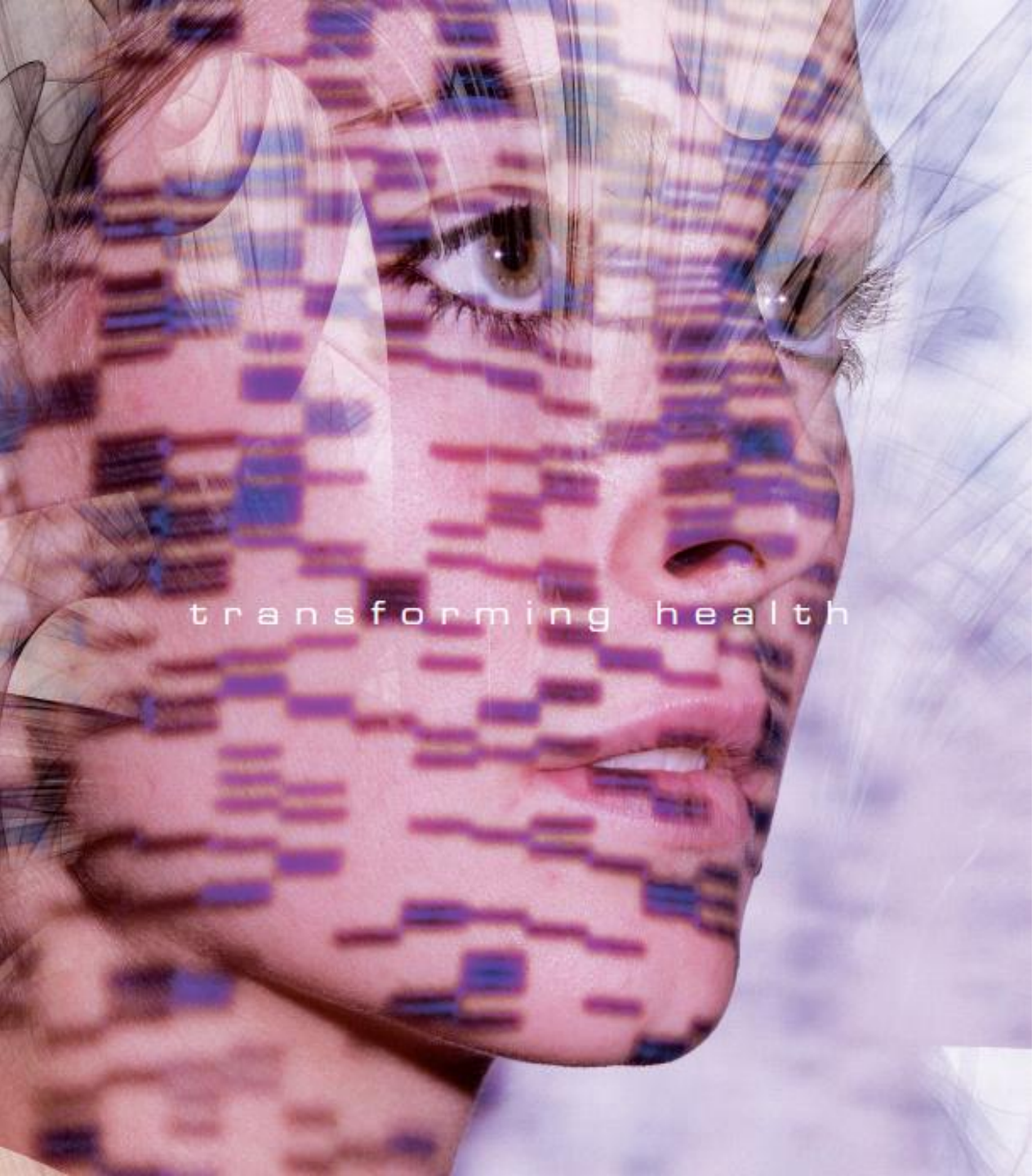


The Institute for Individualized Health™

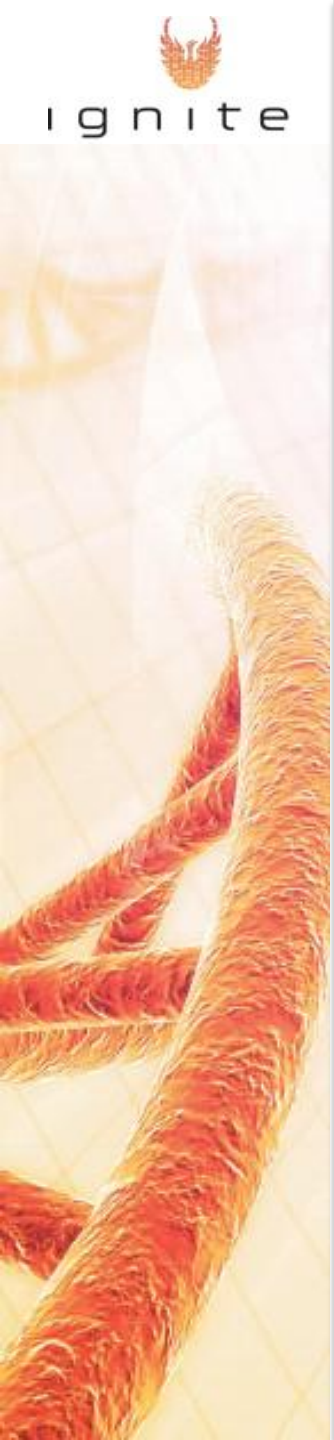
Overview of WGS

Dietrich A. Stephan, PhD
President and CEO

22nd Annual SACGHS Meeting

June 15, 2010

ignite

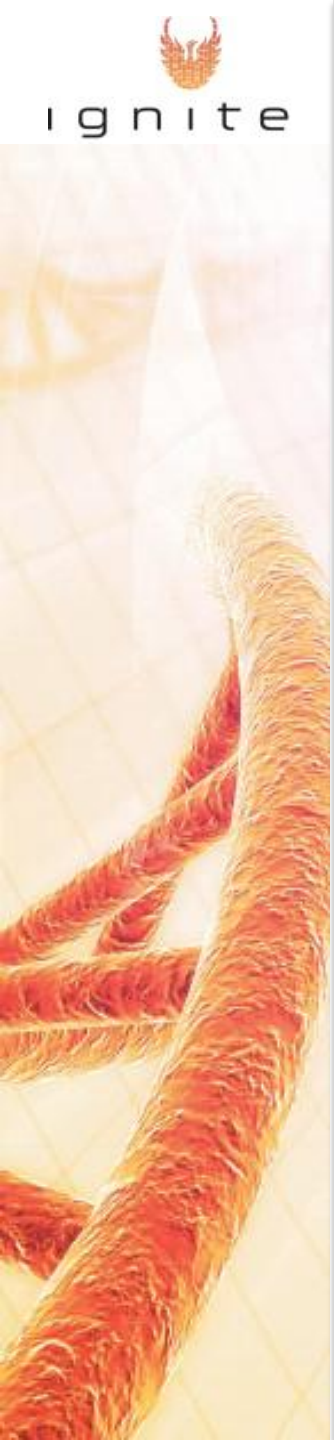


Introduction

- Why would you want to sequence a full genome?
  - Research setting
  - Clinical setting
- Key definitions and performance metrics
- The pace of technology evolution in the WGS space
- Partitioning the genome
- The “generations” of sequencing technology
- Analysis and interpretation

Disclosures

Co-Founder, Navigenics Inc
Co-Founder, Aueon Inc
Co-Founder, Amnestix Inc
Director, Sage BioNetworks
Director, Personalized Medicine Coalition
Director, Virginia Biotechnology Association
Advisor, gnuBIO Inc
Advisor, RainDance Inc



