Cancer Genomes

02-223 Personalized Medicine: Understanding Your Own Genome
Fall 2014
The Economic Burden of Cancer

- The economic cost of cancer exceeds that of any other disease
  - Only accounts for disability and premature death, no direct medical costs
Overview

• What is cancer and how does it develop?

• Genetics of cancer

• Finding cancer mutations

• Personalized medicine in cancer (next class)
Development of Cancer Cells

• Types of cancers
  – Carcinomas: cancers arising from epithelial cells
  – Sarcomas: cancers arising from connective tissue or muscle cells
  – Leukemias and lymphomas: cancers derived from white blood cells and their precursors

• Agents that trigger carcinogenesis
  – Chemical carcinogens (causes local DNA alterations)
  – Radiation such as x-rays (causes chromosome breaks and translocations), UV light (causes DNA base alterations)
  – Viruses: Hepatitis-B, Hepatitis-C virus for liver cancer
Tumors

- Cancer cells
  - Reproduce in defiance of the normal restraints on cell growth and division
  - Invade and colonize territories normally reserved for other cells
Defective Control of Cell Death and Differentiation in Cancer Cells

• Both increased cell division and decreased apoptosis (cell death) can contribute to tumorigenesis
Cancer Progression

a) Primary tumour

b) Proliferation/angiogenesis

c) Detachment/invasion

Lymphatics, venules, capillaries
Cancer Progression

e Extravasation

Adherence to vessel wall

Arrest in organs

Transport

Lung

Heart

f Proliferation/angiogenesis

Metastasis

Establishment of a microenvironment
Pathways of Tumorigenesis

<table>
<thead>
<tr>
<th>Component</th>
<th>Acquired Capability</th>
<th>Example of Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>🟠</td>
<td>Self-sufficiency in growth signals</td>
<td>Activate H-Ras oncogene</td>
</tr>
<tr>
<td>🟢</td>
<td>Insensitivity to anti-growth signals</td>
<td>Lose retinoblastoma suppressor</td>
</tr>
<tr>
<td>🟤</td>
<td>Evading apoptosis</td>
<td>Produce IGF survival factors</td>
</tr>
<tr>
<td>🟦</td>
<td>Limitless replicative potential</td>
<td>Turn on telomerase</td>
</tr>
<tr>
<td>🟧</td>
<td>Sustained angiogenesis</td>
<td>Produce VEGF inducer</td>
</tr>
<tr>
<td>🟩</td>
<td>Tissue invasion &amp; metastasis</td>
<td>Inactivate E-cadherin</td>
</tr>
</tbody>
</table>
Overview

• What is cancer how does it develop?

• Genetics of cancer

• Finding cancer mutations

• Personalized medicine in cancer (next class)
Germline Mutations

- Any detectable and heritable variation in the lineage of germ cells
- Mutations in these cells are transmitted to offspring
- So far, we focused on germline mutations for complex diseases
Somatic Mutations

- Mutations in non germline cells
- These mutations are not transmitted to offsprings
- Cancer can be caused by both germline and somatic mutations
Cancer Progression
Cancer Progression

- Fertilized egg
- Gestation
- Infancy
- Childhood
- Adulthood
- Early clonal expansion
- Benign tumour
- Early invasive cancer
- Late invasive cancer
- Chemotherapy-resistant recurrence

Intrinsic mutation processes

Environmental and lifestyle exposures

Mutator phenotype

Chemotherapy
Cancer Progression
Cancer Progression

- Fertilized egg
- Gestation
- Infancy
- Childhood
- Adulthood
- Early clonal expansion
- Benign tumour
- Early invasive cancer
- Late invasive cancer
- Chemotherapy-resistant recurrence

Intrinsic mutation processes
Environmental and lifestyle exposures
Mutator phenotype
Chemotherapy
Cancer Mutations

- Mutations in genome
- Mutations in epigenome
Epigenome

- Two-meter long DNA sequences are packaged into each cell
Epigenetics: Histone Modification

