Canadian Bioinformatics Workshops

www.bioinformatics.ca
Module 1
Introduction to Cancer Genomics

Trevor Pugh, PhD, FACMG
Bioinformatics for Cancer Genomics
May 30 – June 3, 2016

Learning Objectives of Module

• Describe how cancer genomes differ from normal genomes and various sources of cancer genome variation

• Understand different bioinformatic approaches to detecting cancer genome variation

• Learn strengths and weaknesses of different approaches to cancer genome analysis (whole genome, exome, RNA-seq, targeted sequencing, etc.)

• Inform how cancer genome analysis may be used to guide cancer patient management
We are all made of cells

https://genographic.nationalgeographic.com/science-behind/genetics-overview/

My human genome
My human genome

Cancer is a disease of the genome

Glioma karyogram (GTG-banding) 78, <4n>, XXX, 2, -5, +6, del(6)(q21q23)x2, +7, -8, del(8) (q22q24.1), del(9)(p10)x2, -10, -11, -11, -12, -13, -13, -14, -14, -16, -19, del(19)(p10), -21, +22

www.cityofhope.org/research/support/cytogenetics/
“All happy families are alike; each unhappy family is unhappy in its own way”

Anna Karenina
Leo Tolstoy

Cancer cells accumulate somatic alterations over time

Mutation frequency depends on cancer type
External forces (e.g. drug treatment) can select for specific clones (e.g. cells with resistance mutations)

Mutation burden varies by cancer type, exposure, age of onset, & DNA repair ability


Tumour cells acquire abnormal abilities by co-opting normal cell behaviour

Oncogenic somatic alterations target a core set of biological functions

Differentiating “driver” from “passenger” mutations is a central challenge of cancer genome analysis

Targeted therapies exploit specific mutations

Lung adenocarcinoma with activating mutations of \textit{EGFR} \downarrow
Inhibiting EGFR shrinks tumors

Metastatic melanoma with activating mutations of \textit{BRAF} \downarrow
Inhibiting BRAF shrinks tumors

Resistance inevitably arises
**“Actionable” mutations in 3,299 tumours**

- 37 genes from 4 pathways linked to treatments
- 12 genes disrupted by multiple mechanisms

1) Druggable alterations cut across cancer types

2) Combination therapies may be effective in tumors with compound pathway alterations

3) 50% of tumors have at least two disrupted, druggable pathways

---

**Even within the same tumour mass, different cells may have different mutations,**

- Subclone 1
- Subclone 2
- Subclone 3

Clonal heterogeneity

Cancers are a mix of subclones that can respond differently to therapy


Subclones and metastases are genetically related, diverge to form subpopulations

Reasons for molecular testing for cancer

Treatment
treatment susceptibility, predict adverse side-effects, watch-and-wait versus aggressive treatment

Drug resistance and metabolism
pre-existing resistance mutations, drug dosage

Inherited cancer syndromes
additional primary tumors, at-risk family members

Prognosis
molecular subtyping, benign vs. malignant, determine unknown primary, predict metastatic potential

What are the targets in my cancer...

<table>
<thead>
<tr>
<th>Sequence</th>
<th>mutations, polymorphisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>copy number variation, translocation, loss-of-heterozygosity</td>
</tr>
<tr>
<td>Function</td>
<td>expression, exon-usage</td>
</tr>
<tr>
<td>External</td>
<td>Viruses, bacteria</td>
</tr>
</tbody>
</table>

…and what can be done about them?
Applications of next-generation DNA sequencing technology to cancer

General DNA sequencing workflow

Genomic DNA or RNA → Library of DNA fragments → Sequencing device → Computational analysis
An alignment of reads to a human genome reference

Bwa, Novoalign, Stampy, SOAP, and >50 more listed on Wikipedia under “Short-Read Sequence Alignment”
Multiple types of cancer genome variation may be inferred from sequencing read alignments

DNA sequencing approaches to cancer

Whole genome sequencing (WGS)
Whole exome sequencing (WES)
Targeted gene sequencing
Targeted variant genotyping
Epigenome modification (bisulphite)
Transcriptome sequencing (RNAseq)
miRNA sequencing
Protein/DNA interaction mapping
Epigenome mapping (histones)

DNA
RNA
Protein
(DNA footprint)
Coverage of genome, exome, & RNAseq data

- KRAS, 49 kb
- Genome ~40X (0-92X)
- Exome ~150X (0-307X)
- RNAseq 0-106X

www.broadinstitute.org/igv

Exomes & Genomes don’t measure everything

Fraction of each exon “callable”

