



NEXT GENERATION SEQUENCING OF MULTIPLE GENES IN PARALLEL IN GENETICALLY HETEROGENEOUS DISORDERS AFTER ENRICHMENT

Hans Scheffer

SEQUENCE CAPTURE IN HETEROGENEOUS MONOGENIC DISORDERS



Disease	# (known) Major Genes
Breast cancer	2
Recessive ataxias	>20
Blindness	>200
Mental retardation	~1000



CHALLENGES



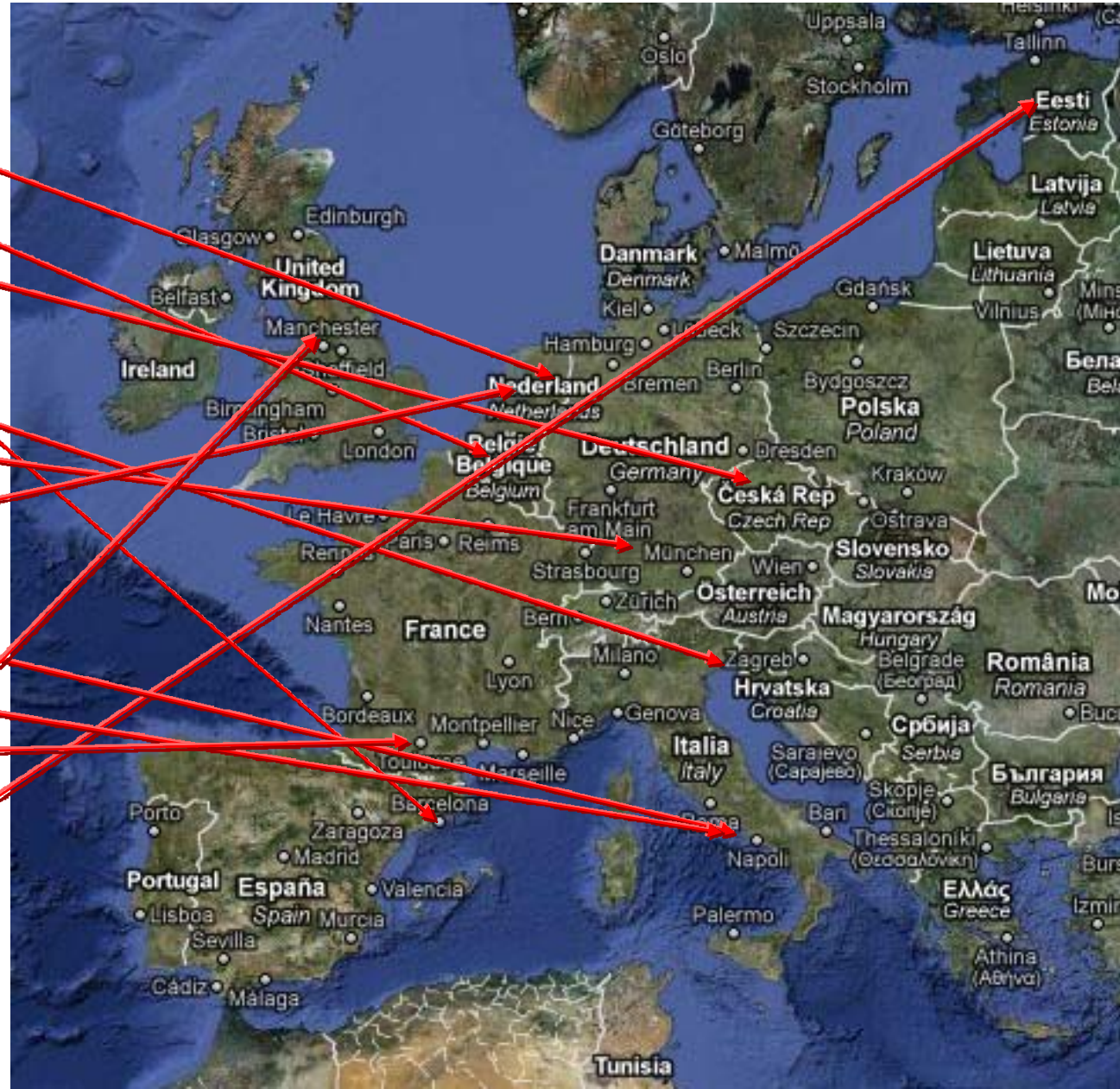
- **Sequencing of multiple genes in parallel for heterogeneous hereditary disorders**
- **Multiplexing of samples for sequencing to increase throughput and reduce costs**
- **Automated data-analysis, including**
- **Interpretation of potential pathogenicity of identified variants**

- **Is a network of ISO15189 accredited clinical molecular genetic laboratories and renowned genome research laboratories**
- **Has been positioned as a technological innovation platform connected with the EUGT Network of Excellence dealing with Quality Standards in Clinical (Laboratory) Genetics**

TECHGENE - WHO ARE WE?



PARTNER
Scheffer/Veltman
Matthijs/Cuppens
Macek
Estivill
Gasparini
Riess
Bakker/Den Dunnen
Banfi
Nigro
Cambon-Thomsen
Payne
Sak



PLATFORMS



COMPANY	TECHNOLOGY	CURRENT READ LENGTH	CURRENT THROUGHPUT
Roche/454 (GS FLX Titanium)	Sequencing by synthesis	300-500 bases	500 Mb/10h
Illumina (Solexa) (Illumina Genome Analyzer System)	Sequencing by synthesis	35 bases	4.5-6.0 Gb/2 days
Life Technologies (SOLiD3)	Sequencing by ligation	50 bases	30 Gb/5 days



