Molecular Diagnosis of Genetic Diseases: From 1 Gene to 1000s’

Maria Iascone
USSD Lab Genetica Medica
Ospedali Riuniti
Bergamo
The problem...

What does the clinician/patient want to know?

Genomics medicine/Personalized medicine: clinical care based on genomic information

Diagnosis, Clinical management & therapeutic choice....
2011 vision for the path towards an era of genomic medicine

Nature, 470 2011
Whole-genome vs Targeted Resequencing

- **Size of target**
  - Accuracy of design
  - Amount of generated data

- **Costs**
  - Coverage

- **Interpretation**
  - Bioinformatics requirements
  - Accuracy of data

**Reduction of genomic complexity**

- Whole exome
- Contiguous genomic region (e.g. linkage region)
- Multiple genes in different loci
- Multiple exons of different genes

**Targeted resequencing requires the knowledge of sequence/genomic position of target DNA**
NGS TARGETED RESEQUENCING WORKFLOW

1. **Genomic DNA Sample**
2. **Genomic DNA Fragment Libraries**
3. **Solution Hybridization & Capture**
4. **Sequencing**
5. **Bioinformatic Data Analysis**
Genomic Analysis of Mutations Extracted by Sequencing

- Mismatches calling
- “Cleaning“
  - Phred score quality, PCR duplicate, non-unique reads

- Annotation
  - gene, chromosome coordinates, exons (all isoforms of the genes), and the position of variation (UTR region, intron, exon junctions and exon)
  - GO ontology, Kegg pathway, ...

- Prediction
  - dbSNP (NCBI)
  - Human Gene Mutation Database (HGMD) and locus-specific databases
  - effect of SNP/indel (silent, missense,...)
  - PhyloP conservation score
  - PolyPhen, SIFT, MutationTaster, splicing prediction

Targeted Resequencing & Diagnostic Yield

- Single gene disorders with high diagnostic yield
  - Cystic fibrosis
  - ... (other disorders)

- Homogeneous disorders where >1 of a large number of genes may be implicated
  - Cardiomyopathies
  - Epilepsies
  - XLMR

- Genomic disorders where an appropriate set of genes to test cannot be clinically defined
  - Autism
  - CHD
  - Malformations
  - ... (other disorders)

Maria Iascone
Roma - June 23, 2011
A panel for ...

Epilepsies

106 genes, 5482 exons 0.74 Mb

Cardiomyopathies

98 genes, 3508 exons 0.69 Mb
Epilepsies

- Diverse etiologies and pathophysiology
- Genetic heterogeneity
- Many known genes
- Syndromic vs familial

Classification of epilepsies on the basis of molecular defects could help choose the right treatment for trial design

The study was blinded

Without any "a priori" knowledge of the clinical characteristics of the enrolled patients
1° mutation
ALDH7A1:p.Arg167Ser*

* Bennet et al, Epilepsia 2009
2° mutation
ALDH7A1:c.1405+5G>A

<table>
<thead>
<tr>
<th>SpliceSiteFinder-like</th>
<th>MaxEntScan</th>
<th>NNSPLICE</th>
<th>Human Splicing Finder</th>
<th>WT</th>
<th>Mut</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'</td>
<td>[0-100]</td>
<td>76.6</td>
<td>0.9</td>
<td>8.1</td>
<td>1.8</td>
</tr>
<tr>
<td>3'</td>
<td>[0-100]</td>
<td>77.6</td>
<td>0.4</td>
<td>3.3</td>
<td>3.0</td>
</tr>
<tr>
<td>Branch Points</td>
<td>[0-100]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Reference Sequence

Mutated Sequence
ALDH7A1
aldehyde dehydrogenase 7 family, member A1

- p.Arg167Ser + c.1405+5G>A
- Pyridoxine-dependent epilepsy (PDE) is a rare autosomal recessive disorder causing intractable seizures in neonates and infants
- PDE patients are typically resistant to anti-epileptic treatment but respond to the administration of pyridoxine
This gene encodes an EF-hand-containing calcium binding protein

