Lecture 23:
Cancer Genomics

Slides credit:
Alvin Shi, Nir Hacohen
<table>
<thead>
<tr>
<th>Project</th>
<th>Sets</th>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
<th>Lec</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS1 out on L1-L5</td>
<td>1</td>
<td>Thu</td>
<td>Sep 8</td>
<td>Introduction</td>
<td>L1</td>
<td>Intro: Biology, Algorithms, Machine Learning, Course Overview</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fri</td>
<td>Sep 9</td>
<td></td>
<td>R1</td>
<td>Recitation 1: Biology and Probability Review</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Tue</td>
<td>Sep 13</td>
<td>Module I: Aligning and Modeling Genomes</td>
<td>L2</td>
<td>Alignment I: Dynamic Programming, Local and global alignment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Sep 15</td>
<td></td>
<td>L3</td>
<td>Alignment II: Database search, Rapid string matching, BLAST, BLOSUM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fri</td>
<td>Sep 16</td>
<td></td>
<td>R2</td>
<td>Recitation 2: Deriving Parameters of Alignment, Multiple Alignment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tue</td>
<td>Sep 20</td>
<td></td>
<td>L4</td>
<td>Hidden Markov Models Part 1: Evaluation/Parsing, Viterbi, Forward algorithms</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Sep 22</td>
<td></td>
<td>L5</td>
<td>Hidden Markov Models Part 2: Posterior Decoding, Learning, Baum-Welch</td>
</tr>
<tr>
<td>Frontiers</td>
<td>3</td>
<td>Thu</td>
<td>Sep 23</td>
<td>No classes - student holiday</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS2 out on L6-R4</td>
<td>4</td>
<td>Thu</td>
<td>Sep 27</td>
<td>Module II: Gene Expression and Networks</td>
<td>L6</td>
<td>Expression Analysis: Clustering/Classification, K-means, Hierarchical, Bayesian</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fri</td>
<td>Sep 30</td>
<td></td>
<td>L7</td>
<td>Transcript structure: GenScan, RNA-seq, Mapping, De novo Assembly, Diff Expr</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tue</td>
<td>Oct 4</td>
<td></td>
<td>R3</td>
<td>Recitation 3: Affinity Propagation Clustering and Random Forest Classification</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Oct 6</td>
<td></td>
<td>L8</td>
<td>Epigenomics: ChIP-Seq, Read mapping, Peak calling, IDR, Chromatin states</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fri</td>
<td>Oct 7</td>
<td></td>
<td>L9</td>
<td>Three-dimensional chromatin interactions: 3C, 5C, HiC, HiChIP-Pet</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Thu</td>
<td>Oct 13</td>
<td>Project Intro: about the projects, self introductions, mentor intro, example projects, teamwork 32D-507</td>
<td>R4</td>
<td>Recitation 4: ENCODE, Epigenome Roadmap, ChromHMM, ChromImpute</td>
</tr>
<tr>
<td>Frontiers</td>
<td></td>
<td>Thu</td>
<td>Oct 17</td>
<td>No classes - Columbus Day Holiday</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS3 out on L10-R6</td>
<td>6</td>
<td>Thu</td>
<td>Oct 25</td>
<td>Module III: Gene Regulation and Epigenomics</td>
<td>L10</td>
<td>Regulatory Motifs: Discovery, Representation, PBM, Gibbs Sampling, EM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fri</td>
<td>Oct 28</td>
<td></td>
<td>L11</td>
<td>Networks I: Neural Networks, Belief Networks, Deep learning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tue</td>
<td>Nov 1</td>
<td></td>
<td>L12</td>
<td>Networks II: Bayesian inference, Variational Bayes, approximate inference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Nov 3</td>
<td></td>
<td>R6</td>
<td>Recitation 6: Networks review, Recommendation systems, EHR, HeWAS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Nov 4</td>
<td></td>
<td>L13</td>
<td>Regulatory Motifs: Discovery, Representation, PBM, Gibbs Sampling, EM</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Thu</td>
<td>Nov 10</td>
<td>Project Planning: research areas, initial ideas, type of project, mentor matching, finding partners 32D-507</td>
<td>R7</td>
<td>Recitation 7: Linkage Disequilibrium, Haplotype Phasing, Genotype Imputation</td>
</tr>
<tr>
<td>Frontiers</td>
<td></td>
<td>Thu</td>
<td>Nov 17</td>
<td>No Recitation, Veterans Day</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS4 out on L13-R8</td>
<td>8</td>
<td>Thu</td>
<td>Oct 27</td>
<td>Module IV: Population and Disease Genetics</td>
<td>L14</td>
<td>Disease Association Mapping, GWAS, organismal phenotypes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fri</td>
<td>Oct 30</td>
<td></td>
<td>R7</td>
<td>Recitation 7: Linkage Disequilibrium, Haplotype Phasing, Genotype Imputation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tue</td>
<td>Nov 1</td>
<td></td>
<td>L15</td>
<td>Quantitative trait mapping, molecular traits, eQTLs, mediation analysis, IMWAS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Nov 3</td>
<td></td>
<td>L16</td>
<td>Missing Heritability, Complex Traits, Interpret GWAS, Rank-based enrichment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Nov 4</td>
<td></td>
<td>R8</td>
<td>Recitation 8: Rare Variants, ExAC</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Thu</td>
<td>Nov 10</td>
<td>Panel Discussion: reconciling critiques, strategies for improvement, feedback to author 32D-507</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Nov 17</td>
<td>No Recitation, Veterans Day</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontiers</td>
<td></td>
<td>Thu</td>
<td>Nov 24</td>
<td>Quiz (the only quiz - the class has no final exam)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS5 out on L17-R9</td>
<td>10</td>
<td>Thu</td>
<td>Nov 29</td>
<td>Module V: Comparative Genomics and Evolution</td>
<td>L17</td>
<td>Comparative genomics and evolutionary signatures</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fri</td>
<td>Dec 1</td>
<td></td>
<td>L18</td>
<td>Genome Scale Evolution, Genome Duplication</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>Thu</td>
<td>Nov 15</td>
<td>Midcourse report due Wed 11/23</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 6</td>
<td></td>
<td>L19</td>
<td>Phylogenetics: Molecular evolution, Tree building, Phylogenetic inference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 9</td>
<td></td>
<td>L20</td>
<td>Phylogenetics: Gene/species trees, reconciliation, coalescent, ARGs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td>R9</td>
<td>Recitation 8: Phylogenetic distance metrics, Coalescent Process</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Thu</td>
<td>Dec 13</td>
<td>No more sets! (work on your final project)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td>Written report due Sun 12/11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 20</td>
<td></td>
<td>Update conference report, conference publication format. As part of your final project, comment on your overall project experience. Written report due Sun 12/11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Thu</td>
<td>Dec 13</td>
<td>Conference report slide pres. Talks on Tue 12/13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td>Conference report slide pres. Talks on Tue 12/13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td>Conference report slide pres. Talks on Tue 12/13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thu</td>
<td>Dec 13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• L21: Single-cell genomics
  • Measuring and analyzing biology at the single-cell level

