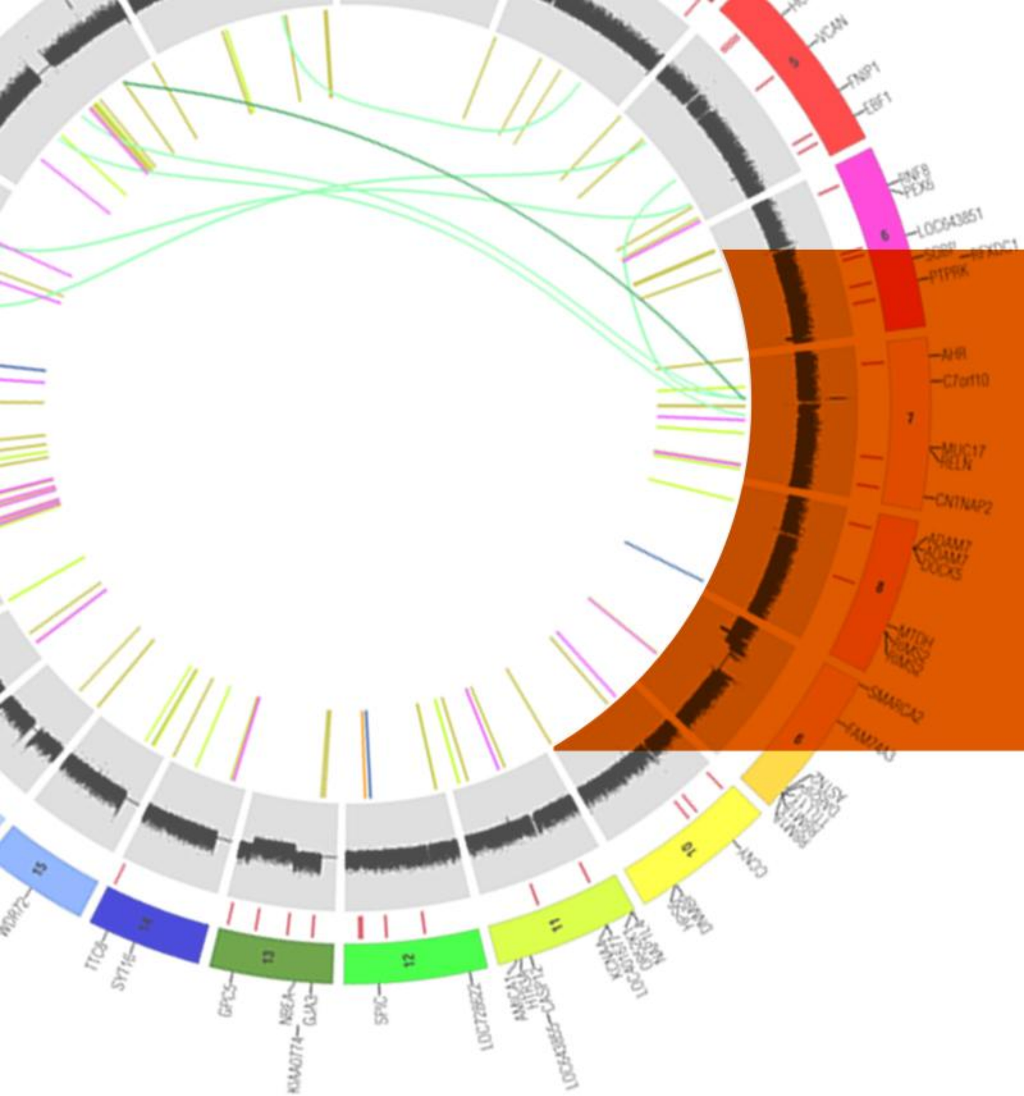
The Cancer Genome Atlas

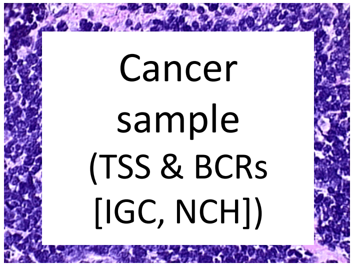


Sequence-based RNA profiling

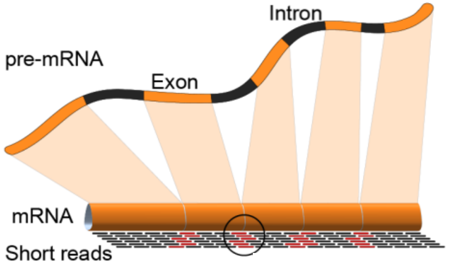
Expression maps at base-pair resolution

TCGA at a glance

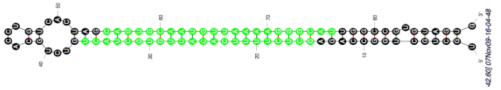
Clinical data (TSS, DCC)



Cancer sample (TSS & BCRs [IGC, NCH])

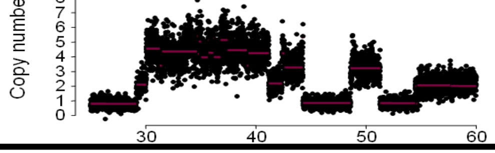


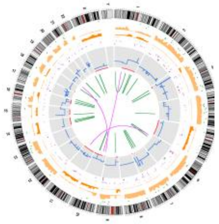
Gene (mRNA) expression (UNC & BCGSC)



miRNA expression (BCGSC)

Copy number (Harvard & Broad)

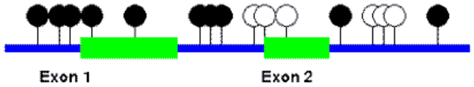




Genome sequencing (Broad, Wash U, Baylor)

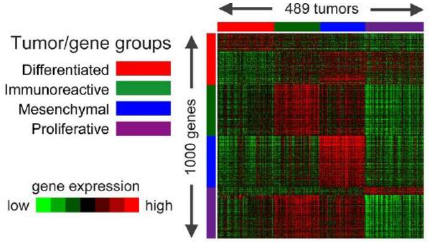
TARGET REGION A
TARGET REGION B
TARGET REGION C

Exome sequencing (Broad, Wash U, Baylor)

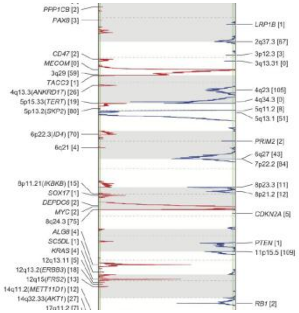


Epigenomics (USC & JHU)

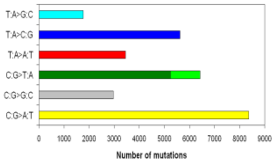
Genome Data Analysis Centres (Broad, ISB, LBNL, MSKCC, UCSC, UNC, UofT/MDACC)



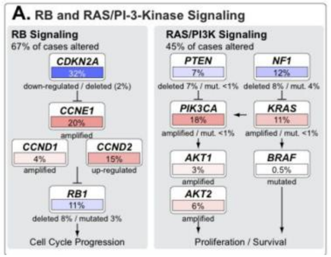
Expression profiles



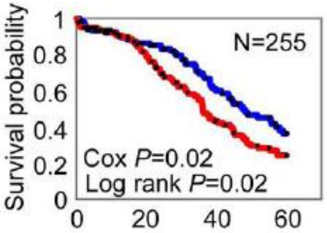
Recurrent events



Mutations



Pathways



Outcomes

The Cancer Genome Atlas

Acknowledgements

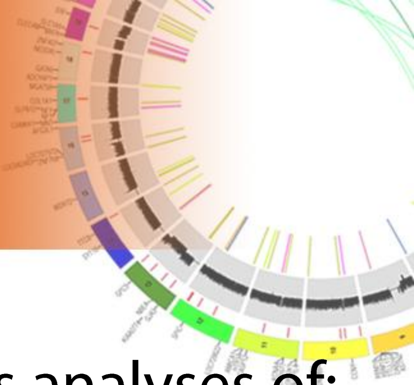
- Gordon Robertson
- Andy Chu
- Andy Mungall
- Dominik Stoll
- Payal Sipahimalani
- Elizabeth Chun
- Jared Slobodan
- Robin Coope
- Yisu Li
- Ryan Morin
- Inanc Birol
- Steven Jones

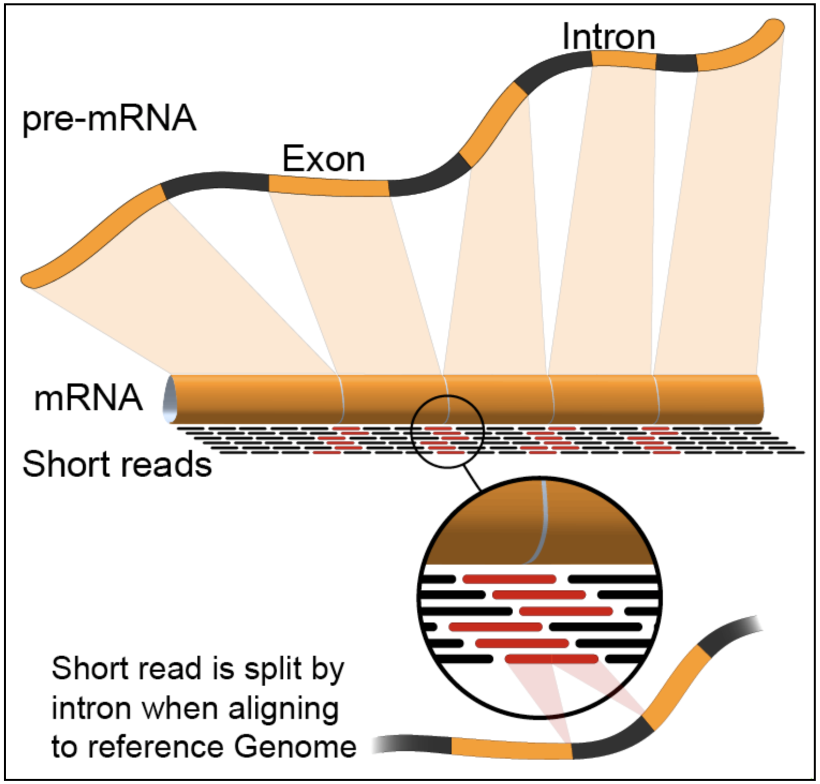
Patients

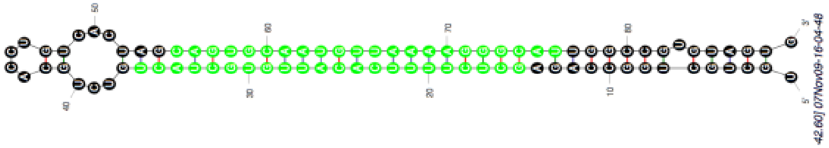
The TCGA Research Network

- Chuck Perou, UNC
- Neil Hayes, UNC
- Katie Hoadley, UNC
- Todd Auman, UNC
- Matt Wilkerson, UNC
- Chad Creighton, BCM
- Angela Hadjipanayis, Harvard
- Sorana Morrissy, Sickkids (TO)
- Malachi Griffith, WashU
- Timothy Ley, WashU
- Li Ding, WashU
- Peter Westervelt, WashU
- Elaine Mardis, WashU
- Richard Wilson, WashU

Applications

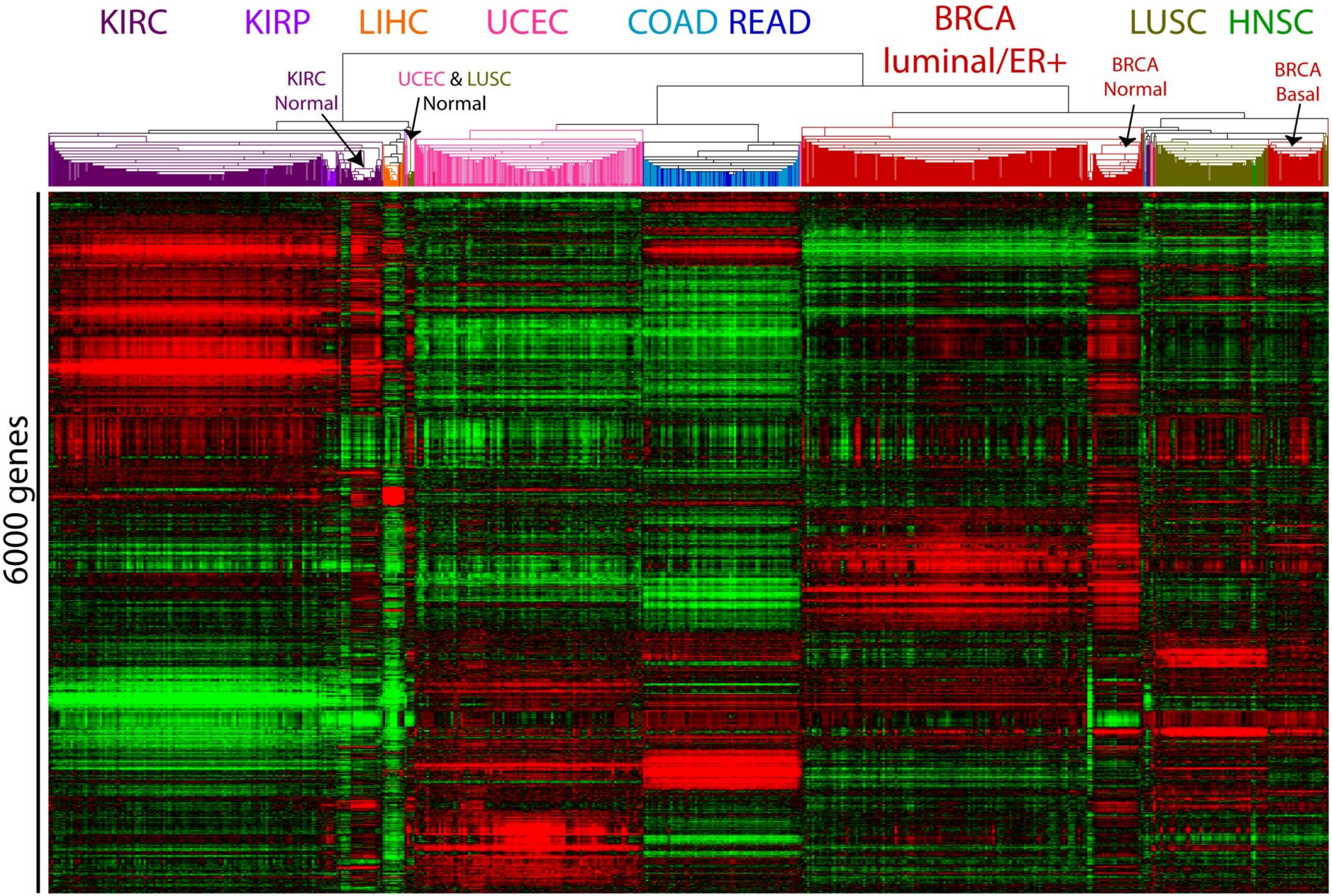






- RNA Seq enables analyses of:
  - gene expression
  - isoform expression
  - gene-fusion detection
  - “expressed mutations”
  - cancer sub-types
  - …
- miRNA Seq enables analyses of:
  - cancer sub-types
  - regulatory networks
  - …

The Cancer Genome Atlas



Chuck Perou

1,530 Samples/lanes (DCC)

Analysis tools (Garber et al., *Nat Meth* 2011)

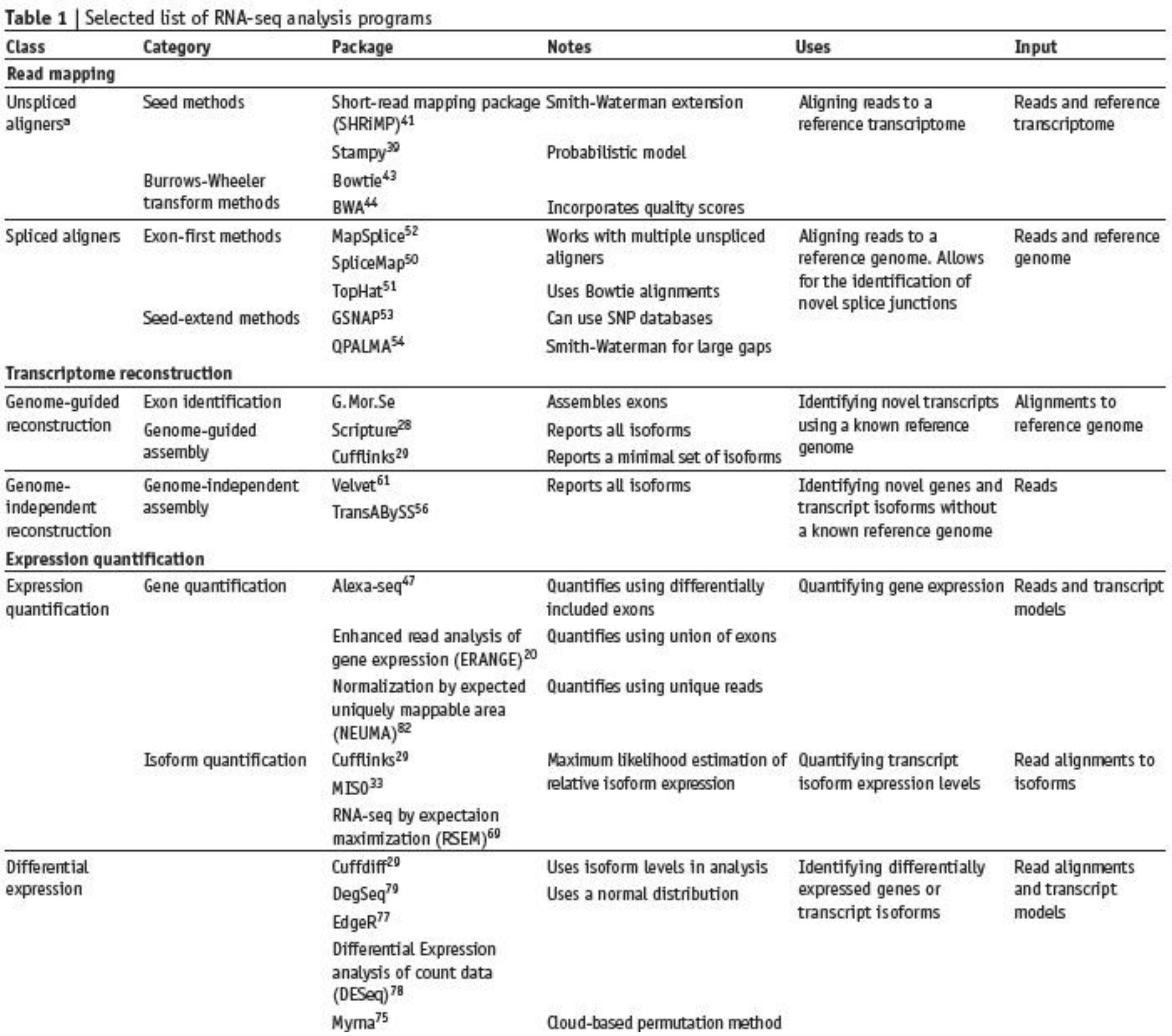
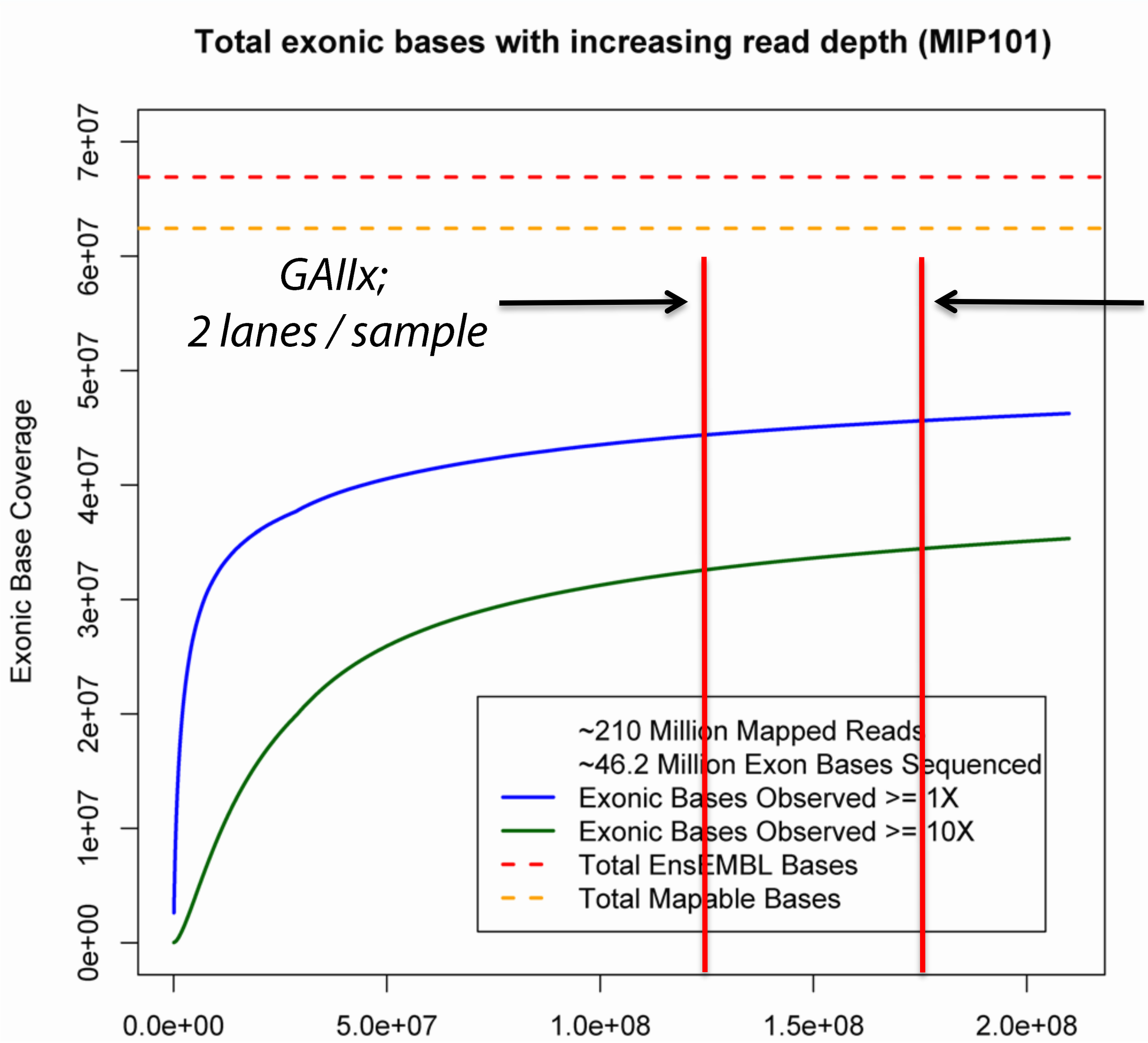


Table 1 | Selected list of RNA-seq analysis programs

Class	Category	Package	Notes	Uses	Input
Read mapping					
Unspliced aligners[a]	Seed methods	Short-read mapping package (SHRiMP)[41]	Smith-Waterman extension	Aligning reads to a reference transcriptome	Reads and reference transcriptome
		Stampy[39]	Probabilistic model		
	Burrows-Wheeler transform methods	Bowtie[43]			
		BWA[44]	Incorporates quality scores		
Spliced aligners	Exon-first methods	MapSplice[52]	Works with multiple unspliced aligners	Aligning reads to a reference genome. Allows for the identification of novel splice junctions	Reads and reference genome
		SpliceMap[50]			
		TopHat[51]	Uses Bowtie alignments		
	Seed-extend methods	GSNAP[53]	Can use SNP databases		
		QPALMA[54]	Smith-Waterman for large gaps		
Transcriptome reconstruction					
Genome-guided reconstruction	Exon identification	G.Mor.Se	Assembles exons	Identifying novel transcripts using a known reference genome	Alignments to reference genome
	Genome-guided assembly	Scripture[28]	Reports all isoforms		
		Cufflinks[29]	Reports a minimal set of isoforms		
Genome-independent reconstruction	Genome-independent assembly	Velvet[61]	Reports all isoforms	Identifying novel genes and transcript isoforms without a known reference genome	Reads
		TransABySS[56]			
Expression quantification					
Expression quantification	Gene quantification	Alexa-seq[47]	Quantifies using differentially included exons	Quantifying gene expression	Reads and transcript models
		Enhanced read analysis of gene expression (ERANGE)[20]	Quantifies using union of exons		
		Normalization by expected uniquely mappable area (NEUMA)[82]	Quantifies using unique reads		
	Isoform quantification	Cufflinks[29]	Maximum likelihood estimation of relative isoform expression	Quantifying transcript isoform expression levels	Read alignments to isoforms
		MISO[33]			
		RNA-seq by expectaion maximization (RSEM)[69]			
Differential expression		Cuffdiff[29]	Uses isoform levels in analysis	Identifying differentially expressed genes or transcript isoforms	Read alignments and transcript models
		DegSeq[79]	Uses a normal distribution		
		EdgeR[77]			
		Differential Expression analysis of count data (DESeq)[78]			
		Myrna[75]	Cloud-based permutation method		

RNA Seq read depth and coverage


Total exonic bases with increasing read depth (MIP101)
Exonic Base Coverage
0e+00
1e+07
2e+07
3e+07
4e+07
5e+07
6e+07
7e+07
0.0e+00
5.0e+07
1.0e+08
1.5e+08
2.0e+08
~210 Million Mapped Reads
~46.2 Million Exon Bases Sequenced
🟦 Exonic Bases Observed >= 1X
🟩 Exonic Bases Observed >= 10X
🟥 Total EnsEMBL Bases
🟧 Total Mapable Bases

GAIIx;
2 lanes / sample

HiSeq v3;
2 samples per lane

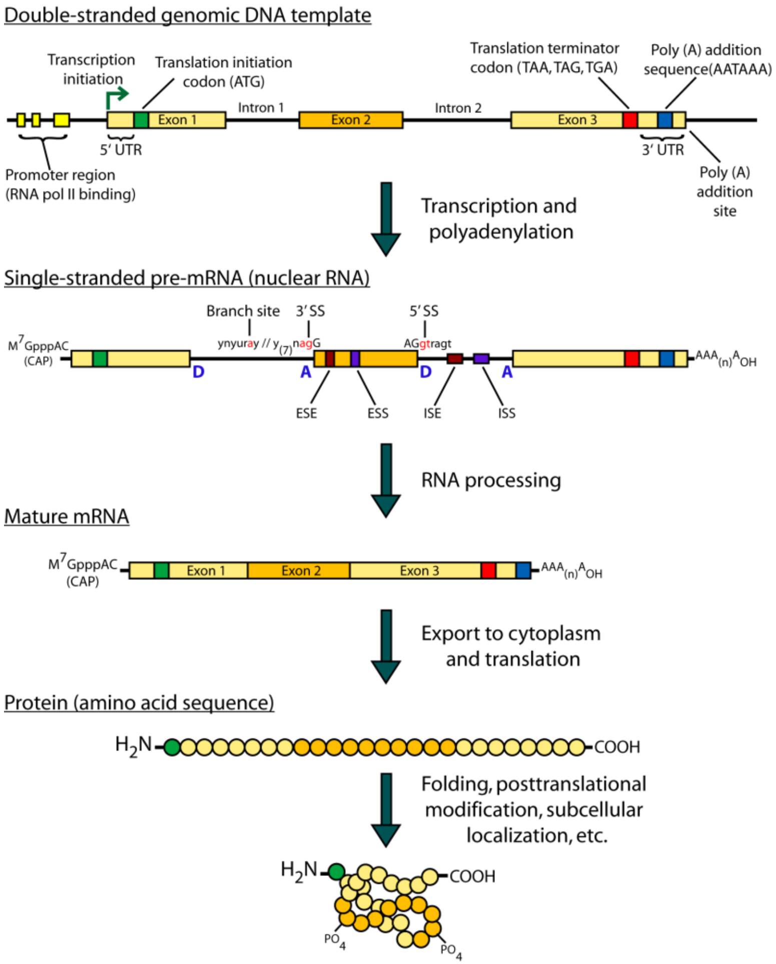
Malachi Griffith, Elizabeth Chun, Yisu Li

The Cancer Genome Atlas

Exon wiring maps

Alternative Expression Modes

Gene expression



Double-stranded genomic DNA template

Transcription initiation
Translation initiation codon (ATG)
Translation terminator codon (TAA, TAG, TGA)
Poly (A) addition sequence(AATAAA)
Intron 1
Intron 2
Exon 1
Exon 2
Exon 3
5′ UTR
3′ UTR
Promoter region (RNA pol II binding)
Poly (A) addition site

Transcription and polyadenylation

Single-stranded pre-mRNA (nuclear RNA)

RNA processing

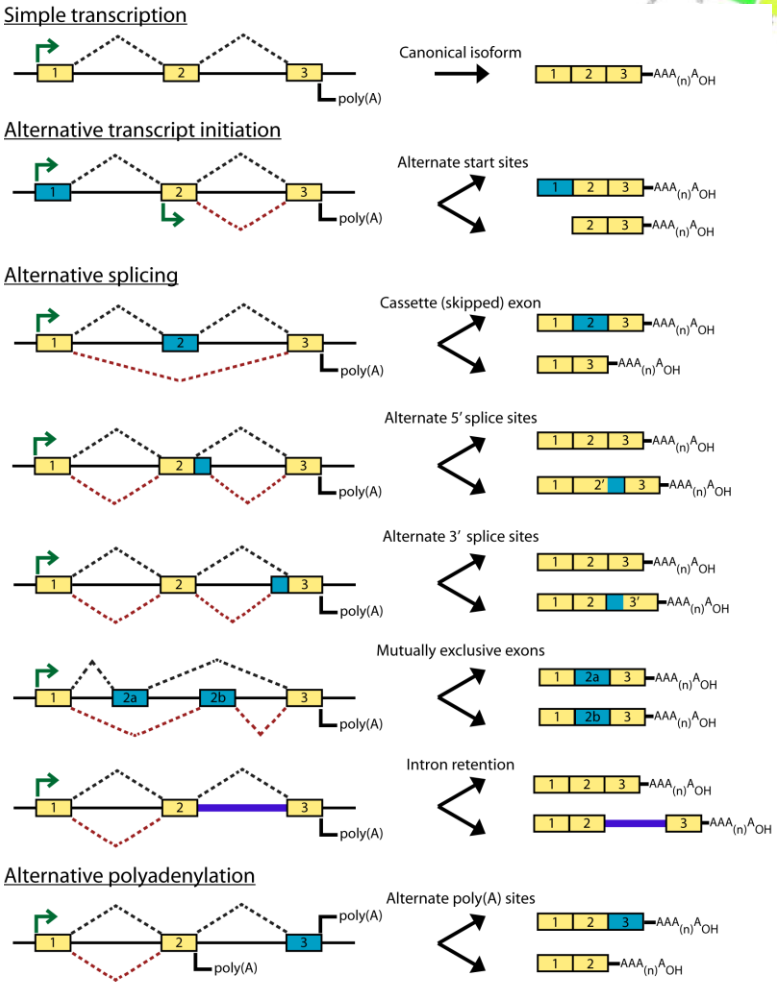
Mature mRNA

Export to cytoplasm and translation

Protein (amino acid sequence)

Folding, posttranslational modification, subcellular localization, etc.

Types of alternative expression



Simple transcription

Alternative transcript initiation

Alternative splicing

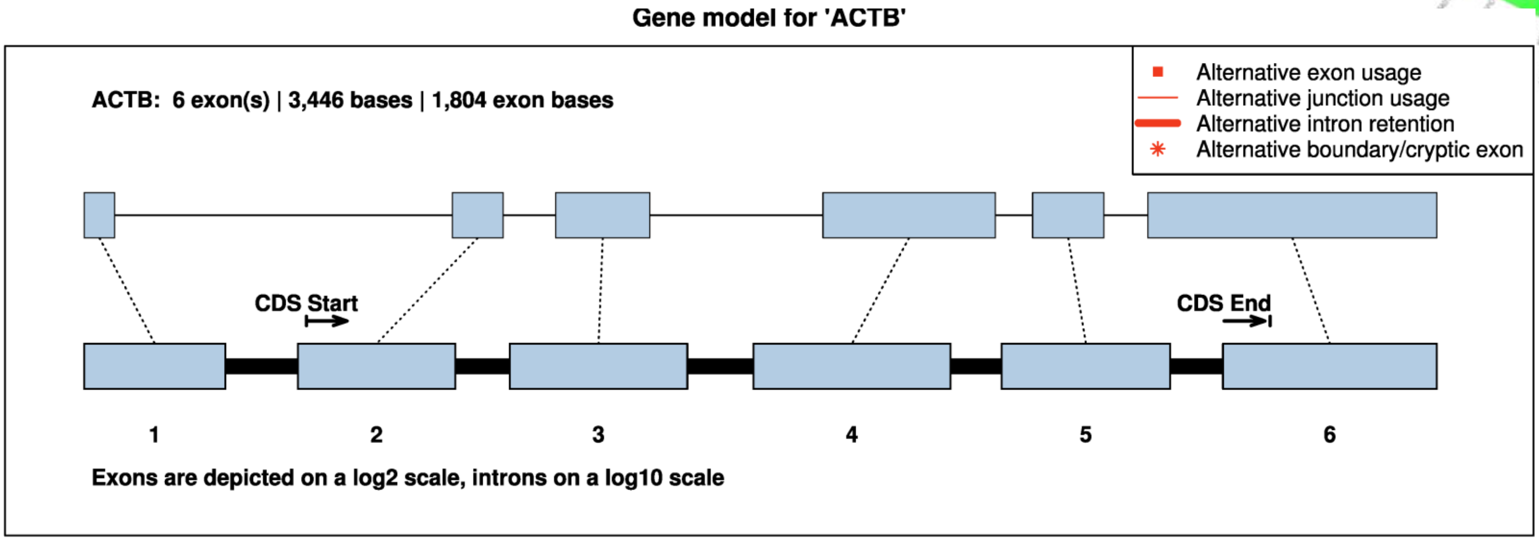
Alternative polyadenylation

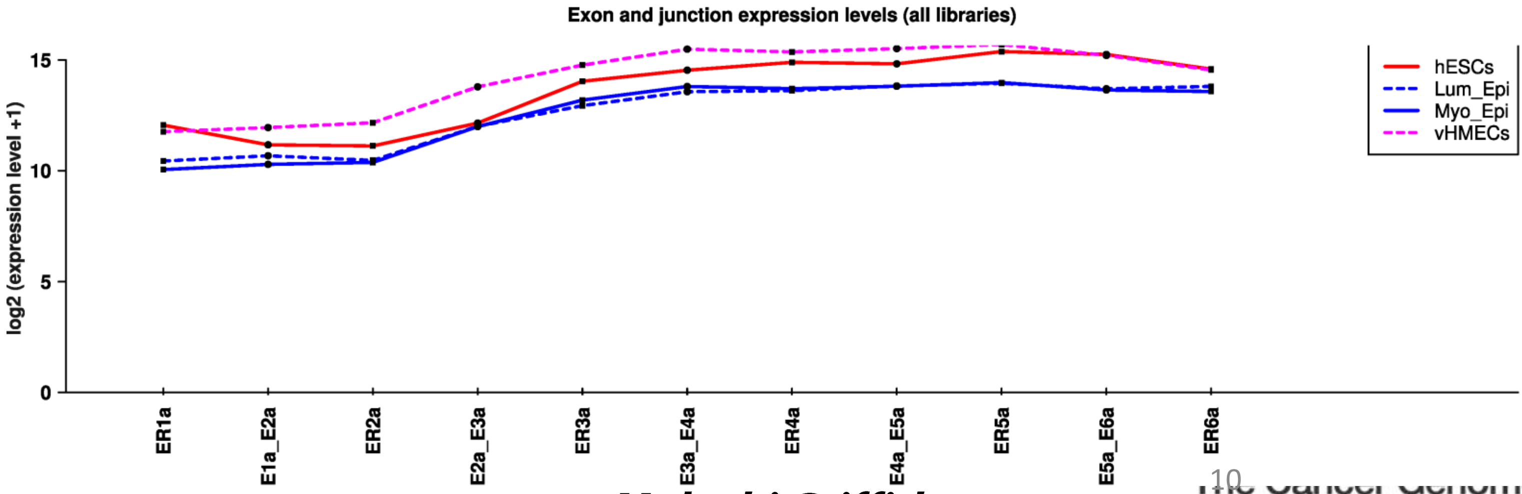
Malachi Griffith

The Cancer Genome Atlas

Actin (cell lines)

www.AlexaPlatform.org/alexa_seq/

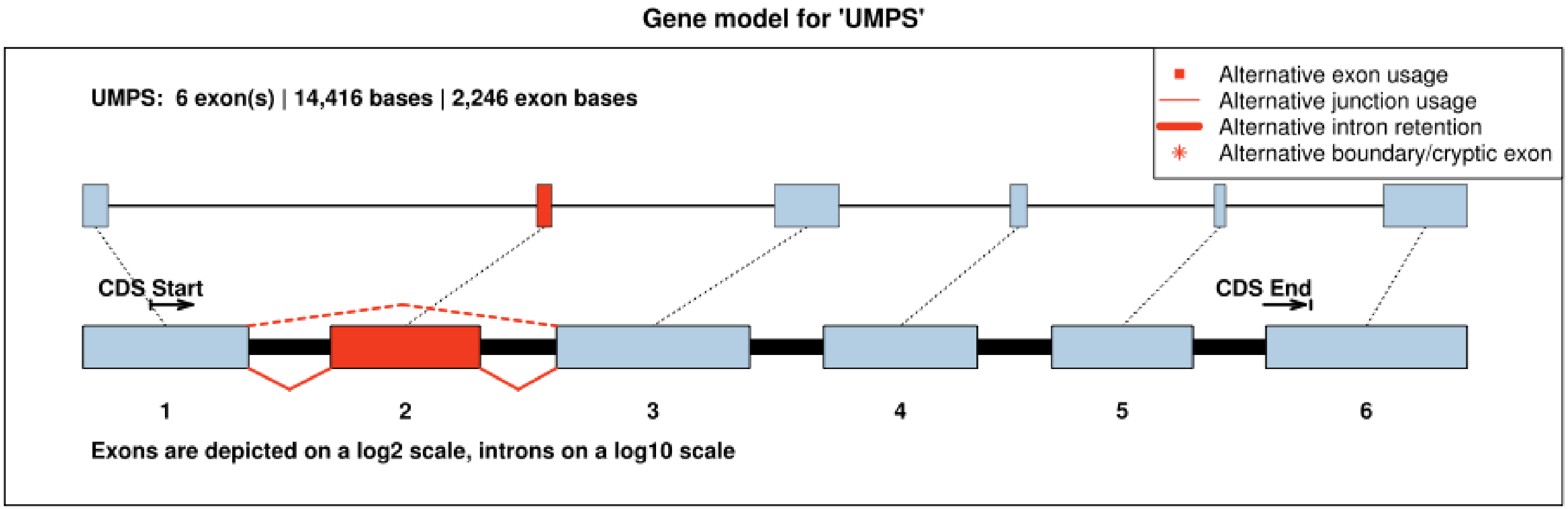

Gene model for 'ACTB'
ACTB: 6 exon(s) | 3,446 bases | 1,804 exon bases
■ Alternative exon usage
Alternative junction usage
Alternative intron retention
✱ Alternative boundary/cryptic exon
CDS Start
CDS End
1
2
3
4
5
6
Exons are depicted on a log2 scale, introns on a log10 scale