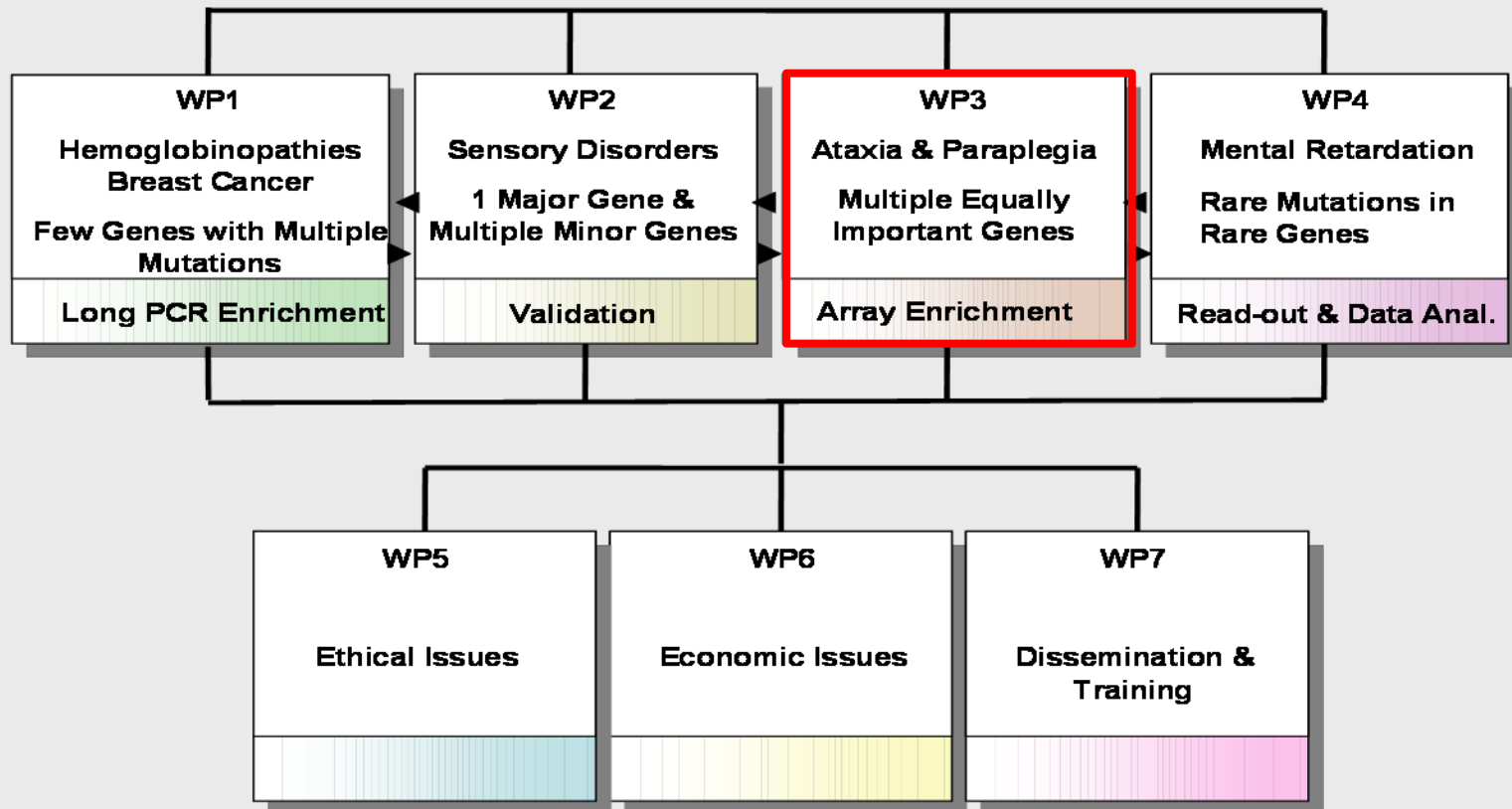
600 million
map-able 50bp reads
→ 30Gb

1 million
map-able 500bp reads
→ 500Mb

TECHGENE - HOW?

Overview of Plan of Investigation

WP8 Management of TECHGENE



→ Next Generation Sequencing in Diagnostics

THE MAIN CHALLENGE: INTERPRETATION!



- Step 1: Identify all potential genetic variation (SNPs,CNVs)
- Step 2: Exclude false positives

Data of first whole genome sequencing results look great but.....

0,1% FP = 6 million bp!

1 mutation may cause disease!

Asian Genome
Wang. Nature 2008

- Step 3: Link true genetic variants to phenotype

**Whole genome sequencing is not
suitable for diagnostics yet**

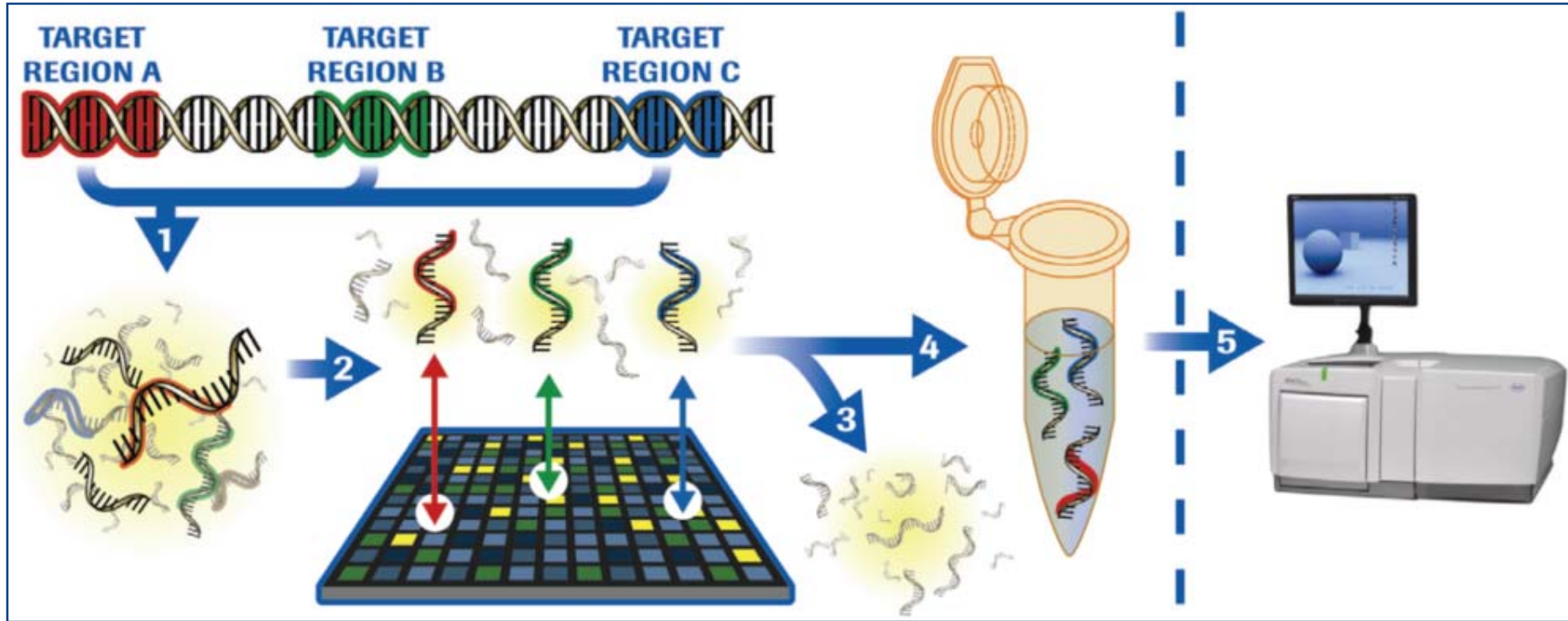
Enrichment Strategies Including Sequence Capture

ENRICHMENT STRATEGIES



- **Long range PCR**
- **Amplicon sequencing**
- **Array-based gene capture (e.g. Nimblegen)**
- **Gene capture in solution (e.g. Agilent SureSelect)**
- **Alternative approaches (e.g. Raindance)**

ARRAY-BASED SEQUENCE CAPTURE



1. DNA preparation (adaptor ligation)
2. Hybridization
3. Stringent washing
4. Elution & ligation mediated (LM)-PCR
5. Sequencing

SEQUENCE CAPTURE WORKFLOW

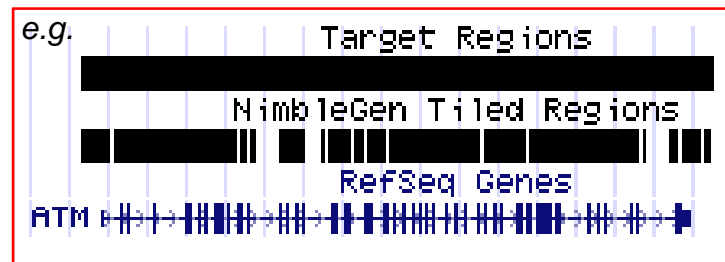


- Select your targets

7 genes:

