Cancer “genomics” – Technological opportunities in cancer biology and management

Joe W. Gray, PhD
Lawrence Berkeley National Laboratory

A brief survey of cancer genomics with illustrations of current technologies that are driving the field – how can we contribute?
Topics for discussion

- What goes wrong during cancer development?
- How can we understand the details?
- How can we use the information to improve cancer treatment?

Opportunities for technology
Topics for discussion

- What goes wrong during cancer development?
- How can we understand the details?
- How can we use the information to improve cancer treatment?

Opportunities for technology
The “hallmarks of cancer”

Unchecked proliferation
Genome instability
Increased angiogenesis
Increased motility
Ability to bind and proliferate in a foreign environment

Hanahan & Weinberg, Cell, 2000
This happens through deregulation of complex regulatory pathways.

We have evidence that this may involve as many as 4000 genes in breast cancer alone.
Topics for discussion

• What goes wrong during cancer development?
• How can we understand the details?
• How can we use the information to improve cancer treatment?

Opportunities for technology
Cancer is a disease of the genes
The normal genome is encoded in $3 \times 10^9$ bp of DNA packaged into the nucleus of a cell.

Genes can be deregulated by aberrations involving:
- DNA sequence
- Copy number or expression
- Genome organization

http://www.accessexcellence.org/AB/GG/chromosome.html
The normal genome is encoded in 3x10^9 bp of DNA packaged into the nucleus of a cell.

Genes can be deregulated by aberrations involving:
- DNA sequence
- Copy number or expression
- Genome organization

Some studies suggest the mutation rate may be as high as 10^-5/bp/cell (i.e. 10^4 mutations per cell).

Cancer genomes need to be scanned at the DNA sequence level to discover the mutation subset that deregulates critical genes and to ID therapy targets.

A $100M cancer genome sequencing project is underway. May expand to $1B if successful.
Current sequencing technology

Analysis by gel electrophoresis
Sequencing Production at the Joint Genome Institute ('05)

- Current cost: $0.1/base = $10^8/genome
- The NIH has an RFA calling for technologies capable of sequencing at $1000/genome
- How is the 5-order of magnitude increase possible?
Sequencing one molecule at a time

**FIGURE 8**

- **DNA Library Preparation and Titration**
  - 4.5 HOURS
  - Anneal sstDNA to an excess of DNA Capture Beads

- **emPCR**
  - 10.5 HOURS
  - Emulsify beads and PCR reagents in water-in-oil microreactors

- **Sequencing**
  - 8 HOURS
  - Clonal amplification occurs inside microreactors

- 4.5 HOURS
  - Break microreactors enrich for DNA-positive beads

**gDNA**

- sstDNA Library

454 Technology
Sequencing one molecule at a time

**Figure 9**

DNA Library Preparation and Titration
- 4.5 hours

emPCR
- 10.5 hours
- 8 hours

Sequencing
- 4.5 hours

- Well diameter: average of 44µm
- 200,000 reads obtained in parallel
- A single cloned amplified ssDNA bead is deposited per well

Amplified ssDNA library beads

Quality filtered bases

454 Technology
Sequencing one molecule at a time

**FIGURE 10**

- DNA Library Preparation and Titration
  - 4.5 HOURS
- emPCR
  - 10.5 HOURS
- Sequencing
  - 8 HOURS
  - 4 bases (TACG) cycled 42 times
  - Chemiluminescent signal generation
  - Signal processing to determine base sequence and quality score

**Amplified ssDNA library beads**

- DNA Capture Bead containing millions of copies of a single clonal fragment
- Sulfurylase
- Luciferase

**Quality filtered bases**
There are many variations on “polony” sequencing

• Future work will involve
  – Increasing polony density
  – Decreasing reagent costs
  – Improving sequencing fidelity
The normal genome is encoded in $3 \times 10^9$ bp of DNA packaged into the nucleus of a cell.