• L22: PheWAS (Phenome-wide associations studies)
  • Multi-phenotype analyses, inferences, association, imputation

• L23: Cancer genomics
  • Mutational heterogeneity, tumor evolution, immune evasion

• L24: Genome engineering & high-throughput biology
  • From reading to writing, CRISPR-Cas9,
Goals for today: Cancer genomics

0. Introduction: oncogenes, tumor suppressors, hallmarks
   - Hallmarks of cancer, tumor-suppressors, proto-oncogenes/oncogenes, mutator phenotype, oncoviruses, and fusion oncogenes.

1. Exome sequencing lessons: recurrence and heterogeneity
   - Recurrence: common signatures across types/patients/tumors/clones
   - Evolutionary dynamics: clonal heterogeneity, computational models

2. WGS lessons: non-coding drivers and convergence
   - Background mutation rate: regions, chromatin states, patients, plexi
   - Convergence: mutations $\rightarrow$ enhancers $\rightarrow$ genes $\rightarrow$ pathways

3. Beyond mutations: epigenomics, functional heterogeneity
   - Epigenomic alterations: reprogramming
   - Single-cell sequencing: functional heterogeneity

4. Tumor immunology, microenvironment, immunotherapy
   - Tumor-microenvironment interactions, cancer immunoediting.
   - Neo-antigens, immune suppression, immunotherapy.
Bob Weinberg and Douglas Hanahan wrote a Cell review in 2000 titled “The Hallmarks of Cancer” that attempted to characterize what differentiates a tumor from a normal cell.

They summarized the acquired capabilities of cancer in six different categories (and four new ones in 2011):

- Self-sufficiency in growth signals
- Insensitivity to anti-growth signals
- Evading apoptosis
- Sustained angiogenesis
- Tissue invasion & metastasis
- Limitless replicative potential
The multiple avenues of tumorigenesis

- Multiple pathways for a tumor to achieve self-sufficiency.

Combination of: germline mutations (common+rare), somatic mutations, gene-regulatory alterations

- Cancer can be thought of as a combination of germline mutations, somatic mutations, epigenomic changes, and gene-regulatory alterations that give rise to complex phenotypes

Driver vs. passenger mutations

- Oncologists often differentiate between driver and passenger mutations.
- ‘Driver’ mutations confer an advantage to the growth of the tumor.
- ‘Passenger’ mutations do not directly contribute to the fitness of a tumor.

Mutated genes that drive cancer emergence

- Mutations that drive tumorgenesis often fall in four classes:
- **Proto-oncogenes** are genes that normally promote and directs normal cell growth; however, when mutated, they become **oncogenes** and stimulate overactive cell growth.
- **Tumor suppressors** are genes which normally functions to slow cell division; when mutated, they exhibit loss of function and allows for unchecked cell growth.
- **Mutator genes** are mutated genes which normally regulate genomic stability.
- **Epi-mutator genes** are genes whose mutation or dysregulation leads to drastic gene-regulatory changes.
p53 as an example of a tumor-suppressor

- Tumor Protein 53 (p53) serves as a tumor suppressor that is commonly known as “guardian of the genome” serves as a key link between DNA damage and repair/apoptosis.
- Mutations cause loss-of-function and promotes tumor emergence and growth.