60X genomes 30X genomes Native exome WGA exome

Pugh et al. Nature 2013
Transcriptome coverage depends on expression

Expression level vs. Genes ordered by rank expression

Epigenetic modifications that regulate gene expression detectable by bisulphite sequencing, even in admixed tissues

Normal tissue
- Unmethylated MLH1 promoter (C>T)
- Endometrial carcinoma
- Methylated MLH1 promoter (C>C)

EPCAM ~ MLH1
## Data types are confirmatory and complementary

<table>
<thead>
<tr>
<th></th>
<th>Genome</th>
<th>Exome</th>
<th>RNAseq</th>
<th>MethylSeq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence mutations</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Copy number</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Rearrangements</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Pathogens</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purity/Ploidy</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Low purity or subclonal mutations</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Gene expression</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Allele- or exon-specific expression</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Somatic Mutations
Germline variants are detectable at low coverage

Heterozygous
Total coverage = 19
C = 10
T = 9
Allele balance = 0.47

Variable DNA quality & quantity from tissues used for routine diagnosis

Tumors are a mix of cancer & normal cells ("purity" or "tumour content")


Tumors can have multiple genome copies ("ploidy")

www.cityofhope.org/research/support/cytogenetics/
Deep coverage is necessary to detect mutations in low purity or high ploidy tumors.

Genome 0/15 reads

Exome 6/139 reads

Deep (e.g. 250X) DNA sequencing coverage enables inference of tumour purity, ploidy, & subclonal structure.

Multiple DNA sequencing technologies enable mutation confirmation & cross-validation

PCR/ PacBio

Hybrid capture/ Illumina


Mutation detection sensitivity is dictated by sequencing depth and is limited by sequencer & polymerase error

Minimum fraction of tumour DNA in detectable with 99% probability

Exome

Clinical lab

Subclonality

Circulating tumour DNA

Sequencer error rate

Coverage

25,544X

E11X
The ultimate complex mixture:
Cell-free DNA dissolved in blood is derived from many normal cells and a few tumour cells

Crowley et al., Nature Reviews Clinical Oncology, 2013

Coverage dependent on laboratory method used to target regions of interest
PCR >100,000X coverage
Hybrid capture ~50,000X

Tam-Seq: Forshew et al.

CAPP-seq: Newman, Bratman et al.
Quantity of circulating tumour DNA reflects cancer burden


Quantity of circulating tumour DNA reflects clinical course

Somatic copy number alterations and rearrangements (structural alterations)

Tumors can have complex structural alterations

Gains and losses evident from sequence data

Chromosome Relative copy number

Tumor microarray Normal exome Tumor exome Normal exome

Re-analysis of existing clinical sequence data provides additional diagnostic value

Homozygous PTEN deletion in a breast invasive ductal carcinoma

EGFR amplification (6 copies) in a lung adenocarcinoma
Additional probes provide richer copy number profiles (183 gene panel, all exons plus select introns)

FGFR1
5 copies, no mutations

Exact breakpoints can be detected by examining reads that span the region

Translocation detection from soft-clipped bases plus large-insert or intrachromosomal read-pairs

PML-RARA

Rearrangements can be highly complex and detectable at base-pair resolution

Some tumor genomes are highly rearranged, some are quiet, some have no rearrangements

Imielinski et al. 2012
Lung adenocarcinoma

Pugh et al. 2012
Neuroblastoma

Molenaar et al. 2012.

Transcriptome sequencing
RNAseq can detect differences in exon usage that underlie resistance to treatment


RNAseq enables expression profiling & subtyping

Primary tissue of origin evident from expression (if normal samples are available for comparison)

- Brain
- Lung
- Thyroid
- Skin
- Heart
- Muscle
- Blood

Four lung tumors of unknown tissue of origin

Pathogens seen in Genomes, RNAseq, sometimes Exomes

Unusual mapping of RNAseq reads to Epstein-Barr virus in lung adenocarcinoma

0.8 mm
ISH shows EBV confined to tumor cells, tumor is actually a lymphoepithelioma-like carcinoma
Germline cancer genetics (Hereditary Cancer Syndromes)

“Second hit” can unmask a pathogenic germline allele
Carrier for a pathogenic SUFU mutation developed medulloblastoma (10-)

Normal, heterozygous

Tumor, hemizygous

Module 1  bioinformatics.ca

Once found, variants (of all types) require annotation to guide interpretation

Module 1  bioinformatics.ca
How do we interpret variants en masse?

