



Genome Sciences Centre

BC Cancer Agency, Vancouver, BC, Canada

Novel methods for transcript variant discovery and alternative expression analysis

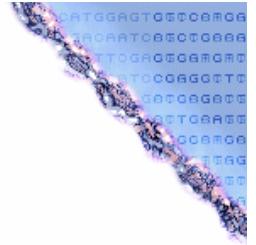
Application to the study of fluorouracil resistance in colorectal
cancer

Malachi Griffith
(Marra lab)

25 September 2009

1

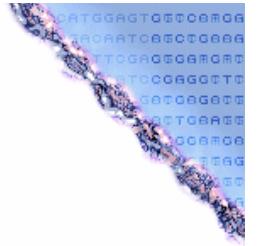
Overview



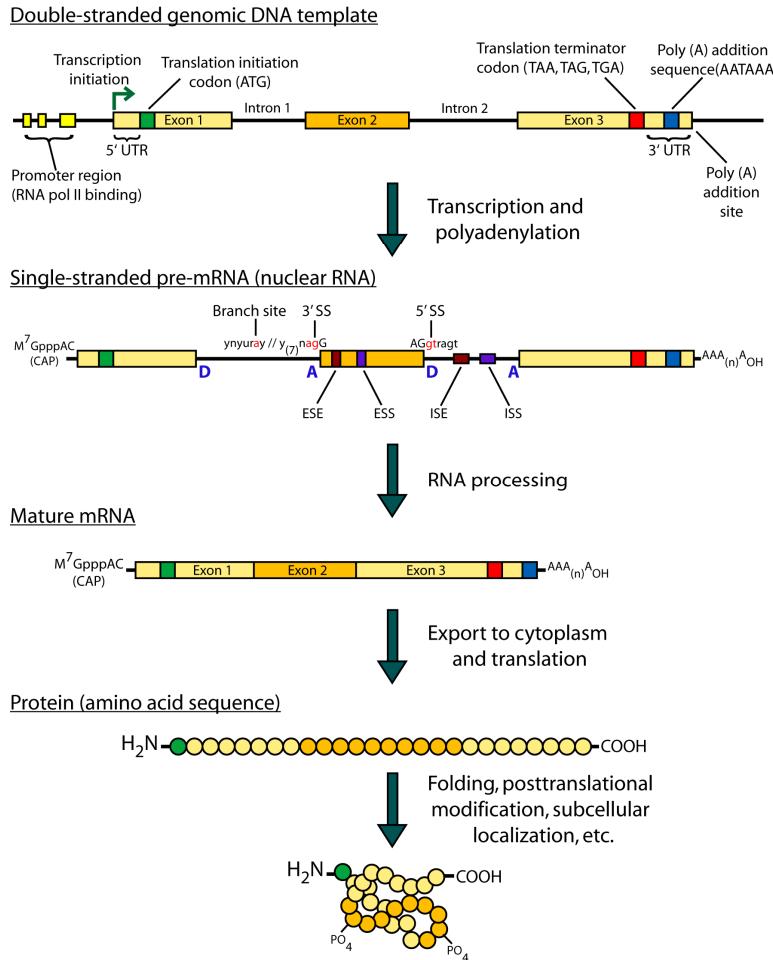
Introduction

1. Alternative expression analysis by microarrays
2. Alternative expression analysis by Illumina paired-end RNA sequencing
3. 5-FU resistance candidate gene, ‘uridine monophosphate synthetase’ (*UMPS*)

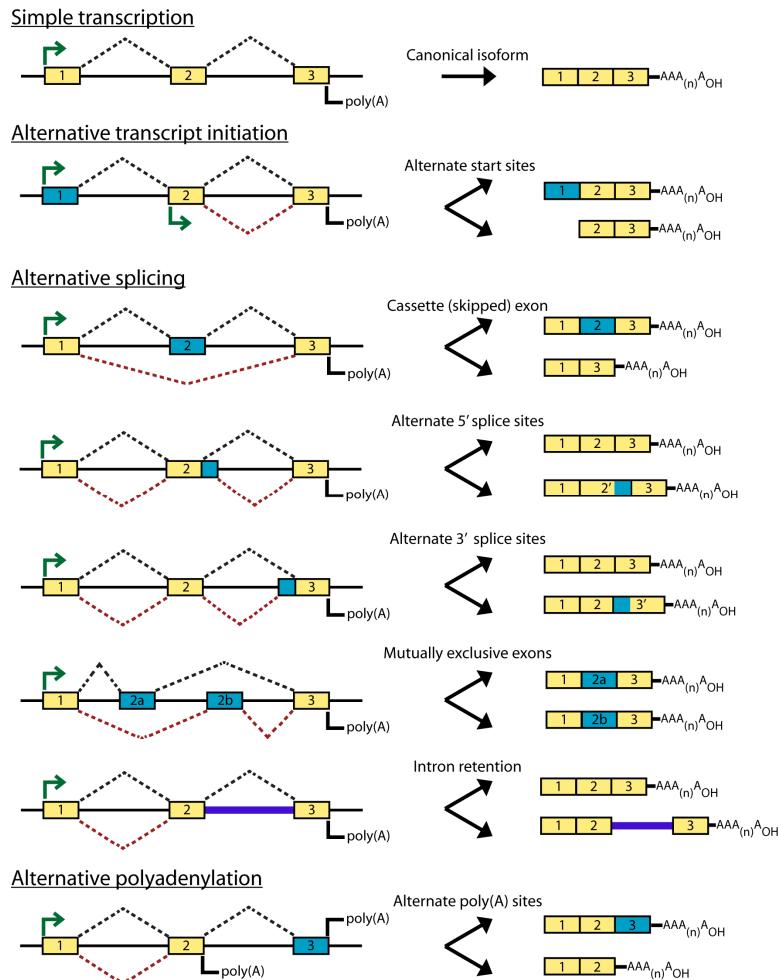
In most genes, transcript diversity is generated by alternative expression



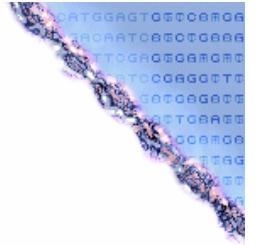
Gene expression



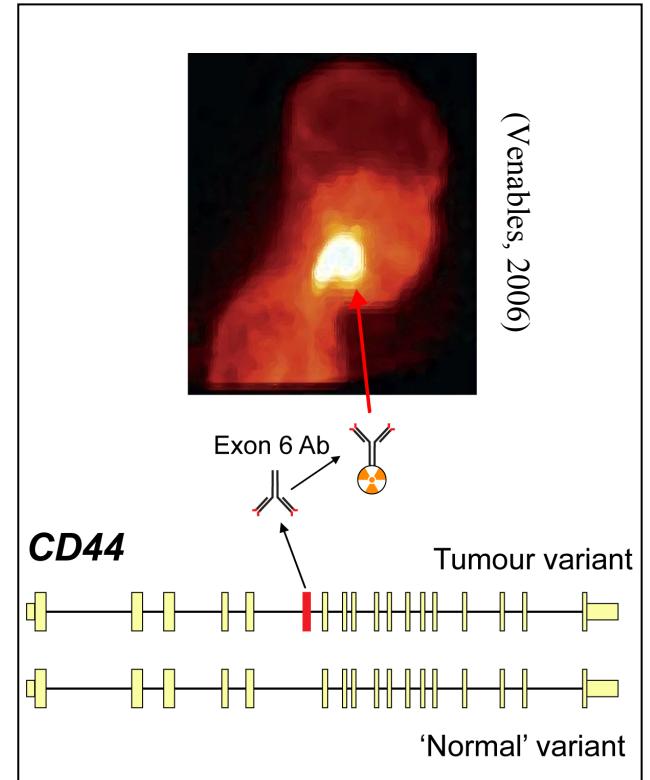
Types of alternative expression



Transcript variation is important to the study of human disease



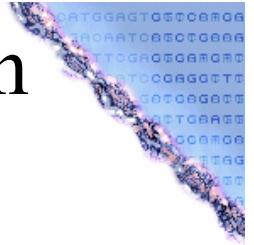
- Specific transcript variants may represent useful therapeutic targets or diagnostic markers
- **Improved methods are needed** to effectively profile transcript variants
 - Advances in microarray and massively parallel sequencing technology...
- Study 5-FU resistance in colorectal cancer



(Venables, 2006)

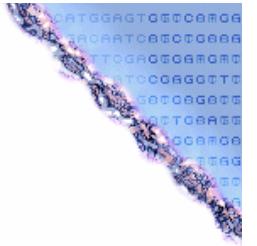


Resistance to fluorouracil treatment is a common theme of colorectal cancer treatment

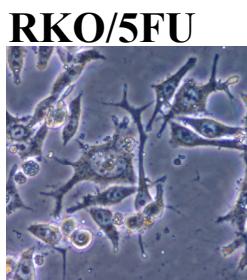
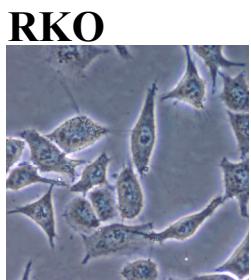
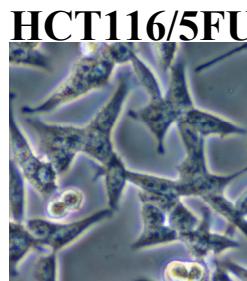
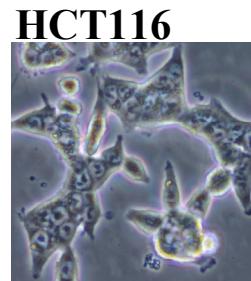
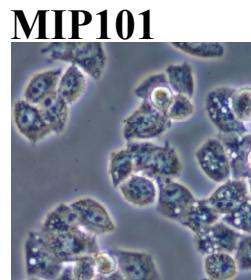


- Fluorouracil (5-FU) is a uracil analog commonly used in treatment of colorectal and other cancers
 - 5-FU & Capecitabine (oral 5-FU)
 - ~1,900 patients last year at BCCA (65% GI)
 - ~1.3 million dollars
- Response rates range from 5% to 50%
 - Intrinsic and acquired resistance...

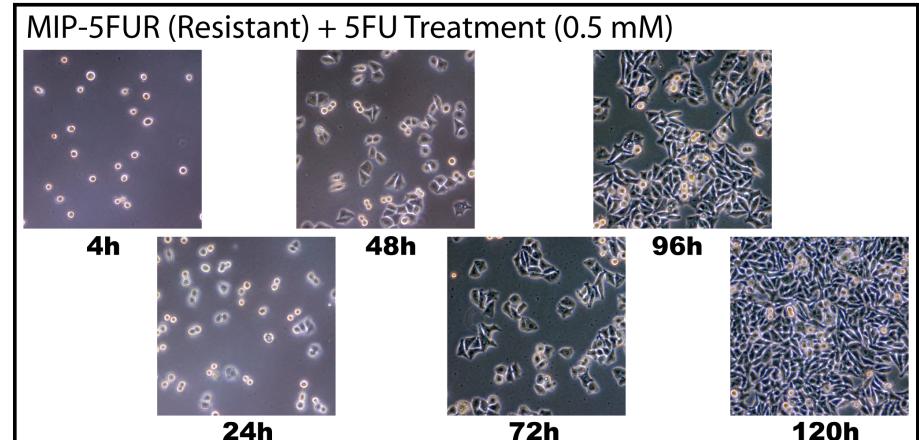
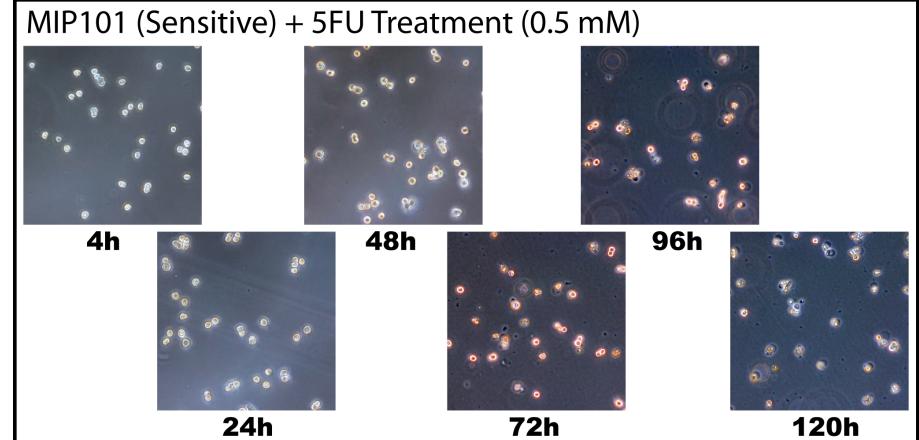
Testing ground for method development: Genome wide screen for 5-FU resistance genes