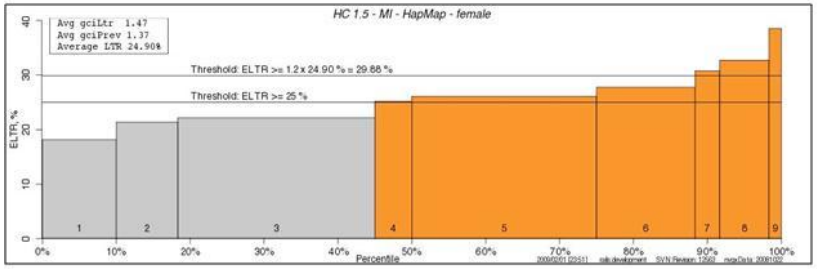
Why Sequence a Full Genome – Research

- All human disease has a genetic component
- Many monogenic diseases left to understand
- Most complex genetic diseases remain a mystery
  - Reduce ability to diagnose early
  - Reduce our ability to treat

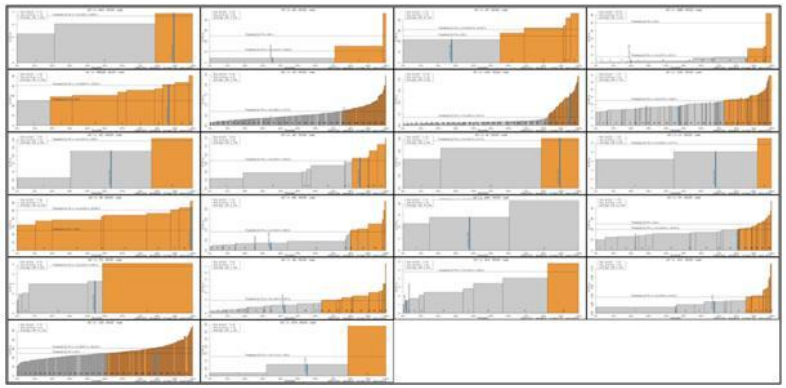
- Using WGS sequencing in a R&D setting we can:
  - Capture the “missing heritability” for pre-symptomatic risk stratification
    - Common variants
    - Rare variants
    - Epigenetic modifications
    - Copy number variants
  - Sub-classify disease at molecular level and re-name
  - Enable a new era of exposure epidemiology
  - Define prioritized screening strategies
  - Understand core pathways for drug targeting and dosing

Risk Classification – Exposure Epidemiology – Screening

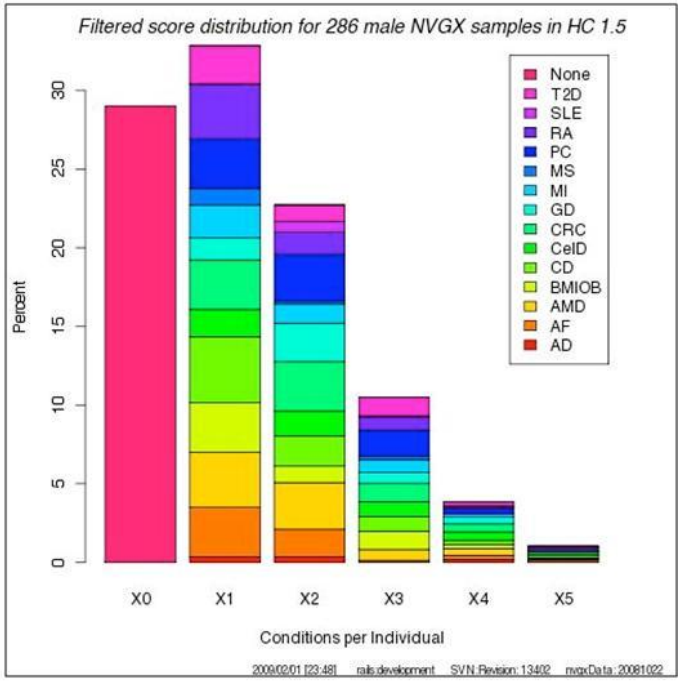
Condition View



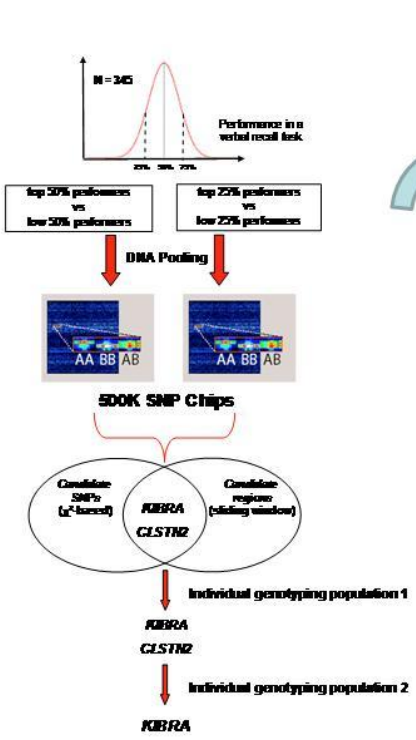
Individual View

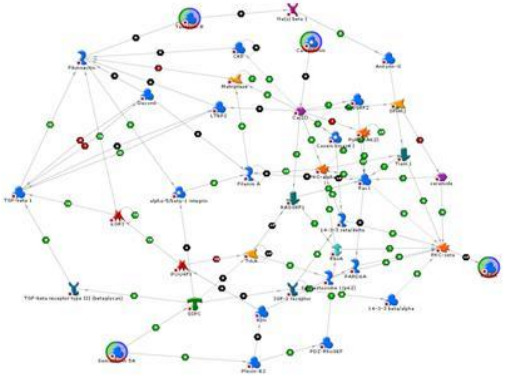


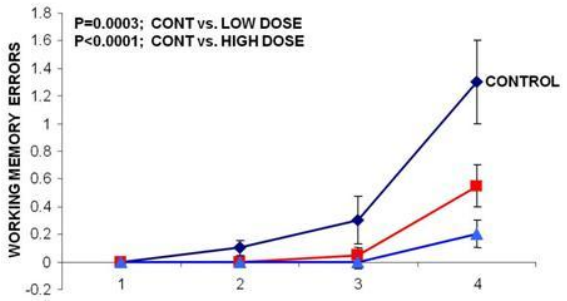
Aggregate View



Today we are OK at New Drug Development







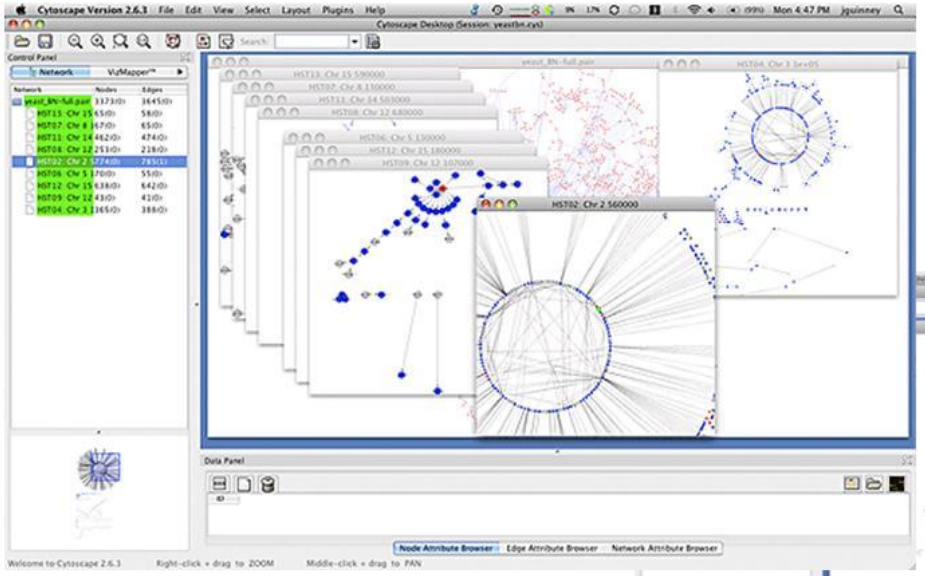
- WGAs resulted in great drug development with high hit-rates, beginnings of the field of Probabilistic Risk Assessment or “Medical Genetics 2.0”