Genes can be deregulated by aberrations involving:
- DNA sequence
- Copy number or expression
- Genome organization

http://www.accessexcellence.org/AB/GG/chromosome.html
Staining DNA with DNA

Labeled DNA

Denatured DNA
Scanning for genomic aberrations that alter gene expression

Normal
Tumor

Array elements

DNA
RNA

Genome copy number profile
Expression level profile
Array manufacture and readout is key

- Array manufacture
  - Spotted DNA
  - Photolithography
  - Micromirror based synthesis
- Hybridization efficiency
- Detection
- Issues
  - Array “probe” density
  - Cost
  - Amount of material required
Scanning the cancer genome – A “typical” breast cancer

**Normal DNA**
- Decreased gene copy

**Tumor DNA**
- Increased gene copy

**DNA**
- Log2 copy num.
  - 3
  - 2
  - 1
  - 0
  - -1
  - -2

**Genome location**
Recurrent aberrations in breast cancers – markers for poor outcome.

High level amplifications are associated with reduced survival.
Technologies are needed to read out at higher resolution and lower cost while using less material.
Array synthesis using photolithography
Current arrays have 8 μm features (>2 million features)

Five micrometer features would generate >6 million features
Synthesis one pixel at a time
Electrostatic hybridization to improve efficiency and specificity

- Discriminates single base pair mismatch
Electrochemical detection to facilitate readout?

![Diagram of electrochemical detection process]

More powerful technologies are clearly possible.

The current challenge is cost ($150/array).
The normal genome is encoded in $3 \times 10^9$ bp of DNA packaged into the nucleus of a cell.

Cancers can arise from aberrations involving:
- DNA sequence
- Copy number or expression
- Genome organization

http://www.accessexcellence.org/AB/GG/chromosome.html
Staining DNA with DNA

Labeled DNA

Denatured DNA
Fluorescence in situ hybridization
Technical need - Advanced microscopy to assess 3D organization

- Currently assessed using confocal microscopy
- Resolution limited to ~0.2 µm
- Limited temporal resolution
- Multicolor analysis
- Software for 3D visualization
Scanned ion probe mass spectrometry

**Characteristics**
- Ion transmission ~ 100%
- Ion microprobe and ion microscope
- Mass range: unlimited
- Mass resolution: \( m/\Delta m \approx 8000 \)

**Secondary ion formation depends on:**
- Primary ions
- Matrix effects
- Sample preparation

---

**Diagram:**
- Liquid Metal Ion Gun (\( \text{Ga}^+ \))
- Spot size ~150nm, \( \tau = 300 \text{ psec} \)
- Three 90° electrostatic energy analyzer (ESA)
ToF-SIMS image analysis of a prostate cancer tissue section

Cancerous, fixed prostate tissue

David Seligson, MD

$M/z = 221$

Total ion image

$M/z = 184$
(phosphocholine head group)
Super-resolution microscopy
Stimulated emission depletion

(a) STED principle

(b) Saturated depletion of state A

(c) STED microscope

SW Hell

Current Opinion in Neurobiology
Topics for discussion

- What goes wrong during cancer development?
- How can we understand the details?
- How can we use the information to improve cancer treatment?

*Opportunities for technology*
The details of gene deregulation differ between individual cancers – even cancers that appear the same to a pathologist.

Detailed molecular analyses may identify molecular features that will predict tumor behavior including response to therapy.
Assessing drug responses

- Identify cell lines with drug target present
- Apply genome analysis technologies to identify molecular features of each cell line
- Measure responses
- Correlate pre-treatment molecular features with cellular responses to identify response predictors

_Hundreds of drugs are now in the development pipeline_  
_We need to know who will respond before we evaluate them in patients_
An in vitro system for assessment of function or Rx response

• ~60 breast cancer cell lines in 2D and 3D culture
• Molecular profiling
  – DNA, RNA, methylation, protein
  – DNA sequence
• Semi-automated cell culture
• High content imaging
  – Apoptosis
  – Motility
  – Proliferation
  – Protein localization

Supported by the NIH, Genentech, GalaxoSmithKline and Affymetrix
Technological opportunities

Current system is too expensive and slow to test thousands of compounds – Microfluidics and detectors (e.g. Luke Lee at UCB)
We need to be able to “see” how gene deregulation affects the signaling pathways.
Cells are not “bags of chemicals”
Super-resolution microscopy for real time assessment of signal propagation

(a) STED principle

(b) Saturated depletion of state A

(c) STED microscope

SW Hell
Conclusions

- Technologies now exist to interrogate the sequence, copy number, structure and expression of essentially all genes.
- This information is driving the development of individualized medicine.
- Challenges now are to reduce cost, increase analysis speed and enable analysis “in tissue context”.