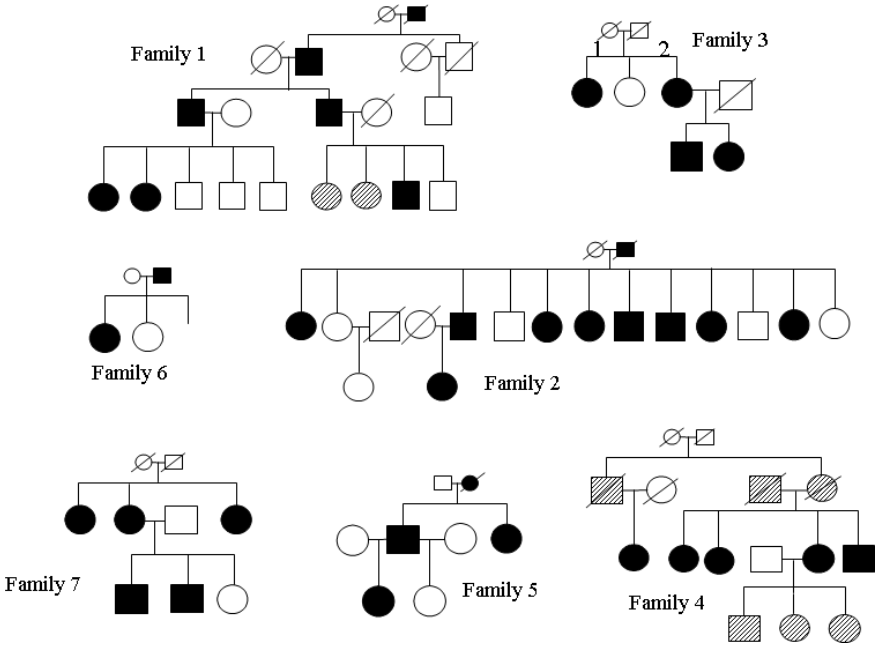
Candidate genes

Hypotheses driven

Genome wide scan

Hypothesis free

Early onset Osteoarthritis families



- Primary osteoarthritis at early ages (~ 40 years)
- No syndromic features
- Dominant Mendelian Inheritance pattern

Candidate linkage

Meulenbelt,I. et al. (1997) Genetic linkage analysis of 14 candidate gene loci in a family with autosomal dominant osteoarthritis without dysplasia. J Med Genet, 34:1024-1027.

Genome wide linkage

Meulenbelt,I et al. (2006): Strong linkage on 2q33.3 to familial early-onset generalized osteoarthritis and a consideration of two positional candidate genes. Eur.J.Hum.Genet., 14:1280-1287

Sequencing linkage area

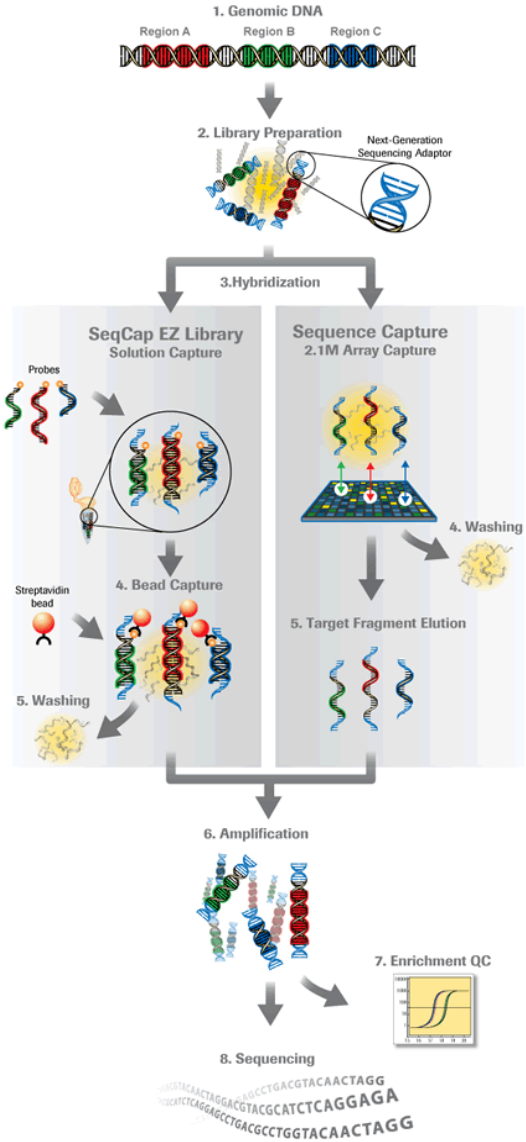
Min,J.L., Meulenbelt,I., et al. (2007): Mutation analysis of candidate genes within the 2q33.3 linkage area for familial early-onset generalised osteoarthritis. Eur.J.Hum.Genet., 15:791-799.

