Welcome!

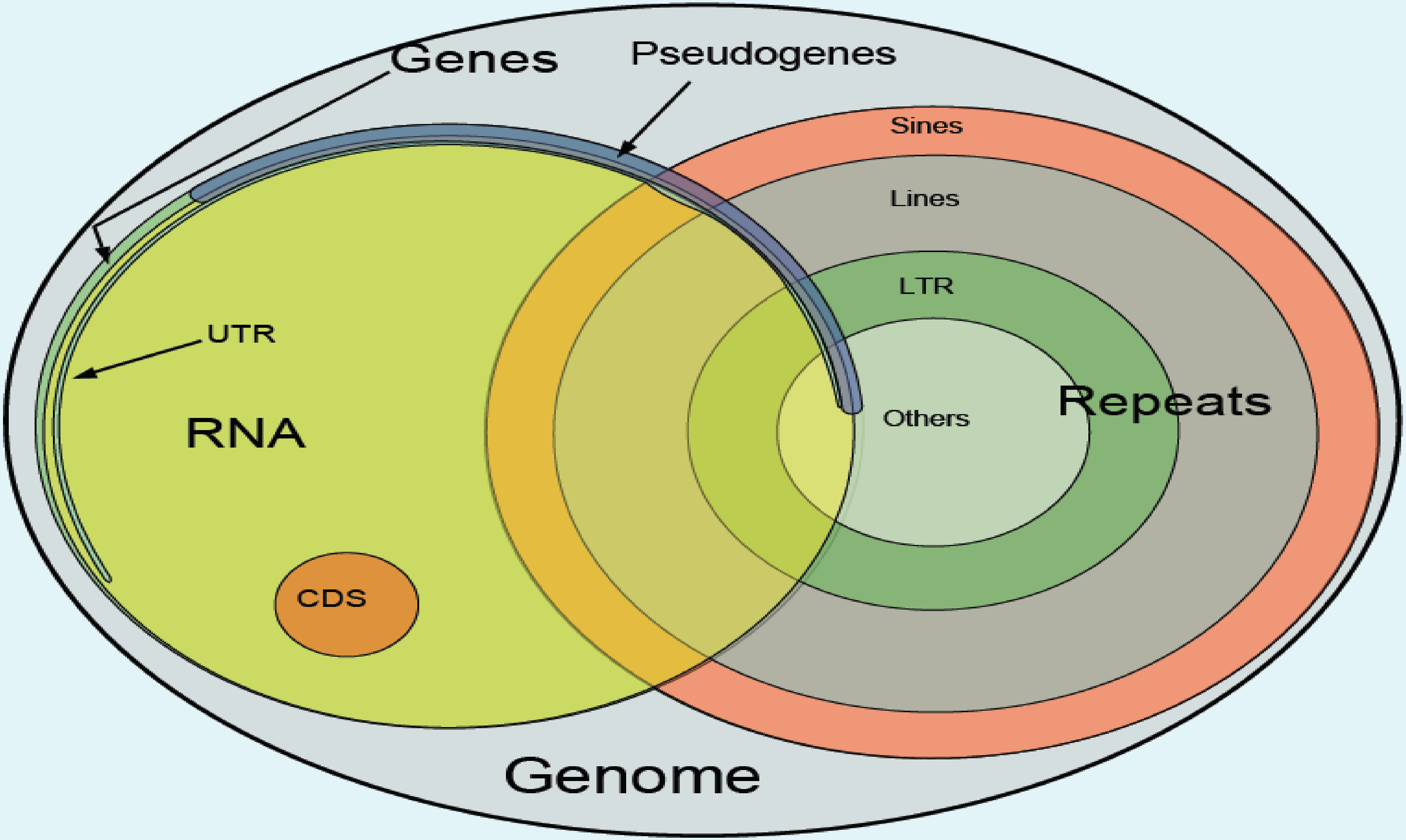
Transcriptome sequencing

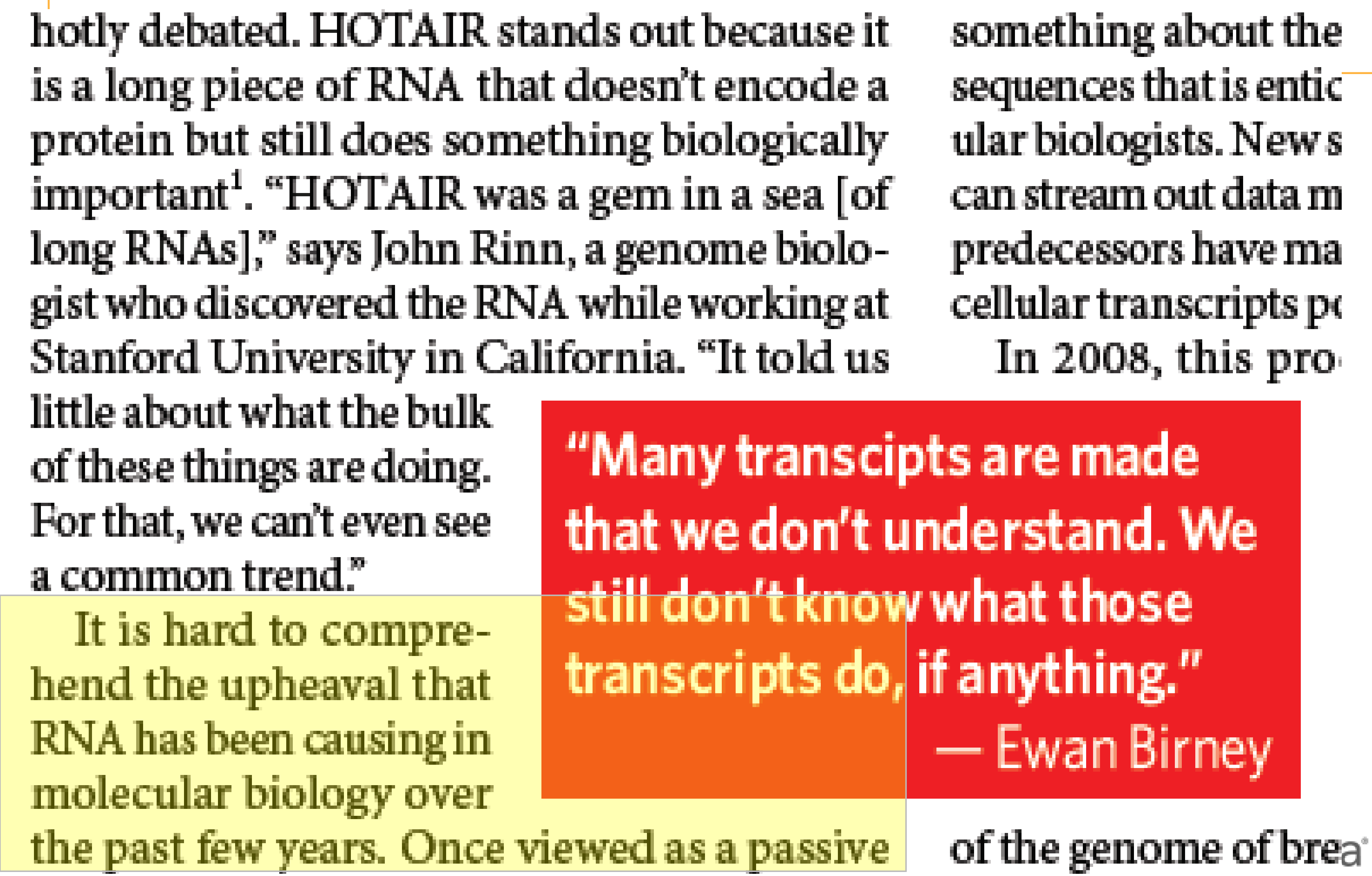
Goodbye!
Arrays

- Pervasive transcription
- Antisense transcription
- Non-coding RNA
- Promoter associated RNA
- Micro-RNA/mRNA targets
- Enhancer RNAs
- Gene fusions

Abizar Lakdawalla, PhD
alakdawalla@illumina.com

Human Genome *circa* 2010





hotly debated. HOTAIR stands out because it is a long piece of RNA that doesn't encode a protein but still does something biologically important[1]. "HOTAIR was a gem in a sea [of long RNAs]," says John Rinn, a genome biologist who discovered the RNA while working at Stanford University in California. "It told us little about what the bulk of these things are doing. For that, we can't even see a common trend."

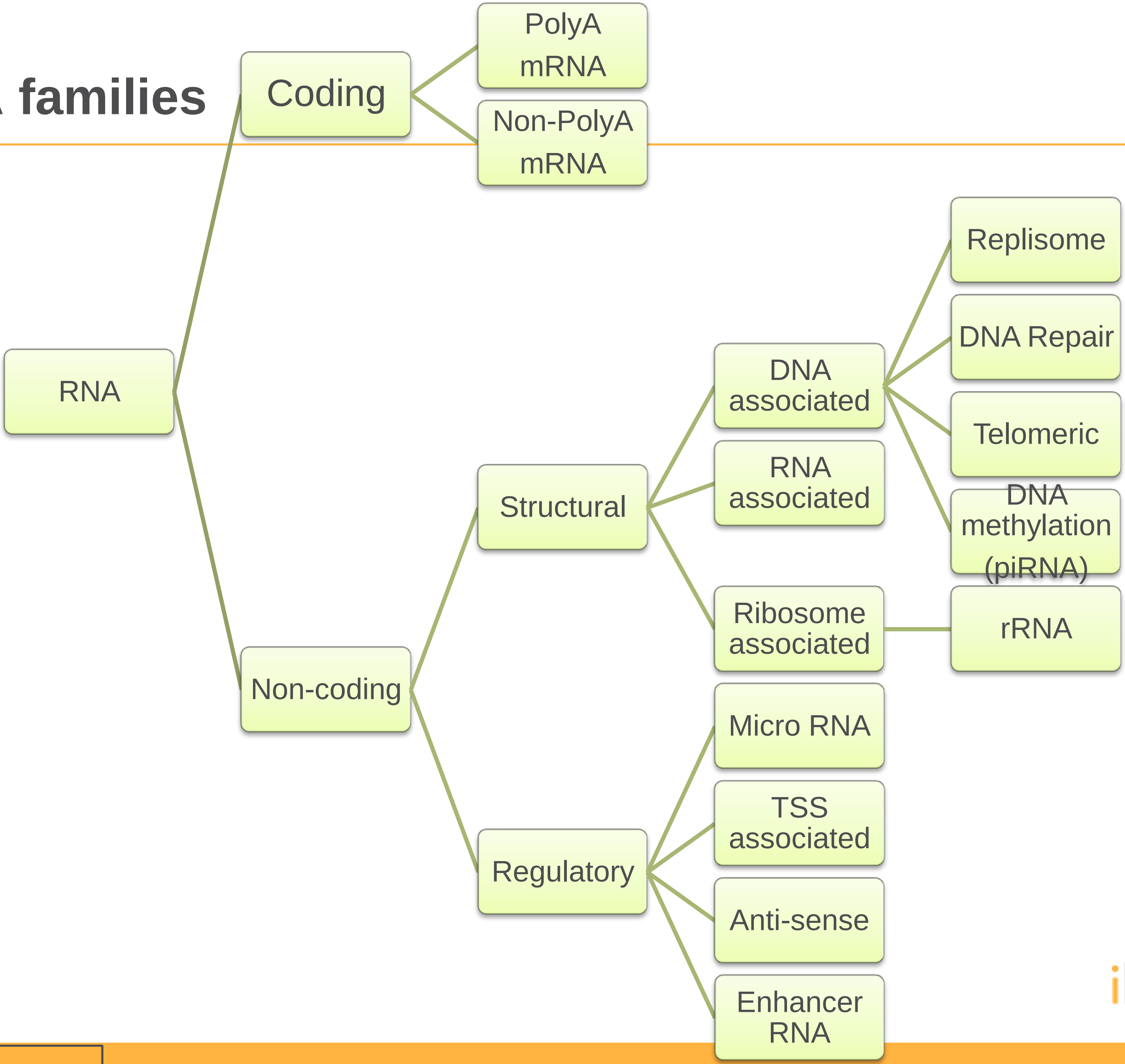
It is hard to comprehend the upheaval that RNA has been causing in molecular biology over the past few years. Once viewed as a passive

something about the
sequences that is entic
ular biologists. New s
can stream out data m
predecessors have ma
cellular transcripts pc
In 2008, this pro

of the genome of brea

> "Many transcipts are made that we don't understand. We still don't know what those transcripts do, if anything."
> — Ewan Birney

RNA families



RNA

Coding

PolyA mRNA

Non-PolyA mRNA

Non-coding

Structural

DNA associated

Replisome

DNA Repair

Telomeric

DNA methylation (piRNA)

RNA associated

Ribosome associated

rRNA

Regulatory

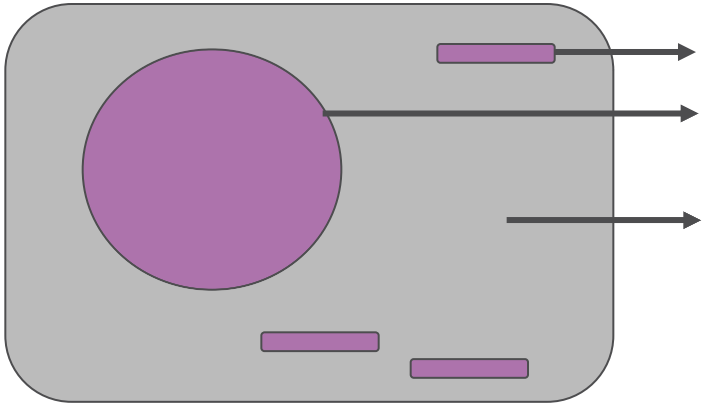
Micro RNA

TSS associated

Anti-sense

Enhancer RNA

Genome Analyzer Transcriptional Analysis



RNA from mitochondria

RNA from nucleus

RNA from cytoplasm

Sub-fractionate the RNA

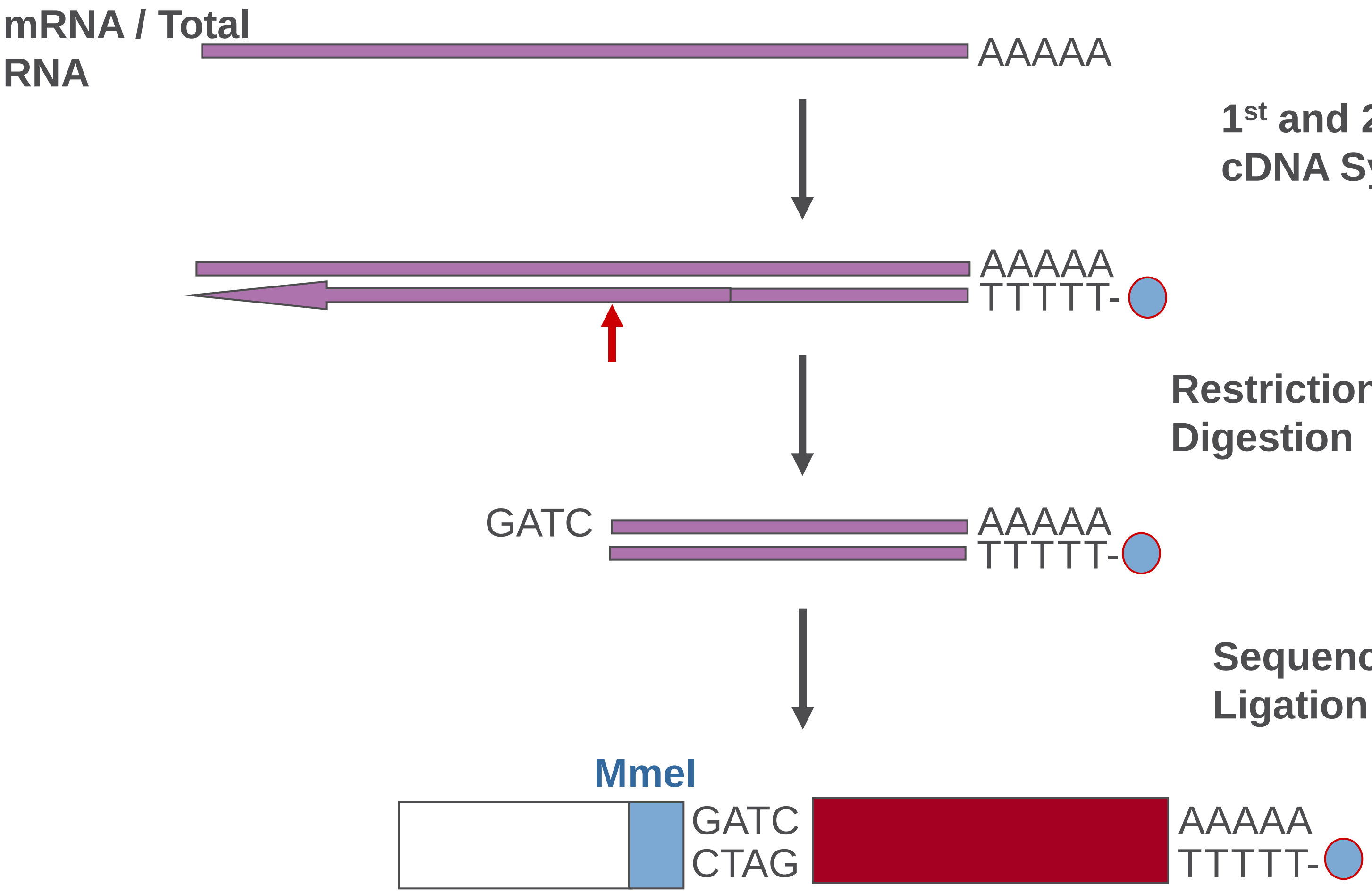
- Poly-A RNA

- RNA without rRNA
- RNA without …

Tag Sequence
(a snippet)

