Lecture 22. Cancer Genetics & Genomics

Michael Schatz

April 19, 2017 JHU 600.749: Applied Comparative Genomics



Presentations!

Recommended outline for your talk (1 minute per slide):

- 1. Title Slide: Who are you, title, date
- 2. Intro 1: Whats the big idea???
- 3. Intro 2: More specifically, what are you trying to learn?
- 4. Methods 1: What did you try?
- 5. Methods 2: What is the key idea?
- 6. Data 1: What data are you looking at?
- 7. Data 2: Anything notable about the data?
- 8. Results 1: What did you see!
- 9. Results 2: Does it work?
- 10. Results 3: How does it compare to other methods/data/ideas?
- 11. Discussion 1: What did you learn from this study?
- 12. Discussion 2: What does this mean for the future?
- 13. Acknowledgements: Who helped you along the way?

I strongly discourage you from trying to give a live demo as they are too unpredictable for a short talk. If you have running software you want to si the important steps.

Presentations!

JHU EN.600.749: Computational Genomics: Applied Comparative Genomics

Project Presentations

Presentations will be a total of 15 minutes: 12 minutes for the presentation, followed by 3 minutes for questions. We will shictly keep to the schedule to ensure that all groups can present in class.

Schedule of Presentations

Day	Time	Team Name	Students	Title
Th 4/26	130-145	Vahwareth	Pranay Vishwanath	Overnitiality on plant genomes
Th 4/26	145-2.00	minikitSA	Nathan Roach	Minimizer accelerated clustering and MSA of highly similar sequences
Th 4/26	2:00-2:15	Chrome	Ashiee Feng	identifying Sequence-Specific Open Chromatik Regulators and Applications on Noncoding Variant Analysia
Th 4/26	215-2.00	ZX Team	Zhowei Na	Analysis of the Performance of MinMash algorithm for Copy Number Analysis by Paciblo reads in C. glabrata
TE 4/28	230-245	The infimPlateors	Roham Razaghi, Timethy Dipatrice.	Using nampore sequencing data to examine Afele-specific expression and methylation patterns
		E.		
11.5,9	100-146	Chesapeake Bay Explorer	Zhuoyue Zheng, Junyeo Go	A metagenomic-based survey of microbial (de)halogenation potential in the Chesapeake Bay
72 5/2	145-2:00	Binomica	Benjamin Kaminow	Analysis of Repeat Bits in Binaries Using DNA Short Read Assembly
7+5/2	2:00-2:15	Cao	Raymon Cao	Evaluating Metagenomic Classification Software for Identifying Food Sample Contaminations
7+5/2	215-2:30	Pathy	Shreyes Padhy	Deep Recursent Nativorks for Base-Calling
7+6/2	230-248	Rey Charles Driving School Class of 2020	Diego Gelsinger, Michael Skan), Andrew Gale	Benchmarking of RNA-seg Analysis Pipelines using Circulating Tumor Expression Data
-	* 45 (Inc.)	********	*	NAMES OF A DESCRIPTION OF A DESCRIPTION OF A DESCRIPTIONO
Th 5/4	1.92-1.45	Araba-cadabra	Micheel Alonge	Pairwise Pan-Genome Structural Variant Detection in Arabidopsis Ittaliana Genomes
Th.5/6	1.45-2:00	30	Wang XI	Systemic evaluation of ChramHMM and Segway, and chromatin state analysis on dipibid human genome
79-5/6	2:00-2:15	Rehtign	Ales Varabytu	Graph genome assisted realignment of short-read sequencing data for small highly divergent genomes.
Th 5/4	215-230	KrispKrame	Srivethee Resumants	On-target activity prediction of sgRNA in CRISPR/CASB using Deep Neural Networks
Th 5/4	230-245	Insertion Short	Michael Kirsche	Short-read Known Insertion Detection
Th 5/4	245-3.00	Gray	Justin Dray	Finding Gene Regulatory Pathways Shared Between Dontrant Tumor Cells

Adventures in Overfitting







Benign vs. Malignant

Benign vs. Malignant Tumors

Benign (not cancer) tumor cells grow only locally and cannot spread by invasion or metastasis Malignant (cancer) cells invade neighboring tissues, enter blood vessels, and metastasize to different sites



Somatic Mutations In Cancer



Signatures of mutational processes in human cancer

Alexandrov et al (2013) Nature. doi:10.1038/nature12477

A firestorm in cancer



Figure 2. Major types of tumor genomic profiles. Segmentation profiles for individual tumors representing each category: (A) simplex; (B) complex type I or sawtooth; (C) complex type II or firestorm. Scored events consist of a minimum of six consecutive probes in the same state. The y-axis displays the geometric mean value of two experiments on a log scale. Note that the scale of the amplifications in C is compressed relative to A and B owing to the high levels of amplification in firestorms. Chromosomes 1–22 plus X and Y are displayed in order from left to right according to probe position.