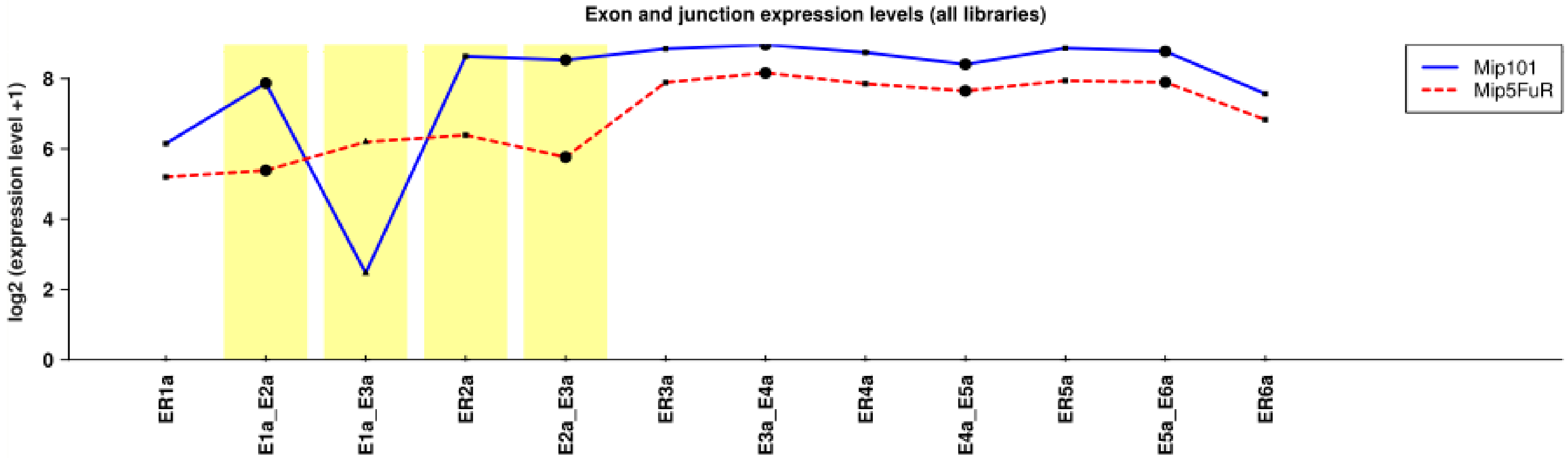


Malachi Griffith

The Cancer Genome Atlas

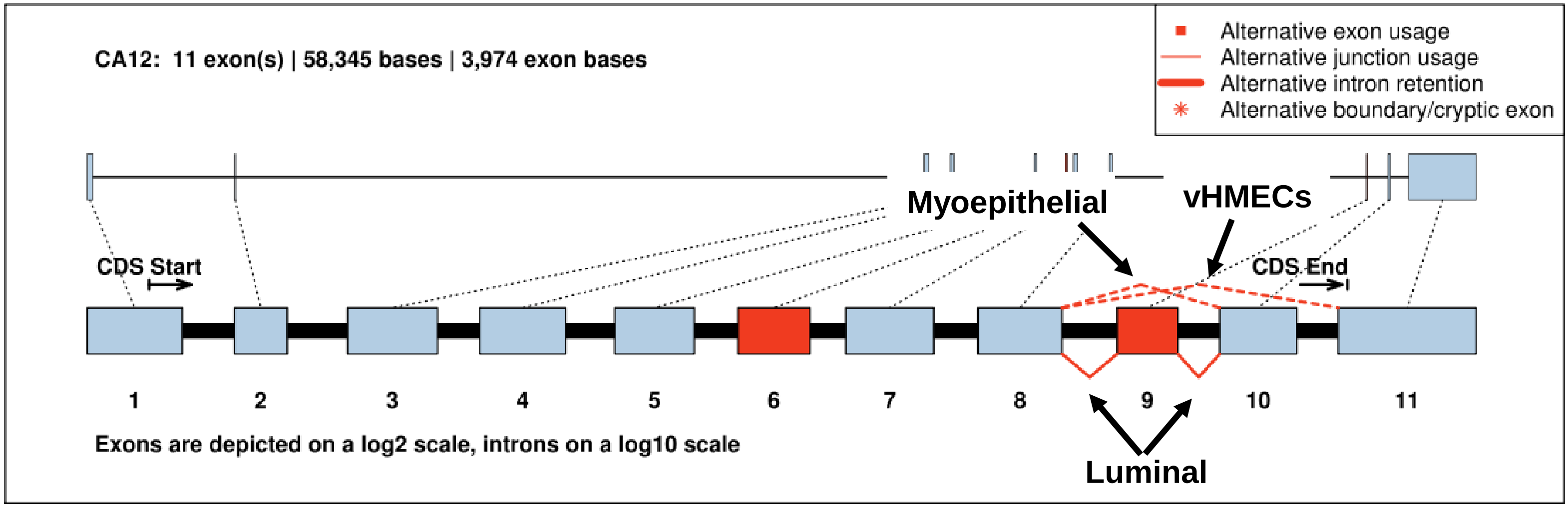
UMPS (cell lines)


Gene model for 'UMPS'
UMPS: 6 exon(s) | 14,416 bases | 2,246 exon bases
■ Alternative exon usage
— Alternative junction usage
▬ Alternative intron retention
✳ Alternative boundary/cryptic exon
CDS Start
CDS End
1
2
3
4
5
6
Exons are depicted on a log2 scale, introns on a log10 scale



CA12 (cell lines)

Gene model for 'CA12'

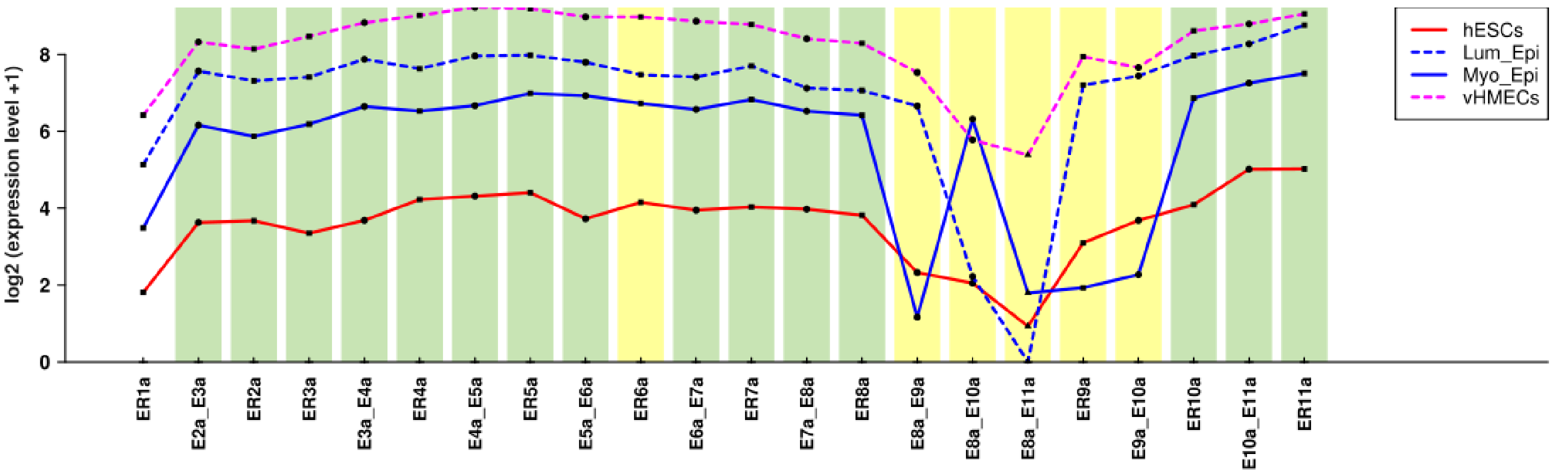

CA12: 11 exon(s) | 58,345 bases | 3,974 exon bases
■ Alternative exon usage
— Alternative junction usage
▬ Alternative intron retention
✳ Alternative boundary/cryptic exon
CDS Start
CDS End
1
2
3
4
5
6
7
8
9
10
11
Exons are depicted on a log2 scale, introns on a log10 scale

Myoepithelial

vHMECs

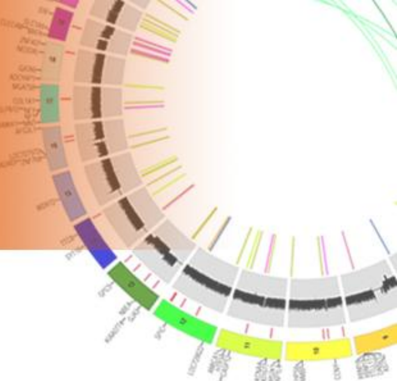
Luminal

Exon and junction expression levels (all libraries)

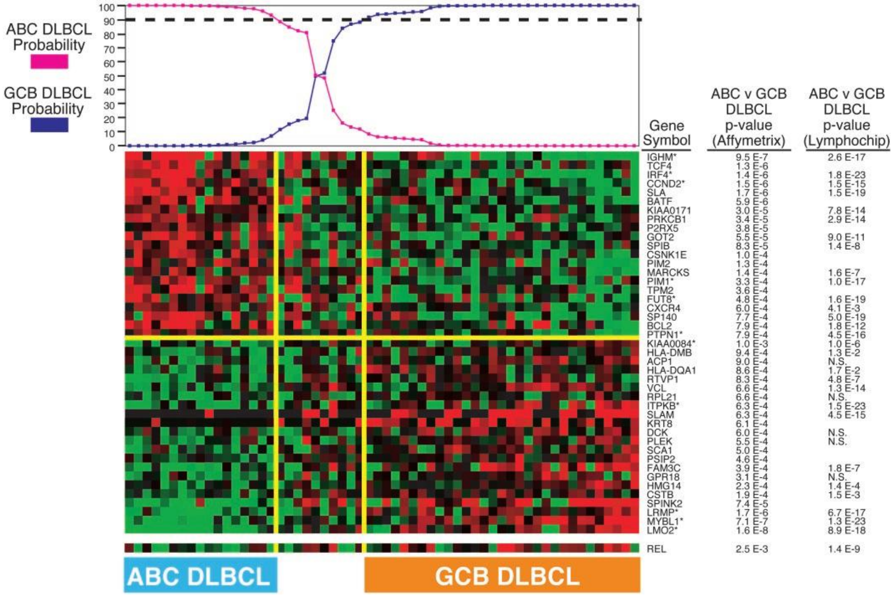


Malachi Griffith

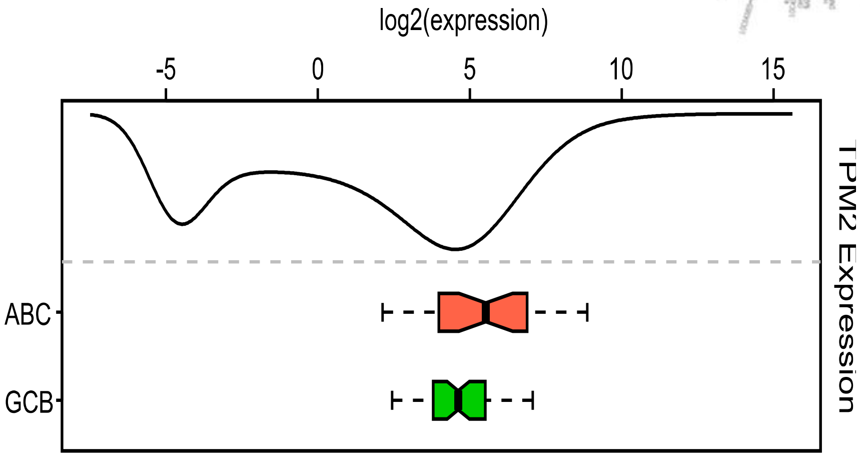
TPM2 (DLBCL)

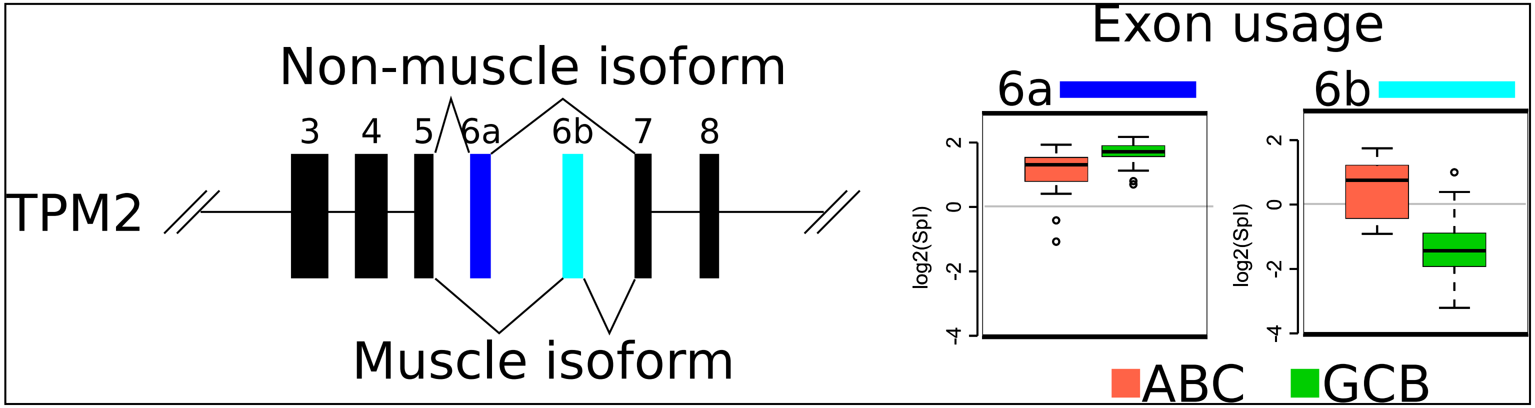


ABC vs. GCB gene expression classifier
Wright et al, 2003



Gene Symbol	ABC v GCB DLBCL p-value (Affymetrix)	ABC v GCB DLBCL p-value (Lymphochip)
IGHM*	9.5 E-7	2.6 E-17
TCF4	1.3 E-6	
IRF4*	1.4 E-6	1.8 E-23
CCND2*	1.5 E-6	1.5 E-15
SLA	1.7 E-6	1.5 E-19
BATF	5.9 E-6	
KIAA0171	3.0 E-5	7.8 E-14
PRKCB1	3.4 E-5	2.9 E-14
P2RX5	3.8 E-5	
GOT2	5.5 E-5	9.0 E-11
SPIB	8.3 E-5	1.4 E-8
CSNK1E	1.0 E-4	
PIM2	1.3 E-4	
MARCKS	1.4 E-4	1.6 E-7
PIM1*	3.3 E-4	1.0 E-17
TPM2	3.6 E-4	
FUT8*	4.8 E-4	1.6 E-19
CXCR4	6.0 E-4	4.1 E-3
SP140	7.7 E-4	5.0 E-19
BCL2	7.9 E-4	1.8 E-12
PTPN1*	7.9 E-4	4.5 E-16
KIAA0084*	1.0 E-3	1.0 E-6
HLA-DMB	9.4 E-4	1.3 E-2
ACP1	9.0 E-4	N.S.
HLA-DQA1	8.6 E-4	1.7 E-2
RTVP1	8.3 E-4	4.8 E-7
VCL	6.6 E-4	1.3 E-14
RPL21	6.6 E-4	N.S.
ITPKB*	6.3 E-4	1.5 E-23
SLAM	6.3 E-4	4.5 E-15
KRT8	6.1 E-4	
DCK	6.0 E-4	N.S.
PLEK	5.5 E-4	N.S.
SCA1	5.0 E-4	
PSIP2	4.6 E-4	
FAM3C	3.9 E-4	1.8 E-7
GPR18	3.1 E-4	N.S.
HMG14	2.3 E-4	1.4 E-4
CSTB	1.9 E-4	1.5 E-3
SPINK2	7.4 E-5	
LRMP*	1.7 E-6	6.7 E-17
MYBL1*	7.1 E-7	1.3 E-23
LMO2*	1.6 E-8	8.9 E-18
REL	2.5 E-3	1.4 E-9





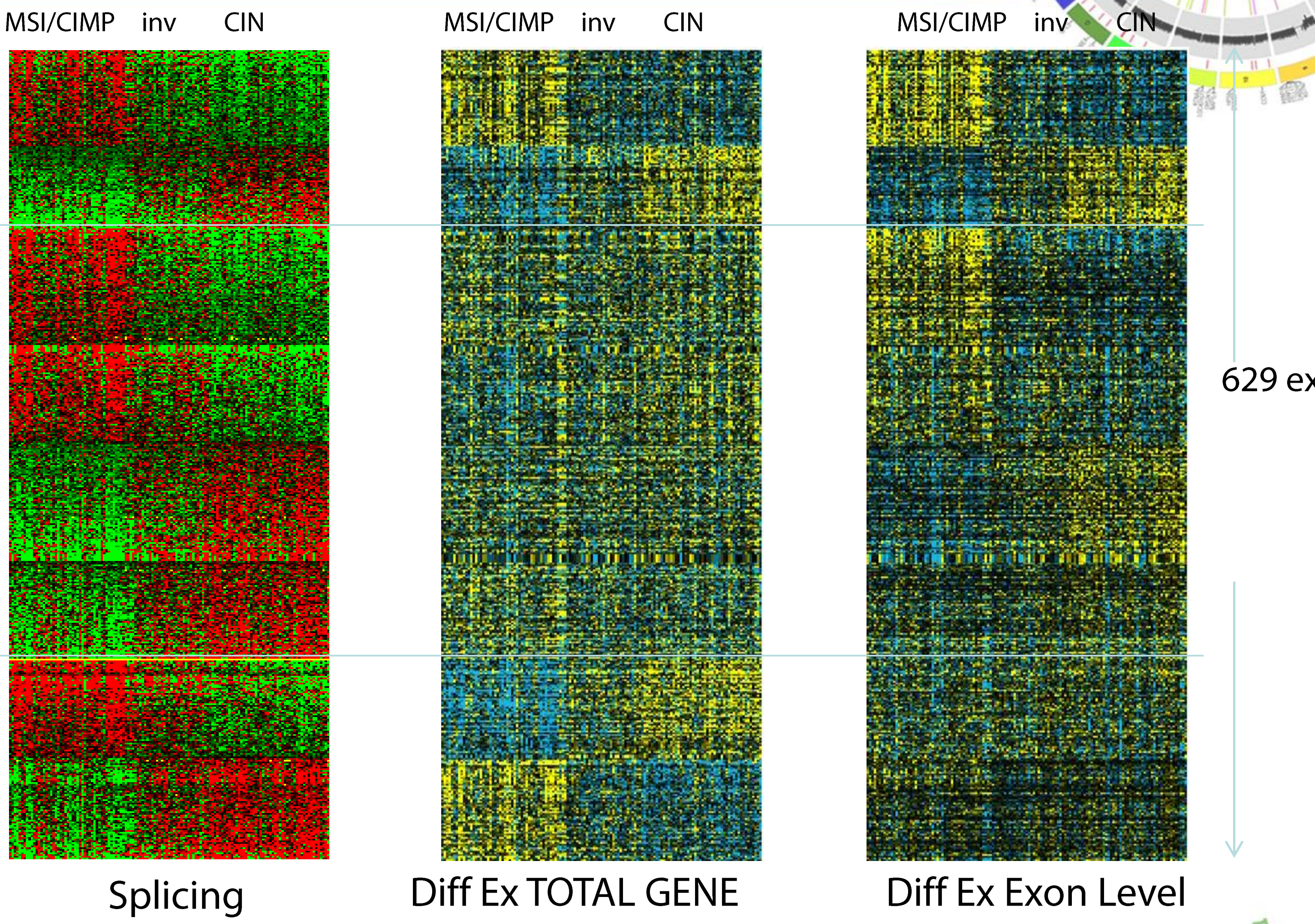
Rodrigo Goya

Exon-level expression in CRC

Splicing patterns CORRELATED with total gene levels

Splicing patterns NOT CORRELATED with total gene levels

Splicing patterns ANTI-CORRELATED with total gene levels



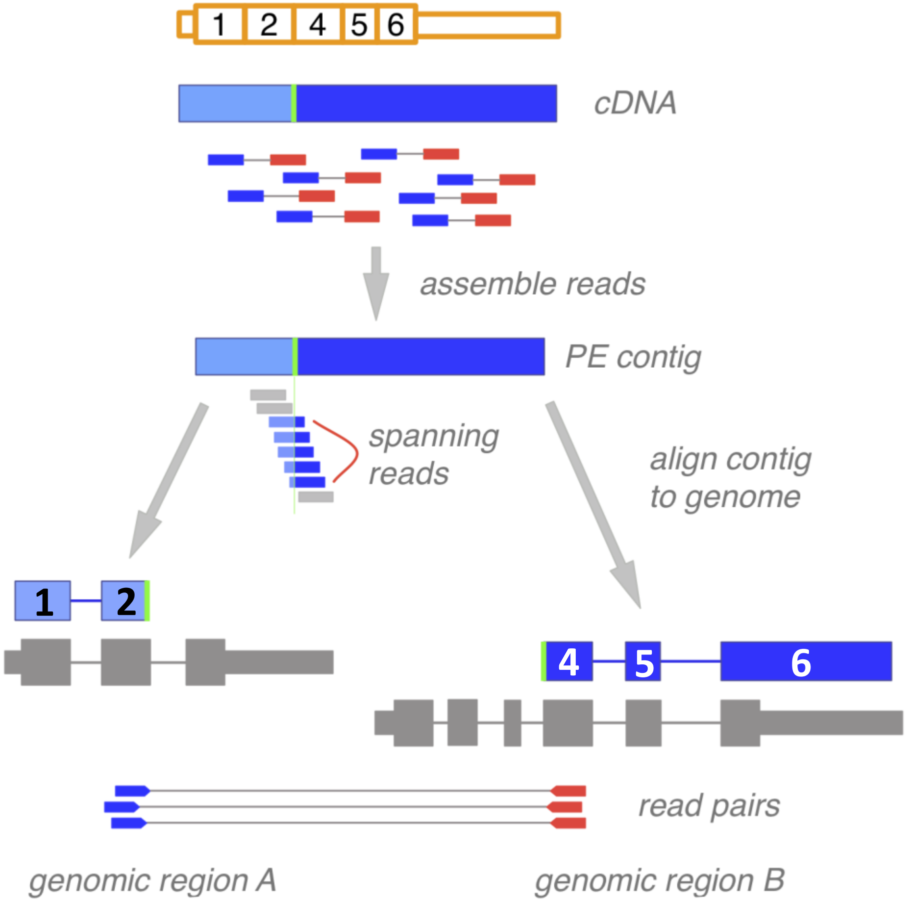
<u>Chad Creighton, BCM</u>

The Cancer Genome Atlas

De novo assembly

Detection of "fusion genes"

Trans-ABySS (Robertson, G. et al. 2010 Nature Methods 7(11):909-12)



•**Alignment-independent detection of**:

•*Gene fusions*
•*Alternative transcripts*
•*Internal tandem duplications*
•*Partial tandem duplications*
•*Insertions / deletions*

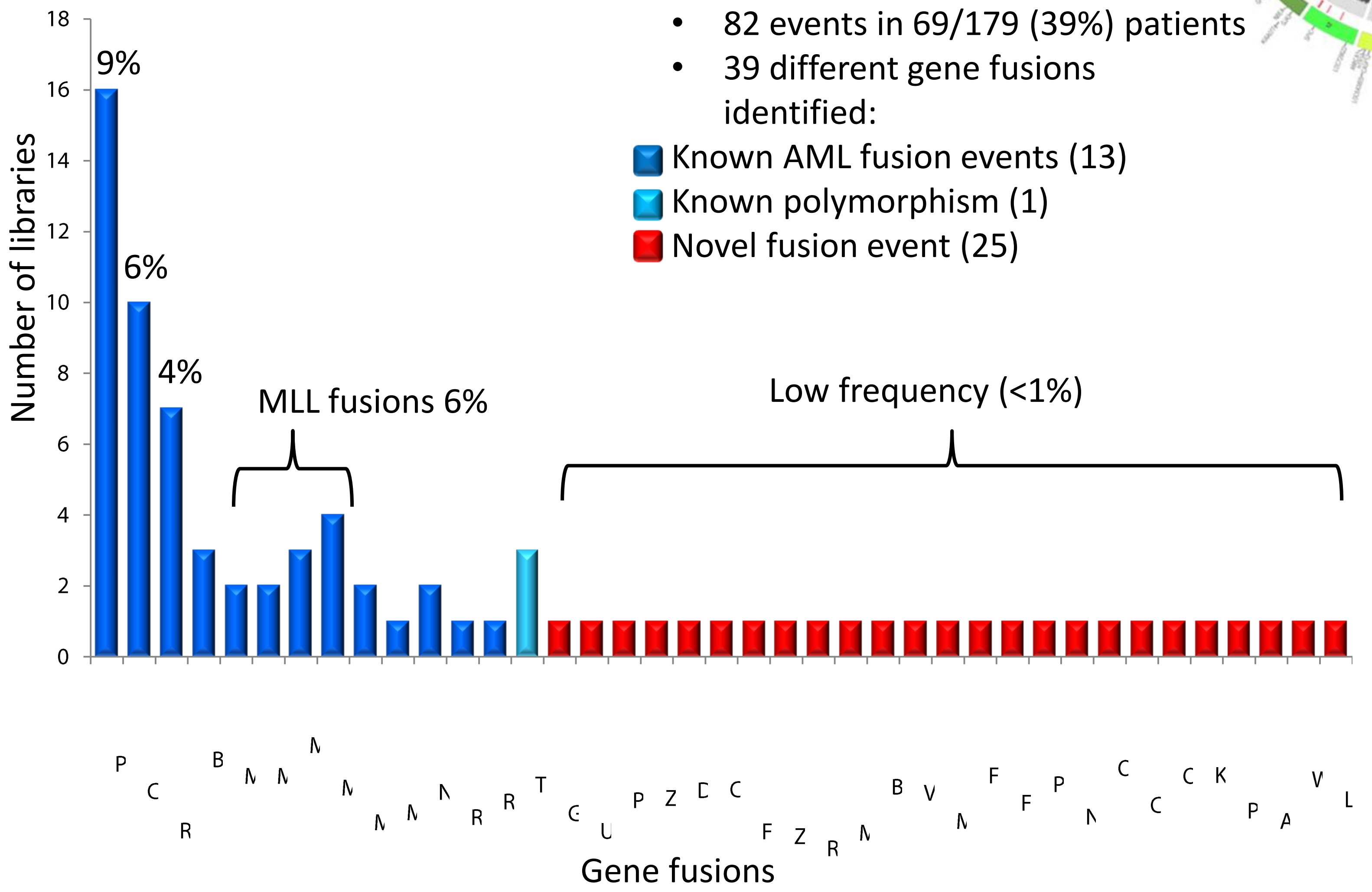
Verified AML gene fusions

- 82 events in 69/179 (39%) patients
- 39 different gene fusions identified:

🟦 Known AML fusion events (13)

🟦 Known polymorphism (1)

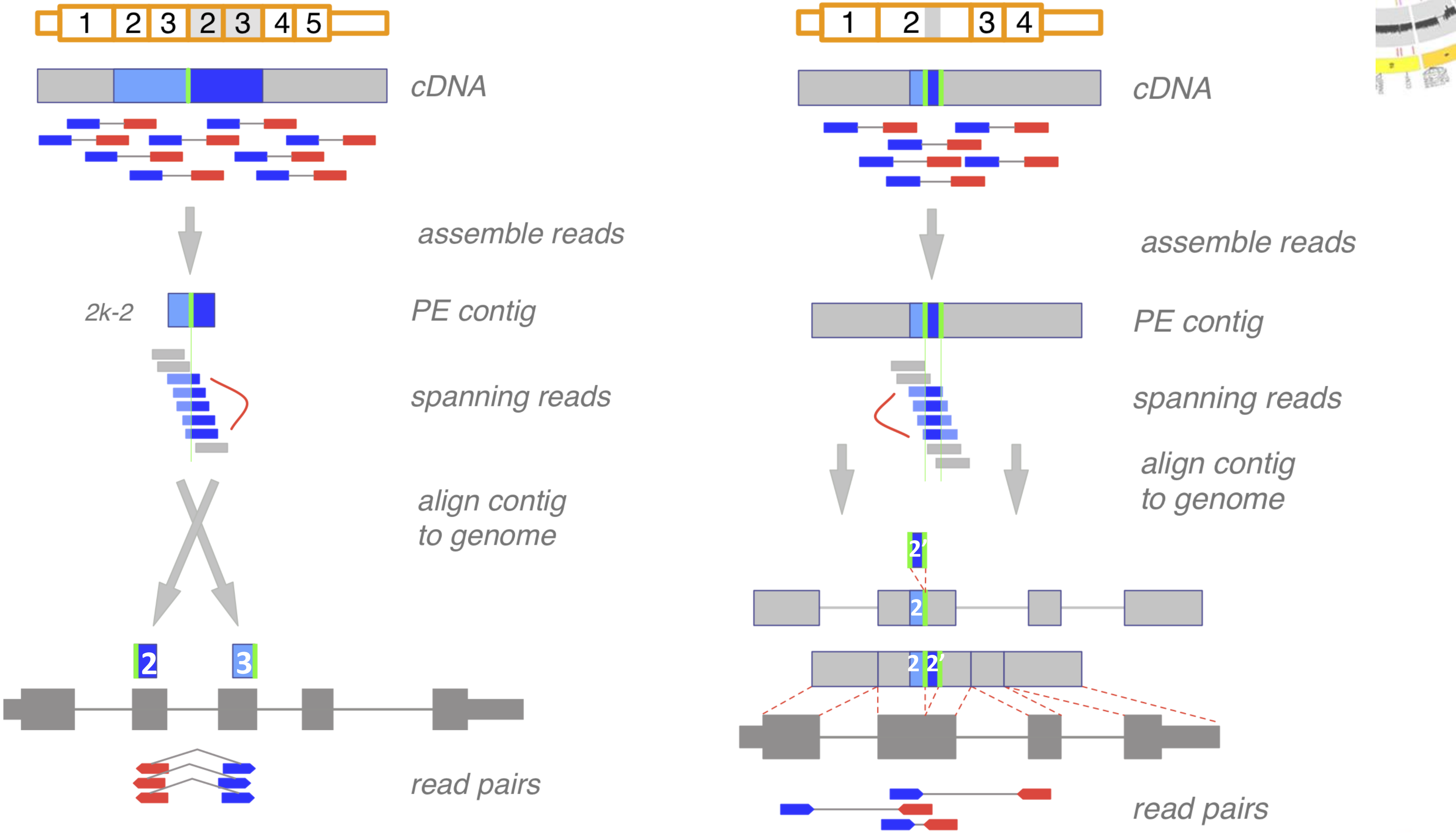
🟥 Novel fusion event (25)



See A. Mungall poster

The Cancer Genome Atlas

Detecting PTDs & ITDs

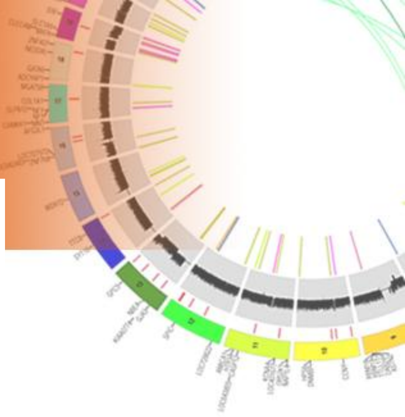


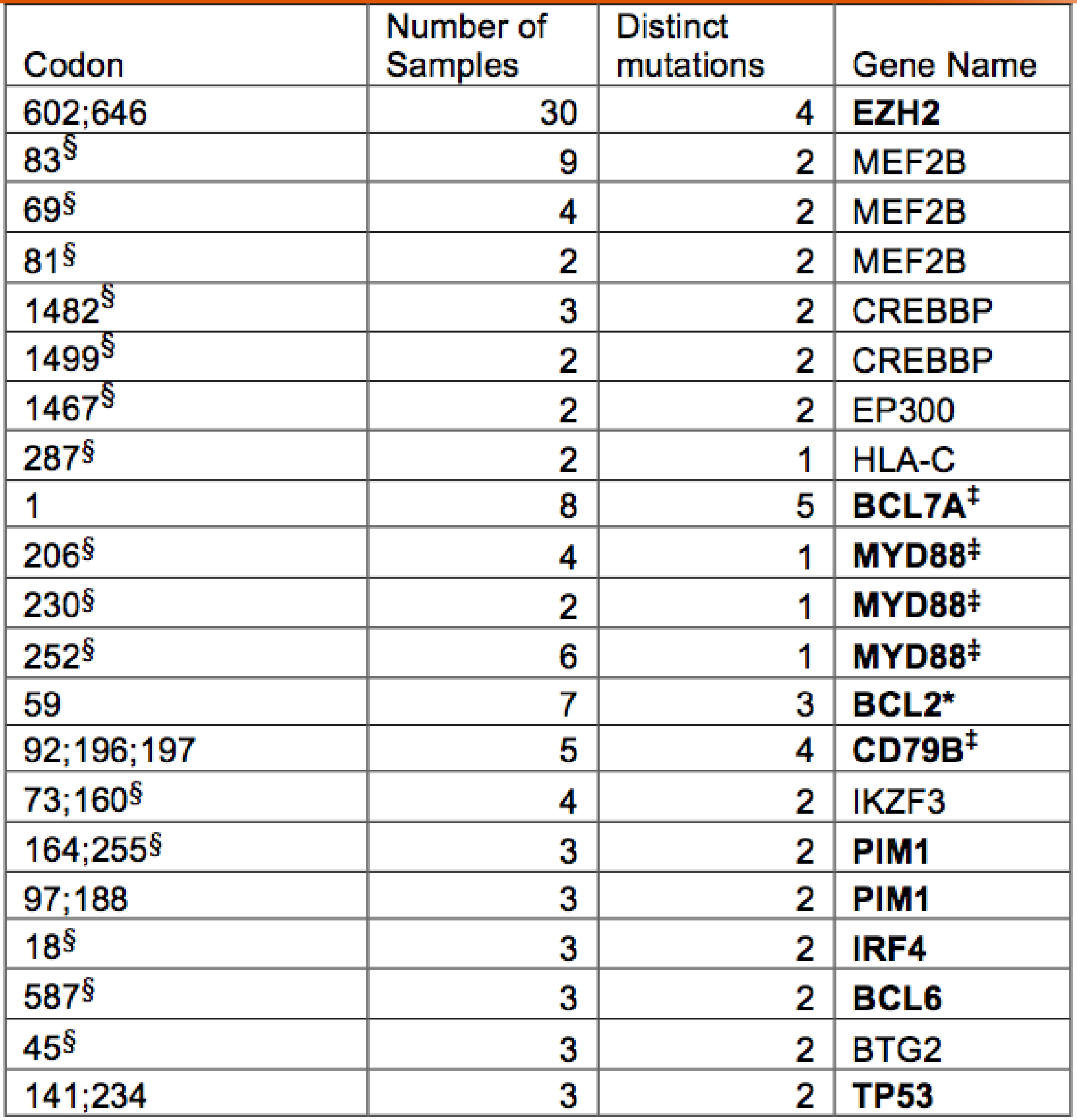
Gordon Robertson

The Cancer Genome Atlas

“Expressed mutations”