*APT*X, *SYNE1*, *TDP1*,
SACS, *SETX*, *ATM*, *FXN*

- all exons & introns
- 886kb



→ Targeted region

→ Tiled region, i.e.
represented on the array

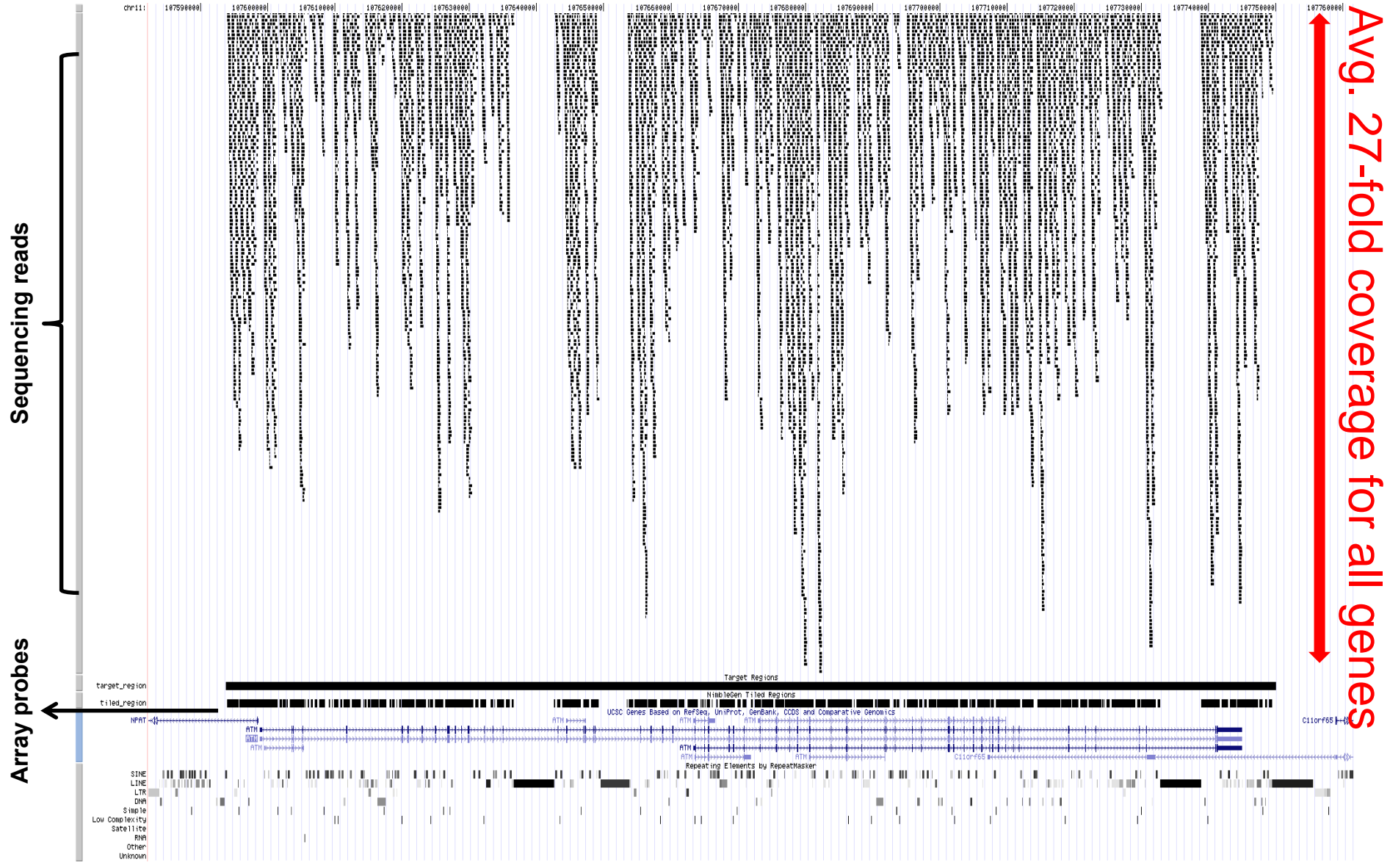
- Enrichment of 10 samples
- 8/10 samples carried known disease causing mutations (10 mutation alleles) for **AR ataxia** (severe movement disorder)
- Roche 454 GS FLX Titanium Sequencing (1/4 run each)
>65Mb map-able sequence

