An introduction to next-generation sequencing

Laura MacConaill PhD
PROFILE, Department of Pathology, Brigham and Women's Hospital, Dana-Farber Cancer Institute

Assistant Professor of Pathology, Harvard Medical School
Boston, MA USA

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

- I have no disclosures to report

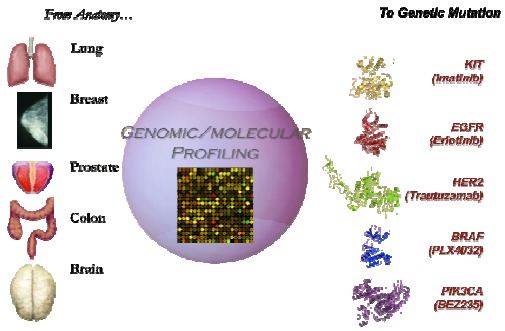
Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

"Personalized" or Precision Cancer Medicine

Principle 1: Molecular pathways involved in tumor survival/progression are often activated by genetic alterations.

Principle 2: Targeted therapies are available for some of these alterations



Match a patient and a driver alteration to a therapy rather than treating solely based on site of origin or histological/pathology tests

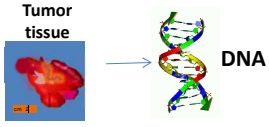
Principle 3: **Genomic technologies** enable comprehensive profiling in the clinic

Dana-Farber Cancer Institute

Garraway, *J. Clin. Oncol.*, 2013

BWH BRIGHAM AND WOMEN'S HOSPITAL

Current Molecular Diagnostic Strategies in the Clinic



Older technologies

Newer technologies

Point mutations (substitutions/indels)

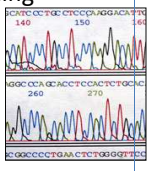
Many oncogenes (EGFR, KIT, BRAF, KRAS…)

Capillary sequencing

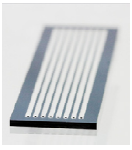
Pyrosequencing

Quantitative PCR

Genotyping



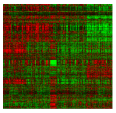
Massively parallel sequencing



454, Illumina, SOLiD, Helicos, Ion Torrent, PacBio….

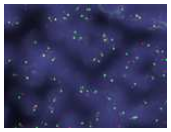
Chromosomal aberrations (copy number gains or losses)

FISH
IHC
Array CGH
SNP array



Translocations, fusion genes

Karyotyping
FISH



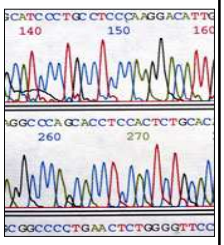
Dana-Farber Cancer Institute

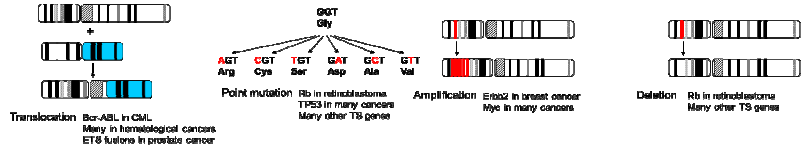
MacConaill *et al.*, 2011 Cancer Discovery

BWH BRIGHAM AND WOMEN'S HOSPITAL

Limitations of older technologies

- Not very sensitive (eg Sanger LOD ~20%)
- Sanger sequencing is analog- composite of signals at a point
- Often test one gene, one alteration, or a small number of nucleotides
- Single gene tests are sequential, time-consuming, use a lot of material
- FFPE material yields short reads
- Cancer genome analysis was *expensive, and limited*