RNA Seq for mutation detection

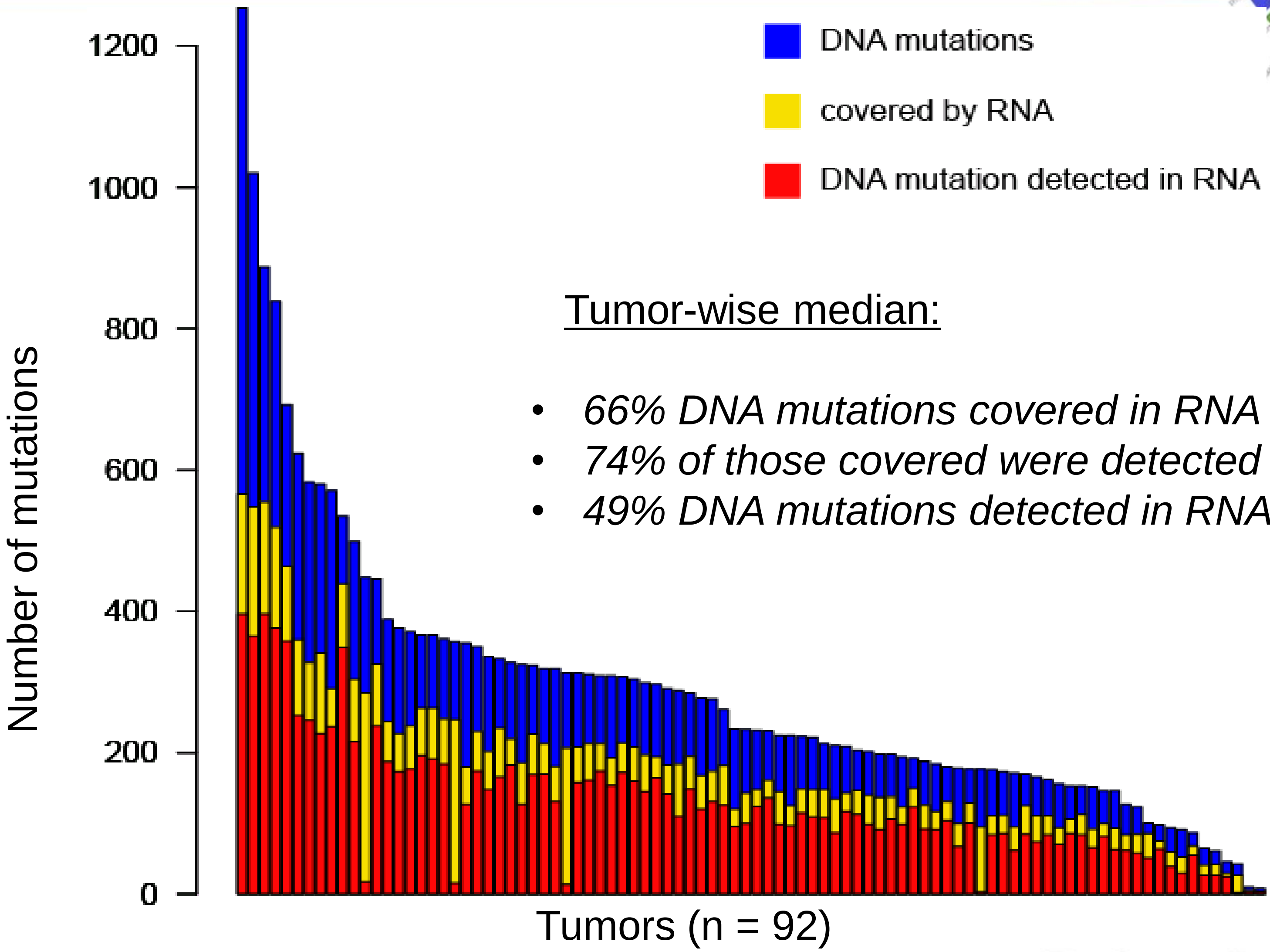




Codon	Number of Samples	Distinct mutations	Gene Name
602;646	30	4	**EZH2**
83[§]	9	2	MEF2B
69[§]	4	2	MEF2B
81[§]	2	2	MEF2B
1482[§]	3	2	CREBBP
1499[§]	2	2	CREBBP
1467[§]	2	2	EP300
287[§]	2	1	HLA-C
1	8	5	**BCL7A**[‡]
206[§]	4	1	**MYD88**[‡]
230[§]	2	1	**MYD88**[‡]
252[§]	6	1	**MYD88**[‡]
59	7	3	**BCL2**[*]
92;196;197	5	4	**CD79B**[‡]
73;160[§]	4	2	IKZF3
164;255[§]	3	2	**PIM1**
97;188	3	2	**PIM1**
18[§]	3	2	**IRF4**
587[§]	3	2	**BCL6**
45[§]	3	2	BTG2
141;234	3	2	**TP53**

Morin et al., Nature 2011

RNA Seq for mutation verification in lung cancer

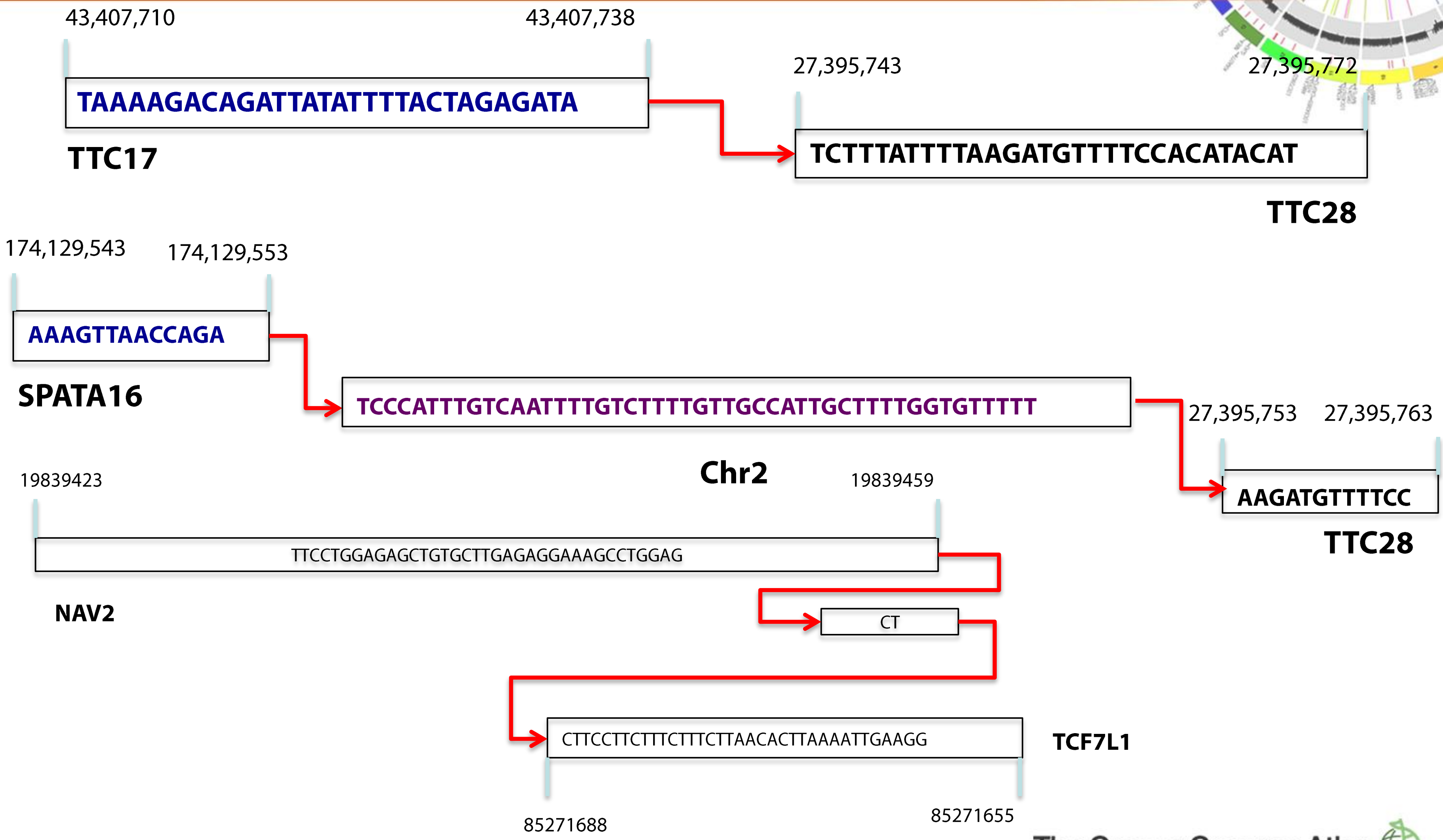


Tumor-wise median:

- *66% DNA mutations covered in RNA*
- *74% of those covered were detected in RNA*
- *49% DNA mutations detected in RNA*

Matt Wilkerson

RNA Seq confirms fusions detected using low pass sequencing of CRCs



43,407,710

43,407,738

TAAAAGACAGATTATATTTTACTAGAGATA

TTC17

27,395,743

27,395,772

TCTTTATTTTAAGATGTTTTCCACATACAT

TTC28

174,129,543

174,129,553

AAAGTTAACCAGA

SPATA16

TCCCATTTGTCAATTTTGTCTTTTGTTGCCATTGCTTTTGGTGTTTTT

Chr2

27,395,753

27,395,763

AAGATGTTTTCC

TTC28

19839423

19839459

TTCCTGGAGAGCTGTGCTTGAGAGGAAAGCCTGGAG

NAV2

CT

CTTCCTTCTTTCTTTCTTAACACTTAAAATTGAAGG

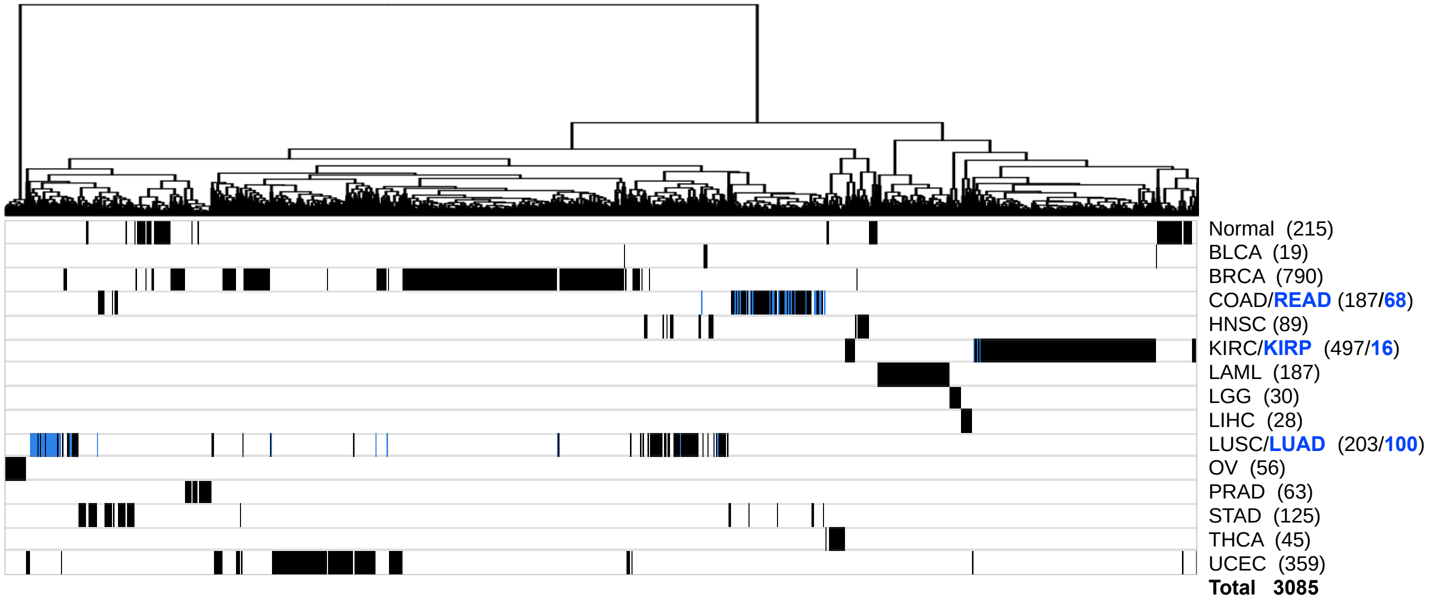
TCF7L1

85271688

85271655

3,085 miRNA-seq profiles at DCC

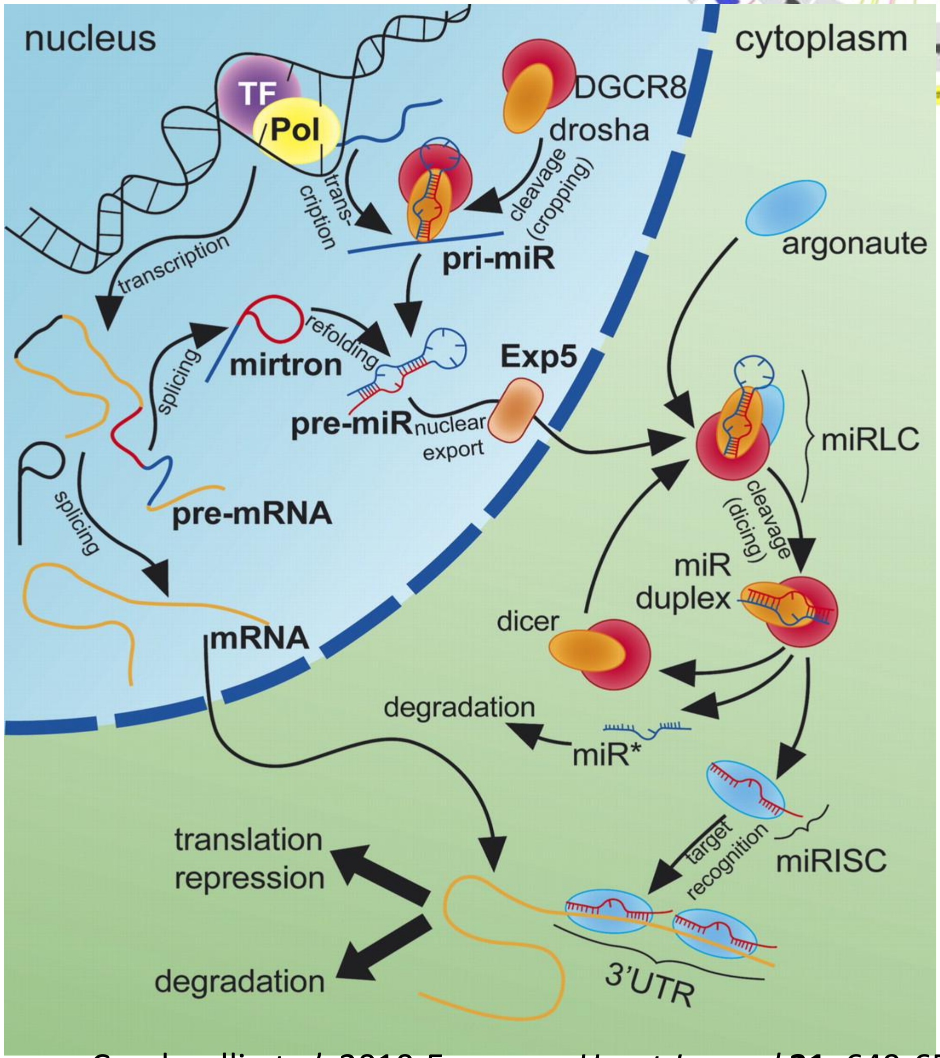
Cases sequenced	**3,536**
Bases sequenced (raw)	1,140,211,885,680
Bases sequenced (pf)	871,388,396,000
Cancer types sequenced	19
Cases submitted to DCC	3,085
Cancer types submitted to DCC	18



Andy Chu

miRNA biogenesis

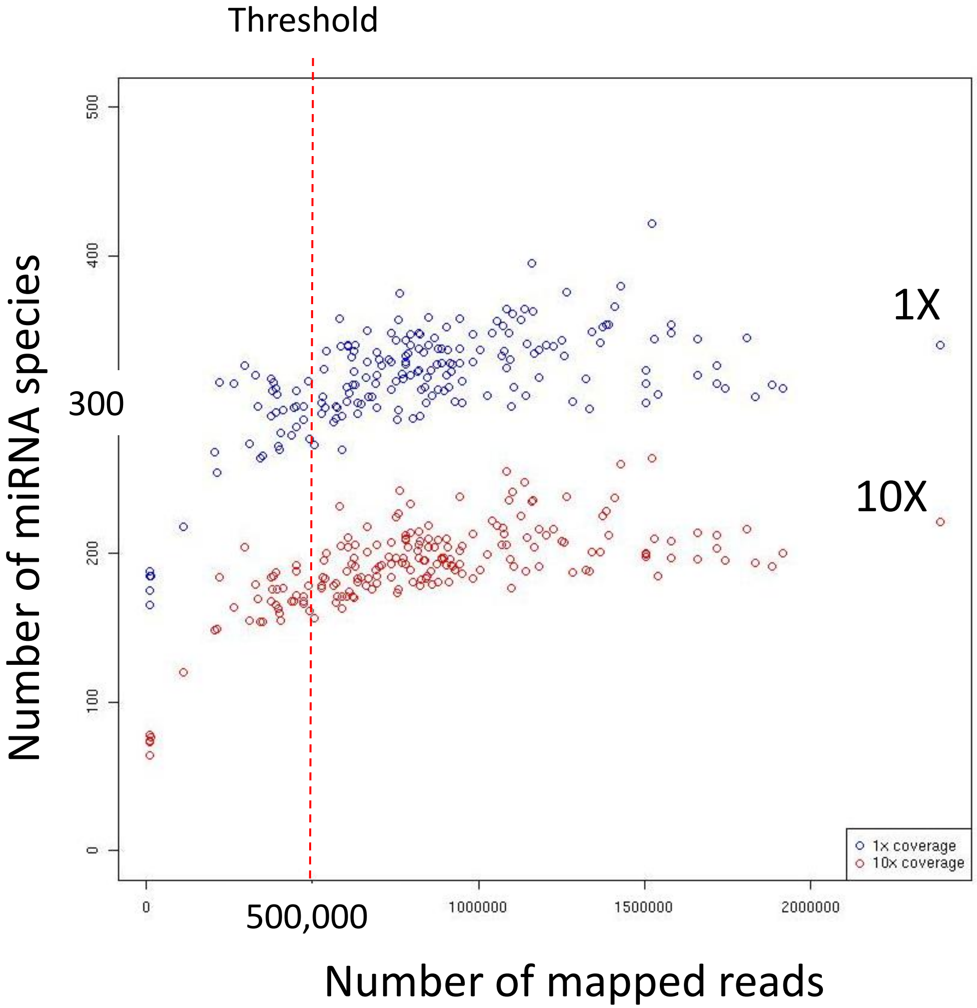
- Products of miRNA biogenesis include mature miRNA and miRNA*.
- Non-canonical miRNA variants ("isomiRs") may further expand target gene repertoire.



Condorelli *et al*. 2010 *European Heart Journal* **31**, 649-658

The Cancer Genome Atlas

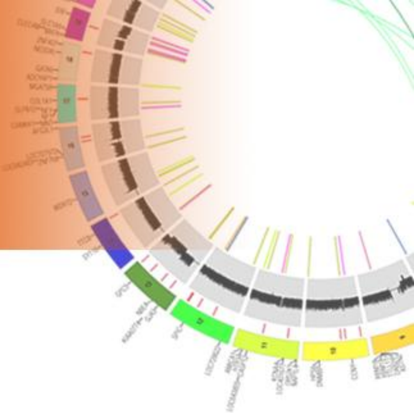
miRNA Seq sampling depth (AML)

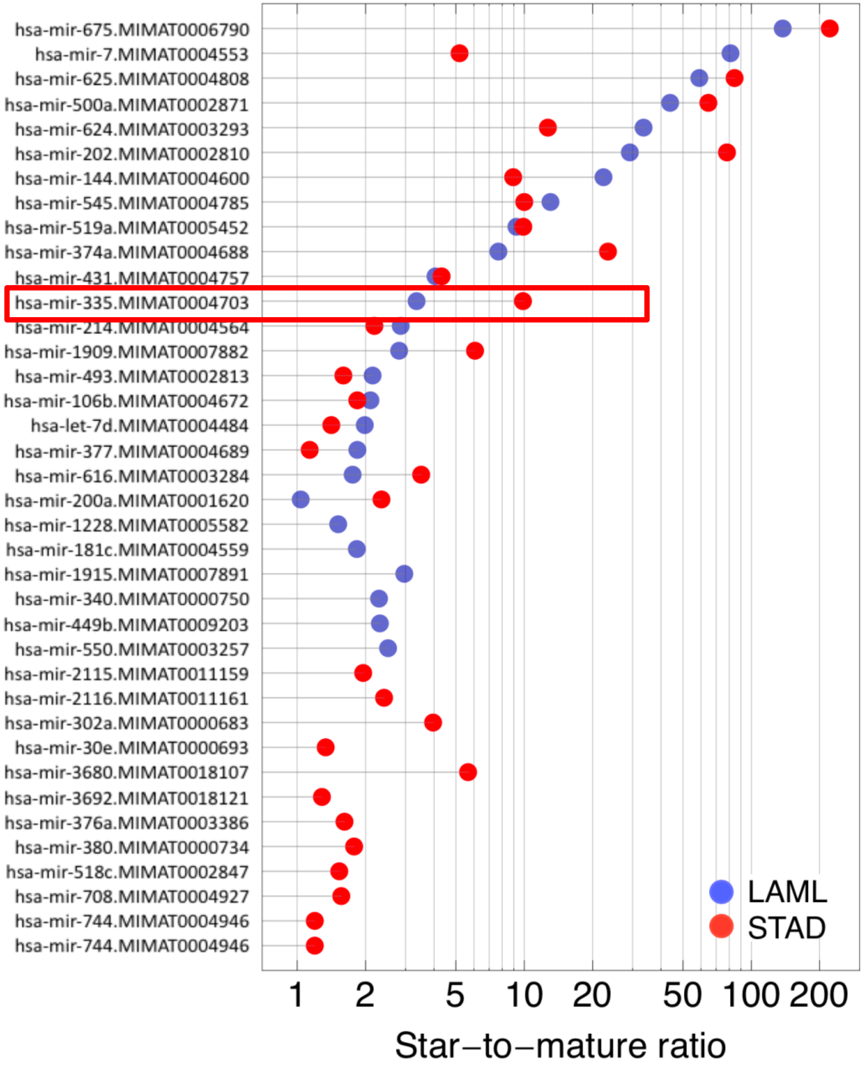


- 191 libraries sequenced.
- Mapped reads avg 0.98M.
- Known miRNAs detected: 270 to 422 (avg 328).
- 16 novel miRNAs detected (*miRBase 13).

Andy Chu

Star vs mature strand expression





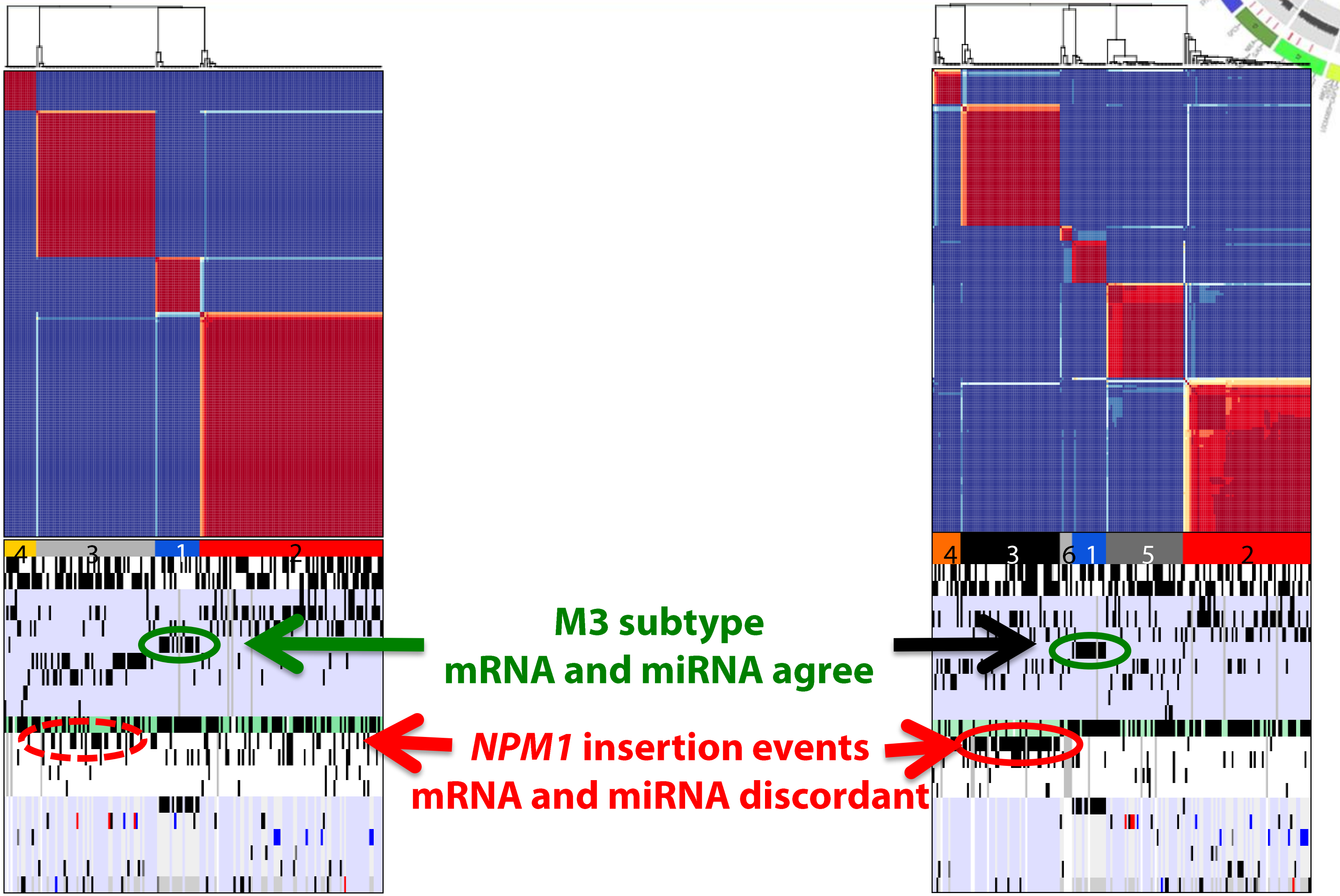
hsa-mir-374a, **star** strand and **mature** Strand TCGA-AB-3008-03A-01T-0736-13

```
TACACAGACAATTACAATACAATCTGATAAGTGCTATAACACTTATCAGGTTGTATTATAATGG
----------ATTACAATACAATCTGATAAG---------------------------------
-----------TTACAATACAATCTGATAAG---------------------------------
------------TACAATACAATCTGATAAG---------------------------------
---------------------------------------CACTTATCAGGTTGTATTATAA---
----------------------------------------ACTTATCAGGTTGTATTATAA---
-----------------------------------------CTTATCAGGTTGTATTATAA---
------------------------------------------TTATCAGGTTGTATTATAA---
```

Andy Chu

The Cancer Genome Atlas

Clustering cancer subtypes



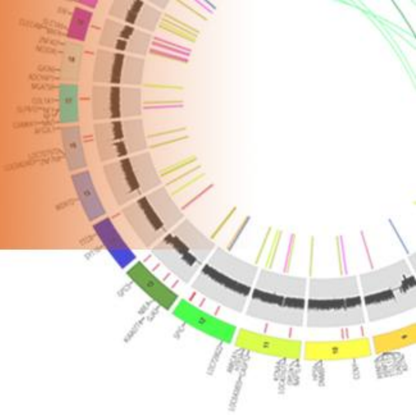
M3 subtype
mRNA and miRNA agree

***NPM1* insertion events**
mRNA and miRNA discordant

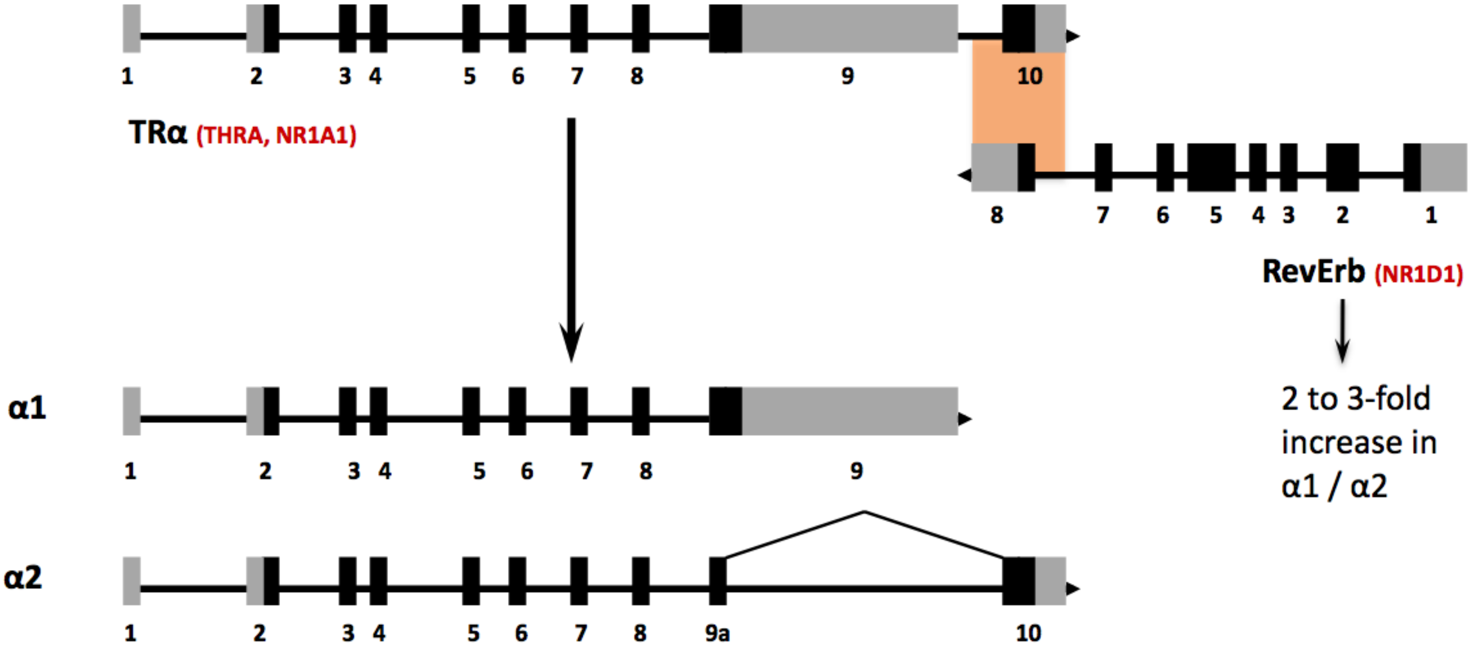
See Andy Chu, Gordon Robertson Poster

The Cancer Genome Atlas

Making sense of antisense



Antisense transcription regulates TRα alternative splicing



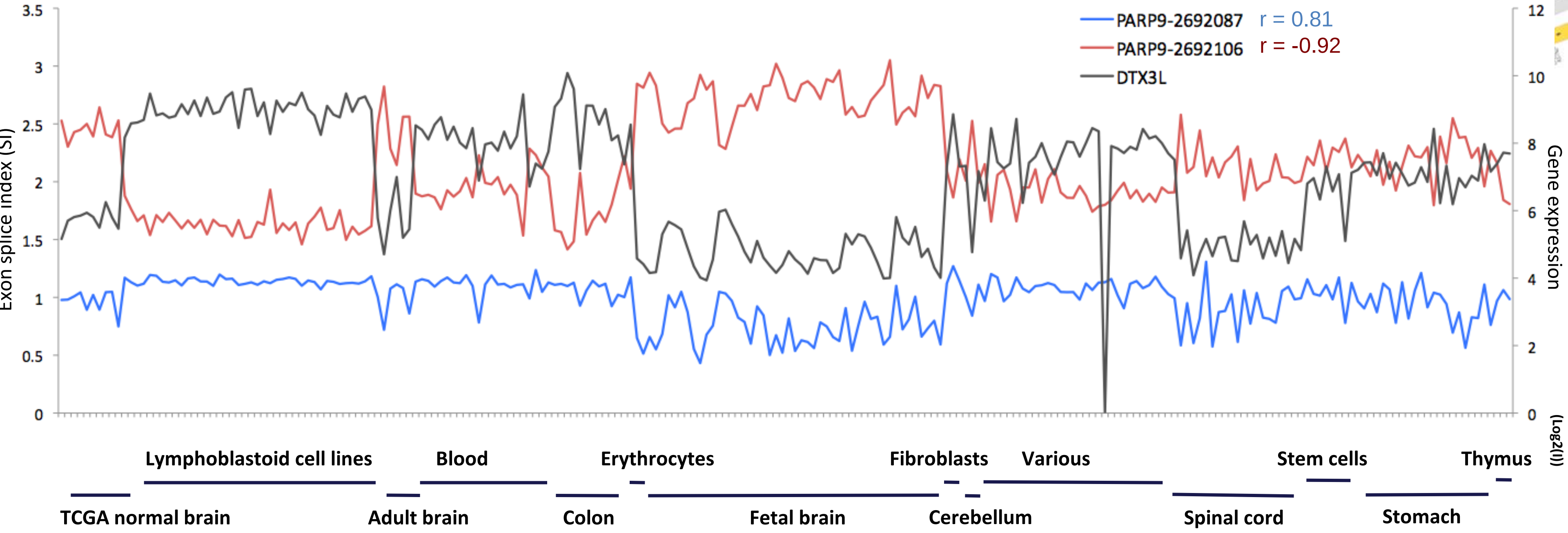
2 to 3-fold increase in α1 / α2

- *Also associated with epigenetic silencing*

Sorana Morrissy

The Cancer Genome Atlas

Antisense - correlated splicing



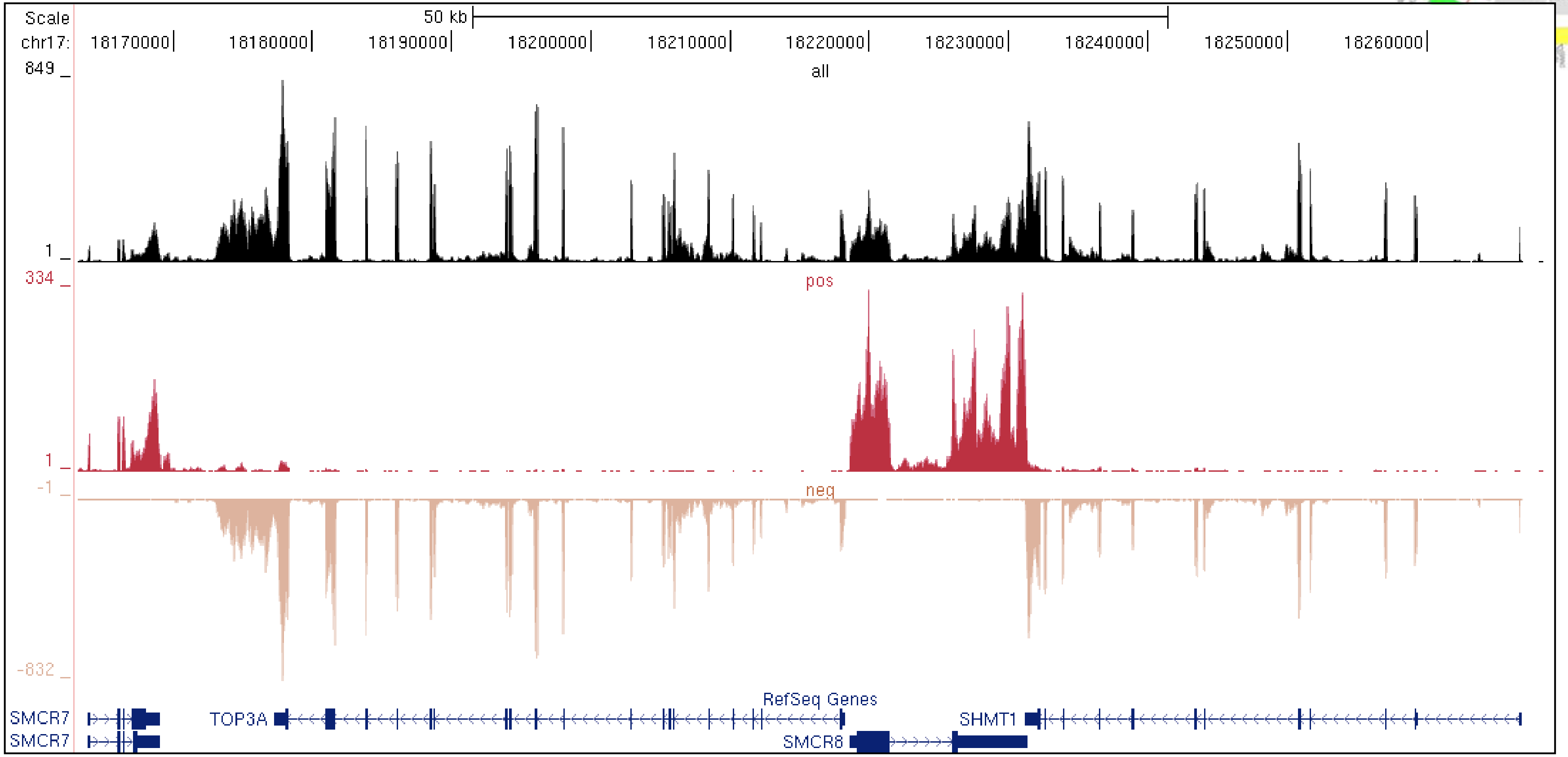
Category	1,014 Arrays	Expressed SAS genes	Expressed SAS probesets	Genes with SAS-correlated splicing	Probesets with SAS-correlated splicing
GBM*	266	4,594	83,646	2,179	9,410
OVC*	518	4,739	90,287	3,099	14,610
Normals**	230	4,801	107,179	3,312	17,420