Nucleosome consists of eight histone proteins
Epigenetics: Histone Modification

[Diagram showing the process of chromatin to nucleosome and histone modifications with Acetylation, Methylation, and Phosphorylation]
Histone Modification
Epigenetics

Gene "switched on"
- Active (open) chromatin
- Unmethylated cytosines (white circles)
- Acetylated histones

Gene "switched off"
- Silent (condensed) chromatin
- Methylated cytosines (red circles)
- Deacetylated histones

Transcription possible

Transcription impeded
Epigenetics: Histone Modification
Methylation

- Methylation of cytosine (C) nucleotides in DNA
Methylation

- Associated with repressing gene expression
  - X-chromosome silencing
  - Promoter regions of expressed genes in a tissue are usually unmethylated
  - Different methylation patterns have been observed in Identical twins – environmental effects
Cancer-Causing Genes

- **Oncogenes**
  - Mutations that confer gain of functions to oncogenes can promote cancer
  - Mutations with growth-promoting effects on the cell
  - Often heterozygous mutation is enough to make cells cancerous. Why?

- **Tumor suppressor genes**
  - Mutations that confer loss of function can contribute to cancer
  - Typically homozygous mutation is required to make cells cancerous. Why?

- **DNA maintenance genes**
  - Indirect effects on cancer development
Mutations in Tumor Suppressor Genes

- Tumor suppressor gene
- Mutation
- Heritable gene silencing in condensed chromatin
- Genetic change
- Epigenetic change
Mutations in Oncogenes

- Deletion or point mutation in coding sequence
  - DNA: hyperactive protein made in normal amounts
  - RNA

- Regulatory mutation
  - DNA: normal protein greatly overproduced
  - RNA

- Gene amplification
  - DNA: normal protein greatly overproduced
  - RNA

- Chromosome rearrangement
  - DNA: nearby regulatory DNA sequence causes normal protein to be overproduced
  - RNA: fusion to actively transcribed gene produces hyperactive fusion protein

proto-oncogene
Replication of DNA Damages

1. Cell enters S phase and replicates its DNA despite unpaired strand break.
2. One daughter cell inherits a chromosome lacking a telomere.
3. Cell enters S phase and replicates its DNA.
4. Sister chromatid ends that lack telomeres fuse.
5. Fused sister chromatids are pulled apart at mitosis, creating breakage at new site.
6. One daughter cell inherits a chromosome with duplicated genes but again lacking a telomere.

Breakage-Fusion-Bridge Cycle
Overview

• What is cancer how does it develop?

• Genetics of cancer

• Finding cancer mutations

• Personalized medicine in cancer (next class)
How to Find Cancer Causing Mutations

• Germline cancer mutations?

• Somatic cancer mutations?
How to Find Cancer Causing Mutations

• Germline cancer mutations?
  – Genotype data are collected for a large number of normals and cancer patients (case/control studies)

• Somatic cancer mutations?
  – Genotype data are collected for blood (normal) and cancer cells for each cancer patient, for a large number of cancer patients.
  – What is the challenge?
Driver and Passenger Mutations

- Fertilized egg
- Gestation
- Infancy
- Childhood
- Adulthood
- Early clonal expansion
- Benign tumour
- Early invasive cancer
- Late invasive cancer
- Chemotherapy-resistant recurrence

Intrinsic mutation processes

- Passenger mutation
- Driver mutation
- Chemotherapy resistance mutation

Environmental and lifestyle exposures

- Mutator phenotype
- Chemotherapy

- 10s–1,000s of mitoses depending on the organ
- 10s–100s of mitoses depending on the cancer
- 10s–100,000 or more passenger mutations
Driver and Passenger Mutations

• Driver mutations
  – Causally implicated in oncogenesis
  – Gives growth advantage to cancer cells
    • E.g., mutations that de-activate tumor suppressor genes
    – positively selected in the microenvironment of the tissue