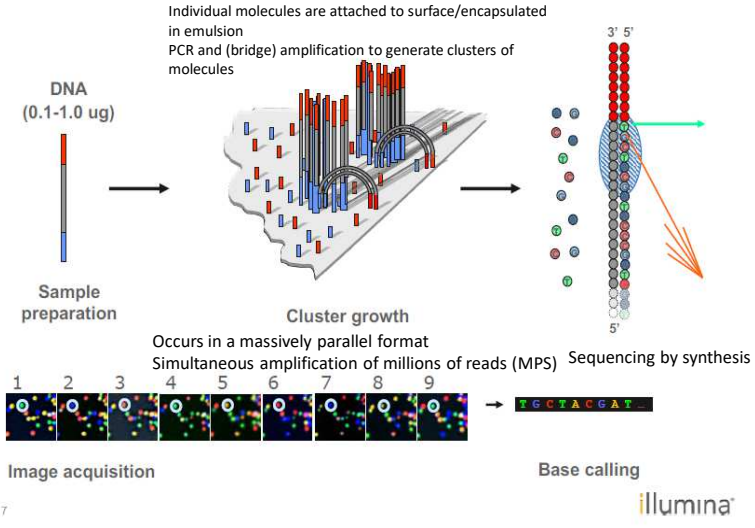




Dana-Farber Cancer Institute

BRIGHAM AND WOMEN'S HOSPITAL

Next Generation or Massively Parallel Sequencing (NGS/MPS)

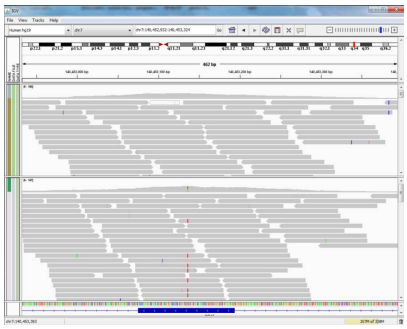


Dana-Farber Cancer Institute

BRIGHAM AND WOMEN'S HOSPITAL

Assembly of sequencing reads

- Sequencing yields billions of reads per run
- Reads are short fragments of sheared DNA (~150bp), paired end
- Reads are assembled/stitched together against a reference sequence
- Reads pile up at a locus (coverage)
- Confidence of call depends on coverage
- Harder to assemble insertions/deletions/structural rearrangements



Dana-Farber Cancer Institute

BRIGHAM AND WOMEN'S HOSPITAL

Flavors of NGS

	Library construction	Sequencing mechanism	Detection mechanism	Maximum read length	Error mode
First generation					
ABI Sanger	bacterial cloning	Dideoxy chain termination	Fluorescence	900 bp	End-of-read errors
Massively parallel sequencing- part 1					
Roche 454	Emulsion PCR on microbead surface	Polymerase-mediated incorporation of unlabelled nucleotides	Photon detection (light)	700 bp	Errors in homopolymer runs
Illumina HiSeq	Amplification on glass surface	Polymerase-mediated incorporation of fluorescent nucleotides	Fluorescence	250 bp	End-of-read errors
Life Technologies SOLiD	Emulsion PCR on microbead surface	Ligase-mediated addition of 2 base encoded fluorescent oligonucleotides	Fluorescence	75 bp	End-of-read errors
Massively parallel sequencing- part 2	Some feature of "3rd gen"- single molecule, more direct detection of signal, or both				
Helicos	NA (single molecule detection)	Polymerase-mediated incorporation of fluorescent nucleotides	Fluorescence	32 bp	End-of-read errors
Thermo Fisher Ion Torrent	Emulsion PCR on microbead surface	Polymerase-mediated incorporation of unlabelled nucleotides	Ion sensing by semiconductors	200 bp	Errors in homopolymer runs
Pacific Biociences	NA (single molecule detectio	Polymerase-mediated incorporation of fluorescent nucleotides	Fluorescence in real-time	>10,000bp	Random errors
Oxford Nanopore	NA (single molecule detectio	Depolymerization and cleavage of individual nucleotides	Ion sensing by nanopore in electrically resistance membrane bilayer	~50 bp	Errors generated by slipping or skipping of DNA

Dana-Farber Cancer Institute

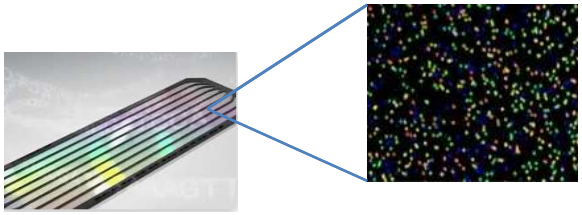
MacConaill, J Clin Oncol, 2013

BRIGHAM AND WOMEN'S HOSPITAL

What are the benefits of NGS?