* TCGA, Nature, 2008

** GEO, Barrett et al., NAR, 2009

Sorana Morrissy

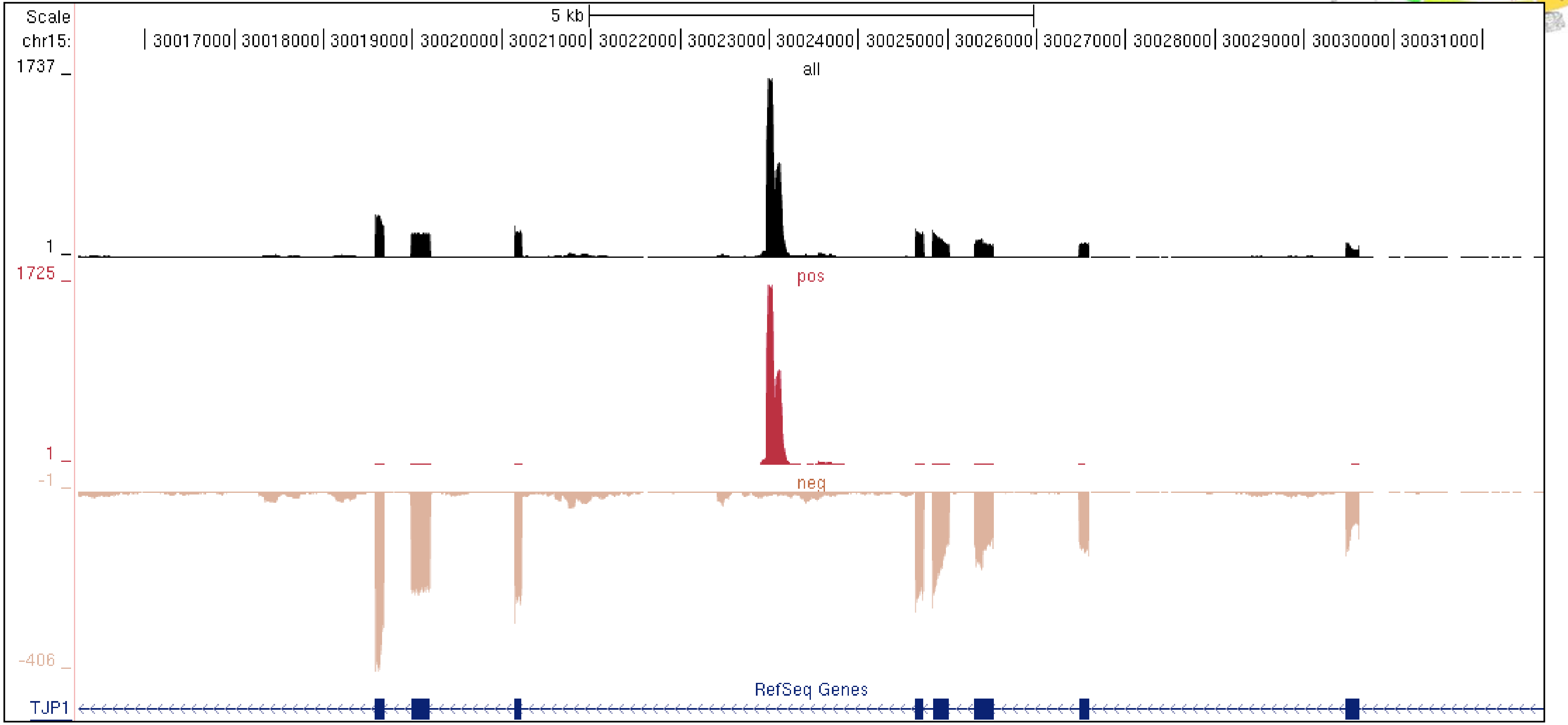
Strand specific RNA Seq



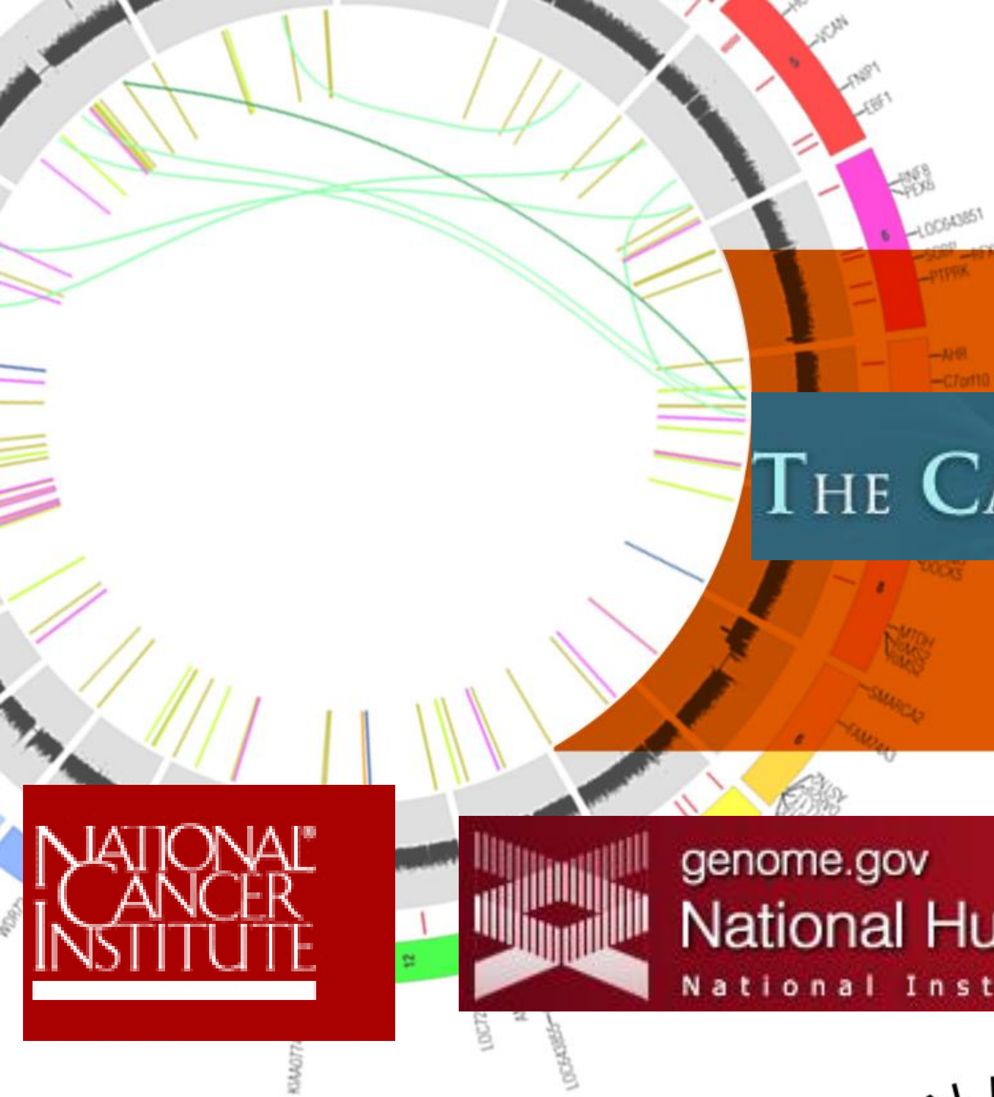
Parkhomchuk et al., *Nucleic Acids Research* 2009
Levin et al., *Nature Methods* 2010

Sorana Morrissy

Strand specific RNA Seq



Sorana Morrissy



The Cancer Genome Atlas

THE CANCER GENOME ATLAS

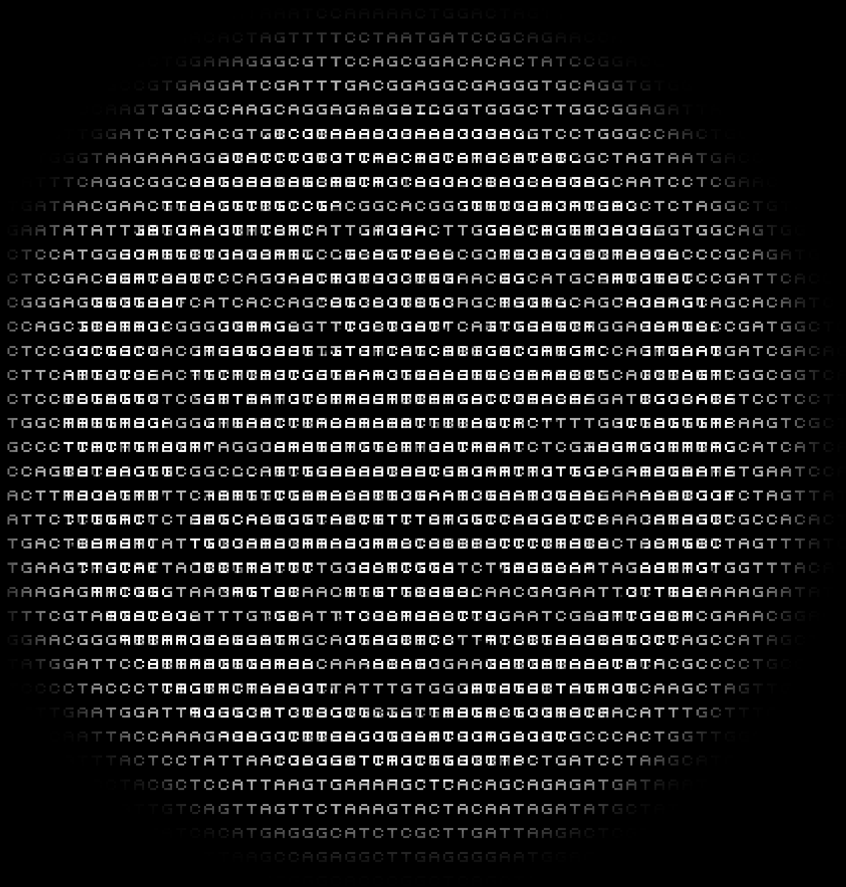
NATIONAL CANCER INSTITUTE

genome.gov
National Human Genome Research Institute
National Institutes of Health

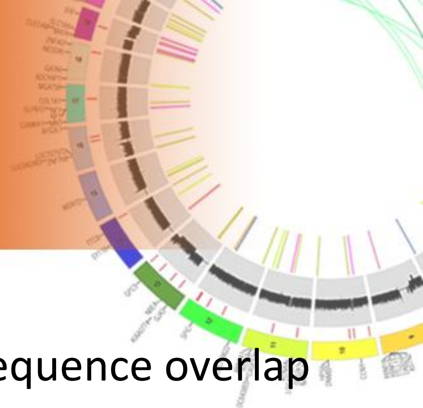
BC Cancer Foundation
Supporting research & care at BC Cancer Agency

NATIONAL INSTITUTES OF HEALTH

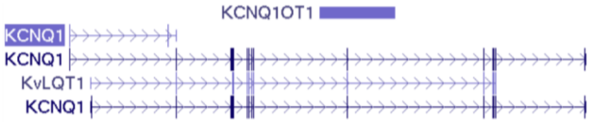
BC Cancer Agency
CARE + RESEARCH
An agency of the Provincial Health Services Authority



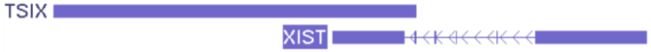
Sense-Antisense Expression



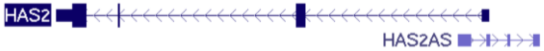
- Sense-antisense (SAS) genes: encoded on opposite strands; share sequence overlap
  - transcription rate, RNA editing, epigenetic state, alternative transcript processing



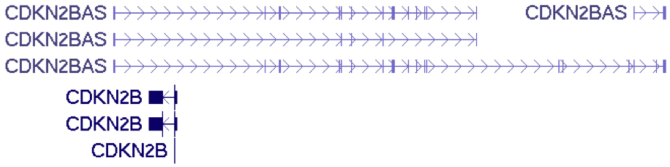
bidirectional spread of epigenetic silencing
neighbouring imprinted genes



escape from X-chromosome inactivation via
Xist promoter silencing (H3K9me3, DNA meth)



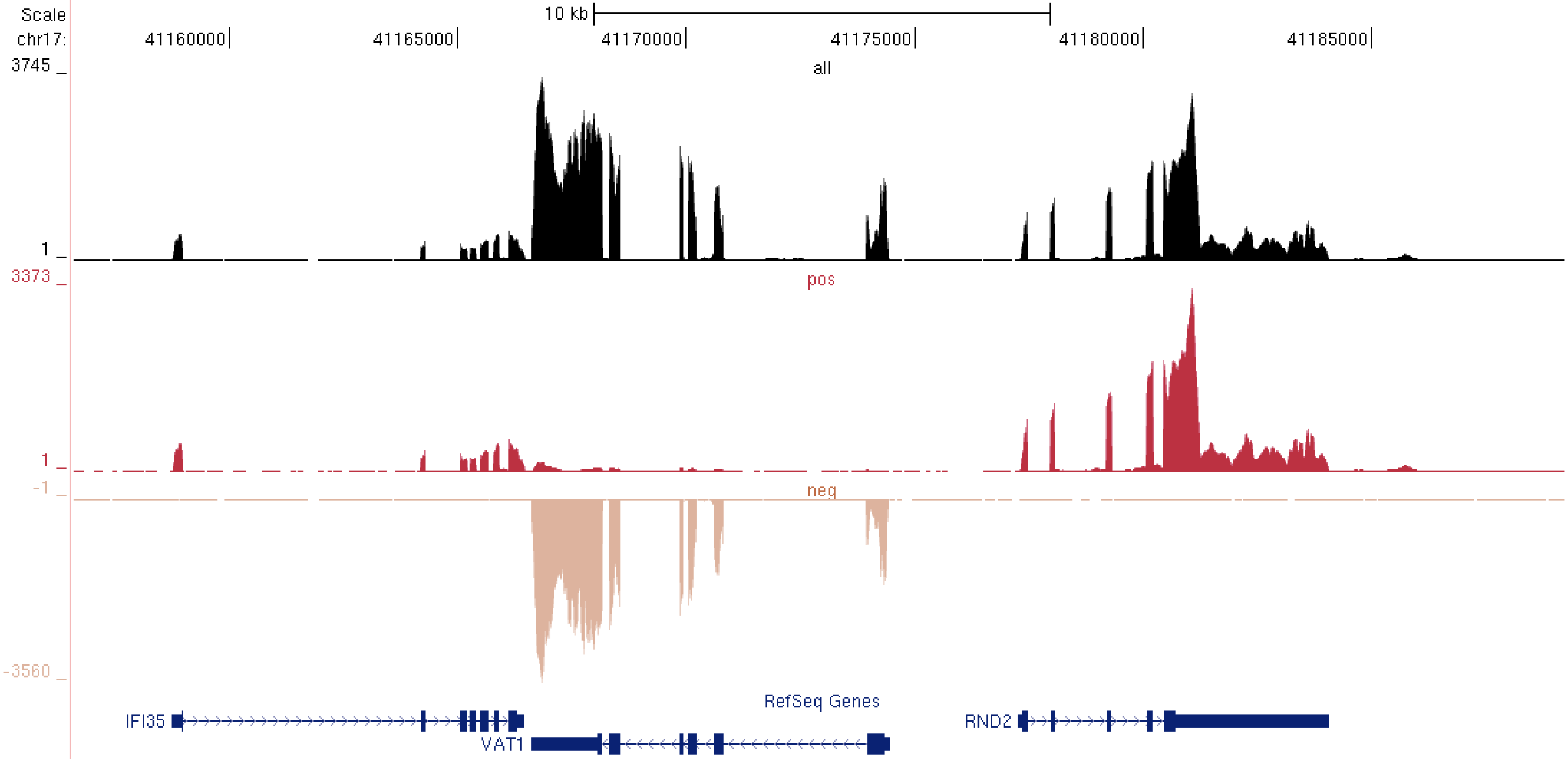
HAS2A down-regulates HAS2 expression
affects: cell proliferation, cell adhesion, migration, differentiation, metastatic spread



epigenetic silencing of CDKN2A (tumor suppressor) via
heterochromatin formation in promoter
(H3K9me2 increased, H3K4me2 decreased)

- Antisense transcription observed at >75% of genes (RIKEN, Science, 2005)

ssRNA Seq

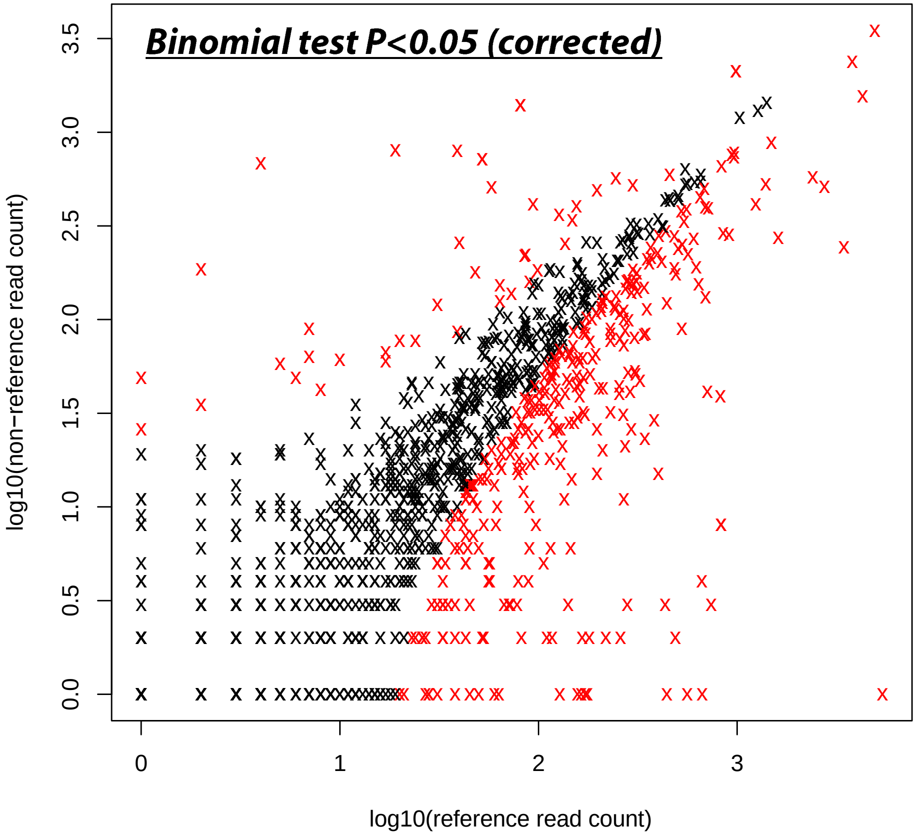


Sorana Morrissy

The Cancer Genome Atlas

Skewed representation of alleles in DLBCL RNA Seq data

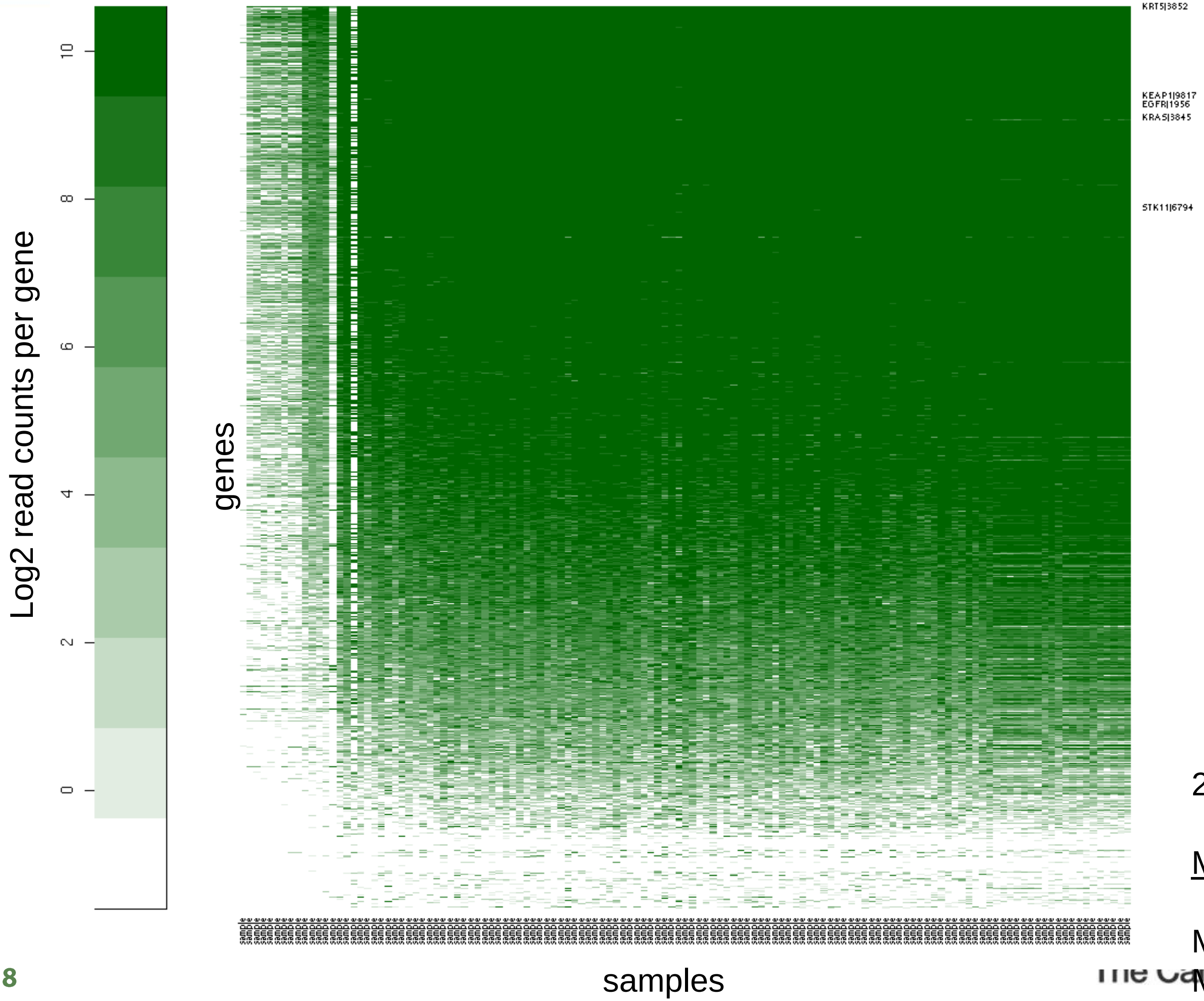
- 27% of somatic mutations exhibit significantly skewed expression (red).
- 25% are skewed in favour of the wild-type, 2% are skewed in favour of the mutant.
- ~50% of these would be undetectable by RNA-seq alone.
- 47% of truncating mutations are significantly skewed.
- Skew observed in favour of mutant allele for some known oncogenes: CD79B, CARD11, BCL2, EZH2.



Ryan Morin

The Cancer Genome Atlas

RNAseq Summary: Coverage



KRT5

KEAP1
EGFR
KRAS

STK11

20,532 genes

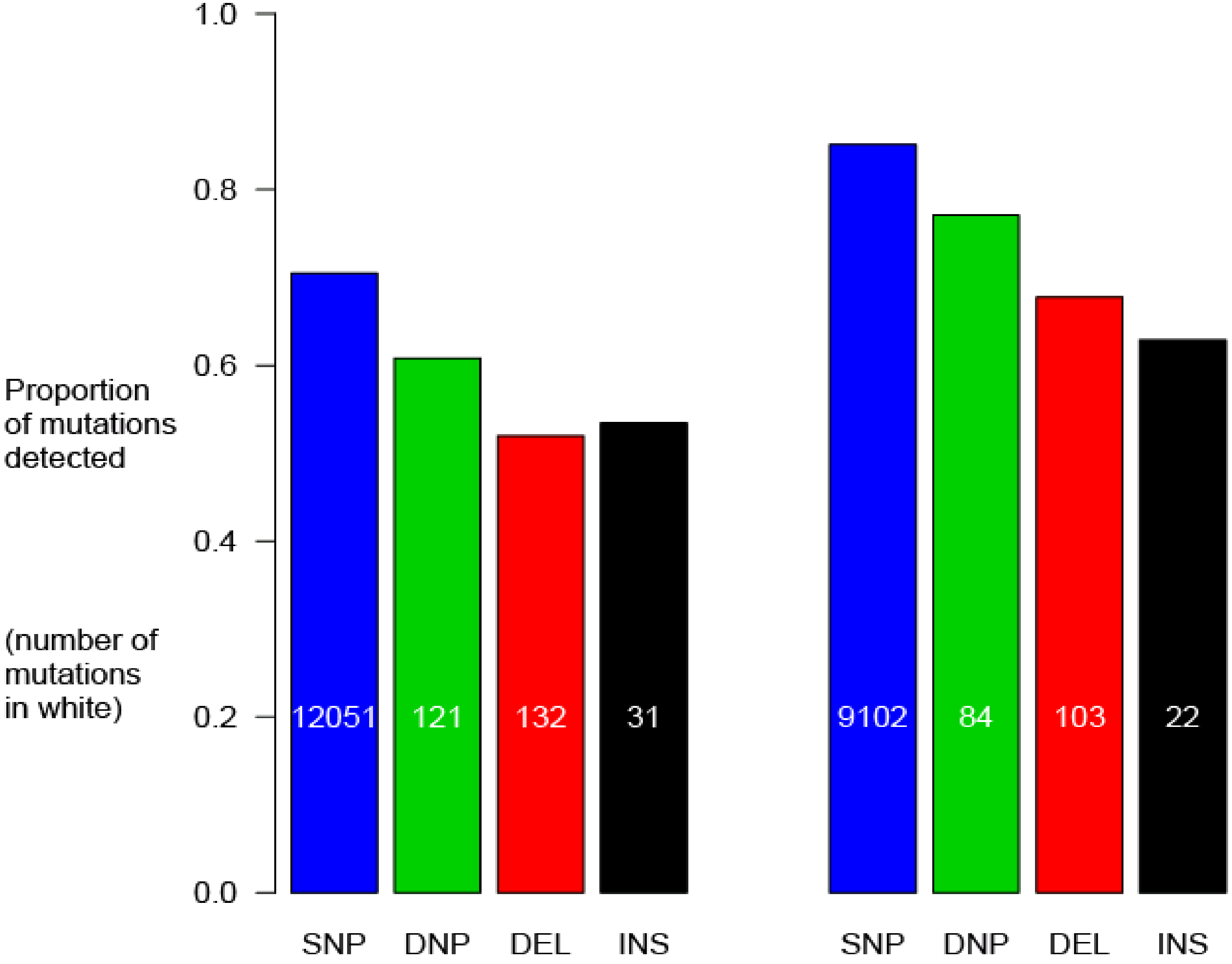
<u>Median gene coverage</u>

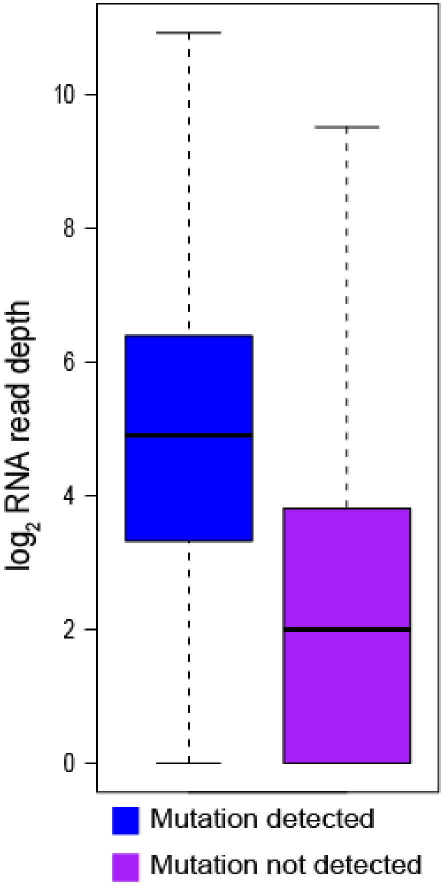
Median: 1,002
Max: 1,436,937

RNA detects major mutation types and is related to RNA read depth

Mutation sites with RNA read depth >=1

Mutation sites with RNA read depth >=10

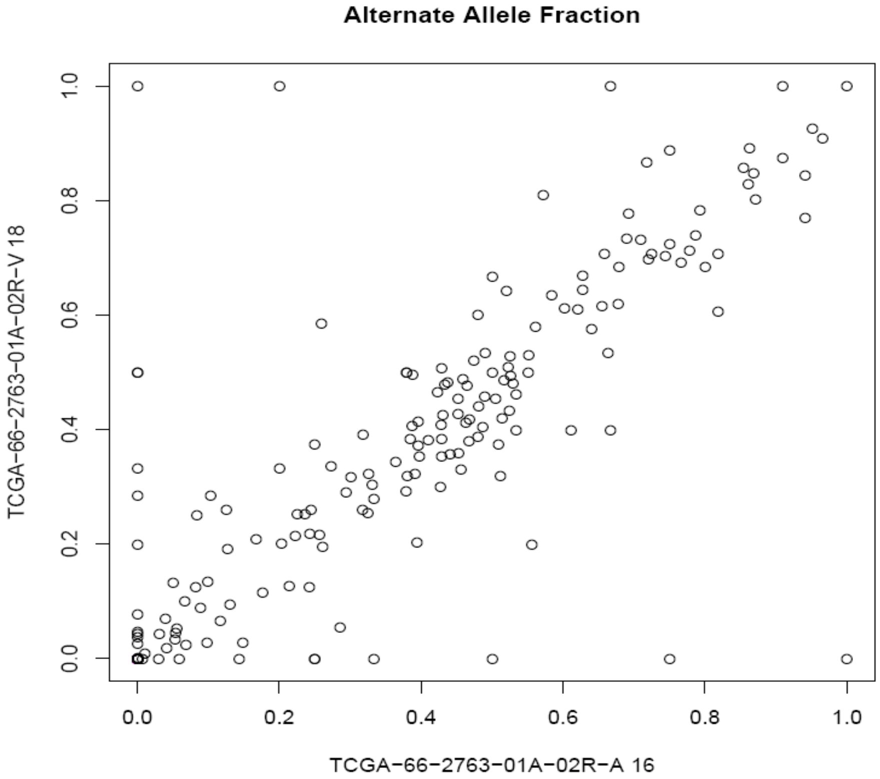




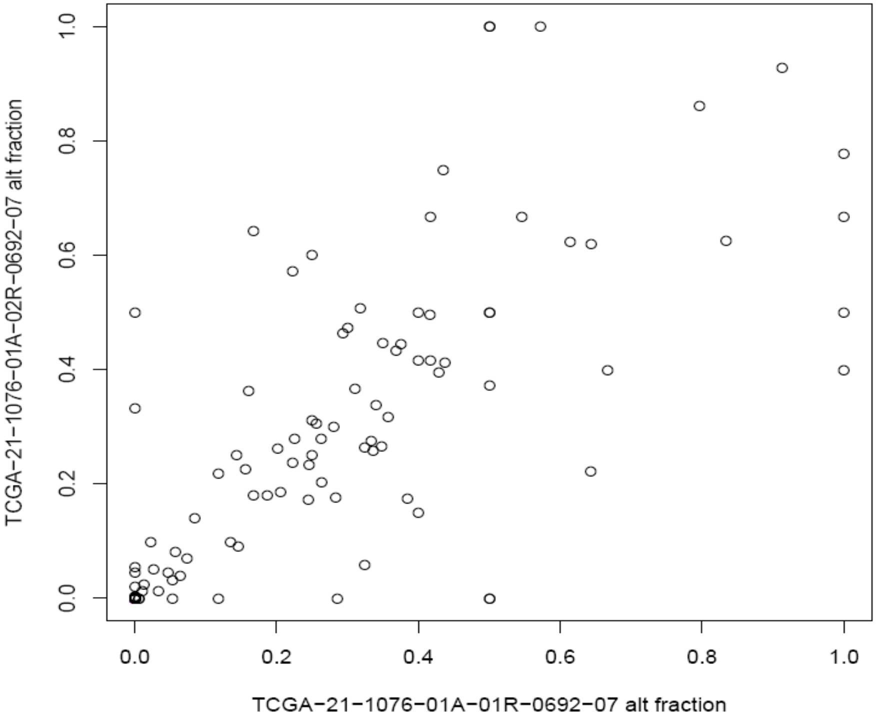
RNA Allelic Fraction for a locus : (mutant allele count / total allele count)

Is it stable among replicates?