ignite

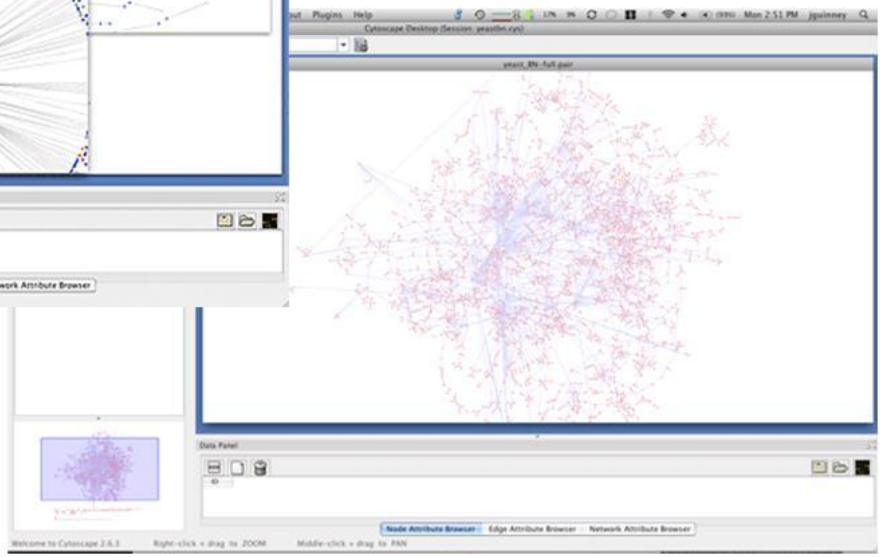
Highly Efficient Genome-Informed Drug Development?

Stephen Friend, Eric Schadt, Lee Harwell, John Wilbanks, Hans Wigzell, Dietrich Stephan

Personalized Drug Discovery Based on Your Genome?



Input: A framework network is informed by personalized variants, clinical data, exposures, etc.

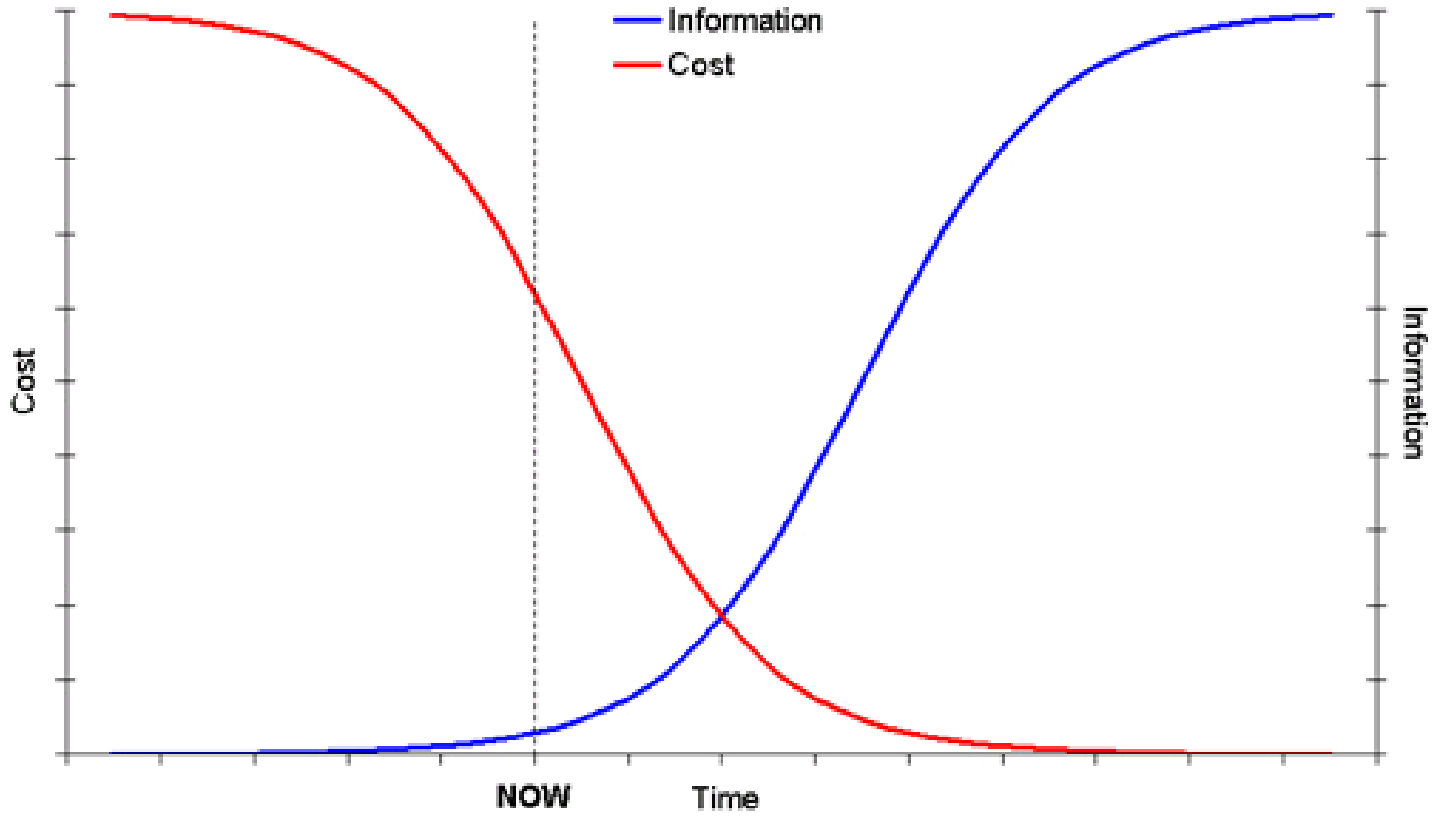


Output: Personalized combination therapies at the point of care

Soon technically feasible with great promise in cancer and others
– will test the limits of FDA!

ignite

Why Sequence a Full Genome – Clinical



Value over cost reaches inflection point in next 2 years

Why Sequence a Full Genome – Clinical

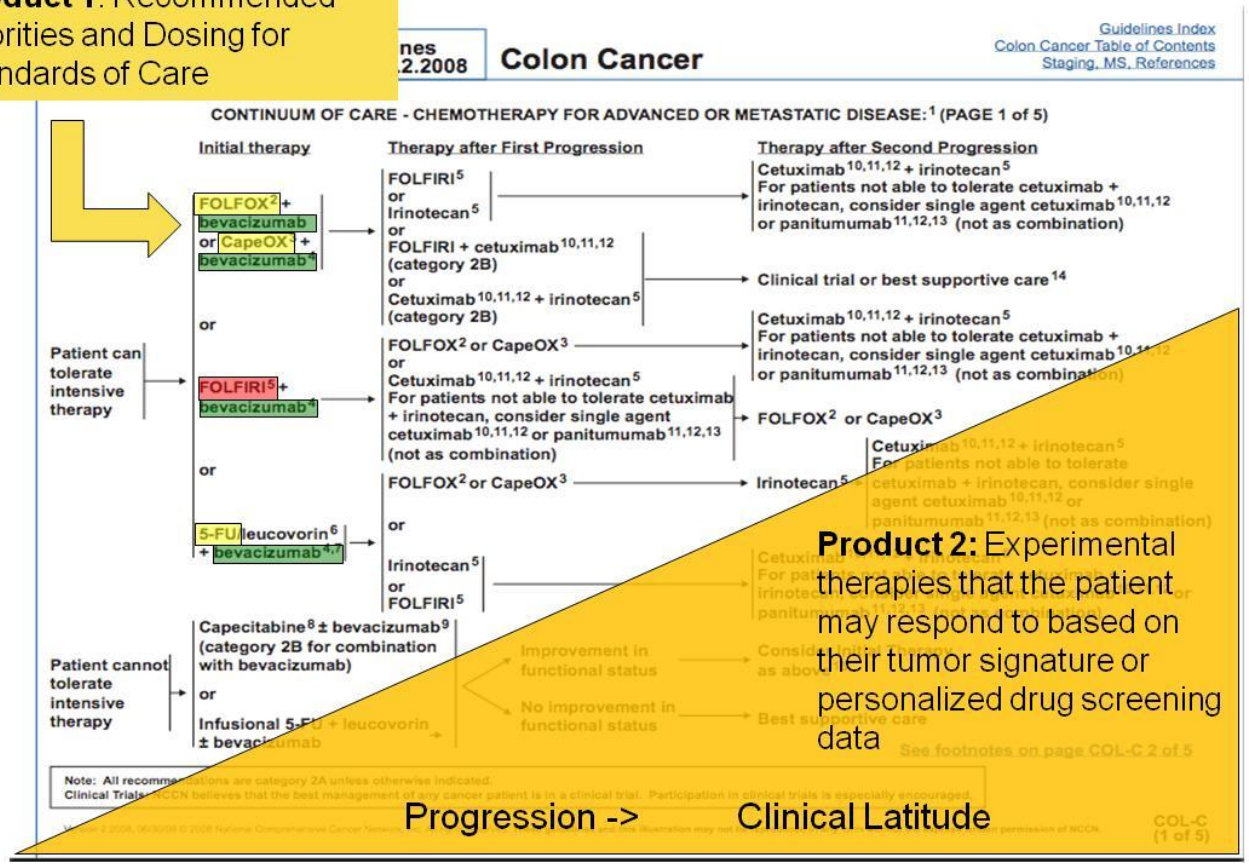
- Better Diagnostics and Drug Targeting/Dosing