Sensitive/resistant cell lines

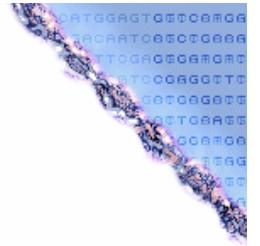


Resistance test

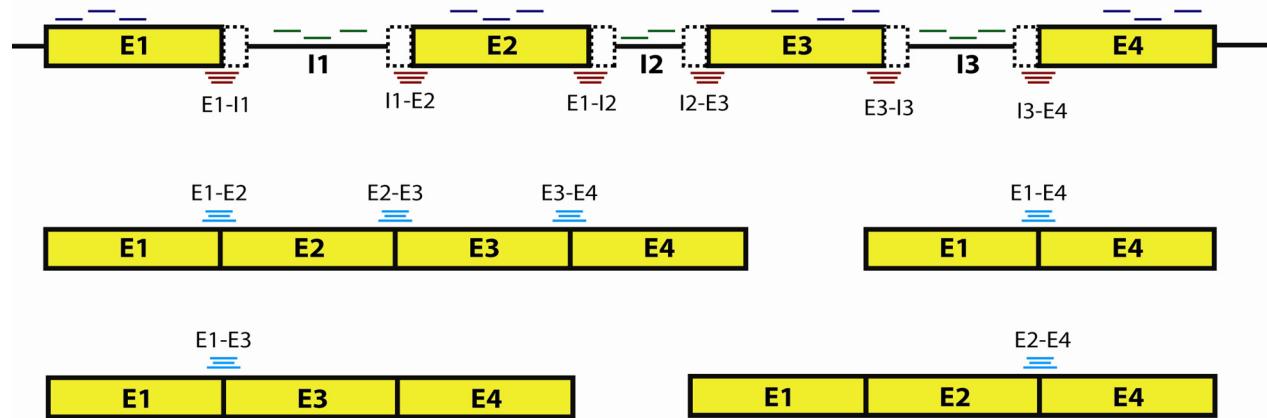


Cell lines developed in Dr. Isabella Tai's lab.

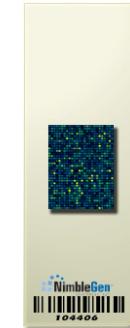
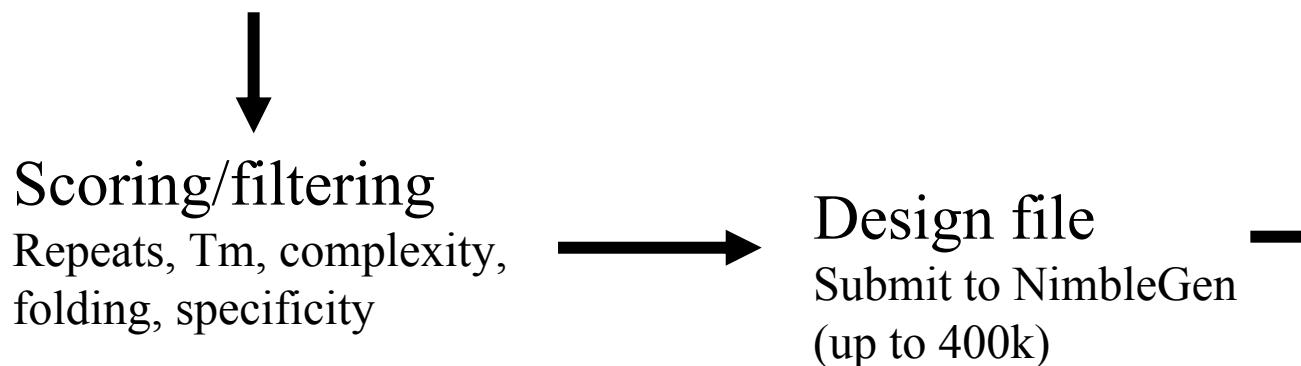
1. Alternative expression analysis by microarrays



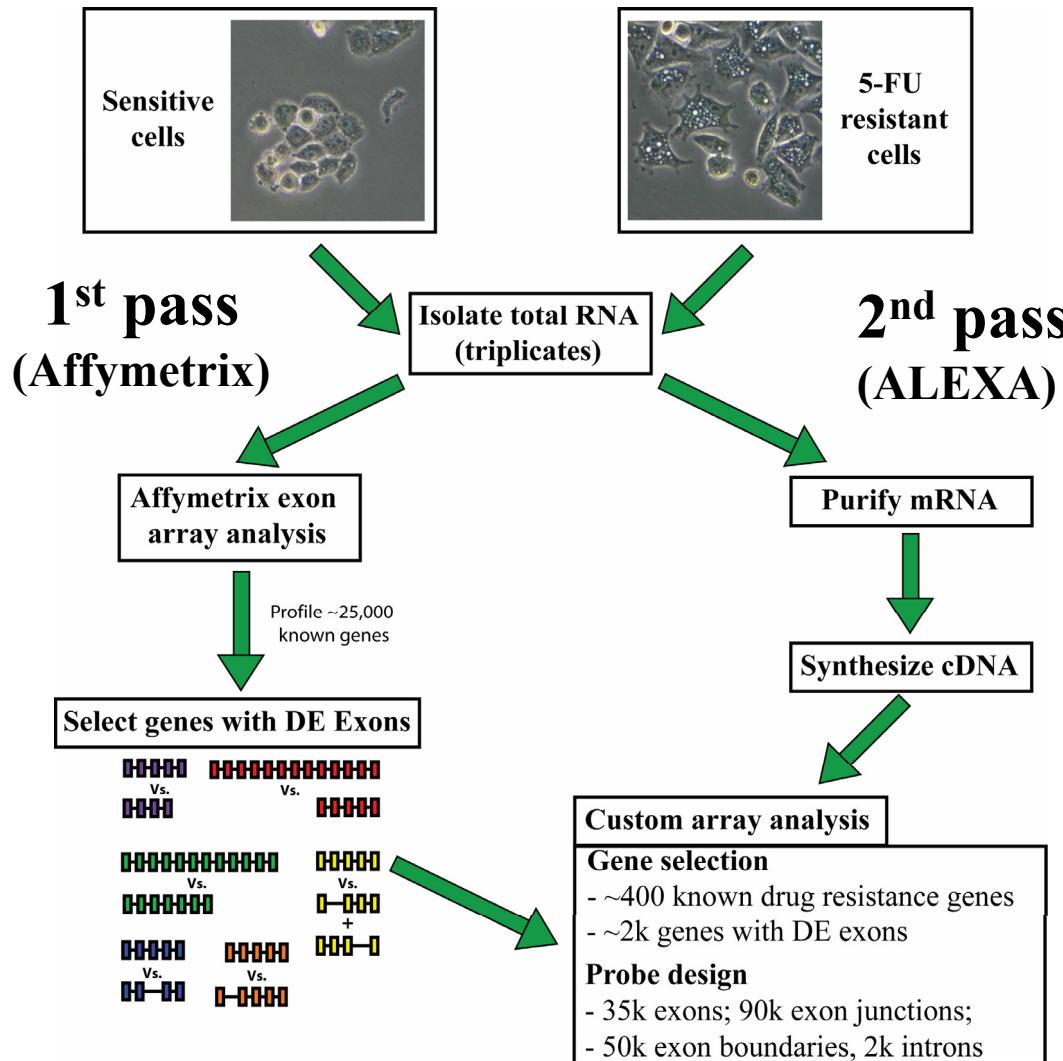
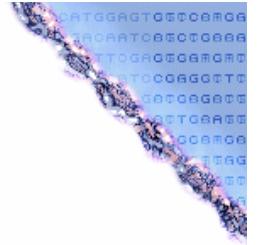
Identify oligonucleotide sequences (20-70-mers)
At least 3 million sequences for all human genes



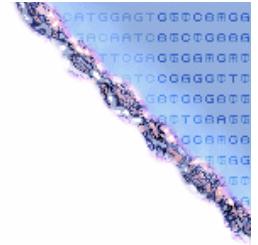
ALEXA Platform:
www.AlexaPlatform.org
Software download
Designs for 10 species



A proof-of-concept validation of the ALEXA approach



ALEXA arrays provide additional information on the structure of expressed transcripts

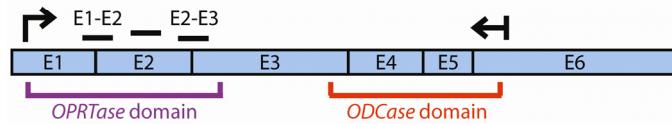


Platform	DE event type	Total events profiled	<u>Significant DE events</u>
Affymetrix	Gene-level	2,507	78
	Exon	49,681	1117
	Intron	65,327	25
	Total	117,515	1,220
ALEXA	Gene-level	2,507	233
	Exon	32,164	2,703
	Canonical junction	27,046	2,310
	Exon skip	69,761	191
	Exon boundary	52,402	253
	Intron	472	0
	Total	184,354	5690