Sequence the whole RNA

Tag Sequencing



mRNA / Total RNA

AAAAA

1st and 2nd Strand cDNA Synthesis

AAAAA
TTTTT-

Restriction Enzyme Digestion

GATC

AAAAA
TTTTT-

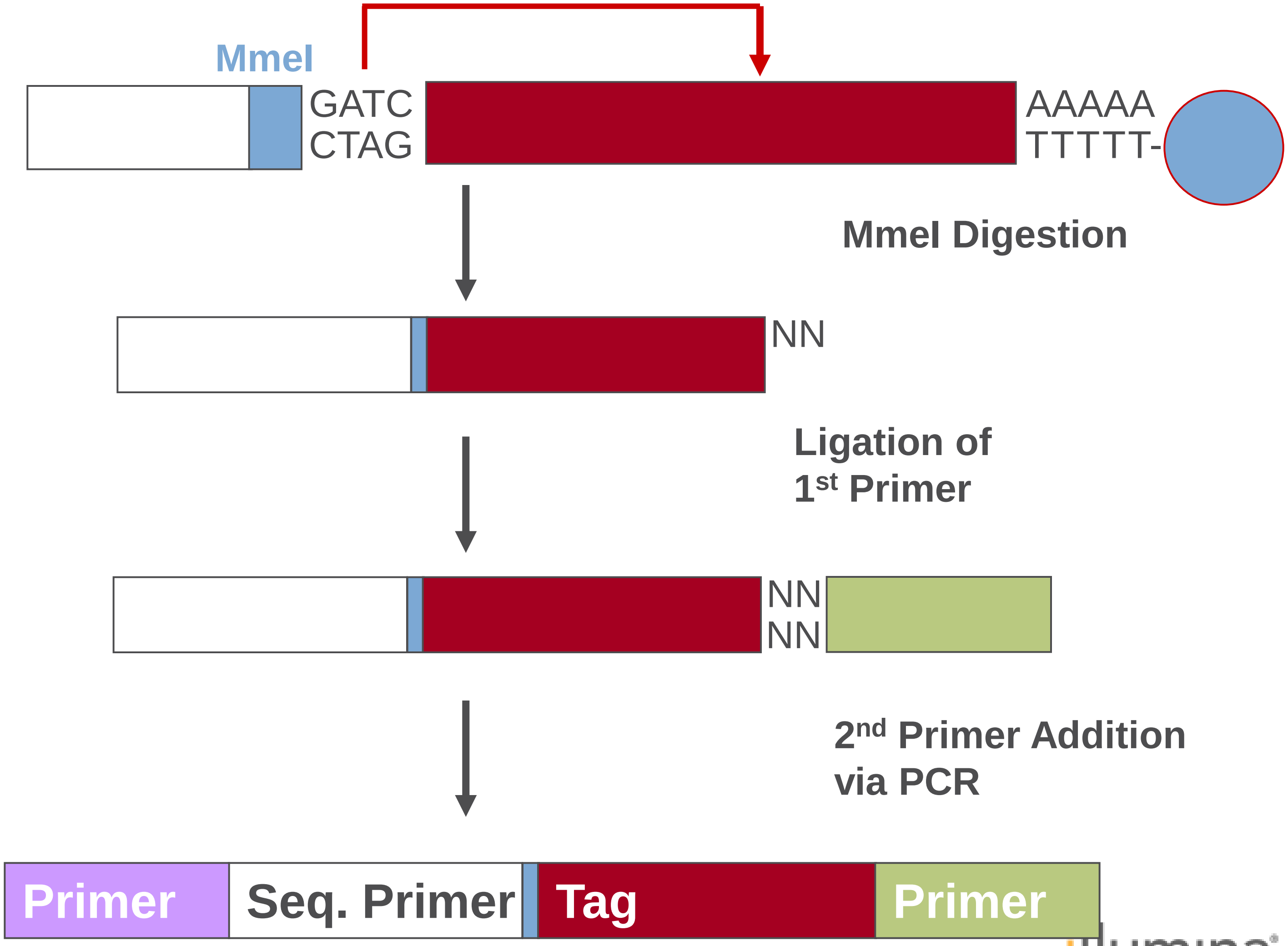
Sequencing Adapter Ligation

MmeI

GATC
CTAG

AAAAA
TTTTT-

Digital Tag mRNA Sample Preparation

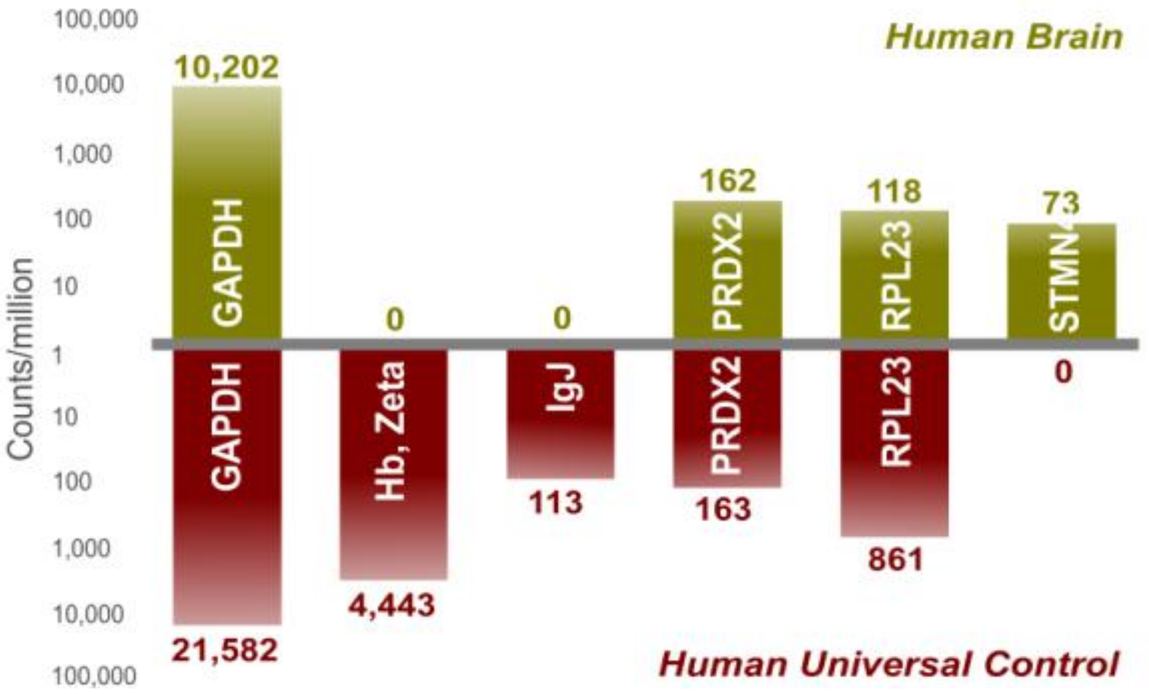


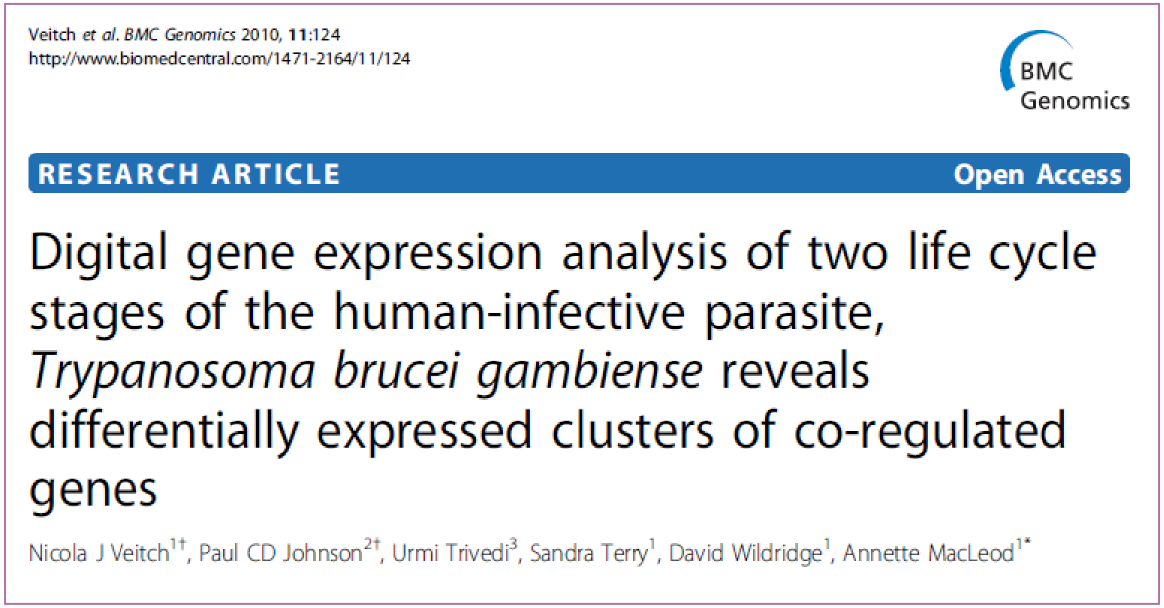
MmeI Digestion

Ligation of 1st Primer

2nd Primer Addition via PCR

Gene Expression





Veitch *et al. BMC Genomics* 2010, **11**:124
http://www.biomedcentral.com/1471-2164/11/124

BMC Genomics

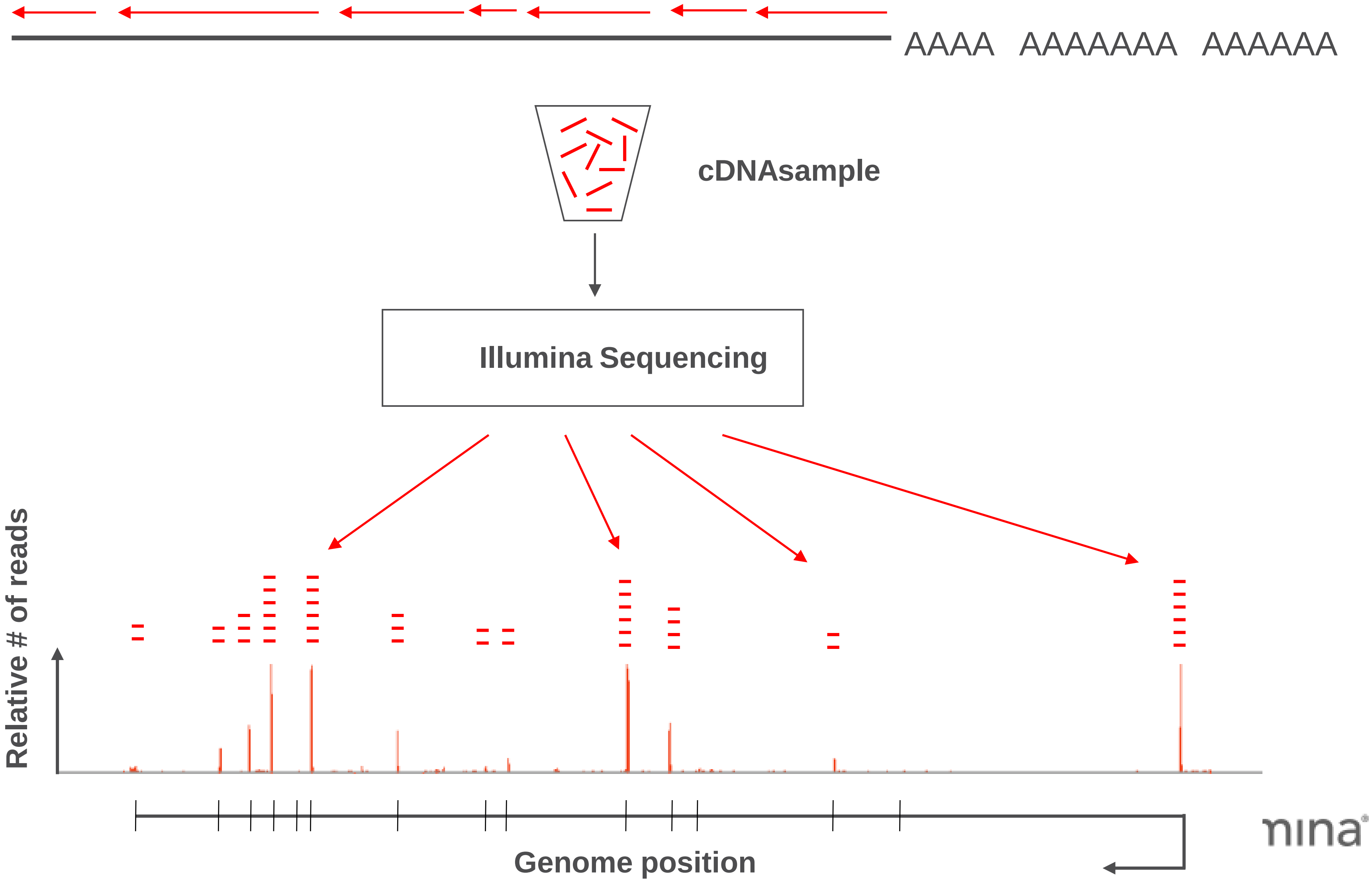
RESEARCH ARTICLE **Open Access**

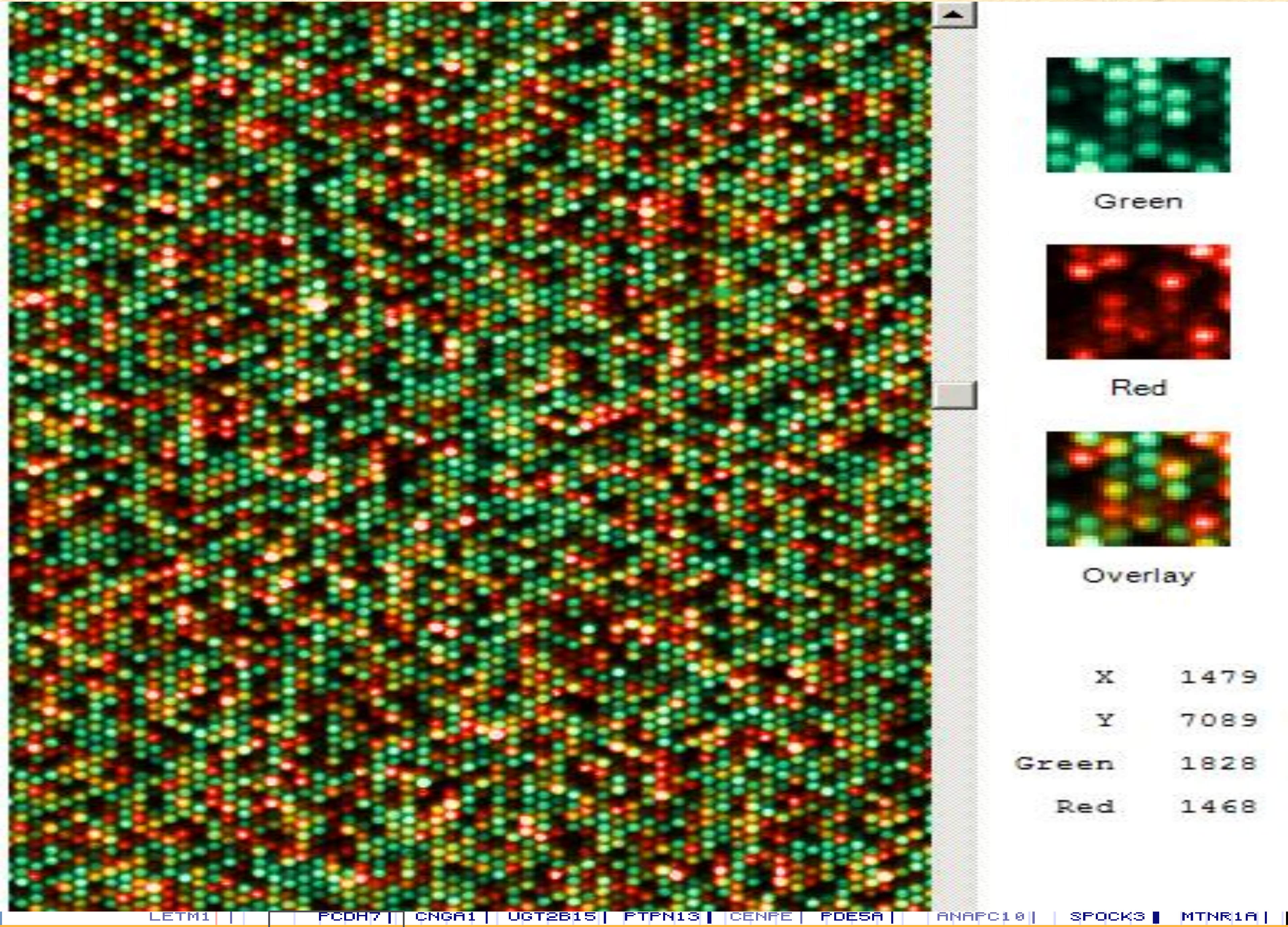
Digital gene expression analysis of two life cycle stages of the human-infective parasite, *Trypanosoma brucei gambiense* reveals differentially expressed clusters of co-regulated genes

Nicola J Veitch[1†], Paul CD Johnson[2†], Urmi Trivedi[3], Sandra Terry[1], David Wildridge[1], Annette MacLeod[1*]

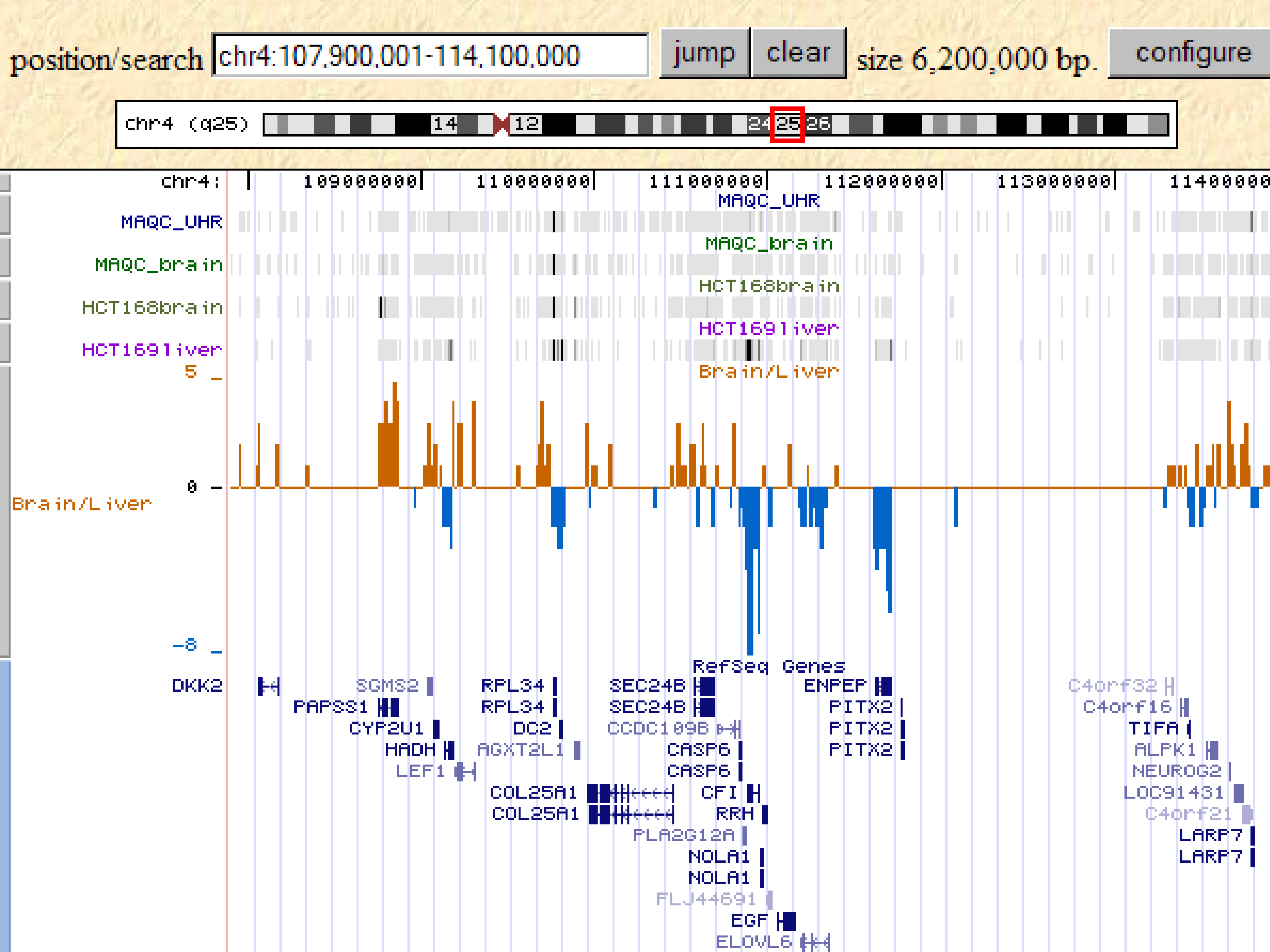
Revolution – RNA-Seq – PCR-free – Ribo-Seq – CLIP-Seq – Normalization - FFPE

RNA-Seq: Method





X	1479
Y	7089
Green	1828
Red	1468



position/search chr4:107,900,001-114,100,000 size 6,200,000 bp.

