Clinical Chemistry
Trainee Council

PEARLS OF LABORATORY MEDICINE

Massively Parallel (or Next-Generation) Sequencing

DOI: 10.15428/CCTC.2013.209544

Melody Caramins, B.Med, PhD, FRCPA, FFSc
New South Wales Health Pathology

www.traineecouncil.org

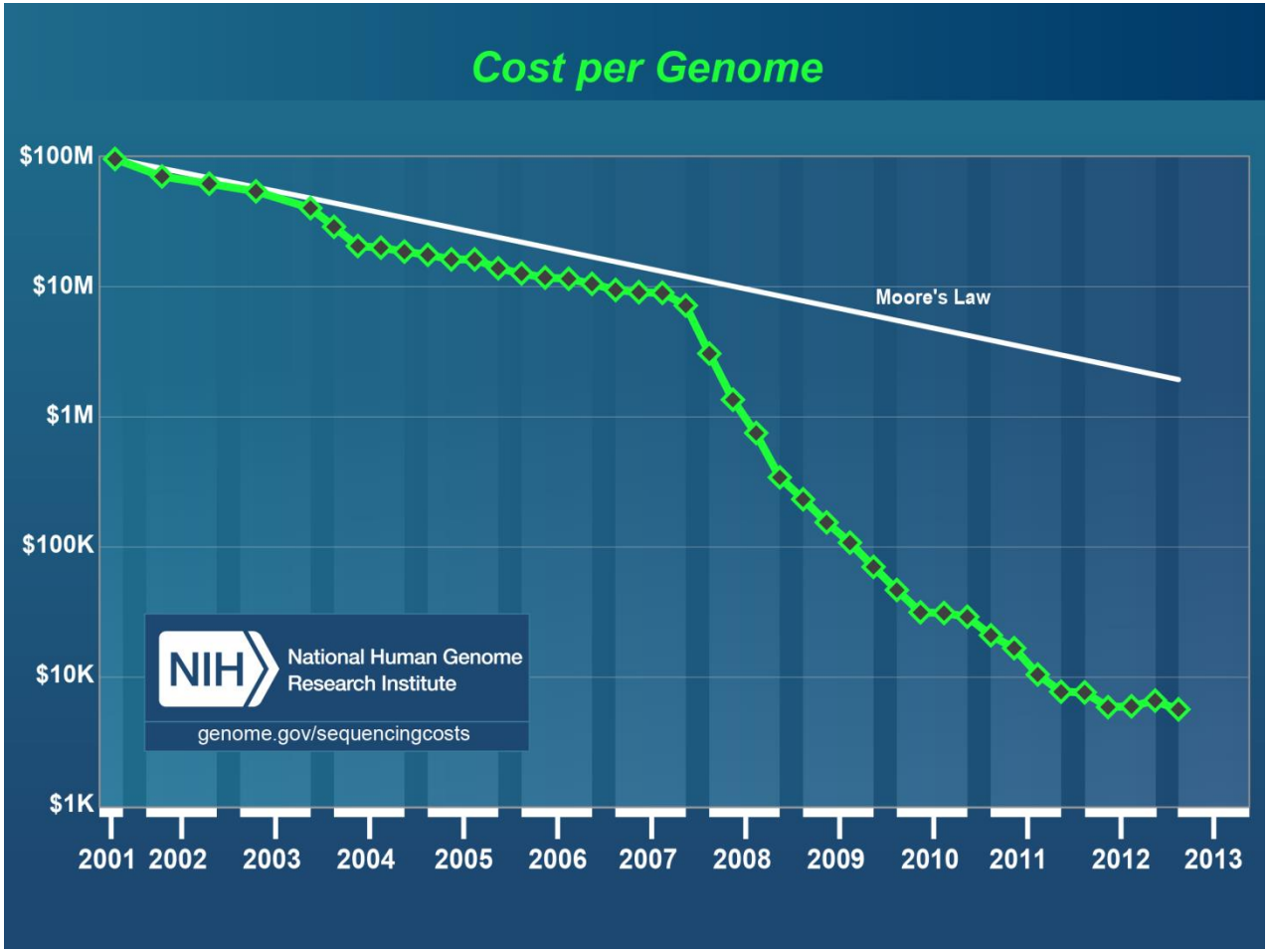
Sequencing, from past to present

- DNA sequencing
  - Involves determining the order of nucleotides in a given DNA molecule
  - Described by Maxim and Gilbert and Sanger et al. independently in 1977
  - Sanger sequencing has been the method of choice for DNA sequencing in the diagnostic setting until recently
  - Cost of sequencing when human genome project proposed (1985) ~$10/base. When finished (2003) ~$0.01/base, and is ~ millionth of $1/base today

Old capacity vs. New capacity

- All sequencing platforms generate sequence data in the form of independent reads, which then must be assembled to form a complete sequence
- Sanger sequencing read lengths ~800-1000 bp, with 1-96 capillaries (max. 96 reads); so the typical output would be ~ 1Mb/day
- Massively parallel methods produce much larger quantities of reads, but typically in shorter read lengths, with total outputs of up to 800 Gb/run