- Better Value in Healthcare: Collapse all DNA-based MDx onto a single assay that can be mined and ad value across a lifetime

- When? The value amortized across a lifetime across the aggregate is more than the cost of a sequence for each – 2 years

Aueon: Cancer Genomes can add Immense Value Today

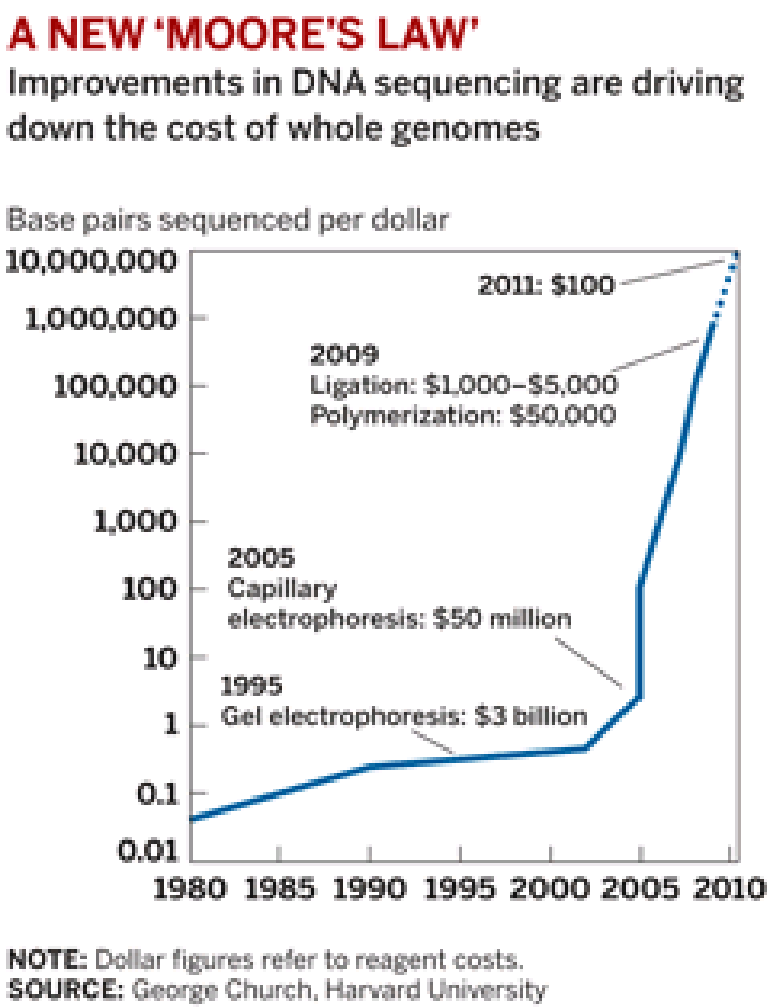
Product 1: Recommended Priorities and Dosing for Standards of Care



- Aueon Diagnostics, Inc.
- Cancer Care Alliance (Life Technologies, US Oncology, Ignite, Etc)

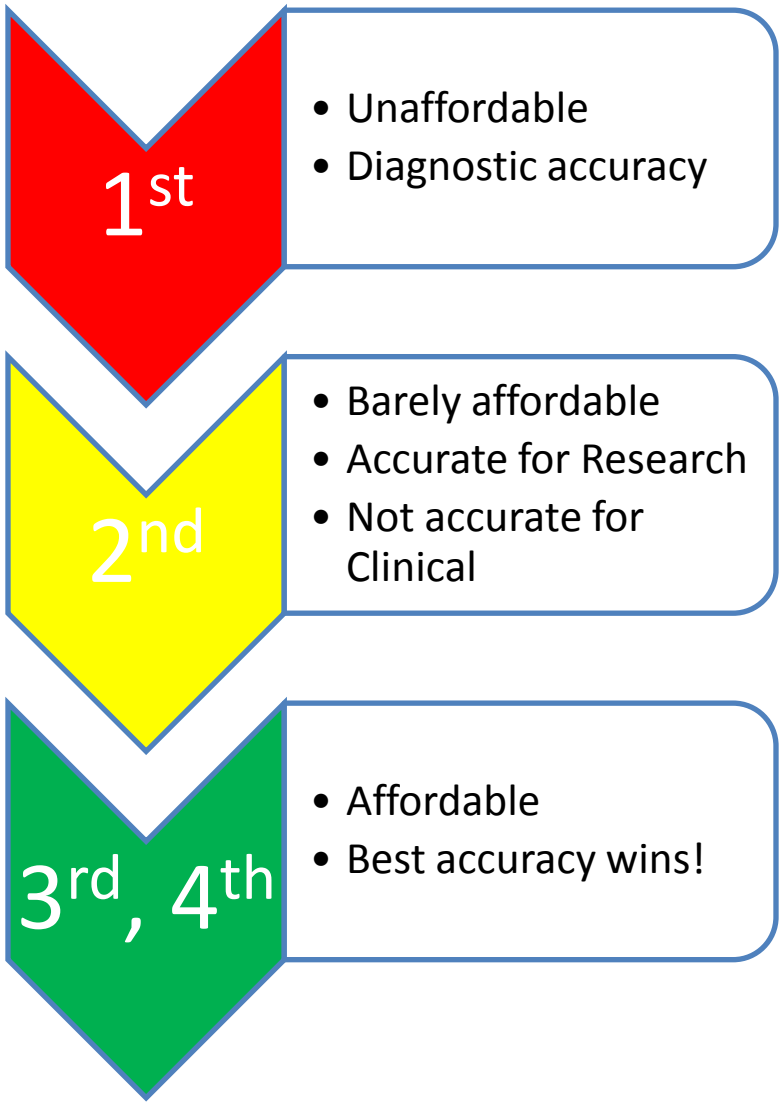
ignite

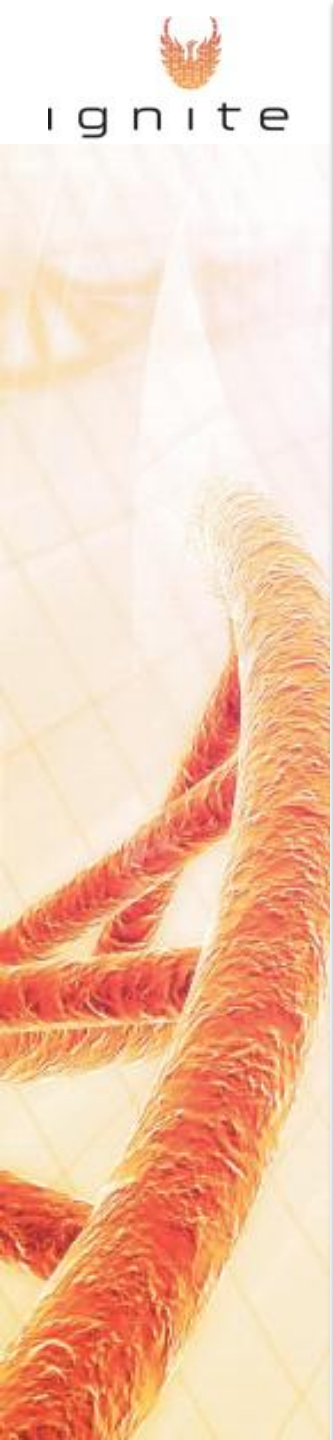
The Pace of Technology Evolution in the WGS Space



The "Generations" of Sequencing Technology

- 1st Generation
  - ABI Capillary Electrophoresis
- 2nd Generation
  - Life Technologies/ABI SOLiD4HQ
  - Complete Genomics
  - Illumina HiSeq
  - 454
  - Polonator
- 3rd Generation
  - Life Technologies SM
  - Pacific Biosciences
  - Helicos
- 4th Generation
  - Oxford NanoPore
  - Ion Torrent
  - GnuBio
  - Genia
  - Etc.