• Passenger mutations
  – Somatic mutations with no functional consequences
  – Does not give growth advantage to cancer cells
  – However, the passenger mutations are propagated to daughter cells just because they co-exist with the driver mutation in the same cell

• Key Scientific Question: How can we distinguish between driver and passenger mutations?
Identifying Driver Mutations

• Compare the tumor genome with the normal genome of the same individual

• Compare the tumor genome with reference genome

• Other known DNA polymorphisms

• Signatures of driver mutations
  – Frequently observed mutations across tumors.
  – Mutations that cluster in subset of genes. Passenger mutations are more randomly distributed across genomes
Challenges

• Challenges unique to cancer genome analysis
  – Sequence alignment and assembly can be significantly more challenging because of highly rearranged chromosomes and high variation across cancer genomes
  – Somatic mutation calling is more challenging
    • the impurity of the sample
      – Normal genomes have allele copies of 0, 1, or 2
      – Cancer genomes can have allele copies of fractions of 0, 1, or 2
    • Most somatic mutations are rare

• Different cancer types have different rates of mutations. Mutator phenotype may or may not present.

• Infrequently occurring driver mutations are hard to identify.
Methods for Detecting Driver Mutations

- **SIFT**
  - A tool that predicts whether an amino acid substitution affects protein function
  - Classifies a substitution into tolerated or deleterious ones

- **PolyPhen**
  - Software for predicting damaging effects of nonsynonymous mutations
## PolyPhen Features