AVERAGE MAPPING DATA

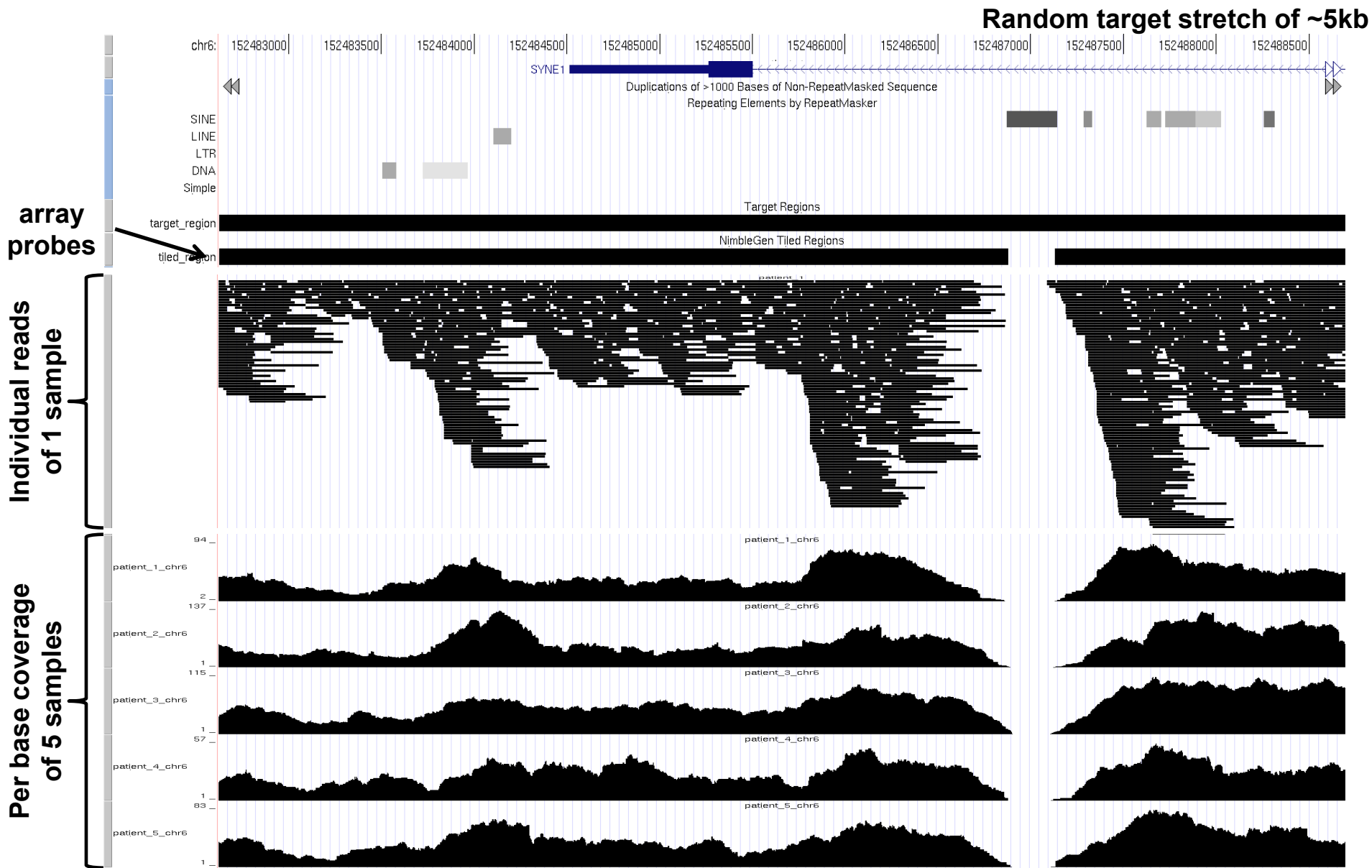


Target bases covered	98.05%
Mean sequence depth in target regions	27 fold
Mean sequence depth in coding regions	31 fold
Mean sequence depth in non-coding regions	24 fold
Target region bases not covered	1.95%

ENRICHMENT WORKS !



ENRICHMENT AND SEQUENCING IS REPRODUCIBLE !



ARE THERE REGIONS NOT COVERED ?

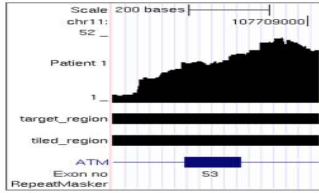


Only two exons (exon 1 of FXN, and 2nd coding exon of SACS) are not covered at all, due to very high CG content (~65%)

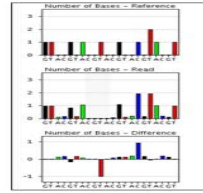
ALL MUTATIONS IDENTIFIED



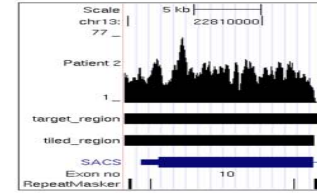
Patient 1: ATM, c.del7875-7876insGC, homozygous



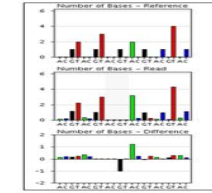
chr11	107709795	Sequence
cont19		T - G - A T - C - C - T T - A T
READ 01		T - G - A - C - C - C - T T - A T
READ 02		T - G - A - C - C - C - C T - A T
READ 03		T - C - A - C - C - C - C T T - A T
READ 04		T - C - A - C - C - C - C T T - A T
READ 05		T - C - A - C - C - C - C T T - A T
READ 06		T - C - A - C - C - C - C T T - A T
READ 07		T - C - A - C - C - C - C T T - A T
READ 08		T - C - A - C - C - C - C T T - A T
READ 09		T - C - A - C - C - C - C T T - A T
READ 10		T - C - A - C - C - C - C T T - A T
READ 11		T - C - A - C - C - C - C T T - A T
READ 12		T - C - A - C - C - C - C T T - A T
READ 13		T - C - A - C - C - C - C T T - A T
READ 14		T - C - A - C - C - C - C T T - A T
READ 15		T - C - A - C - C - C - C T T - A T
READ 16		T - C - A - C - C - C - C T T - A T
READ 17		T - C - A - C - C - C - C T T - A T
READ 18		T - C - A - C - C - C - C T T - A T
READ 19		T - C - A - C - C - C - C T T - A T
READ 20		T - C - A - C - C - C - C T T - A T
READ 21		T - C - A - C - C - C - C T T - A T
READ 22		T - C - A - C - C - C - C T T - A T



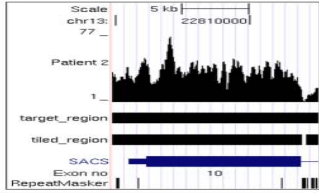
Patient 2: SACS, c.12160C>T



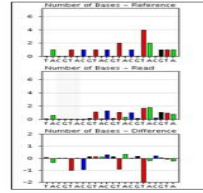
chr19	22809855	Sequence
cont19		T C T T T C - A A - G C - T T
READ 01		T C T T T - A A G - T T
READ 02		T C T T T - A A G C - T T
READ 03		T C T T T - A A A G C - T T
READ 04		T C T T T - A A - G C - T T
READ 05		T C T T T T C - A A - G C - T T
READ 06		T C T T T C - A A - G C - T T
READ 07		T C T T T - A A A G C - T T
READ 08		T C T T T - A A A G C - T T
READ 09		T C T T T C - A A - G C - T T
READ 10		T C T T T - A A A G C - T T
READ 11		T C T T T - A A A G C - T T
READ 12		T C T T T - A A A G C - T T
READ 13		T C T T T - A A A G C - T T
READ 14		T C T T T - A A A G C - T T
READ 15		T C T T T - A A A G C - T T
READ 16		T C T T T C - A A - G C - T T
READ 17		T C T T T C - A A - G C - T T
READ 18		T C T T T C - A A - G C - T T
READ 19		T C T T T - A A A G C - T T
READ 20		T C T T T C - A A - G C - T T
READ 21		T C T T T - A A A G C - T T
READ 22		T C T T T - A A A G C - T T



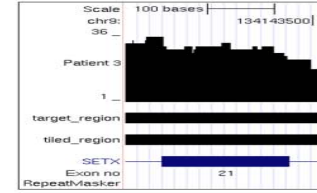
Patient 2: SACS, c.5998-6002del



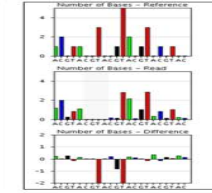
chr13	22810033	Sequence
cont19		T T - C T T T T A A - C T A T
READ 01		T T - C T T T T A A - C T A T
READ 02		T T - C T T T T A A - C T A T
READ 03		T T - C T T T T A A - C T A T
READ 04		T T - C T T T T A A - C T A T
READ 05		T T - C T T T T A A - C T A T
READ 06		T T - C T T T T A A - C T A T
READ 07		T T - C T T T T A A - C T A T
READ 08		T T - C T T T T A A - C T A T
READ 09		T T - C T T T T A A - C T A T
READ 10		T T - C T T T T A A - C T A T
READ 11		T T - C T T T T A A - C T A T
READ 12		T T - C T T T T A A - C T A T
READ 13		T T - C T T T T A A - C T A T
READ 14		T T - C T T T T A A - C T A T
READ 15		T T - C T T T T A A - C T A T
READ 16		T T - C T T T T A A - C T A T
READ 17		T T - C T T T T A A - C T A T
READ 18		T T - C T T T T A A - C T A T
READ 19		T T - C T T T T A A - C T A T
READ 20		T T - C T T T T A A - C T A T
READ 21		T T - C T T T T A A - C T A T
READ 22		T T - C T T T T A A - C T A T