10 years of research no definitive mutation found

Next generation sequencing
Technique

- E.g. Illumina HiSeq2000 technology
- Generated as 100 bp pair-end reads
- ~ 44 MB exonic sequences

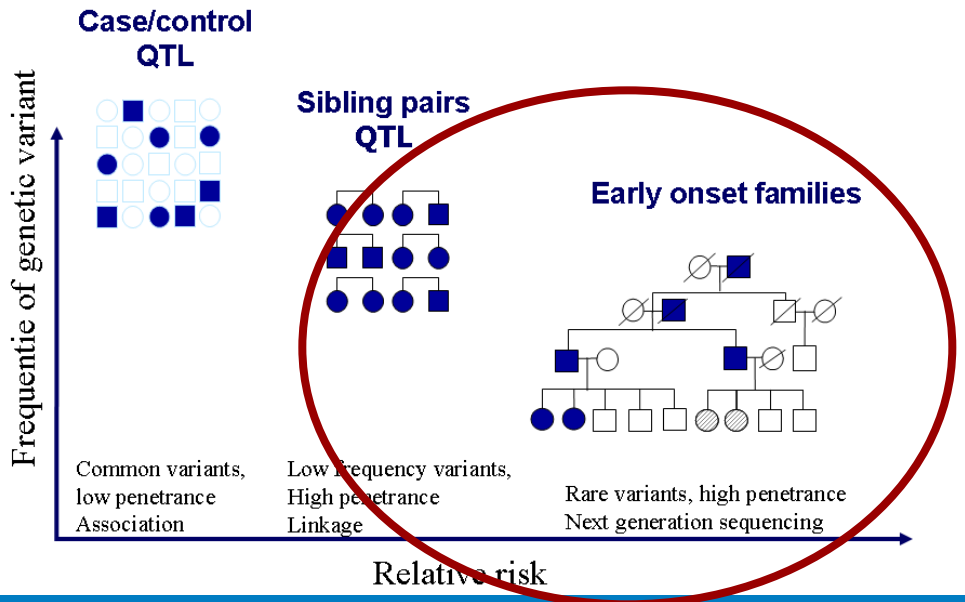
Allows sensitive sequencing of whole exomes or genomes!!!



Exome sequencing
Design

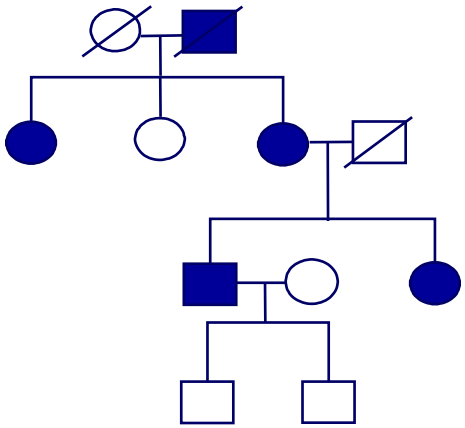
Most successful when applied to:

- Family-based phenotypes (to exploit parent-child transmission patters)
- Extreme spectrum of phenotypes to increase the efficiency of the detection of rare variants with larger effect sizes.



What to expect?