Novel patterns of genome rearrangement and their association with survival in breast cancer Hicks et al (2006) *Genome Research*. Doi: 10.1101/gr.5460106

SK-BR-3

Most commonly used Her2-amplified breast cancer cell line



(Davidson et al, 2000)

80+ chromosomes,

Many are a patchwork of fragments of other chromosomes

Aberrations in cancer genomes



Chromothripsis, which literally means 'chromosome shattering', is a phenomenon that has recently been reported to occur in cells harbouring complex genomic rearrangements (CGRs). Has 3 defining characteristics:

- (1) Occurrence of remarkable numbers of rearrangements in localized chromosomal regions;
- (2) Low number of copy number states (generally between one or two) across the rearranged region;
- (3) Alternation in the chromothriptic areas of regions where heterozygosity is preserved with regions presenting loss of heterozygosity (LOH).

Chromothripsis and cancer: causes and consequences of chromosome shattering Forment et al (2012) Nature Reviews Cancer. doi:10.1038/nrc3352

Hypomethylation distinguishes genes of some human cancers from their normal counterparts

Andrew P. Feinberg & Bert Vogelstein

Cell Structure and Function Laboratory, The Oncology Center, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

It has been suggested that cancer represents an alteration in DNA, heritable by progeny cells, that leads to abnormally regulated expression of normal cellular genes; DNA alterations such as mutations1.2, rearrangements3-6 and changes in methylation⁶⁻⁸ have been proposed to have such a role. Because of increasing evidence that DNA methylation is important in gene expression (for review see refs 7, 9-11), several investigators have studied DNA methylation in animal tumours, transformed cells and leukaemia cells in culture^{8,12-30}. The results of these studies have varied; depending on the techniques and systems used, an increase¹²⁻¹⁹, decrease²⁰⁻²⁴, or no change²⁵⁻²⁹ in the degree of methylation has been reported. To our knowledge, however, primary human tumour tissues have not been used in such studies. We have now examined DNA methylation in human cancer with three considerations in mind: (1) the methylation pattern of specific genes, rather than total levels of methylation, was determined; (2) human cancers and adjacent analogous normal tissues, unconditioned by culture media, were analysed; and (3) the cancers were taken from patients who had received neither radiation nor chemotherapy. In four of five patients studied, representing two histological types of cancer, substantial hypomethylation was found in genes of cancer cells compared with their normal counterparts. This hypomethylation was progressive in a metastasis from one of the patients.

and (3) HpaII and HhaI cleavage sites should be present in the regions of the genes.

The first cancer studied was a grade D (ref. 43), moderately well differentiated adenocarcinoma of the colon from a 67-yrold male. Tissue was obtained from the cancer itself and also from colonic mucosa stripped from the colon at a site just outside the histologically proven tumour margin. Figure 1 shows the pattern of methylation of the studied genes. Before digestion with restriction enzymes, all DNA samples used in the study had a size >25,000 base pairs (bp). After HpaII cleavage, hybridization with a probe made from a cDNA clone of human growth hormone (HGH) showed that significantly more of the DNA was digested to low-molecular weight fragments in DNA from the cancer (labelled C in Fig. 1) than in DNA from the normal colonic mucosa (labelled N). In the hybridization conditions used, the HGH probe detected the human growth hormone genes as well as the related chorionic somatotropin

Table 1 Quantitation of methylation of specific genes in human cancers and adjacent analogous normal tissues

Patient	Carcinoma	Probe	Enzyme	% Hypomethylated fragments		
				N	с	м
	Colon	HGH	∫Hpa1I	<10	35	-
1.8			Uthel	<10	39	-
		y-Globin	fHpall	<10	52	-
			Lithal	<10	39	-
	Colon	a-Globin	[Hpall	<10	<10	-
			UHhal	<10	<10	
		HGH	\$ Hpa II	<10	76	-
- 2			Lithal	<10	85	-
		y-Gilobin	f Hpall	<10	58	-
			Uthal	<10	23	-
	Colon	a-Globin	∫ Hpa11	<10	<10	-
			LHhal	<10	<10	-
		нон	[Hpall	<10	41	
			LHha1	<10	38	-
		y-Globin	(Hpall	<10	50	
			Libert	~10	22	

Methylation changes in cancer detected by Nanopore Sequencing



Comparison of bisulfite sequencing and nanopore-based R7.3 data in reduced representation data sets from cancer and normal cells. (a) Raw data (points) and smoothed data (lines) for methylation, as determined by bisulfite sequencing (top) and nanopore-based sequencing using an R7.3 pore (bottom), in a genomic region from the human mammary epithelial cell line MCF10A (green) and metastatic mammary epithelial cell line MDA-MB-231 (orange). (b) Same region as in a but with individual nanopore reads plotted separately. Each CpG that can be called is a point. Blue indicates methylated; red indicates unmethylated.

