UHN Microarray Centre

Next-generation sequencing: Synergy with microarrays

Next-generation sequencing (NGS), also called massively parallel sequencing, has been commercially available since 2004. Unlike dye-terminator sequencing using capillary electrophoresis, which can process 96 samples at a time, NGS increases sequencing throughput by laying millions of DNA fragments on a single chip and sequencing all fragments in parallel (1). DNA fragments, which can be collected from a number of upstream processes, are used to build DNA fragment libraries that are subsequently used as sequencing templates. DNA fragment libraries are prepared for sequencing by ligating specific adaptor oligonucleotides to both ends of each fragment (1,2). Following sequencing, informatics allows each sequencing read to be mapped to a reference genome. The goal of this report is to describe the impact NGS will have on microarrays and provide a summary of three popular NGS platforms.

NGS platforms can be used for transcriptome profiling, miRNA profiling, DNA-protein interaction studies using chromatin immunoprecipitation (ChIP), and DNA methylation studies, thus challenging the use of microarrays. However, several studies have concluded that microarrays and NGS are actually complementary platforms, rather than competitive alternatives, that should be used together to gain the maximum results (3-5). Euskirchen *et al*. compared ChIP-chip with ChIP-sequencing and found that each method detected targets that were missed by the other method (3). For transcriptome profiling, Oudes compared Affymetrix GeneChip Array data with Massively Parallel Signature Sequencing (MPSS) and found that each method detected genes that the other did not and concludes that transcriptome profiling with a single methodology will not fully assess the expression of all genes (4). Researchers considering replacing microarrays with sequencing platforms should be aware of the advantages and disadvantages associated with both methods (Table 1). NGS has some advantages that cannot be rivaled by current microarray platforms, including lower background, better sensitivity, and quantitative measures (6). However, a sophisticated information technology infrastructure is necessary to manage the

Table 1. The strengths and weaknesses of NGS and microarrays.
(This table was taken from Asmann *et al*. Gastroenterology 2008, 135:1466)

Microarray Analysis		Massively parallel, or next-generation, sequencing	
Pros	*Cons*	*Pros*	*Cons*
Relatively inexpensive	High background, low sensitivity	Low background, very sensitive	Expensive
Easy sample preparation	Limited dynamic range	Large dynamic range	Complex sample preparation
Mature informatics and statistics	Not quantitative	Quantitative	Limited bioinformatics
	Competitive hybridisation		Massive information technology infrastructure required
	Annotation of probes		

vast amount of sequence data (6,7), and the future success of NGS will depend on the development of new algorithms to perform alignments and handle large amounts of data efficiently (7).

NGS platforms

Three commercially available NGS platforms, including Roche (454) FLX Genome Sequencer, Illumina Genome Analyzer II, and Applied Biosystems' SOLiD™ (Sequencing by Oligo Ligation and Detection) Sequencer, will be briefly described. Table 2 compares the specifications of each platform. The information presented in this report was collected in November 2008 (8-10) and was intended to provide a general comparison. We highly recommend that you contact each manufacturer for current performance specifications.

It should also be mentioned that single molecule sequencers, so-called "third-generation sequencers", have been developed. Helicos' HeliScope, which employs True Single Molecule Sequencing (tSMS) technology to sequence samples without amplification, is capable of producing over 10 Gb of sequence data per 8 day run (11). Pacific Biosciences has developed Single Molecule Real Time (SMRT™) sequencing technology which involves proprietary surface and nucleotide chemistries (12). This sequencer, which promises longer reads, shorter run times, and higher quality data, is expected to be released in 2010 (12).

Sequencing chemistry

Pyrosequencing (Roche platform) involves the use of a pyrophosphate molecule, released following nucleotide incorporation by DNA polymerase, to propagate reactions that ultimately produce light. Illumina's sequencing-by-synthesis involves the use of four differently labelled fluorescent nucleotides that have their 3'-OH groups chemically inactivated to ensure only a single base is incorporated per cycle. Each base incorporation cycle is followed by an imaging step to identify the base incorporated, and a chemical step that removes the fluorescent group and deblocks the 3' end for the next base incorporation cycle. The SOLiD sequencer (Applied Biosystems) uses a ligation-based sequencing process that starts by annealing a universal sequencing primer that is complementary to the SOLiD-specific adaptors on the library fragments. Then, a limited set of semi-degenerate 8-mer oligonucleotides (and DNA ligase) is added. When matching 8-mers hybridise to DNA fragment sequences adjacent to the universal primer, DNA ligase seals the phosphate backbone and a fluorescent readout identifies the fixed base of the 8-mer. A subsequent cleavage step removes bases 6-8 of the ligated 8-mer, removing the fluorescent group and enabling another round of 8-mer ligation, and so on. The advantage of ligation-based sequencing chemistry is the "built-in" quality check of read accuracy (2).