Same tissue; two RNA isolations

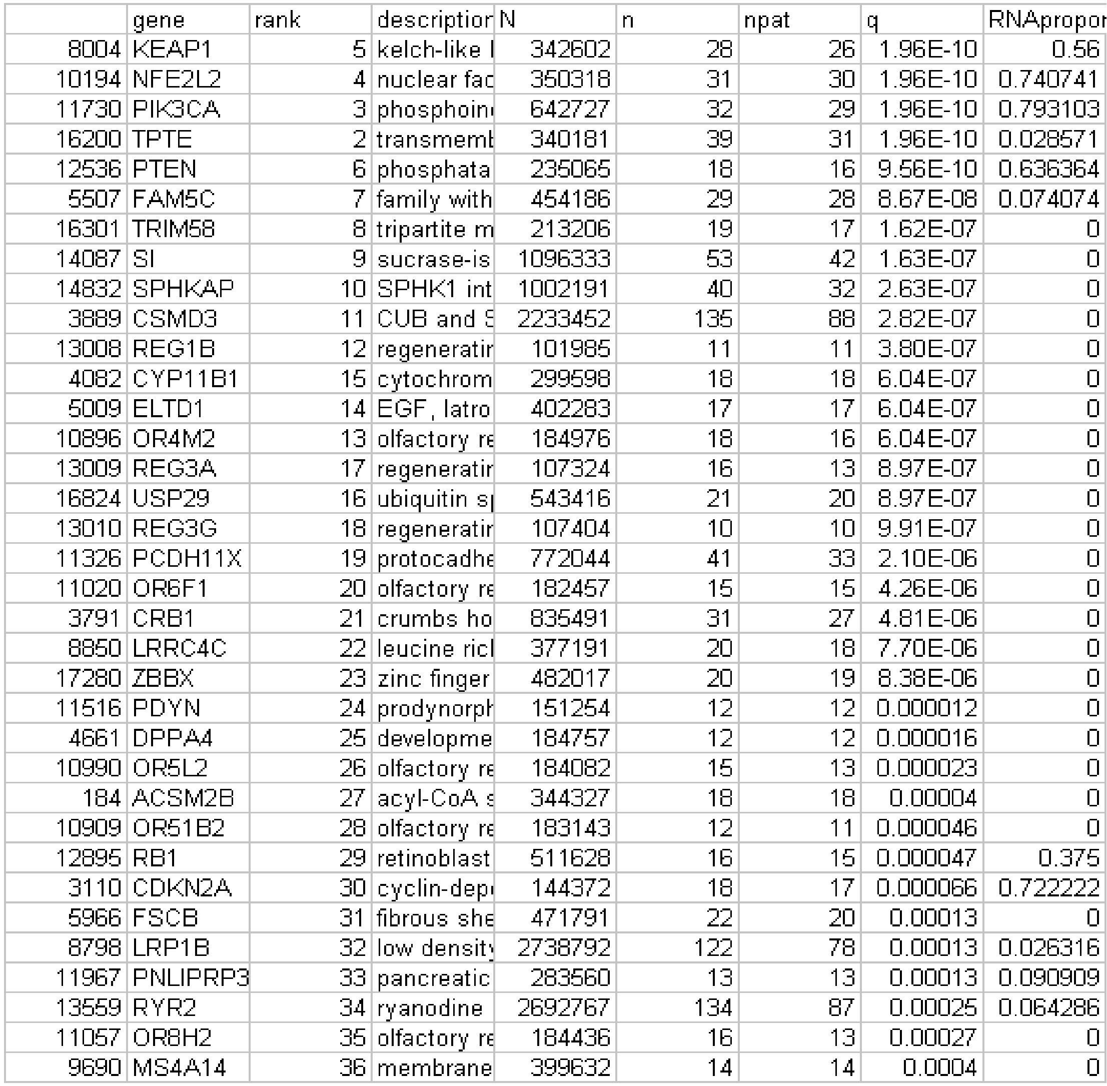


Two pieces of tissues; two RNA isolations



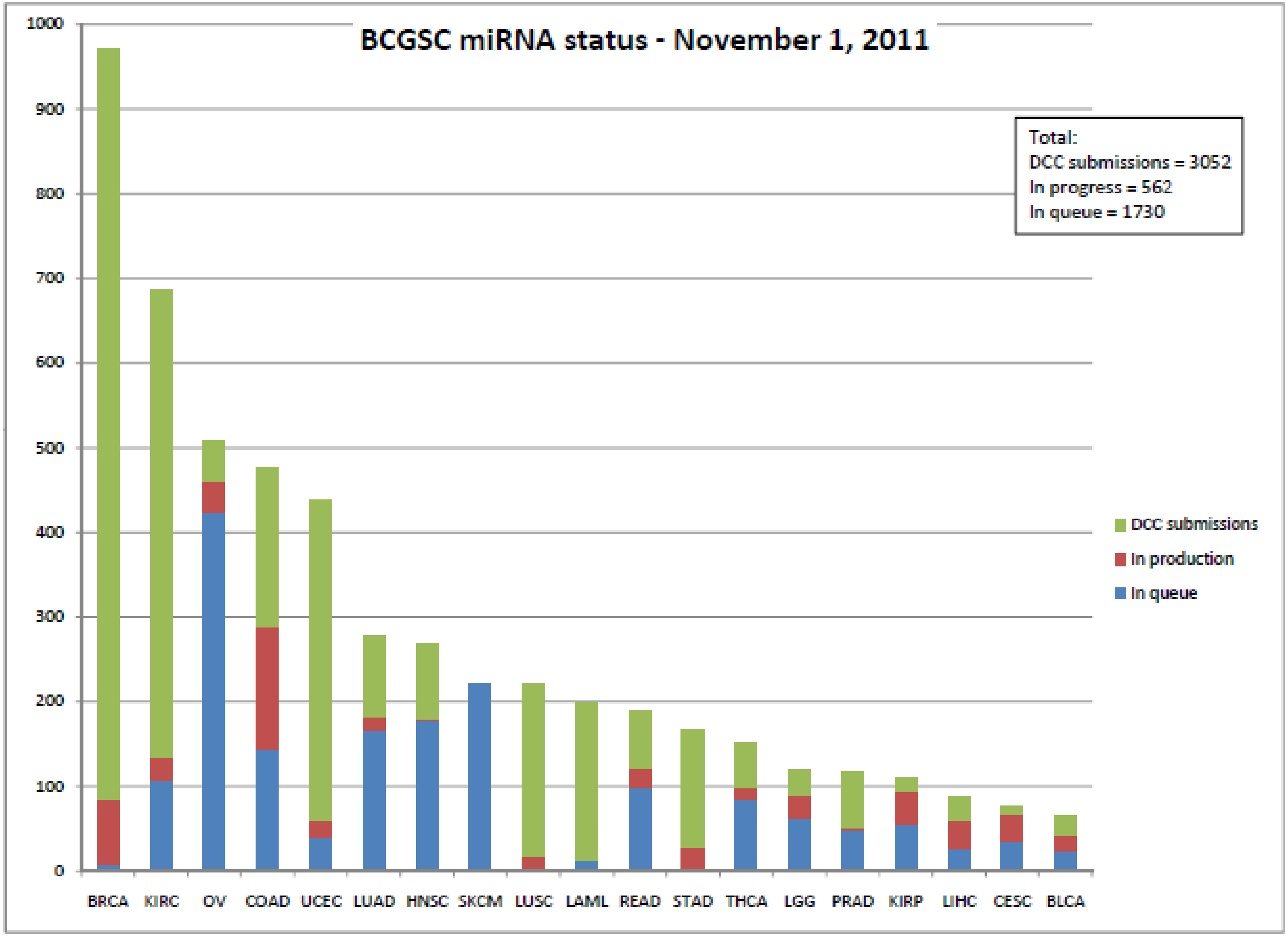
The Cancer Genome Atlas

RNA mutation detection helps determination of significantly mutated genes across LUSC



	gene	rank	description	N	n	npat	q	RNApropor
8004	KEAP1	5	kelch-like I	342602	28	26	1.96E-10	0.56
10194	NFE2L2	4	nuclear fac	350318	31	30	1.96E-10	0.740741
11730	PIK3CA	3	phosphoin	642727	32	29	1.96E-10	0.793103
16200	TPTE	2	transmemb	340181	39	31	1.96E-10	0.028571
12536	PTEN	6	phosphata	235065	18	16	9.56E-10	0.636364
5507	FAM5C	7	family with	454186	29	28	8.67E-08	0.074074
16301	TRIM58	8	tripartite m	213206	19	17	1.62E-07	0
14087	SI	9	sucrase-is	1096333	53	42	1.63E-07	0
14832	SPHKAP	10	SPHK1 int	1002191	40	32	2.63E-07	0
3889	CSMD3	11	CUB and S	2233452	135	88	2.82E-07	0
13008	REG1B	12	regeneratir	101985	11	11	3.80E-07	0
4082	CYP11B1	15	cytochrom	299598	18	18	6.04E-07	0
5009	ELTD1	14	EGF, latro	402283	17	17	6.04E-07	0
10896	OR4M2	13	olfactory re	184976	18	16	6.04E-07	0
13009	REG3A	17	regeneratir	107324	16	13	8.97E-07	0
16824	USP29	16	ubiquitin s	543416	21	20	8.97E-07	0
13010	REG3G	18	regeneratir	107404	10	10	9.91E-07	0
11326	PCDH11X	19	protocadhe	772044	41	33	2.10E-06	0
11020	OR6F1	20	olfactory re	182457	15	15	4.26E-06	0
3791	CRB1	21	crumbs ho	835491	31	27	4.81E-06	0
8850	LRRC4C	22	leucine ricl	377191	20	18	7.70E-06	0
17280	ZBBX	23	zinc finger	482017	20	19	8.38E-06	0
11516	PDYN	24	prodynorph	151254	12	12	0.000012	0
4661	DPPA4	25	developme	184757	12	12	0.000016	0
10990	OR5L2	26	olfactory re	184082	15	13	0.000023	0
184	ACSM2B	27	acyl-CoA s	344327	18	18	0.00004	0
10909	OR51B2	28	olfactory re	183143	12	11	0.000046	0
12895	RB1	29	retinoblast	511628	16	15	0.000047	0.375
3110	CDKN2A	30	cyclin-dep	144372	18	17	0.000066	0.722222
5966	FSCB	31	fibrous she	471791	22	20	0.00013	0
8798	LRP1B	32	low density	2738792	122	78	0.00013	0.026316
11967	PNLIPRP3	33	pancreatic	283560	13	13	0.00013	0.090909
13559	RYR2	34	ryanodine	2692767	134	87	0.00025	0.064286
11057	OR8H2	35	olfactory re	184436	16	13	0.00027	0
9690	MS4A14	36	membrane	399632	14	14	0.0004	0

Likely passenger mutations (e.g. olfactory receptors) removed



BCGSC miRNA status - November 1, 2011

Total:
DCC submissions = 3052
In progress = 562
In queue = 1730

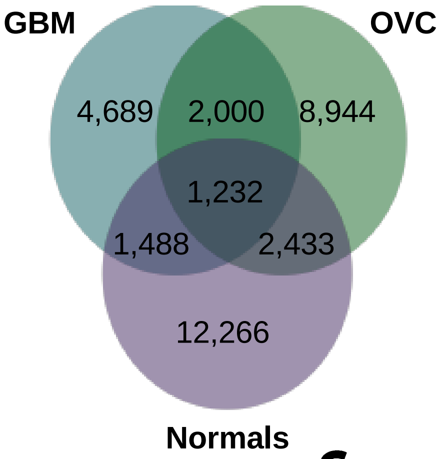
Antisense-correlated splicing events in brain and ovarian cancers

Category	28 Tissues	1,014 Arrays	Expressed SAS genes	Expressed SAS probesets	Genes with SAS-correlated splicing	Probesets with SAS-correlated splicing
GBM*	1	266	4,594	83,646	2,179	9,410
OVC*	1	518	4,739	90,287	3,099	14,610
Normals**	26	230	4,801	107,179	3,312	17,420

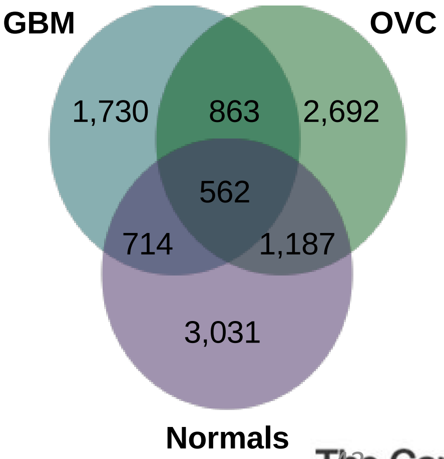
* TCGA, Nature, 2008

** GEO, Barrett et al., NAR, 2009

Probesets with antisense-correlated splicing



Genes with antisense-correlated splicing events

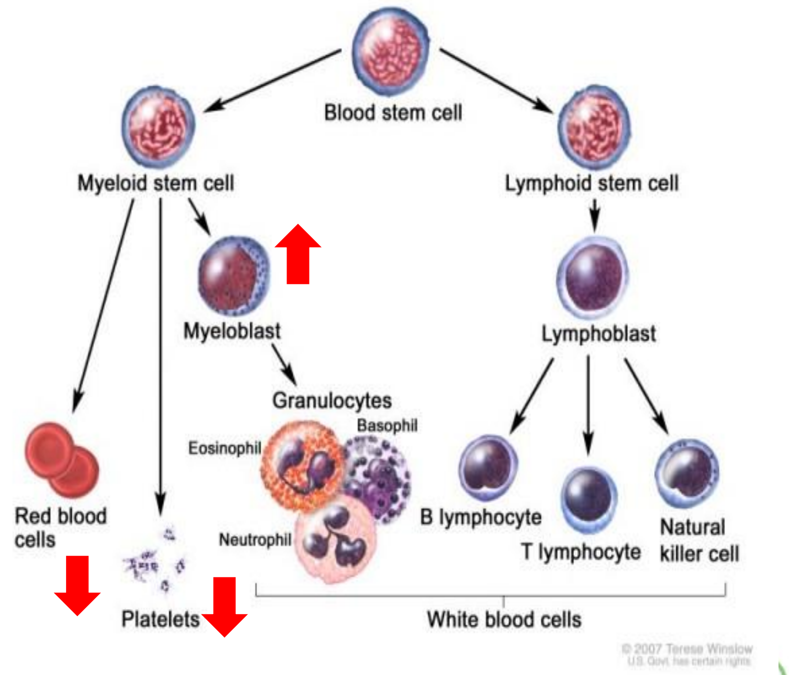


Sorana Morrissy

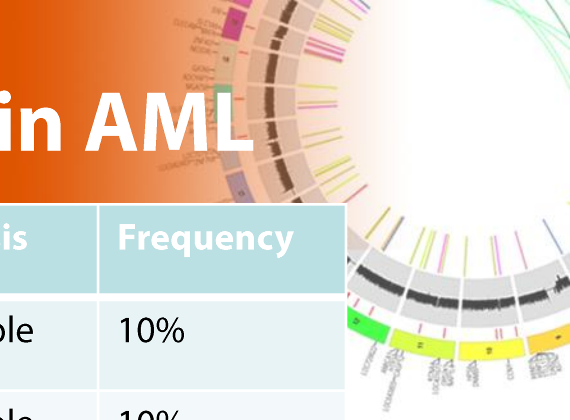
The Cancer Genome Atlas

Acute Myeloid Leukemia

- Selected for study by The Cancer Genome Atlas (TCGA)
- Haematopoietic stem cell disorder
- Most common acute adult leukemia
- World Health Organization identifies 4 subtypes
- Characterized by abnormal myeloblasts that do not mature into healthy WBC
- Abnormal cells build up in bone marrow, decreasing available space for healthy blood cells
- Possible causes: smoking, previous chemotherapy, radiation exposure



Known molecular abnormalities in AML



Rearrangement(s)	Fusion protein	FAB	Prognosis	Frequency
t(15;17)	PML-RARα	M3	Favourable	10%
t(8;21)	RUNX1-RUNX1T1	M2	Favourable	10%
Inv(16)	CBFβ-MYH11	M4	Favourable	5%
der(11q23)	MLL-fusions	M4/M5	Variable	4%
t(9;22)	BCR-ABL1	M1/M2	Adverse	2%
Others	Multiple	Multiple	Variable	<1%

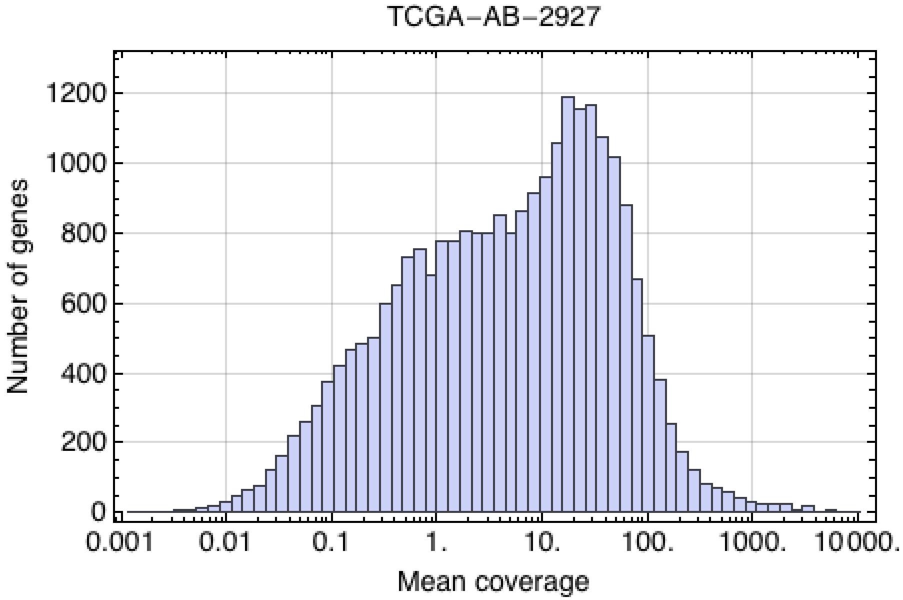
Martens and Stunnenberg (2010) *FEBS Letters* 584:2662-9

- Partial tandem duplications (PTDs) and internal tandem duplications (ITDs) are relatively common in AML:
  - ➢ MLL and FLT3
- Insertion/Deletions & point mutations have also been identified in e.g.:
  - ➢ *ASXL1, CBFB, DNMT3A, FLT3, IDH1&2, JAK2, NPM1, RAS, RUNX1, TET2, WT1*

RNA Sequencing in AML

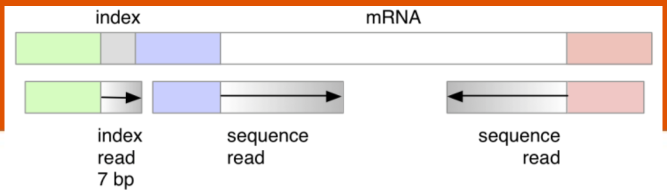
- 191 AML samples received; 179 sequenced and submitted to SRA/dbGaP/DCC
- Sequence 2 Illumina GAIIx lanes per sample with 50 bp paired reads
- Average 125 million reads, 6.26 Gb (filtered) per sample

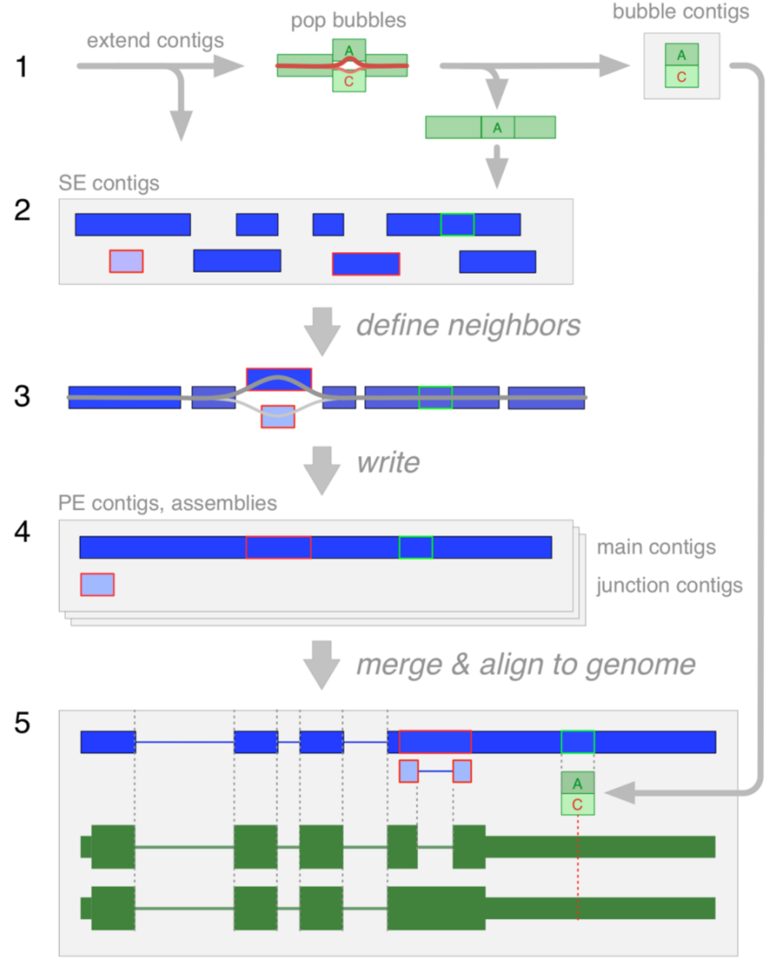
- Gene detection per sample:
  - 25,426 genes detected
  - 18,413 with ≥ 1X coverage
  - 13,254 with ≥ 5X coverage
  - 1,607 with ≥ 100X coverage

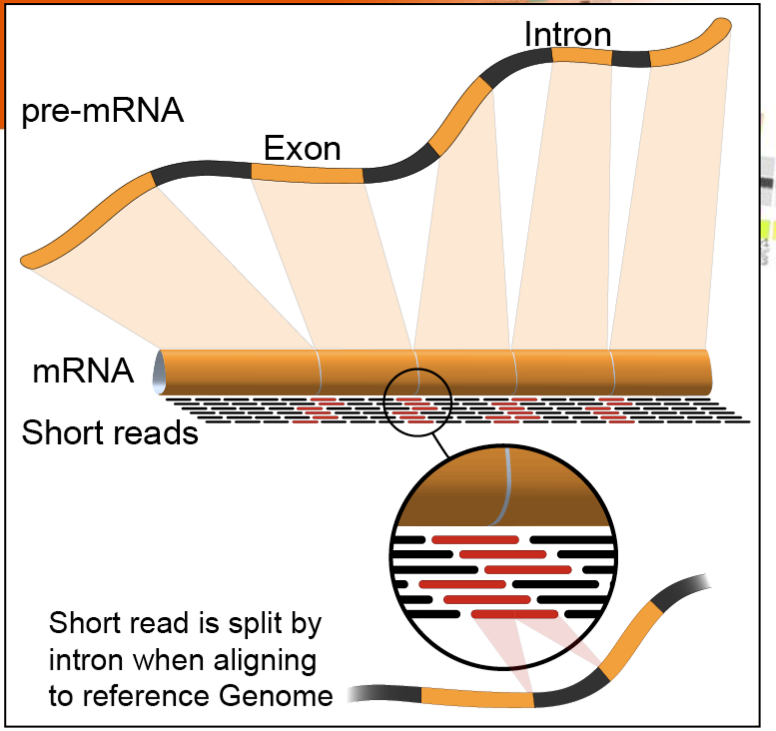


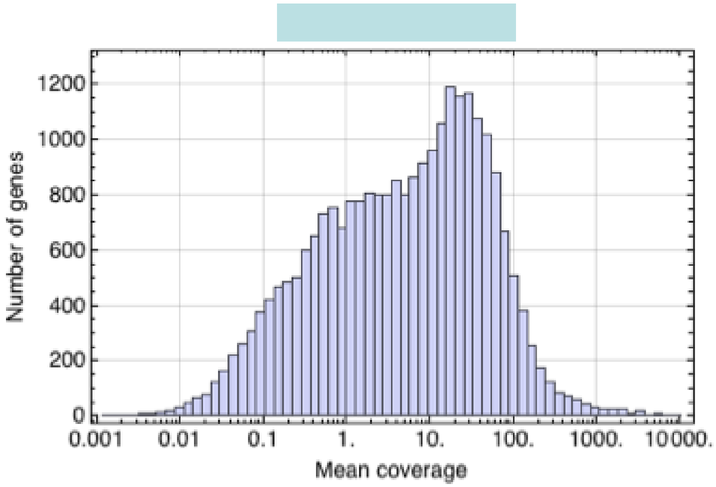
Erin Pleasance, Gordon Robertson

The Cancer Genome Atlas


index
mRNA
index
read
7 bp
sequence
read
sequence
read

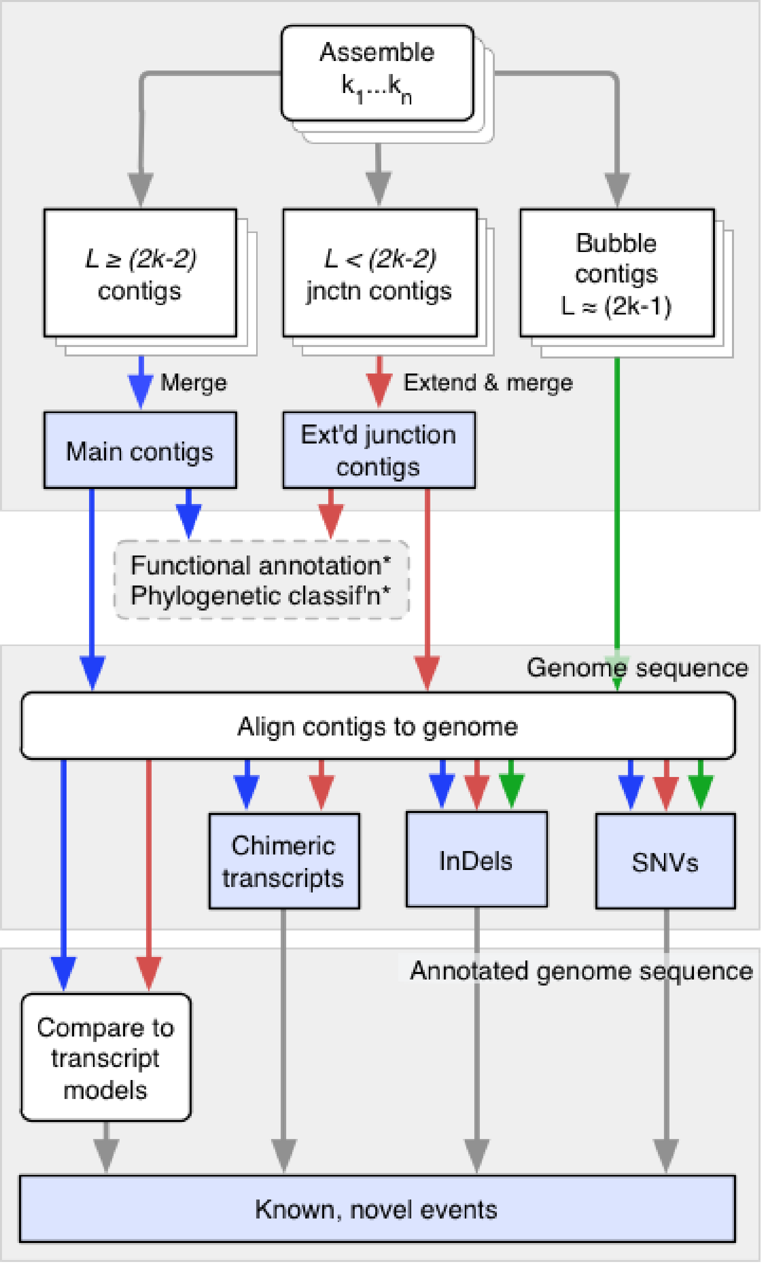

1
extend contigs
pop bubbles
A
C
bubble contigs
A
C
A
2
SE contigs
define neighbors
3
write
PE contigs, assemblies
4
main contigs
junction contigs
merge & align to genome
5
A
C

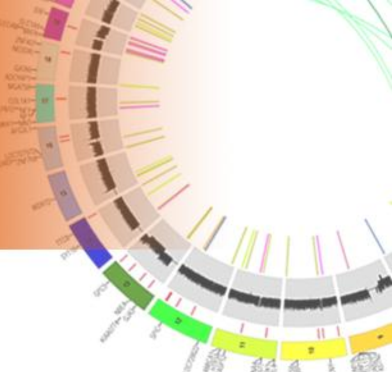

Intron
pre-mRNA
Exon
mRNA
Short reads
Short read is split by
intron when aligning
to reference Genome


Number of genes
1200
1000
800
600
400
200
0
0.001
0.01
0.1
1.
10.
100.
1000.
10000.
Mean coverage

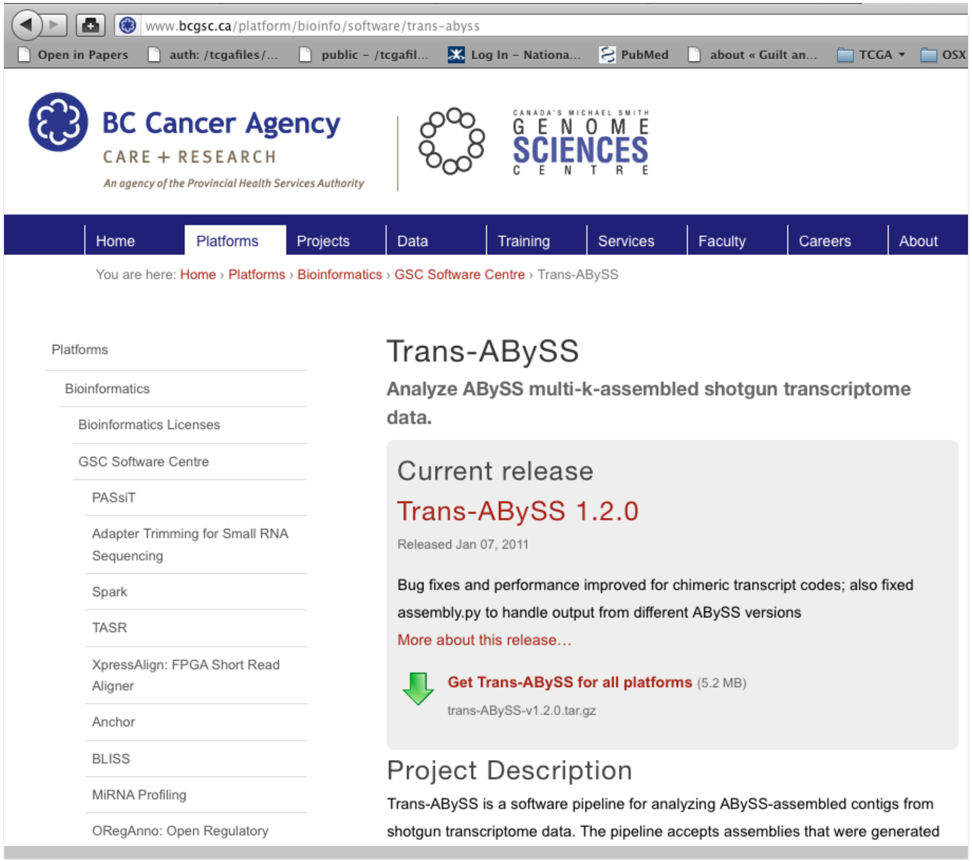
Atlas

Trans-ABySS pipeline





www.bcgsc.ca/platform/bioinfo/software/trans-abyss



You are here: Home › Platforms › Bioinformatics › GSC Software Centre › Trans-ABySS

- Platforms
- Bioinformatics
- Bioinformatics Licenses
- GSC Software Centre
- PASsiT
- Adapter Trimming for Small RNA Sequencing
- Spark
- TASR
- XpressAlign: FPGA Short Read Aligner
- Anchor
- BLISS
- MiRNA Profiling
- ORegAnno: Open Regulatory

Trans-ABySS

Analyze ABySS multi-k-assembled shotgun transcriptome data.

Current release

Trans-ABySS 1.2.0

Released Jan 07, 2011

Bug fixes and performance improved for chimeric transcript codes; also fixed assembly.py to handle output from different ABySS versions

More about this release…

Get Trans-ABySS for all platforms (5.2 MB)

trans-ABySS-v1.2.0.tar.gz

Project Description

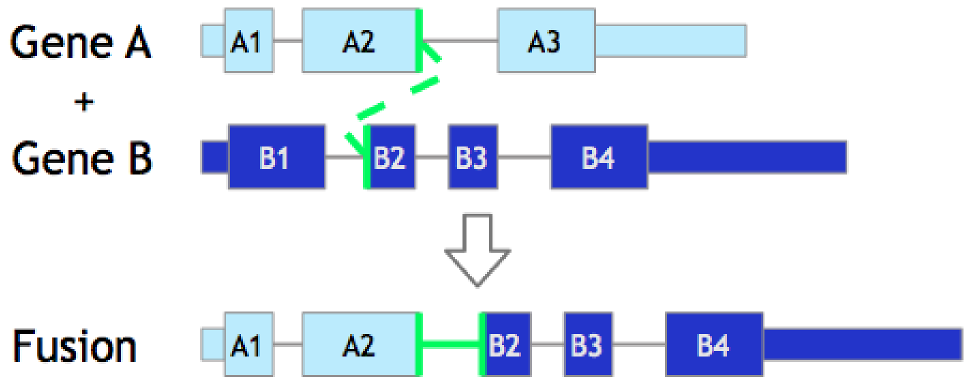
Trans-ABySS is a software pipeline for analyzing ABySS-assembled contigs from shotgun transcriptome data. The pipeline accepts assemblies that were generated

The Cancer Genome Atlas

Chimeric transcripts

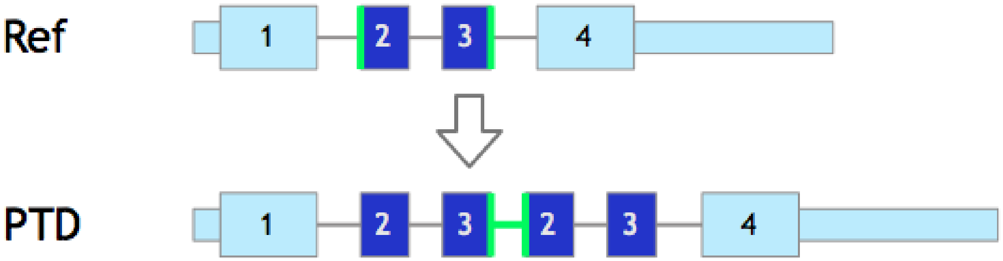
Fusions

Medves S, Demoulin J-B: **Tyrosine kinase gene fusions in cancer: translating mechanisms into targeted therapies.** *J Cell Mol Med* 2011, [Epub ahead of print]



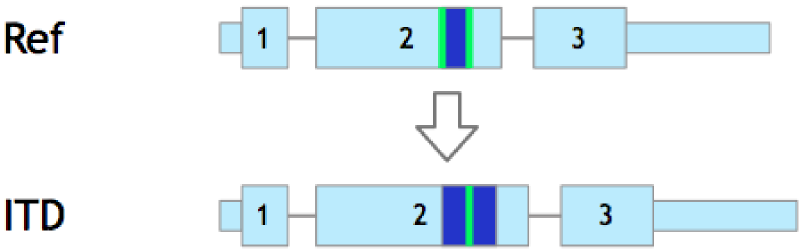
Partial tandem duplications

Liu HC, Shih LY, May Chen MJ, Wang CC, Yeh TC, Lin TH, Chen CY, Lin CJ, Liang DC. **Expression of HOXB genes is significantly different in acute myeloid leukemia with a partial tandem duplication of MLL vs. a MLL translocation: a cross-laboratory study**. Cancer Genet. 2011 204(5):252-9.

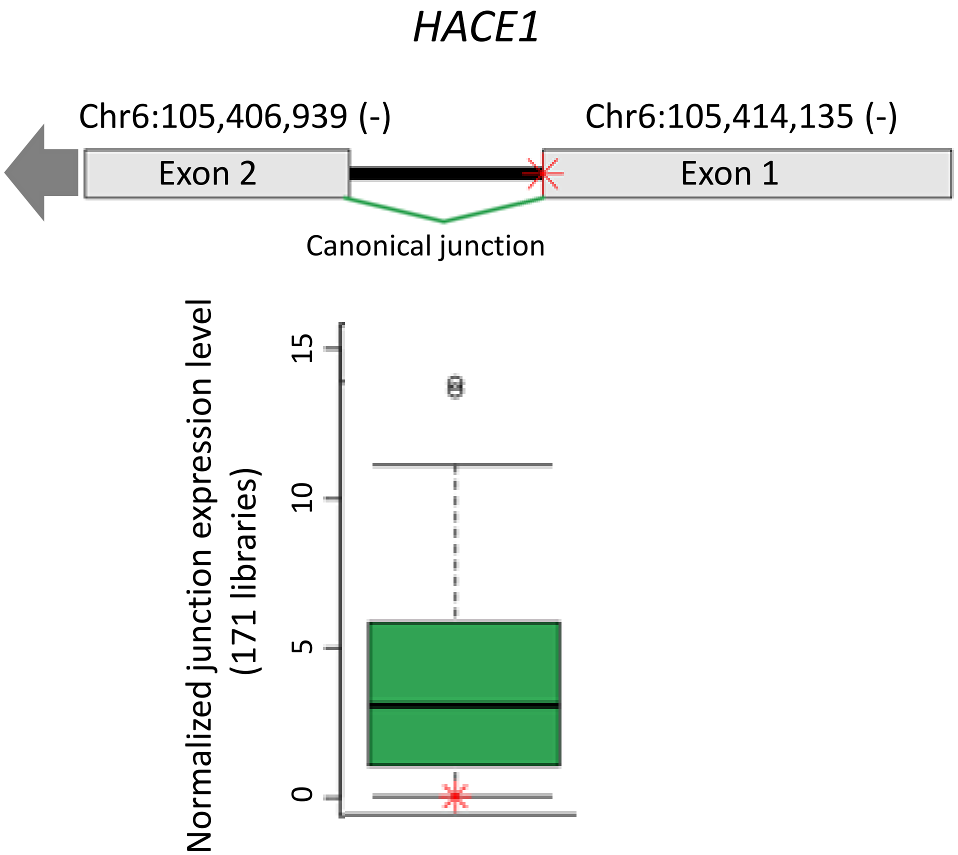


Internal tandem duplications

Fathi AT, Arowojolu O, Swinnen I, Sato T, Rajkhowa T, Small D, Marmsater F, Robinson JE, Gross SD, Martinson M, Allen S, Kallan NC, Levis M. **A potential therapeutic target for FLT3-ITD AML: PIM1 kinase**. Leuk Res. 2011 [Epub ahead of print]



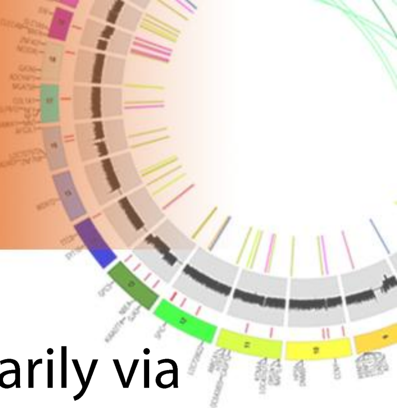
Splice donor site mutation alters *HACE1* exon expression



- Donor mutation in *HACE1** gene of sample TCGA-AB-2986 (chr6:105,414,133)
- 6q tumour suppressor gene (Thelander *et al*. 2008 *Leuk & Lymphoma)*
- *HACE1* is a putative Wilms Tumour susceptibility gene (Slade *et al*. 2010 *J Med Genet)*
- Lack of *HACE1* expression in RNA-seq data is consistent with nonsense-mediated decay

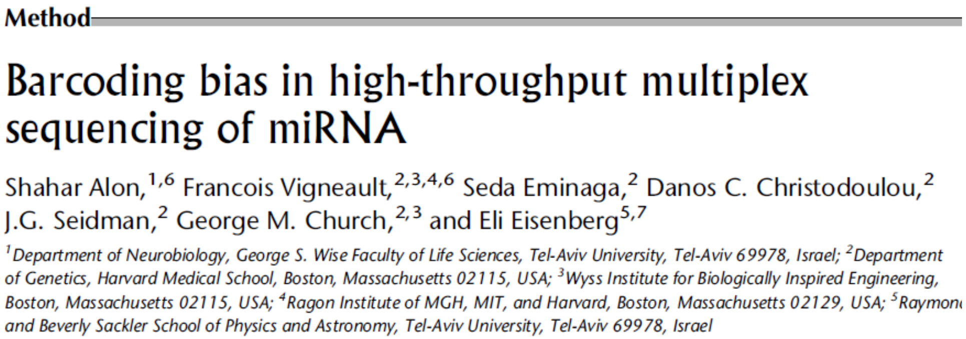
*HECT domain and ankyrin repeat containing, E3 ubiquitin protein ligase 1

A role for microRNAs in AML?