Detecting DNA cytosine methylation using nanopore sequencing

Simpson, Workman, Zuzarte, David, Dursi, Timp (2017) Nature Methods. doi:10.1038/nmeth.4184

Causes of Cancer



Cancer is a Preventable Disease that Requires Major Lifestyle Changes

Anand et al (2008) Pharmaceutical Research. doi: 10.1007/s11095-008-9661-9



FAP = Familial Adenomatous Polyposis
HCV = Hepatitis C virus
HPV = Human papillomavirus
CLL = Chronic lymphocytic leukemia
AML = Acute myeloid leukemia

Fig. 1. The relationship between the number of stem cell divisions in the lifetime of a given tissue and the lifetime risk of cancer in that tissue. Values are from table S1, the derivation of which is discussed in the supplementary materials.

Variation in cancer risk among tissues can be explained by the number of stem cell divisions Tomasetti and Vogelstein (2015) Science. DOI: 10.1126/science.1260825



Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention Tomasetti, Li, and Vogelstein (2017) Science. DOI: 10.1126/science.aaf9011

The Six Hallmarks of Cancer



Hallmarks of Cancer

Hanahan and Weinberg (2000) Cell. http://doi.org/10.1016/S0092-8674(00)81683-9

Oncogenes



Proto-oncogene to oncogene

- *HER-2/neuHER-2/neu:* encodes for a cell surface receptor that can stimulate cell division. The HER-2/neu gene is amplified in up to 30% of human breast cancers.
- **RAS:** The Ras gene products are involved in kinase signaling pathways that ultimately control transcription of genes, regulating cell growth and differentiation.
- **MYC:** The Myc protein is a transcription factor and controls expression of several genes.
- **SRC:** First oncogene ever discovered. The Src protein is a tyrosine kinase, which regulates cell activity.
- *hTER:* Codes for an enzyme (telomerase) that maintains chromosome ends.

Tumor Suppressors



- **TP53:** a transcription factor that regulates cell division and cell death.
- *Rb:* alters the activity of transcription factors and therefore controls cell division.
- **APC:** controls the availability of a transcription factor.
- **PTEN:** acts by opposing the action of PI3K, which is essential for anti-apoptotic, pro-tumorogenic Akt activation.

TP53:The first and most important tumor suppressor

Mechanism of inactivating p53	Typical tumours	Effect of inactivation		
Amino-acid-changing mutation in the DNA- binding domain	Colon, breast, lung, bladder, brain, pancreas, stomach, oesophagus and many others	Prevents p53 from binding to specific DNA sequences and activating the adjacent genes		
Deletion of the carboxy- terminal domain	Occasional tumours at many different sites	Prevents the formation of tetramers of p53		
Multiplication of the MDM2 gene in the genome	Sarcomas, brain	Extra MDM2 stimulates the degradation of p53		
Viral infection	Cervix, liver, lymphomas	Products of viral oncogenes bind to and inactivate p53 in the cell, in some cases stimulating p53 degradation		
Deletion of the p14 ^{ARF} gene	Breast, brain, lung and others, expecially when p53 itself is not mutated	Failure to inhibit MDM2 and keep p53 degradation under control		
Mislocalization of p53 to the cytoplasm, outside the nucleus	Breast, neuroblastomas	Lack of p53 function (p53 functions only in the nucleus)		

Figure 1 The many ways in which p53 may malfunction in human cancers.