- Rare mutations; private
- Large effect sizes
- Changing protein functions



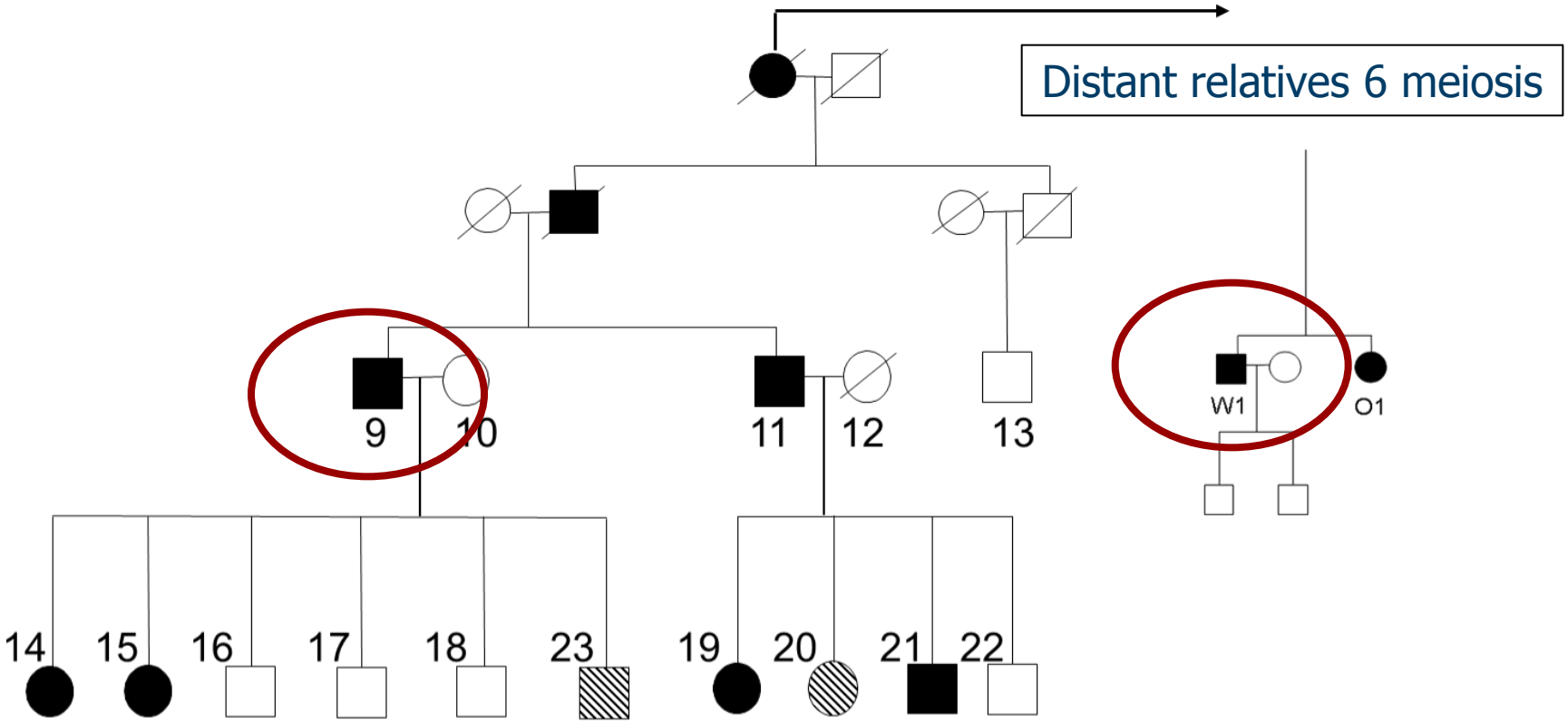
Genetic Codes

5' End	2nd Position				3' End
	T	C	A	G	
T	Phe	Ser	Tyr	Cys	T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	OCH	OPA	A
	Leu	Ser	AMB	Trp	G
C	Leu	Pro	His	Arg	T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Stop COdes: AMBer, OCHer, OPA

Exome sequencing

An extended early onset families



Prioritization scheme, based on identification of novel, high impact pathogenic mutation

	FOA1	FOA2
Numbers of variants identified	57018	60562
Exonic (including 5'utr and 3'utr) and canonical splice sites	15130	15183
Missense, nonsense, readthrough, frame error, splice site	7735	7699
Novel	618	636

Damaging protein function?

Effect of DNA variant on protein function

Depending on:

Codon change

Amino acid change (polarity, molarity)

Protein domain

Splice site

Genetic Codes

5' End	2nd Position				3' End
	T	C	A	G	
T	Phe	Ser	Tyr	Cys	T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	OCH	OPA	A
	Leu	Ser	AMB	Trp	G
C	Leu	Pro	His	Arg	T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Stop COdes: AMBer, OCHer, OPA