- miRNAs are key players in gene regulation, acting primarily via target mRNA degradation and/or translational repression.
- Clinically relevant biomarkers include:
  - miR-126/126* increased expression is associated with t(8;21) and inv(16) and inhibits apoptosis [Li *et al*. 2008 *PNAS* **105**:15535-40]
  - miR-29b targeting DNMT3A and associated with improved clinical response to decitabine (DNMTi) [Blum *et al*. 2010 *PNAS* **107**:7473-8]
  - miR-223- and miR-181b-like binding sites created by somatic mutation of the TNFAIP2 3'UTR leading to translational repression of this gene [Ramsingh *et al*. 2010 *Blood* **116**:5316-5326]
  - miR-17-92 cluster members are over-expressed as a direct result of promoter binding by MLL fusion proteins [Mi *et al*. 2010 *PNAS* **107**:3710-5]
- miRNA expression profiling may therefore have important roles in cancer prognosis and therapeutics

Multiplexed small RNA sequencing – the problem



Method

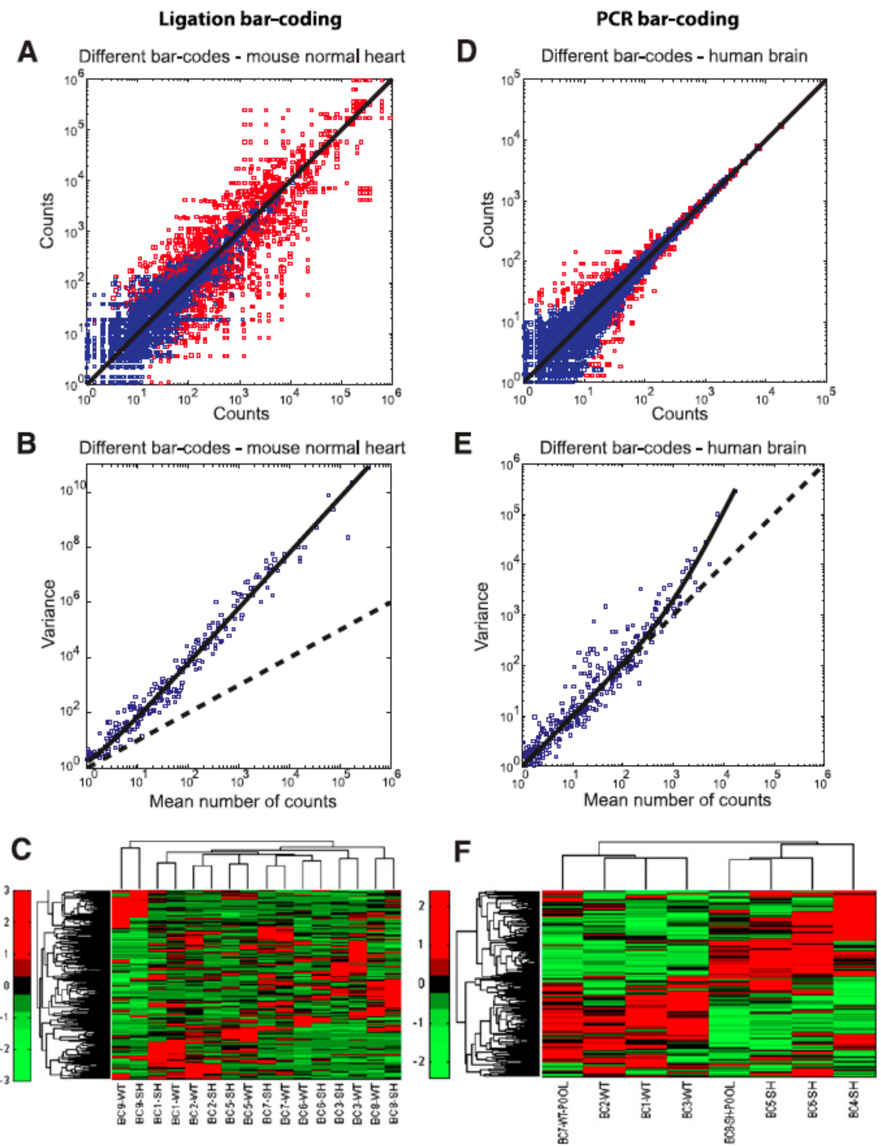
Barcoding bias in high-throughput multiplex sequencing of miRNA

Shahar Alon,[1,6] Francois Vigneault,[2,3,4,6] Seda Eminaga,[2] Danos C. Christodoulou,[2] J.G. Seidman,[2] George M. Church,[2,3] and Eli Eisenberg[5,7]

[1]Department of Neurobiology, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel; [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA; [3]Wyss Institute for Biologically Inspired Engineering, Boston, Massachusetts 02115, USA; [4]Ragon Institute of MGH, MIT, and Harvard, Boston, Massachusetts 02129, USA; [5]Raymon and Beverly Sackler School of Physics and Astronomy, Tel-Aviv University, Tel-Aviv 69978, Israel

http://www.genome.org/cgi/doi/10.1101/gr.121715.111

“Here we report that barcodes introduced through adapter ligation confer significant bias on miRNA expression profiles.”



Multiplexed small RNA sequencing – the solution

- Adding barcodes during PCR amplification minimizes the bias we and others (Alon *et al*. 2011 *Gen. Res*. Epub Aug 4 & Hafner *et al*. 2011 *RNA* Epub July 20) observe when employing bar-coding by ligation.
- Illumina GAIIx/HiSeq 2000 platforms

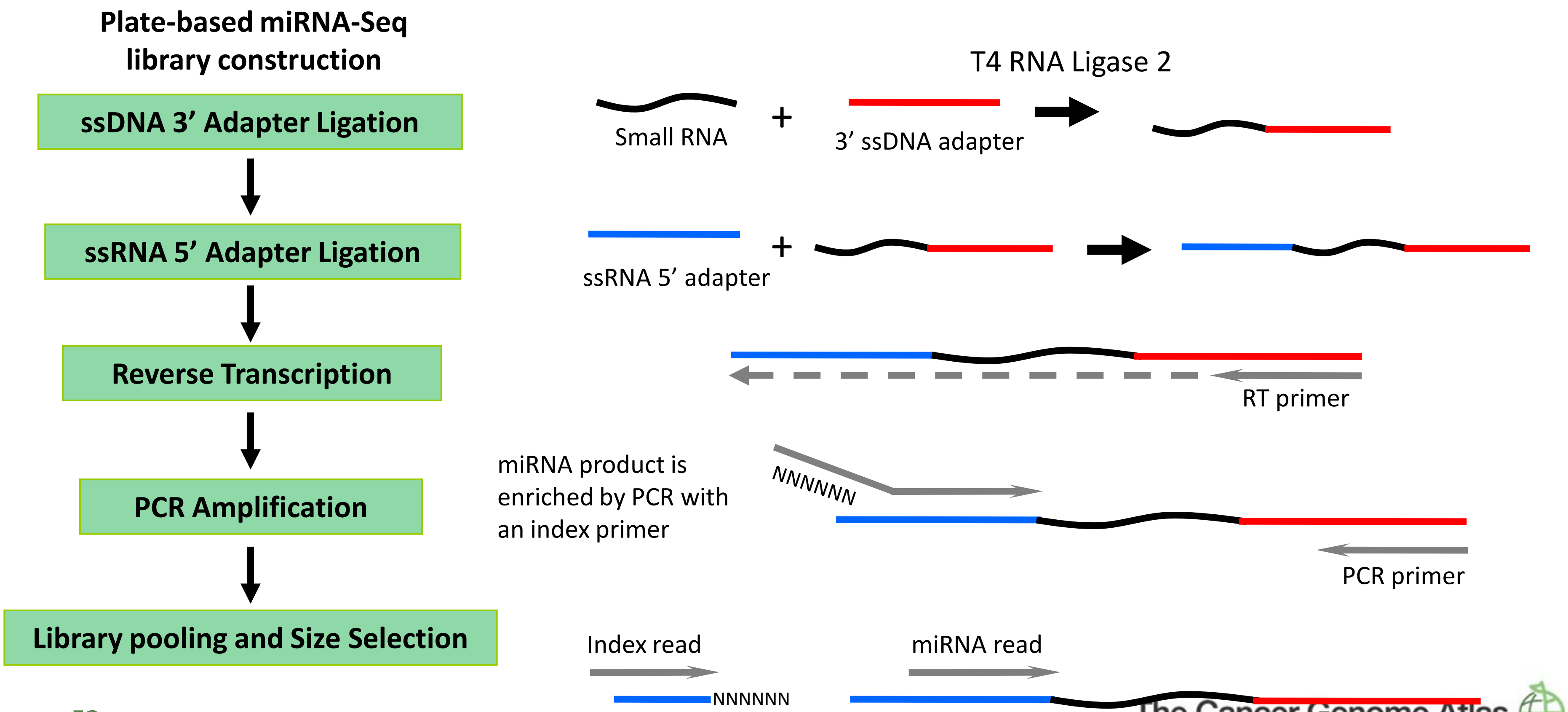


Plate-based miRNA-Seq library construction

ssDNA 3’ Adapter Ligation → **ssRNA 5’ Adapter Ligation** → **Reverse Transcription** → **PCR Amplification** → **Library pooling and Size Selection**

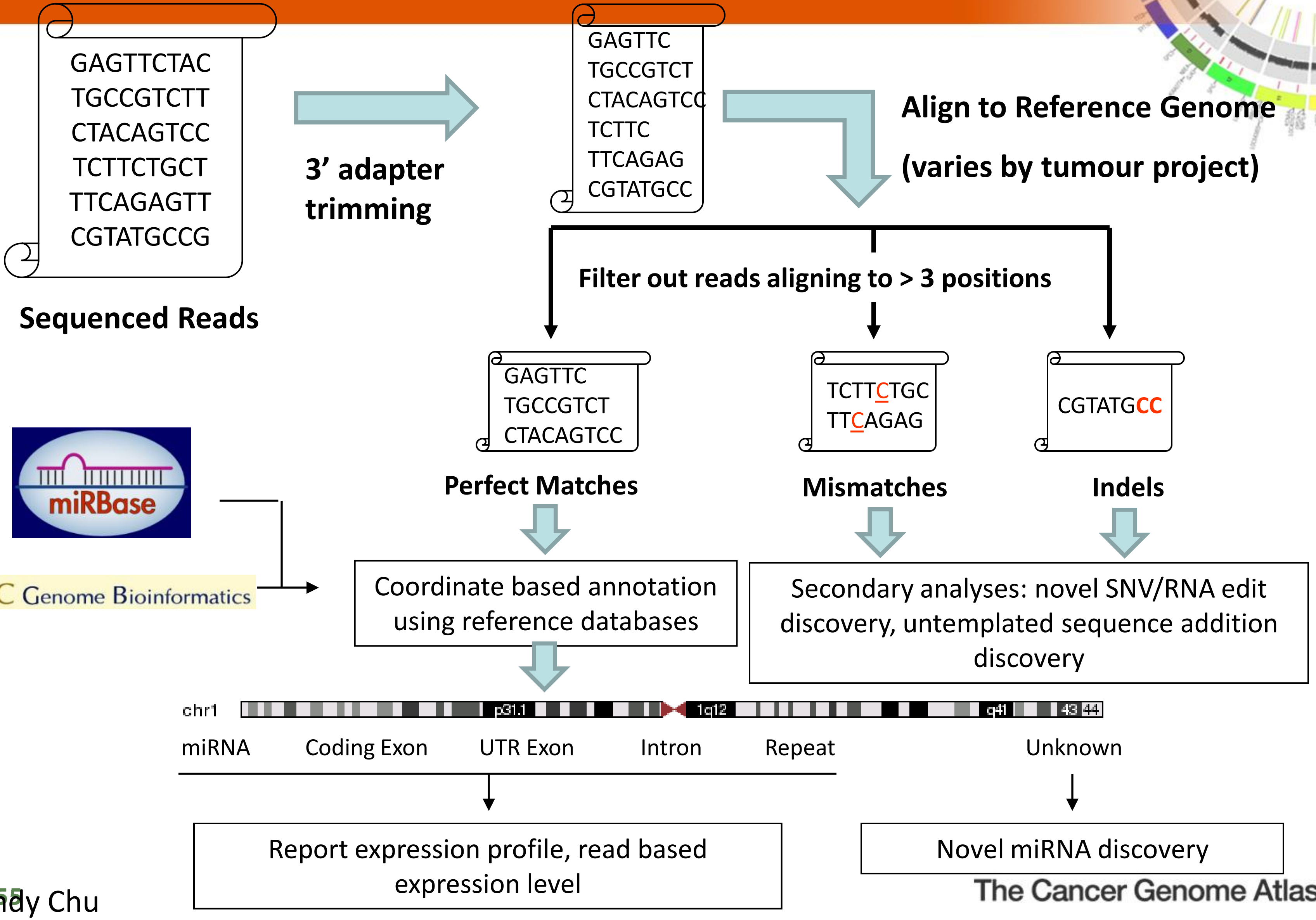
miRNA product is enriched by PCR with an index primer

The Cancer Genome Atlas

Sequence analysis pipelines

- ***Profile miRNA expression***
- ***Library quality assessment***
- ***Hierarchical clustering***
- ***Consensus clustering***
- ***MicroRNA prediction***
- ***RNA edits &/or mutations***
- ***3' untemplated additions***

miRNA sequence analysis pipeline



GAGTTCTAC
TGCCGTCTT
CTACAGTCC
TCTTCTGCT
TTCAGAGTT
CGTATGCCG

Sequenced Reads

3' adapter trimming

GAGTTC
TGCCGTCT
CTACAGTCC
TCTTC
TTCAGAG
CGTATGCC

Align to Reference Genome

(varies by tumour project)

Filter out reads aligning to > 3 positions

GAGTTC
TGCCGTCT
CTACAGTCC

Perfect Matches

TCTTCTGC
TTCAGAG

Mismatches

CGTATGCC

Indels

miRBase

UCSC Genome Bioinformatics

Coordinate based annotation using reference databases

Secondary analyses: novel SNV/RNA edit discovery, untemplated sequence addition discovery

miRNA Coding Exon UTR Exon Intron Repeat

Unknown

Report expression profile, read based expression level

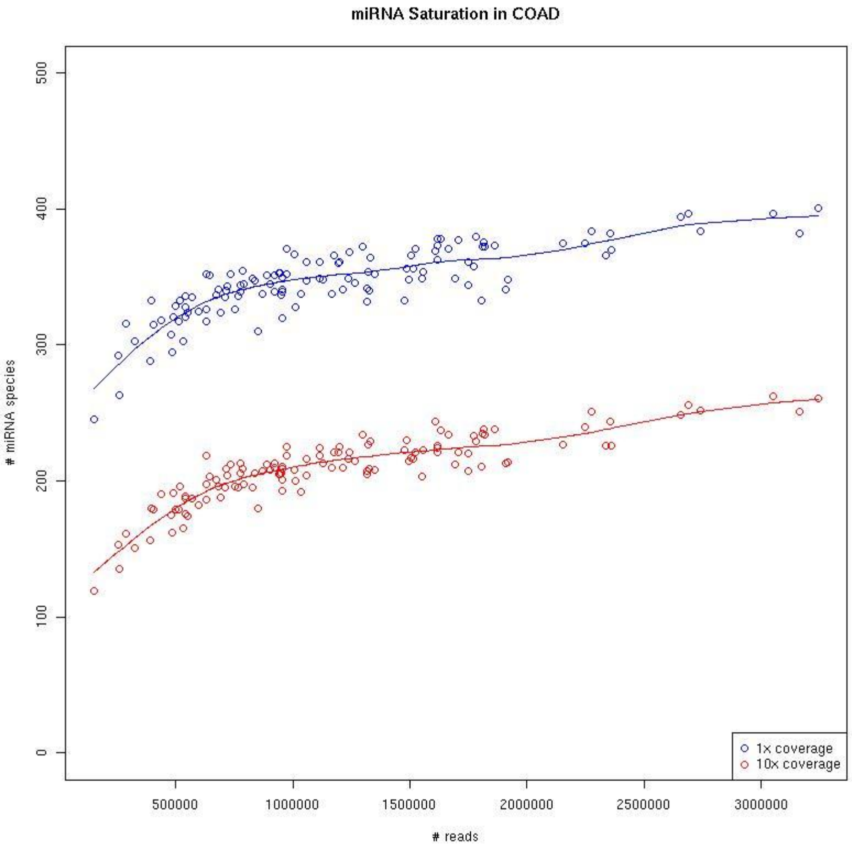
Novel miRNA discovery

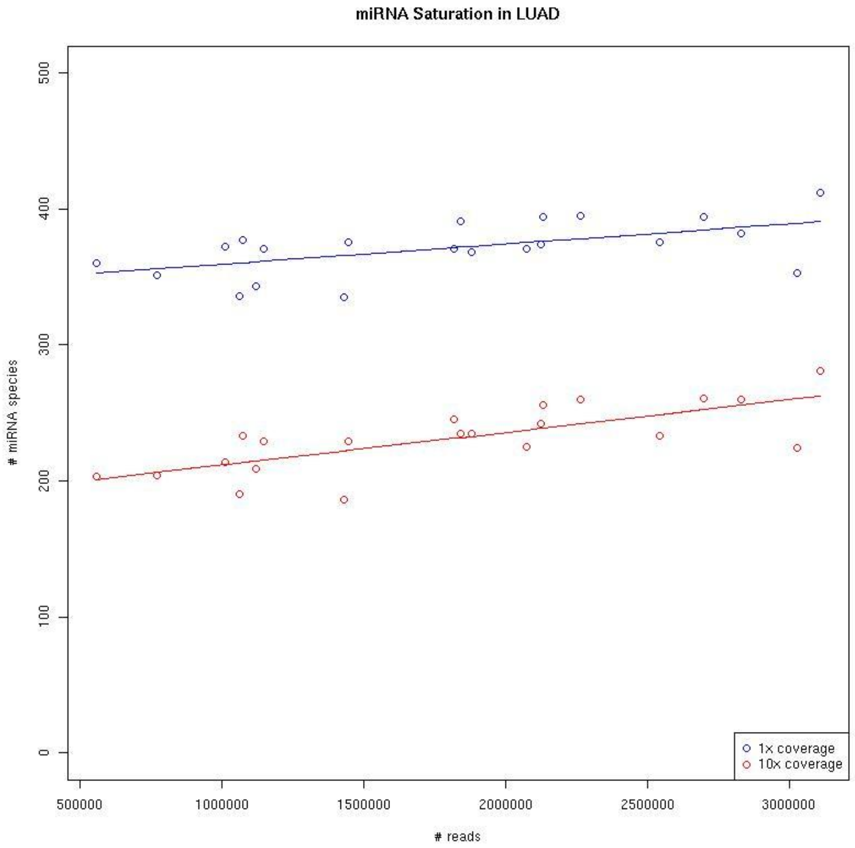
Andy Chu

The Cancer Genome Atlas

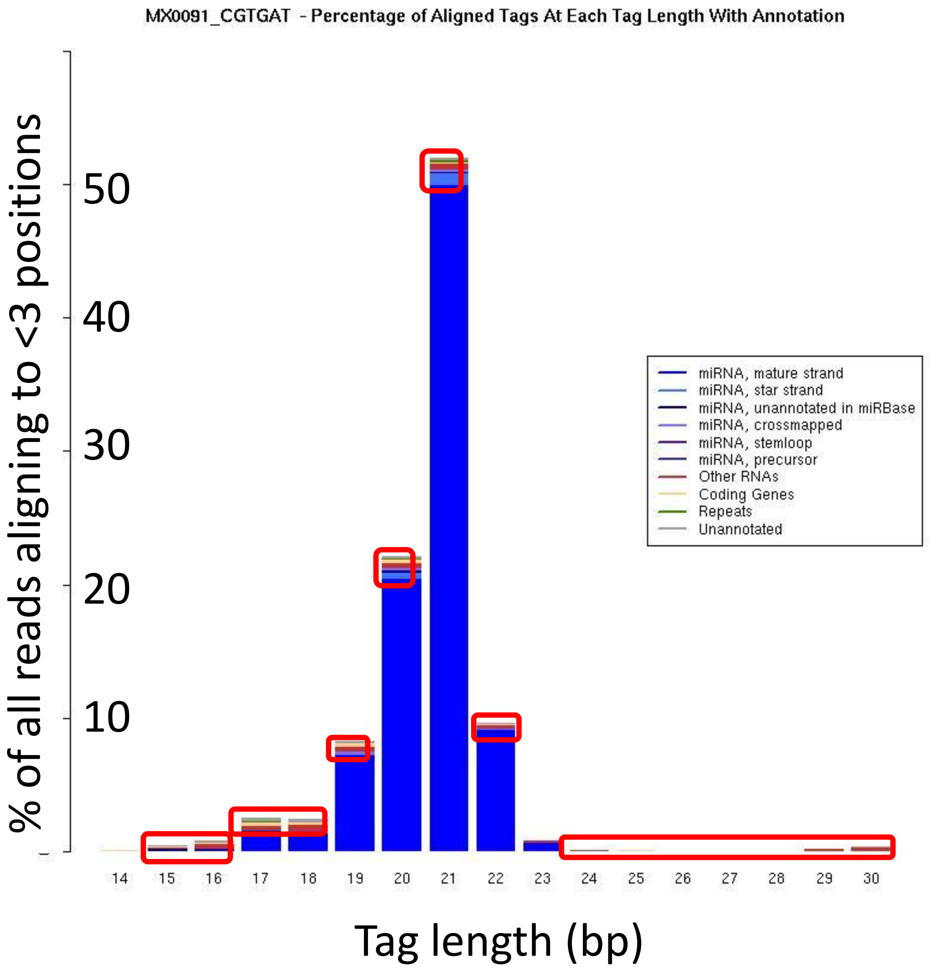
Cross-library miRNA saturation plots

By plotting the number of miRNA reads against number of miRNA species found in all samples in a given tissue, we can see the when we've captured most of the miRNAs we'd expect to see in the sample.



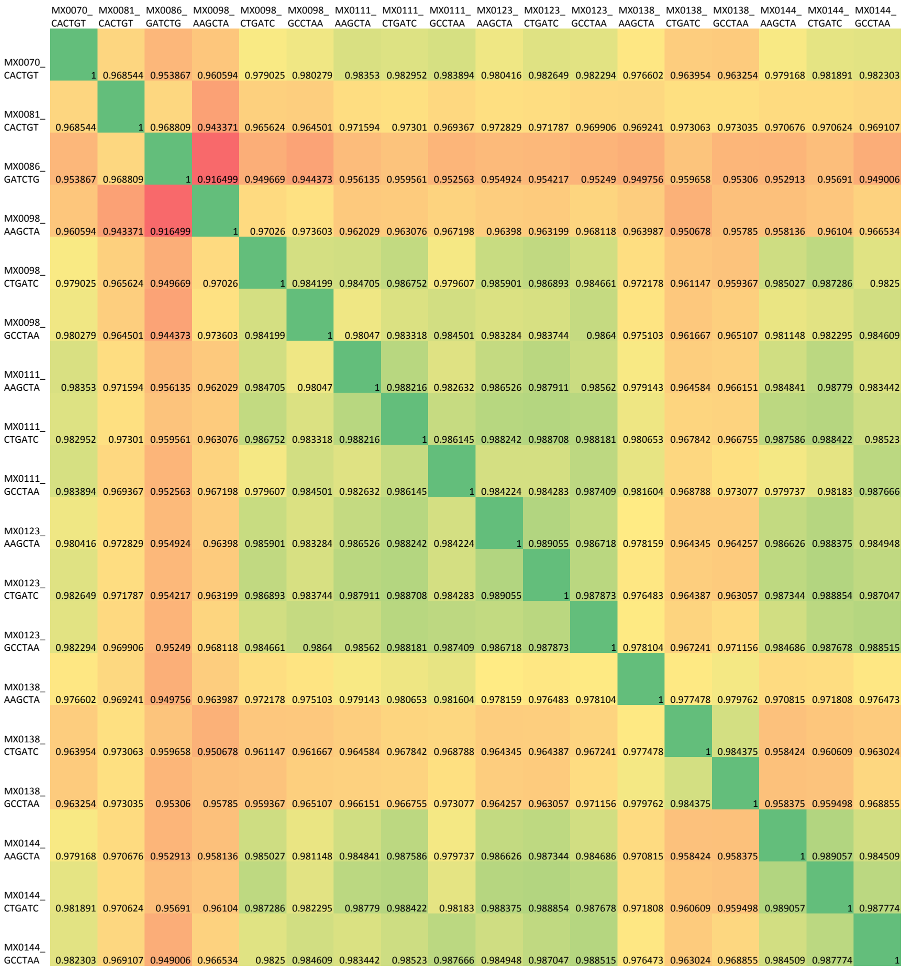


miRNA quality assurance



- Profile miRNA expression of samples through read counts to all known miRNAs
- Quality assessment – what other RNA species are present?
- Comparison of miRNA expression across multiple samples

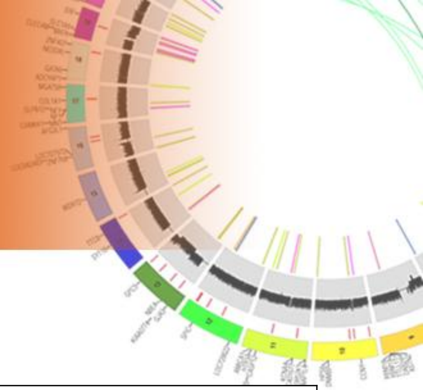
miRNA quality assurance



	MX0070_ CACTGT	MX0081_ CACTGT	MX0086_ GATCTG	MX0098_ AAGCTA	MX0098_ CTGATC	MX0098_ GCCTAA	MX0111_ AAGCTA	MX0111_ CTGATC	MX0111_ GCCTAA	MX0123_ AAGCTA	MX0123_ CTGATC	MX0123_ GCCTAA	MX0138_ AAGCTA	MX0138_ CTGATC	MX0138_ GCCTAA	MX0144_ AAGCTA	MX0144_ CTGATC	MX0144_ GCCTAA
MX0070_ CACTGT	1	0.968544	0.953867	0.960594	0.979025	0.980279	0.98353	0.982952	0.983894	0.980416	0.982649	0.982294	0.976602	0.963954	0.963254	0.979168	0.981891	0.982303
MX0081_ CACTGT	0.968544	1	0.968809	0.943371	0.965624	0.964501	0.971594	0.97301	0.969367	0.972829	0.971787	0.969906	0.969241	0.973063	0.973035	0.970676	0.970624	0.969107
MX0086_ GATCTG	0.953867	0.968809	1	0.916499	0.949669	0.944373	0.956135	0.959561	0.952563	0.954924	0.954217	0.95249	0.949756	0.959658	0.95306	0.952913	0.95691	0.949006
MX0098_ AAGCTA	0.960594	0.943371	0.916499	1	0.97026	0.973603	0.962029	0.963076	0.967198	0.96398	0.963199	0.968118	0.963987	0.950678	0.95785	0.958136	0.96104	0.966534
MX0098_ CTGATC	0.979025	0.965624	0.949669	0.97026	1	0.984199	0.984705	0.986752	0.979607	0.985901	0.986893	0.984661	0.972178	0.961147	0.959367	0.985027	0.987286	0.9825
MX0098_ GCCTAA	0.980279	0.964501	0.944373	0.973603	0.984199	1	0.98047	0.983318	0.984501	0.983284	0.983744	0.9864	0.975103	0.961667	0.965107	0.981148	0.982295	0.984609
MX0111_ AAGCTA	0.98353	0.971594	0.956135	0.962029	0.984705	0.98047	1	0.988216	0.982632	0.986526	0.987911	0.98562	0.979143	0.964584	0.966151	0.984841	0.98779	0.983442
MX0111_ CTGATC	0.982952	0.97301	0.959561	0.963076	0.986752	0.983318	0.988216	1	0.986145	0.988242	0.988708	0.988181	0.980653	0.967842	0.966755	0.987586	0.988422	0.98523
MX0111_ GCCTAA	0.983894	0.969367	0.952563	0.967198	0.979607	0.984501	0.982632	0.986145	1	0.984224	0.984283	0.987409	0.981604	0.968788	0.973077	0.979737	0.98183	0.987666
MX0123_ AAGCTA	0.980416	0.972829	0.954924	0.96398	0.985901	0.983284	0.986526	0.988242	0.984224	1	0.989055	0.986718	0.978159	0.964345	0.964257	0.986626	0.988375	0.984948
MX0123_ CTGATC	0.982649	0.971787	0.954217	0.963199	0.986893	0.983744	0.987911	0.988708	0.984283	0.989055	1	0.987873	0.976483	0.964387	0.963057	0.987344	0.988854	0.987047
MX0123_ GCCTAA	0.982294	0.969906	0.95249	0.968118	0.984661	0.9864	0.98562	0.988181	0.987409	0.986718	0.987873	1	0.978104	0.967241	0.971156	0.984686	0.987678	0.988515
MX0138_ AAGCTA	0.976602	0.969241	0.949756	0.963987	0.972178	0.975103	0.979143	0.980653	0.981604	0.978159	0.976483	0.978104	1	0.977478	0.979762	0.970815	0.971808	0.976473
MX0138_ CTGATC	0.963954	0.973063	0.959658	0.950678	0.961147	0.961667	0.964584	0.967842	0.968788	0.964345	0.964387	0.967241	0.977478	1	0.984375	0.958424	0.960609	0.963024
MX0138_ GCCTAA	0.963254	0.973035	0.95306	0.95785	0.959367	0.965107	0.966151	0.966755	0.973077	0.964257	0.963057	0.971156	0.979762	0.984375	1	0.958375	0.959498	0.968855
MX0144_ AAGCTA	0.979168	0.970676	0.952913	0.958136	0.985027	0.981148	0.984841	0.987586	0.979737	0.986626	0.987344	0.984686	0.970815	0.958424	0.958375	1	0.989057	0.984509
MX0144_ CTGATC	0.981891	0.970624	0.95691	0.96104	0.987286	0.982295	0.98779	0.988422	0.98183	0.988375	0.988854	0.987678	0.971808	0.960609	0.959498	0.989057	1	0.987774
MX0144_ GCCTAA	0.982303	0.969107	0.949006	0.966534	0.9825	0.984609	0.983442	0.98523	0.987666	0.984948	0.987047	0.988515	0.976473	0.963024	0.968855	0.984509	0.987774	1

- Profile miRNA expression of samples through read counts to all known miRNAs
- Quality assessment – what other RNA species are present?
- Comparison of miRNA expression across multiple samples

miRNA-Seq expression in AML



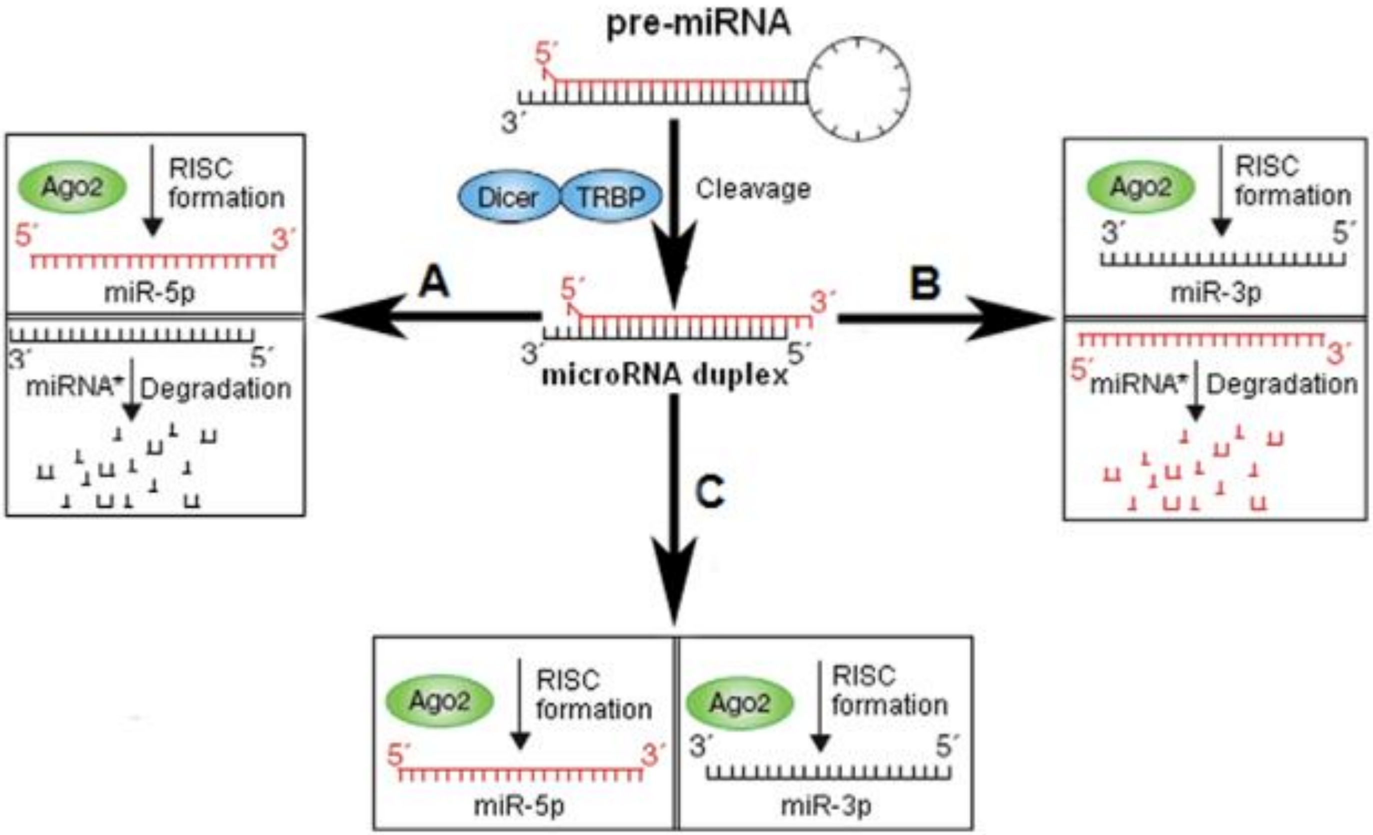
- Top 10 expressed miRNAs are leukemia

Top 10 miRNAs	Average tags per million	Role in cancers
hsa-miR-21	119,563	Overexpressed in many tumours including leukemias
hsa-miR-142	96,583	Aberrant expression in leukemia
hsa-miR-92a-2	96,005	Overexpressed in many tumours including leukemias
hsa-miR-10a	89,865	Down regulated in chronic myeloid leukemia
hsa-miR-223	39,032	Aberrant expression in AML; *CEBPA* target
hsa-miR-181a-1	38,565	Aberrant expression in leukemia and other cancers; HOX regulator
hsa-miR-30e	35,442	Metastasis related in hepatocellular carcinoma
hsa-miR-25	32,725	Aberrant expression in many tumours
hsa-miR-148a	31,990	Hypermethylated in breast cancer; differentiates T & B cell leukemias; targets *DNMT3*
hsa-let-7b	28,928	Highly discriminatory between Acute Lymphocytic Leukemia and Acute Myeloid Leukemia (over-expressed)

Jiang Q. *et al*. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nuc. Acids Res* 37:D98-104

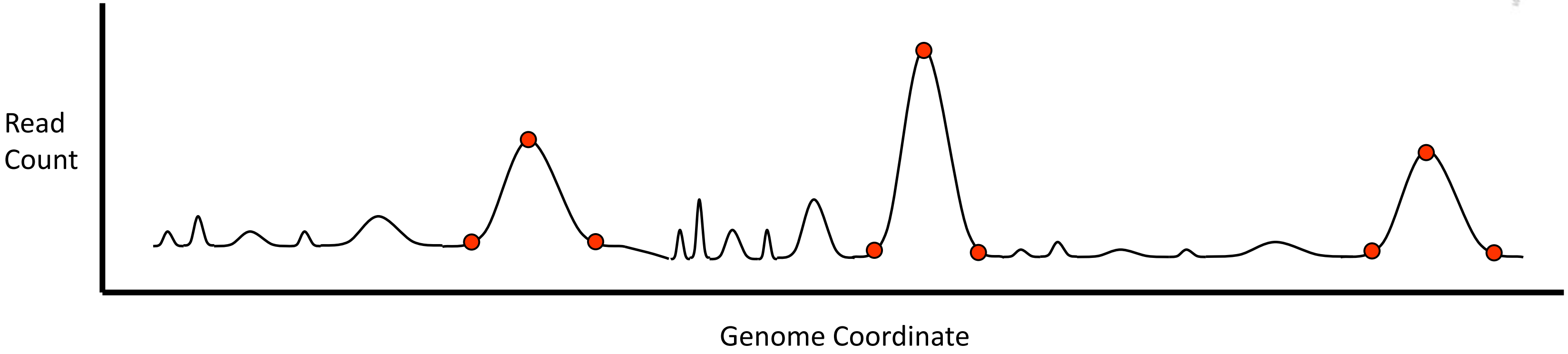
The Cancer Genome Atlas

Mature vs star miRNAs



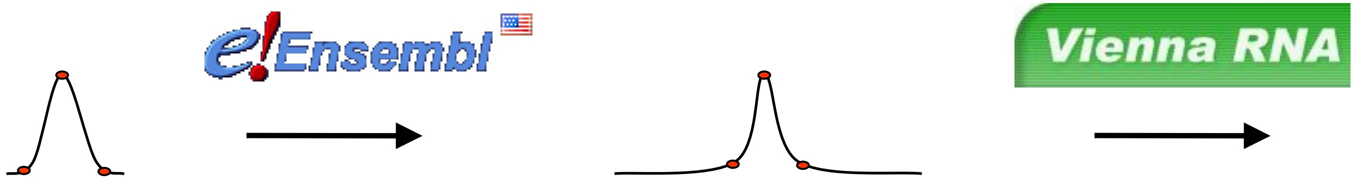
Novel microRNA prediction

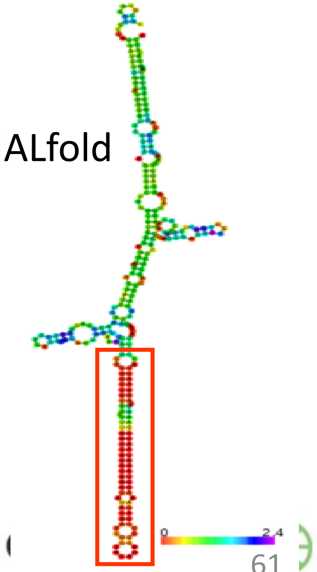
Aggregate all filtered reads from a set of samples.
Use FindPeaks to find relative expression “hotspots”.