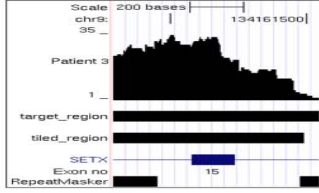
Patient 3: SETX, c.6831-6836del



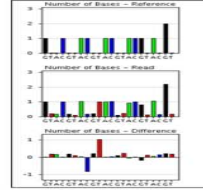
chr9	134143285	Sequence
cont19		T A T T T C T T T T T T A A - G
READ 01		T A T T T C T T T T T T A A - G
READ 02		T A T T T C T T T T T T A A - G
READ 03		T A T T T C T T T T T T A A - G
READ 04		T A T T T C T T T T T T A A - G
READ 05		T A T T T C T T T T T T A A - G
READ 06		T A T T T C T T T T T T A A - G
READ 07		T A T T T C T T T T T T A A - G
READ 08		T A T T T C T T T T T T A A - G
READ 09		T A T T T C T T T T T T A A - G
READ 10		T A T T T C T T T T T T A A - G
READ 11		T A T T T C T T T T T T A A - G
READ 12		T A T T T C T T T T T T A A - G
READ 13		T A T T T C T T T T T T A A - G
READ 14		T A T T T C T T T T T T A A - G
READ 15		T A T T T C T T T T T T A A - G
READ 16		T A T T T C T T T T T T A A - G
READ 17		T A T T T C T T T T T T A A - G



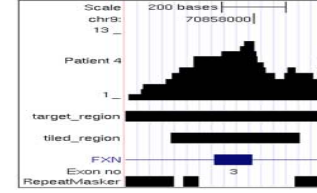
Patient 3: SETX, c.6017C>T



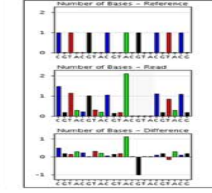
chr9	134161170	Sequence
cont19		A G C T G C A - C A C A C G A
READ 01		A G C T G C A - C A C A C G A
READ 02		A G C T G C A - C A C A C G A
READ 03		A G C T G C A - C A C A C G A
READ 04		A G C T G C A - C A C A C G A
READ 05		A G C T G C A T - A C A C G A
READ 06		A G C T G C A - C A C A C G A
READ 07		A G C T G C A - C A C A C G A
READ 08		A G C T G C A - C A C A C G A
READ 09		A G C T G C A T - A C A C G A
READ 10		A G C T G C A - C A C A C G A
READ 11		A G C T G C A T - A C A C G A
READ 12		A G C T G C A - C A C A C G A
READ 13		A G C T G C A - C A C A C G A
READ 14		A G C T G C A T - A C A C G A
READ 15		A G C T G C A - C A C A C G A
READ 16		A G C T G C A - C A C A C G A
READ 17		A G C T G C A T - A C A C G A
READ 18		A G C T G C A - C A C A C G A
READ 19		A G C T G C A T - A C A C G A
READ 20		A G C T G C A - C A C A C G A
READ 21		A G C T G C A T - A C A C G A
READ 22		A G C T G C A - C A C A C G A



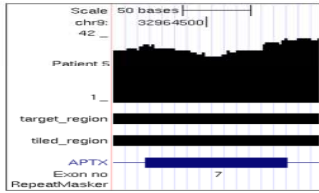
Patient 4: FXN, c.264-1G>A



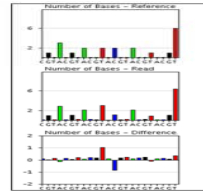
chr9	70857875	Sequence
cont19		C T C T C - A - G C T C T C
READ 01		C T C T C - A - G C T C T C
READ 02		C T C T C - A - G C T C T C
READ 03		C T C T C - A - G C T C T C
READ 04		C T C T C - A - G C T C T C
READ 05		C T C T C - A - G C T C T C
READ 06		C T C T C - A - G C T C T C
READ 07		C T C T C - A - G C T C T C
READ 08		C T C T C - A - G C T C T C
READ 09		C T C T C - A - G C T C T C



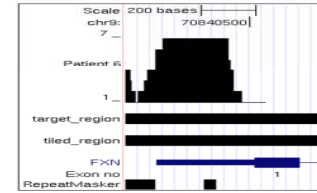
Patient 5: APTX, c.837C>T homozygous



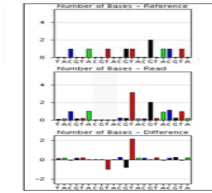
chr9	32964495	Sequence
cont19		G - A A - T T C - C A A - T
READ 01		G - A A - T T T C - A A - T
READ 02		G - A A - T T T C - A A - T
READ 03		G - A A - T T T C - A A - T
READ 04		G - A A - T T T C - A A - T
READ 05		G - A A - T T T C - A A - T
READ 06		G - A A - T T T C - A A - T
READ 07		G - A A - T T T C - A A - T
READ 08		G - A A - T T T C - A A - T
READ 09		G - A A - T T T C - A A - T
READ 10		G - A A - T T T C - A A - T
READ 11		G - A A - T T T C - A A - T
READ 12		G - A A - T T T C - A A - T
READ 13		G - A A - T T T C - A A - T
READ 14		G - A A - T T T C - A A - T
READ 15		G - A A - T T T C - A A - T
READ 16		G - A A - T T T C - A A - T
READ 17		G - A A - T T T C - A A - T
READ 18		G - A A - T T T C - A A - T
READ 19		G - A A - T T T C - A A - T
READ 20		G - A A - T T T C - A A - T
READ 21		G - A A - T T T C - A A - T
READ 22		G - A A - T T T C - A A - T



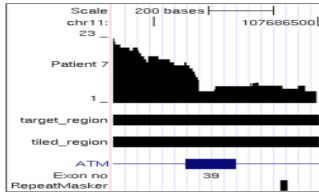
Patient 6: FXN, c.3G>T



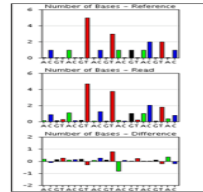
chr9	70840520	Sequence
cont19		G C A G C A T - T G G A C T
READ 01		G C A G C A T T - T G G A C T



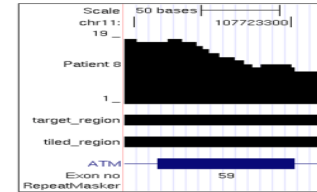
Patient 7: ATM, c.5763-2A>T



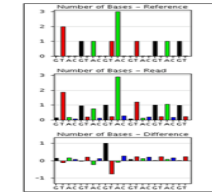
chr11	107686095	Sequence
cont19		T T C T T T - A C - A C C T T
READ 01		T T C T T T - A C - A C C T T
READ 02		T T C T T T - A C - A C C T T
READ 03		T T C T T T - A C - A C C T T
READ 04		T T C T T T - A C - A C C T T
READ 05		T T C T T T - A C - A C C T T
READ 06		T T C T T T - A C - A C C T T
READ 07		T T C T T T - A C - A C C T T
READ 08		T T C T T T - A C - A C C T T
READ 09		T T C T T T - A C - A C C T T
READ 10		T T C T T T - A C - A C C T T
READ 11		T T C T T T - A C - A C C T T
READ 12		T T C T T T - A C - A C C T T
READ 13		T T C T T T - A C - A C C T T
READ 14		T T C T T T - A C - A C C T T
READ 15		T T C T T T - A C - A C C T T