Determine *in silico* effect of DNA variant e.g. SIFT

Exome sequencing
Early onset family

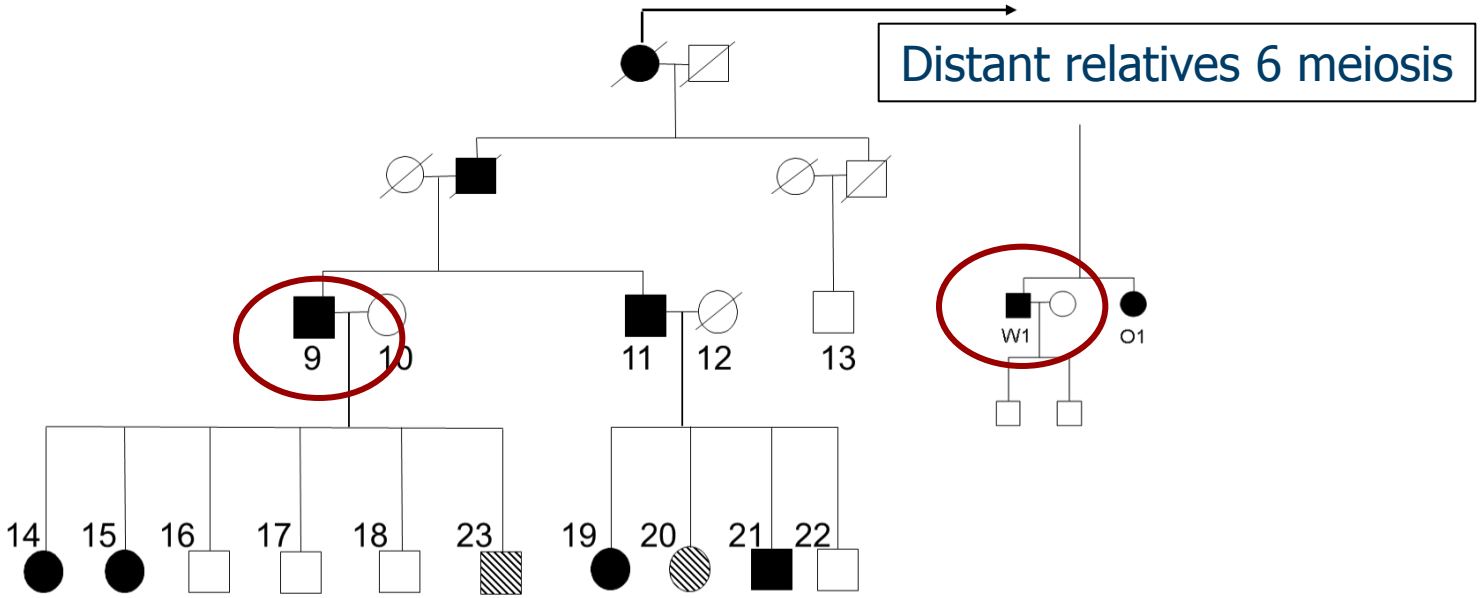
Prioritization scheme

	FOA1	FOA2
Numbers of variants identified	57018	60562
Exonic (including 5'utr and 3'utr) and canonical splice sites	15130	15183
Missense, nonsense, readthrough, frame error, splice site	7735	7699
Novel	618	636
Missense damaging, nonsense, readthrough, frame error, splice sites	168	176

Each individual carriers around > 100 novel, rare, damaging (Sift) mutations.

You need additional ways to identify causal gene

Exome sequencing
An extended early onset families



If variant is causal then:

- Distant affected relatives should share novel, missense probably damaging variant.
- Absent in population at large.

Prioritization scheme; novelty, distant family members

	FOA1	FOA2
Numbers of variants identified	57018	60562
Exonic (including 5'utr and 3'utr) and canonical splice sites	15130	15183
Missense, nonsense, readthrough, frame error, splice site	7735	7699
Novel	618	636
Missense damaging, nonsense, readthrough, frame error, splice sites	168	176
Not present in in-house whole genome sequence data (N=472)	110	115
Not present in in-house exome sequence data (N=43)	102	105
Shared among distant family members	6	6

What are the details of these 6 variants?