Re-annotate the peaks themselves, rather than using the read annotation.
This allows greater stringency (eg. bp overlapped) than the original annotation.

Add flanking sequence around each peak and attempt to fold the RNA using RNALfold (ViennaRNA package), then extract structure information using RNAfold.

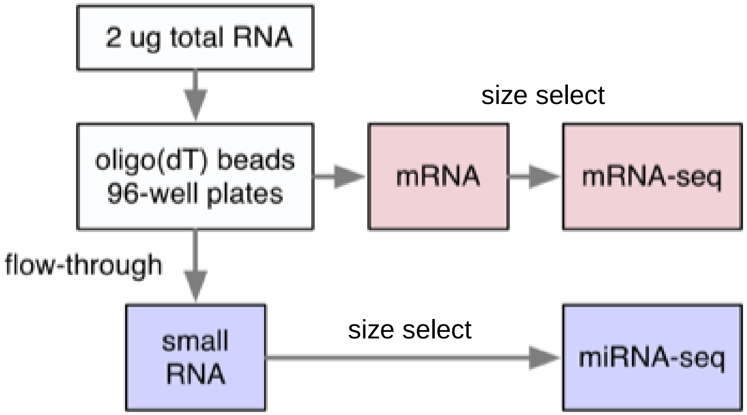


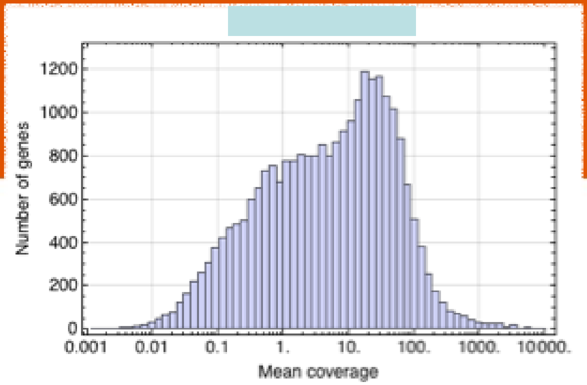


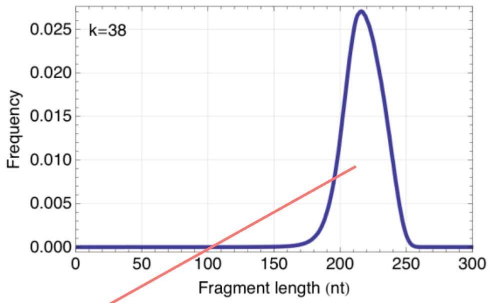
The Cancer

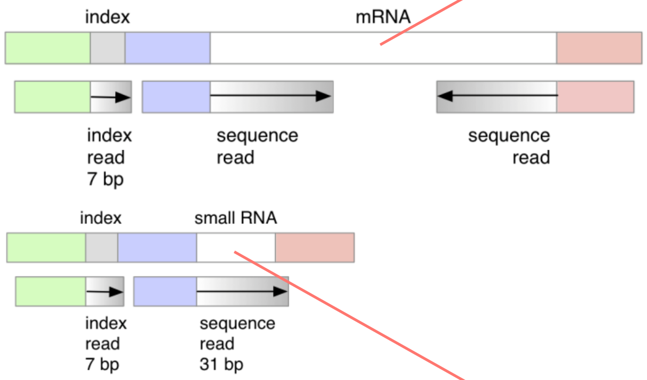
mRNA-seq and miRNA-seq data

Library construction



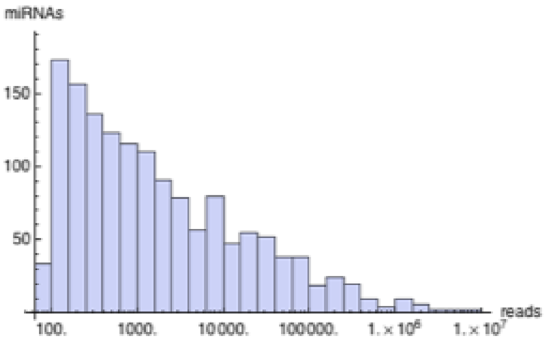


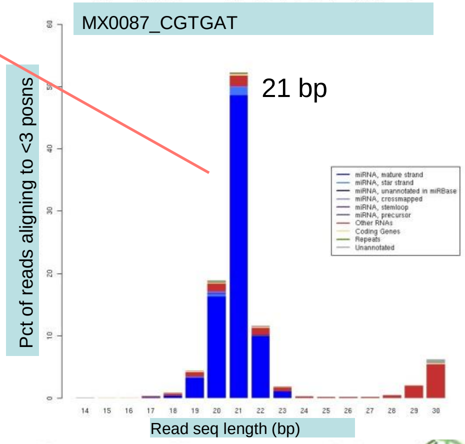




HiSeq 2000 v3
2 libraries/lane

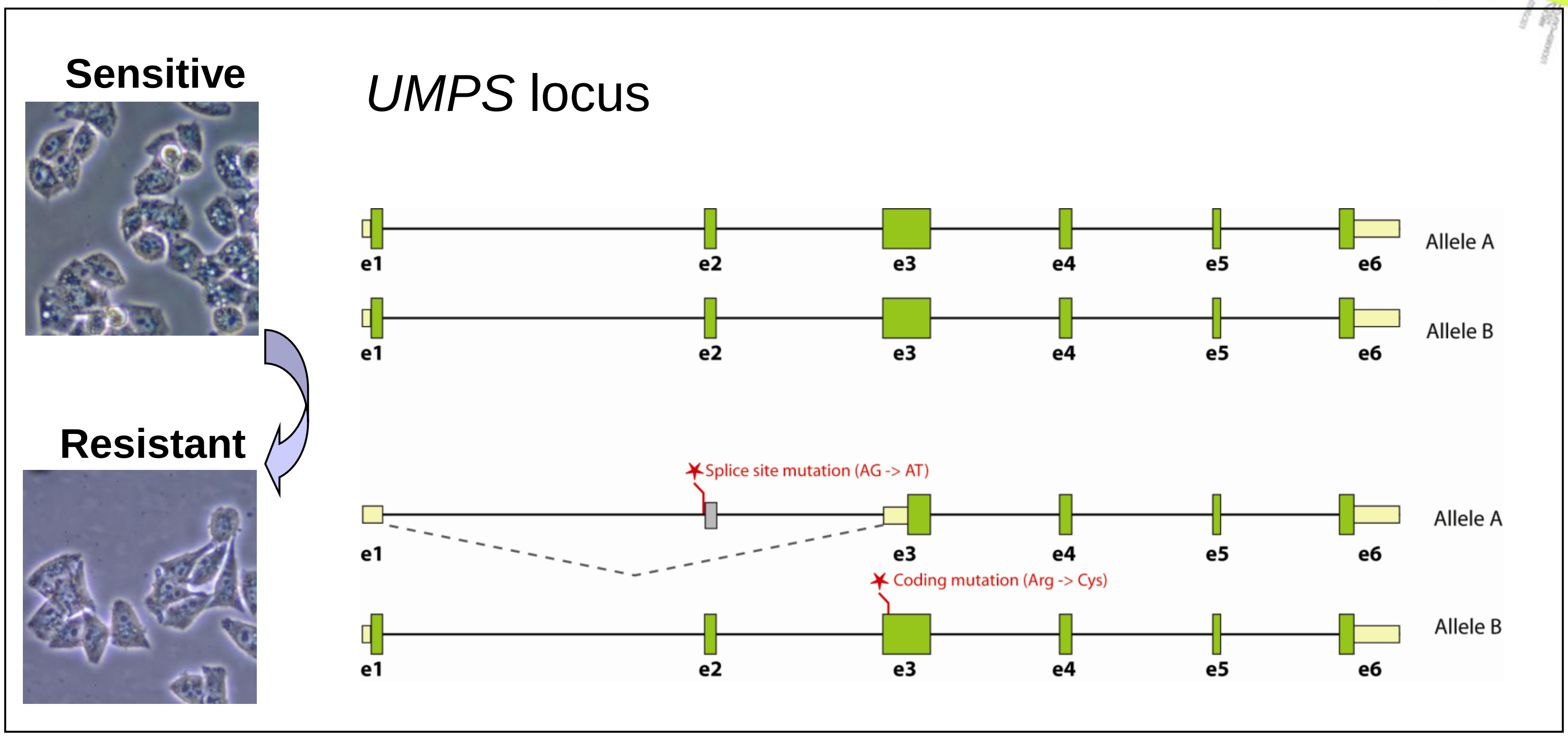
HiSeq 2000 v3
N libraries/lane





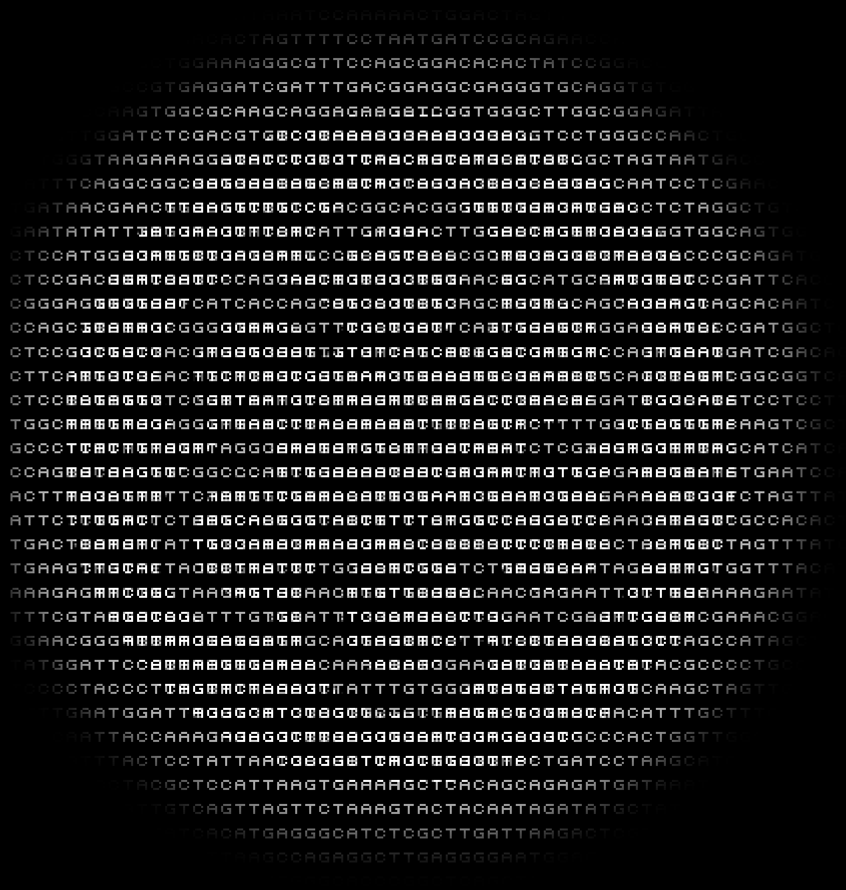
The Cancer Genome Atlas

UMPS mutations affect

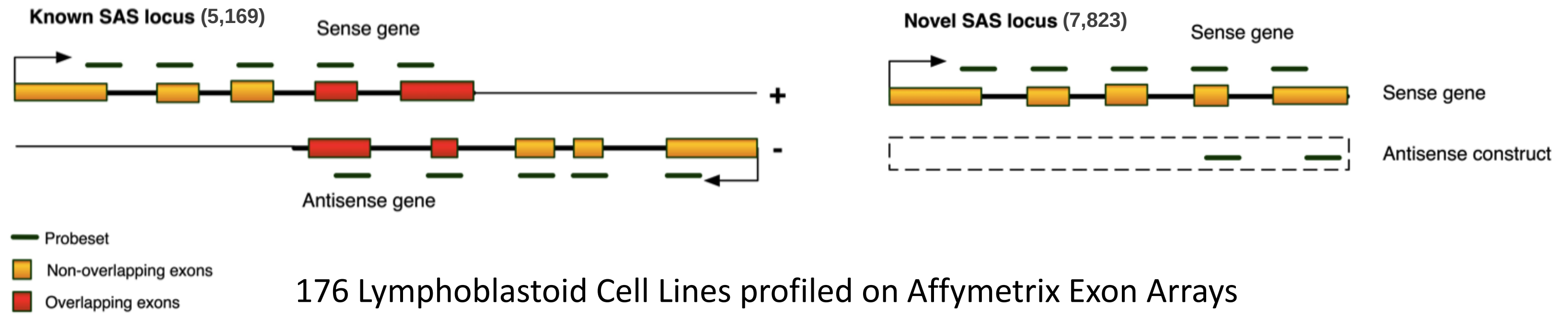


UMPS catalyses the the last step in the pyrimidine nucleotide synthesis pathway: conversion of orotate to UMP.

UMPS is required for 5 FU induced cell death.



Correlating alternative expression and antisense transcription



176 Lymphoblastoid Cell Lines profiled on Affymetrix Exon Arrays

$$\textbf{Probeset splice index (SI)} = \frac{\text{Probeset expression}}{\text{Sense gene expression}}$$

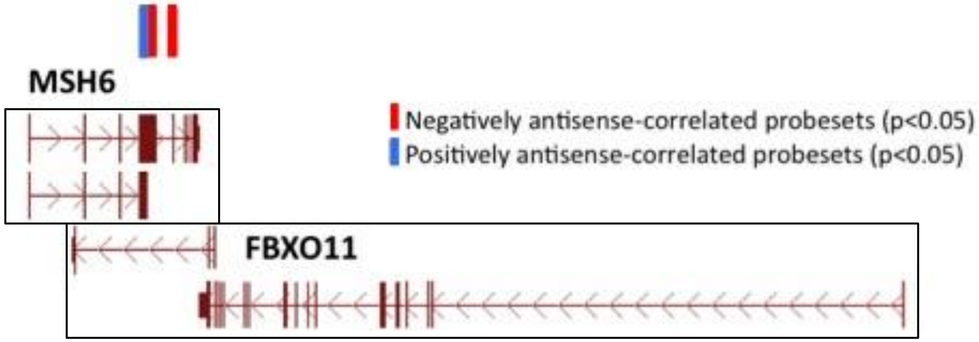
Large SI: probeset inclusion in mRNA
Small SI: probeset exclusion in mRNA

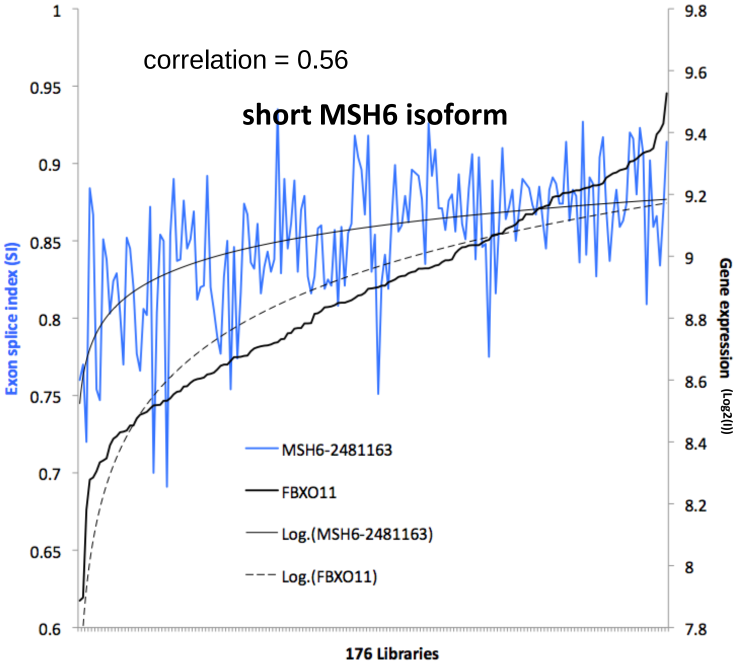
Correlations: SI vs Antisense gene expression
Bonferroni correction of correlation p-values

The expression of **24%** of 7,162 probesets (402 genes) is significantly correlated to antisense gene expression

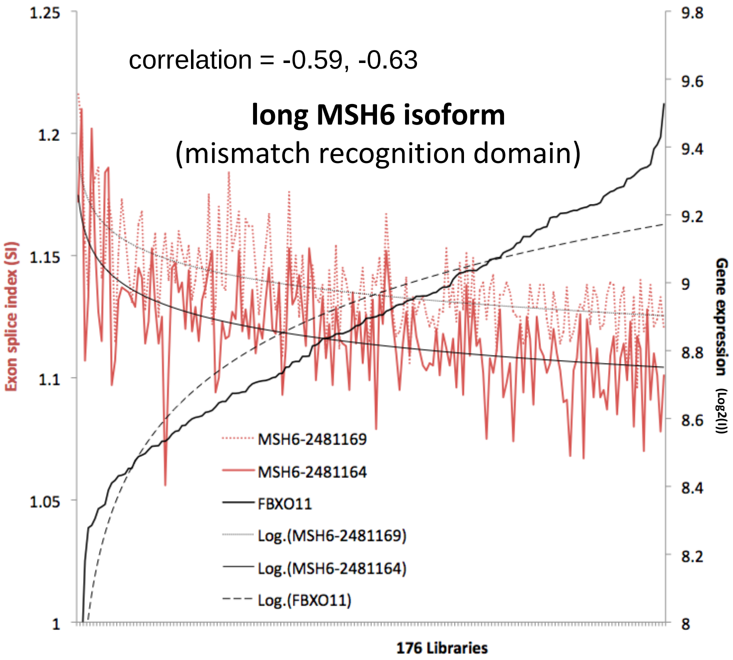
Morrissy, Griffith, and Marra, 2010, *Genome Research, in revision*

Antisense-correlated probe set expression: MSH6


MSH6
🟥 Negatively antisense-correlated probesets (p<0.05)
🟦 Positively antisense-correlated probesets (p<0.05)
FBXO11


correlation = 0.56

short MSH6 isoform

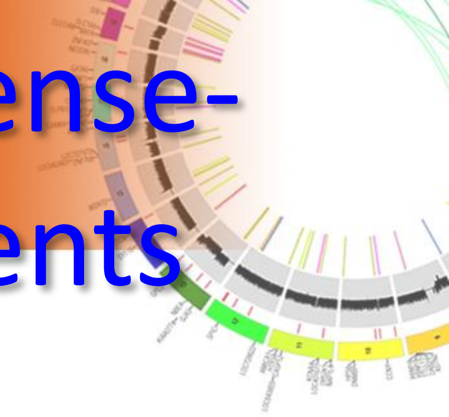

correlation = -0.59, -0.63

long MSH6 isoform
(mismatch recognition domain)

85% of expressed SAS loci (n = 402) have significant correlations between antisense transcription and sense gene probeset inclusion & exclusion events (i.e. splicing)

Morrissy, Griffith, and Marra, 2010, *Genome Research, in revision*

Cancer-associated antisense-correlated splicing events



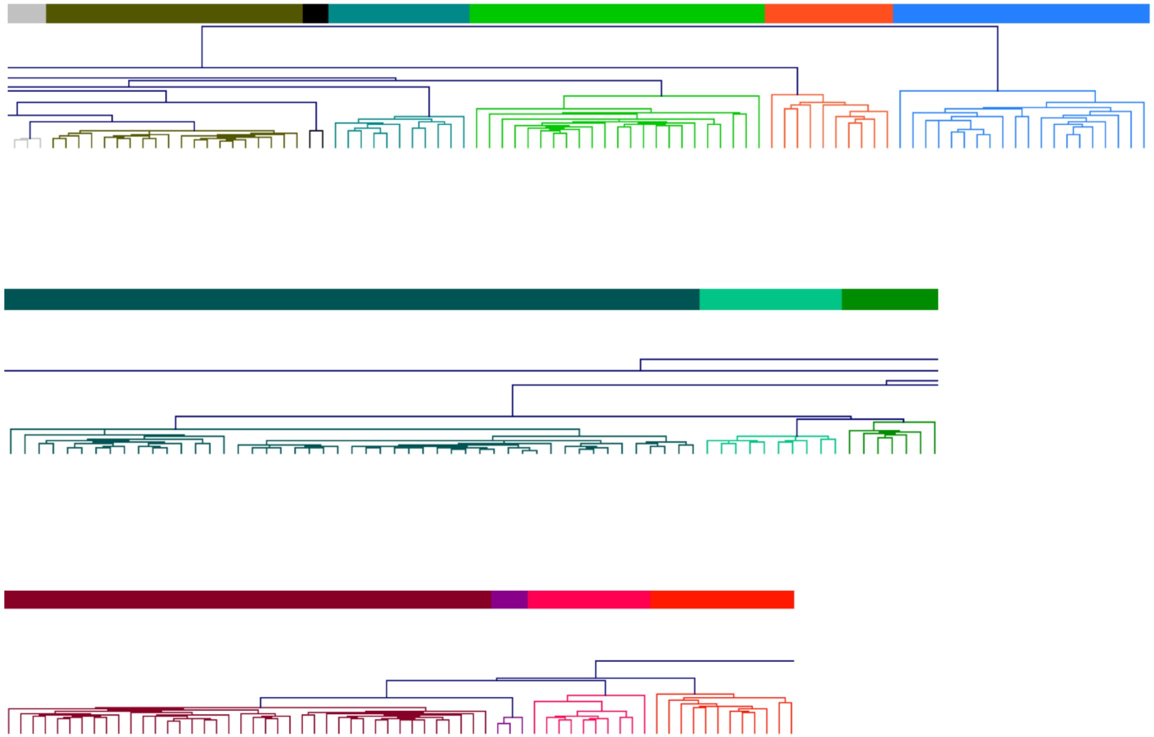
- Known SAS gene pairs have altered expression ratios in cancer (Chen et al., TiG, 2005)
- Intronic antisense transcripts correlate to the degree of tumor differentiation in prostate cancer (Reis et al., Oncogene, 2004)
- Many known cancer-related genes have novel antisense transcription
  - ex. p15, Yu et al., Nature, 2008
  - 215 of 389 Cancer Gene Census genes (p-value=4.2×10^{-9})

Goal: Assess cancer-specific antisense-correlated splicing events using exon array data

Focus: 266 Glioblastoma multiforme samples from The Cancer Genome Atlas (TCGA)

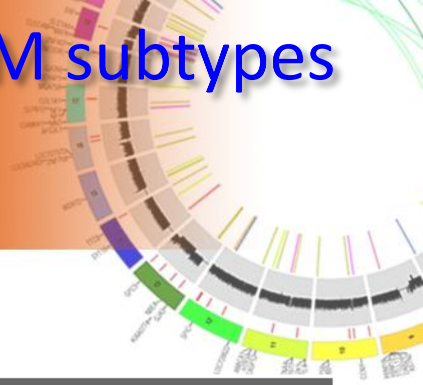
Antisense-correlated splicing events have tissue-specific patterns

- inclusion & exclusion of probesets is tissue specific
- like gene expression values, SI values can be used to group samples
- unsupervised hierarchical clustering of all 17,420 probesets expressed in normal samples recapitulates groups of normal tissues



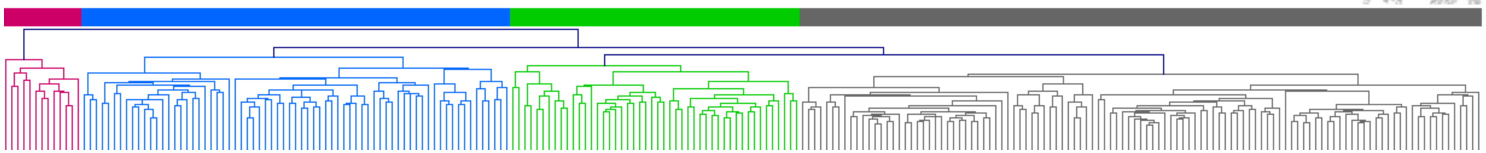
adult brain and cerebellum	7	🟩
GBM controls (brain)	10	🟩
blood (MS patients)	10	🟥
erythrocytes	3	🟪
LCLs	40	🟥
blood (MS patients)	12	🟥
spinal cord	20	🟦
fetal brain	49	🟦
prostata	3	⬜
lung	20	🟫
thymus	2	⬛
stem cells and fibroblasts	11	🟦
stomach	23	🟩
colon	10	🟧

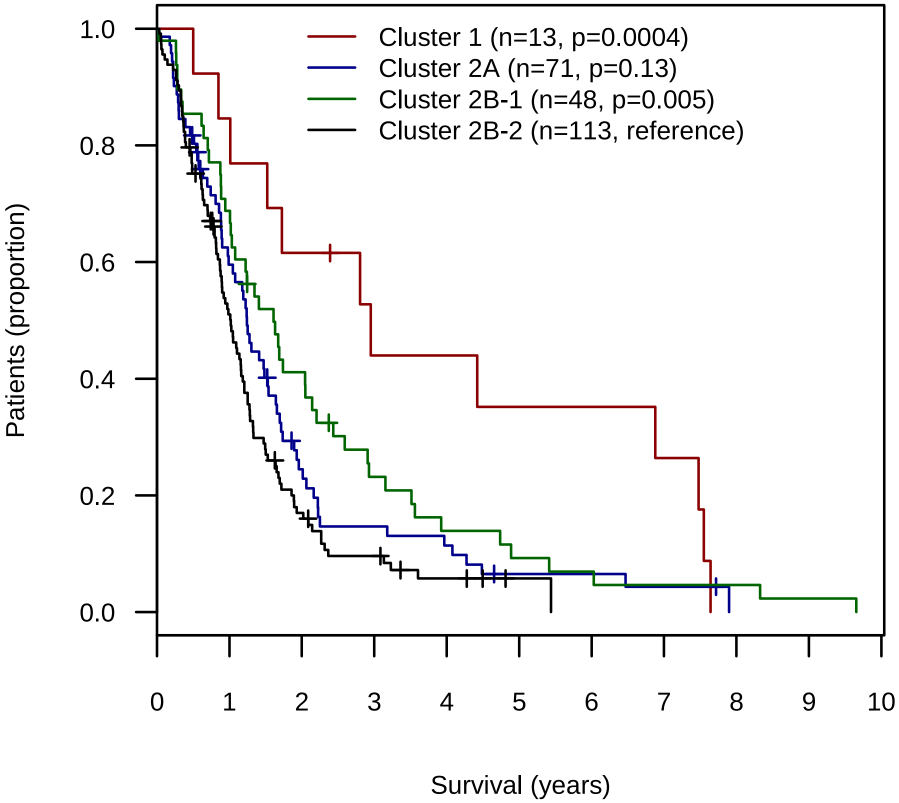
Antisense-correlated splicing events reveal GBM subtypes



1,000 probesets (629 genes) with cancer-specific alternative inclusion can be used to find GBM sub-types

Cluster 1	13	🟥
Cluster 2A	71	🟦
Cluster 2B1	48	🟩
Cluster 2B2	113	⬛




Cluster 1 (n=13, p=0.0004)
Cluster 2A (n=71, p=0.13)
Cluster 2B-1 (n=48, p=0.005)
Cluster 2B-2 (n=113, reference)
Patients (proportion)
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
8
9
10
Survival (years)

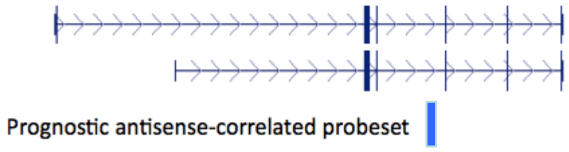
Known GBM candidate driver genes have prognostic splicing events

Expressed in GBM	Antisense-correlated splicing	Cancer-specific isoforms	GBM-specific isoforms
A2M	Y	Y	Y
AKT3	Y	Y	Y
AVIL	Y	Y	Y
CCND2	Y	Y	Y
CDKN2C	Y	Y	Y
EGFR	Y	Y	Y
PIK3R1	Y	Y	Y
PTEN	Y	Y	Y
SPRY2	Y	Y	Y
APC	Y	Y	Y
FOXO1	Y	Y	Y
PLCL2	Y	Y	Y
TSC1	Y	Y	Y
CCND1	Y	Y	
FGFR1	Y	Y	
KLF6	Y	Y	
PLCB1	Y	Y	
EPHA3	Y		
PTPN11	Y		
FGFR2			
IFNW1			
SH3GL2			
CBL			
FOXO3			
PTPRB			
TUBGCP2			
TBP			
PIK3C2B			
TP53			
FRS2			
CRK			
IRS1			
BNC2			

- 33 of 82 candidate driver genes are expressed SAS genes
  - 19 / 33 had antisense-correlated splicing
  - 17 / 19 cancer-specific splicing, 13 / 19 GBM-specific
- 6 of these genes have exons found within the set of 1,000 exons used to generate the patient clusters

Identifying prognostic splicing events using driver genes

PLCL2: phospholipase C-like 2



- intronic probeset associated with survival (corrected P = 0.038)
  - **inclusion**: 484 days median survival (109 patients)
  - **exclusion**: 682 days median survival (136 patients)

The Cancer Genome Atlas

Antisense-correlated splicing events in cancer

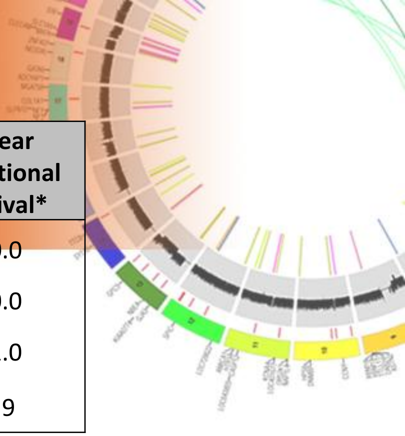
- Antisense transcription is highly correlated to the alternative processing of sense genes in both normal and disease states

- Probesets with antisense-correlated splicing can be used to find clinically-relevant groups of GBM patients, differing in median survival and in response to therapy
  - this is a new approach to addressing the molecular heterogeneity of human cancers

Goal: Identify signature of antisense-correlated events prognostic of survival or chemotherapy response

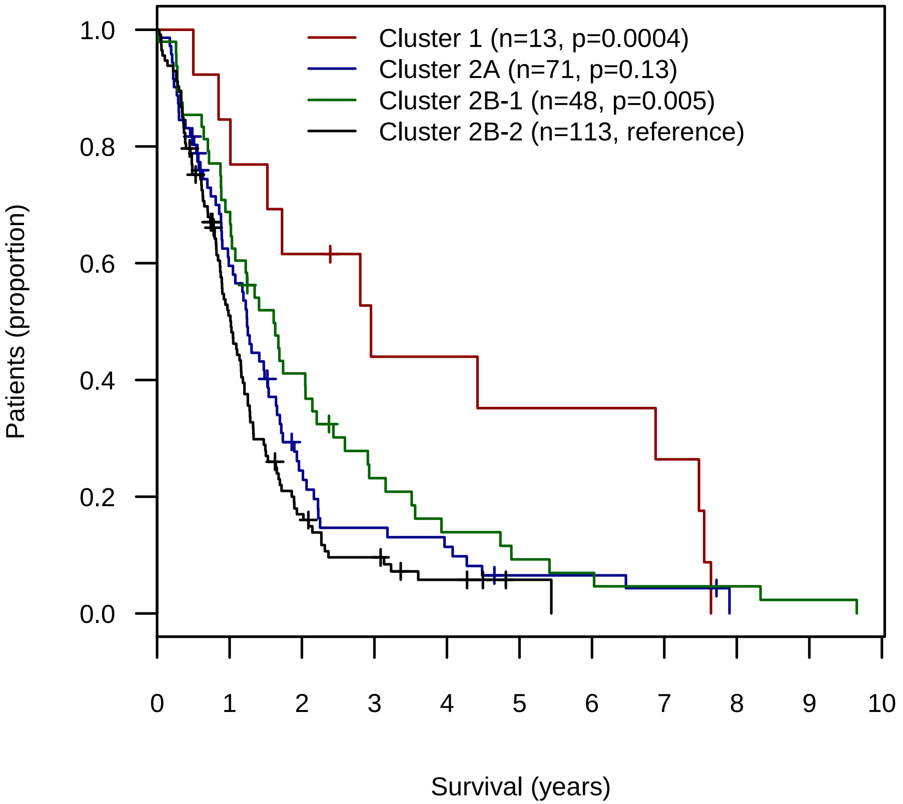
  - these events represent a shortlist of genes whose alternative expression is relevant to cancer biology, and which have putative antisense-mediated regulation
  - the focus on cancer-specific events is designed to identify novel putative targets for therapeutics or diagnostics

Clinical features of GBM subtypes



	Number of patients	Median survival (days)	Median age	1-Year Survival	2-Year Survival	5-Year conditional survival*
Cluster 1	13	1,024	33	84.6	61.5	50.0
Cluster 2A	71	447	56	56.3	21.1	20.0
Cluster 2B1	48	551	58.5	68.8	39.6	21.0
Cluster 2B2	113	345	57	47.8	15.0	5.9

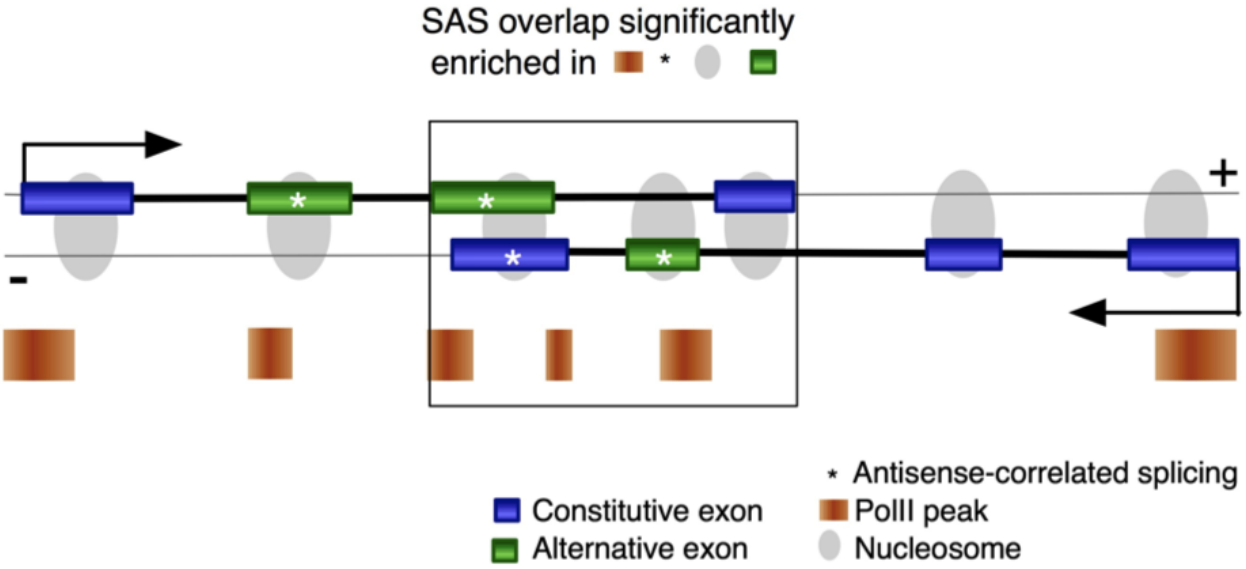
* 5-year survival rate was calculated for the subset of patients still alive at 2 years



Treatment differences?