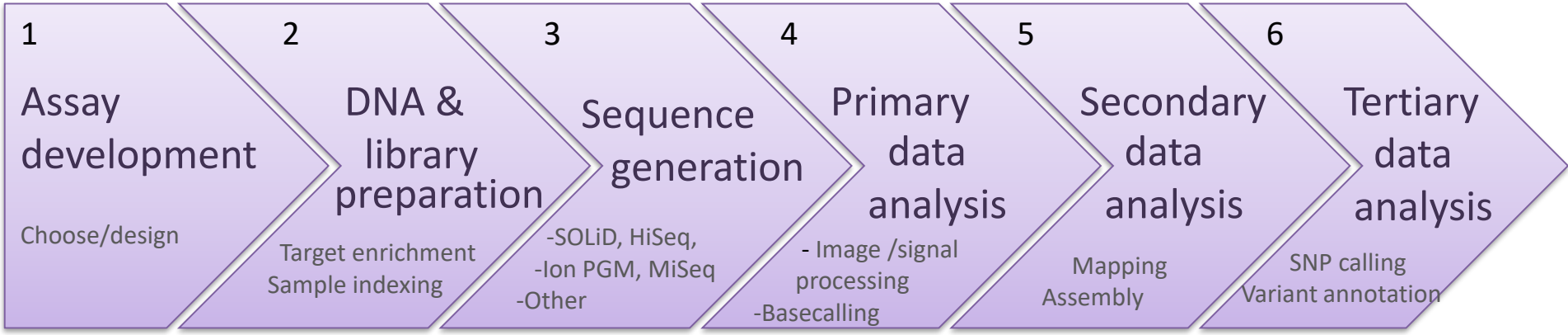
Falling costs of sequencing



"Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: www.genome.gov/sequencingcosts. Accessed April 22, 2013."

Overview of the MPS workflow

- The general steps involved in the workflow for MPS:
  1. Library preparation
  2. Sequence generation
  3. Three tiers of data analysis



Library preparation

- There are four major sub-steps for library prep (About 6 hours of lab work)
  - Fragment DNA
  - Repair ends
  - Ligate adapters
  - Select ligated DNA
- Enrichment PCR

Sequencing

- There are several competing MPS technologies
- The main technologies currently on the market:
  - Illumina
  - Roche 454
  - Life Technologies
  - Pacific Biosciences
  - Complete Genomics

Analysis

- Primary data analysis
  - Refers to base calling from raw output
  - Most of this is under vendor control , using reasonably robust algorithms
  - “Raw data” produced is already compressed 1000x
- Secondary analysis
  - Involves the alignment of reads to the reference sequence and the identification of sequence variants
- Tertiary analyses
  - Describes the annotation of sequence variants