Exome sequencing
Details of detected variants

CHR	BP	SNP_Type	base_change	Codons	Substitution	gene_name	DEFINITION
1	13,036,596	missense	T/C	ATA-AcA	I223T	PRAMEF22	preferentially expressed antigen in melanoma
8	119,936,614	**readthrough**	T/A	'TAA=>T	'*=>L'	TNFRSF11B	tumor necrosis factor receptor superfamily, member 11b
11	17,191,207	missense	T/C	AAA-gAA	K28E	PIK3C2A	phosphoinositide-3-kinase, class 2, alpha polypeptide
12	71,523,146	missense	C/A	GTT-tTT	V209F	TSPAN8	tetraspanin 8
12	124,281,300	missense	C/A	TCT-TaT	S577Y	DNAH10	dynein, axonemal, heavy chain 10
16	79245514	missense	G/A	AGC-AaC	S197N	WWOX	WW domain containing oxidoreductase

With 6 variants you can:

- Check for linkage within extended family
- Expression gene in disease relevant tissue
- Responsiveness to OA process in tissue

Exome sequencing
Expression, differential expression linkage

CHR	BP	gene_name	Cartilage expression	Differential Expression (FDR)	positional_FOA	Erik_genes
1	13,036,596	PRAMEF22				
8	**119,936,614**	**TNFRSF11B**	**10.9**	**2.02E-05**	**1**	
11	17,191,207	PIK3C2A	7.0	8.30E-01		
12	71,523,146	TSPAN8	7.2	9.94E-01		
12	124,281,300	DNAH10	7.1	7.05E-01		
16	79,245,514	WWOX	6.8	9.77E-01		1.00

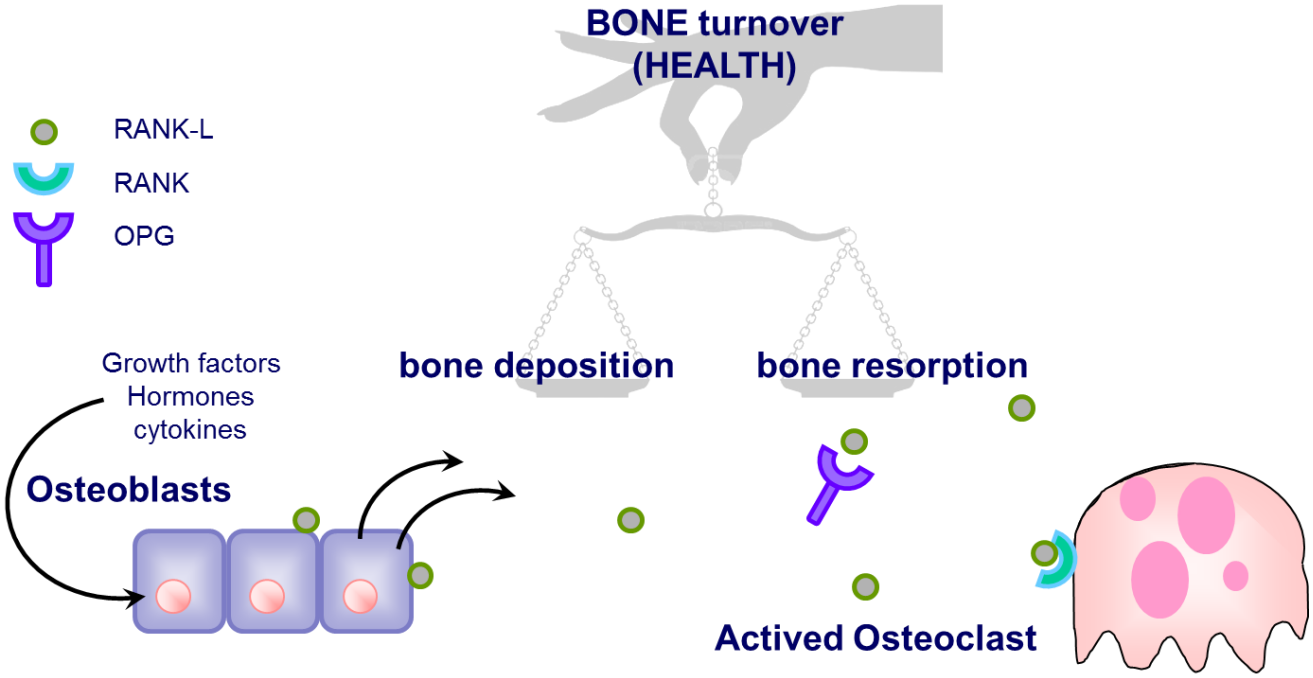
Expression
- Micro-array mRNA expression dataset generated on Illumina V3 Human-12 chips
- Human matched OA and preserved articular cartilage from N=33 subjects.
- Average expression of the genes in the dataset (7.4 with range 6,6-14,9).

Positional FOA
Maximal NPL-score D8S1132 LOD-score 3.47; Family 1
D8S1132 = 107,228,821
TNFRSF11B = 119,936,614
D8S1179 = 126,007,405

Readthrough mutation:
19 additional amino acids

TNFRSF11B encoding Osteoprotegerin (OPG)!

