Module 6: Current research directions

- 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/colonies
   - 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.

The Hallmarks of Cancer: A Framework for Understanding Cancer Biology

- 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):
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 tumorigenesis often fall in four classes:
  - Proto-oncogenes are genes that normally promote and direct normal cell growth; however, when mutated, they become oncogenes and stimulate overactive cell growth
  - Tumor suppressors are genes which normally function 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 drives chronic myelogenous leukemia (CML) progression


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
- Lead to rearrangements = structural variants (SV)
  - Gene-regulatory changes, genes in new context


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. 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 → 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 immunoeediting.
   - 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

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
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

- 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 ↔ phylogenetic history

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).

Non-coding mutations: (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

Outside proteins: WGS and non-coding mutations
- 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
- Counting events relative to background
- Statistical enrichment and driver calling
- Functional prioritization
- Experimental validation
(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) Mutations rate varies with DNA accessibility

- Repair machinery optimized for accessible sites (which would otherwise become hypermuted 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 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

(6) Mutation rate varies by patient

- 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 the 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

Convergence: enhancers → target gene → 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

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)
  - 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

Recurrence features of the cancer epigenome

- Global epigenomic changes
  - Activation: Loss of DNA methylation across hundred 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
**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 \(ightarrow\) 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 immunoeediting.
   - 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 rewires 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)

\(\rightarrow\) 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

Tumors can downregulate antigen presentation

On promote immunosuppression by activating suppressive signals in T-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**

- Match single-cell DNA seq + RNA seq: infer clonality
- Single-cell sequencing: functional heterogeneity

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

- Match single-cell DNA seq + RNA seq: infer clonality
- Single-cell sequencing: functional heterogeneity
Recurrent mutations in HLA genes repress MHC antigen presentation

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

Tumors create immunosuppressive env by inhibiting T-cell function

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

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, clonotypies 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 antigen presentation system 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 immuno-geic 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 specific recurrent tumor antigens) fused with the signaling component of a T-cell receptor

Towards a complete understanding of the cancer microenvironment

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 → 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.