Patient 8 (carrier): ATM, c.8633T>G



chr11	107723265	Sequence
cont19		T C T T C - A - T A A A - T C
READ 01		T C T T C - A - T A A A - T C
READ 02		T C T T C - A - T A A A C T C
READ 03		T C T T C - A - T A A A - T C
READ 04		T C T T C - A G - T A A A - T C
READ 05		T C T T C - A - T A A A - T C
READ 06		T C T T T C A G - T A A A - T C
READ 07		T C T T C - A - T A A A - T C
READ 08		T C T T C - A - T A A A - T C
READ 09		T C T T C - A - T A A A - T C
READ 10		T C T T C - A - T A A A - T C
READ 11		T C T T C - A C - T A A A - T C
READ 12		T C T T C - A - T A A A - T C
READ 13		T C T T C - A C - A A A - T C
READ 14		T C T T C - A G - A A A - T C



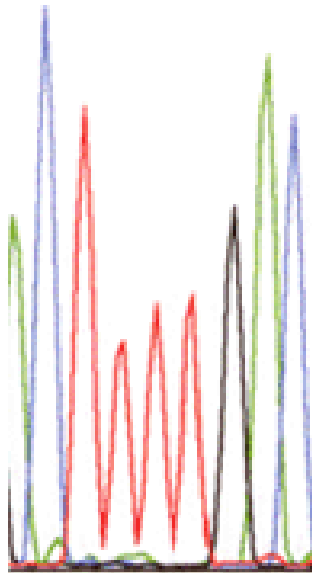
CAN WE DETECT PATHOGENIC MUTATIONS?



Wild-type



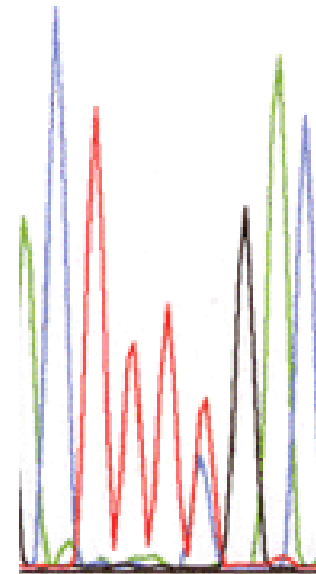
A C T T T T G A C
110



Point
Mutation



↓
A C T T T N G A C
110

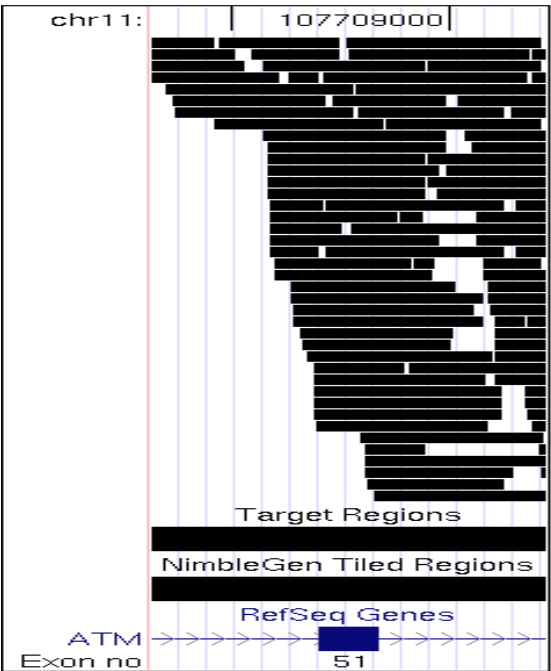
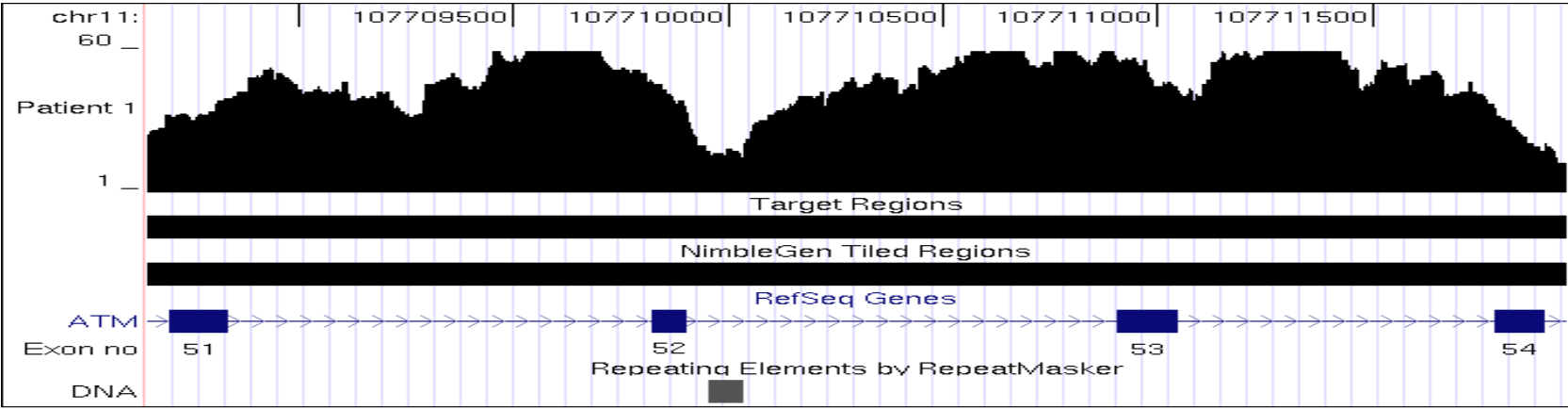


PRIORITIZATION HIERARCHIES FOR THE DISEASE CAUSING MUTATIONS



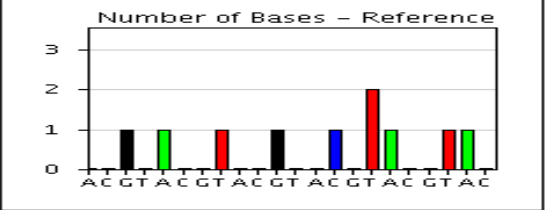
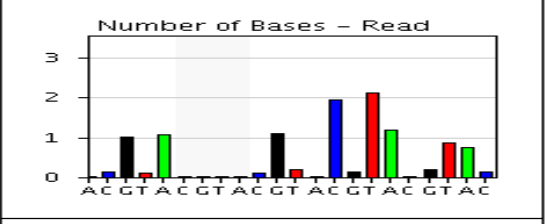
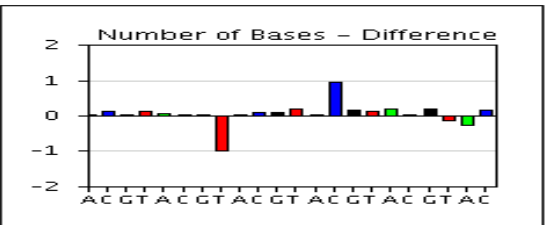
- Mutations that could give rise to premature protein truncating mutations *i.e.* stop mutations, variants at potential splice sites and exonic deletions/insertions potentially resulting in frame shifts
 - All exonic missense variants not known in dbSNP with at least 20% of variant reads. (This reduced the number of variants to an average of 7 per sample)
 - Ranking these variants based upon evolutionary conservation of the affected nucleotide using vertebrate PhyloP and PhastCons scores
- All known pathogenic mutations were ranked on the top position