The encoded protein likely plays a role in calcium homeostasis

Mutations in this gene have been associated with juvenile myoclonic epilepsy
It was shown that hypertrophic, dilated and restrictive cardiomyopathy may be caused by mutations in same genes, in particular in sarcomeric genes.
Hypertrophic Cardiomyopathy (HCM)

M.M., female, 27 years
- Severe obstructive HCM
- LVOT gradient 95 mmHg
- IVS 30 mm
- Family history
- NYHA class IV
HCM is the most common inherited cardiovascular disease (1:500) and the most common cause of sudden death in young <30 yrs.
MYBPC3:Pro330fs
- Novel, absent in 570 control alleles

DSC2:Leu732Val
- Bhuiyan Circ Cardiovasc Genet 2009
- Absent in 570 control alleles

DSG2:Val392Ile
- Syrris Eur Heart J 2007; Bhuiyan Circ Cardiovasc Genet 2009
- Absent in 570 control alleles

**Dominant mutations in genes encoding Desmoglein-2 (DSG2) and Desmocollin-2 (DSC2) have been reported in up to 12% and 5% of ARVD/C patients, respectively!**
NGS in diagnostic settings

- **Accuracy & sensitivity**
  - the false-positive and -negative rates must be kept as low as possible

- **Reproducibility**

- **Speed of data generation & analysis**

- **Link true genetic variants to phenotype**
MYBPC3 Deletion
Exon 28-35
~ 3 kb

1 exon - window
How many variants do we find in targeted genes?

AVERAGE OF 40 HCM + 10 DCM SAMPLES

- Detected variants: 2625
  - Known-SNP variants: 2132

- Unreported Variants: 493
  - Intronic variants: 473
  - Coding variants: 20
    - INDELS: 1
    - MISSENSE: 4
    - SPLICE SITE: 1

...Link true genetic variants to phenotype
Standard interpretation of nucleotide variation

- Confirming the findings by an independent method
- Effect of mutation
- Evolutionary conservation
- Occurrence in normal population
- Family analysis
- ...

Maria Iascone
Roma - June 23, 2011
Work in progress...

- Analysis of myocardial tissues
  - Targeted RNA NGS analysis
  - Expression of mutated protein
  - Electron microscopy
  - ...

- Analysis of normal individuals by cardiomyopathies panel
  - Measurements of pick-up rate of rare benign variants

- ...

Roma - June 23, 2011

Maria Iascone
Targeted Resequencing & Diagnostic Yield

SINGLE GENE DISORDERS WITH HIGH DIAGNOSTIC YIELD

HOMOGENEOUS DISORDERS WHERE >1 OF A LARGE N° OF GENES MAY BE IMPLICATED

GENETIC DISORDERS WHERE AN APPROPRIATE SET OF GENES TO TEST CANNOT BE CLINICALLY DEFINED

Whole Exome NGS
The problem

- Only recently survived children begin to reach the reproductive age
  - Small families

- Most forms occur sporadically
  - Families with clear Mendelian inheritance are scarce and characterized by variable penetrance and a wide range of expressivity
  - Very few cases have been described involving low-penetrant mutations of single genes

Improving the effectiveness of healthcare means that genetic diagnosis HAS TO BE obtained for most diseases in short times thus cutting the diagnostic odyssey of patients
REMARKS

- There is no comprehensive and affordable diagnostic platform for several diseases
  - Substantial delay in the diagnosis

- We are developing different approaches and different projects to cover a large set of disorders
  - Shortening time and costs to diagnosis
  - Therapeutic decisions & clinical trials

We are ready for clinical applications of targeted NGS!
Acknowledgements

Università di Pavia
Orsetta Zuffardi
Roberto Ciccone
Annalisa Vetro
Erika Della Mina
Ivan Limongelli

OORR Bergamo
Paolo Ferrazzi
Anna Rita Lincesso
Daniela Marchetti
Laura Pezzoli
Maria Elena Sana

Università di Bergamo
Paolo Malighetti

Università di Parma
Federico Quaini

Università di Ferrara
Stefano Volinia

miascone@ospedalieriuniti.bergamo.it