Moving to simultaneous detection of nucleotides from arrayed, amplified DNA products that originate from single DNA molecules –

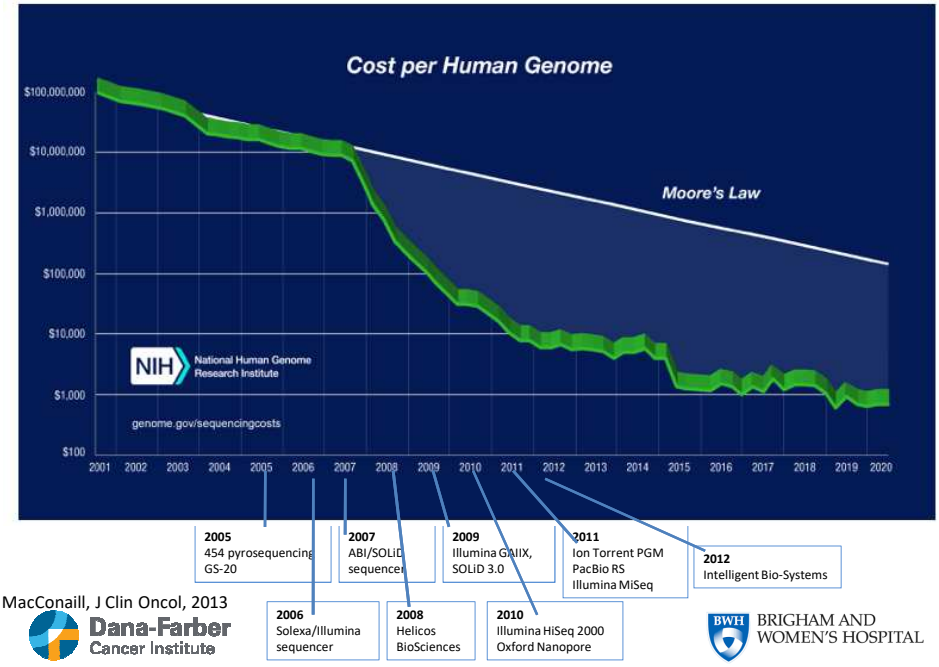
1. Cost has decreased rapidly
2. Increase in sensitivity and scalability
3. Can detect many types of alterations simultaneously
4. Enables new applications e.g. tumor mutational burden, mutational signatures, single cell sequencing, cell free tumor DNA etc



Dana-Farber Cancer Institute

BRIGHAM AND WOMEN'S HOSPITAL

1. Cost has decreased rapidly



MacConaill, J Clin Oncol, 2013

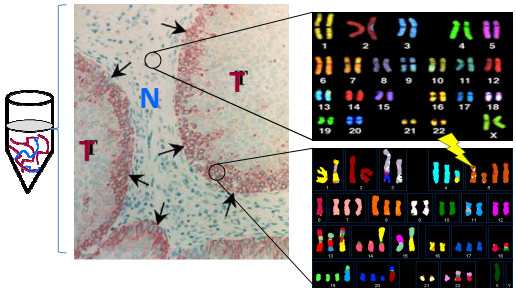
Dana-Farber Cancer Institute

BRIGHAM AND WOMEN'S HOSPITAL

2. Increase in *sensitivity* and scalability – important in cancer

By sequencing in massively parallel format, can sequence many more genes and sequence deeper
The ability to detect mutations depends on the **coverage** (number of reads) and **mutation allelic fraction** (the expected fraction of reads that support a mutation)
Whole genome ~30X, whole exome 50-150X, targeted sequencing >200X, cfDNA 5000X (lower AF)

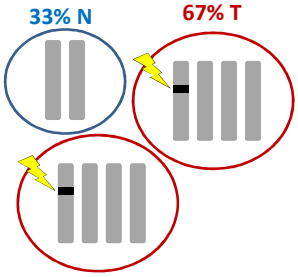
DNA from tumors is a mix of normal genomes and complex tumor genomes



Purity
% tumor cells

Ploidy
mass of DNA in tumor cells

Gaddy Getz, Scott Carter et al.