Mutation in *TNFRSF11B* encoding osteoprotegerin (OPG)



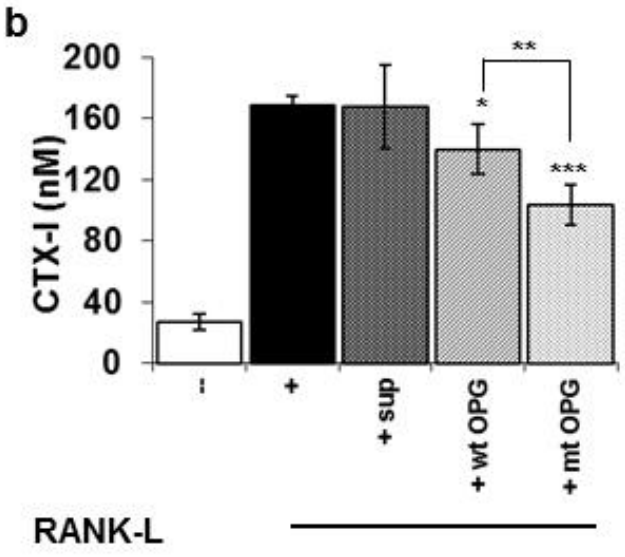
Unfavourable (subchondral) bone balance likely causal to OA in family. What is effect of mutation?

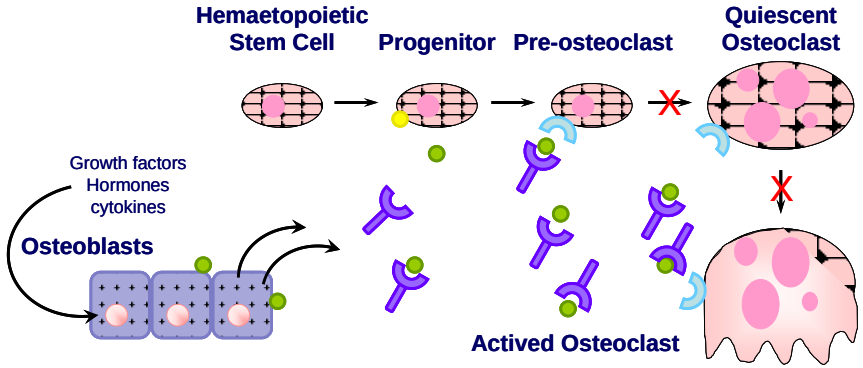
Functional genomics on mutation

Gain of function:

- ➔ more binding to RANK-L
- ➔ more inhibition of osteoclastogenesis
- ➔ less bone resorption
- ➔ more bone

⟶ less CTX-I





Gain of function of *TNFRSF11B* encoding OPG

Ramos et al. (2015);74(9):1756-62

Take home messages

- Most efficient prioritization based on novelty, predicted effect, distant relatives.
- Each individuals has large number of protein damaging mutations without obvious phenotype.
- Functional studies are necessary to determine effect mutation, linkage in family, expression in disease relevant tissuetissue and .

Familial OA, specific & penetrant phenotype

Screen genome by next generation sequencing

Identification causal genes, generalisability to age related OA

Leiden University
Medical Center

Introduction to the practical

Exome sequencing analysis of an early onset osteoarthritis patient

Ingrid Meulenbelt

Molecular Epidemiology

LEIDEN, THE NETHERLANDS

Exercise: Finding the causal variant

Aim:

Find mechanisms of the disease process by identification of the underlying genes and/or pathways

Practical aspects:

Biological insight into etiology may lead to new druggable targets and early diagnosis

Exercise: Finding the causal variant

About 15 years ago we included early onset OA families:

- Symptomatic OA ≥2 joints
- Age onset ~20
- Dominant Mendelian inheritance pattern

Linkage-analysis was performed, but causal gene was not identified.
New techniques, new analysis: **Next Generation Exome Sequencing**

In this exercise, by using the data that were generated you will be searching for the mutation most likely responsible for development of OA in one of the families.

Exercise: Finding the causal variant

Integration of multiple analyses:

- Whole Exome Sequencing
- Gene expression in disease relevant tissue (cartilage)
- Online databases (e.g. Human Gene Mutation Database)
- Co-segregation between mutant and phenotype within family (LINKAGE)