Ras-family proteins as examples of proto-oncogenes/oncogenes

- Ras-family members are small GTPases that are involved in cell-growth and cell-cycle pathways
- Mutations in Ras-family members can cause rampant growth and proliferation.

Ras-family proteins as examples of proto-oncogenes/oncogenes

- Ras-family members are small GTPases that are involved in cell-growth and cell-cycle pathways.
- Mutations in Ras-family members can cause rampant growth and proliferation.

Fusion events can create chimeric oncogenes

- Certain tumors are driven by recurrent fusion-events that create chimeric proteins that serve as oncogenes.
  - Famous example include the BCR-ABL fusion gene that drive chronic myelogenous leukemia (CML) progression

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Number of gene fusions*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>Haematological disorders</td>
<td></td>
</tr>
<tr>
<td>Undifferentiated and biphenotypic leukaemia</td>
<td>25</td>
</tr>
<tr>
<td>Acute myeloid leukaemia</td>
<td>302</td>
</tr>
<tr>
<td>Myelodysplastic syndromes</td>
<td>52</td>
</tr>
<tr>
<td>Myeloproliferative neoplasms, including chronic myeloid leukaemia</td>
<td>72</td>
</tr>
<tr>
<td>Acute lymphoblastic leukaemia</td>
<td>237</td>
</tr>
<tr>
<td>Plasma cell neoplasms</td>
<td>22</td>
</tr>
<tr>
<td>Mature B cell neoplasms</td>
<td>181</td>
</tr>
<tr>
<td>Mature T cell and natural killer cell neoplasms</td>
<td>28</td>
</tr>
<tr>
<td>Hodgkin disease</td>
<td>13</td>
</tr>
<tr>
<td>Benign solid tumours</td>
<td></td>
</tr>
<tr>
<td>Benign epithelial tumours</td>
<td>14</td>
</tr>
<tr>
<td>Benign mesenchymal tumours</td>
<td>57</td>
</tr>
<tr>
<td>Malignant solid tumours</td>
<td></td>
</tr>
<tr>
<td>Respiratory system</td>
<td>2,110</td>
</tr>
<tr>
<td>Digestive system</td>
<td>522</td>
</tr>
<tr>
<td>Breast</td>
<td>3,856</td>
</tr>
<tr>
<td>Female genital organs</td>
<td>432</td>
</tr>
<tr>
<td>Male genital organs</td>
<td>676</td>
</tr>
<tr>
<td>Urinary tract</td>
<td>626</td>
</tr>
<tr>
<td>Endocrine system</td>
<td>158</td>
</tr>
<tr>
<td>Nervous system</td>
<td>738</td>
</tr>
<tr>
<td>Skin</td>
<td>193</td>
</tr>
<tr>
<td>Bone</td>
<td>28</td>
</tr>
<tr>
<td>Soft tissue</td>
<td>104</td>
</tr>
</tbody>
</table>

Therapeutic hypothesis: Oncogene addiction

1. Despite multitude of genetic/epigenetic alterations, cancer cells are dependent on a few select oncogenes

2. Targeting these oncogenes could provide an “Achilles heel” for cancer, and enable targeted therapies

3. Combination treatments can drive into oncogene addiction evo space

Mutator genes: Master mutational switches

• Some mutations lead to lower repair efficiency, increasing overall tumor mutation rate
• Mutator genes: involved in DNA repair pathways and genes involved in controlling chromatin stability and movement during the M phase of the cell cycle.

Beyond single proteins: CNVs & rearrangements

- Tumors often exhibit many signatures of genomic instability
  - Polyploidy (cf. WGD)
  - Aneuploidy (gain & loss)
  - Chromosomal duplications/deletions
- Lead to copy number variations (CNV) and
  - Dosage effects (ecDNA)
- Lead to rearrangements = structural variants (SV)
  - Gene-regulatory changes, genes in new context

0. Introduction: oncogenes, tumor suppressors, hallmarks
   - Hallmarks of cancer, tumor-suppressors, proto-oncogenes/oncogenes, mutator phenotype, oncoviruses, and fusion oncogenes.

1. GWAS/Exome lessons: recurrence and heterogeneity
   - Before NGS: cancer predisposition genetics, germline variants, GWAS
   - Recurrence: common signatures across types/patients/tumors/clones
   - Evolutionary dynamics: clonal heterogeneity, computational models

2. WGS lessons: non-coding drivers and convergence
   - Background mutation rate: regions, chromatin states, patients, plexi
   - Convergence: mutations \( \rightarrow \) enhancers \( \rightarrow \) genes \( \rightarrow \) pathways

3. Beyond mutations: epigenomics, functional heterogeneity
   - Epigenomic alterations: reprogramming
   - Single-cell sequencing: functional heterogeneity

4. Tumor immunology, microenvironment, immunotherapy
   - Tumor-microenvironment interactions, cancer immunoediting.
   - Neo-antigens, immune suppression, immunotherapy (checkpoint blockade therapy, CAR T-cells, cancer vaccines).
Identifying cancer driver genes / driver mutations

• Key goal of exome/whole-genome profiling to identify “driver” mutations, positive fitness benefit for the tumor
  – Mutations can be common, rare, or somatic variants
  – Discovered genes revealed hallmarks of cancer biology