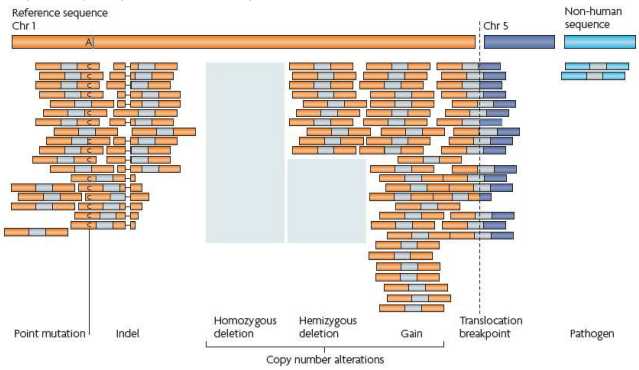
Purity = 67%
Absolute copy number in tumor = 4
Mutation multiplicity = 1
➔ Allelic fraction = 2/10 = 0.2

Detection of resistance mutations, low abundance drivers, subclones etc

Dana-Farber Cancer Institute

3. Can detect many types of alterations- enables *comprehensive* analysis of cancer genomes

Meyerson, Gabriel, Getz, Nat. Rev. Genetics, 2010

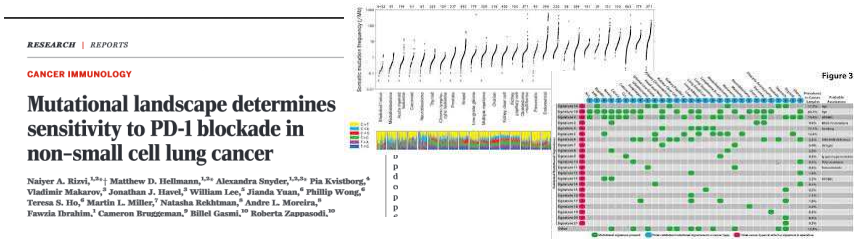


- Rare/common mutations in all exons (not just high frequency, recurrent)
- Can identify other genetic aberrations - amplifications and deletions, translocations/ fusion genes, epigenetic alterations (ChipSeq), RNA Seq (transcript expression, allele-specific expression, splicing..)

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

4. Enables new applications- tumor mutational burden, mutational signatures, single cell sequencing, cfDNA etc



RESEARCH | REPORTS

CANCER IMMUNOLOGY

Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer

Naiyer A. Rizvi,[1,2]* Matthew D. Hellmann,[1,2]* Alexandra Snyder,[1,3,5]* Pia Kvistborg,[4] Vladimir Makarov,[3] Jonathan J. Havel,[3] William Lee,[6] Jianda Yuan,[6] Phillip Wong,[6] Teresa S. Ho,[6] Martin L. Miller,[7] Natasha Rekhtman,[8] Andre L. Moreira,[8] Fawzia Ibrahim,[1] Cameron Bruggeman,[9] Billel Gasmi,[10] Roberta Zappasodi,[10]

Figure 3

June 2019: FDA approved pembrolizumab for treatment in patients with tumors (high) TMB: >10 mutations/Mb.
First pan-cancer approval
HRD signature → response to PARP inhibition

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

Considerations in using NGS for cancer genomics

It all depends on what you are looking for

(point mutation, structural change, copy number, chromosomal conformation, all?)