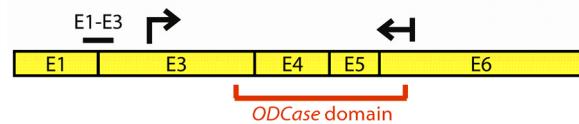
UMPS isoforms are DE in resistant cells

UMPS isoforms, protein domains, probesets

Isoform A

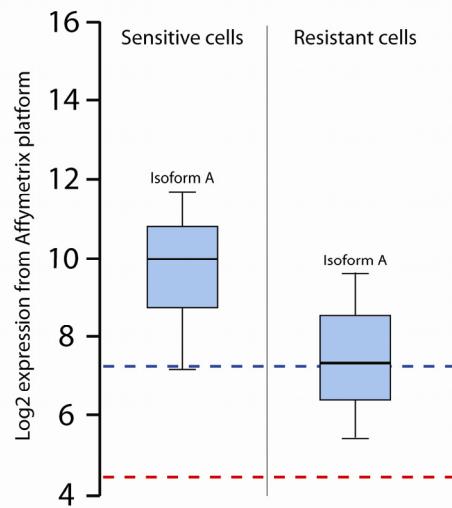


Isoform B

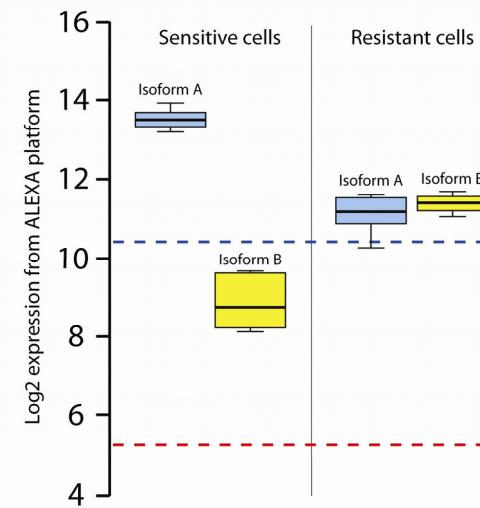


Don't forget to
describe the
effect on the
protein

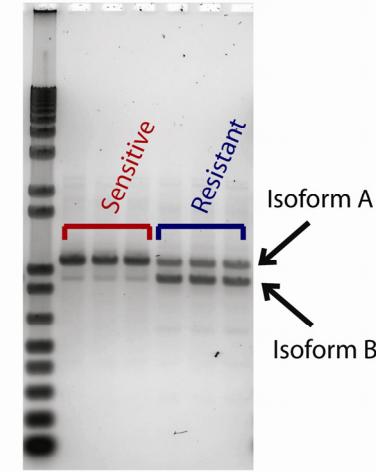
Affymetrix microarray data



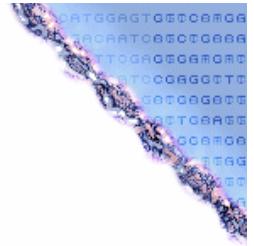
ALEXA microarray data



RT-PCR data

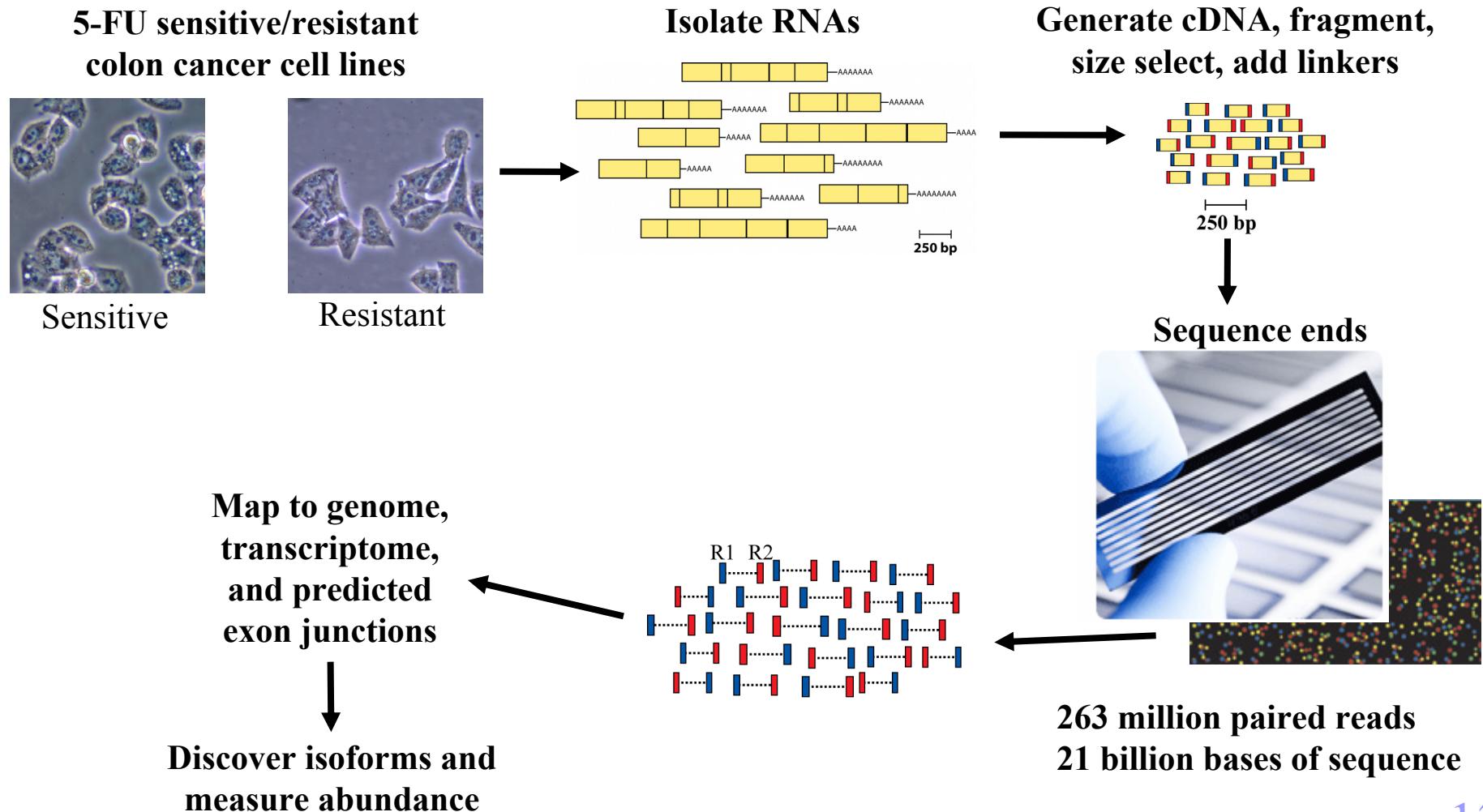
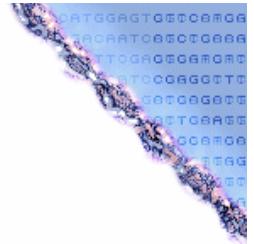


Limitations of microarray approach

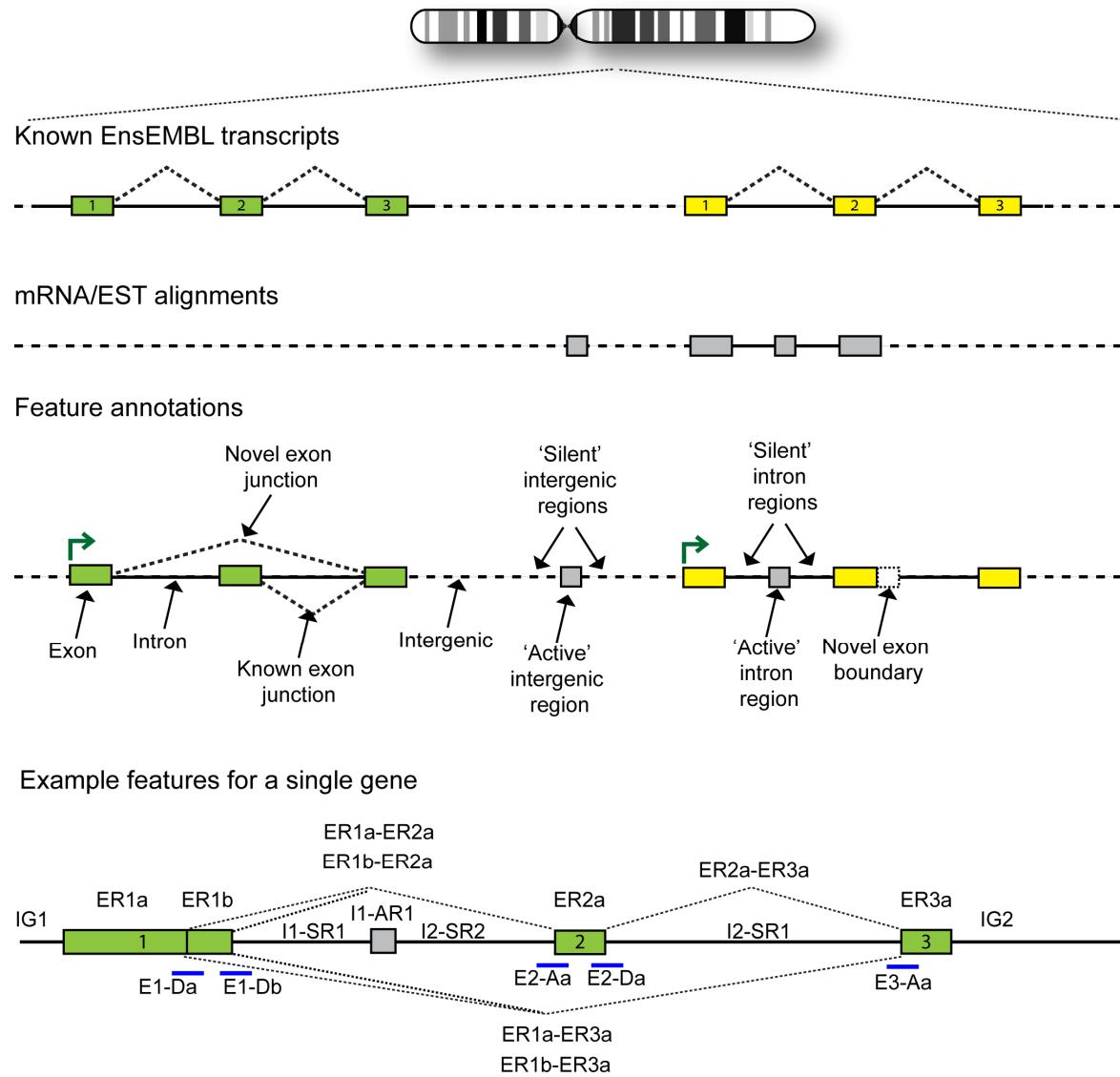


- Parameters are frozen at design time
- Limited by quality and completeness of gene annotations
- Insensitive to certain sequence features
 - Highly similar sequences
 - Small exons
- Dynamic range
- Number of array features available
- Next-generation sequencing to the rescue?

2. Alternative expression analysis by Illumina paired-end RNA sequencing



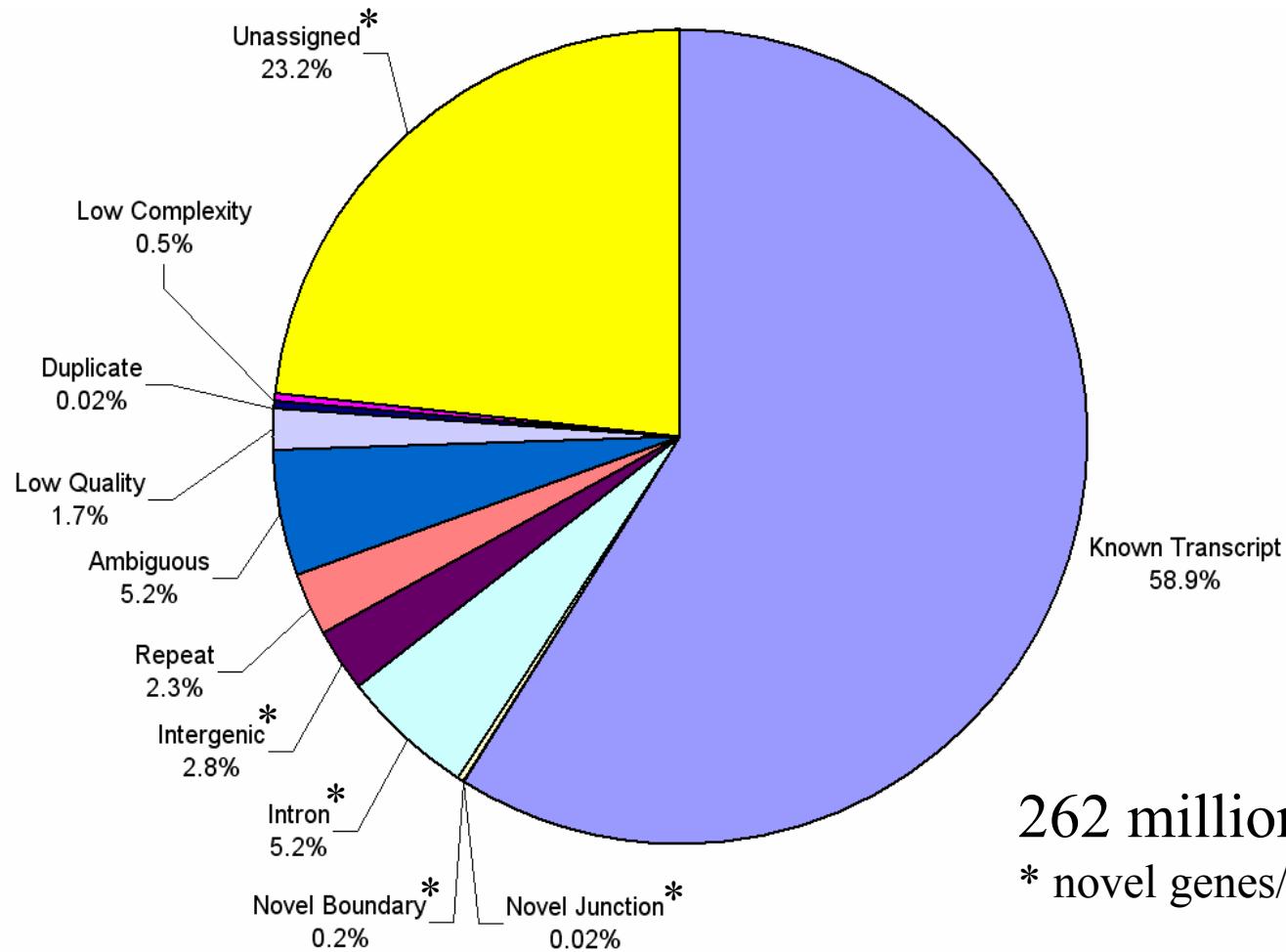
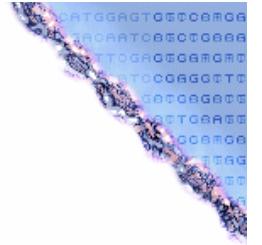
Measuring specific isoforms requires identification of distinctive ‘features’