• Three types of mutations / three types of analyses:
  – GWAS: weak-effect non-coding mutations common in the population
  – Genetic linkage analysis: strong-effect Mendelian mutations in families
  – Sequencing (exome, genome): strong-effect somatic mutations that arise during mitotic cell divisions
1. Common variants: GWAS, diverse cancer types

- 400+ GWAS hits, diverse cancer types
- Polygenic risk, weaker effects, no visible Mendelian inheritance
2. Rare variants: Mendelian genetics, linkage mapping

- Family history, pedigrees: map cancer driver genes w/ linkage analysis
- Mary-Claire King maps BRCA1 on Chromosome 17q21 in 1990, using genetic linkage analysis in families at inherited risk for breast cancer
- Each family carries a different set of inherited mutations (recurrence)

3. Somatic mutations: exome sequencing

- Paired sequencing of normal and tumor tissue
- Recurrent mutations, hotspots implicated in cancer.
- Clonal heterogeneity: mutational diversity within tumor
  - Heterogeneity leads to lower allelic fractions
  - Only small fraction of seq reads show mutation
  - Need highly sensitive methods to call mutations
Calling somatic point mutations (MuTect)

- MuTecT: Bayesian classifier, determine if mutation is tumor-specific in paired tumor/normal sequencing, even at low allelic fraction (down to 0.1 frequency)

Annotating somatic variants (Oncotator)

- Somatic mutations from MuTect need to be annotated
- Oncotator: annotate using info from several databases

<table>
<thead>
<tr>
<th>Annotation Category</th>
<th>Resource</th>
<th>URL</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic</td>
<td>GENCODE</td>
<td><a href="http://www.gencodegenes.org/">http://www.gencodegenes.org/</a></td>
<td>GENCODE/ENSEMBL transcripts and annotations for hg19</td>
</tr>
<tr>
<td></td>
<td>ref_context</td>
<td></td>
<td>Can be used for artifact inference</td>
</tr>
<tr>
<td></td>
<td>gc_content</td>
<td></td>
<td>Can be used for artifact inference</td>
</tr>
<tr>
<td></td>
<td>Human DNA Repair Genes</td>
<td><a href="http://sciencepaku.mdanderson.org/laby/wood/DNA_Repair_Genes.html">http://sciencepaku.mdanderson.org/laby/wood/DNA_Repair_Genes.html</a></td>
<td>Alteration in such genes can help explain higher overall mutation rates in specific samples</td>
</tr>
<tr>
<td>Protein</td>
<td>UniProt</td>
<td><a href="http://www.uniprot.org/">http://www.uniprot.org/</a></td>
<td>Includes Drugbank &amp; GO annotations</td>
</tr>
<tr>
<td></td>
<td>dbNSFP</td>
<td><a href="https://sites.google.com/site/jpopgen/dbNSFP">https://sites.google.com/site/jpopgen/dbNSFP</a></td>
<td>Contains pre-computed conservation scores, prediction classifications, and other information</td>
</tr>
<tr>
<td>Cancer Variant</td>
<td>COSMIC</td>
<td><a href="http://www.sanger.ac.uk/genetics/CGP/cosmic/">http://www.sanger.ac.uk/genetics/CGP/cosmic/</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cancer Gene Census</td>
<td><a href="http://www.sanger.ac.uk/genetics/CGP/Census/">http://www.sanger.ac.uk/genetics/CGP/Census/</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCLE</td>
<td><a href="http://www.broadinstitute.org/ccle/home">http://www.broadinstitute.org/ccle/home</a></td>
<td>Cancer cell line annotations. Can be used to identify cell line models containing variants of interest</td>
</tr>
<tr>
<td></td>
<td>Familial Cancer Database</td>
<td><a href="http://www.familialcancerdatabase.nl/">http://www.familialcancerdatabase.nl/</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1000 Genomes</td>
<td><a href="http://www.1000genomes.org/data">http://www.1000genomes.org/data</a></td>
<td>Phase 3 variant set</td>
</tr>
<tr>
<td></td>
<td>NHLBI GO Exome</td>
<td><a href="https://esp.gs.washington.edu/drupal/">https://esp.gs.washington.edu/drupal/</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sequencing Project (ESP)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mutational recurrence identifies cancer driver genes

- Recurrence at nucleotide level: specific AA alterations
- Recurrence at gene level: e.g. multiple deactivations
- BRCA1/2: recurrently mutated, indiv. mutations recurrent

**BRCA2**
- On chromosome 13
- Autosomal dominant transmission
- Protein has role in genomic stability
- ~300 different mutations reported

**BRCA1**
- On chromosome 17
- Autosomal dominant transmission
- Protein has role in genomic stability
- ~500 different mutations reported
Large-scale cancer exome analysis reveals recurrently mutated genomic hotspots

- Analysis of mutational profiles from 11,119 tumors across 41 cell types
  - 2 million somatic hypermutations total
- Average # of somatic mutations in exome = 57
- 19,223 human genes harbored hotspots
- 470 hot-spots affecting 275 protein-coding genes
Two sides of cancer: Recurrence vs. Heterogeneity

• **Recurrence:**
  – Small set of pathway alterations necessary for cancer (hallmarks)
  – Oncogenes and tumor suppressors as points of recurrence