A. Type of input material: DNA, RNA, chromatin
B. What you target (whole genome, WES, transcriptome, other)
C. How you target (*hybrid capture*, PCR)

Analytical/bioinformatic considerations for all of these

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

Whole Genome Sequencing

Sequence 3 billion base pairs
Average coverage 50X
Can detect copy number alterations, structural rearrangements and non-coding alterations
Most expensive
Most non-coding variants are uninterpretable

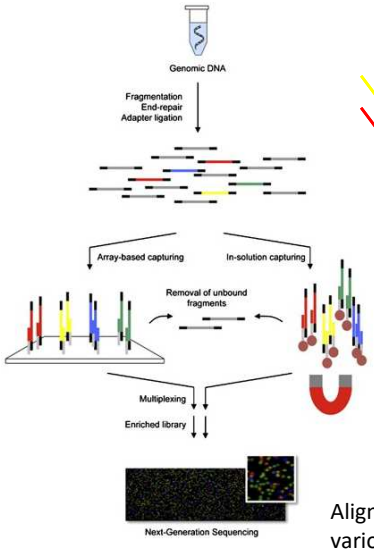
RNA Sequencing

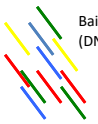
Sequence expressed mRNA regions of genome
Can detect changes to gene expression and splicing
Need the right reference tissue, good reference sequence, good pipeline and analysis

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

Targeted Sequencing- hybrid capture





Baitset of regions of interest (DNA or RNA)

Regions of interest can be:
Whole exome (WES)
Cancer specific (OncoPanel, MSK-IMPACT, Foundation One etc etc)

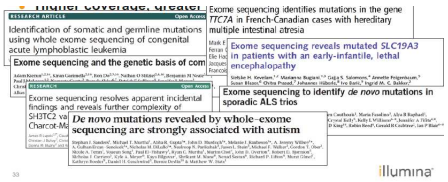
Align to reference genome; apply algorithms to call various types of alterations

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

Whole Exome Sequencing (WES) & targeted cancer panels

- Exons are about 1.5% of human genome
- 85% of disease-causing mutations in exons
- Sequence only coding regions of genome
- Reduced analysis compared to WGS
- Higher coverage, greater confidence of calls (at lower allele fractions)



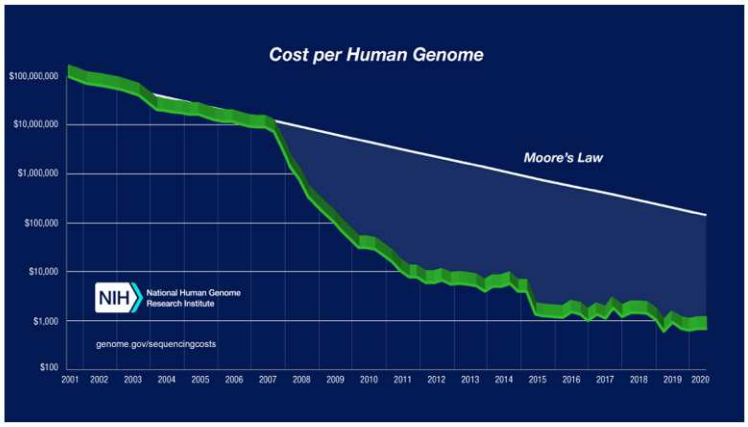
- Targeted cancer panels further reduce the real-estate investigated
- Can sequence deeper (~300-1000X)– detection of lower AF variants
- Reduced cost and analysis compared to WES
- Many academic cancer centers have developed targeted cancer panels

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

Precision cancer medicine enabled by genomic technologies

Massively parallel sequencing has enabled cost-effective profiling of genomes, exomes, and targeted exomes: real world examples



Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

Comprehensive reference genomes by NGS

- Any individual genome differs from a reference genome by 4-5M SNPs
- "SNP" minor allele must have frequency of >1% in a population

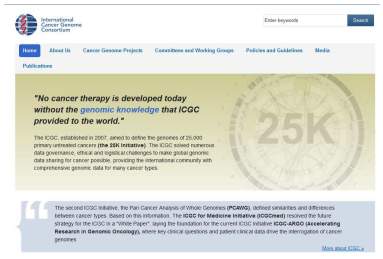
- Large numbers of human references genomes are needed for population studies, ancestry mapping, identification of monogenic/polygenic causes of rare diseases
- More populations sampled, the better we can filter population SNPs that do not contribute to disease
- Several international consortia developed to create available knowledgebases

