Clonal sequencing technologies: considerations for diagnostic service delivery

Chris Mattocks

National Genetics Reference Laboratory (Wessex)

Applications

- Mutation scanning
- Copy number analysis by counting single reads (array comparative hybridisation equivalent)
- Genome architecture using paired end reads (including balanced variations)
- Non-invasive prenatal diagnosis
- Methylation/epigenetic analysis
- Mitochondrial DNA analysis
- Deep sequencing / tumour profiling
- Expression analysis
- cDNA (RNA) sequencing

Genome analysis

Strategies

- Whole genome
  - Sequence the whole genome regardless of presentation and extract required info.
- Fixed targeting
  - Define limited set of test types with specific pre-defined targeting regime e.g. cardiac, hearing loss, X-linked MR etc
- Flexible targeting
  - Targeting dependant on referral

Potential diagnostic approaches

Strategy	Technologies	For	Against
Whole genome	PacBio Complete genomics Oxford Nanopore ??	•Simplicity •Cost? •Turnaround? •Pre-emptive	•Current feasability •Colateral findings •Data analysis •Cost?
Fixed targeting	Nimblegen Febit Agilent SureSelect	•Simplicity •Availability •Pre-emptive?	•Flexibility •Collateral findings •Data analysis
Flexible targeting	PCR/LR-PCR Gene Collector Gene Selector Long Padlock probes Patch PCR	•Flexibility •Patient based •Efficiency	•Development •Sample processing

Targeting

- Targeting is economic if:
  - targeting $ ≤ whole genome $ (simplistically)
  - Nr targeted analyses ≡ whole genome > ~10
- Targeting efficiency = selectivity x uniformity
  - Selectivity = proportion of sequencing is on target
  - Uniformity = min representation / ave. representation
- Targeting sensitivity
  - For diagnostics it is critical all regions are suitably covered

Uniformity

Fragment	Relative rep.	Sequence required	Actual sequence
A	1	500	500
B	80	500	40000
C	28	500	14000
D	18	500	9000
sequence for 50x coverage		100000	3175000

Uniformity = 1/ave.[1,80,28,18] = 0.031
Actual sequence (i.e. capacity usage) = sequence required / uniformity

Uniformity is measured at level of sequencing - About 5x difference in (representation is seen without targeting and is due to sequencing itself.

Typical values for current technologies range from 0.05 – 0.005
Direct sequencing U~0.25

Capacity

- Usable capacity
  - Platform
  - Application
  - Read length
  - Required read depth
  - Loss of capacity due to partitioning?
  - Targeting efficiency

- Required capacity (Mutation scanning)
  - Colorectal cancer 10 genes ~45kb
  - Hypertrophic cardiomyopathy 22 genes ~200kb

Analysis

- Base calling
- Alignment
- Application specific analysis
- Interpretation
- Data archiving
- Infrastructure

Process models

Sample collection > DNA/RNA extraction > test > analyse > interpret > report >follow-up

- Local - Self contained
- Hub and spoke
- Centralised

- Process control
  - Audit trail / sample tracking
  - Sample exchange

Other issues

- Political
- Ethical
- Financing
- Accreditation

- Implementation (ACCE)
  - Analytical / technical Validation
  - Clinical validation
  - Clinical utility
  - Ethical, legal and social considerations

Summary / conclusions

- Targeting is likely to be useful for diagnostics even with availability of 3rd gen sequencing
- Ideal targeting methodology would be:
  - ✓Cheap
  - ✓Simple
  - ✓Flexible (in capacity and content)
  - ✓100% sensitivity, specificity & uniformity
  - ✓Quantitative
- There are many other considerations for diagnostic implementation of NGS
- Overhaul of service delivery and IT infrastructure

Complex interaction of many factors