>10,000 known mutations >17,000 publications

Surfing the p53 network

Volgelstein et al (2000) Nature. DOI: 10.1038/35042675

DNA Repair Genes



• BRCA1 and BRCA2 (breast cancer type 1/2 susceptibility genes): Normally expressed in the cells of breast and other tissue, where they help repair damaged DNA, or destroy cells if DNA cannot be repaired. They are involved in the repair of chromosomal damage with an important role in the error-free repair of DNA double-strand breaks

Tumor Evolution



The Clonal Evolution of Tumor Cell Populations Peter C. Nowell (1976) *Science*. 194(4260):23-28 DOI: 10.1126/science.959840

Tumor Evolution



Evolution of the cancer genome

Yates & Campbell (2012) Nature Review Genetics. doi:10.1038/nrg3317

Tumor Heterogeneity



The evolution of tumour phylogenetics: principles and practice

Schwarz and Schaffer (2017) Nature Reviews Genetics. doi:10.1038/nrg.2016.170

Cancer Mutation Analysis



Vazquez M, de la Torre V, Valencia A (2012) Chapter 14: Cancer Genome Analysis. PLOS Computational Biology 8(12): e1002824. https://doi.org/10.1371/journal.pcbi.1002824

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002824



First Cancer Genome

nature

Vol 456 6 November 2008 doi:10.1038/nature07485

ARTICLES

DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome

Timothy J. Ley^{1,2,3,4}*, Elaine R. Mardis^{2,3}*, Li Ding^{2,3}, Bob Fulton³, Michael D. McLellan³, Ken Chen³, David Dooling³, Brian H. Dunford-Shore³, Sean McGrath³, Matthew Hickenbotham³, Lisa Cook³, Rachel Abbott³, David E. Larson³, Dan C. Koboldt³, Craig Pohl³, Scott Smith³, Amy Hawkins³, Scott Abbott³, Devin Locke³, LaDeana W. Hillier^{3,8}, Tracie Miner³, Lucinda Fulton³, Vincent Magrini^{2,3}, Todd Wylie³, Jarret Glasscock³, Joshua Conyers³, Nathan Sander³, Xiaoqi Shi³, John R. Osborne³, Patrick Minx³, David Gordon⁸, Asif Chinwalla³, Yu Zhao¹, Rhonda E. Ries¹, Jacqueline E. Payton⁵, Peter Westervelt^{1,4}, Michael H. Tomasson^{1,4}, Mark Watson^{3,4,5}, Jack Baty⁶, Jennifer Ivanovich^{4,7}, Sharon Heath^{1,4}, William D. Shannon^{1,4}, Rakesh Nagarajan^{4,5}, Matthew J. Walter^{1,4}, Daniel C. Link^{1,4}, Timothy A. Graubert^{1,4}, John F. DiPersio^{1,4} & Richard K. Wilson^{2,3,4}

Acute myeloid leukaemia is a highly malignant haematopoietic tumour that affects about 13,000 adults in the United States each year. The treatment of this disease has changed little in the past two decades, because most of the genetic events that initiate the disease remain undiscovered. Whole-genome sequencing is now possible at a reasonable cost and timeframe to use this approach for the unbiased discovery of tumour-specific somatic mutations that alter the protein-coding genes. Here we present the results obtained from sequencing a typical acute myeloid leukaemia genome, and its matched normal counterpart obtained from the same patient's skin. We discovered ten genes with acquired mutations; two were previously described mutations that are thought to contribute to tumour progression, and eight were new mutations present in virtually all tumour cells at presentation and relapse, the function of which is not yet known. Our study establishes whole-genome sequencing as an unbiased method for discovering cancer-initiating mutations in previously unidentified genes that may respond to targeted therapies.

First Melanoma Genome



- Insertions (light-green rectangles);
- Deletions (dark-green rectangles);
- Heterozygous (light-orange bars) and Homozygous (darkorange bars) Substitutions
- Coding substitutions (coloured squares: silent in grey, missense in purple, nonsense in red and splice site in black);
- Copy number (blue lines); regions of LOH (red lines);
- Intrachromosomal rearrangements (green lines);
- Interchromosomal rearrangements (purple lines).

A comprehensive catalogue of somatic mutations from a human cancer genome Pleasance et al (2010) Nature. doi:10.1038/nature08658

Mutations in Breast Cancer



Comprehensive molecular portraits of human breast tumours

Cancer Genome Atlas Network (2012) Nature. doi:10.1038/nature11412

Finding Driving Mutations



Integrative Annotation of Variants from 1092 Humans: Application to Cancer Genomics Khurana et al (2013) Science. DOI: 10.1126/science.1235587

Regulatory mutations in PDAC



Coding alterations of PDAC are now fairly well established but non-coding mutations (NCMs) largely unexplored

•Developed GECCO to analyze the thousands of somatic mutations observed from hundreds of tumors to find potential drivers of gene expression and pathogenesis

- •NCMs are enriched in known and novel pathways
- •NCMs correlate with changes in gene expression
- •NCMs can demonstrably modulate gene expression
- •NCMs correlate with novel clinical outcomes

NCMs are an important mechanism for tumor genome evolution

Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma Feigin, M, Garvin, T et al. (2017) Nature Genetics. doi:10.1038/ng.3861