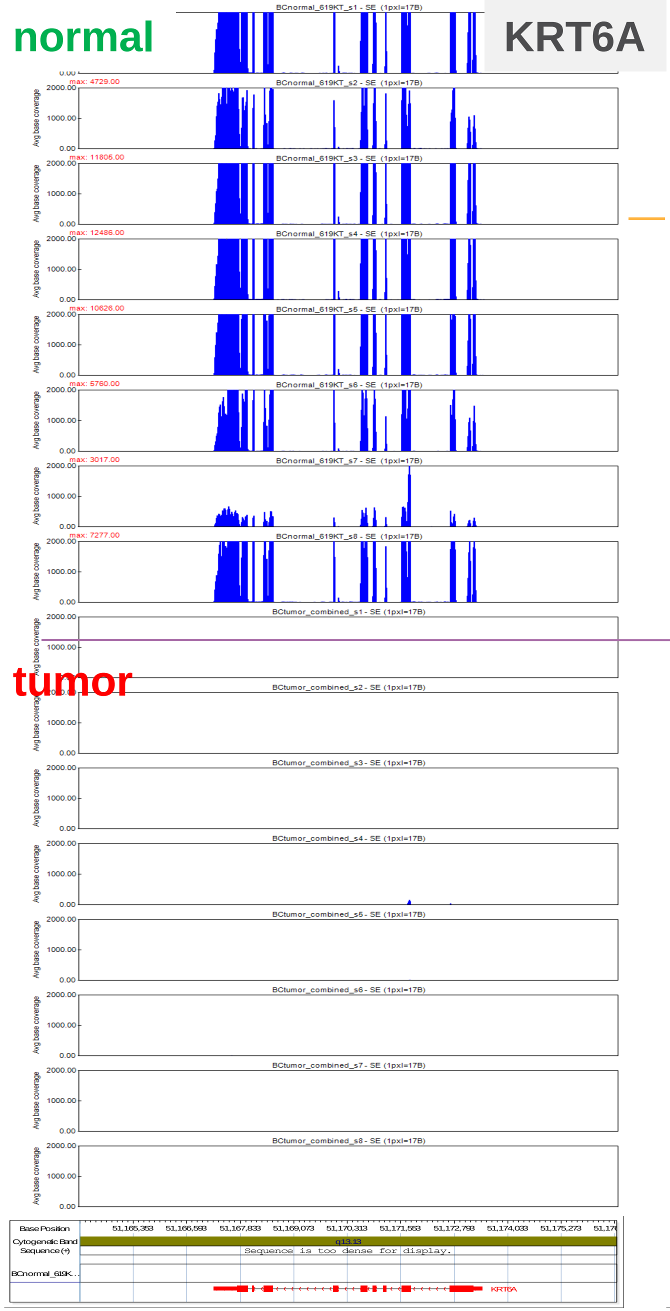
mRNA-Seq Data is Information Rich

- mRNA Expression Profiling
- Alternative Splicing Analysis
- Analysis of expressed SNPs and mutations
- Analysis of Allelic-specific Expression
- Chimeric Transcript Discovery
- Gene Discovery and Annotation

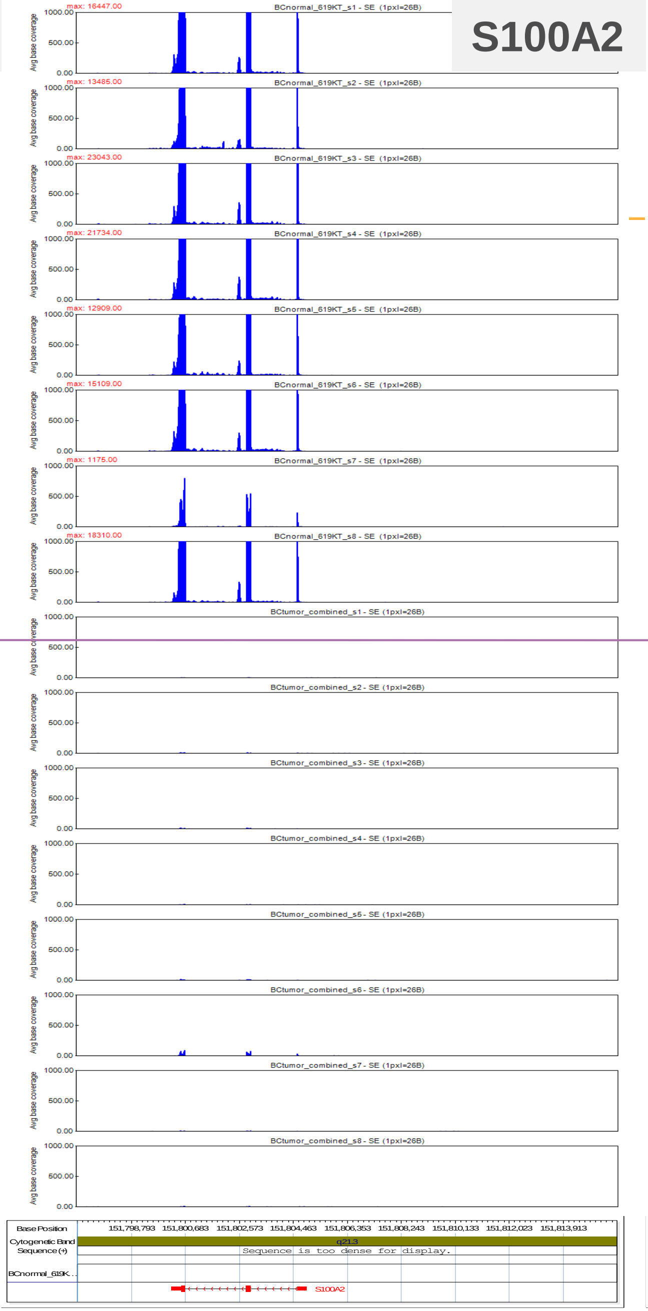
normal

tumor

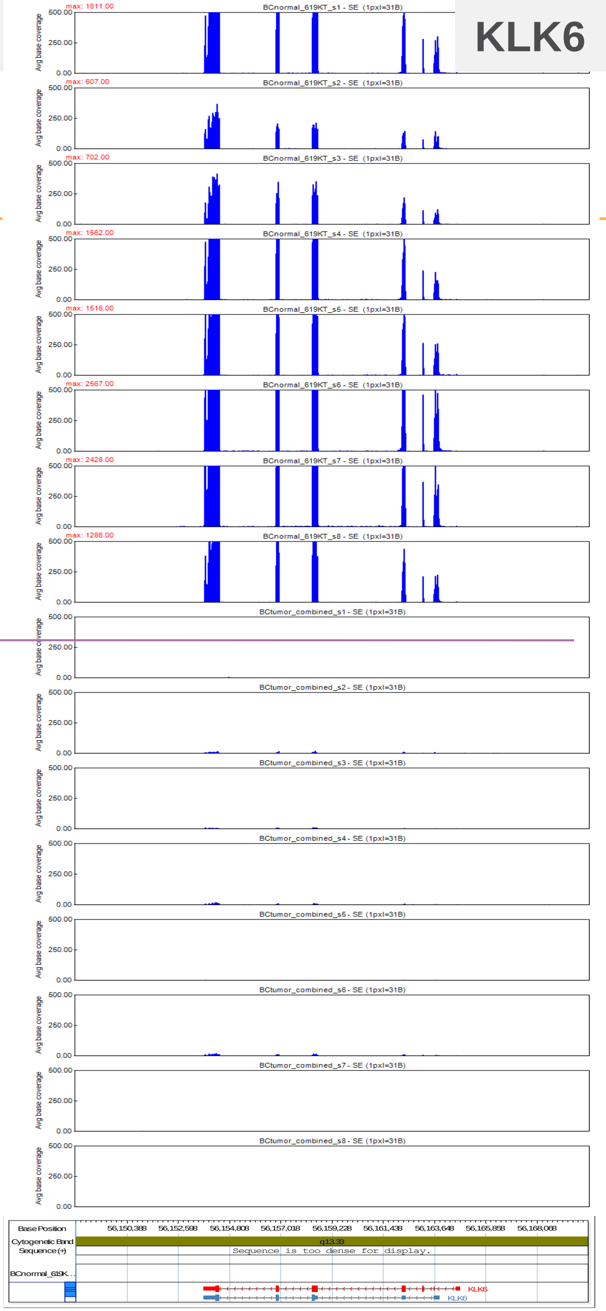
KRT6A



S100A2



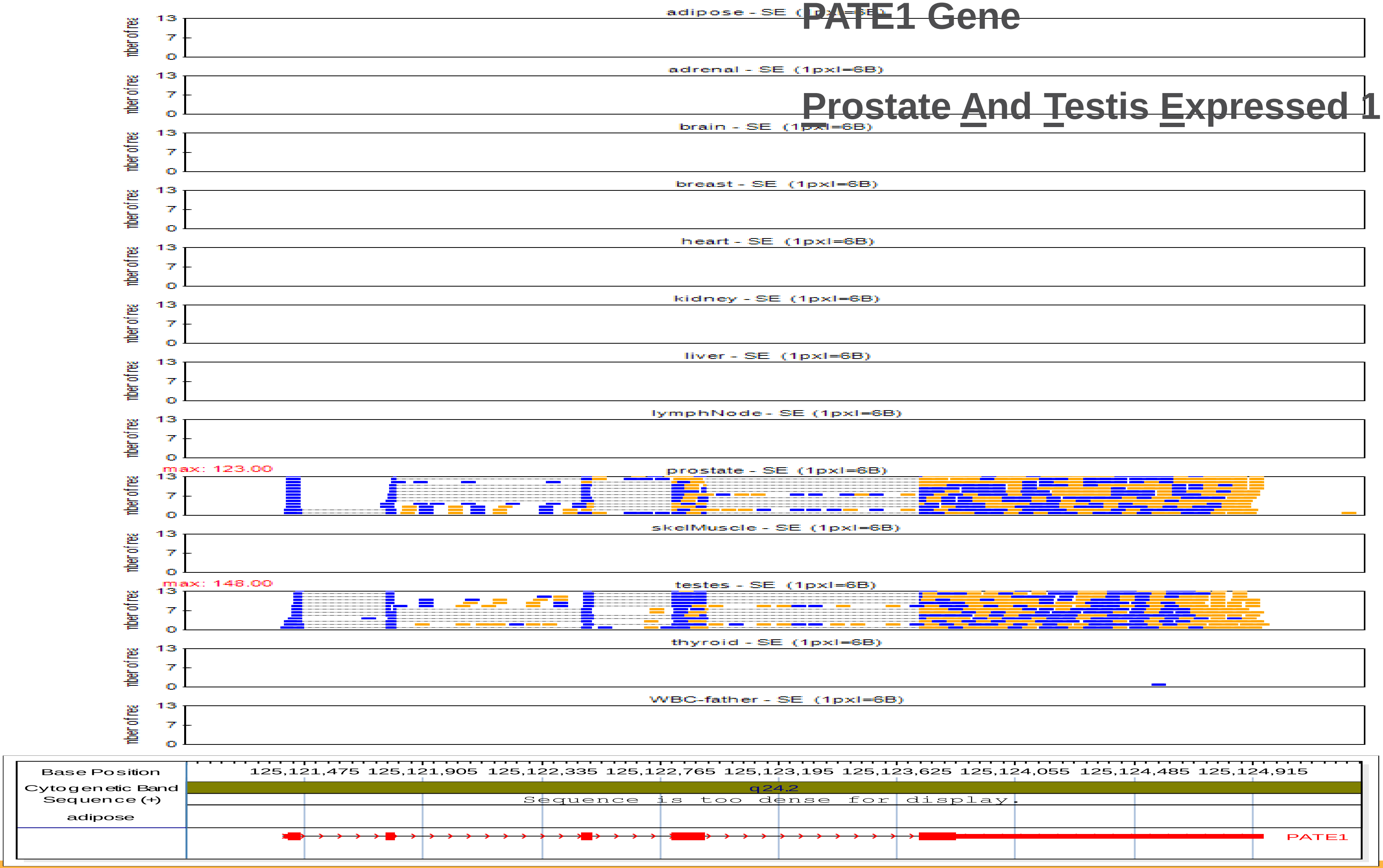
KLK6



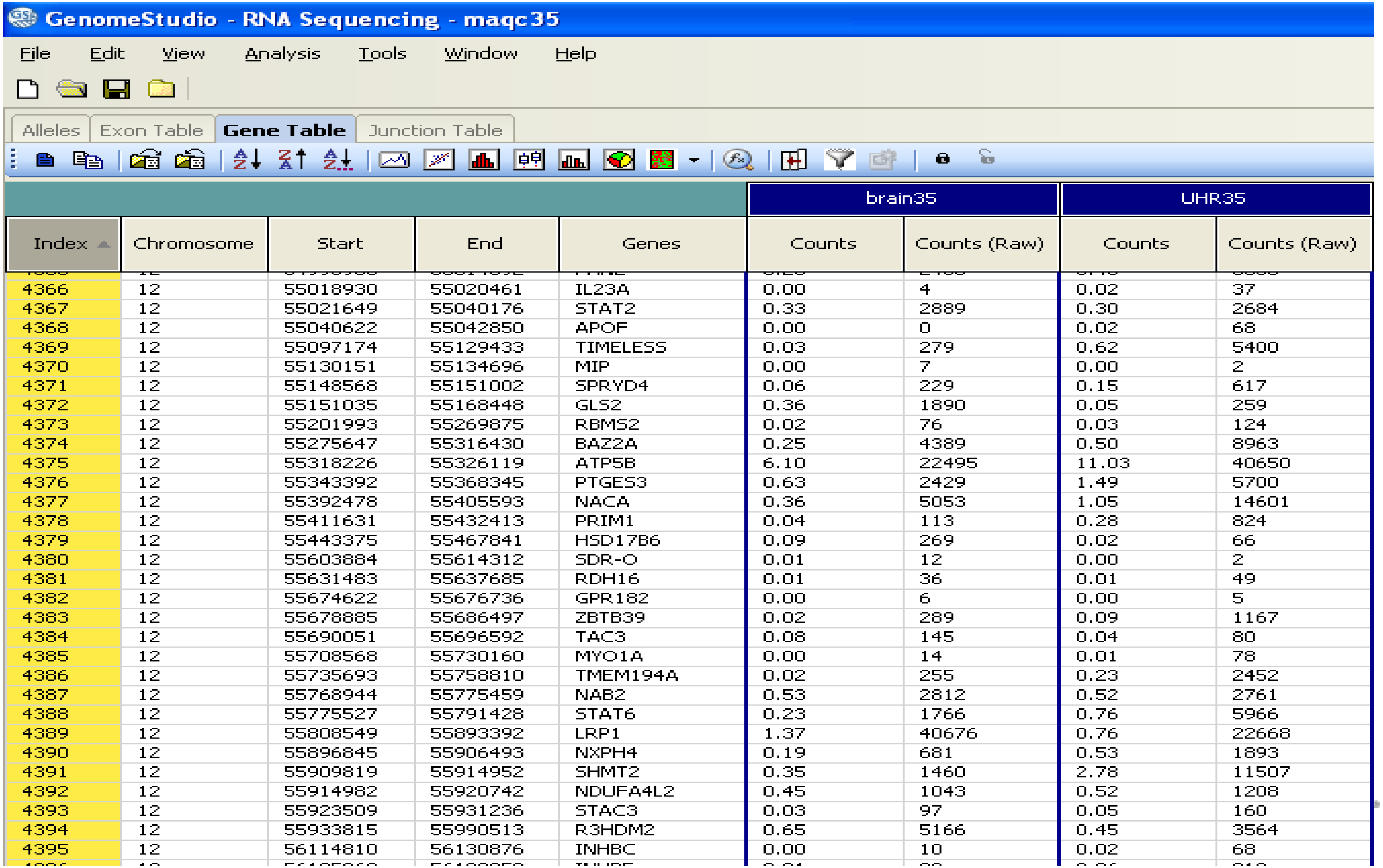
Tissue-specific genes

PATE1 Gene

Prostate And Testis Expressed 1



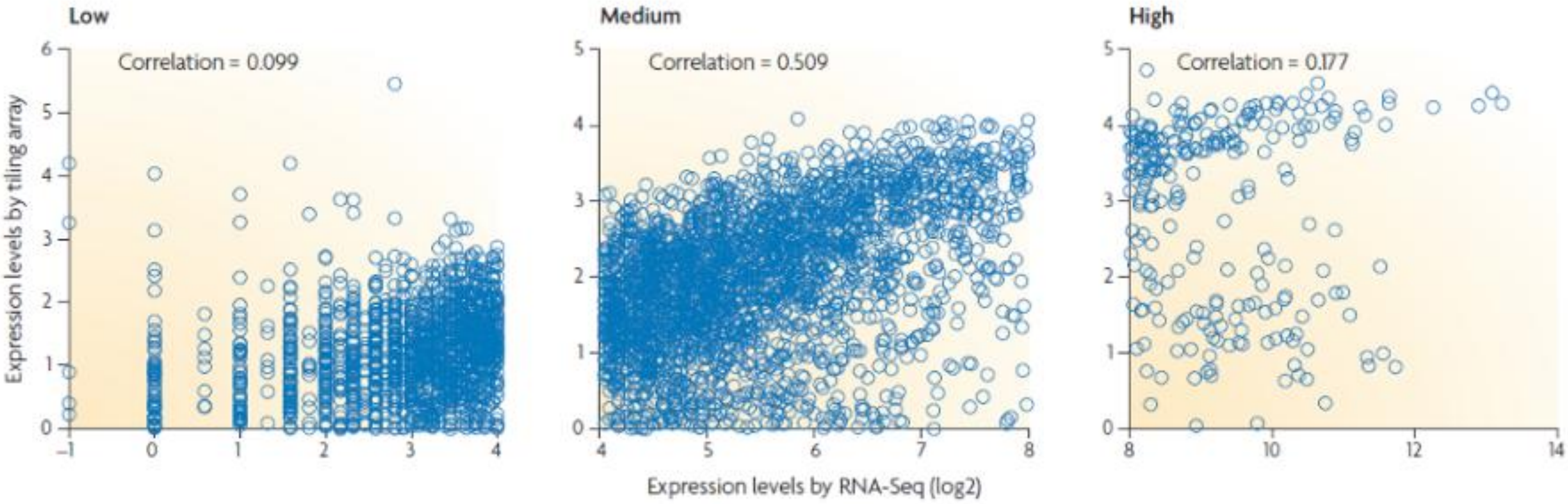
Read Counts for All RefSeq Genes



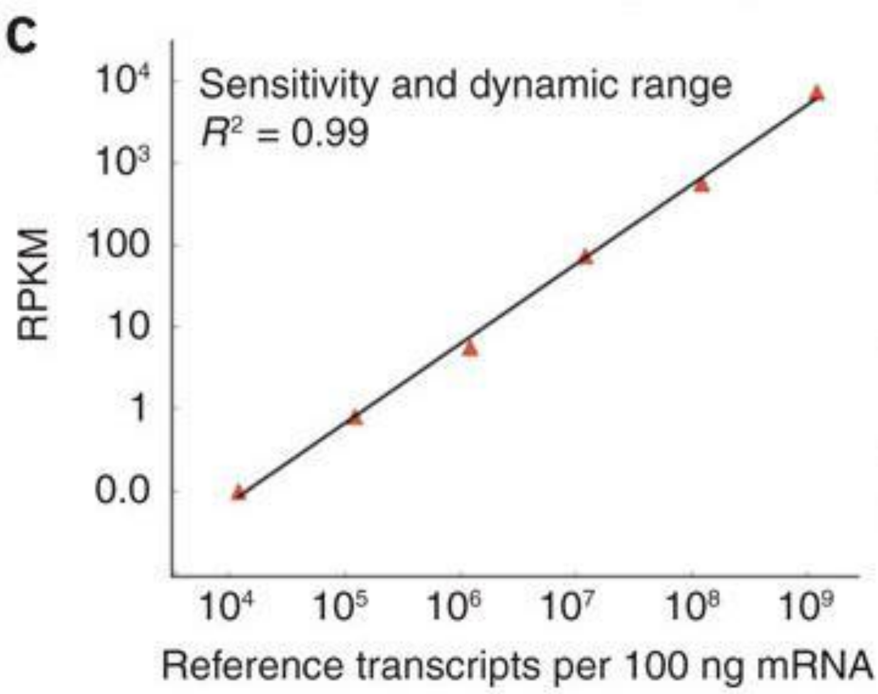
					brain35		UHR35	
Index	Chromosome	Start	End	Genes	Counts	Counts (Raw)	Counts	Counts (Raw)
4366	12	55018930	55020461	IL23A	0.00	4	0.02	37
4367	12	55021649	55040176	STAT2	0.33	2889	0.30	2684
4368	12	55040622	55042850	APOF	0.00	0	0.02	68
4369	12	55097174	55129433	TIMELESS	0.03	279	0.62	5400
4370	12	55130151	55134696	MIP	0.00	7	0.00	2
4371	12	55148568	55151002	SPRYD4	0.06	229	0.15	617
4372	12	55151035	55168448	GLS2	0.36	1890	0.05	259
4373	12	55201993	55269875	RBMS2	0.02	76	0.03	124
4374	12	55275647	55316430	BAZ2A	0.25	4389	0.50	8963
4375	12	55318226	55326119	ATP5B	6.10	22495	11.03	40650
4376	12	55343392	55368345	PTGES3	0.63	2429	1.49	5700
4377	12	55392478	55405593	NACA	0.36	5053	1.05	14601
4378	12	55411631	55432413	PRIM1	0.04	113	0.28	824
4379	12	55443375	55467841	HSD17B6	0.09	269	0.02	66
4380	12	55603884	55614312	SDR-O	0.01	12	0.00	2
4381	12	55631483	55637685	RDH16	0.01	36	0.01	49
4382	12	55674622	55676736	GPR182	0.00	6	0.00	5
4383	12	55678885	55686497	ZBTB39	0.02	289	0.09	1167
4384	12	55690051	55696592	TAC3	0.08	145	0.04	80
4385	12	55708568	55730160	MYO1A	0.00	14	0.01	78
4386	12	55735693	55758810	TMEM194A	0.02	255	0.23	2452
4387	12	55768944	55775459	NAB2	0.53	2812	0.52	2761
4388	12	55775527	55791428	STAT6	0.23	1766	0.76	5966
4389	12	55808549	55893392	LRP1	1.37	40676	0.76	22668
4390	12	55896845	55906493	NXPH4	0.19	681	0.53	1893
4391	12	55909819	55914952	SHMT2	0.35	1460	2.78	11507
4392	12	55914982	55920742	NDUFA4L2	0.45	1043	0.52	1208
4393	12	55923509	55931236	STAC3	0.03	97	0.05	160
4394	12	55933815	55990513	R3HDM2	0.65	5166	0.45	3564
4395	12	56114810	56130876	INHBC	0.00	10	0.02	68

RNA-seq and Microarray Compared

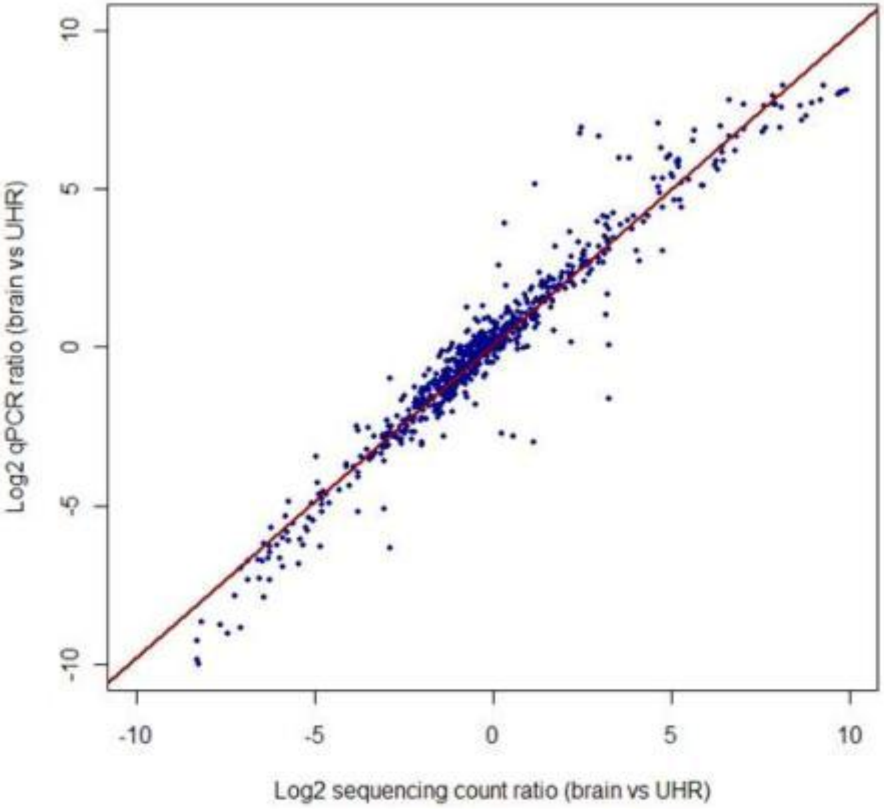
- **E**xpression levels are shown, as measured by RNA-Seq and tiling arrays for *Saccharomyces cerevisiae* cells. Agree for genes with medium levels of expression, but correlation is very low for genes with either low or high expression levels.