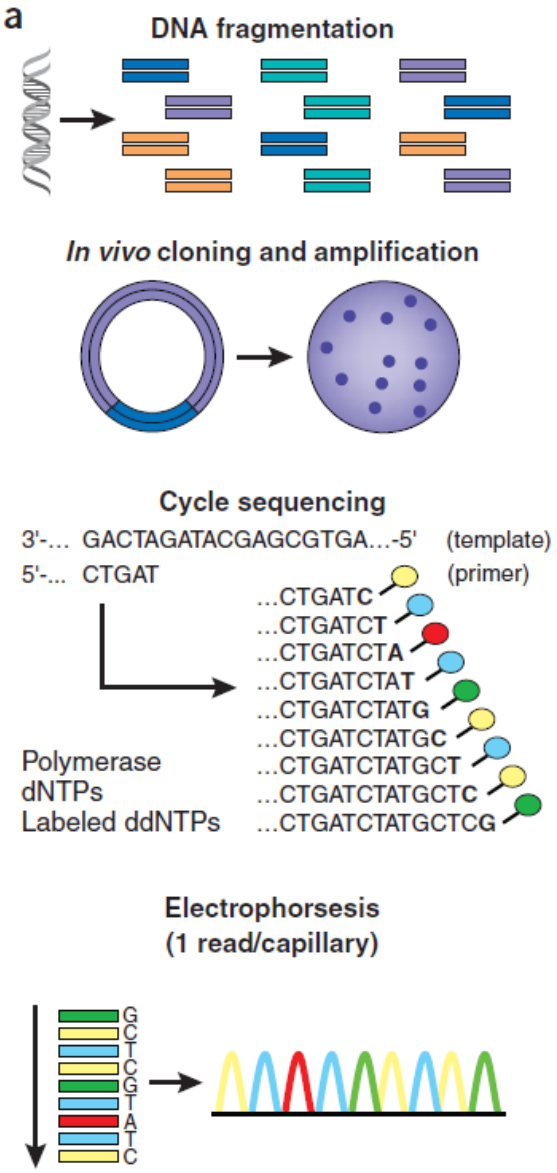


Definitions

- DNA extension reaction
- DNA polymerase
- Fluorescent nucleotide
- Optical imaging
- Fold-coverage
- Read length
- Paired-end reads
- Library
- Partitioning or selection
- Cost per base
- Cost per genome
- Accuracy
- Throughput
- Assembly
- Storage
- Interpretation

First Generation Technology

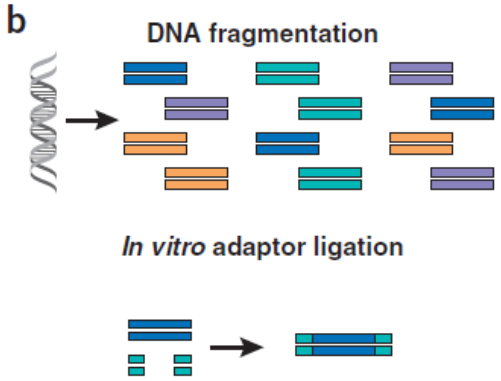
- Capillary electrophoresis
- Life Technologies/ABI sequencer:
  - Long reads
  - Very targeted (single gene)
  - Very expensive
  - Clinical/diagnostic accuracy



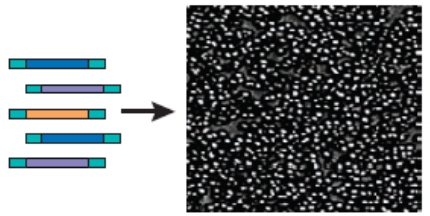
Adapted from a presentation by: FRANCIS OUELLETTE Informatics on High Throughput Sequencing Data July 2009

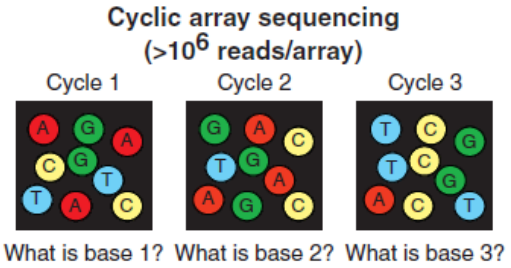
Second Generation Technologies

- Parallel extension reactions on a single array
  - Life Technologies/ABI SOLiD
  - Illumina HiSeq
  - Polonator
  - 454



- Sequence libraries of fragments (Mb-Gb of sequence)
- Slow (days)
- Short reads make assembly difficult
- Moderate error rates (research grade tools)

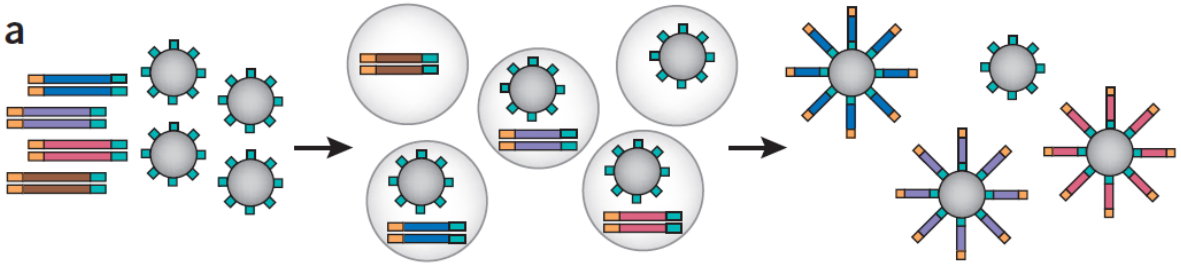




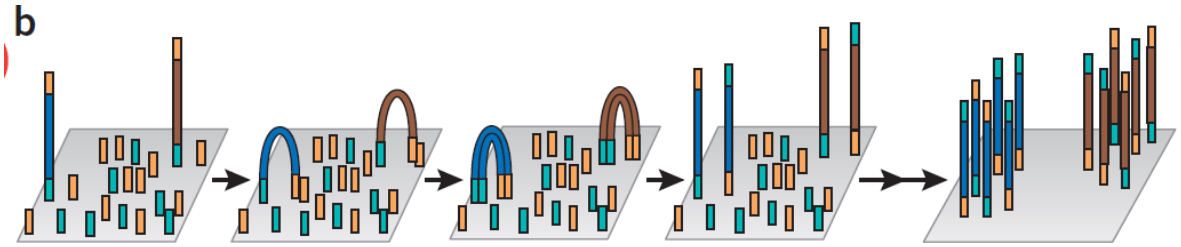
Adapted from a presentation by: FRANCIS OUELLETTE Informatics on High Throughput Sequencing Data July 2009

Various Assays

- Emulsion PCR
  - Fragments, with adaptors, are PCR amplified within a water drop in oil
  - One primer is attached to the surface of a bead
  - Used by 454, Polonator and SOLiD



- Bridge PCR
  - DNA fragments are flanked with adaptors
  - A flat surface coated with two types of primers, corresponding to the adaptors
  - Amplification proceeds in cycles, with one end of each bridge tethered to the surface
  - Used by Illumina