Clinical indication & history (family members, de novo)
Locus-specific and internal variant databases
Population frequency
Public databases (disease-specific & general)
Publications
Amino acid & biochemistry conservation
Substitution prediction tools
(Polyphen2, SIFT, AlignGVGD, SarcomerePolyphen)
Splicing prediction tools
(SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer, Human Splice Finder)
Druggable targets and pathways
Drug Gene Interaction Database (DGIdb), Drug Bank

Command-line versus Manual Annotation

Oncotator – plain text table with >150 columns

Alamut – Interactive desktop software

Varient Effect Predictor, ANNOVAR, SNPeff, VAT, & many others
Novel Variant Assessment is still a manual enterprise, aided greatly by sharing interpretations

How do we report and share these results?

**Overall result: Somatic variants identified:**
- **BRAF** (NM_004333.4) Heterozygous, c.1406G>T (p.G469V)
- **RB1** (NM_000321.2) Heterozygous, c.607+1G>A (splice)
- **TP53** (NM_000546.4) Heterozygous, c.488A>G (p.Y163C)

**Class 4A mutation:** BRAF p.Gly469Val (58% of reads)
BRAF p.G469V is a rare mutation across cancer (76 of >193,000 cases in the Catalogue of Somatic Mutations in Cancer), and has not been previously reported in over 4,300 ovarian cancers. BRAF is recurrently mutated in ovarian cancers, but is rarely mutated in high grade serous carcinoma (mycancergenome.org).

The most common variant found in ovarian tumours is p.V600E (mycancergenome.org). Currently, the impact of this mutation on patient prognosis or treatment in this tumour site is unknown.
AACR Project GENIE is an international, multiphase, multiyear project that will provide the “critical mass” of genomic and clinical data necessary to improve clinical decision making and catalyze new clinical and translational research.

GENIE will aggregate existing and ongoing genotyping efforts from the seven phase 1 project participants into a single registry and link these data to select clinical outcomes, ultimately making these data publicly available.

www.aacr.org/genie

Molecular report of the future?
Concise, systematic (crowd-sourced?) genome interpretation alongside clinical annotations

www.cbioportal.org

Courtesy of Niki Schultz
Initial presentation

- 78 year-old Caucasian man
- Fit and active
- Presented in August 2007 with throat discomfort
- Examination found a 2 cm mass at left base of tongue
- Minimal comorbidities
- No obvious risk factors for oropharyngeal malignancy
PET-CT scan & subsequent biopsy


Primary tongue mass, H&E stain, 20X objective

Surgery & further pathology

Laser resection of tumor & lymph nodes

**Primary:** 1.5 cm poorly differentiated adenocarcinoma with micropapillary & mucinous features

**Lymph nodes:** 3 of 21 neck nodes contain metastatic adenocarcinoma

60 Gy of adjuvant radiation therapy completed in Feb.

Good quality of life, returned to work for four months
Then...numerous small (< 1.2 cm) bilateral pulmonary metastases
**EGFR IHC positive expression**

+2 EGFR expression by Zymed immunohistochemistry protein expression assay

*Treatments?*

**6 week trial of erlotinib**

All pulmonary nodules grew while on erlotinib
Largest lesion grew from 1.5 to 2.1 cm
Erlotinib discontinued in August

*Palliative care? What next?*
What are the targets in our case?

Having exhausted standard of care, BC Cancer Agency oncologist turned to the Genome Sciences Centre (a large research group) for new leads

Special meeting of the BCCA Research Ethics Board approved a single-case analysis for this patient

Patient consented to full genomic sequencing and analysis with the understanding that novel treatment options may be suggested

Fresh frozen biopsy taken for RNA-seq

Fine-needle aspirates taken from large lung lesion
Pathologist reviewed, used samples with highest tumor content (~80%) for RNA sequencing
FFPE DNA from surgical resection used for DNA
Somatic Mutations

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>p.D259Y</td>
<td>Located in region of LOH. Known somatic mutation in 7 cancer types. (16 entries + 4 alternate mutations across 28 tumors in the Catalogue of Somatic Mutations in Cancer)</td>
</tr>
<tr>
<td>RB1</td>
<td>p.L234*</td>
<td>Novel truncating mutation. Results in loss of 75% of protein sequence. Loss of RB1 and PTEN result in gefitinib resistance. (Albitar et al. Gynecol Oncol. 2007 Jul;106(1):94-104.)</td>
</tr>
</tbody>
</table>

Confirmed by secondary sequencing method (Sanger)

Copy number alterations

18q loss (SMAD4) with focal amps.
TP53 LOH
MAPK3 amplification

10q loss (PTEN) Focal RET amplification
EGFR amplification
FISH confirmation of copy number alterations

- **RBBP8 amplification**
- **PTEN single copy loss**
- **Focal RET amplification**