Performance of RNA sequencing is superior to other gene expression profiling methods



Dynamic Range



Accuracy and Sensitivity

illumina

Overview of mRNA-Seq Assays

- Start with 1 ug of Total RNA
- Purify poly-A mRNA
- Randomly Fragment RNA

Random Priming

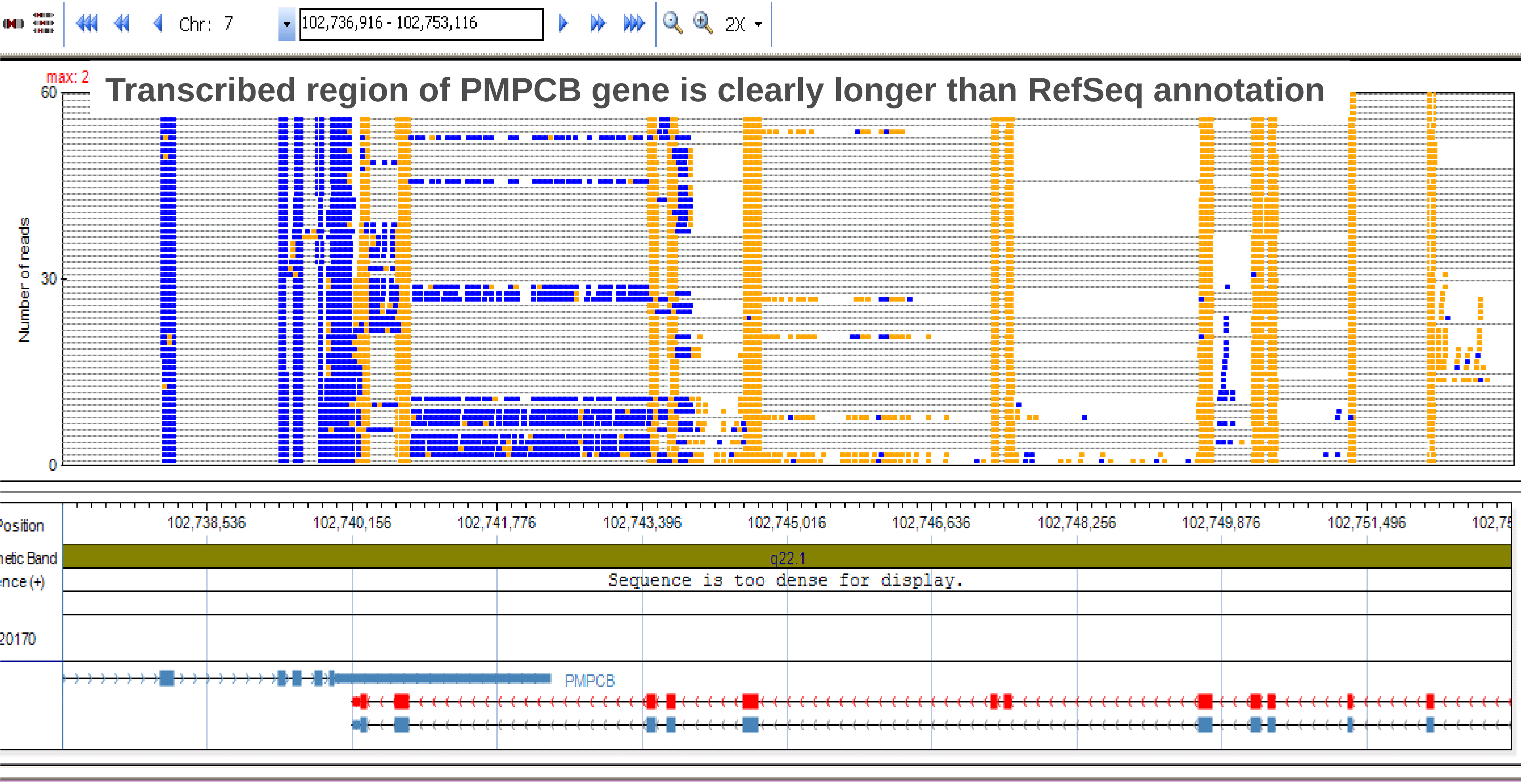
- Random Prime mRNA → cDNA
- Make 2nd Strand cDNA
- Ligate Sequencing Adapters

RNA Ligation

- CIP Treatment
- PNK Treatment
- Ligate 3'- Small RNA Adapters
- Ligate 5'- Small RNA Adapters
- RT to make cDNA

- PCR Amplify 15 Cycles
- Purify, Quantify
- Grow Clusters
- Sequence on Genome Analyzer
  - → Single or Paired-end Reads
  - → 35 to 100 base read lengths

Overlapping of PMPCB and DNAJC2 genes on two strands



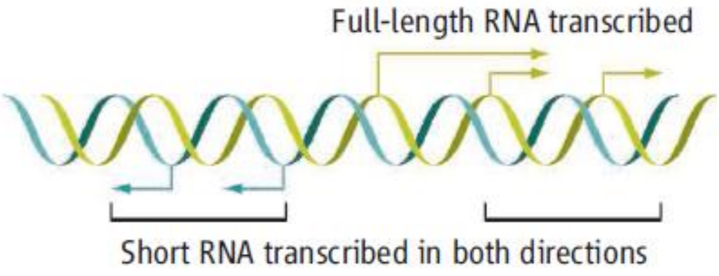
Transcribed region of PMPCB gene is clearly longer than RefSeq annotation

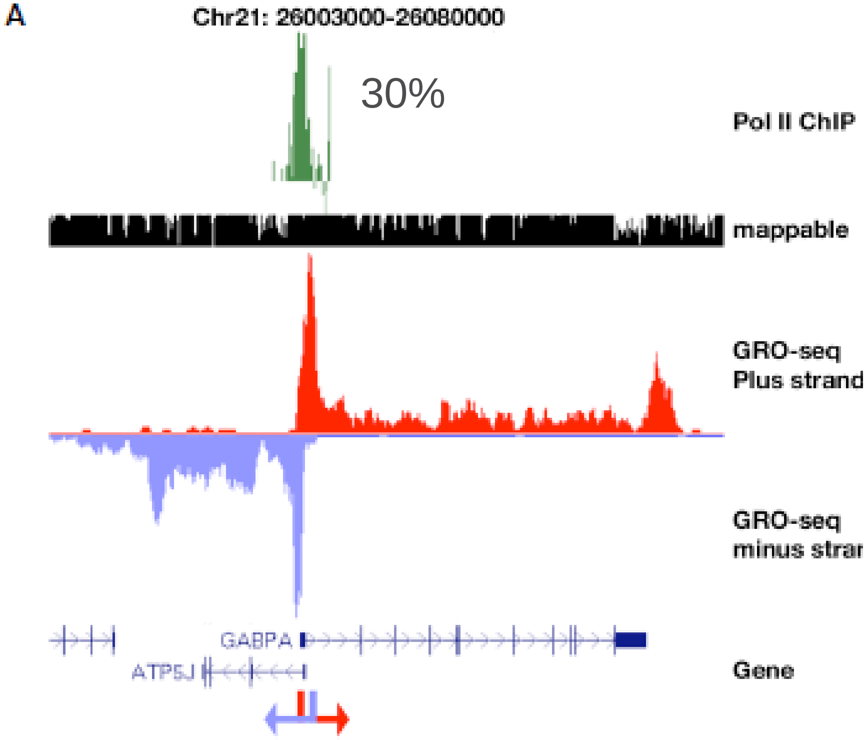
illumina

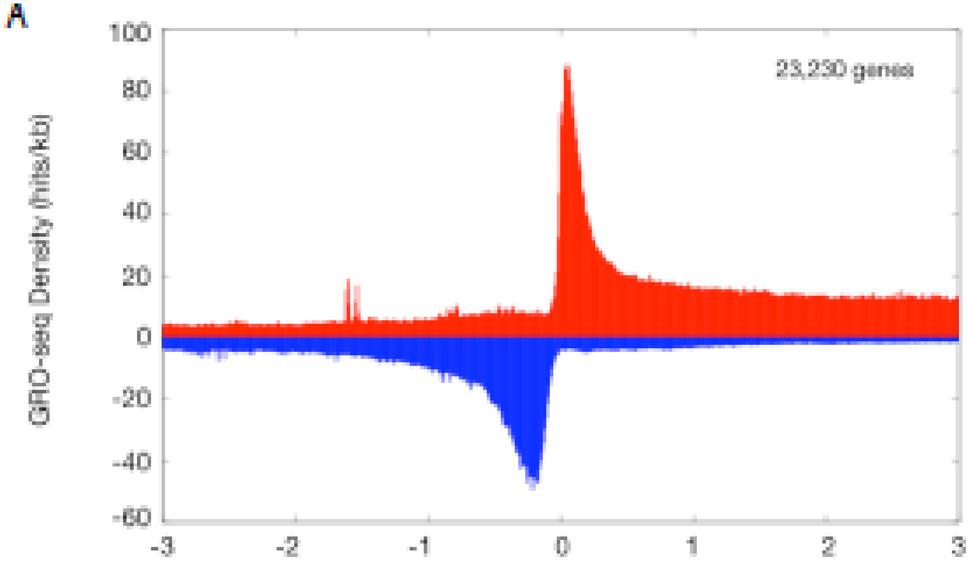
Revolution – **RNA-Seq** – PCR-free – Ribo-Seq – CLIP-Seq – Normalization - FFPE

Divergent Transcription from Active Promoters
Amy C. Seila, *et al.*
Science **322**, 1849 (2008);
DOI: 10.1126/science.1162253

Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters
Leighton J. Core, *et al.*
Science **322**, 1845 (2008);
DOI: 10.1126/science.1162228

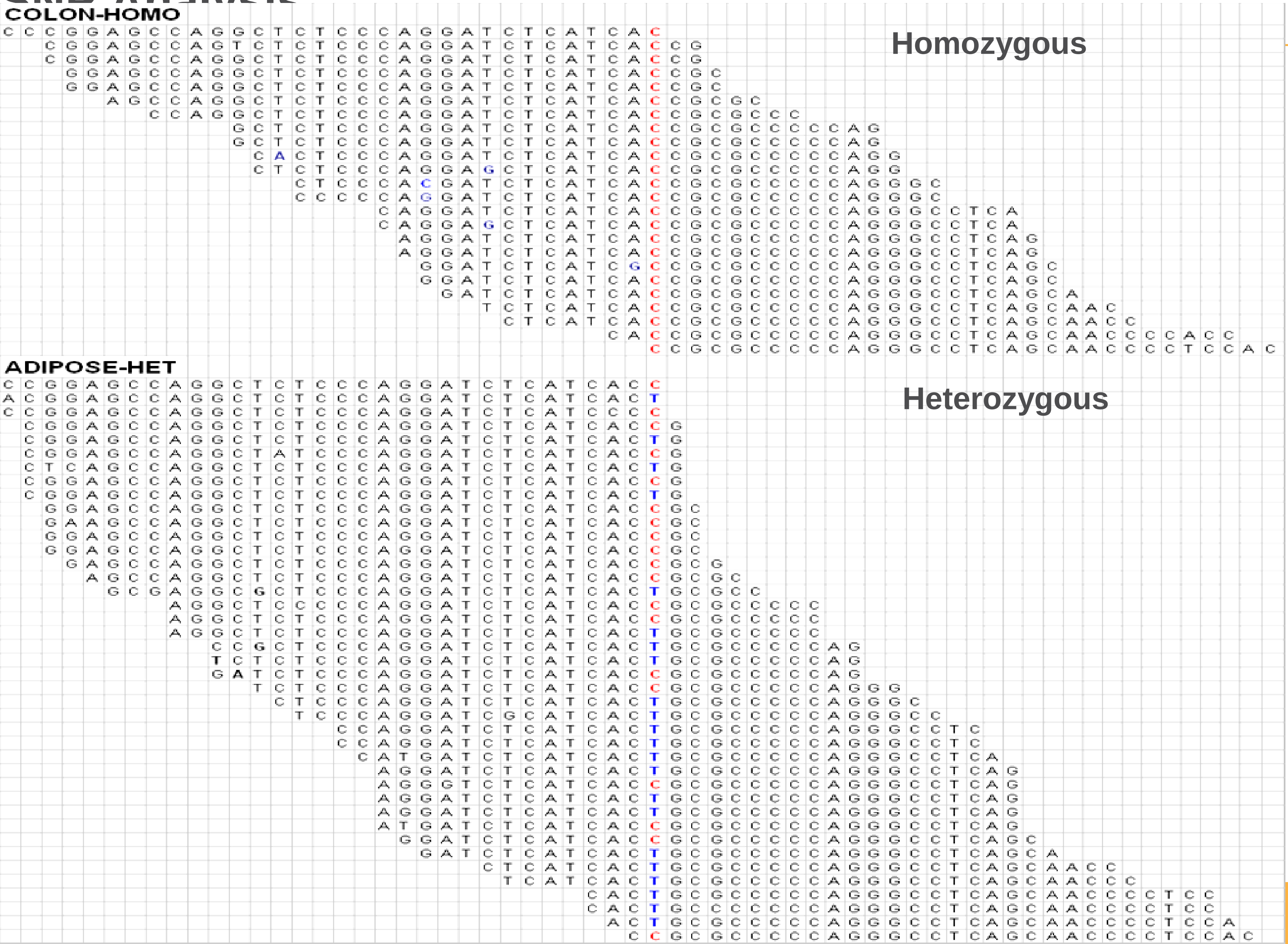






illumina

cSNP Analysis



COLON-HOMO

Homozygous

```
CCCGGAGCCAGGCTCTCCCAGGATCTCATCAC
  CGGAGCCAGTCTCTCCCAGGATCTCATCACCG
  CGGAGCCAGGCTCTCCCAGGATCTCATCACCG
   GGAGCCAGGCTCTCCCAGGATCTCATCACCGC
   GGAGCCAGGCTCTCCCAGGATCTCATCACCGC
     AGCCAGGCTCTCCCAGGATCTCATCACCGCGC
       CCAGGCTCTCCCAGGATCTCATCACCGCGCCC
           GCTCTCCCAGGATCTCATCACCGCGCCCCCAG
           GCTCTCCCAGGATCTCATCACCGCGCCCCCAG
            CACTCCCAGGATCTCATCACCGCGCCCCCAGG
            CTCTCCCAGGAGCTCATCACCGCGCCCCCAGG
              CTCCCACGATCTCATCACCGCGCCCCCAGGGC
              CCCCCAGGATCTCATCACCGCGCCCCCAGGGC
                  CAGGATCTCATCACCGCGCCCCCAGGGCCTCA
                  CAGGAGCTCATCACCGCGCCCCCAGGGCCTCA
                   AGGATCTCATCACCGCGCCCCCAGGGCCTCAG
                   AGGATCTCATCACCGCGCCCCCAGGGCCTCAG
                    GGATCTCATCGCCGCGCCCCCAGGGCCTCAGC
                    GGATCTCATCACCGCGCCCCCAGGGCCTCAGC
                     GATCTCATCACCGCGCCCCCAGGGCCTCAGCA
                       TCTCATCACCGCGCCCCCAGGGCCTCAGCAAC
                        CTCATCACCGCGCCCCCAGGGCCTCAGCAACC
                             CACCGCGCCCCCAGGGCCTCAGCAACCCCACC
                               CCGCGCCCCCAGGGCCTCAGCAACCCCTCCAC
```

ADIPOSE-HET

Heterozygous

```
CCGGAGCCAGGCTCTCCCAGGATCTCATCACC
ACGGAGCCAGGCTCTCCCAGGATCTCATCACT
CCGGAGCCAGGCTCTCCCAGGATCTCATCCCC
 CGGAGCCAGGCTCTCCCAGGATCTCATCACCG
 CGGAGCCAGGCTCTCCCAGGATCTCATCACTG
 CGGAGCCAGGCTATCCCAGGATCTCATCACCG
 CTCAGCCAGGCTCTCCCAGGATCTCATCACTG
 CGGAGCCAGGCTCTCCCAGGATCTCATCACCG
 CGGAGCCAGGCTCTCCCAGGATCTCATCACTG
  GGAGCCAGGCTCTCCCAGGATCTCATCACCGC
  GAAGCCAGGCTCTCCCAGGATCTCATCACCGC
  GGAGCCAGGCTCTCCCAGGATCTCATCACCGC
  GGAGCCAGGCTCTCCCAGGATCTCATCACCGC
   GAGCCAGGCTCTCCCAGGATCTCATCACCGCG
    AGCCAGGCTCTCCCAGGATCTCATCACCGCGC
     GCGAGGCGCTCCCAGGATCTCATCACTGCGCC
        AGGCTCCCCCAGGATCTCATCACCGCGCCCCC
        AGGCTCTCCCAGGATCTCATCACCGCGCCCCC
        AGGCTCTCCCAGGATCTCATCACTGCGCCCCC
          CCGCTCCCAGGATCTCATCACTGCGCCCCCAG
          TCTCTCCCAGGATCTCATCACTGCGCCCCCAG
          GATCTCCCAGGATCTCATCACCGCGCCCCCAG
            TCTCCCAGGATCTCATCACCGCGCCCCCAGGG
             CTCCCAGGATCTCATCACTGCGCCCCCAGGGC
              TCCCAGGATCGCATCACTGCGCCCCCAGGGCC
                CCAGGATCTCATCACTGCGCCCCCAGGGCCTC
                CCAGGATCTCATCACTGCGCCCCCAGGGCCTC
                 CATGATCTCATCACTGCGCCCCCAGGGCCTCA
                  AGGATCTCATCACTGCGCCCCCAGGGCCTCAG
                  AGGGTCTCATCACCGCGCCCCCAGGGCCTCAG
                  AGGATCTCATCACTGCGCCCCCAGGGCCTCAG
                  AGGATCTCATCACTGCGCCCCCAGGGCCTCAG
                  ATGATCTCATCACCGCGCCCCCAGGGCCTCAG
                   GGATCTCATCACCGCGCCCCCAGGGCCTCAGC
                    GATCTCATCACTGCGCCCCCAGGGCCTCAGCA
                       CTCATCACTGCGCCCCCAGGGCCTCAGCAACC
                        TCATCACTGCGCCCCCAGGGCCTCAGCAACCC
                            CACTGCGCCCCCAGGGCCTCAGCAACCCCTCC
                            CACTGCCCCCCCAGGGCCTCAGCAACCCCTCC
                             ACTGCGCCCCCAGGGCCTCAGCAACCCCTCCA
                              CCGCGCCCCCAGGGCCTCAGCAACCCCTCCAC
```

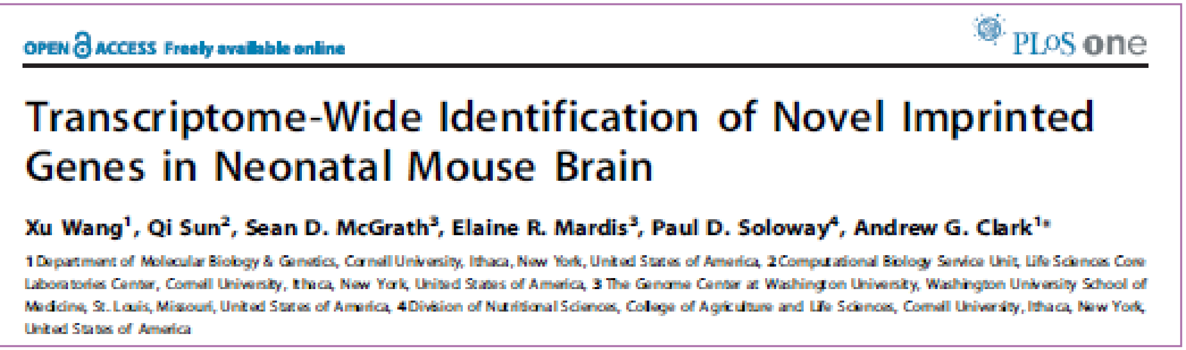
Differential Allelic Expression

Advances in the identification and analysis of allele-specific expression

Christopher G Bell and Stephan Beck

Address: Medical Genomics, University College London Cancer Institute, Huntley Street, London WC1E 6BT, UK.