HOMOZYGOUS MUTATION IN THE ATM GENE



chr11	107708785																					
contig	GT	-	G	-	A	T	G	-	C	-	T	T	-	A	T	A						
READ 01	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 02	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 03	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 04	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 05	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 06	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 07	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 08	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 09	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 10	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 11	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 12	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 13	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 14	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 15	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 16	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 17	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 18	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 19	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 20	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 21	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 22	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 23	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 24	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 25	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 26	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 27	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 28	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 29	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 30	GT	-	G	-	A	-	G	-	C	-	C	T	T	-	A	T	A					
READ 31	GT	-	G	-	A	-	G	-	C	-	G	C	T	T	-	A	T	A				
READ 32	GT	-	G	-	A	-	G	-	C	-	G	C	T	T	-	A	T	A				
READ 33	GT	-	G	-	A	-	G	-	C	-	A	-	G	-	C	-	T	T	-	A	T	A

TG → GC



MUTATION IDENTIFICATION AGAINST A BACKGROUND OF APPR. 800 HIGH CONFIDENTIALITY (HC) DIFFERENCES



AT Patient: c.7875-7576 TG>GC (homozygous mutation *ATM* gene)

Roche Mapping software (HCdifferences) output:

- * exclude variants outside coding regions
- * exclude known SNPs (dbSNP)
- * sort for conservation (phastCons & phyloP)

chromosome	start	stop	reference	mutation	reads	variation reads	% variation	ref amino acid	mutation amino acid	coding frame	gene name	phast Cons	phyloP
chr11	107,708,785	107,708,786	TG	GC	22	22	100	DA	EP	1	ATM	1	3.8459
chr13	22,812,629	22,812,629	T	C	21	2	14	K	R	-2	SACS	1	1.7722
chr6	152,735,877	152,735,877	T	C	12	6	50	E	E	-1	SYNE1	1	1.1568
chr14	63,491,288	63,491,288	T	A	28	3	11	M	K	3	SYNE2	0.992	4.4954
chr6	152,699,834	152,699,835	CT	AC	15	9	60	K	S	-1	SYNE1	0.992	1.9699

→ Top 1 on the list is the known mutation!

HETEROZYGOUS MUTATION IN THE SACS GENE



```
chr13|51511729|ref... AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
contig00082 AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04YFR55 AA
FP1U9CL02J18R5 AA--C
FP1U9CL02JXKR0 A---CC
FP1U9CL04Y2AT1 AA--CCAA-C-A-T
FP1U9CL04XY85Z AA--CCAA-C-A-T
FP1U9CL02J753C AA--CCAA-C-A-T-CT-CTT-A
FP1U9CL03RUE3M AA--CCAA-C-A-T-CT-CTT-A
FP1U9CL02I7MCD AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04YF93C AA--C-A--C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAA--TCTT
FP1U9CL02JZXWY AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL02ILXER AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL02IGM1E AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04Y2EFG AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04XJDW4 AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL02IB4YM AA--CCAA-C-A-T-CT-CTT-G-----A-G-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03Q4VPA AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04Y9M0Q AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03RK0U4 AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03RR37I A---CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL02KCV4C AA--CCA--C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL02JZW4Y AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL02JBR7G AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL02J0R44 AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03PWYCT AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL02JNWSH AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03Q89YX AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04X1EGX AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03P0009 AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04Y759L AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04YTF82 AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03RCLGU AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03QK5W1 AA--CCAA-C-A-T-CT-CTT-A-----AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL02I2FFU AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAA--TCTT
FP1U9CL04X5FHU AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03Q6IFQ AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04YGP3L AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03RM1H0 AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03Q0A53 AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAA--TCTT
FP1U9CL03RI3UC AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04XE8F2 AA--CCAA-C-A-T-CT-CTT-G-----A-G-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04YD5D5 AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04YGK4X AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04XS5CYP AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04XS5JTQ AA--CCAA-C-A-T-CT-CTT--AAG-GTA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03R3P16 AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03QMRMB AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04YBX3G AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT-G-AGAAA--TCTT
FP1U9CL02IJNLN AA--CCAA-C-A-T-CT-CTT--AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL03P2L6K AA--CCAA-C-A-T-CT-CTT-A-----T-TTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL02JLII7 AA--CCAA-C-A-T-CT-CTT-A-----T-TTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
FP1U9CL04XTT16 AA--CCAA-C-A-T-CT-CTT--CTTTT-AAG-TA-TAGAGT-C-ATCT--AGAAA--TCTT
```

ARSACS patient:

A 5 bp heterozygous CTTT deletion in exon 5 of the SACS gene

MUTATION IDENTIFICATION AGAINST A BACKGROUND OF APPR. 800 HIGH CONFIDENTIALITY (HC) DIFFERENCES



ARSACS patient: compound heterozygous for SACS c.5559-5563 delCTTTT / c. 11719C>T (p.Q3907X)

Roche Mapping software (HC differences) output:

- * exclude variants outside coding regions
 - * exclude known SNPs (dbSNP)
 - * sort for conservation (phastCons & phyloP)
- Top 2 on the list are the known mutations!

chromosome	start	stop	reference	mutation	reads	variation reads	% variation	ref amino acid	mutation amino acid	coding frame	gene name	phast Cons	phyloP
chr13	22,803,855	22,803,855	G	A	35	21	60	Q	*	-2	SACS	1	6.5854
chr13	22,810,013	22,810,017	CTTTT	-	27	17	63			-2	SACS	1	5.3822
chr17	42,576,285	42,576,285	C	T	4	4	100	S	N	-3	CDC27	1	5.2978
chr14	63,498,047	63,498,047	A	G	30	3	10	Q	R	2	SYNE2	1	5.1516
chr14	63,695,793	63,695,793	A	G	25	3	12	D	G	1	SYNE2	1	3.4163
chr14	63,700,048	63,700,048	T	A	28	3	11	F	Y	1	SYNE2	1	3.3456

PROOF OF CONCEPT



- combination Roche 454 GS FLX Titanium and NimbleGen array works
- 7 disease genes
- Enrichment works, ~80% on or near target
- Mean coverage: 24x
- All known mutations were identified
- Detection of the disease causing mutation depends on the coverage
- Coverage: depending on the size of the targeted region, GC content, repeat elements
- At least 15x coverage
- Prioritization due to conservation

METHODS

Human Mutation

OFFICIAL JOURNAL



Massively Parallel Sequencing of Ataxia Genes after Array-Based Enrichment

Alexander Hoischen,^{1*}† Christian Gilissen,^{1†} Peer Arts,¹ Nienke Wieskamp,¹ Walter van der Vliet,¹ Sascha Vermeer,¹ Marloes Steehouwer,¹ Petra de Vries,¹ Rowdy Meijer,¹ Jorge Seiquerios,² Nine V.A.M. Knoers,¹ Michael F. Buckley,¹ Hans Scheffer,¹ and Joris A. Veltman¹

¹Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; ²UnIGENE and CGPP, IBMC—Instituto de Biologia Molecular e Celular and ICBAS, University Porto, Porto, Portugal

Communicated by Graham R. Taylor

Received 16 October 2009; accepted revised manuscript 28 January 2010.