Driving Non-Coding Mutations

CRR (MUT#)	Nearest gene	MUT allele	WT allele	Fold change	p-value	q-value
MAX (5)	PTPRN2	0.82	10.92	0.075	0.00593	0.09689
FOSL2 (7)	KCNQ1	0.85	6.39	0.133	0.02456	0.18212
TAF7 (9)	SNRPN	0.46	3.4	0.135	0.00818	0.11818
NFKB1 (7)	GYPC	1.08	7.29	0.148	0.01845	0.15157
TAF1 (6)	PDPN	2.09	13.08	0.160	0.03544	0.22016
BCLAF1 (5)	PRSS12	1.07	6.46	0.166	0.01107	0.14144
MAFK (3)	SOX5	0.29	1.63	0.178	0.02851	0.20379
POU2F2 (6)	MIR4420	8.16	40.24	0.203	0.01773	0.15157
WRNIP1 (3)	IKZF1	0.64	3.15	0.203	0.01811	0.15157
GATA3 (3)	PCLO	0.35	1.67	0.210	0.01113	0.14144
JUND (3)	TUSC7	0.98	4.53	0.216	0.02909	0.20560
REST (3)	MTERF4	1.46	5.78	0.253	0.02209	0.16542
GATA1 (3)	FNIP2	7.59	18.32	0.414	0.02588	0.18929
CEBPB (3)	PNPLA8	5.69	13.62	0.418	0.01726	0.15157
EGR1 (5)	SLC12A8	4.34	7.99	0.542	0.04185	0.23823
SIN3A (3)	FAM192A	20.31	30.48	0.666	0.01788	0.15157

a NCMs correlate with gene expression changes



Recurrent noncoding regulatory mutations in pancreatic ductal adenocarcinoma Feigin, M, Garvin, T et al. (2017) Nature Genetics. doi:10.1038/ng.3861

Tumor Heterogeneity



The evolution of tumour phylogenetics: principles and practice

Schwarz and Schaffer (2017) Nature Reviews Genetics. doi:10.1038/nrg.2016.170

Tumor Heterogeneity





Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples Cibulskis et al (2013) Nature Biotech. doi:10.1038/nbt.2514



THetA: inferring intra-tumor heterogeneity from high-throughput DNA sequencing data Oesperet al (2013) Genome Biology. DOI: 10.1186/gb-2013-14-7-r80

Tumor Heterogeneity



Evolution of the cancer genome

Yates & Campbell (2012) Nature Review Genetics. doi:10.1038/nrg3317



Tumor Heterogeneity



The evolution of tumour phylogenetics: principles and practice

Schwarz and Schaffer (2017) Nature Reviews Genetics. doi:10.1038/nrg.2016.170

Tumour evolution inferred by single-cell sequencing

Nicholas Navin^{1,2}, Jude Kendall¹, Jennifer Troge¹, Peter Andrews¹, Linda Rodgers¹, Jeanne McIndoo¹, Kerry Cook¹, Asya Stepansky¹, Dan Levy¹, Diane Esposito¹, Lakshmi Muthuswamy³, Alex Krasnitz¹, W. Richard McCombie¹, James Hicks¹ & Michael Wigler¹

LETTER



Gingko http://qb.cshl.edu/ginkgo

Interactive Single Cell CNV analysis & clustering

- Easy-to-use, web interface, parameterized for binning, segmentation, clustering, etc
- Per cell through project-wide analysis in any species

Compare MDA, DOP-PCR, and MALBAC

DOP-PCR shows superior resolution and consistency

Available for collaboration

- Analyzing CNVs with respect to different clinical outcomes
- Extending clustering methods, prototyping scRNA





Interactive analysis and assessment of single-cell copy-number variations. Garvin et al. (2015) Nature Methods doi:10.1038/nmeth.3578

Realtime CNV Analysis



illumina®



Chromosome

Single Cell RNA-seq of Cancer





Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma Tirosh et al (2016) Nature. doi:10.1038/nature20123

Tumor Heterogeneity and Treatment



Clonal evolution in relapsed acute myeloid leukemia revealed by whole genome sequencing Ding et al (2012) Nature. doi:10.1038/nature10738

Liquid Biopsies



Liquid biopsies come of age: towards implementation of circulating tumour DNA Wan et al (2017) Nature Review Cancer. doi:10.1038/nrc.2017.7

Liquid Biopsies



Liquid biopsies come of age: towards implementation of circulating tumour DNA Wan et al (2017) Nature Review Cancer. doi:10.1038/nrc.2017.7