Summary of features for human:

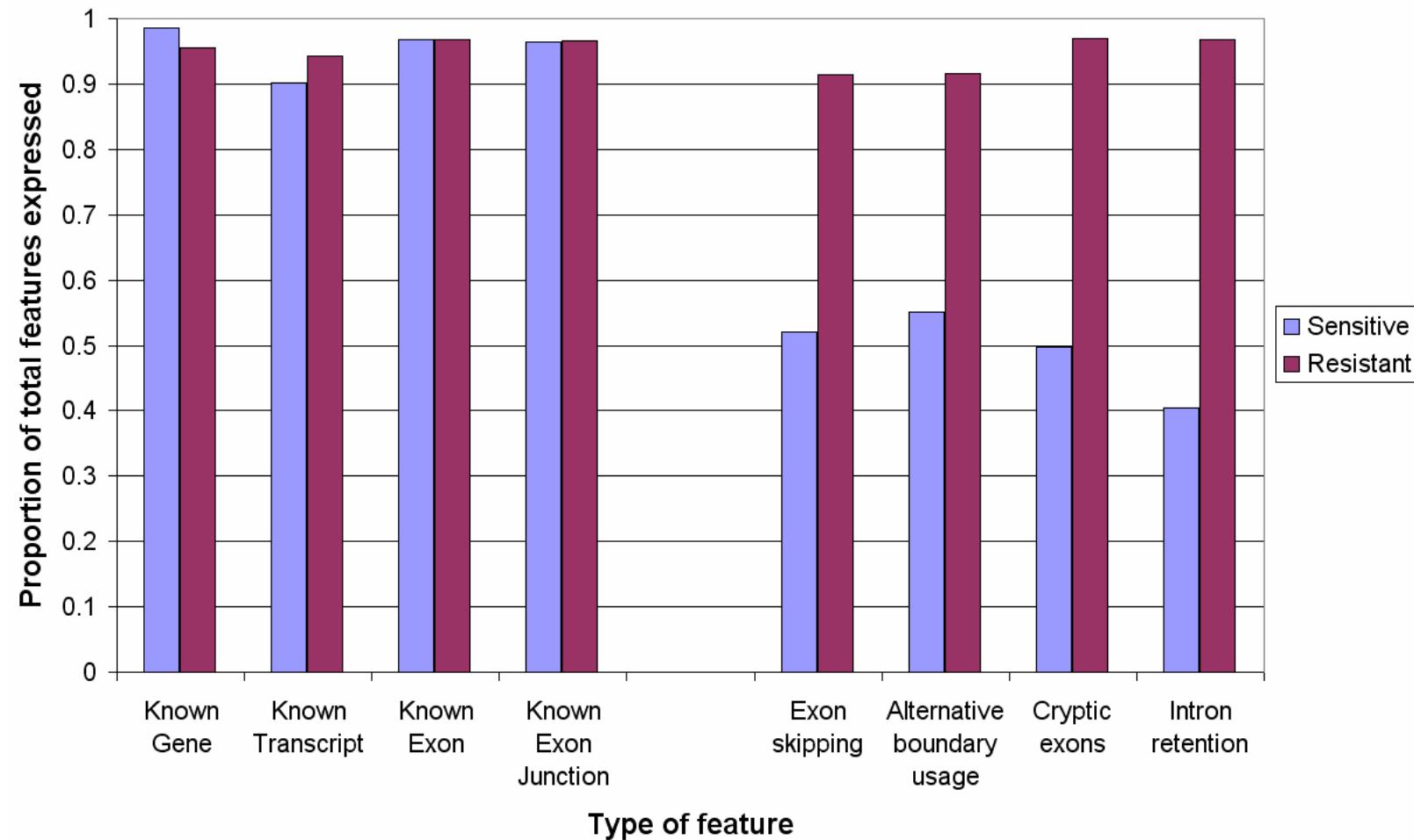
- ~4 million total (14% ‘known’)
- 37k Genes
- 62k Transcripts
- 278k exons
- 2,210k exon junctions
- 407k alternative exon boundaries
- 560k intron regions
- 227k intergenic regions

59% of reads map to known transcripts

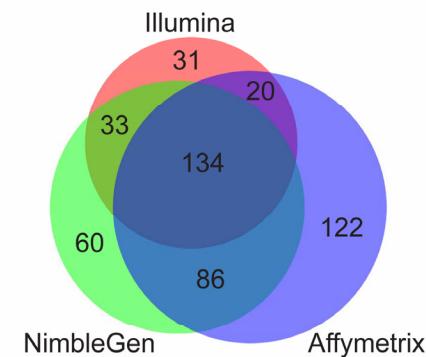
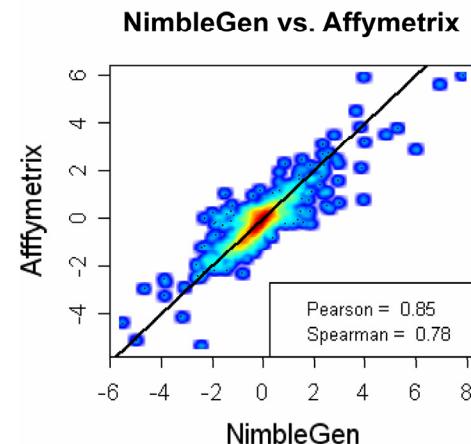
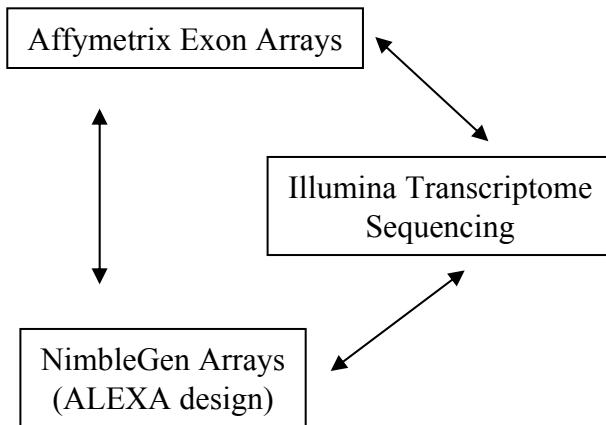
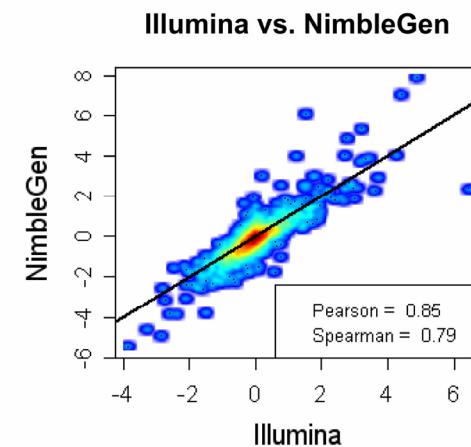
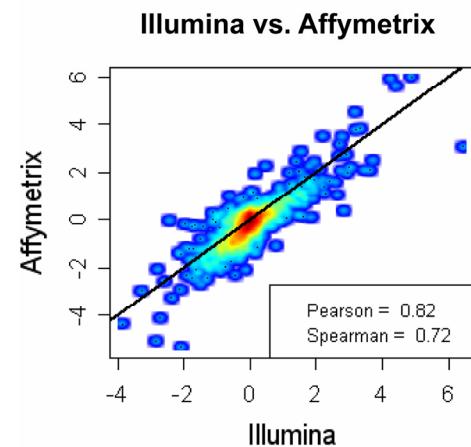
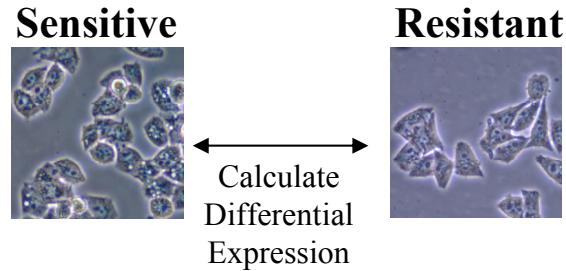
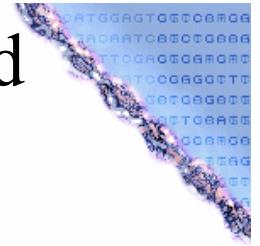


262 million paired reads
* novel genes/isoforms?

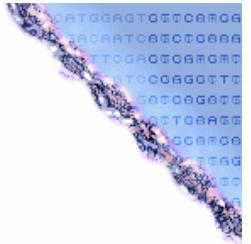
Expression of novel isoforms was increased in resistant cells



Accurate differential expression estimates were derived (comparison to Affymetrix and NimbleGen arrays)

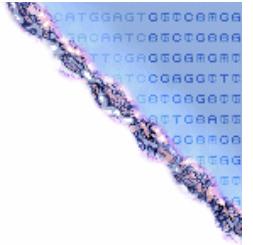


Summary of RNA sequencing



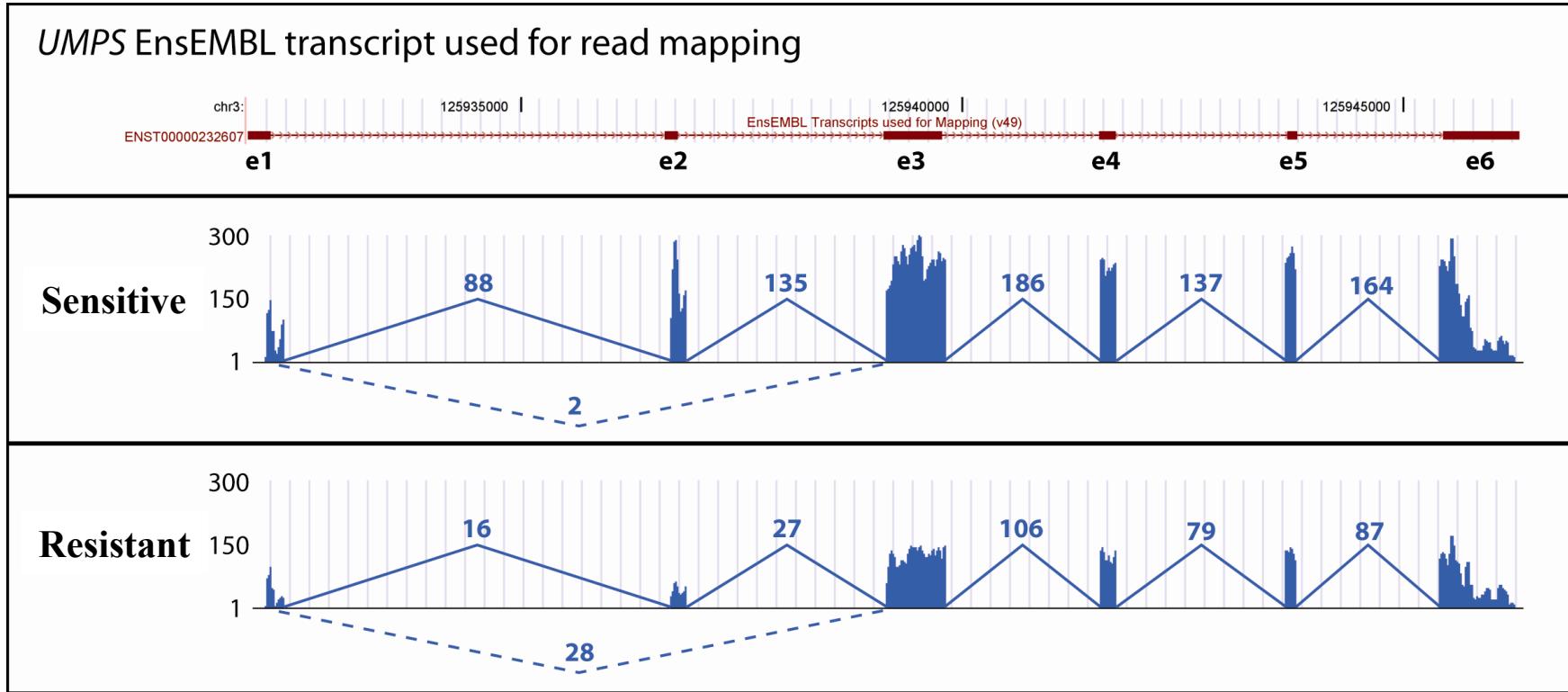
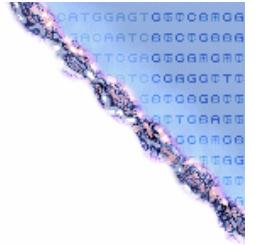
- 262 million reads (21 billion bases)
- ~25,000 genes detected
 - ~12,000 above ‘background’
 - 68% expressing multiple isoforms
- 35.6 million bases of transcriptome covered at 10x or greater sequence depth
 - 146k exons, 115k exon junctions, 3k novel junctions, etc.
- Comparison of 5-FU sensitive and resistant cells
 - 259 differentially expressed genes
 - 306 alternatively expressed genes