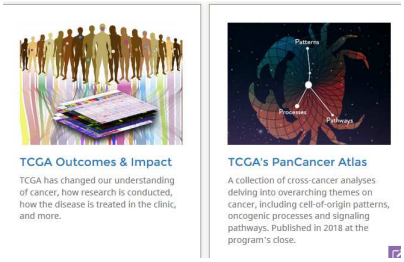
- ExAC (Exome Aggregation Consortium) >60,000 exomes
- gnomAD (Genome Aggregation Database) ~125K exomes, ~16K genomes
    - (build GRCh38 has >76K genomes)

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

Comprehensive *cancer* genome profiling enabled by NGS





- **ICGC and TCGA- large international sequencing projects across over 10 years, >20 cancer types, >12,000 tumors sequenced- DNA, RNA….**
- Characterized DNA mutations, indels, fusions, copy number alterations and other complex structural variations.
- Characterized gene expression, epigenetics, protein expression etc., implicating different functional consequences.
- Described functional pathways and mutational signatures
- Enabled new biomarker trials, subtyping of cancers, links to prognosis etc

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

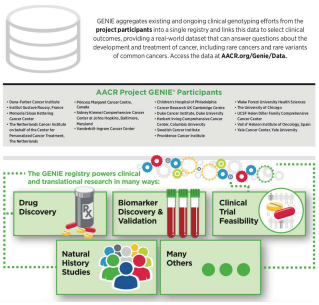
Profile: Precision Medicine in Oncology

- Many major academic cancer centers worldwide have implemented targeted NGS sequencing for their patients
- Profile is a joint initiative between Dana-Farber Cancer Institute, Brigham and Women's Hospital and Boston Children's Hospital launched in 2011
- "Enterprise-level" genomic testing- collect and analyze tumor genomic data from **all** patients, performed in CLIA environment; OncoPanel 447 genes, ~60 regions for SV detection
- Data reported to clinician and stored in a research knowledgebase, enable linkage of specimen results to clinical arm of cohort study
- Research enabled by cbioportal, OncDRS
- Automated trial matching (MatchMiner), molecular tumor boards –increase patient impact
- 50,000 cases completed in past 10 years

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

Comprehensive cancer genomic profiling enabled by NGS



Project GENIE

- Initiated in 2014, 8 initial members
- Created to enable sharing of genomic data across institutions
- Data sharing between 18 of the leading cancer centers in the world- all using NGS
- Publicly accessible, international cancer registry
- Genomic, pathologic and (some) clinical data
- Over 100,000 samples sequenced
- Power in numbers: decision making on rare cancers, rare variants in common cancers
- Will enable new discoveries, clinical trials, drug development

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL

Acknowledgements

CAMD, Profile
Elizabeth Garcia
Yonghui Jia
Danielle Manning
Christine Bertrand
Charles DiRienzo
Amanda Jastram
Mei Lin
Mike Liu
Andrew Miller
Jeffrey Sanborn
David Szeto
Angelina Topalidou
Autumn Duchesne
Eric Reed
Courtenay Brown

Bioinformatics
Athena Petrides
Matthew Ducar
Priyanka Shivdasani
Neil Patel
Phani Davineni
Monica Manam
Leah Bialic
Ethan Cerami
Caitlin Perry

CCG
Aaron Thorner
Paul van Hummelen
Anwesha Nag
Mike Slevin
Matthew Meyerson

Pathology
Jeff Golden
Lynette Sholl
Fei Dong
Jonathan Nowak
Janina Longtine
Jacqueline Bruce
Vanesa Rojas-Rodilla
Alanna Church
Michele Baltay

Leadership
Neal Lindeman
Laura MacConaill
Bruce Johnson
Barrett Rollins

Dana-Farber Cancer Institute

BWH BRIGHAM AND WOMEN'S HOSPITAL