Tumour gene expression levels compared against reference set of 50 tumors & matched blood sample

- **SMAD4** expression 43X lower vs compendium deleted & down-regulated, associated with metastasis of colorectal cancer
- **RET** in top 5% of expressed genes, 34X vs compendium most highly expressed oncogene
- **PTEN** in bottom 5% of expressed genes, significantly under-expressed
Tumor driven by RET up-regulation, PTEN loss, not AKT

- Up-regulation of the MAPK pathway increases cell proliferation.
- RET, a validated thyroid cancer target, and its growth factors are amplified and overexpressed.
- AQP5, a known activator of this pathway, is overexpressed.
- MAPK3 (ERK1) is amplified.
- PTEN, a suppressor of this pathway, is highly down-regulated.

Drug | Known mechanism & indications
--- | ---
Sunitinib | Targets PDGFRs, VEGFRs, RET, KIT, CSF1R, FLT3. Approved for GIST and RCC. In trials for thyroid cancer.
Motesanib | Targets VEGFRs, PDGFRs, KIT, RET. In trials for thyroid cancer, GIST, NSCLC.
Sorafenib | Targets BRAF, RAF1, RET, VEGFRs, PDGFRB, KIT, FLT3. Approved for RCC and HCC. In trials for thyroid cancer.
Sulindac | An NSAID COX inhibitor for inflammation but also inhibits MAPK3 (ERK1).
**22% decrease after 4 weeks on sunitinib**

*(16% growth prior to sunitinib)*

<table>
<thead>
<tr>
<th>Nodule 1</th>
<th>Nodule 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct 1st 2008 Biopsy taken</td>
<td>Oct 29th 2008 Sunitinib begun</td>
</tr>
<tr>
<td>Oct 29th 2008 Sunitinib begun</td>
<td>Dec 9th 2008 4 wks 50mg daily Sunitinib</td>
</tr>
</tbody>
</table>

**Stabilization for 7 months**

Sunitinib dose reduced due to known side-effects, otherwise excellent quality of life

Repeated scans showed no new nodules and disease stabilization and 4 months

...then, existing lung mets began to grow

Switched to sorafenib & sulindac
Disease again stabilized within 4 weeks and continued for 3 months
Recurrent disease after 7 months

Recurrent disease at primary site on tongue
New neck skin nodule
Progressive & new metastases in lung
Deteriorating quality of life

What changed?

Skin metastasis contains new mutations

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZMYM4</td>
<td>p.Q317H</td>
<td>Zinc-finger</td>
</tr>
<tr>
<td>DNAH7</td>
<td>p.V2590L</td>
<td>Force generating protein of respiratory cilia</td>
</tr>
<tr>
<td>CXCL13</td>
<td>p.R56H</td>
<td>B-cell attracting chemokine</td>
</tr>
<tr>
<td>HSD17B8</td>
<td>p.A141T</td>
<td>hydroxysteroid (17-beta) dehydrogenase 8</td>
</tr>
<tr>
<td>PCLO</td>
<td>p.T2759A</td>
<td>presynaptic cytomatrix protein</td>
</tr>
<tr>
<td>GRIA4</td>
<td>p.R872C</td>
<td>glutamate receptor</td>
</tr>
<tr>
<td>OR4K2</td>
<td>p.L197F</td>
<td>olfactory receptor</td>
</tr>
<tr>
<td>SYNE2</td>
<td>p.A302G</td>
<td>tethers the nucleus to the cytoskeleton</td>
</tr>
<tr>
<td>PTPRM</td>
<td>p.A929T</td>
<td>cell-cell adhesion, possibly signal transduction &amp; growth</td>
</tr>
</tbody>
</table>

No evidence of these in the pre-treatment biopsy, even at low frequency
Resistance due to intensified RET and new AKT signaling?

Overcoming mechanisms of resistance

Cocktail of targeted drugs against multiple members of activated and parallel pathways?
- RET, EGFR, mTOR, Akt, etc.
- untested, risk of adverse side-effects

Detection of resistance mechanisms pre-treatment
→ none of the new mutations were evident pre-treatment

Can resistance be modeled & monitored?
- serial biopsies? blood test?
Timeline

August
Initial presentation

November
Surgical resection

February
Radiation complete

June
Lung mets, erlotinib begun

August
Erlotinib failure

Nov
Sunitinib begun

May
Primary recurrence, 2nd biopsy

Oct
1st biopsy for sequencing

Feb
Lung mets grow, switched to sorafenib & sulindac

Palliative care until death

Module 1 bioinformatics.ca