Specific examples



- 5-FU metabolism/action
 - *CDKN1A (P21)*, *ERCC2*, *FDXR*, *NFKB1*, *TPMT*, *TYMS*, *UCK2*, *UGT1A8*, and *UMPS*
- Drug efflux
 - *ABCC2*, *ABCC3*, *ABCC4*, *ABCG2*
- *H19*
 - 113-fold less abundant in resistant cells
- *KRT20 (CD20)*
 - 14-fold more abundant in resistant cells
- *LAMA3*
 - Exons 1-39 are reduced in abundance, while exons 40-77 are increased
- *OCIAD1*
 - 100-fold increase in expression of novel exon-skipping isoforms (E4-E6, E4-E8, E4-E9)
- Genes of unknown function
- Data for any gene can be viewed at:
 - www.AlexaPlatform.org

UMPS isoforms are DE in resistant cells

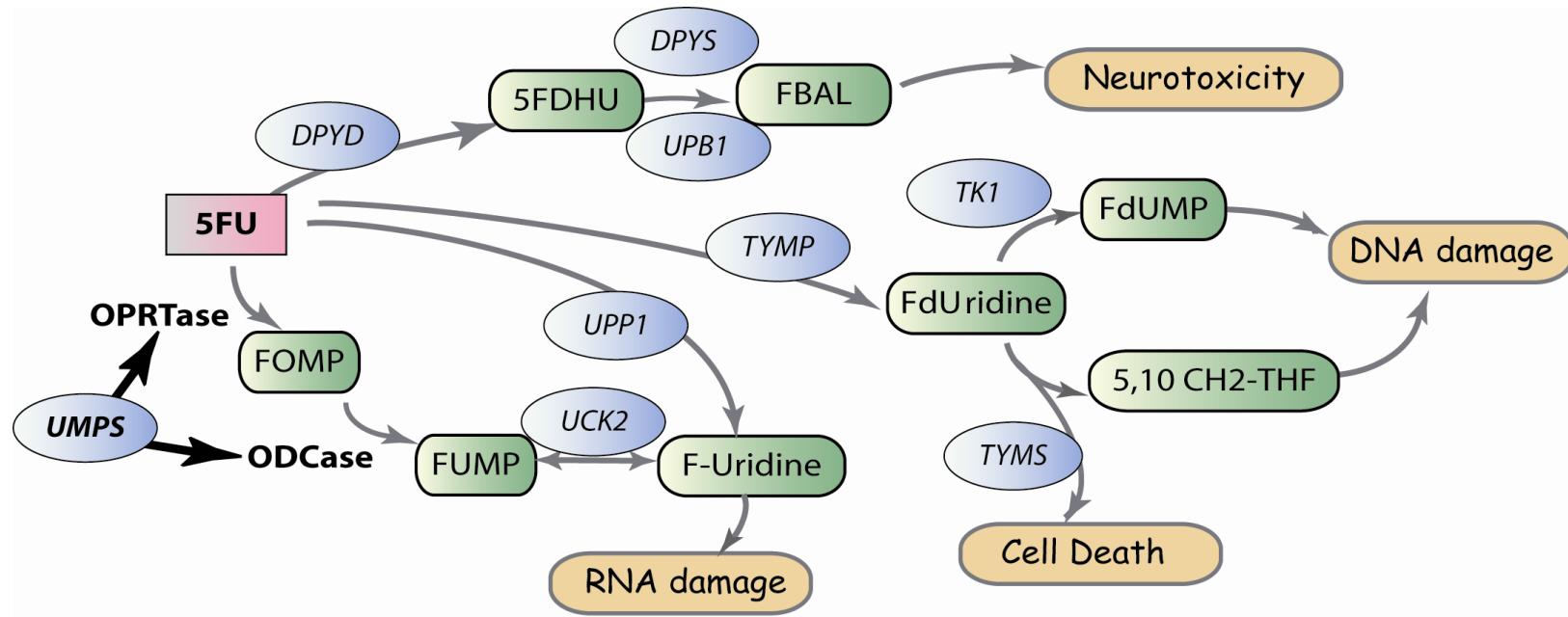
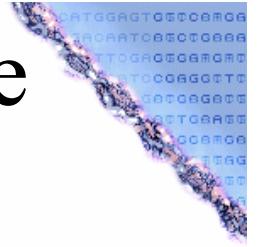


20,156 reads mapped to *UMPS*

Exon 2 skipping isoform is 15-fold more abundant

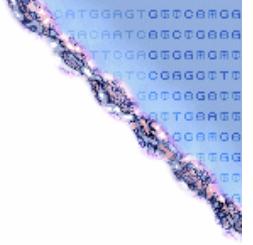
Canonical isoform is 6-fold less abundant

3. 5-FU resistance candidate gene, ‘uridine monophosphate synthetase’ (*UMPS*)



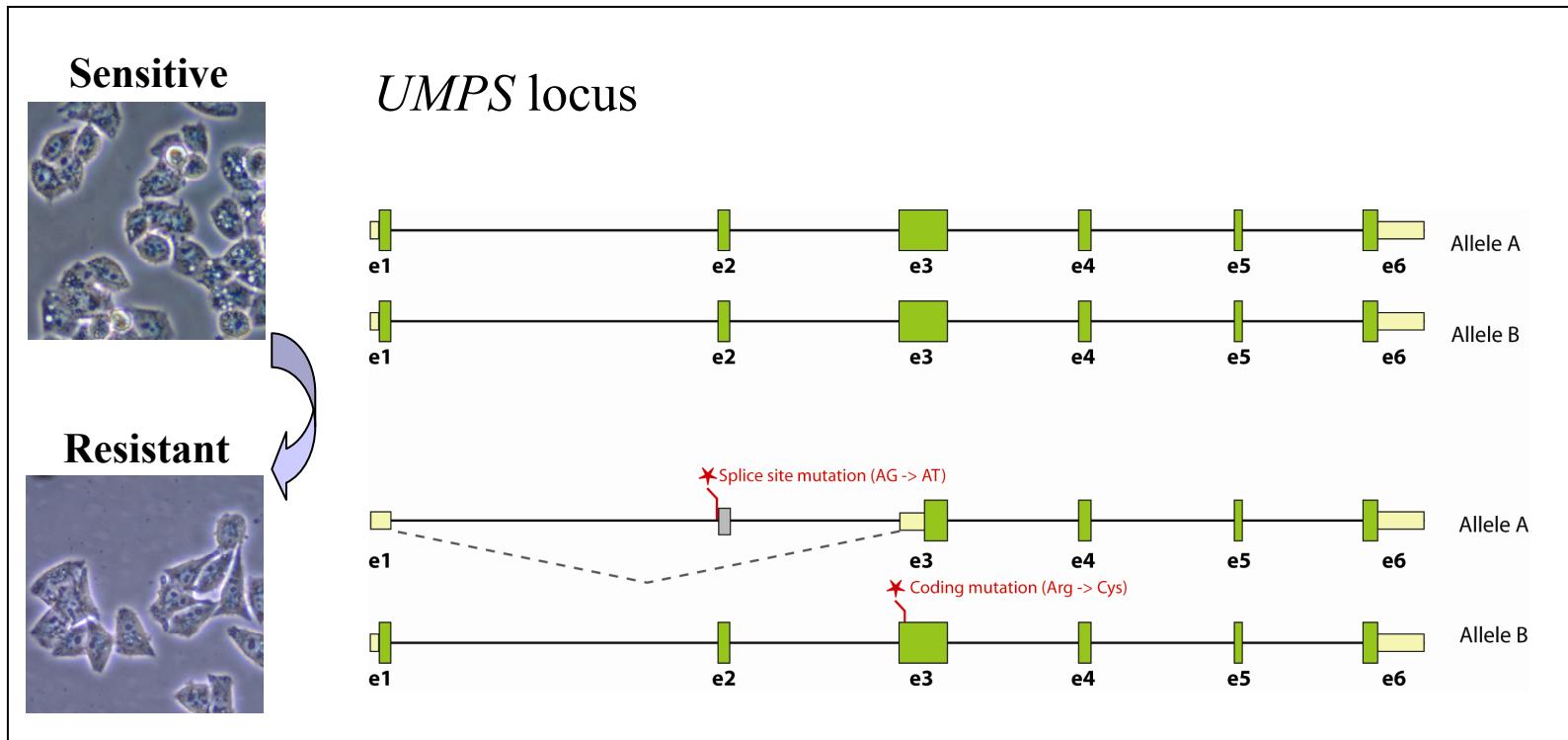
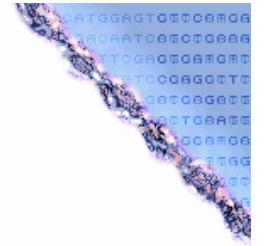
5-FU action depends on the activity of several enzymes

Hypotheses



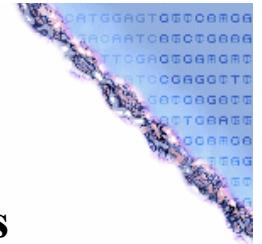
1. Increased abundance of exon 2 skipping *UMPS* isoform is caused by mutation(s) near exon 2
2. Mutation or reduced abundance of *UMPS* reduces conversion of 5-FU to active anti-tumour metabolites → leads to resistance
3. These and/or other mutations of *UMPS* are a feature of post-treatment colorectal tumours

UMPS was recurrently disrupted in 5-FU resistant cell lines

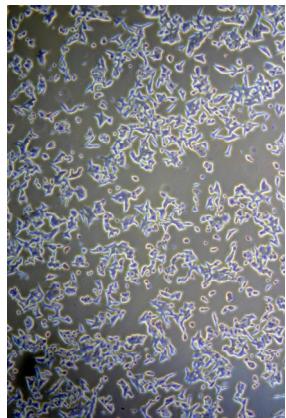


Differential expression, aberrant splicing and mutation of *UMPS* was observed in three 5-FU resistant cell lines

Colorectal cancer sample sources

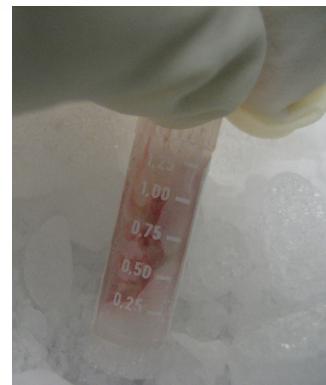


Cell lines



- Isabella Tai
- **6** cell lines
- 5-FU sensitive/resistant pairs

Fresh frozen tissue



- Ontario tumour bank
- **26** primary tumour/normal
- Treatment naïve
- Adjuvant 5-FU treatment