Correspondence: Stephan Beck. Email: s.beck@ucl.ac.uk

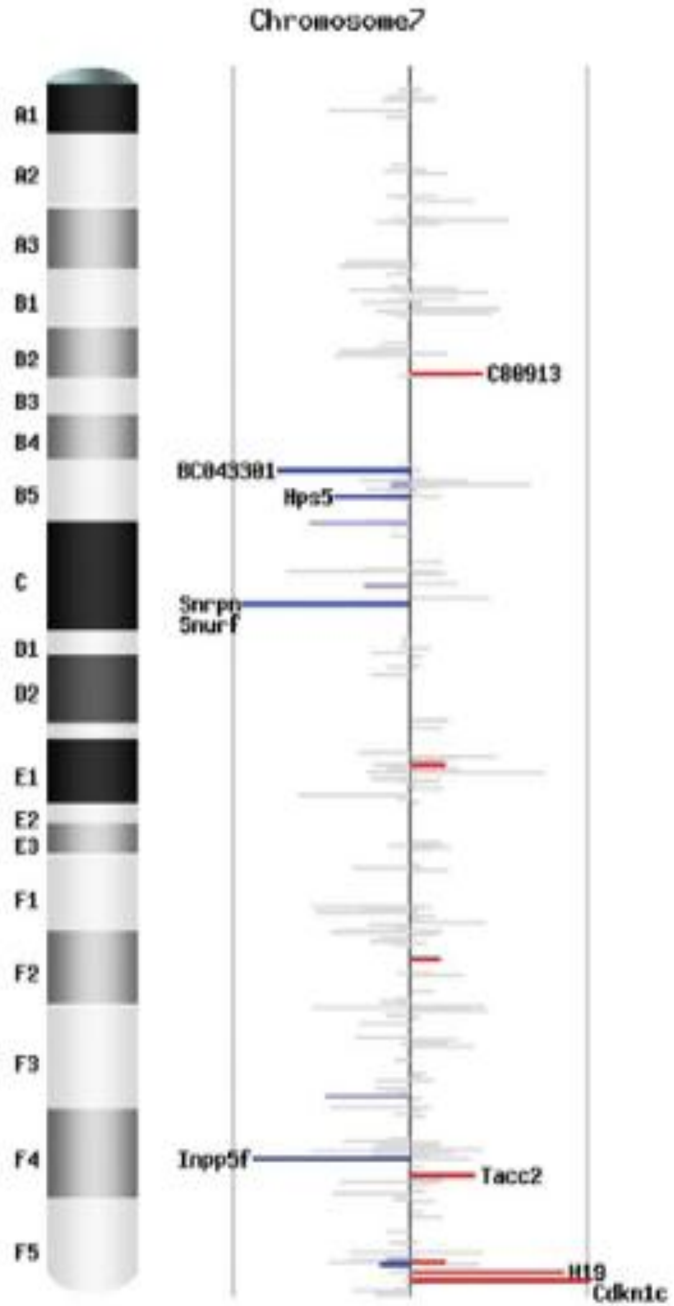


OPEN ACCESS Freely available online PLoS one

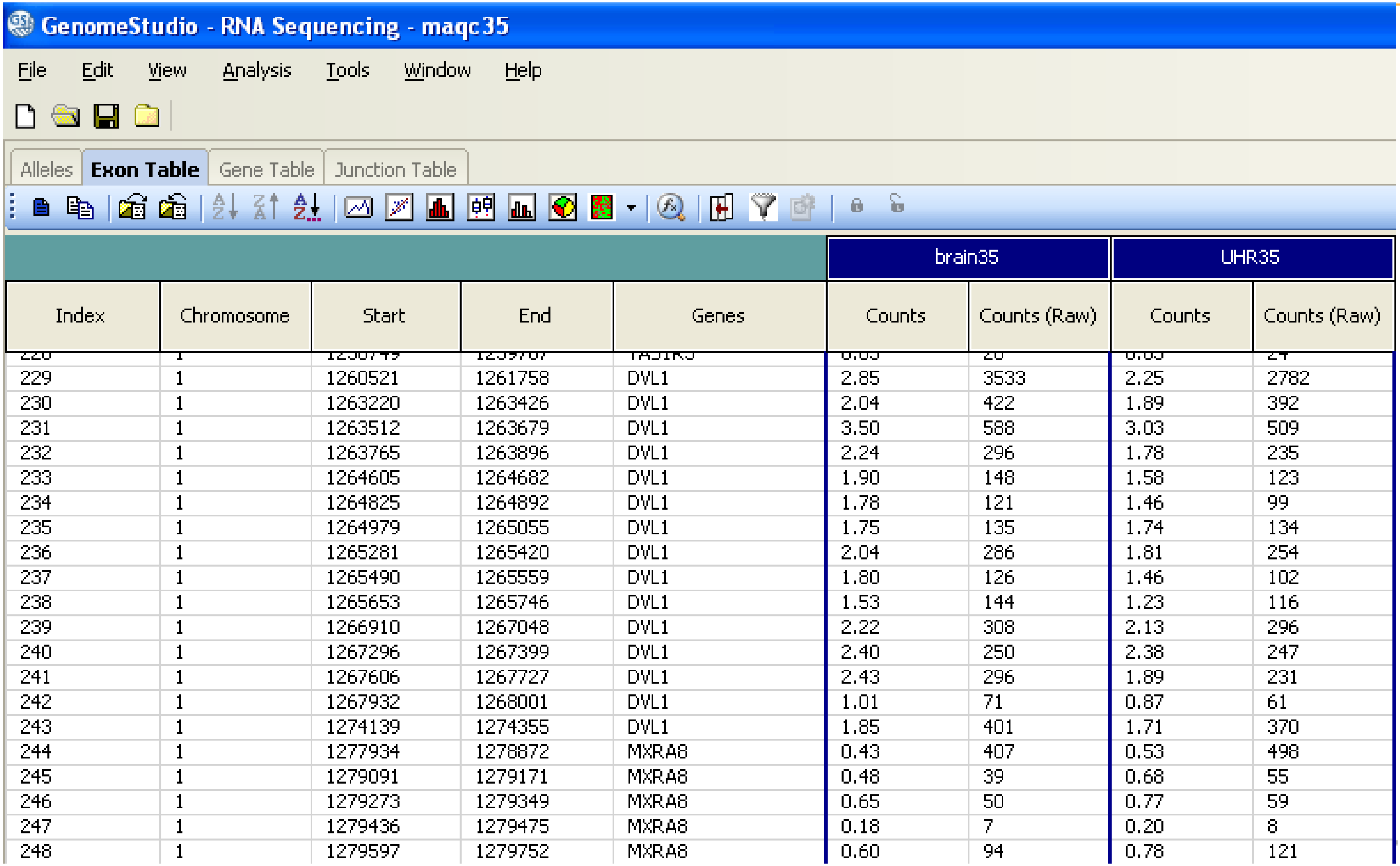
Transcriptome-Wide Identification of Novel Imprinted Genes in Neonatal Mouse Brain

Xu Wang[1], Qi Sun[2], Sean D. McGrath[3], Elaine R. Mardis[3], Paul D. Soloway[4], Andrew G. Clark[1*]

1 Department of Molecular Biology & Genetics, Cornell University, Ithaca, New York, United States of America, 2 Computational Biology Service Unit, Life Sciences Core Laboratories Center, Cornell University, Ithaca, New York, United States of America, 3 The Genome Center at Washington University, Washington University School of Medicine, St. Louis, Missouri, United States of America, 4 Division of Nutritional Sciences, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York, United States of America

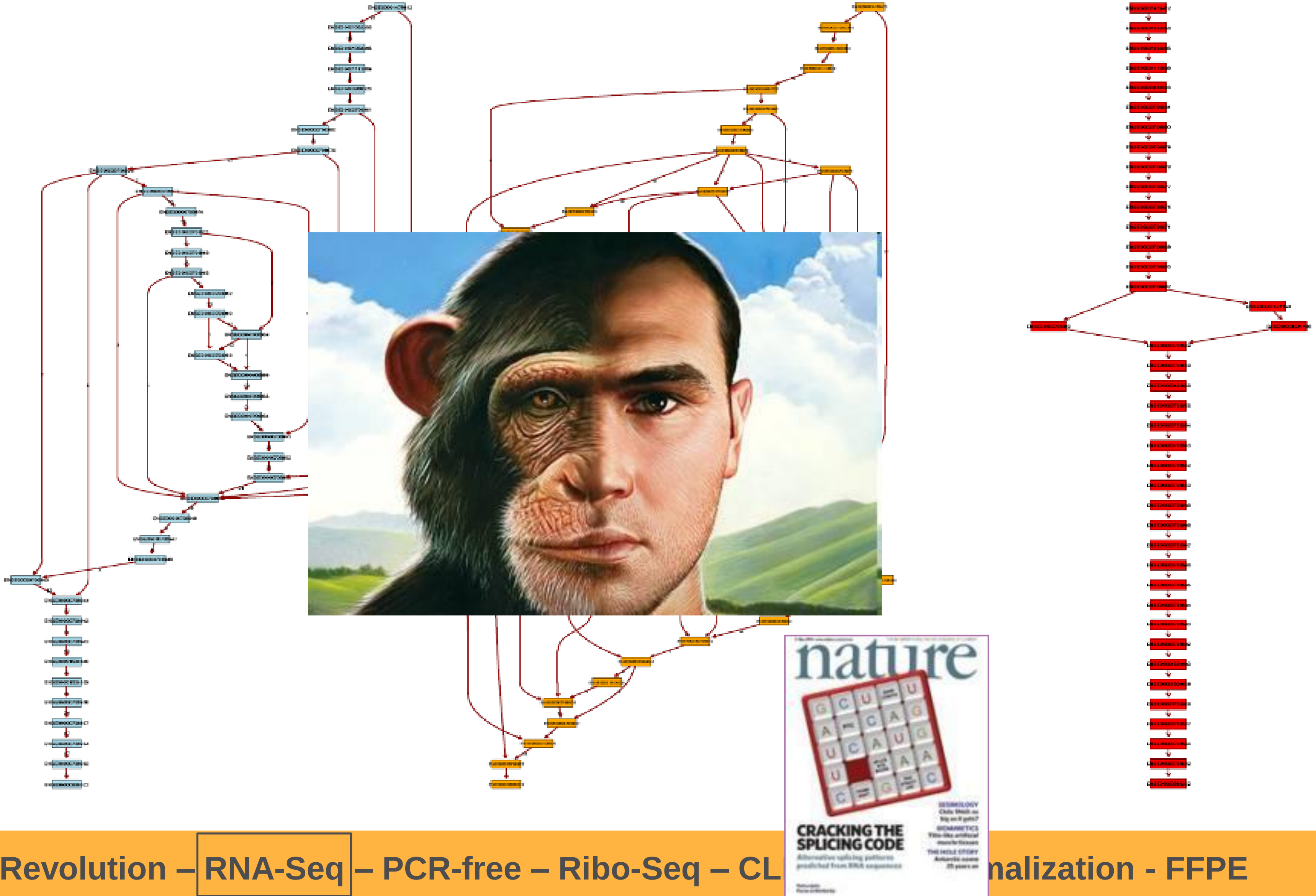


Read Counts for All RefSeq Exons



					brain35		UHR35	
Index	Chromosome	Start	End	Genes	Counts	Counts (Raw)	Counts	Counts (Raw)
229	1	1260521	1261758	DVL1	2.85	3533	2.25	2782
230	1	1263220	1263426	DVL1	2.04	422	1.89	392
231	1	1263512	1263679	DVL1	3.50	588	3.03	509
232	1	1263765	1263896	DVL1	2.24	296	1.78	235
233	1	1264605	1264682	DVL1	1.90	148	1.58	123
234	1	1264825	1264892	DVL1	1.78	121	1.46	99
235	1	1264979	1265055	DVL1	1.75	135	1.74	134
236	1	1265281	1265420	DVL1	2.04	286	1.81	254
237	1	1265490	1265559	DVL1	1.80	126	1.46	102
238	1	1265653	1265746	DVL1	1.53	144	1.23	116
239	1	1266910	1267048	DVL1	2.22	308	2.13	296
240	1	1267296	1267399	DVL1	2.40	250	2.38	247
241	1	1267606	1267727	DVL1	2.43	296	1.89	231
242	1	1267932	1268001	DVL1	1.01	71	0.87	61
243	1	1274139	1274355	DVL1	1.85	401	1.71	370
244	1	1277934	1278872	MXRA8	0.43	407	0.53	498
245	1	1279091	1279171	MXRA8	0.48	39	0.68	55
246	1	1279273	1279349	MXRA8	0.65	50	0.77	59
247	1	1279436	1279475	MXRA8	0.18	7	0.20	8
248	1	1279597	1279752	MXRA8	0.60	94	0.78	121

Transcriptome isoforms –



Making Sense

- "… studying transcription for a long time and never seen this kind of transcription before … but we also see a polymerase that appears to be pointing in the wrong direction." - Seila/Sharp, MIT

"The gene, in other words, is in an identity crisis."

illumina

Gene Fusions

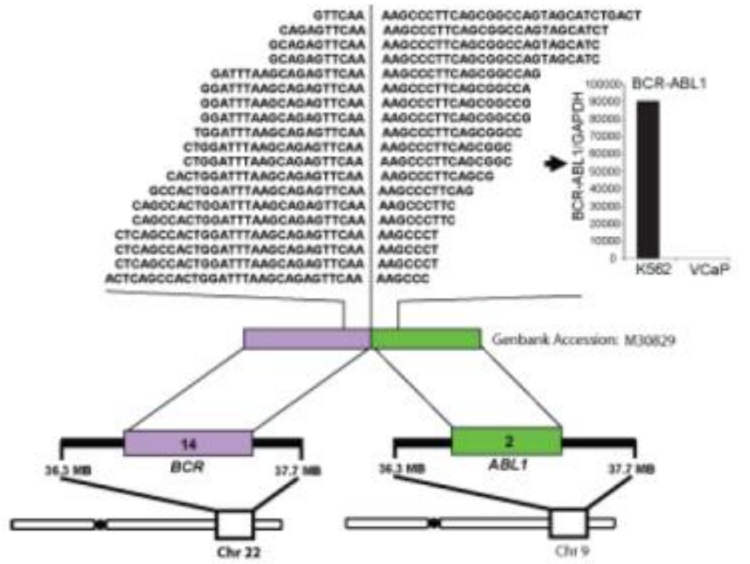
doi:10.1038/nature07638 nature

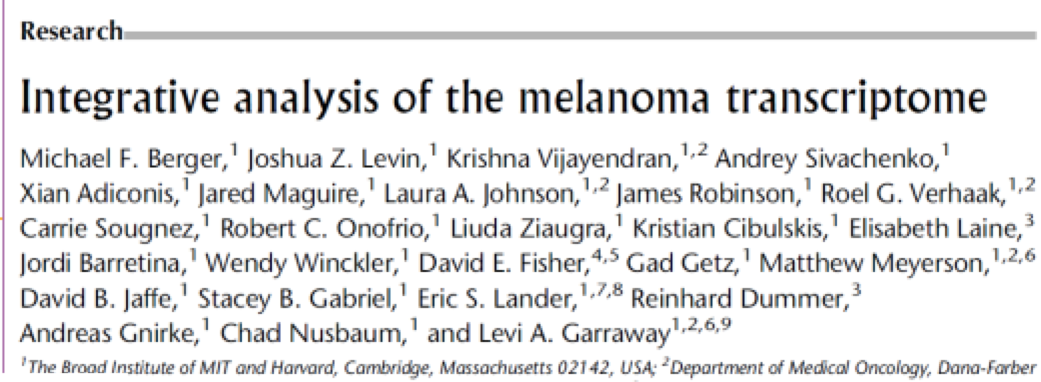
LETTERS

Transcriptome sequencing to detect gene fusions in cancer

Christopher A. Maher[1,3]*, Chandan Kumar-Sinha[1,3]*, Xuhong Cao[1,2], Shanker Kalyana-Sundaram[1,3], Bo Han[1,3], Xiaojun Jing[1,3], Lee Sam[1,3], Terrence Barrette[1,3], Nallasivam Palanisamy[1,3] & Arul M. Chinnaiyan[1,2,3,4,5]

Chimera	Chimera Class	Location	5' Gen				
BCR-ABL1	Class I: Translocation	22q11.23	BCR, breakpoint c				
MRPS10-HPR	Class I: Translocation	6p21.1	MRPS10, mitochond protein S				
MIPOL1-DGKB	Class II: Inter-chromosomal complex	14q13.3-q21.1	MIPOL1, mirror-image polydactyly 1	7p21.2	DGKB, diacylglycerol kinase, beta 90kDa	LNCaP	Long read, Short read, qRT-PCR. FISH
*TMPRSS2-ERG**	Class III: Interstitial Deletion	21q22.3	TMPRSS2, transmembrane protease, serine 2	21q22.3	ERG, v-ets erythroblastosis virus E26 oncogene homolog (avian)	VCaP, VCaP-Met	Long read, Short read, qRT-PCR, FISH
*USP10-ZDHHC7**	Class III: Interstitial Deletion	16q24.1	USP10, ubiquitin specific peptidase 10	16q24.1	ZDHHC7, zinc finger, DHHC-type containing 7	VCaP, VCaP-Met	Long read, Short read, qRT-PCR, aCGH
*STRN4-GPSN2**	Class IV: Intra-chromosomal complex	19q13.2	STRN4, striatin, calmodulin binding protein 4	19p13.12	GPSN2, glycoprotein, synaptic 2	Met-3	Short read, qRT-PCR
LMAN2-AP3S1	Class IV: Intra-chromosomal complex	5q35.3	LMAN2 lectin, mannose-binding 2	5q22			CR
*HJURP-EIF4E2**	Class IV: Intra-chromosomal complex	2q37.1	HJURP, Holliday junction recognition protein	2q37.1			-PCR, FISH
*INPP4A-HJURP**	Class II: Intra-chromosomal complex	2q11.2	INPP4A, inositol polyphosphate-4-phosphatase, type I	2q37.1			-PCR, FISH
RC3H2-RGS3	Class IV: Intra-chromosomal complex	9q34	RC3H2, ring finger and CCCH-type zinc finger domains 2	9q32			CR
ZNF649-ZNF577	Class V: Read-through	19q13.33	ZNF649, zinc finger protein 649	19q13.3			RT-PCR
*MBTPS2-YY2**	Class V: Read-through	Xp22.1-p22.2	MBTPS2, membrane-bound transcription factor peptidase, site 2	Xp22.2-p2			RT-PCR
C19ORF25-APC2	Class V: Read-through	19p13.3	C19ORF25, chromosome 19 open reading frame 25	19p13.3			RT-PCR
WDR55-DND1	Class V: Read-through	5q31.3	WDR55, WD repeat domain 55	5q31.3	DND1, dead end homolog 1 (zebrafish)	RWPE	Long read, Short read, qRT-PCR
*SLC45A3-ELK4**	Class V: Read-through	1q32.1	SLC45A3, Solute carrier family 45 member 3	1q32.1	ELK4, ETS domain-containing protein	*Met-4*	Short read, qRT-PCR