- Temozolomide: 100 / 249 patients

Antisense transcription: a model of alternative splicing regulation

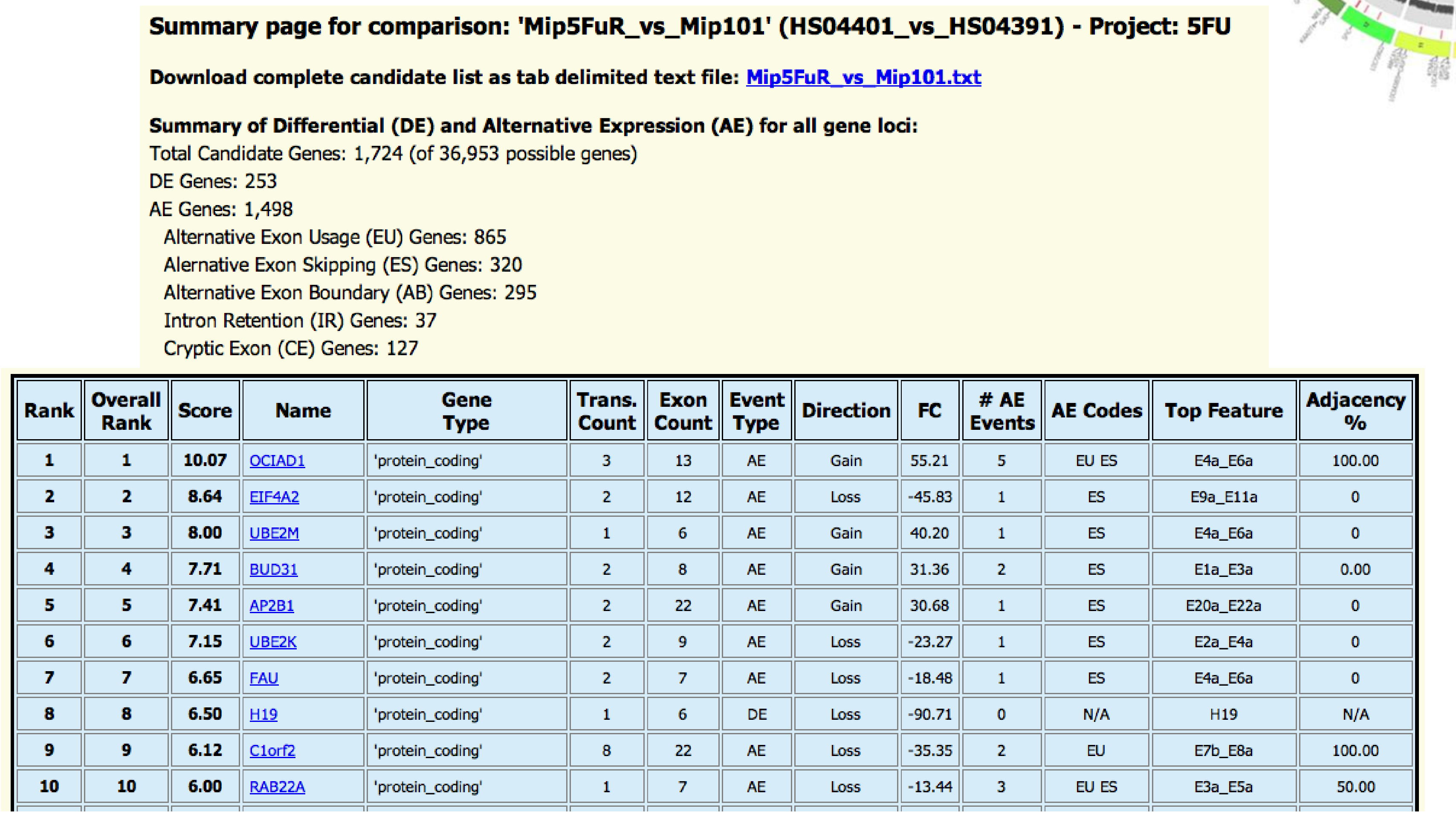

SAS overlap significantly enriched in ▨ * ● 🟩
+
-
*
Antisense-correlated splicing
🟦 Constitutive exon
🟩 Alternative exon
▨ PolII peak
● Nucleosome

↑Exons : ↑Nucleosomes : ↓PolII speed : ↑alternative splicing

Morrissy, Griffith, and Marra, 2010, *Genome Research, in revision*

Data browsing and access

www.AlexaPlatform.org/alexa_seq/



Summary page for comparison: 'Mip5FuR_vs_Mip101' (HS04401_vs_HS04391) - Project: 5FU

Download complete candidate list as tab delimited text file: Mip5FuR_vs_Mip101.txt

Summary of Differential (DE) and Alternative Expression (AE) for all gene loci:

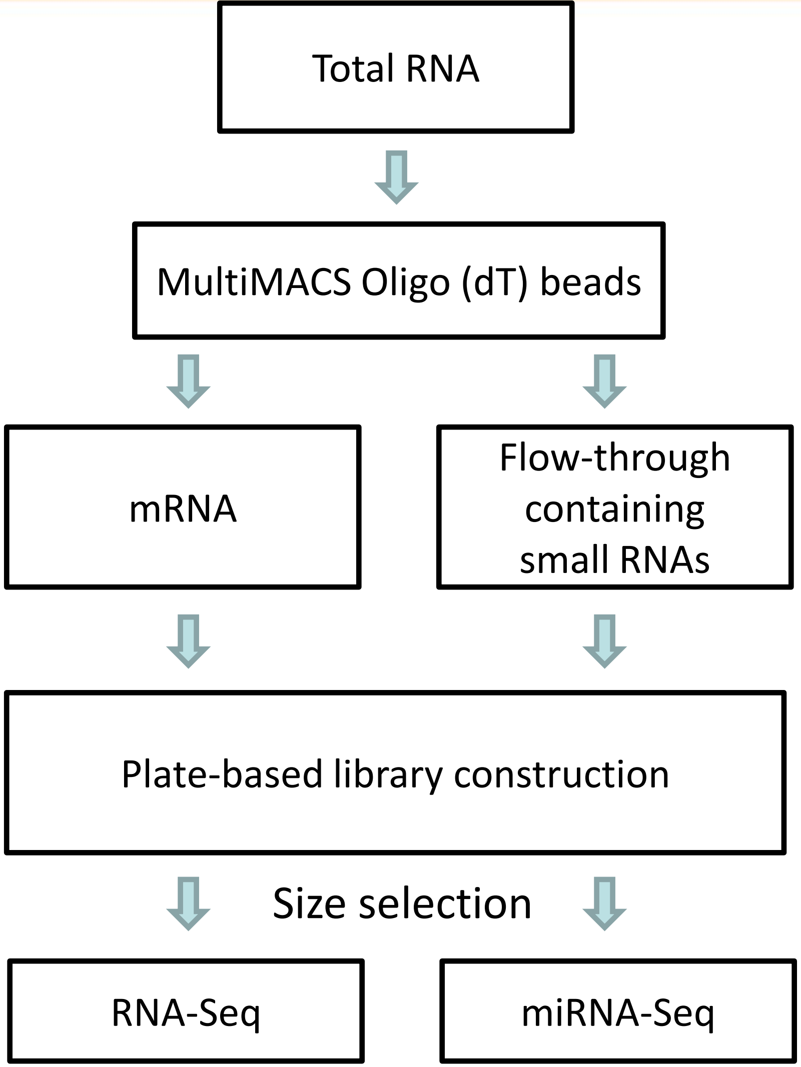
Total Candidate Genes: 1,724 (of 36,953 possible genes)
DE Genes: 253
AE Genes: 1,498
Alternative Exon Usage (EU) Genes: 865
Alernative Exon Skipping (ES) Genes: 320
Alternative Exon Boundary (AB) Genes: 295
Intron Retention (IR) Genes: 37
Cryptic Exon (CE) Genes: 127

Rank	Overall Rank	Score	Name	Gene Type	Trans. Count	Exon Count	Event Type	Direction	FC	# AE Events	AE Codes	Top Feature	Adjacency %
1	1	10.07	OCIAD1	'protein_coding'	3	13	AE	Gain	55.21	5	EU ES	E4a_E6a	100.00
2	2	8.64	EIF4A2	'protein_coding'	2	12	AE	Loss	-45.83	1	ES	E9a_E11a	0
3	3	8.00	UBE2M	'protein_coding'	1	6	AE	Gain	40.20	1	ES	E4a_E6a	0
4	4	7.71	BUD31	'protein_coding'	2	8	AE	Gain	31.36	2	ES	E1a_E3a	0.00
5	5	7.41	AP2B1	'protein_coding'	2	22	AE	Gain	30.68	1	ES	E20a_E22a	0
6	6	7.15	UBE2K	'protein_coding'	2	9	AE	Loss	-23.27	1	ES	E2a_E4a	0
7	7	6.65	FAU	'protein_coding'	2	7	AE	Loss	-18.48	1	ES	E4a_E6a	0
8	8	6.50	H19	'protein_coding'	1	6	DE	Loss	-90.71	0	N/A	H19	N/A
9	9	6.12	C1orf2	'protein_coding'	8	22	AE	Loss	-35.35	2	EU	E7b_E8a	100.00
10	10	6.00	RAB22A	'protein_coding'	1	7	AE	Loss	-13.44	3	EU ES	E3a_E5a	50.00

Transcriptome library construction

Cryoport

Biomek FX
(Beckman Coulter)


Total RNA

MultiMACS Oligo (dT) beads

mRNA

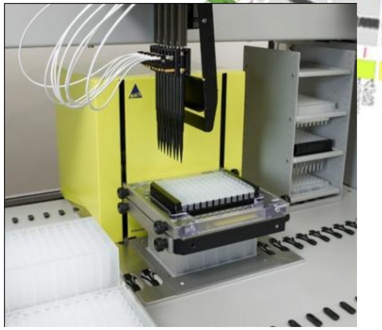
Flow-through containing small RNAs

Plate-based library construction

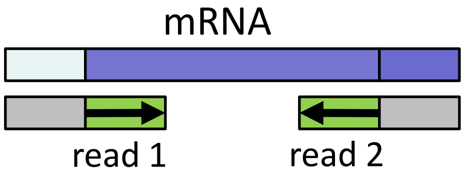
Size selection

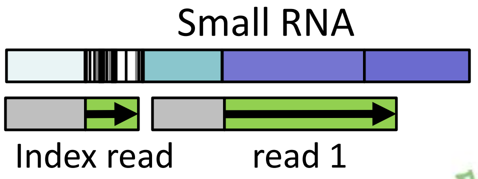
RNA-Seq

miRNA-Seq

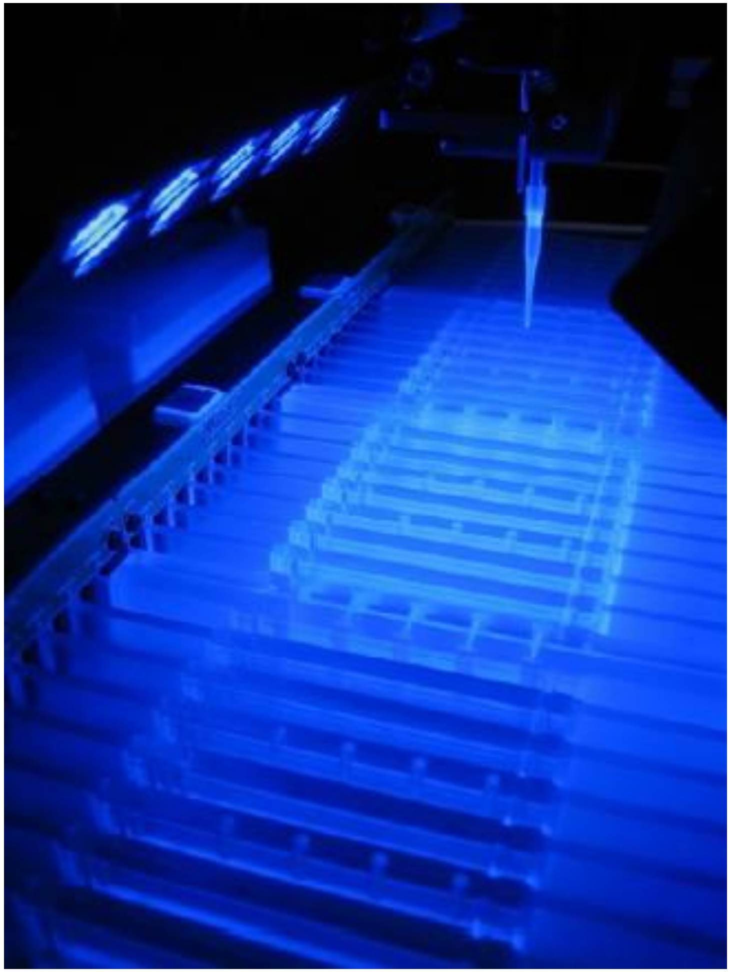

MultiMACS separator
(Miltenyi Biotech)

Caliper GX QC

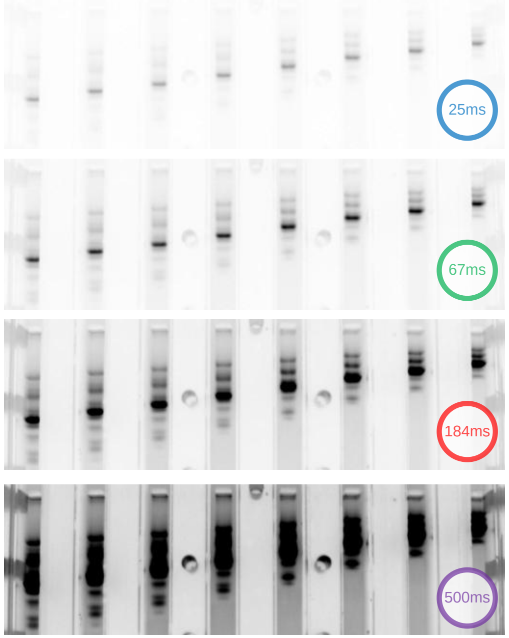




Automated size-selection

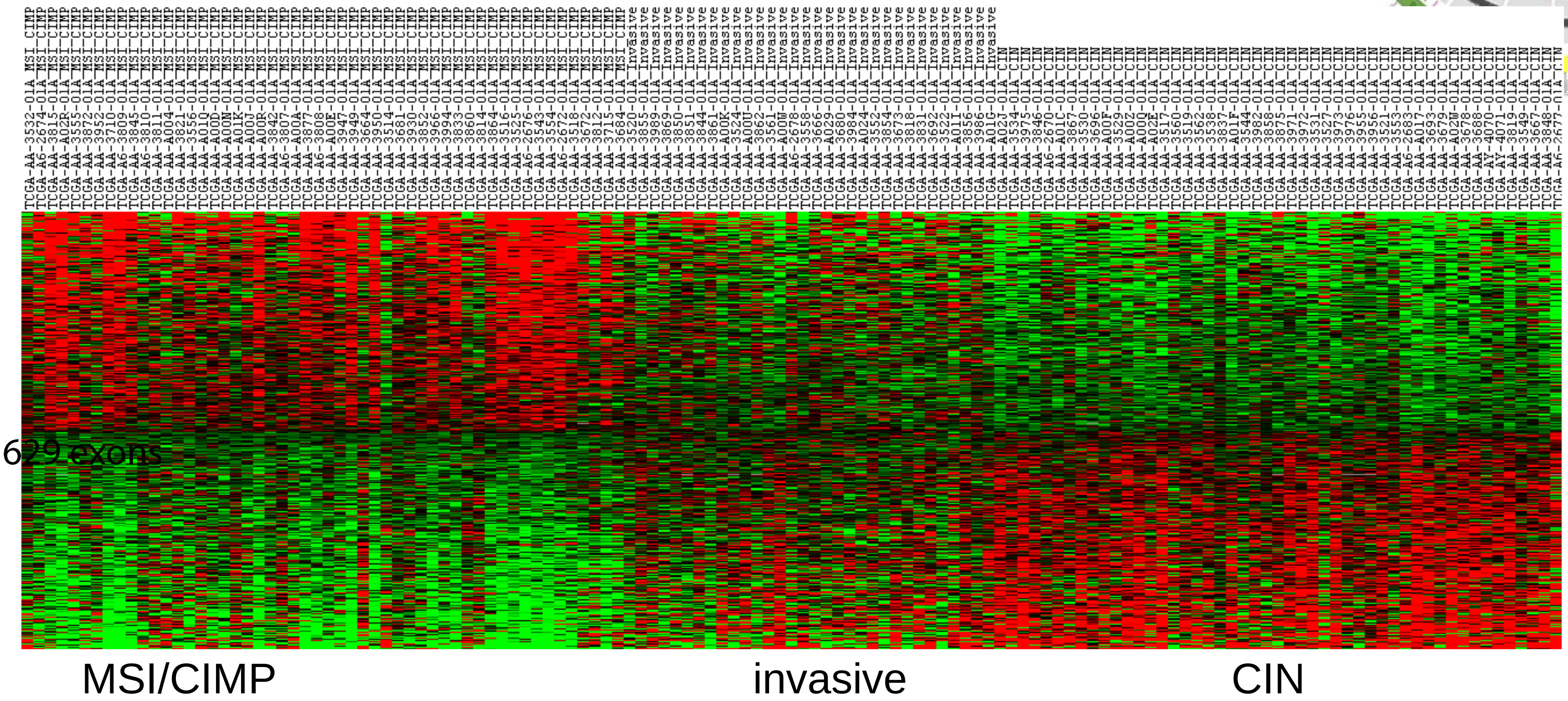


- Individual channel voltage control
- In-channel band sizing
- Optimized for miRNA



__See R. Coope poster__

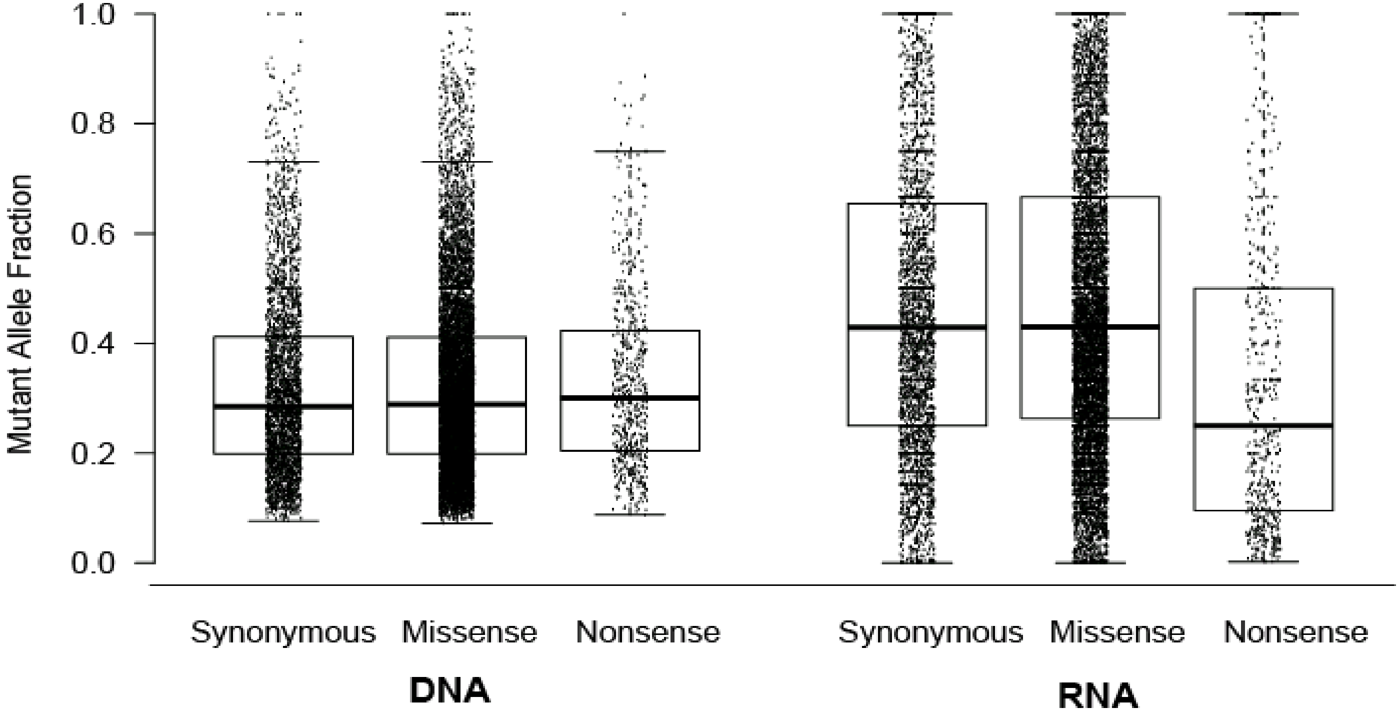
Exon-level differential expression in CRC



- Exons differentially spliced between MSI and CIN expression subtypes (P<0.0001)
- Out of ~155K probe, detected more differences among the tumors over chance expected (found: 629, chance est ~15 at P<0.0001)

Chad Creighton, BCM

Nonsense mutations have reduced mutant allelic fraction

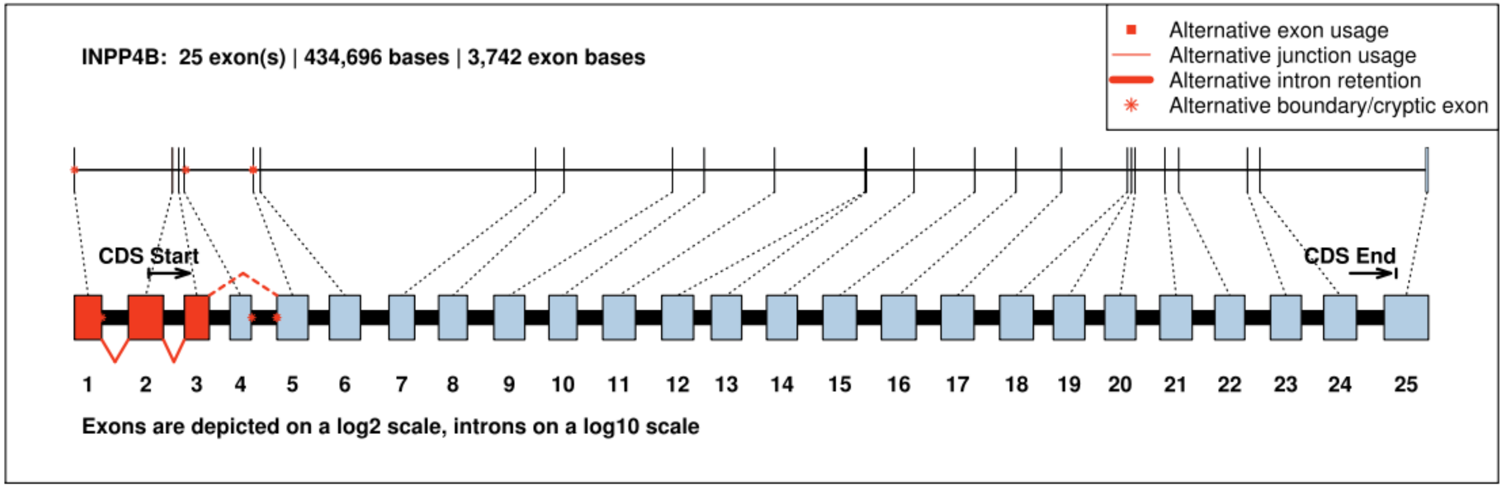


Matt Wilkerson

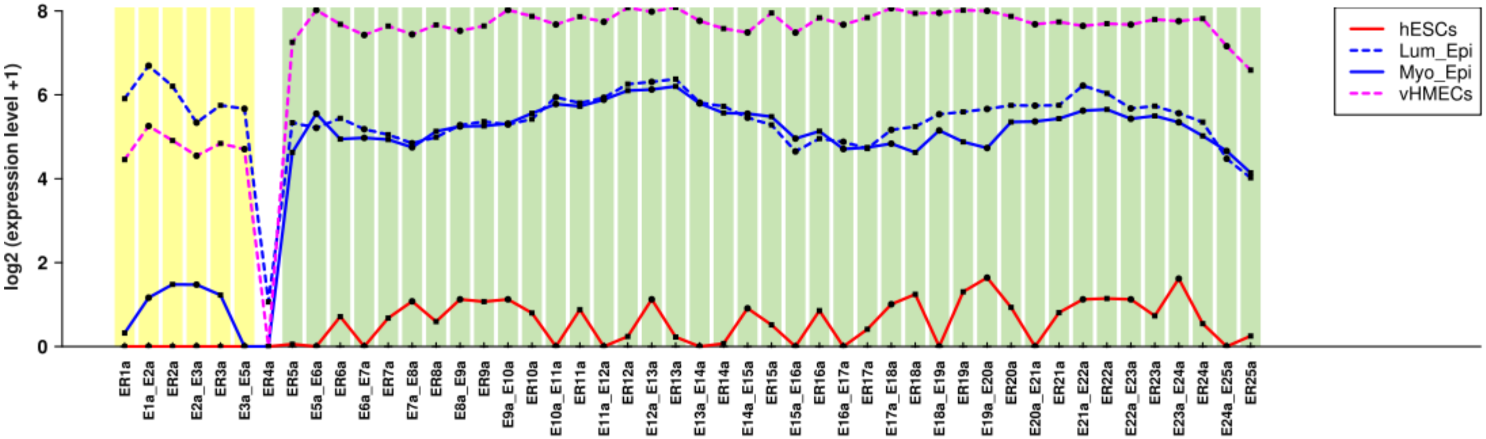
Alternative first exons of *INPP4B*

www.AlexaPlatform.org/alexa_seq/

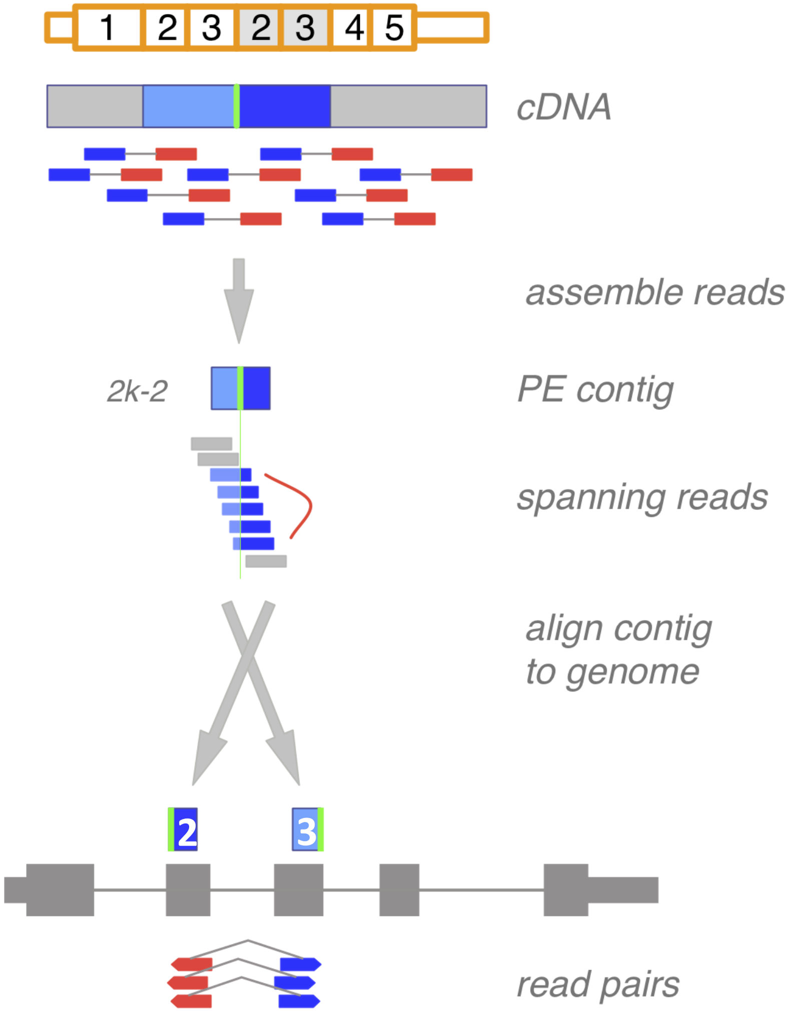
Gene model for 'INPP4B'


INPP4B: 25 exon(s) | 434,696 bases | 3,742 exon bases
■ Alternative exon usage
Alternative junction usage
Alternative intron retention
✱ Alternative boundary/cryptic exon
CDS Start
CDS End
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
Exons are depicted on a log2 scale, introns on a log10 scale

Exon and junction expression levels (all libraries)

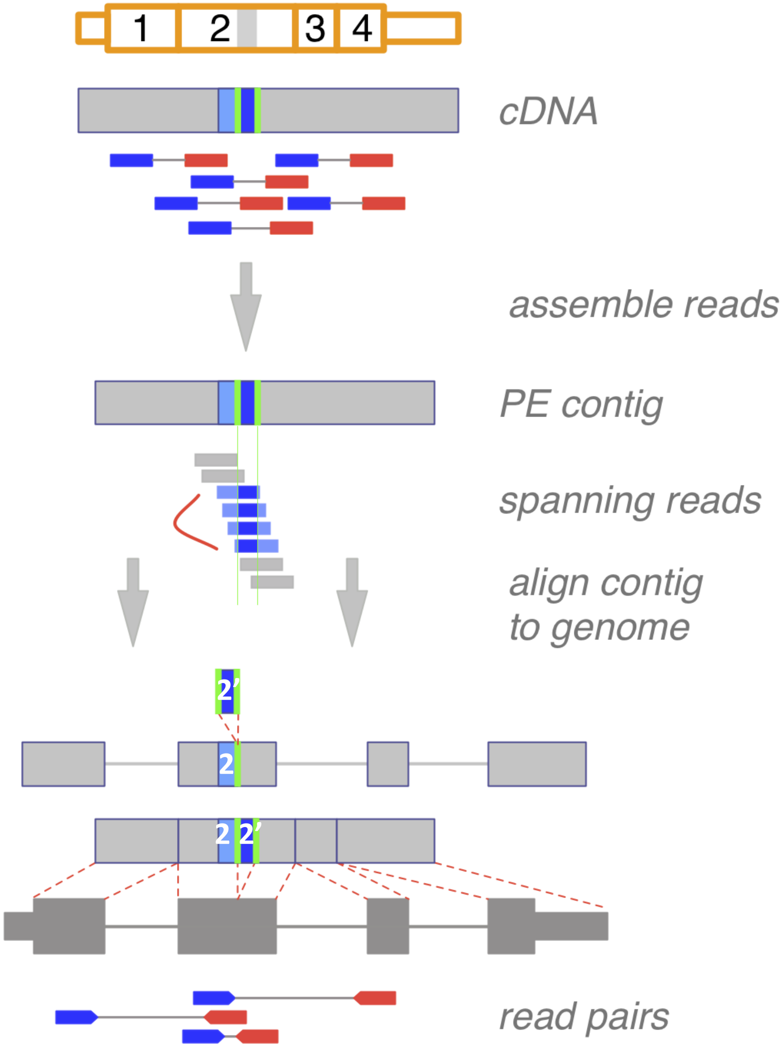


Detecting PTDs & ITDs



- Partial (gene) tandem duplications (PTDs):
  - 10/173 pts (5.8%) harbour duplication of MLL exons (2-10)
  - 181 other PTDs identified
- Internal tandem duplications (ITDs)
  - 29/173 (17%) harbour partial FLT3 exon 14 duplication
  - 6/173 (3.5%) harbour partial WT1 exon 7 duplication

Detecting PTDs & ITDs



cDNA

assemble reads

PE contig

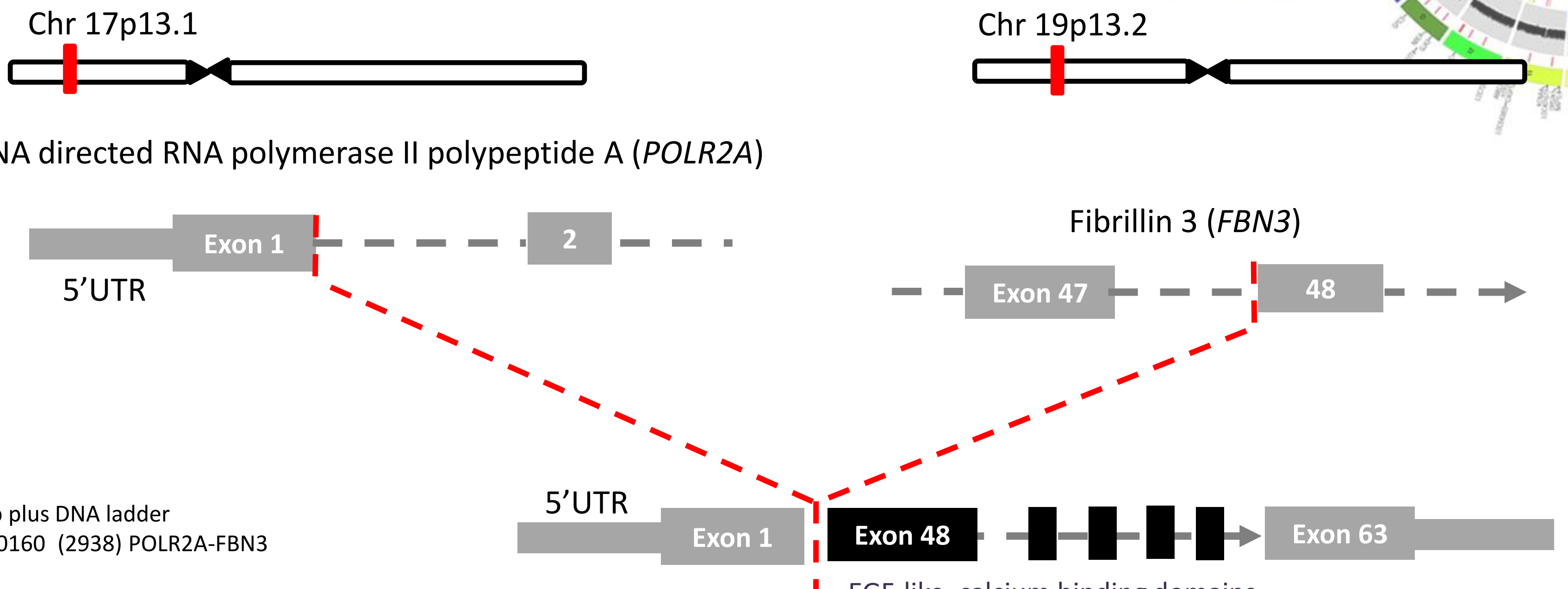
spanning reads

align contig
to genome

read pairs

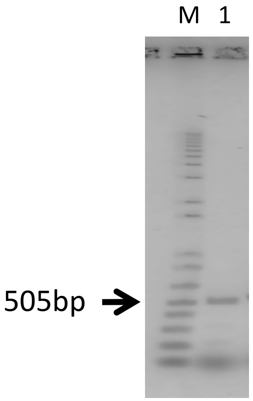
- Partial (gene) tandem duplications (PTDs):
  - 10/173 (5.8%) harbour duplication of MLL exons (2-10)
  - 181 other PTDs identified
- Internal tandem duplications (ITDs)
  - 29/173 pts (17%) harbour partial FLT3 exon 14 duplication
  - 6/173 (3.5%) harbour partial WT1 exon 7 duplication

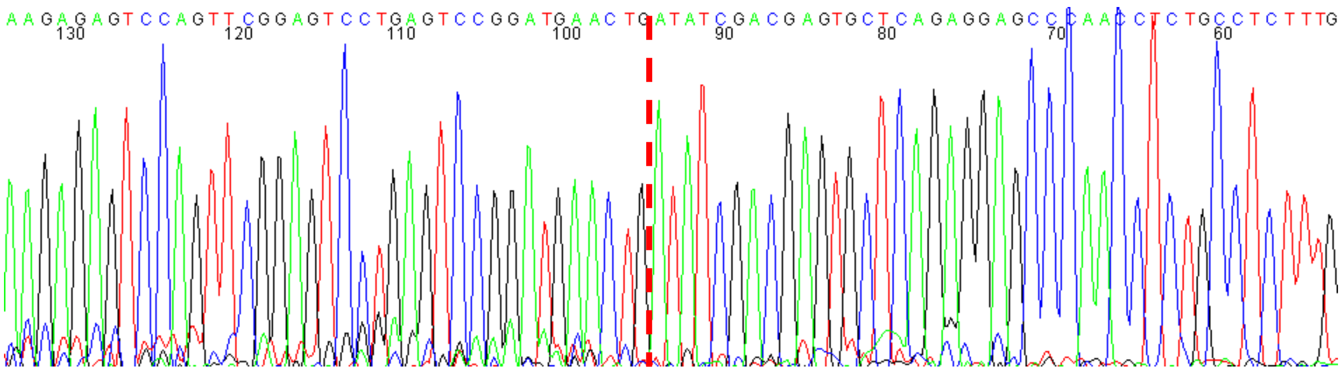
Verification of novel fusion events



EGF-like, calcium binding domains

M: 1kb plus DNA ladder
1: A00160 (2938) POLR2A-FBN3





See A. Mungall Poster