Archival samples

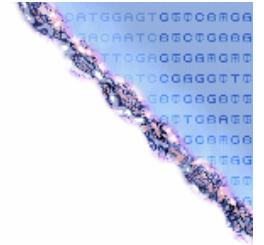


- Sharlene Gill (BCCA)
- **44** liver metastases
- 50% exposed to 5-FU

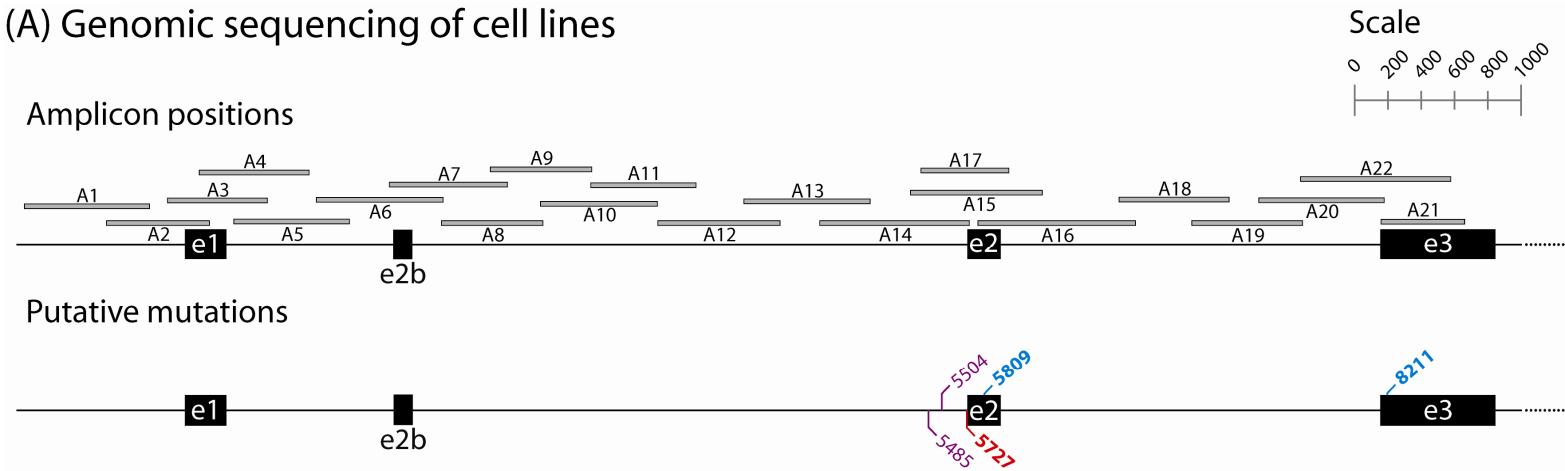
- Carl Brown (St. Paul's)
- **20** primary tumours
- Neoadjuvant 5-FU



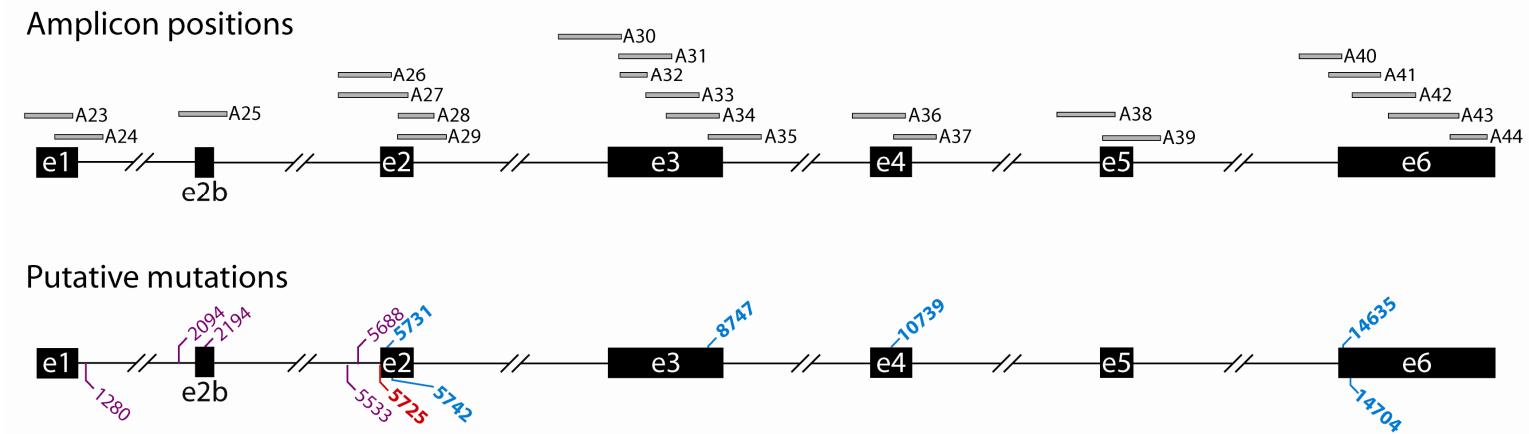
Mutation analysis (6 cell lines, 90 patients)



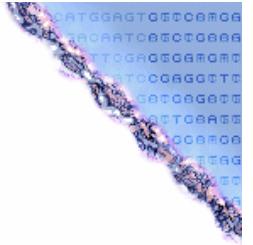
(A) Genomic sequencing of cell lines



(B) Genomic sequencing of patient samples



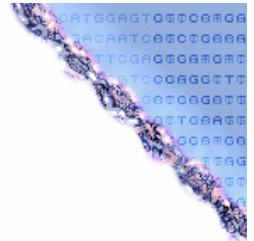
Conclusions and future work



- ALEXA microarrays
 - Effective for alternative expression analysis
 - Improvement over Affymetrix exon arrays
- RNA sequencing by Illumina
 - Overcomes some limitations of arrays
 - Improved dynamic range, signal-to-noise ratio, sensitivity and specificity
- Drug resistance
 - Mutation, differential expression, alternative expression
 - metabolism and drug efflux genes
 - genes/isoforms of unknown function
 - *UMPS*. Possible marker of 5-FU resistance?



Acknowledgements



Supervisor

Marco Marra

Committee

Joseph Connors
Stephane Flibotte
Steve Jones
Gregg Morin

Bioinformatics

Obi Griffith
Ryan Morin
Allen Delaney

Laboratory

Trevor Pugh
Tesa Severson

GI cancer

Isabella Tai
Michelle Tang

Illumina

Martin Hirst
Thomas Zeng
Yongjun Zhao
Helen McDonald

Affymetrix

Agnes Baross
Susanna Chan
Jennifer Asano
Adrian Ally

Students

Alison Lee
Jessica Paul
Lisa Miao
Pierre Cheung

BCCA

Sharlene Gill
Hagen Kennecke
Susan O'Reilly

VGH Pathology

David Owen
Margaret Luk

St. Paul's Hospital

Carl Brown
Aleksandra Kuzmanovic
Sue Finley
Douglas Filipenko

BCCA Pharmacy

Riva Pickering

Ontario Tumour Bank

Vanessa Ballin
Jeffrey Yung
Christina Rodriguez
Victoria Shelep

Patents

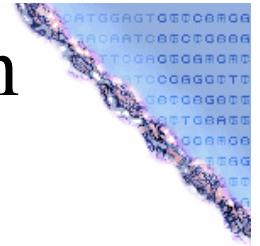
Daniel Polonenko
Erin Pisko

Technology Development

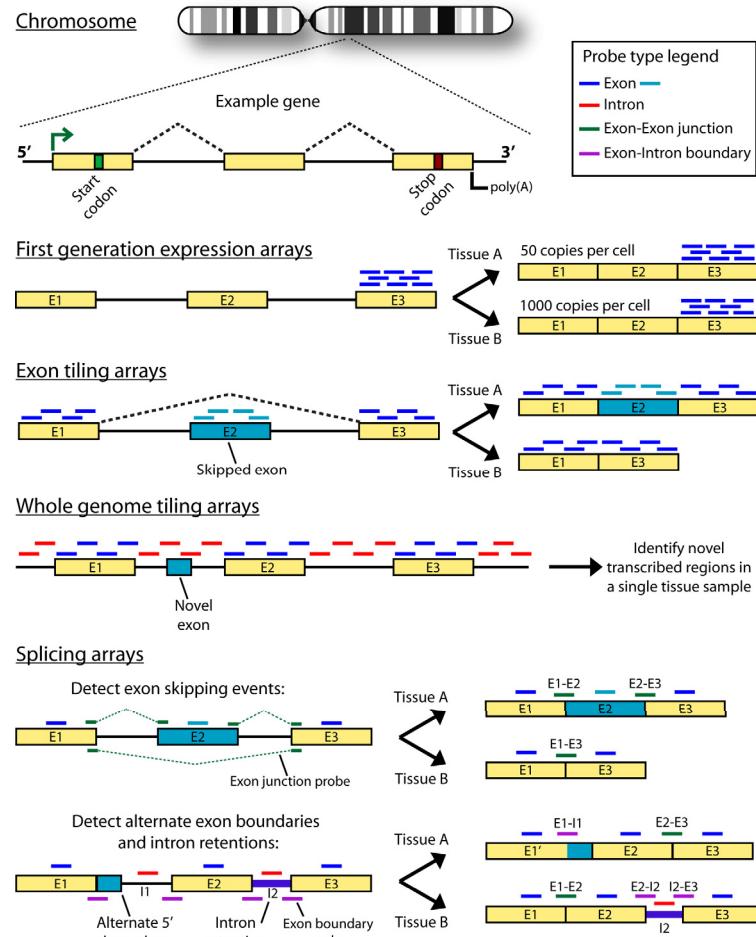
Patrick Rebstein
Kimberly Richmond



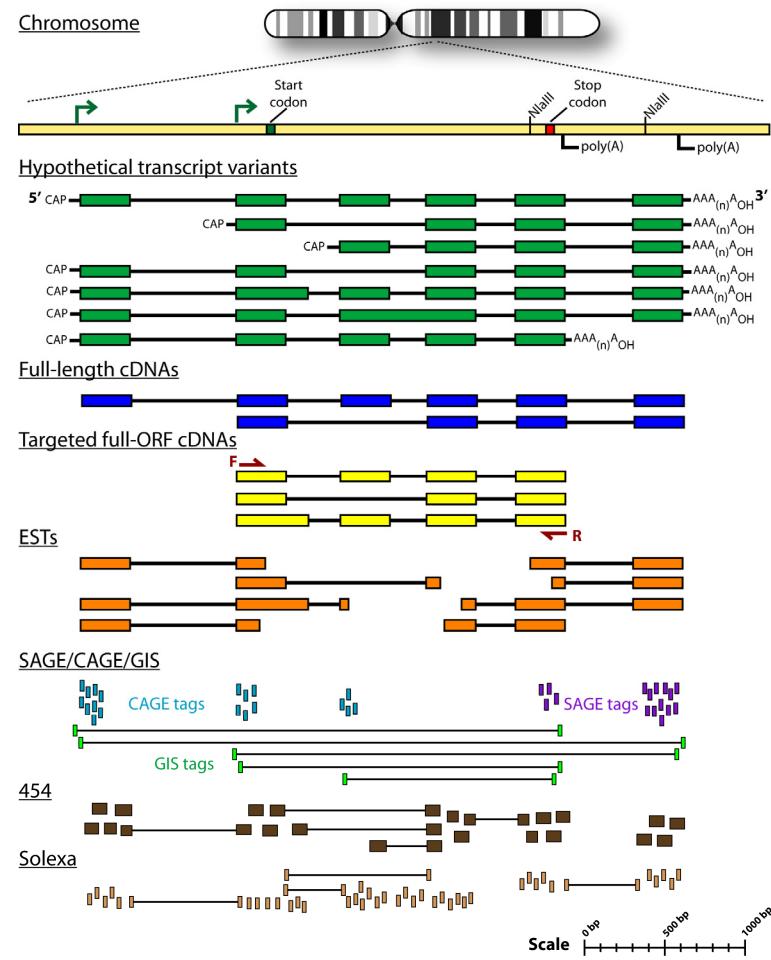
Microarray and sequencing platforms have been modified for alternative expression analysis