**However**

• **Heterogeneity:**
  – Cancer is an evolutionary process driven by positive selection
  – Large number of pre-cancerous cells, constantly subjected to selection
  – Many ways that any oncogenic pathway can be hit

**Deeper look at clonal evolutionary properties**
Clonal genomic heterogeneity within a single tumor

- Intratumor heterogeneity: driven by evolutionary dynamics
- Positive selection for mutations with fitness benefits (to cancer)
- Depends on mutation rate, number of cell divisions, cancer type

Genomic instability as a source of functional heterogeneity

- Abnormalities $\rightarrow$ instability $\rightarrow$ more mutations $\rightarrow$ increased genetic diversity $\rightarrow$ increased heterogeneity
- Instability can be detrimental to individual tumor cells, but also help escape bottlenecks by increasing # of paths to cancer (e.g. emergence, or resistance to chemotherapy/treatment)
- Tumors play the numbers game, little purifying selection

Tumor heterogeneity in both emergence & resistance

- Heterogeneity drives cancer emergence
- Heterogeneity also drives resistance
- Some clones survive therapy, later result in relapse

Tracing clonal history of multiple metastatic sites

- Gerlinger et al. sequenced both metastatic lesions and different locations in the primary tumor in four renal-cell carcinoma patients.
- Mutations are regionally distributed and can be partitioned into private, shared (primary/metastatic), and ubiquitous.
- Can infer phylogenetic relationship between tumor regions.
- Ploidy profiling also shows differences in genomic instability.
Computational models of clonal substructure (PyClone)

- Complex models typically needed for discovering underlying clonal structures.
- Multiple measurements across either time or spatial location is also typically necessary.
- Mutational prevalence $\Leftrightarrow$ phylogenetic history

Tumor heterogeneity within a patient across time

Genevieve Boland, MGH

- Single patient, tracked for 3 years, multiple treatments, multiple lesions, remission
- Dozens of samples from multiple lesions, multiple time-points, across the body
- Use passenger mutations to infer phylogenetic relationships of all tumor samples
- Multiple tumor lineages co-exist simultaneously (at least 5), metastasize, evolve
- Recognize driver events driving tumor resistance and metastasis
- Convergence: genome doubling, C2 mutations found in multiple independent lineages

MAP2K1
IDH1

Genome doubling, Chr2q partial loss
CDH24 nonsense

C10 Mutations (ZNF43)

Chr1p partial loss

CDKN2A/B homozygous del
Additional deletions

PTEN homozygous del
Goals for today: Cancer genomics

0. Introduction: oncogenes, tumor suppressors, hallmarks
   - Hallmarks of cancer, tumor-suppressors, proto-oncogenes/oncogenes, mutator phenotype, oncoviruses, and fusion oncogenes.

1. Exome sequencing lessons: recurrence and heterogeneity
   - Recurrence: common signatures across types/patients/tumors/clones
   - Evolutionary dynamics: clonal heterogeneity, computational models

2. WGS lessons: non-coding drivers and convergence
   - Background mutation rate: regions, chromatin states, patients, plexi
   - Convergence: mutations $\rightarrow$ enhancers $\rightarrow$ genes $\rightarrow$ pathways

3. Beyond mutations: epigenomics, functional heterogeneity
   - Epigenomic alterations: reprogramming
   - Single-cell sequencing: functional heterogeneity

4. Tumor immunology, microenvironment, immunotherapy
   - Tumor-microenvironment interactions, cancer immunoediting.
   - Neo-antigens, immune suppression, immunotherapy (checkpoint blockade therapy, CAR T-cells, cancer vaccines).
Outside proteins: WGS and non-coding mutations

Analysis of mutation types across 436 patients in 8 cell types and overlap with ENCODE/regulatory annotations

- **WGS**: Vast majority of somatic mutations are non-coding
- Most of them are passenger mutations: finding drivers is hard
- How can we make sense of non-coding mutations?
  1. Recurrence: similar to coding, BUT: boundaries not known, background mutation rates vary by region/patient/state/DNA
  2. Convergence: scattered mutations target common genes/pathways

Non-coding mutations: detection & interpretation

Detection

Counting events relative to background

Statistical enrichment and driver calling

Interpretation

Functional prioritization

Experimental validation

Non-coding drivers: (1) Correct for background mutation rate

- Background mutation rates vary greatly: melanoma ~ 100 x sarcoma
- Specific types of mutations vary greatly: C→T vs. C→A vs. C→G


http://doi.org/10.1038/nature12213
(2) Distinct mutational profiles in different cancer type

- Reflect differences in mutator genes
- Different repair genes most active in different tissues
  - e.g. lung cancer will target lung-active repair genes → distinct mutational profiles

(3) Regional mutation rate: replication timing, expression

- Mutation rate varies greatly across genome
- Correlates with:
  - replication timing
  - gene expression level
- e.g. olfactory receptors, long genes both have the most mutations
  - But they are not cancer drivers.
  - They’re just highly mutated!
  - Disappear once you adjust for background mutation rate

(4) Mutation rate varies with DNA accessibility

• Repair machinery optimized for accessible sites (which would otherwise become hypermutated due to access)