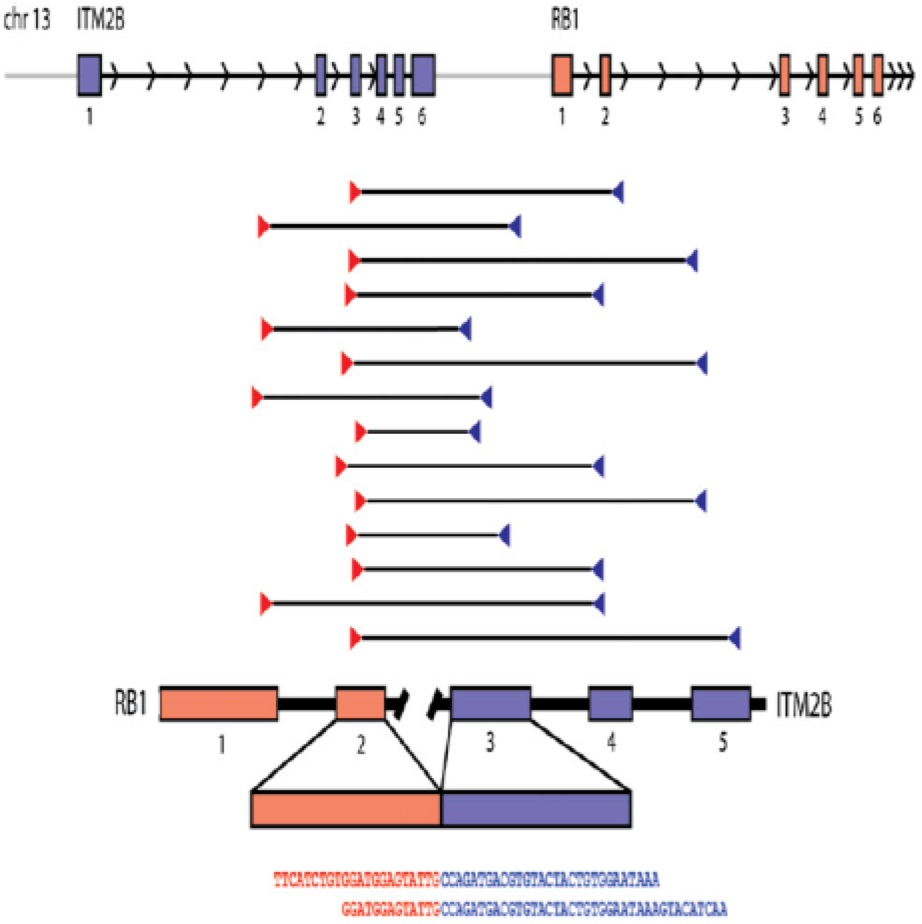


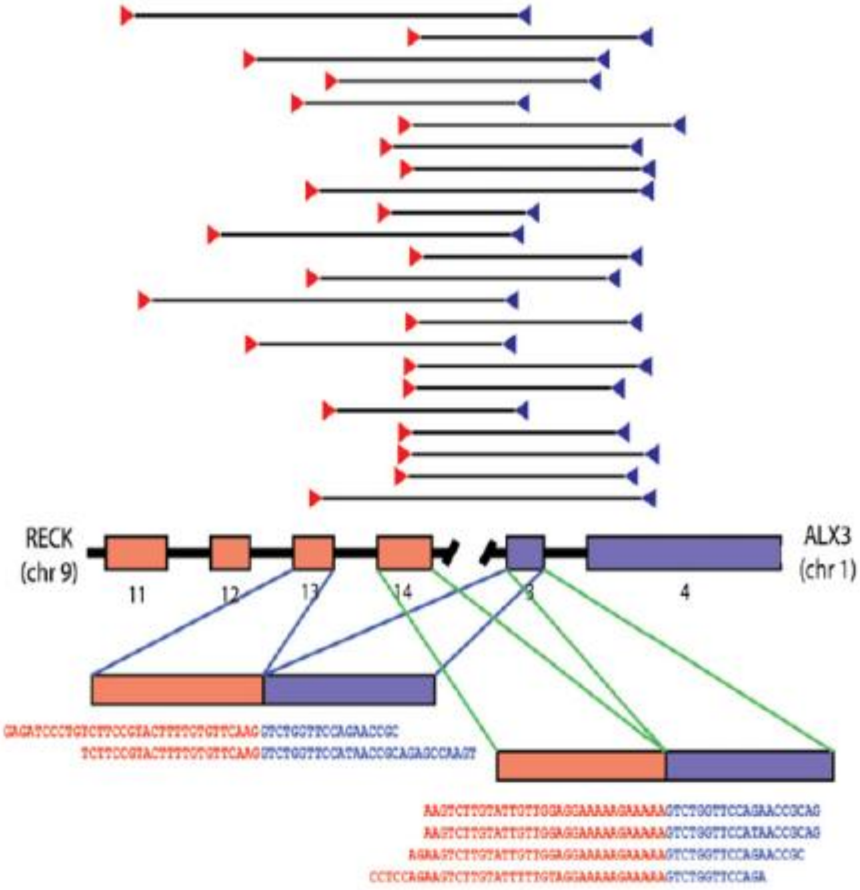
Research

Integrative analysis of the melanoma transcriptome

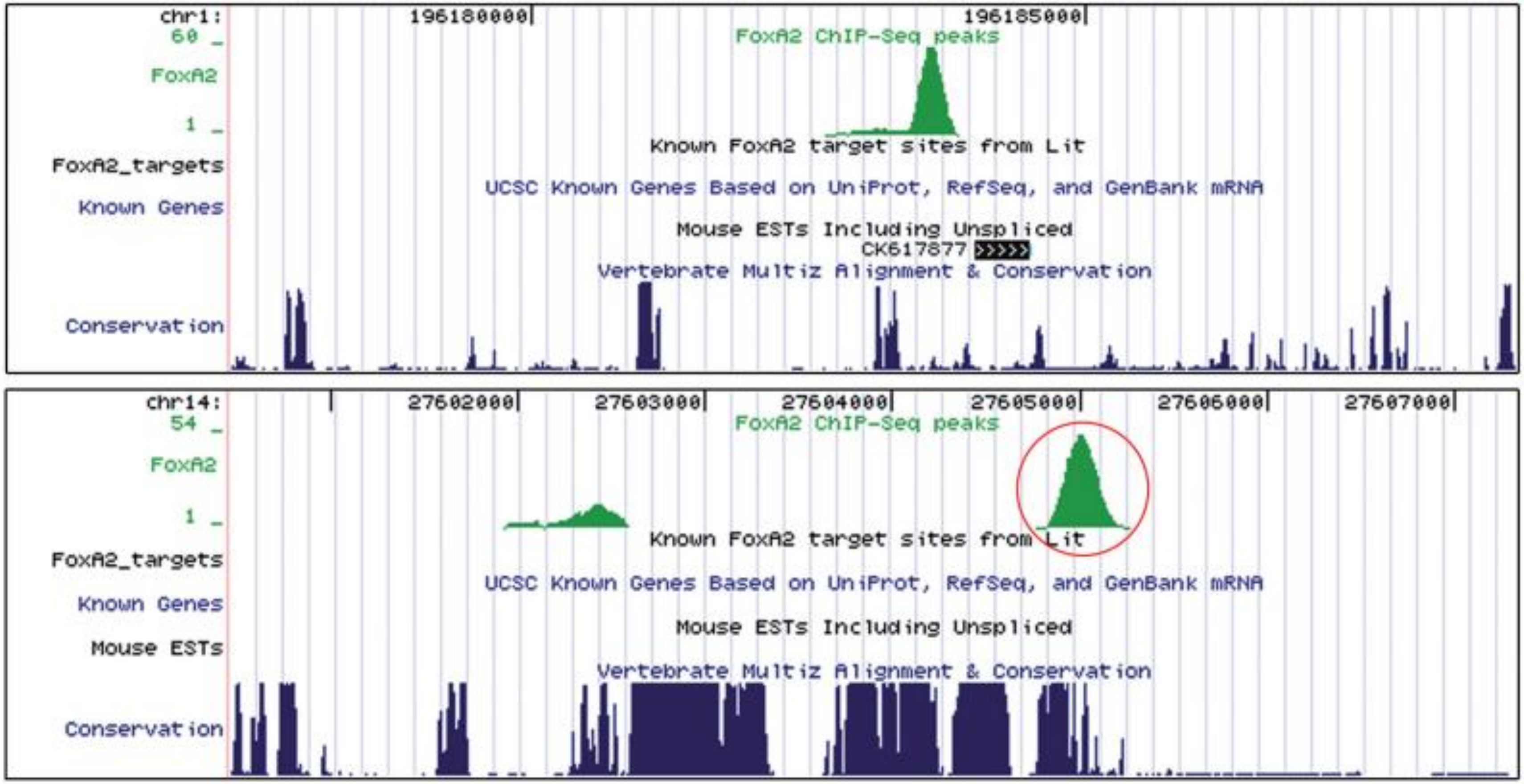
Michael F. Berger,[1] Joshua Z. Levin,[1] Krishna Vijayendran,[1,2] Andrey Sivachenko,[1]
Xian Adiconis,[1] Jared Maguire,[1] Laura A. Johnson,[1,2] James Robinson,[1] Roel G. Verhaak,[1,2]
Carrie Sougnez,[1] Robert C. Onofrio,[1] Liuda Ziaugra,[1] Kristian Cibulskis,[1] Elisabeth Laine,[3]
Jordi Barretina,[1] Wendy Winckler,[1] David E. Fisher,[4,5] Gad Getz,[1] Matthew Meyerson,[1,2,6]
David B. Jaffe,[1] Stacey B. Gabriel,[1] Eric S. Lander,[1,7,8] Reinhard Dummer,[3]
Andreas Gnirke,[1] Chad Nusbaum,[1] and Levi A. Garraway[1,2,6,9]

[1]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA; [2]Department of Medical Oncology, Dana-Farber





Foxa2 Binds in Unknown Regions



… two remote peaks in regions containing no known genes. The first peak (unknown #4) lies close to a mouse EST, while the second peak (unknown #5) overlaps and is close to highly conserved sequence regions.

Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals

Mitchell Guttman[1,2], Ido Amit[1], Manuel Garber[1], Courtney French[1], Michael F. Lin[1], David Feldser[3], Maite Huarte[1,6], Or Zuk[1], Bryce W. Carey[2,8], John P. Cassady[2,8], Moran N. Cabili[7], Rudolf Jaenisch[2,8], Tarjei S. Mikkelsen[1,4], Tyler Jacks[2,3], Nir Hacohen[1,9], Bradley E. Bernstein[1,10,11], Manolis Kellis[1,5], Aviv Regev[1,2], John L. Rinn[1,6,11]* & Eric S. Lander[1,2,7,8]*

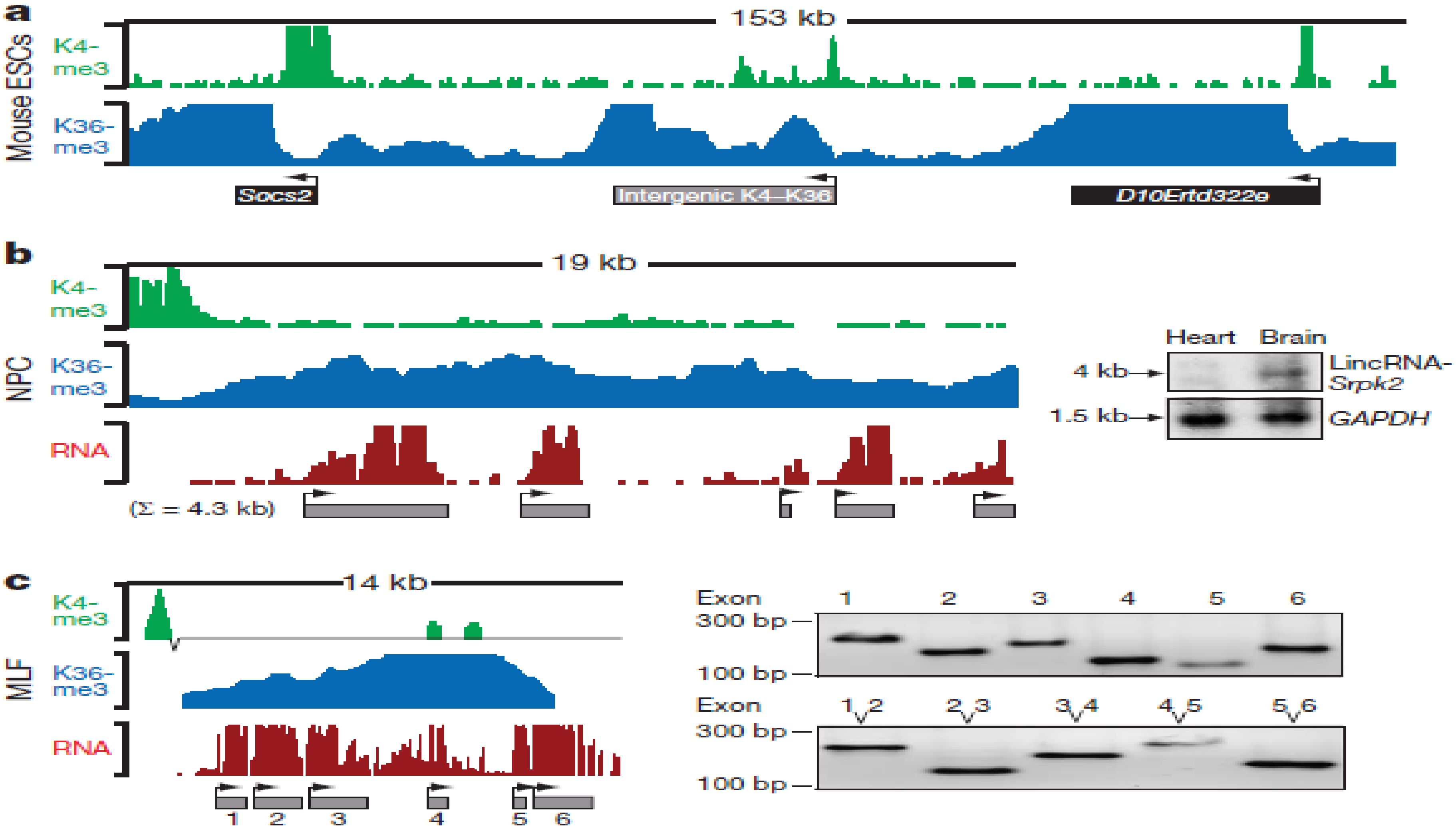
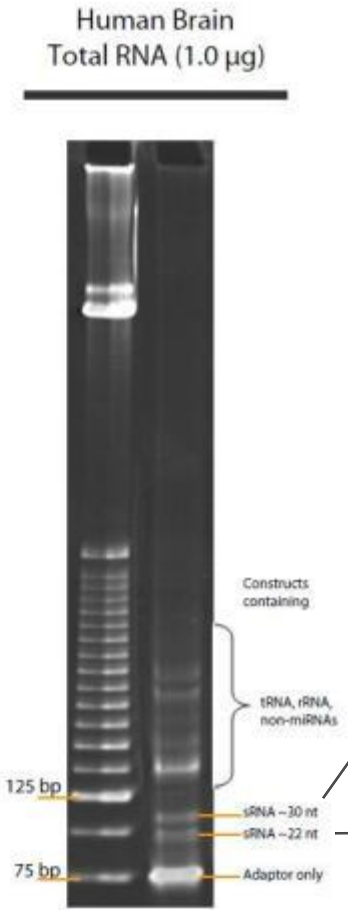
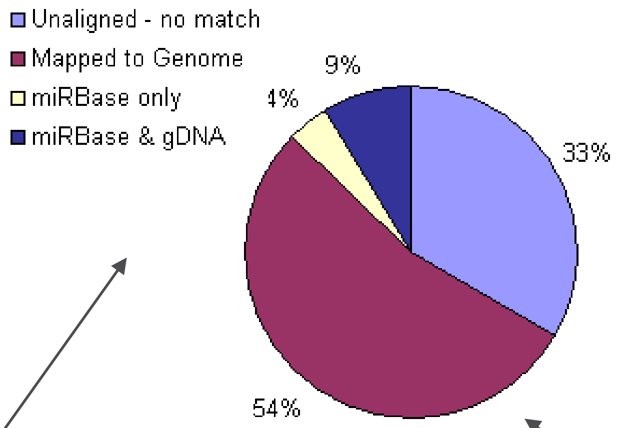


Figure 1 | Intergenic K4–K36 domains produce multi-exonic RNAs.

More to small RNA than known microRNAs

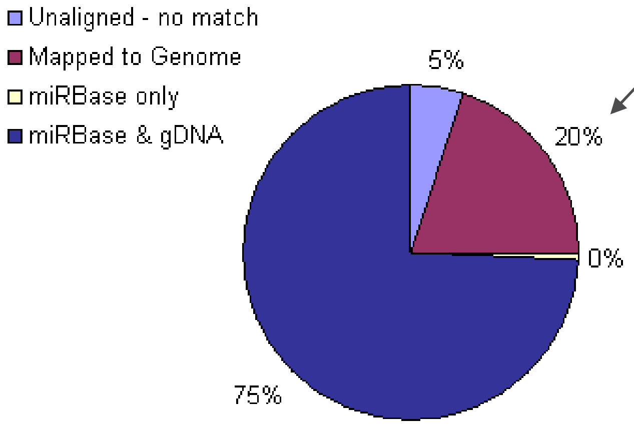


Non-coding RNA Fragment



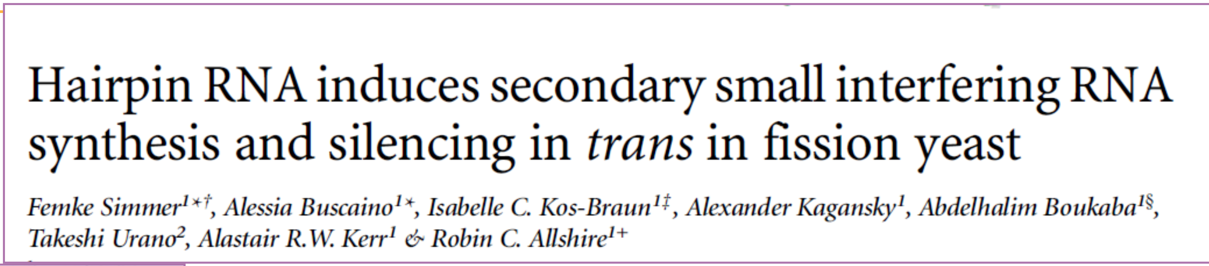
Potential
Novel Content

miRNA Fragment



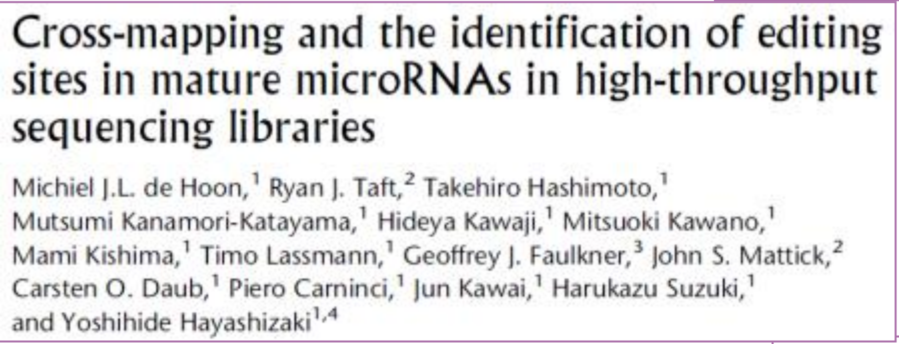
rRNA, tRNA, adaptor contamination	
Standard Protocol:	<3%
sRNA v1.5 30 nt band:	6%
sRNA v1.5 22 nt band:	8%

More to small RNAs ...