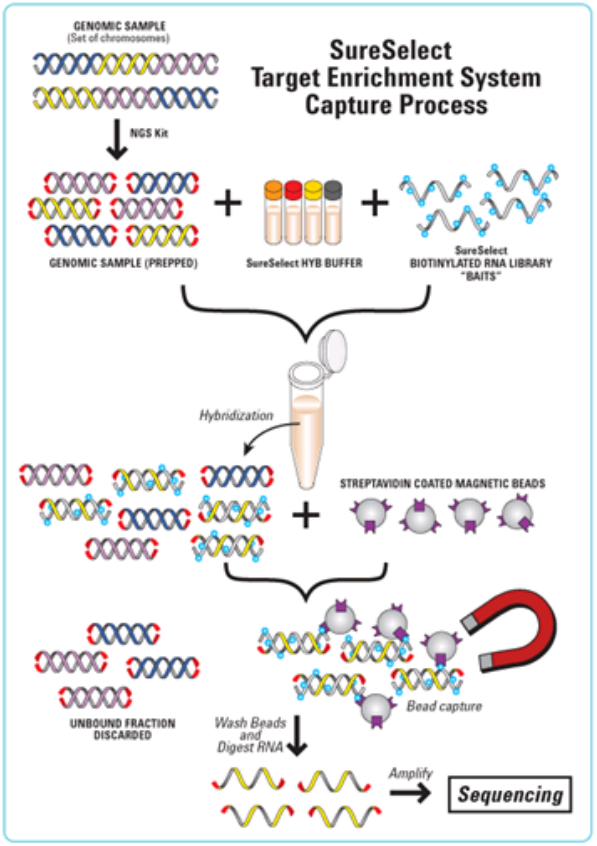
Adapted from a presentation by: FRANCIS OUELLETTE Informatics on High Throughput Sequencing Data July 2009

ignite

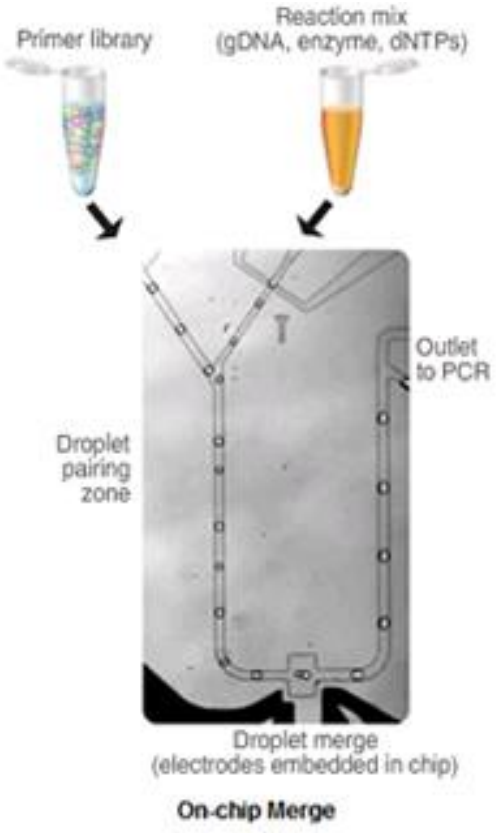
Extracting Value from 2nd Generation - Partitioning

- Identify regions of the genome that are of interest and selectively sequence those to reduce cost