• In cancer however, repair machinery is often disrupted

• Cancer mutations are enriched for DNase I hypersensitive sites

• Cancer-type-specific

• Likely due to promoter activity and nucleotide excision repair

1,161 cancer genomes across 14 cell types

(5) Mutation rate varies with minor groove orientation

- Somatic and germline mutation rates show a 10-bp periodicity in nucleosome-occupied DNA

- This periodicity tracks DNA minor groove facing toward and away from the histones

- The orientation of the periodicity depends on the mutational processes active in the tissue

- This has contributed to the AT/CG 10-bp periodicity in eukaryotic genomes

(6) Mutation rate varies by chromatin state

- Distinguish accessible regions as separate class
  - Lowest mutation rate, accessibility of repair machinery?
- Outside DNase, chromatin states vary greatly
  - E.g. promoter regions >> enhancer regions >> transcribed
• In prostate cancer: 2-fold change in mutation rate between patients
• Need to control for patient-specific mutation rate when calling drivers
MutSigCV calls significant somatic mutations (relative to background) by computing aggregate gene scores

- MutSigCV (Gad Getz/Broad) accounts for many covariates (including patient specific effects, gene-specific effects, conservation, transcriptional activity, DNA replication timing, chromatin state) to construct a background model and calls significant somatic mutations.

- It then aggregates gene scores across a tumor and establishes a significant threshold to control FDR.

Detection of non-coding driver regions needs to account for many details regarding the mutational background (ncdDetect)

Non-coding driver detection model - ncdDetect

Multinomial logistic regression framework

\[
\log \left( \frac{\pi_{i,sam}}{1 - \pi_{i,sam}} \right) \sim \sum_{v=1}^{n} \alpha_v x_{v,i} + \beta y_{sam} + 1
\]

ncdDetect performs significance evaluation of potential driver regions.

Courtesy: Malene Rasmussen
Effects of non-coding mutations on tumorigenesis

- Non-coding mutations can affect tumorigenesis in several ways.
  - Gain/loss of motif (Ba + Bb)
  - Altered binding effects of TFs (Bc)
  - Structural variation bring oncogenes closer to promoter/enhancers (C)
  - Mutations in miRNA binding sites prevent or modulate miRNA binding (D)
  - Pseudogene deletion leads to miRNA binding to parent gene (E)

1. Motif gain
2. Motif loss
3. TF changes
4. Regulatory rewiring
5. Post-tx (miRNA)
6. ‘Sponge’ loss (pseudog)
Convergence: enhancers $\rightarrow$ target gene $\rightarrow$ pathway

- Convergence of heterogeneous driver events into common functions:
  - Many regulatory motifs within a common enhancer
  - Many enhancers targeting a common gene
  - Many genes acting in the same pathway
  - Multiple pathways resulting in convergent functions

- Hierarchical model: aggregates mutations across multiple levels

![Regulatory plexus convergence model](Courtesy: Karthik Murugadoss and Richard Sallari)
Recurrent non-coding mutations in regulatory plexus

1. Construct regulatory plexus for each gene
2. Map non-coding driver regions onto regulatory plexus
3. Recognize driver target genes, even in absence of protein-altering mutations

Courtesy: Karthik Murugadoss and Richard Sallari
Experimental validation of non-coding driver

- High-throughput functional dissection and validation of non-coding mutations involves:
  - Synthesizing the mutated sequence
  - High throughput reporter screens (STARR-seq, luciferase reporters)
  - Single-cell-level genetic screens using CRISPRi
  - Validation in model organisms
Goals for today: Cancer genomics

0. Introduction: oncogenes, tumor suppressors, hallmarks
   – Hallmarks of cancer, tumor-suppressors, proto-oncogenes/oncogenes, mutator phenotype, oncoviruses, and fusion oncogenes.

1. Exome sequencing lessons: recurrence and heterogeneity
   – Recurrence: common signatures across types/patients/tumors/clones
   – Evolutionary dynamics: clonal heterogeneity, computational models

2. WGS lessons: non-coding drivers and convergence
   – Background mutation rate: regions, chromatin states, patients, plexi
   – Convergence: mutations → enhancers → genes → pathways

3. Beyond mutations: epigenomics, functional heterogeneity
   – Epigenomic alterations: reprogramming
   – Single-cell sequencing: functional heterogeneity

4. Tumor immunology, microenvironment, immunotherapy
   – Tumor-microenvironment interactions, cancer immunoediting.
   – Neo-antigens, immune suppression, immunotherapy (checkpoint blockade therapy, CAR T-cells, cancer vaccines).
Epigenomic alterations as cancer driver events

A regulatory alteration can activate/repress driver gene, even in absence of mutational changes.
Epi-mutator genes: Global epigenomic alterations

- Epi-mutator genes: control global epigenome maintenance / wiring
- Can be exploited in cancer for global gene-regulatory alterations
  - e.g. EZH2 (Polycomb repressor complex)
  - DNA methylases/demethylases
- Result in global epigenome-wide changes
Recurrent features of the cancer epigenome

- **Global epigenomic changes**
  - Activation: Loss of DNA methylation across hundreds of genes (blue)
  - Repression: Gain of DNA methylation in promoter region CpG islands (red)
  - Repression: Gain of repressive histone modifications (red)
  - State changes: Epigenomic remodeling, nucleosome positioning (green)

- **Epigenomic driver detection (by exome sequencing)**
  - Recurrent mutations disrupting key epigenetic modifiers
  - Vary by cancer type