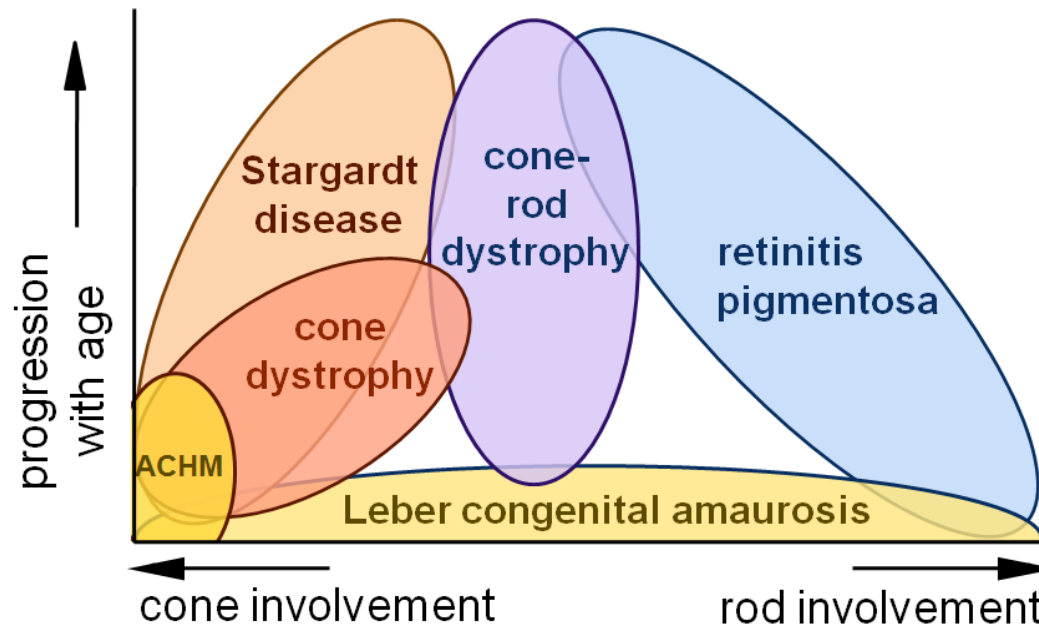
Published online 11 February 2010 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/humu.21221

NEXT PHASE



- **Design of seq cap tools for parallel sequencing of appr. 100 genes, in hereditary blindness, in hereditary movement disorders, in mitochondrial disorders**
- **Validation of amplicon sequencing-based NGS approach for BRCA1/2**

GENE CAPTURE FOR HEREDITARY BLINDNESS DIAGNOSTICS?

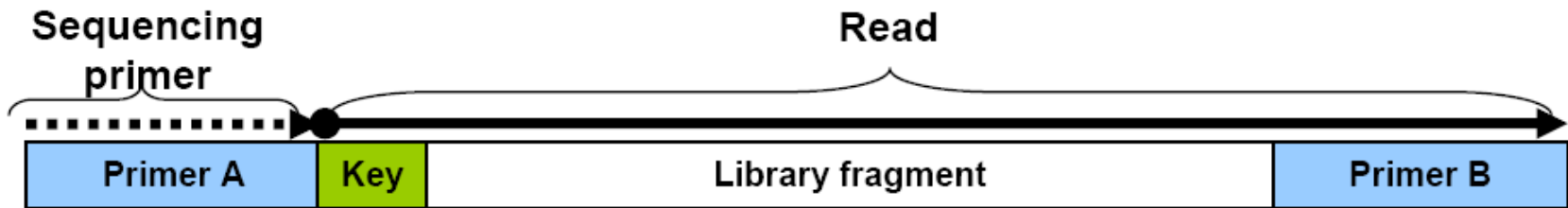


- clinically and genetically heterogeneous
- focus on RP and LCA
- mutations described for >100 genes
- only known mutations detectable on microarrays from Asper
- diagnostic yield by current Sanger sequencing-based strategy is very low!

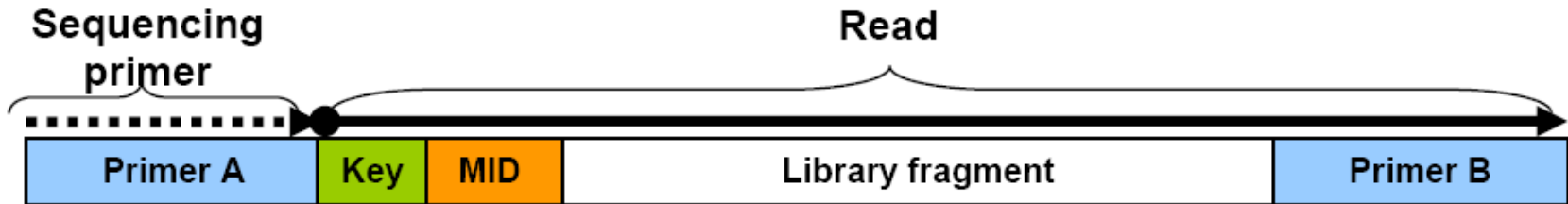
BARCODING/MULTIPLEXING

(12 samples)

a)



b)



MID = multiplex identifiers

DATA ANALYSIS/INTERPRETATION OF SAMPLES WITH UNKNOWN MUTATIONS?



- Design makes data analysis challenging
- Many variations found
- **Every** gene is a candidate gene
- **Every** variation potentially pathogenic
- Several known blindness mutations are in dbSNP
- Include known blindness-causing mutations (e.g. HGMD) in analysis pipeline

Data analysis/interpretation major bottleneck
(taking into account also possibility of digenic
inheritance/modifier genes...)



CONCLUSIONS

- **NGS will be suitable for diagnostics**
- **We need good coverage!**
- **Increase throughput (e.g. multiplexing)**
- **Decrease costs**
- **Robust handling for technicians in routine diagnostics**
- **Possibility for automation**
- **Careful data analysis/interpretation (further development of pathogenic mutation identification pipeline)**

We can distinguish known mutations from 800-1000 variants; sequence capture almost suitable for diagnostics

**Exome sequencing successfully
applied in gene identification
(e.g. Schinzel-Giedion*)**

-

**Soon also as diagnostic test (e.g.
for very heterogeneous disorders
caused by *de novo* mutations)**

*** Hoischen et al. (2010) De novo mutations of SETBP1 cause Schinzel-Giedion syndrome.
Nat Genet 42(6):483-5**

Future perspectives in molecular genetics



Research:



Diagnostics:



**Diagnostic testing first, then
detailed clinical investigation?**



NGS DIAGNOSTICS KNOWLEDGE NETWORK

More info at: www.techgene.eu or www.techgene.org

Register at: info@techgene.org



ACKNOWLEDGEMENTS

Alexander Hoischen

Kornelia Neveling

Peer Arts

Christian Gilissen

Nienke Wieskamp

Joep de Ligt

Michael Buckley

Joris Veltman

Hans Scheffer

Sascha Vermeer

Nine van Slobbe-Knoers

Berry Kremer

Jorge Sequeiros



www.techgene.eu

www.techgene.org