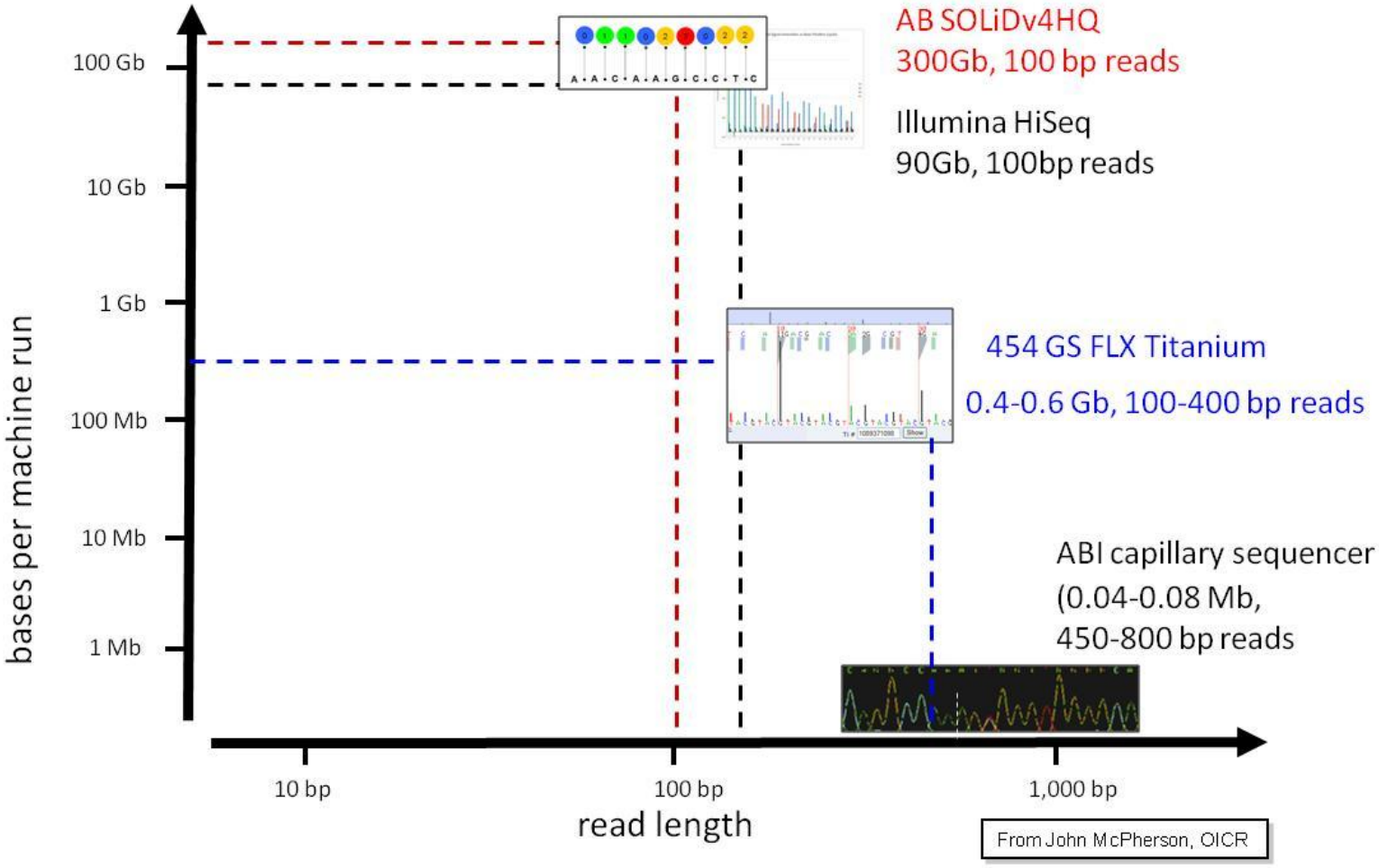
Agilent – Hybdrization



RainDance – NanoReactor/PCR



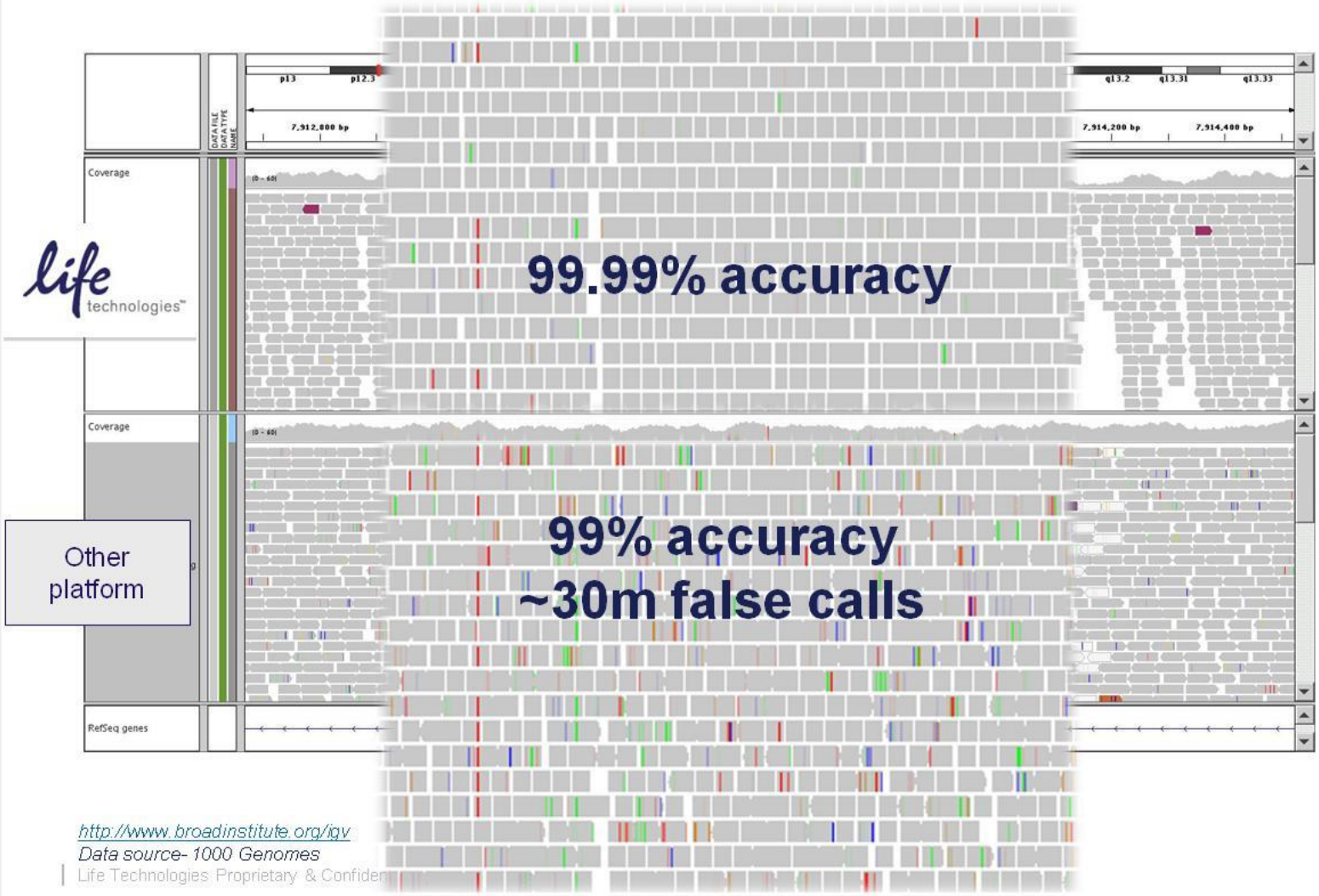
Current Generation Sequencers (1st and 2nd Gen)



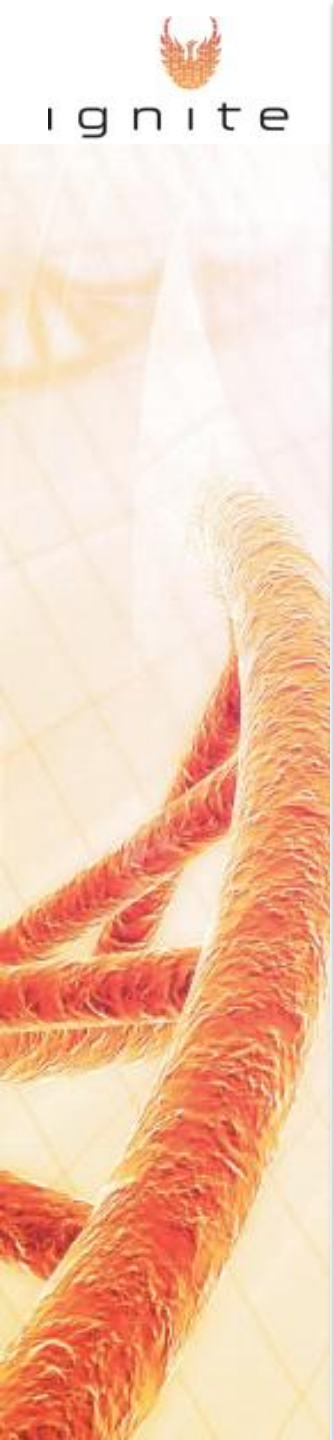
Adapted from a presentation by: FRANCIS OUELLETTE Informatics on High Throughput Sequencing Data July 2009

ignite

Accuracy of 2nd Generation Platforms



http://www.broadinstitute.org/igv
Data source- 1000 Genomes



Comparison of Current Technologies (1st and 2nd Gen)

	Feature generation	Sequencing by synthesis
454	Emulsion PCR	Polymerase (pyrosequencing)
Solexa	Bridge PCR	Polymerase (reversible terminators)
SOLiD	Emulsion PCR	Ligase (octamers with two-base encoding)
Polonator	Emulsion PCR	Ligase (nonamers)
HeliScope	Single molecule	Polymerase (asynchronous extensions)

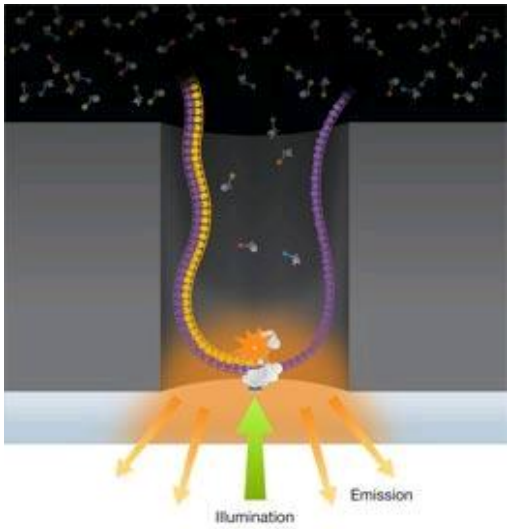
	Cost per megabase	Cost per instrument	Paired ends?	1° error modality	Read-length
454	~$60	$500,000	Yes	Indel	250 bp
Solexa	~$2	$430,000	Yes	Subst.	36 bp
SOLiD	~$2	$591,000	Yes	Subst.	35 bp
Polonator	~$1	$155,000	Yes	Subst.	13 bp
HeliScope	~$1	$1,350,000	Yes	Del	30 bp

Adapted from a presentation by: FRANCIS OUELLETTE Informatics on High Throughput Sequencing Data July 2009

Third Generation Technologies

- Single Molecule Sequencing
  - Pacific Biosciences
  - Life Technologies
  - Helicos

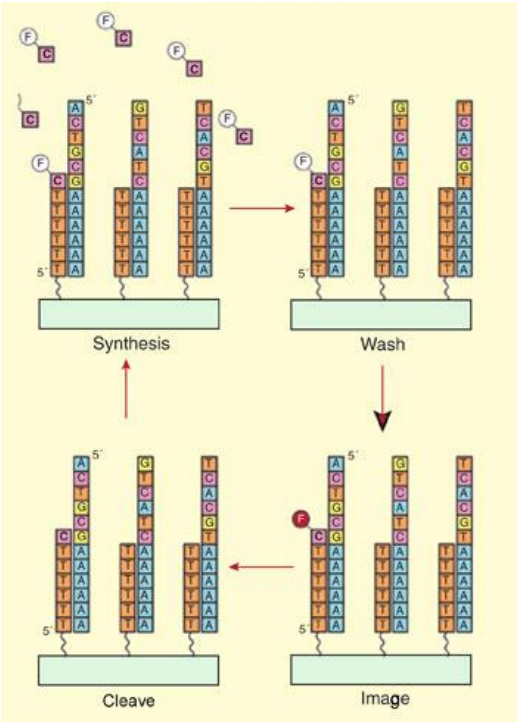
PacBio – ZMW



- Single strand of DNA
- Measured with Fluorescence

- Long Reads (50bp-10kb)
- Inexpensive
- Higher error rates, possibly overcome with redundancy