Current applications

- Whole genome vs. targeted regions
  - Targeted region can be as small as a few (e.g. BRCA1/2) genes, multiple genes (e.g. >20 known AD retinitis pigmentosa genes) or the entire exome
- Research vs. Diagnostics
  - Diagnosis and gene discovery
  - Cohort genomics
- Hereditary variants vs. somatic variants

Disease gene discovery & MPS in Mendelian disorders

- Methods include:
  - Sequencing multiple affected individuals from the same family to identify a shared novel variant and diagnose rare disease
  - Sequencing multiple unrelated, affected individuals for variants in the same gene or pathway
  - Sequencing parent–child trios for identifying de novo mutations
  - Sequencing and comparison of extremes of a phenotype distribution to identify variants for quantitative traits

MPS and Cancer

- Single patient samples
  - Diagnosis – Somatic gene or mutation panels to inform physician decision-making with respect to treatment
  - Research – might generate hypotheses
- Cohort genomics
  - Research – aiming to catalogue the large majority of recurrently somatically mutated genes found in a cancer type
- Cohorts incorporating multiple “omics”
  - Research - E.g. exome, transcriptome, epigenome

Challenges for clinical implementation

- Pre-analytical challenges
  - Hardware implementation
  - Sample quality/quantity - less of an issue with advancing protocols
- Analytical challenges
  - Clinically useful turnaround times
  - Data pipelines and validation
  - Data analysis
- Post analytical challenges
  - Interpretation and Reporting
  - Data Storage

Summary

- Sequencing has become a commodity technology: the exact technology used is not as important as what you want to do with it and whether it will do what you want
- An important consideration for clinical implementation is how much of the genome is sequenced, how much of the sequence is interrogated, and how much of that is reported
- Panel testing for disorders with the moderate number of disease genes will likely be important for a while but ultimately it may be more efficient to perform exome sequencing
- In disorders with low numbers of high penetrance genes – e.g. BRCA1/2 the reduction in genomic complexity allows increased throughput

References

- **Reviews of platforms, laboratory, and analytical methods:**
  - Quail MA, Smith M, Coupland P, Otto TD, Harris SR, et. al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 2012;13:341
  - Liu L, Li Y, Li S, Hu N, He Y, Pong R, et. al. Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012.
  - Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11:31-46.
  - Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efremova M, et. al. A survey of tools for variant analysis of next-generation genome sequencing data. Brief Bioinform. 2013 Jan 21. (epub)
  - Altmann A, Weber P, Bader D, Preuss M, Binder EB, Müller-Myhsok B. A beginners guide to SNP calling from high-throughput DNA-sequencing data. Hum Genet. 2012;131:1541-54.
- **Applications of next-generation sequencing:**

  http://www.nature.com/nrg/series/nextgeneration/index.html

Disclosures/Potential Conflicts of Interest

Upon Pearl submission, the presenter completed the Clinical Chemistry disclosure form. Disclosures and/or potential conflicts of interest:

- **Employment or Leadership**:
  - Staff Specialist Genetic Pathologist, NSW Health Pathology
  - Chair, RCPA Pathology Update 2014 - Genetics Scientific Programming Committee
- **Consultant or Advisory Role:**
  - Member, Genetics Advisory Committee, Royal College of Pathologists of Australasia
  - Member, Federal Department of Health and Ageing, Genetics Working Party
  - Member, Federal Department of Health and Ageing, Pathology Services Advisory Committee
- **Stock Ownership:** None declared
- **Honoraria:** None declared
- **Research Funding:** None declared
- **Expert Testimony:** None declared
- **Patents:** None declared

Thank you for participating in this
Clinical Chemistry Trainee Council
Pearl of Laboratory Medicine.

Find our upcoming Pearls and other
Trainee Council information at
www.traineecouncil.org

Download the free *Clinical Chemistry* app
on iTunes today for additional content!

Follow us: