WHOLE EXOME SEQUENCING

Final Degree Project

Francesc Bou De Pieri, Genetics Degree, UAB

UAB Universitat Autònoma de Barcelona

BACKGROUD

1st. generation sequencing

Irreversible termination

Advantages	Disadvantages
Long reads Excellent accuracy High quality	High cost Limited throughput Time consuming Gel-based size-fractionation

Ex: Sanger

2nd. generation sequencing

Synthesis, ligation, hybridization

Advantages	Disadvantages
High throughput Increased speed Produce over 100 more data Reduced cost Fewer preparatory steps No gel-based size-fractionation	Short reads PCR needed Higher error rates High redundancy required Computational demands Comprehensive data analysis

Ex: 454, HiSeq, SOLiD

3rd. generation sequencing

Single DNA molecule

Advantages	Disadvantages
Less initial DNA No PCR Better accuracy	Lower throughput Short reads

Ex: Heliscope, SMRT, Nanopore

Sequencing platforms

	1st	Next generation			
		2nd			3rd
	Sanger	454	HiSeq	SOLiD	Heliscope
Mechanism	Dideoxy termination	Pyrosequencing	Seq. by synthesis	Seq. by ligation	Single molecule seq.
Read length	400-900bp	400-600bp	100-200bp	50-100bp	25-55bp
Accuracy	99.999%	99.5%	98.5%	99.94%	>99%
Output data/run	1.9-84Kb	0.7Gb	600Gb	120Gb	50Gb
Time/run	20min-3h	24h	3-10 days	7-14 days	8 days
Cost/Mb	S 2400	S 10	S 0.07	S 0.13	>2nd. gen.
+	High quality Long reads	Long reads Fast	High throughput	High accuracy (twice)	Easy and fast preparation
-	High cost Low throughput	High cost Low throughput	Short reads	Short reads	Short reads Low throughput

Exome sequencing

- Targeted sequencing enrichment focused on all protein-coding subsequences.
- Hybrid capture (microarray), solution capture (proves).
- Cost-effective, reproducible and robust strategy for the identification of variants causing protein-coding changes in individual human genomes.

Neural tube defects

- An opening in the spinal cord or brain that occurs very early in human development as a result of the failure in neural tube closure (1/1,000 live births).
- Multifactorial inheritance.
- Complex disease model.

OBJECTIVES

1. Carry out a comparative analysis between genome sequencing and whole exome sequencing techniques.
2. Demonstrate the importance and the applications of whole exome sequencing.
3. Use whole exome sequencing as an approach to identify genetic factors causing complex diseases, such as neural tube defects.

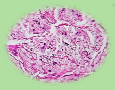
MATERIALS AND METHODS

Bibliographic research 1,2

Papers published in PubMed.
Words used: "Next Generation Sequencing", "Whole Genome Sequencing", "Whole Exome Sequencing", "Comparison" and "Application".

Samples 3

Central nervous system tissue from diagnosed fetuses.
- Normal karyotype by G-banding
- No CNVs clearly linked to the phenotype



WES 3

-Targeted capture

SureSelect 50Mb human exome kit (Agilent)
- 635k proves
- High efficiency

-Sequencing

SOLiD 5500 sequencing (Applied Biosystems)
- High accuracy

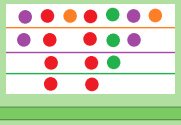
-Mapping

SOLiD LifeScope software (Life Technologies)
- Iterative mapping approach

Human reference genome hg19 (UCSC)

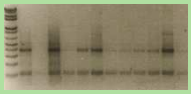
Variant analysis 3

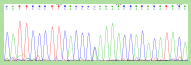
All variants observed should be filtred by:
- Sample recurrence (present in 2 or +)
- Presence of frameshift and nonsense mutations or predicted pathogenicity (Condel)
- Conservation (PhyloP value)



Validation 3

Sanger sequencing of PCR amplicons from genomic DNA:
- Design a pair of primers for each variation
- Do the PCR tuning
- Put the PCR with the samples
- Check the PCRs with agarose gel
- Purify the DNA from PCR
- Do the sequencing PCR and the precipitation of DNA
- Analyze the sequences obtained from Sanger sequencing





RESULTS AND DISCUSSION

Whole genome sequencing (WGS) 1

Advantages	Disadvantages
Wide availability Improved cost efficiency Detects large structural variants Detects coding and non-coding variants Produces 100 times more data	High cost Short read length Low interpretability High computational needs (filtering, storage, software and hardware)

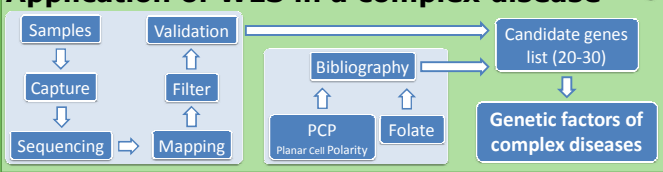
- Results can be very sensitive to false positives and false negatives
- Needs a validation process
- Relies on the controls or the reference genome used to map the reads obtained from the sequencing process

Whole exome sequencing (WES) 1,2

Advantages	Disadvantages
Lower cost Less raw sequence needed Allows bigger sample size Increased sequence coverage Detects coding variants Sensitive and specific identification	Does not detect structural variants Not all targets are captured (80%) Difficult to capture ↑GC% regions Limited view

- Make sequencing feasible for a huge variety of laboratories in terms of costs and infrastructures
- Protein coding genes constitute only 1% of the human genome but harbor 85% of the mutations with large effects on disease traits
- Limited view (coding regions)
- WES coding SNP detection has as sensitivity as WGS
- Efficient approach to identify genes causing mendelian disorders.

Application of WES in a complex disease 3



CONCLUSIONS

- WGS and WES, shows that both approaches can be used to identify genetic factors causing human diseases
- WES has been a very useful and efficient approach to identify the genetic causes of mendelian disorders but has a limited view.
- WES can be also used to identify the genetic causes of multifactorial disorders.

Despite the true advantages that WES offers now, in the future, WGS is predicted to be more economical than WES because the capture process is skipped entirely. Meanwhile, WES is considered the best approach because it provides most of the benefits of WGS but with lower costs and higher clinical interpretability.

BIBLIOGRAPHY

1- Lin, B., Wang, J. & Cheng, Y. **Recent Patents and Advances in the Next-Generation Sequencing Technologies.** *NIH Public Access* 2008, 60–67 (2008).
2- Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A. & Jabado, N. **What can exome sequencing do for you?** *J Med Genet.* 48, 580–589 (2011).
3- Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. a. **Unlocking Mendelian disease using exome sequencing.** *Genome Biol.* 12, 228 (2011).
4- De Marco, P., Merello, E., Mascelli, S. & Capra, V. **Current perspectives on the genetic causes of neural tube defects.** *Neurogenetics.* 7, 201–221 (2006).