Epigenetic reprogramming as a guide for therapeutics


- **Diagnose:** Cancer as altered differentiation
- **Treat:** Reprogram cancer cells towards normal development
Functional heterogeneity at the single-cell level

- scRNA-seq captures diversity of clonal groups: CNV ↔ expression diffs
- Additional diversity beyond genomic alterations: role of stochastic variation

Heterogeneity in the tumor microenvironment

- scRNA-seq also reveals diversity of cells that surround the tumor
- Tumor microenvironment heterogeneity: immune and stromal cells

Heterogeneity in cell cycle stage, cellular state

- scRNA-seq allows us to investigate cells at differing stages of the cell cycle
- scRNA-seq reveals hierarchical grouping of tumor heterogeneity:
  - Between different lesions in same patient
  - Between cell types in same lesion
  - Between cells of the same type
  - During cell cycle for a given cell

Match single-cell DNA seq + RNA seq: infer clonality

Goals for today: Cancer genomics

0. Introduction: oncogenes, tumor suppressors, hallmarks
   - Hallmarks of cancer, tumor-suppressors, proto-oncogenes/oncogenes, mutator phenotype, oncoviruses, and fusion oncogenes.

1. Exome sequencing lessons: recurrence and heterogeneity
   - Recurrence: common signatures across types/patients/tumors/clones
   - Evolutionary dynamics: clonal heterogeneity, computational models

2. WGS lessons: non-coding drivers and convergence
   - Background mutation rate: regions, chromatin states, patients, plexi
   - Convergence: mutations → enhancers → genes → pathways

3. Beyond mutations: epigenomics, functional heterogeneity
   - Epigenomic alterations: reprogramming
   - Single-cell sequencing: functional heterogeneity

4. Tumor immunology, microenvironment, immunotherapy
   - Tumor-microenvironment interactions, cancer immunoediting.
   - Neo-antigens, personalized cancer vaccines, immune suppression, immunotherapy (checkpoint blockade therapy, CAR T-cells).
Tumor dependent on & acting on microenvironment

- Tumors must be understood in context of microenvironment
  - Each tumor has a variety of cells in its microenvironment that provide key interactions that either repress or promote growth
Tumor immunoediting: tumor rewrites its own immune profile

- Early during tumor emergence
- Tumor evades adaptive immune system by altering its own immune profile
  - Simply Darwinian evolution: positive selection for tumor cells that go unrecognized
- Tumor represses immune recognition through multiple means:
  - Lowering antigenic profile of tumor cells (favoring less antigenic mutations)
  - Antigen-presentation (repression of Major Histocompatibility Complex, MHC)
  - Establishing immunosuppressive immune microenvironment
    - Down-regulatory cytokine signaling (repress T-cell activity, reprogram T-regulatory cells and dendritic cells)
    - Direct repression of T-cell effectors (by up-regulating T-cell-inhibitory ligands)
    - Recruitment of suppressive immune cell types (recruit immunosuppressive myeloid-derived suppressor cells MDSCs)

→ escape immune surveillance

Tumor-immune interactions: progression, therapy, relapse

- Novel coding mutations generates new antigens
  - Immune system can now recognize tumor as non-self, target it
- Tumors down-regulate their own anti-gen presentation
  - Tumor mutations down-regulating MHC
  - Epigenetic alterations down-regulating MHC
- Tumors down-regulate immune system
  - Immunosuppressive environment
  - Tumor-immune interactions
- Potential therapeutics targeted towards this
- Immune system helps select clones that are not recognized, by attacking highly antigenic clones
- Tumor decoy cells can potentially overwhelm immune system and let other cells survive

Recurrent mutations in HLA genes repress MHC antigen presentation

- Tumor generate inhibitory ligands that suppress T-cell activity
- Immunotherapies can reverse this effect and promote T-cell activity
- Auto-immune diseases can help trigger immune response against cancer

Mutations in the HLA Class I genes in several tumor types


Tumors create immunosuppressive env by inhibiting T-cell function

- Tumor generate inhibitory ligands that suppress T-cell activity
- Immunotherapies can reverse this effect and promote T-cell activity
- Auto-immune diseases can help trigger immune response against cancer

Current targets of checkpoint blockade immunotherapy programs

WAKING UP THE BODY'S DEFENCES
Tumour cells can inhibit the body’s immune response by binding to proteins, such as PD-1, on the surface of T cells. Antibody therapies that block this binding reactivate the immune response.

Neo-antigen diversity predicts immunotherapy success (higher mutational load more likely to succeed)

- Mutational frequency and neo-antigen load is correlated with response to checkpoint blockade immunotherapy
- Intuition: even if T-cells aren’t immunosuppressed anymore, they still need targets!