Helicos – Anchored single molecules

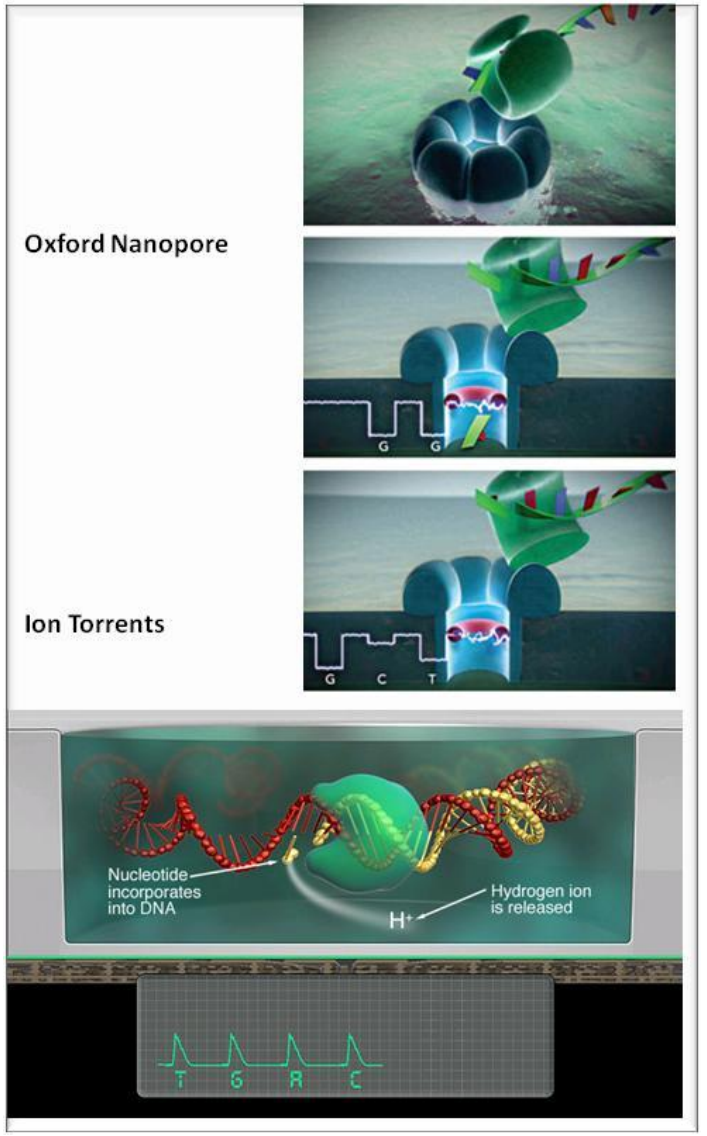


Fourth Generation Technologies (12-18 months)

- Single Molecule Sequencing using “solid state” strategies
  - Oxford Nanopore
  - Ion Torrent
  - Genia
  - gnuBIO
  - NABsys

- Single strand of DNA
- Measured <u>without</u> fluorescence

- Long Reads (50bp-10kb)
- Inexpensive
- Lower error rates, possibly overcome with redundancy
- Unproven



ignite

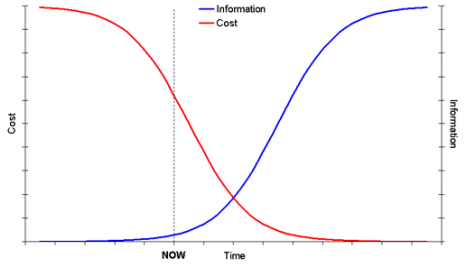
Non-Trivial Assembly and Interpretation

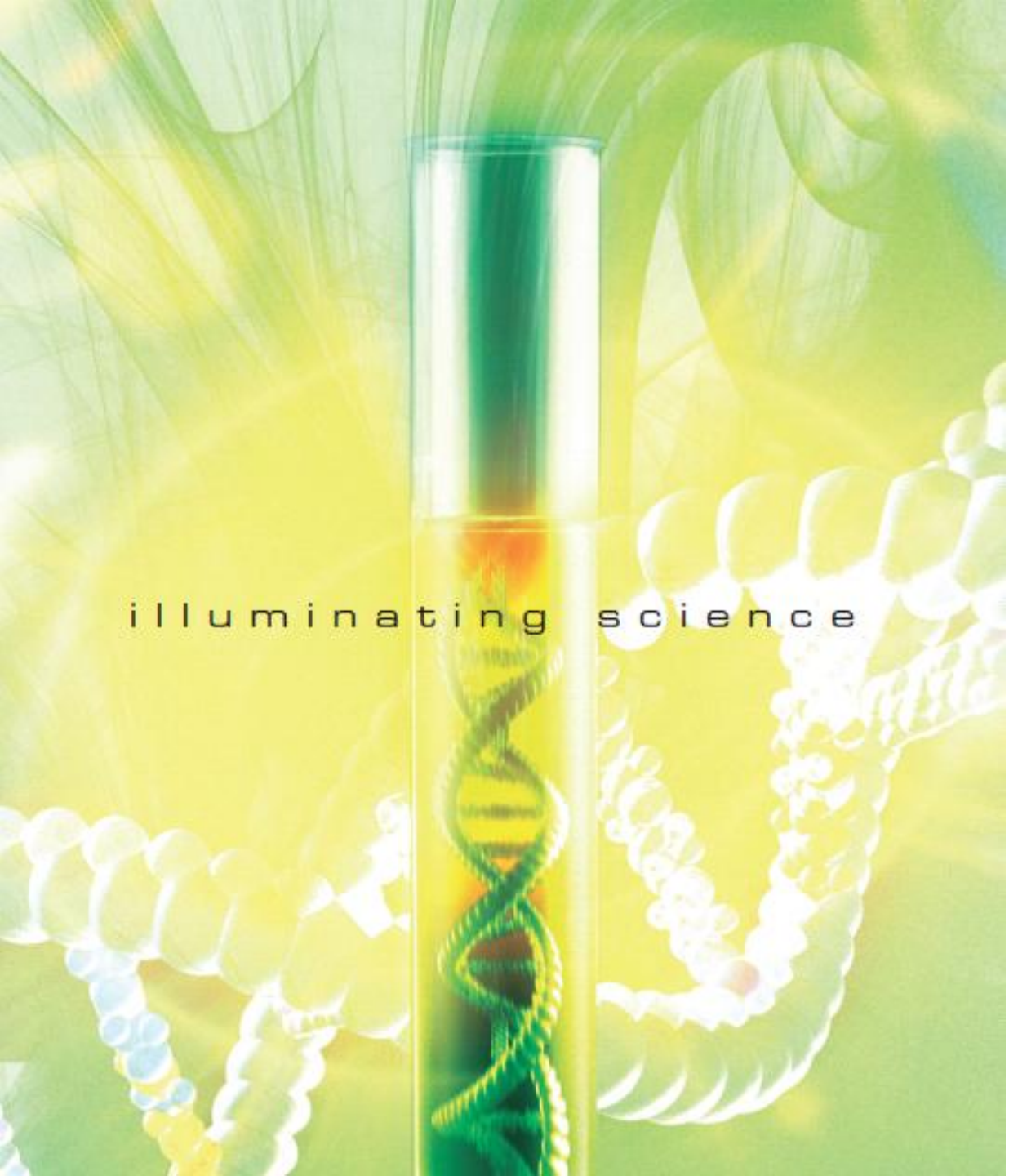
Collapse output into just a variant file for delivery to interpretation engine and/or professional

Delivery of Genome to Point of Care and Clinical Interpretation

With a health care professional using medical genetics principles

- Navigenics
- Omicia
- Knome
- PGP





The Institute for Individualized Health™

Leveraging the Affordable and Accurate Genome Across a Lifetime to Improve health

Dietrich A. Stephan, PhD
President and CEO

dstephan@igniteinstitute

22nd Annual SACGHS Meeting
June 15, 2010