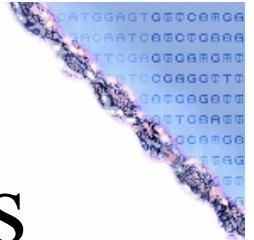
Array based methods



Sequence based methods

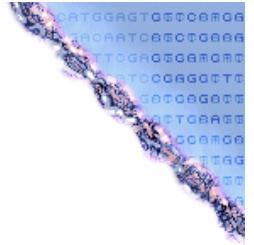


Comparison of UMPS Isoform A/B measurements from different platforms



Platform	A/B Ratio (MIP101)	A/B Ratio (MIP/5FU)
ALEXA	25.62 ± 0.56	0.85 ± 0.13
Semi-quantitative PCR	27.36 ± 9.94	0.24 ± 0.01
Real-time qPCR	22.97 ± 2.64	1.03 ± 1.56
Illumina	51.4	0.80

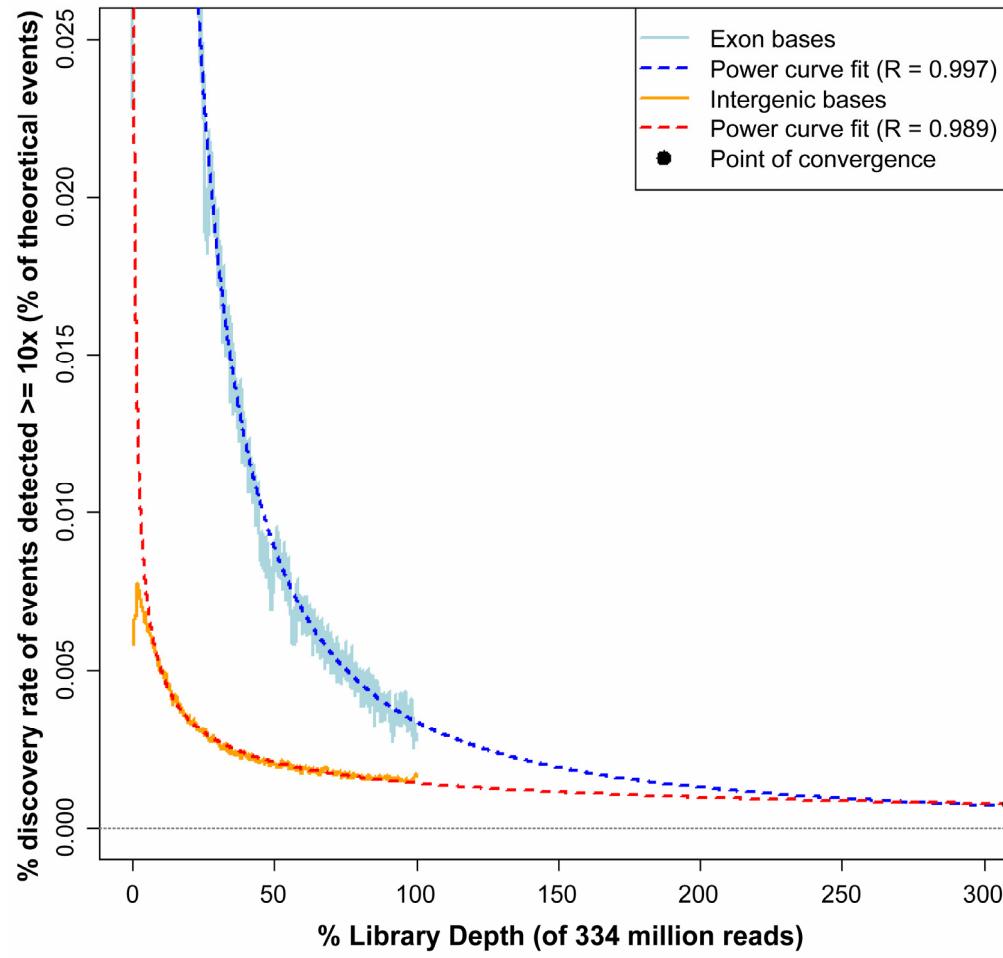
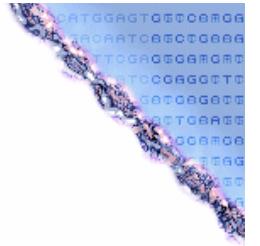
Affymetrix vs. ALEXA vs. RNA-Seq



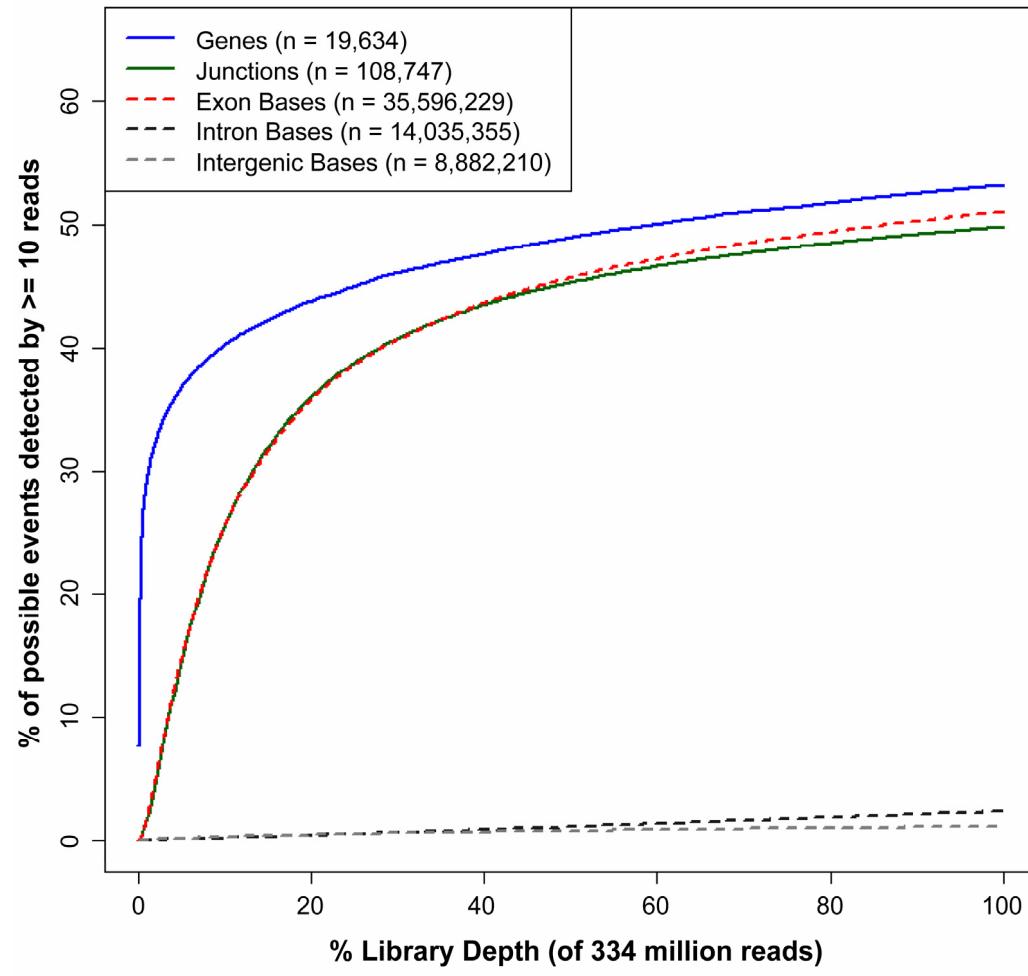
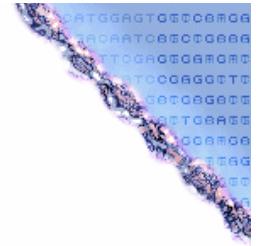
	Affymetrix exon Arrays	NimbleGen ALEXA arrays	Illumina WTSS
Theoretical dynamic range (log2 scale)	16	16	Unlimited
Observed dynamic range (log2 scale)	8.5	9.2	30.1
Signal-to-Noise Ratio	20.8 ± 0.42	56.5 ± 2.5	381.1 ± 44.7
Specificity	86.5%	95.8%	99.0%
Sensitivity	83.5%	86.9%	92.6%

Sensitivity/specificity based on analysis of exons and introns of 100 housekeeping genes

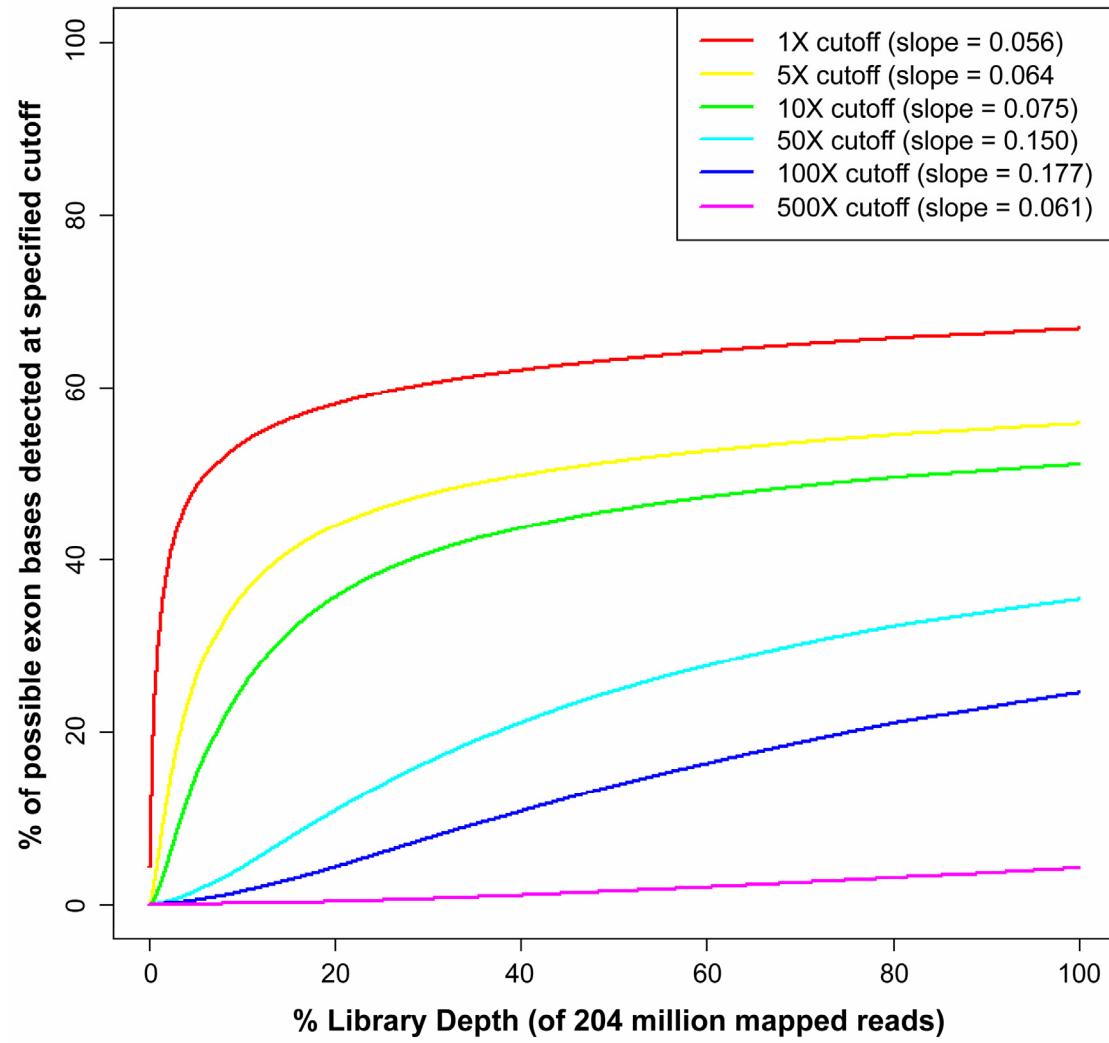
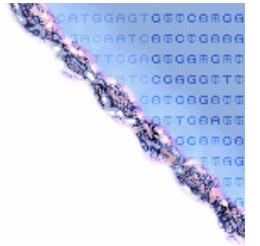
Estimation of transcriptome sequence saturation