Hairpin RNA induces secondary small interfering RNA synthesis and silencing in *trans* in fission yeast

Femke Simmer[1†], Alessia Buscaino[1*], Isabelle C. Kos-Braun[1‡], Alexander Kagansky[1], Abdelhalim Boukaba[1§], Takeshi Urano[2], Alastair R.W. Kerr[1] & Robin C. Allshire[1+]*



Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries

Michiel J.L. de Hoon,[1] Ryan J. Taft,[2] Takehiro Hashimoto,[1] Mutsumi Kanamori-Katayama,[1] Hideya Kawaji,[1] Mitsuoki Kawano,[1] Mami Kishima,[1] Timo Lassmann,[1] Geoffrey J. Faulkner,[3] John S. Mattick,[2] Carsten O. Daub,[1] Piero Carninci,[1] Jun Kawai,[1] Harukazu Suzuki,[1] and Yoshihide Hayashizaki[1,4]

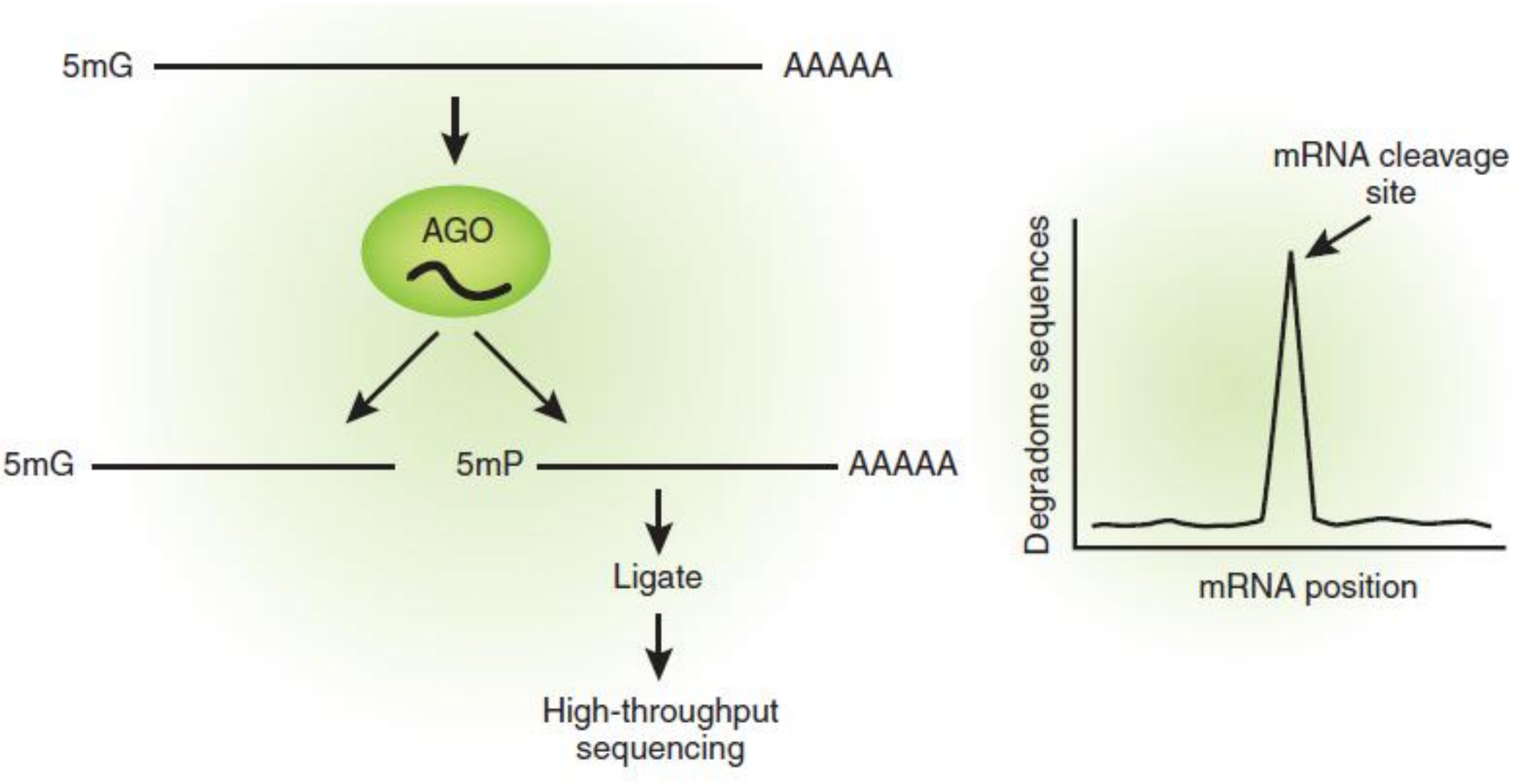
Tiny RNAs associated with transcription start sites in animals

Ryan J Taft[1], Evgeny A Glazov[2], Nicole Cloonan[1], Cas Simons[1], Stuart Stephen[1], Geoffrey J Faulkner[1], Timo Lassmann[3], Alistair R R Forrest[3,4], Sean M Grimmond[1], Kate Schroder[1], Katharine Irvine[1], Takahiro Arakawa[3], Mari Nakamura[3], Atsutaka Kubosaki[3], Kengo Hayashida[3], Chika Kawazu[3], Mitsuyoshi Murata[3], Hiromi Nishiyori[3], Shiro Fukuda[3], Jun Kawai[3], Carsten O Daub[3], David A Hume[1,5], Harukazu Suzuki[3], Valerio Orlando[6,7], Piero Carninci[3], Yoshihide Hayashizaki[3] & John S Mattick[1]

Post-transcriptional regulation of miR-27 in murine cytomegalovirus infection

AMY H. BUCK,[1,2] JONATHAN PEROT,[3] MICHAEL A. CHISHOLM,[1,2] DIWAKAR S. KUMAR,[1,2] LEE TUDDENHAM,[3] VALÉRIE COGNAT,[4] LISA MARCINOWSKI,[5] LARS DÖLKEN,[5] and SÉBASTIEN PFEFFER[3]

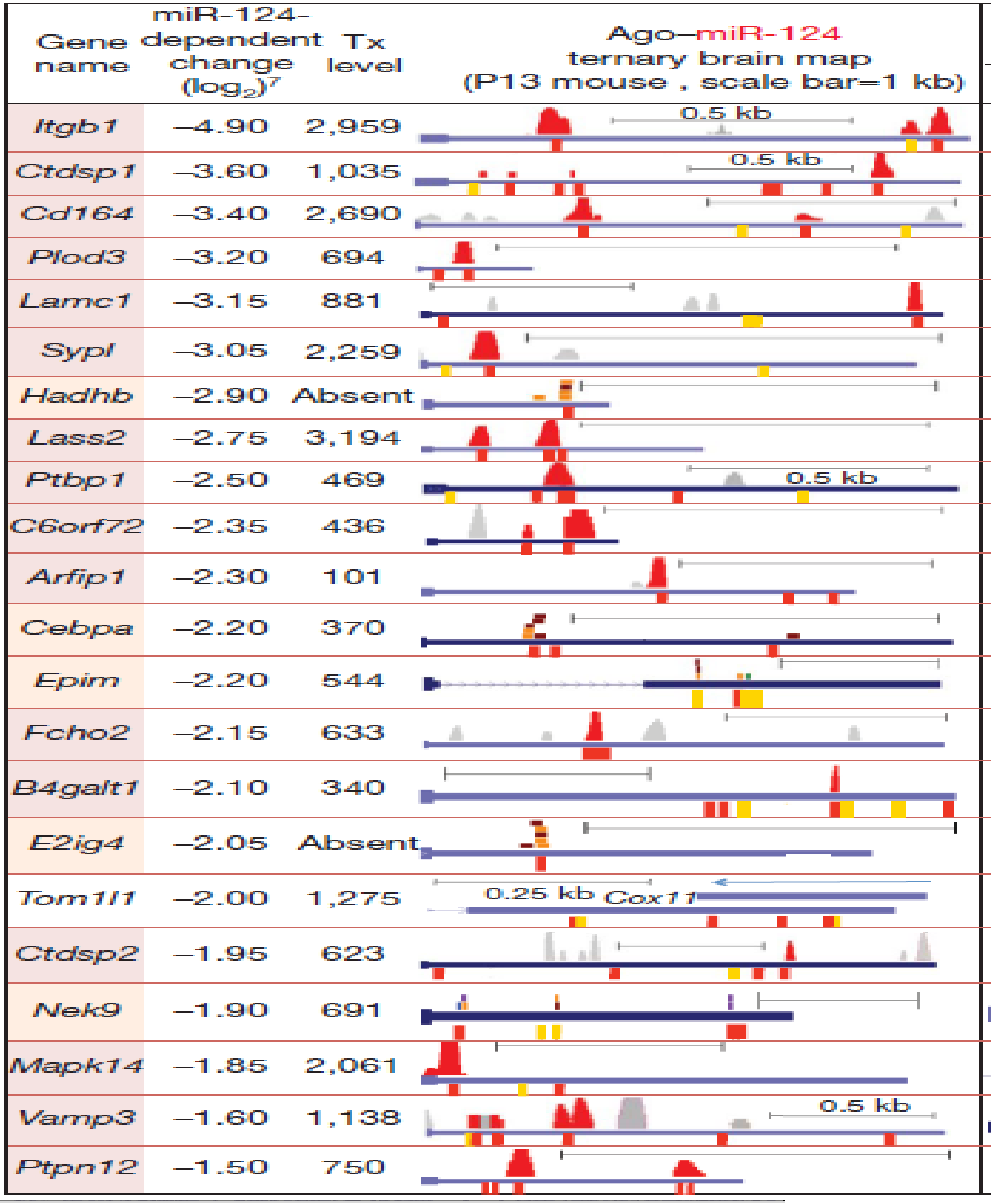
Identification of micro-RNA targets



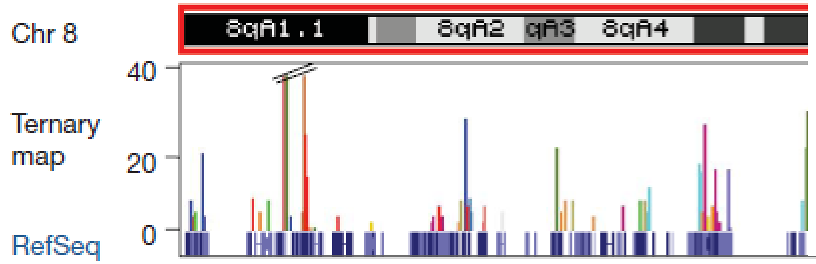
Henderson, 2009

Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Chi SW, Zang JB, Mele A, Darnell RB. *Nature (2009) 460: 479-86.*

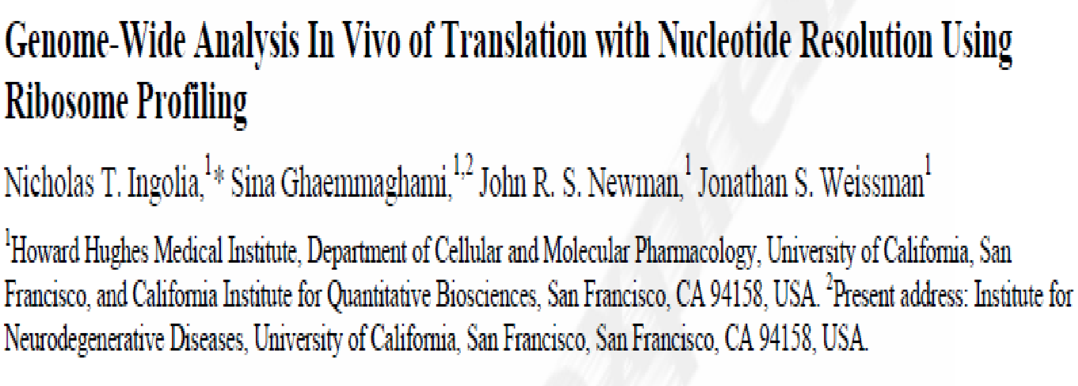
Covalently crosslink native argonaute protein–RNA gives you Ago–miRNA and Ago–mRNA binding sites.



Gene name	miR-124-dependent change $(\log_2)^7$	Tx level	Ago–miR-124 ternary brain map (P13 mouse, scale bar=1 kb)
Itgb1	–4.90	2,959	0.5 kb
Ctdsp1	–3.60	1,035	0.5 kb
Cd164	–3.40	2,690	
Plod3	–3.20	694	
Lamc1	–3.15	881	
Sypl	–3.05	2,259	
Hadhb	–2.90	Absent	
Lass2	–2.75	3,194	
Ptbp1	–2.50	469	0.5 kb
C6orf72	–2.35	436	
Arfip1	–2.30	101	
Cebpa	–2.20	370	
Epim	–2.20	544	
Fcho2	–2.15	633	
B4galt1	–2.10	340	
E2ig4	–2.05	Absent	
Tom1l1	–2.00	1,275	0.25 kb *Cox11*
Ctdsp2	–1.95	623	
Nek9	–1.90	691	
Mapk14	–1.85	2,061	
Vamp3	–1.60	1,138	0.5 kb
Ptpn12	–1.50	750	



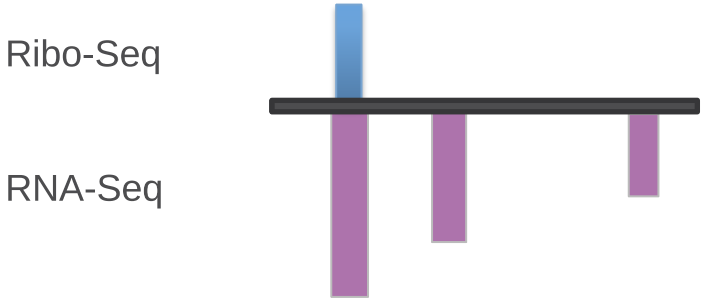
Ribo-Seq.
RNA engaged in translation

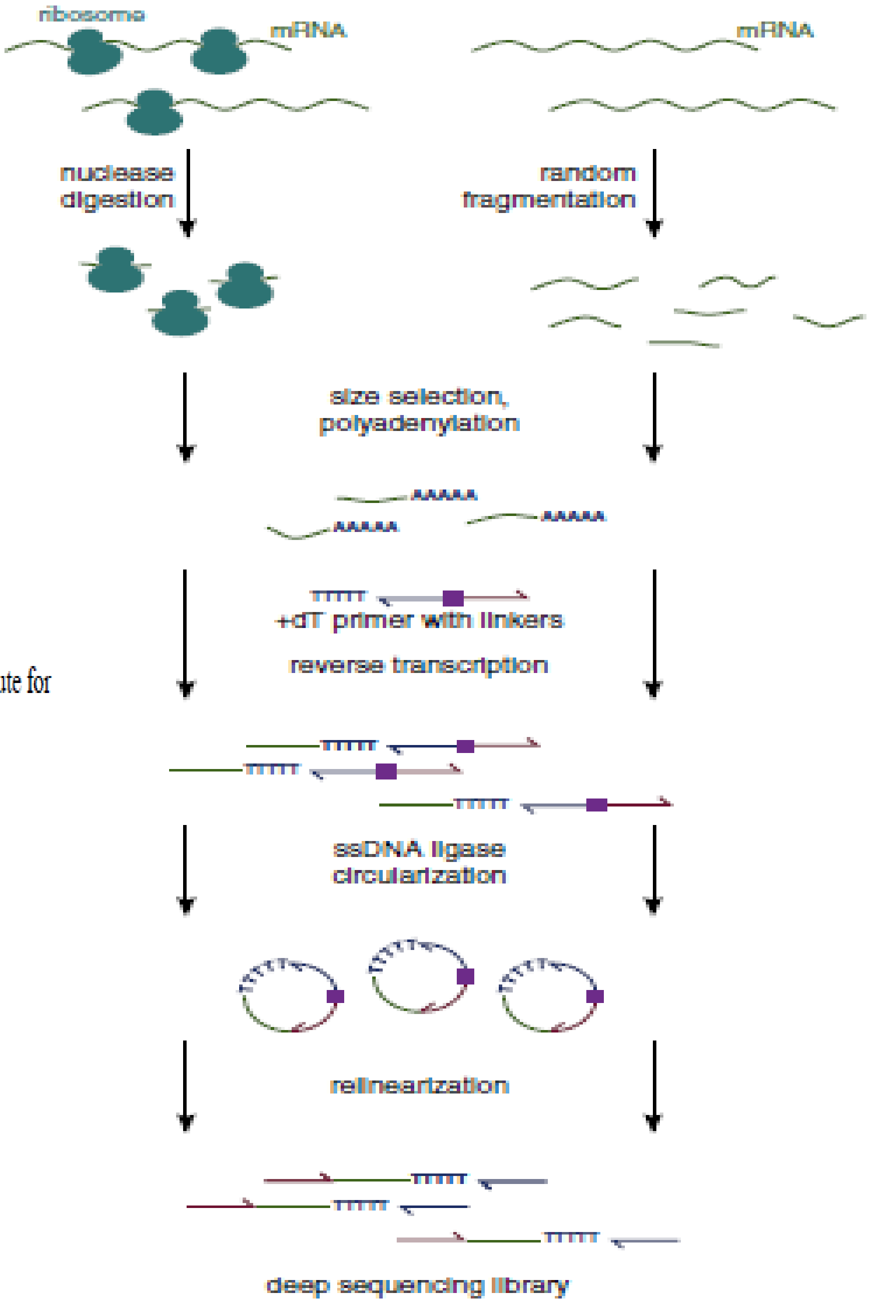


Genome-Wide Analysis In Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling

Nicholas T. Ingolia,[1]* Sina Ghaemmaghami,[1,2] John R. S. Newman,[1] Jonathan S. Weissman[1]

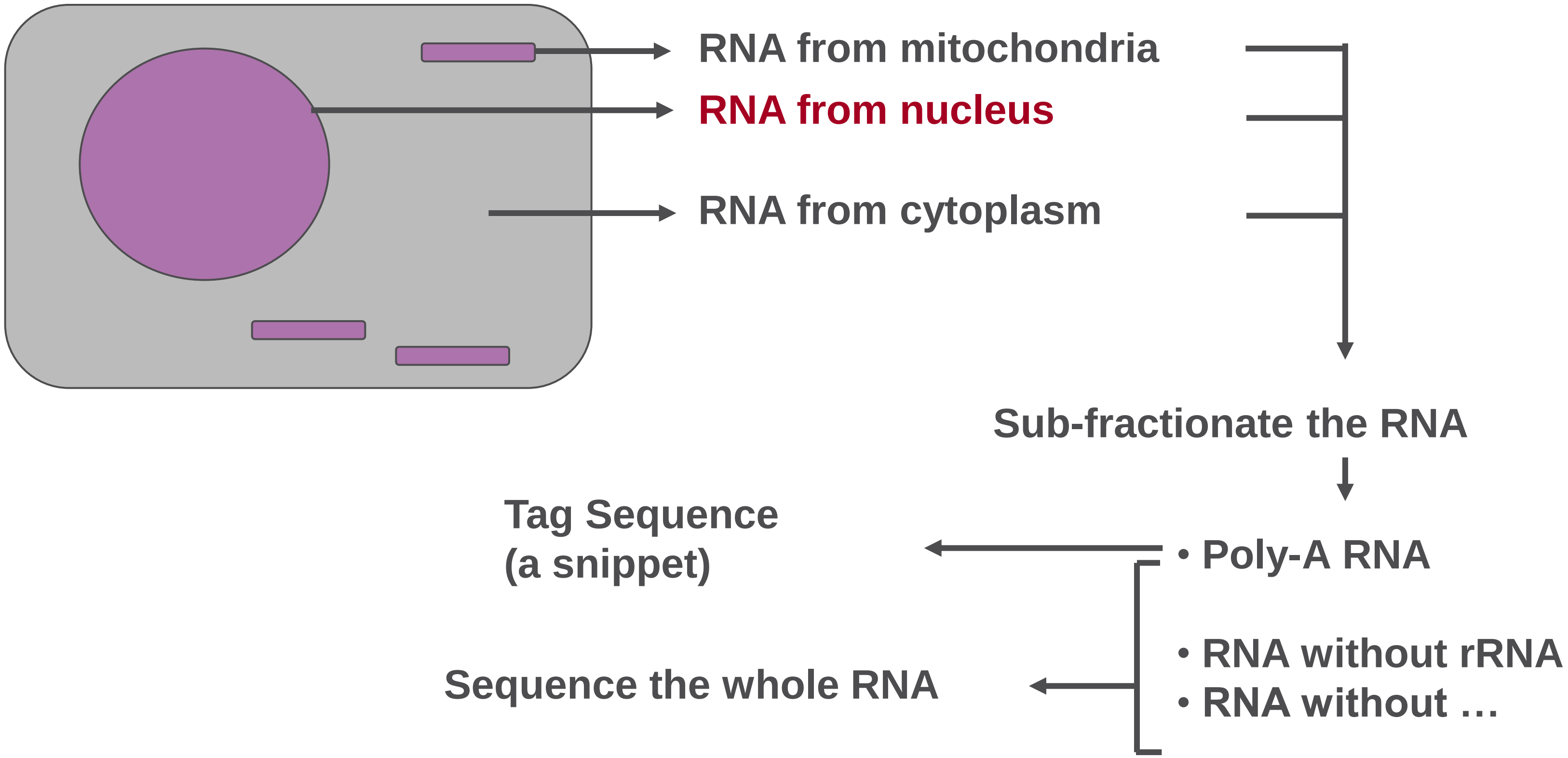
[1]Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, University of California, San Francisco, and California Institute for Quantitative Biosciences, San Francisco, CA 94158, USA. [2]Present address: Institute for Neurodegenerative Diseases, University of California, San Francisco, San Francisco, CA 94158, USA.