Computationally predict neoantigen / neoepitope recognition

- How do we actually know if a coding mutation will become a neo-epitope?
- We can predict the likelihood of presentation of antigens through NetMHCpan (position-specific scoring matrix, newer versions use neural networks)

Tumor types differ in their neo-antigen landscape

- Neo-antigen abundance is highly correlated with mutational rate of the tumor type
- Neo-antigens can be oncogene or passenger mutations
Intratumor heterogeneity of neo-antigen landscape predicts immunotherapy success

- Clonality of the neo-antigen space is also important in determining the effects of blockade immunotherapy
- Separating out neo-antigens by an heterogeneity threshold shows that more clonal neo-antigen profiles separate out survival profiles better

T-cell repertoires and immunotherapy

- How do T-cells recognize the multitude of antigenic peptides that’s being presented on MHC?
- Each T-cell generates a unique T-cell receptor (TCR) through V(D)J somatic recombination; the region that interfaces with the MHC is known as the CDR3

Infiltrating T-cells have distinct repertoires

- TCR sequences can be called directly from RNA-seq data using a novel approach from Shirely X. Liu’s lab at Harvard.
- Can determine the diversity of TCR (CPK, clonotypes per kilo reads) across different cancer types, distinct profiles of TCR repertoires in tumors, and using correlations between neo-antigen and TCR sequences, identify matched neo-epitopes and TCR combinations.

Detection of antigen-specific T-cells using DNA-barcoded peptide-MHC-I multimers

- Recent *Nature Biotech* paper showed that we can directly barcode individual synthetic peptide-MHC I complexes with a DNA barcode and then sequence to determine whether a patient’s immune system is reacting towards a particular antigenic peptide.
- Can track over cancer progression to see how the T-cell repertoire reacts with particular neo-epitopes over time.

Designing personalized neo-antigen cancer vaccines

- Goal: utilize patient-specific neo-antigen profile to develop cancer vaccines to assist with anti-tumor response
- Hypothesis: neo-antigen specific vaccines will promote anti-tumor response by T-cells with T-cell receptors that recognize the neo-antigens.

2 major challenges:
- Identify patient-specific HLA-allele specific peptides
- Validate that synthetic peptides assists tumor response in real patients
Challenge I: Predicting HLA-allele specific peptides

• The types of small peptides presented to the immune system by professional antigen presenting cells depends heavily on the germline HLA (Human Leukocyte Antigen) alleles unique to each patient.

• The HLA locus is highly diverse. MHC class I alleles:
  - HLA-A: 4,340
  - HLA-B: 5,212
  - HLA-C: 3,930

• Solution: combine high-throughput HLA-binding assay peptide w/neural network predictive model
Challenge II: Validating clinical efficacy of synthetic peptide mutations

- Clinical trial design
  - 16 personalized neo-antigen vaccines per patient
  - Primary endpoints: safety, feasibility
  - Secondary endpoints: Immune response, 2 year progression-free survival

**LETTER**

An immunogenic personal neoantigen vaccine for patients with melanoma

Vaccine induces T-cells against almost all neo-antigen pools

Patient T-cells recognizes mutated epitopes more efficiently than wild-type epitopes

Complete response in conjunction with anti-PD1 for 2 patients with progression after Neovax

** Anti-PD1 complete response rate in metastatic melanoma is 5%.

CAR T-cell therapies

- Another popular immunotherapy that uses the targeting elements of a monoclonal antibody (raised against recurrent tumor antigens) fused with the signaling component of a T-cell receptor to direct cytotoxic T-cells against tumor cells.
- These artificial receptors are directly integrated into a patient’s immune cells through adoptive transfer.
- Allows for robust polyclonal T-cell response without antigen presentation and presentation! (unlike checkpoint blockade therapies)
Bispecific T-cell Engagers

- Evolution of CAR T-cell therapies - BiTE utilizes fusion protein with 2 antibody domains. The second antibody scFv domain allows for more specific targeting of antigen-experienced T-cells (with anti-CD3 domain) or multiple tumor antigens.

Towards a complete understanding of the cancer microenvironment

Gene-level regulatory convergence

- Regulator
  - Enhancer
  - Promoter
  - Gene body
  - 3'UTR
  - Repressive

- Exonic mutations
- miRNA

H3K27ac

motif analysis

RNA-Seq expression

DNA methylation

miRNA profiling

Tumor

Tumor-Immune interface

T-cell Receptor neoantigen profiling

Cytokinome profiling

Exosomal profiling

Proteomic profiling

CD4

Immune cells

CD8

Treg

Dendritic

Macrophage

Courtesy: Manolis Kellis
Immunotherapies are just one part of the drug treatment picture

Combination therapies often show the best performance

Goals for today: Cancer genomics

0. Introduction: oncogenes, tumor suppressors, hallmarks
   - Hallmarks of cancer, tumor-suppressors, proto-oncogenes/oncogenes, mutator phenotype, oncoviruses, and fusion oncogenes.

1. Exome sequencing lessons: recurrence and heterogeneity
   - Recurrence: common signatures across types/patients/tumors/clones
   - Evolutionary dynamics: clonal heterogeneity, computational models

2. WGS lessons: non-coding drivers and convergence
   - Background mutation rate: regions, chromatin states, patients, plexi
   - Convergence: mutations $\rightarrow$ enhancers $\rightarrow$ genes $\rightarrow$ pathways

3. Beyond mutations: epigenomics, functional heterogeneity
   - Epigenomic alterations: reprogramming
   - Single-cell sequencing: functional heterogeneity

4. Tumor immunology, microenvironment, immunotherapy
   - Tumor-microenvironment interactions, cancer immunoediting.
   - Neo-antigens, immune suppression, immunotherapy.