- **Black**: candidates, **blue**: selected

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Values with ranges in HumDiv</th>
</tr>
</thead>
<tbody>
<tr>
<td>nt1</td>
<td>wild type allele nucleotide</td>
<td>A,C,G,T</td>
</tr>
<tr>
<td>nt2</td>
<td>mutation allele nucleotide</td>
<td>A,C,G,T</td>
</tr>
<tr>
<td>site</td>
<td>SITE annotation from UniProt/Swiss-Prot</td>
<td>Yes, No</td>
</tr>
<tr>
<td>region</td>
<td>REGION annotation from UniProt/Swiss-Prot</td>
<td>NO, PROPER, SIGNAL, TRANSMEM</td>
</tr>
<tr>
<td>phat</td>
<td>PHAT matrix element in the TRANSMEM region</td>
<td>[-8.0, 4.0], mean = -0.04</td>
</tr>
<tr>
<td>score1</td>
<td>PSIC score for the wild type allele</td>
<td>[-1.1], mean = 1.07</td>
</tr>
<tr>
<td>score2</td>
<td>PSIC score for the mutant allele</td>
<td>[-1.39, 2.64], mean = .166</td>
</tr>
<tr>
<td>score_delta</td>
<td>difference of PSIC scores (Score1-Score2)</td>
<td>[-3.23, 4.57], mean = .905</td>
</tr>
<tr>
<td>num_observ</td>
<td>number of residues observed at the position of the multiple alignment</td>
<td>[1, 432], mean 69.3</td>
</tr>
<tr>
<td>delta_volume</td>
<td>change in residue side chain volume</td>
<td>[-167, 167], mean = -1.93</td>
</tr>
<tr>
<td>transv</td>
<td>mutation origin by transversion or transition</td>
<td>Yes, No</td>
</tr>
<tr>
<td>CpG</td>
<td>mutation origin in the CpG hypermutable context</td>
<td>Yes, No</td>
</tr>
<tr>
<td>pfam_hit</td>
<td>position of the mutation within/outside a protein domain as defined by Pfam</td>
<td>Yes, No</td>
</tr>
<tr>
<td>id_p_max</td>
<td>congruency of the mutant allele to the multiple alignment</td>
<td>[0, 95.5], mean = 24</td>
</tr>
<tr>
<td>id_q_min</td>
<td>congruency of the closest homologue deviating from wild type allele</td>
<td>[1.56, 95.5], mean 68.76</td>
</tr>
<tr>
<td>cpqVar1Var2</td>
<td>sequence identity with the closest homologue deviating from wild type allele</td>
<td>NO, AA1_AA2</td>
</tr>
<tr>
<td>cpq_transition</td>
<td>whether variant happened as transition in CpG context</td>
<td>No, Transition, Transversion</td>
</tr>
<tr>
<td>charge_change</td>
<td>change in electrostatic charge</td>
<td>0,1,2</td>
</tr>
<tr>
<td>hydroph_change</td>
<td>change in hydrophobicity</td>
<td>[0, 2.85], mean 0.80</td>
</tr>
<tr>
<td>ali_id</td>
<td>alignment length with the closest homolog with known 3D structure</td>
<td>[0, 0.33]</td>
</tr>
<tr>
<td>ali_len</td>
<td>alignment length with the closest homolog with known 3D structure</td>
<td>[0, 1213], mean 130.0</td>
</tr>
<tr>
<td>acc_normed</td>
<td>normalized accessible surface area of amino acid residue</td>
<td>[0, 1.55], mean .35</td>
</tr>
<tr>
<td>sec_str</td>
<td>secondary structure</td>
<td>HELIX, SHEET, OTHER</td>
</tr>
<tr>
<td>map_region</td>
<td>region of the Ramachandran map</td>
<td>ALPHA, BETA, OTHER</td>
</tr>
<tr>
<td>delta_prop</td>
<td>change in accessible surface area propensity</td>
<td>[-2.89, 2.89], mean -0.07</td>
</tr>
<tr>
<td>bfact</td>
<td>crystallographic beta-factor</td>
<td>[-1.85, 5.17], mean 0.0</td>
</tr>
<tr>
<td>hct_cont_ave_num</td>
<td>average number of contact with heteroatoms</td>
<td>Yes, No</td>
</tr>
<tr>
<td>hct_cont_min_dist</td>
<td>minimal distance to a heteroatom</td>
<td>Yes, No</td>
</tr>
<tr>
<td>inter_cont_ave_num</td>
<td>average number of interchain contacts in a protein complex</td>
<td>Yes, No</td>
</tr>
<tr>
<td>inter_cont_min_dist</td>
<td>average minimal interchain distance</td>
<td>Yes, No</td>
</tr>
<tr>
<td>delta_volume_new</td>
<td>change in residue volume for buried residues</td>
<td>[-119, 138], mean -0.5</td>
</tr>
<tr>
<td>delta_prop_new</td>
<td>change in accessible surface area propensity for buried residues</td>
<td>[-1.83, 2.89], mean 0.0026</td>
</tr>
</tbody>
</table>
Four Subtypes of Stomach Cancer Identified

Researchers with the TCGA Research Network have found that stomach cancers, also called gastric cancers or gastric adenocarcinomas, fall into four distinct molecular subtypes.

Learn More ▶

http://cancergenome.nih.gov/
# Cancer Tissues Being Collected for Potential Study

Last Updated: May 16, 2014

## Breast

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Sample Collection Complete</th>
<th>Data Publicly Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast Ductal Carcinoma</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Breast Lobular Carcinoma</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

## Central Nervous System

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Sample Collection Complete</th>
<th>Data Publicly Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glioblastoma Multiforme</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Lower Grade Glioma</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

## Endocrine

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Sample Collection Complete</th>
<th>Data Publicly Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adrenocortical Carcinoma</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Papillary Thyroid Carcinoma</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Paragangioma &amp; Pheochromocytoma</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

## Gastrointestinal

## Gynecologic

## Head and Neck

## Hematologic

## Skin

## Soft Tissue

## Thoracic

## Urologic

## Gastrointestinal
The Cancer Genome Atlas

- Other types of data are also collected along with genome data
  - Tumor gene expressions
  - Tumor microRNA expression
  - Epigenetic data
  - Clinical data
Summary

• The genetic causes of cancer include both heritable germline mutations and somatic mutations

• In cancer genome study, genome data are collected for both normal and cancer cells for the same individual

• The key challenge in studying the genetic causes of cancer is to identify driver mutations from background passenger mutations