High-throughput Genomic Paper Notes

Jonathan Yu

October 2, 2017

Tumor Analysis Best Practices Working Group, 2004

- Tumor Analysis Best Practices Working Group. "Expression Profiling–Best Practices for Data Generation and Interpretation in Clinical Trials." Nature Reviews. Genetics 5, no. 3 (March 2004): 229-37. doi:10.1038/nrg1297.

Microarrays is a widely accepted technological way to analyze mRNA transcript levels for a genome. For this technique, there are a variety of methods methods for data generation and analysis for different experimental platforms: cDNA (sets of plasmids of specific cDNA in gridded liquid aliquots), spotted oligonucleotide (concentration of a known single-stranded sequence obtained from liquid handling on glass slides) and Affymetrix arrays (probes that are synthesized using light-activated chemistry and photolithography to find signals). The Tumor Analysis Best Practices Working Group deteremined the best practices for experimental design, probe-set analysis algorithms, signal/noise assessments and biostatistical methods. For human trials, longitudinal or corss-sectional design was determined best protocol as an experimental design for best power. On the note of technical variabliiltiy, reproducibility as well many other problems should be met with some thresholds mentioned such as 2 standard deviation from mean for scaling factor to normalize chips and percentage of present calls among samples should be within 10%. For signal/noise, they determined that each project will have its own signal/noise optimum and their own method of best analysis and compared the different algorithms (Table 1) on a number of criterions. The note that since 'feature or gene selection is vitally important when microarrays ar eused for differential diagnosis', they recommend users try different statistical methods such as standard parametric tests, non parametric methods, and global/local shrinkage methods. With aggregate gene expressions can help reduce dimension - a prevalent problem for analyzing gene related data - as well as gene selection, multiple testing and collinearity. Finally, they state back-end statistical methods such as data visualization and time-series studies can help with circumvent problems as well.

Lipshutz et al., 2005

- Liphsutz, Robert J., Stephen PA Fodor, Thomas R. Gingeras, and David J. Lockhart. "High density synthetic oligonucleotide arrays." Nature genetics 21, no. 1 (1999): 20-24

Lipshutz et al. have developped a tool to collect analyze vast amounts of genetic and cellular information simultaneously from nucleic acid strands. Through 'fabrication of hundred of thousands of polynucleotides at high spatial resolution on a precise location' and 'laser confocal fluorescence scanning', they were able to design a DNA probe array to get complmentary sequences. For example, the array is coated with a chemistry protecting group to prevent DNA deposition and then a mask is placed on top to expose specified regions. A light is used to knocked off exposed protecting group and then add certain solution of nucleitide incubations to hybridize to the nucleitide at that location. This cycle is repeated until a synthesized polynucleotides of about 300,000 are created at specific locations of the 1.28 cm $\times$ 1.28 cm array. The array that can contain approximately 40,000 human genes are used for expression monitoring to understand a gene function. From these arrays, fluorescence intensity image obtained from the array are compared with the reference sequence to be deemed perfect match (PM) and mismatch (MM). The difference of PM and MM can help reduce background noise and cross-hybridization.

This technique is useful as there is a need for 'monitoring expression levels of a large number of genes repeatedly, routinely, and reproducibily' that do not need physcial intermediaries such as cDNA, PCR products or clones and the process that come with it to prepare, verify and cataloque the large number of sequence. Again, for a single position on the DNA, one thousand sets of four probe pairs (1 complementary

probe at the specific region and 3 similar complementary probes at anywhere but the specific position) are conducted to detect variants in DNA sequence - by identifing the difference (given a specific place) or positions. Using photolithography and light-activated chemistry, they are able to take advantage of the complementary properties of genes and the target design of the probes to monitor a large amount of expression levels. Companies such as Affymetrix are developping software tools to manage, monitor, genotype, and sequence analyze these huge datasets.

Watson & Crick, 1953

- Watson, J. D., and F. H. Crick. "Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid." Nature 171, no. 4356 (April 25, 1953):737-38

After discovering some inconsistensies with current proposed structures of nucleic acid, Watson and Crick proposed a two helical chain structure where each are coiled around the same axis. Using the same chemical assumption of the 3', 5' linkages, the chains run in opposite directions and have bases on the insdie and the phosphates on the outside. In other words, the proposed DNA strucutre is a double-stranded helical model with two sugar-phosphate as backbones on the outside and hydrogen bonds between pairs of nitrogenous bases on the inside. The new feature includes having the two chains held by purine and pyrimidine bases - joined together in pairs by hydrogen-bond. Specifically, regarding bases, specific pairs bond together: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine). They thank Dr Jerry Donohue and Dr. M. H. F. Wilkins & Dr. R. E. Franklin for their criticisms and expirments for inspirations.