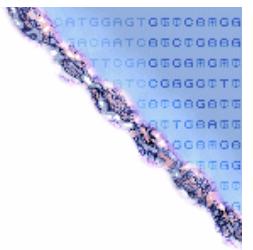
Detection versus depth



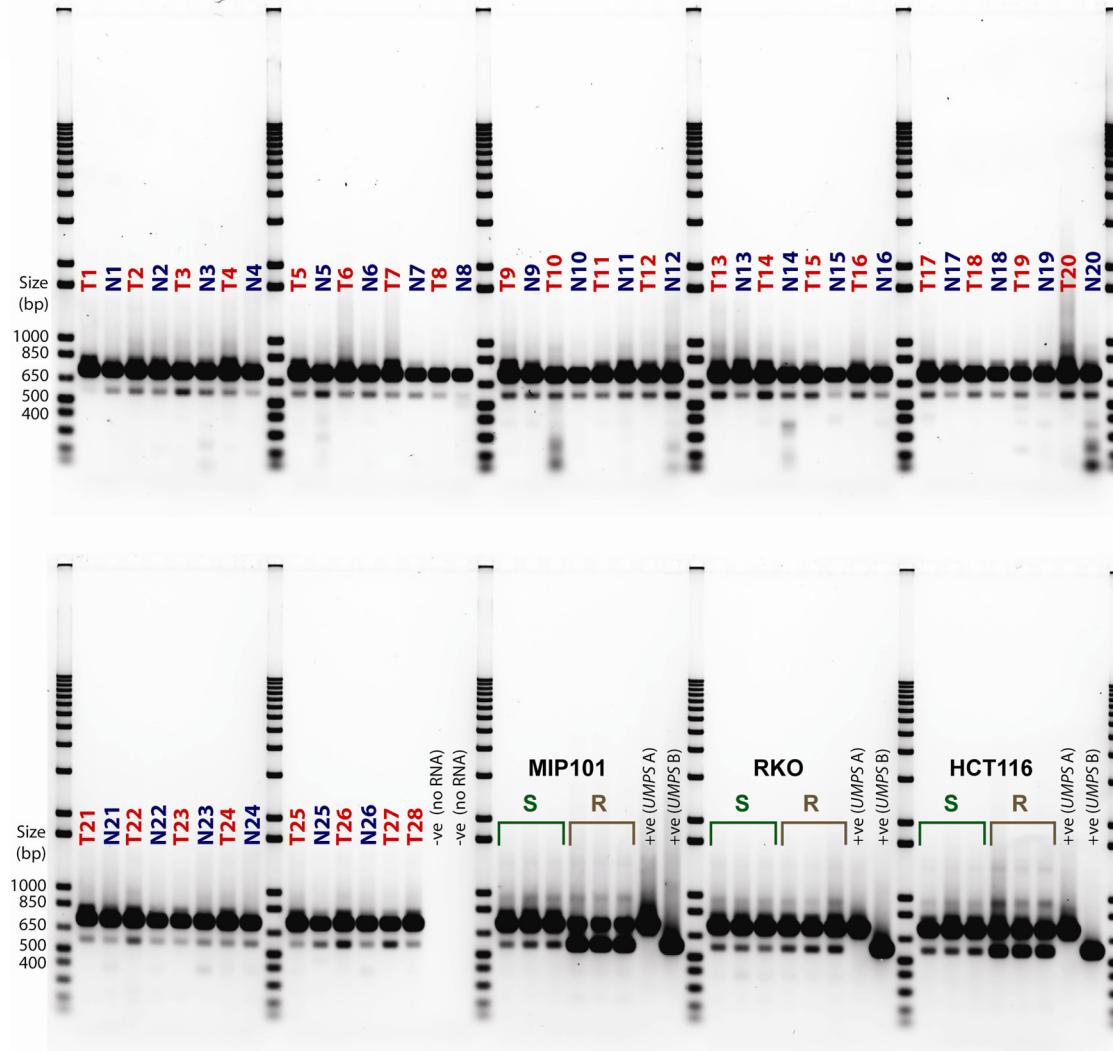
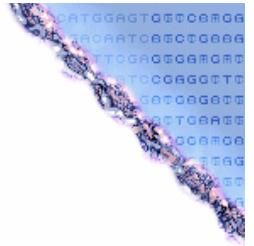
Detection versus depth versus cutoff



ALEXA website traffic

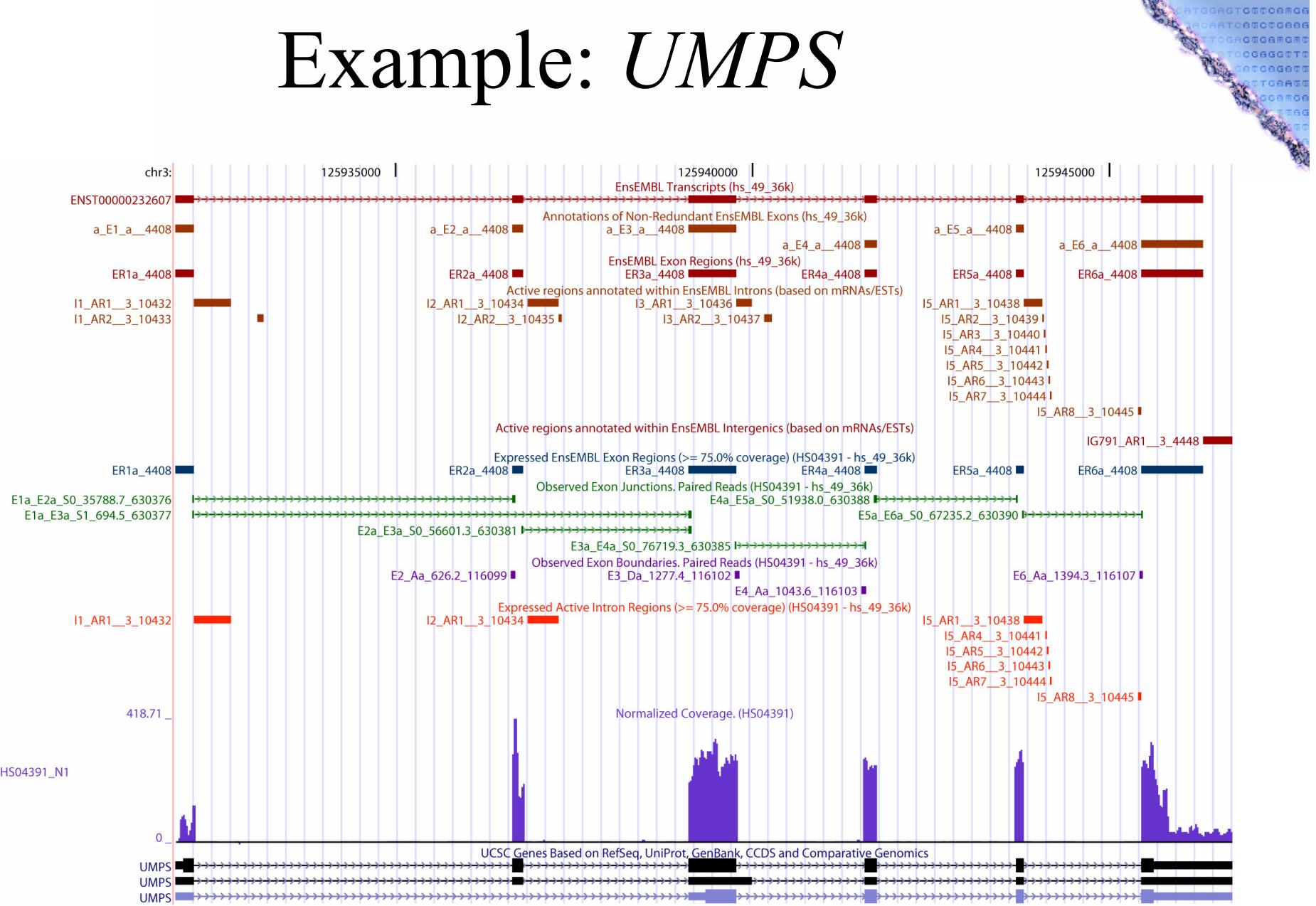


UMPS expression in treatment naïve samples

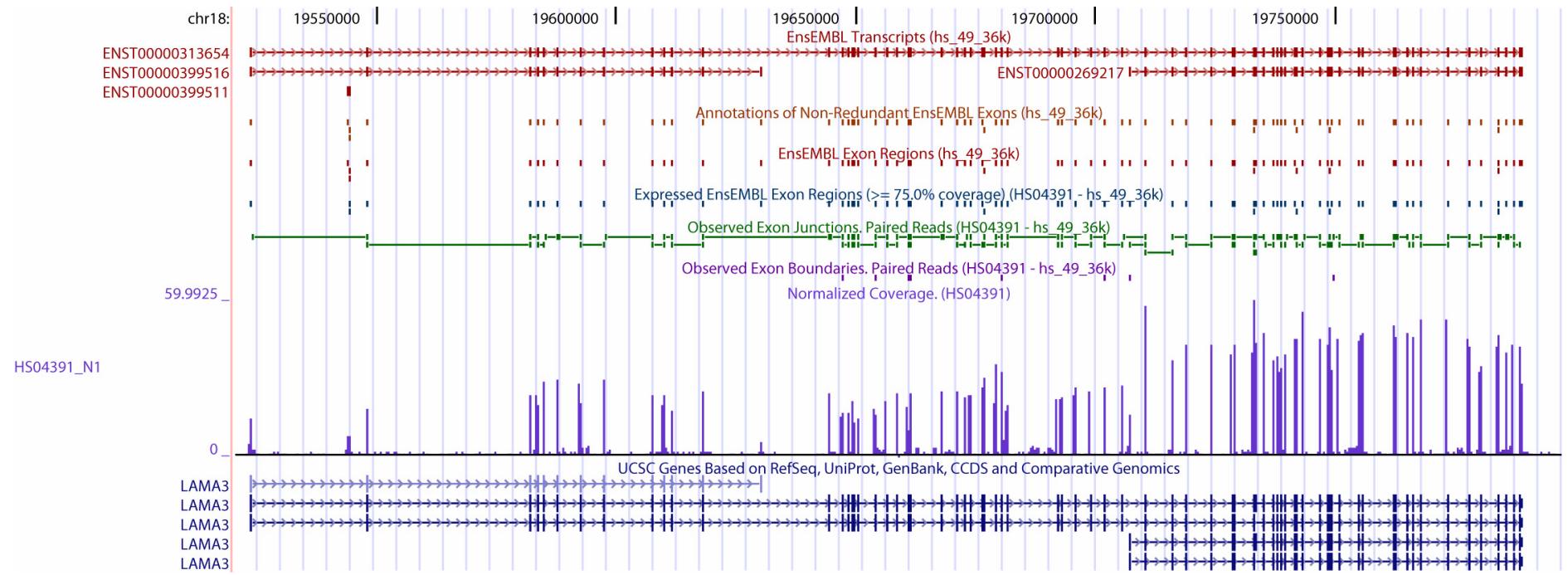
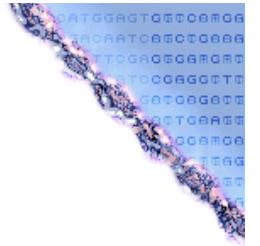


Expected product sizes: 720 bp (UMPS Isoform A) and 566 bp (UMPS Isoform B)
S = Replicates of a 5-FU sensitive cell line
R = Replicates of a 5-FU resistant cell line

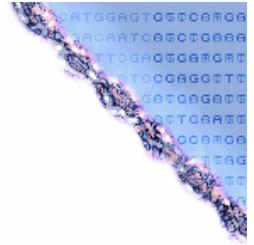
Example: *UMPS*



Example: *lama3*



Colorectal cancer sample sources



- Ontario Tumour Bank (OTB) – *sample isolation complete*
 - 26 primary tumour/normal pairs
 - Mostly treatment naïve
 - Fresh frozen tissue
 - Clinical and follow-up data supplied by OTB
- Dr. Sharlene Gill (BCCA/VGH) - *sample isolation complete*
 - 50 liver metastases of colorectal cancer
 - Varying degrees of 5-FU exposure
 - FFPE sections
 - Clinical and follow-up data extracted from Cancer Agency Info System
- Dr. Carl Brown (St. Paul's Hospital) - *pending*
 - 30 primary tumours of colorectal cancer
 - Neo-adjuvant chemotherapy + radiation
 - FFPE sections
 - Clinical and follow-up data extracted from Cancer Agency Info System