Table 2. Comparison of Roche, Illumina, and Applied Biosystems NGS platforms.

	Roche (454): Titanium series reagents (run on FLX Genome Sequencner)	Illumina: Genome Anlayser II	Applied Biosystems: SOLiD™
Sequencing chemistry	pyrosequencing	polymerase-based sequencing-by-synthesis	Ligation-based sequencing
Amplification approach	Emulsion PCR	Bridge amplification	Emulsion PCR
MB/run	400-600 MB	1300 Mb	3000 MB
Time/run	10 hr	4 days	5 days
Read length	400 bp	up to 75 bp	35 bp

Amplification approach

Emulsion PCR is the amplification approach used by both Roche and Applied Biosystems. For the Roche platform, emulsion PCR is carried out on the DNA fragments attached to the surfaces of agarose beads. On the Applied Biosystems sequencer, the DNA fragements are attached to the surfaces of magnetic beads. Bridge amplification (Illumina) is a PCR reaction that occurs within a discrete area of the flow cell surface.

Read length/Mb per run/Time per run/Approximate cost

NGS produce shorter reads (25-400 bp) than capillary sequencers (650-800 bp), but the read length varies considerably among NGS. Roche provides the longest read (400bp with the Titanium system) while Illumina and Applied Biosystems are much shorter (32-75 bp and 35 bp, respectively). The advantageous read length and run time of the Roche platform is offset by its cost (the Roche GS FLX sequencer's "cost-per-run" is more expensive than the other two; 2) and the lower run yield (roughly between 20% and 46% of the sequence data that could be obtained by using the Applied Biosystems and Illumina systems).

Be sure to read "Cost, Data Analysis, and Throughput Keep Array Users from Switching to DGE, Some Say" by Justin Petrone, December 23, 2008, on GenomeWeb BioArray News. This article discusses digital gene expression (DGE) performed on second-generation sequencing platforms and the impact DGE could have on microarrays.

References

1. Coppee J-Y. Do DNA microarrays have their future behind them? Microbes and Infection 2008, 10:1067-1071
2. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends in Genetics 2007, 24(3):133-141
3. Asmann YW, *et al*. Transcriptome profiling Using Next-Generation Sequencing. Gastroenterology 2008, 135:1466-1468
4. Pop M, Salzberg SL. Bioinformatics challenges of new sequencing technology. (Review) Trends in Genetics 2008, 24(3):142-149
5. Euskirchen GM, *et al*. Mapping of transcription factor binding regions in mammalian cells by ChIP: Comparison of array- and sequencing-based technologies. Genome Res 2007, 17:898-909
6. Oudes AJ, *et al*. Application of affymetrix array and massively parallel signature sequencing for identification of genes involved in prostate cancer progression. BMC Cancer 2005, 5:86
7. Liu F, *et al*. Comparison of hybridization-based and sequencing-based gene expression technologies on biological replicates. BMC Genomics 2007, 8:153
8. Roche 454 Sequencing. System Features for GS FLX Titanium Series. (November 24, 2008) http://www.454.com/products-solutions/system-features.asp
9. Illumina Genome Analyzer II Specification Sheet. (November 24, 2008) http://tucf.org/nextgen_datafile_genome_analyzer2_nov08.pdf
10. Applied Biosystems Specification Sheet SOLiD™ System 2.0. (November 24, 2008) http://www3.appliedbiosystems.com/cms/groups/mcb_marketing/documents/generaldocuments/cms_057562.pdf
11. Helicos Biosciences Corporation. (December 22, 2008) http://www.helicosbio.com/Products/HelicostradeGeneticAnalysisSystem/HeliScopetradeSequencer/tabid/87/Default.aspx
12. Pacific BioSciences (News Page). (January 3, 2009) http://www.pacificbiosciences.com/index.php?q=news