RNA Normalization

Genome Analyzer Transcriptional Analysis



RNA from mitochondria

RNA from nucleus

RNA from cytoplasm

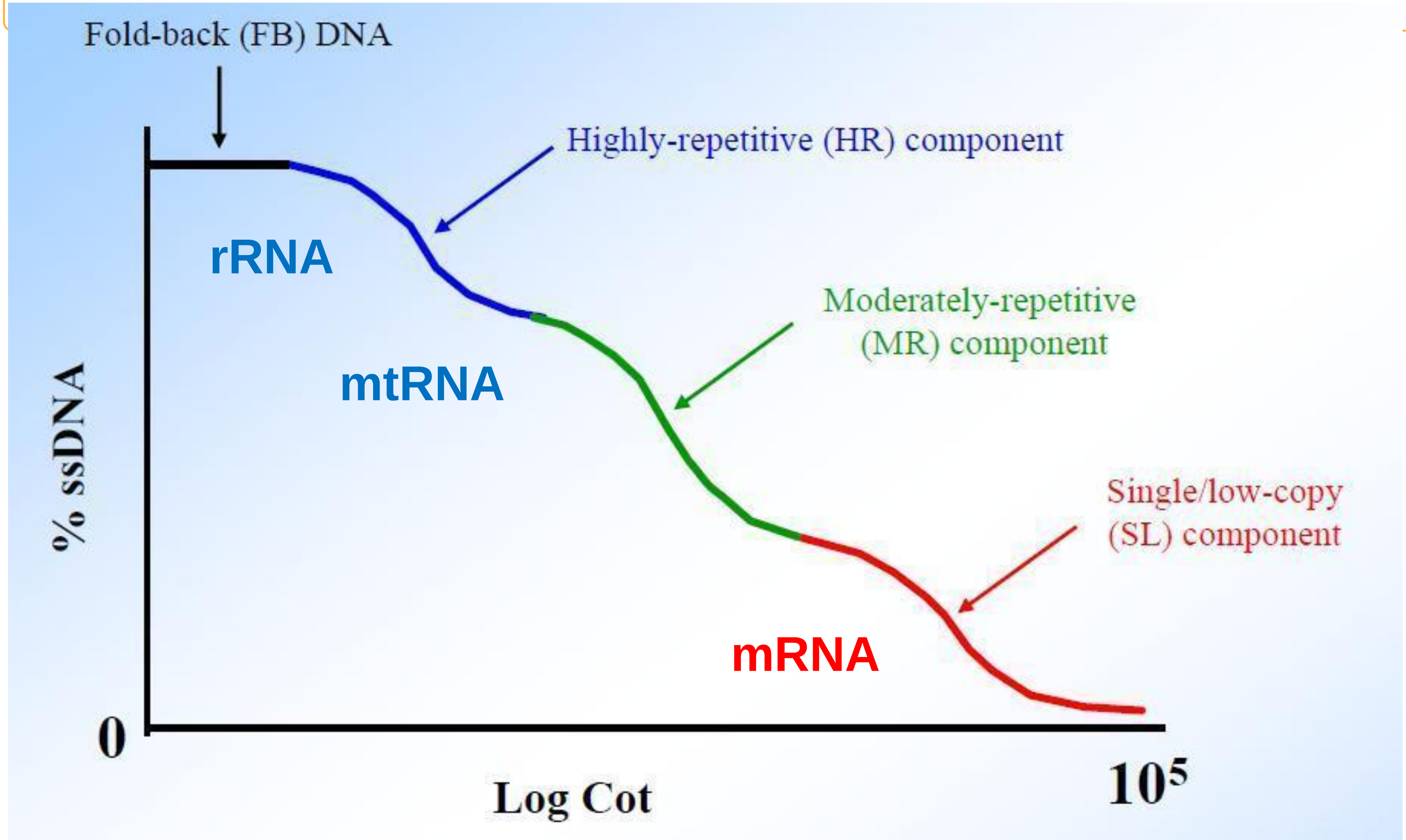
Sub-fractionate the RNA

- Poly-A RNA
- RNA without rRNA
- RNA without …

Tag Sequence
(a snippet)

Sequence the whole RNA

C_ot Curves and cDNA Normalization

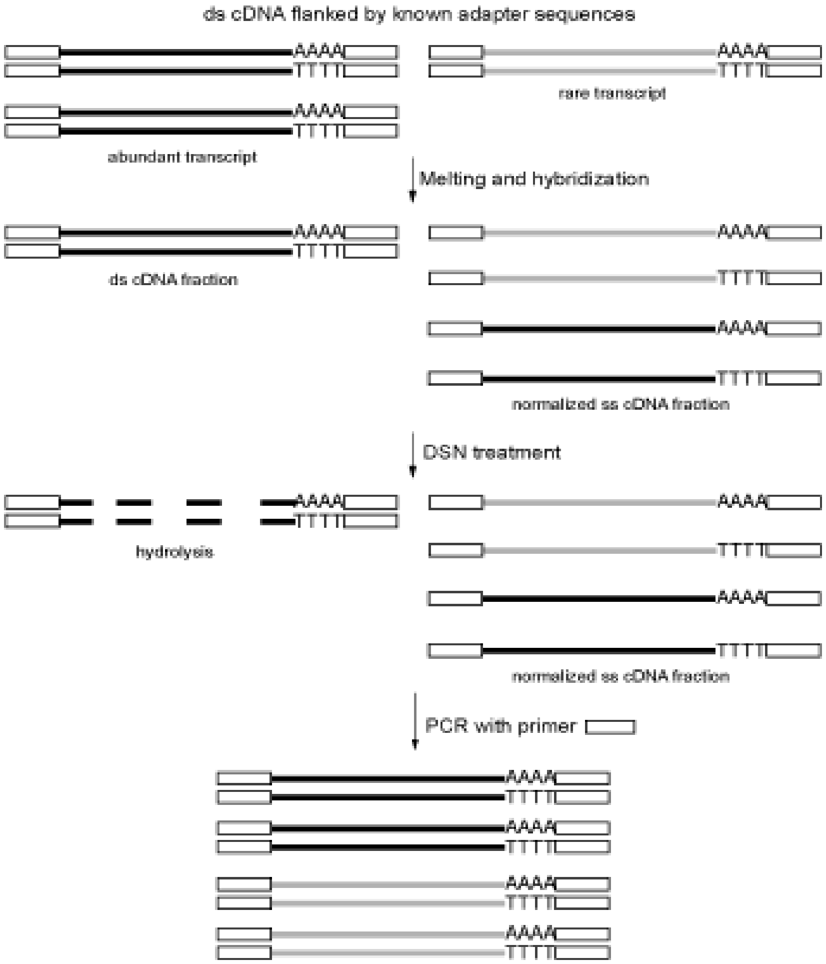


Double strand nuclease normalization

mRNA-Seq	***Total* RNA-Seq**
Purified Total RNA	**Purified Total RNA**
Poly-A Selection	⇩
RNA Fragmentation	**RNA Fragmentation**
cDNA Synthesis*	**cDNA Synthesis***
Adapter Ligation & PCR	**Adapter Ligation & PCR**
	DSN Normalization

DSN Library Normalization Process Overview

- Library is melted at 98ºC
- Library is hybridized at 68ºC
- DSN Cleavage
- PCR to enrich library
- Grow Clusters and Sequence

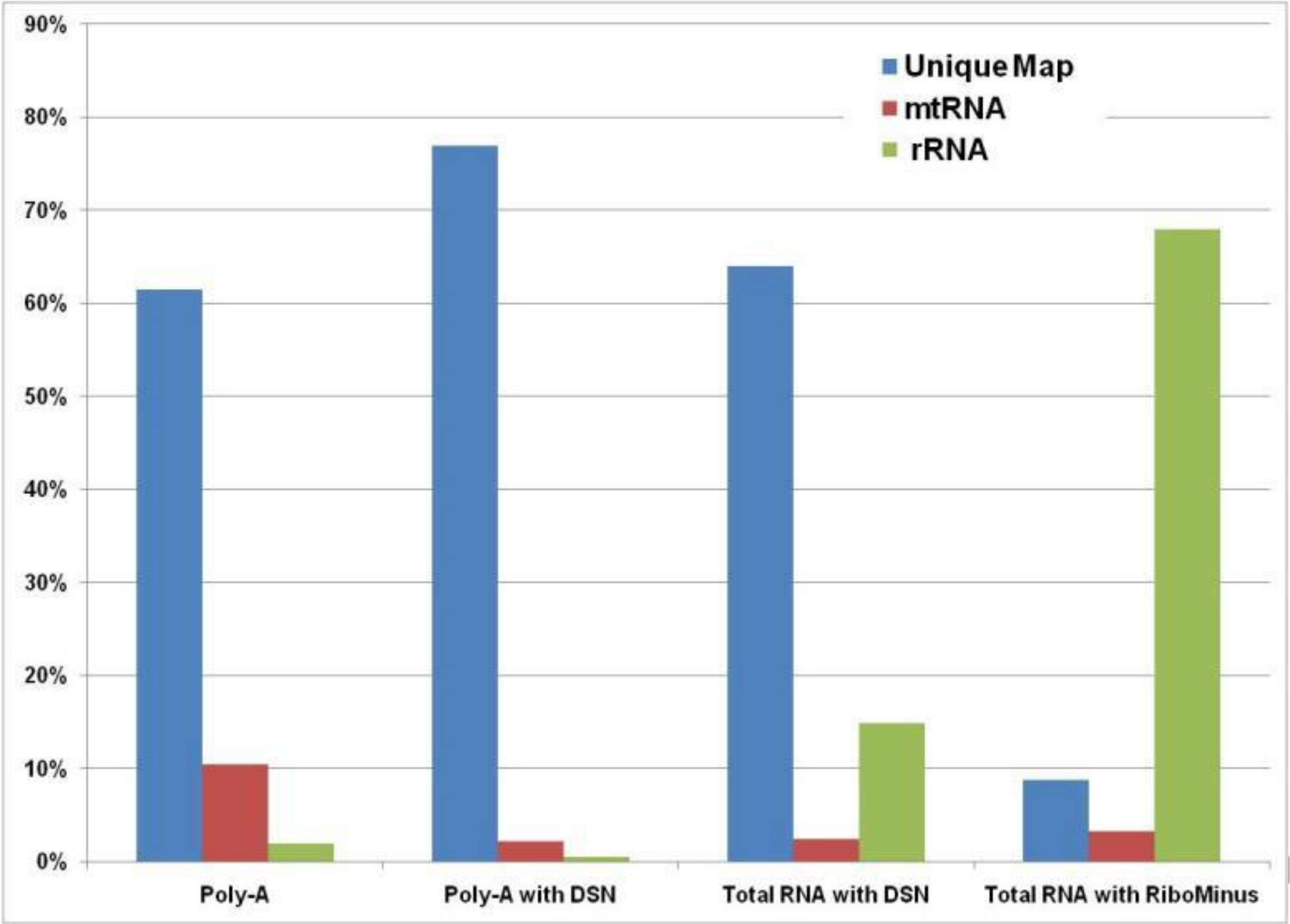


From Evrogen.com

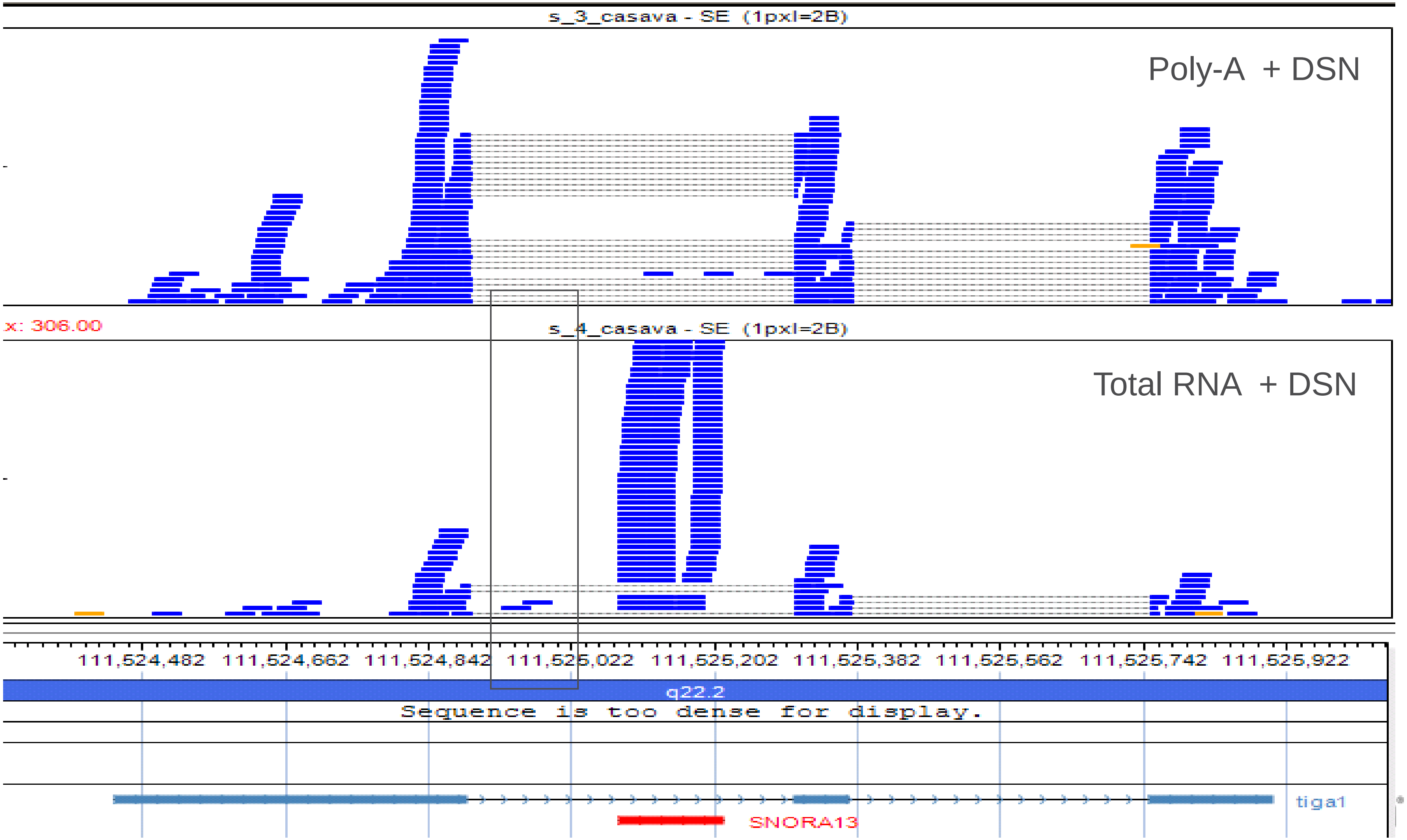
Total RNA-Seq Method Overview

- **RNA from any organism – bacteria, plants, animals**
- **Less than 100 nanograms of total RNA**
- **Full-length cDNA coverage of *all* RNA molecules**
- **Works with low quality RNA – even FFPE preparations**
- **Normalization protocol can be used with most RNA-Seq protocols**

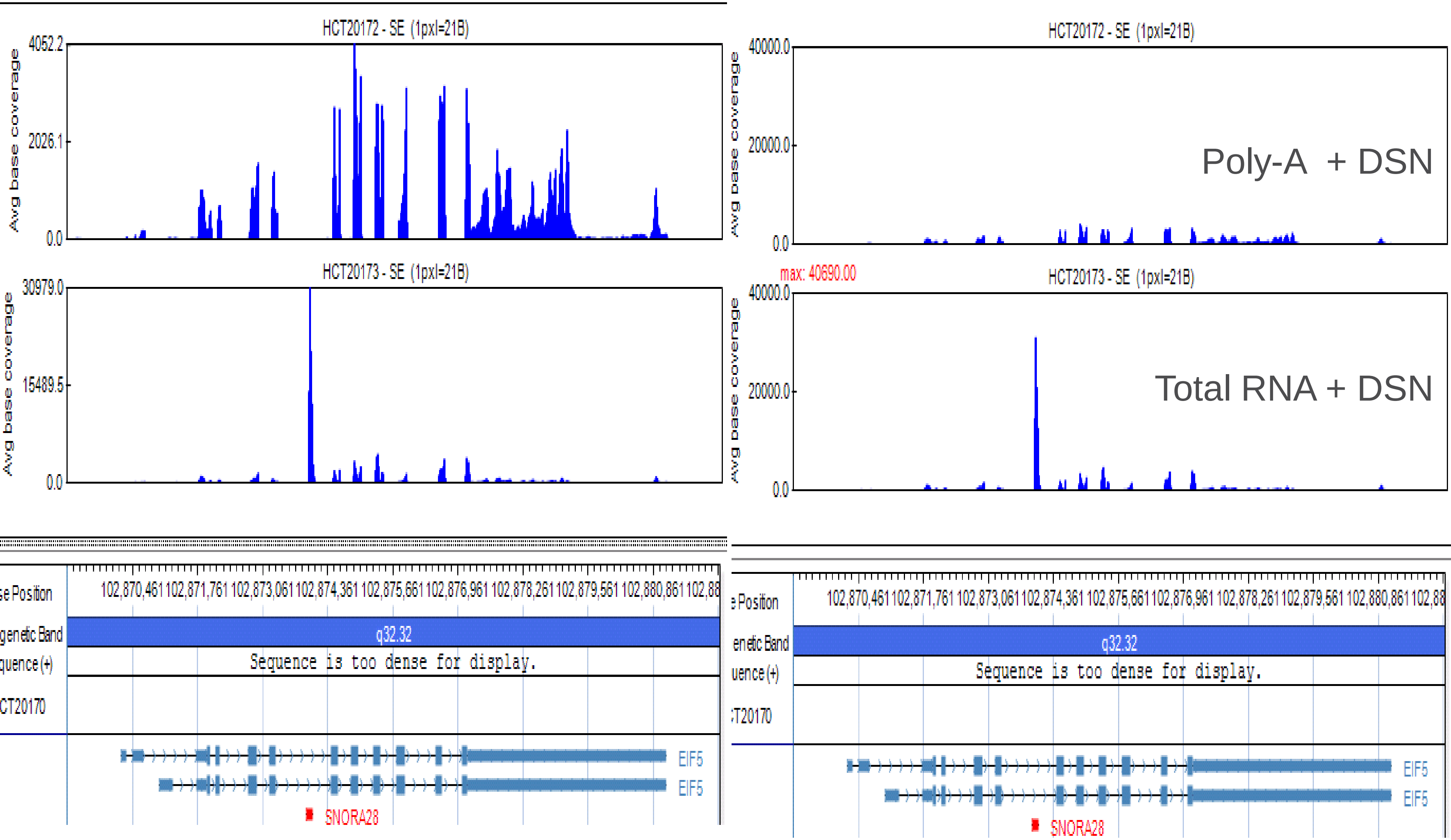
Comparison of Output from Different Protocols



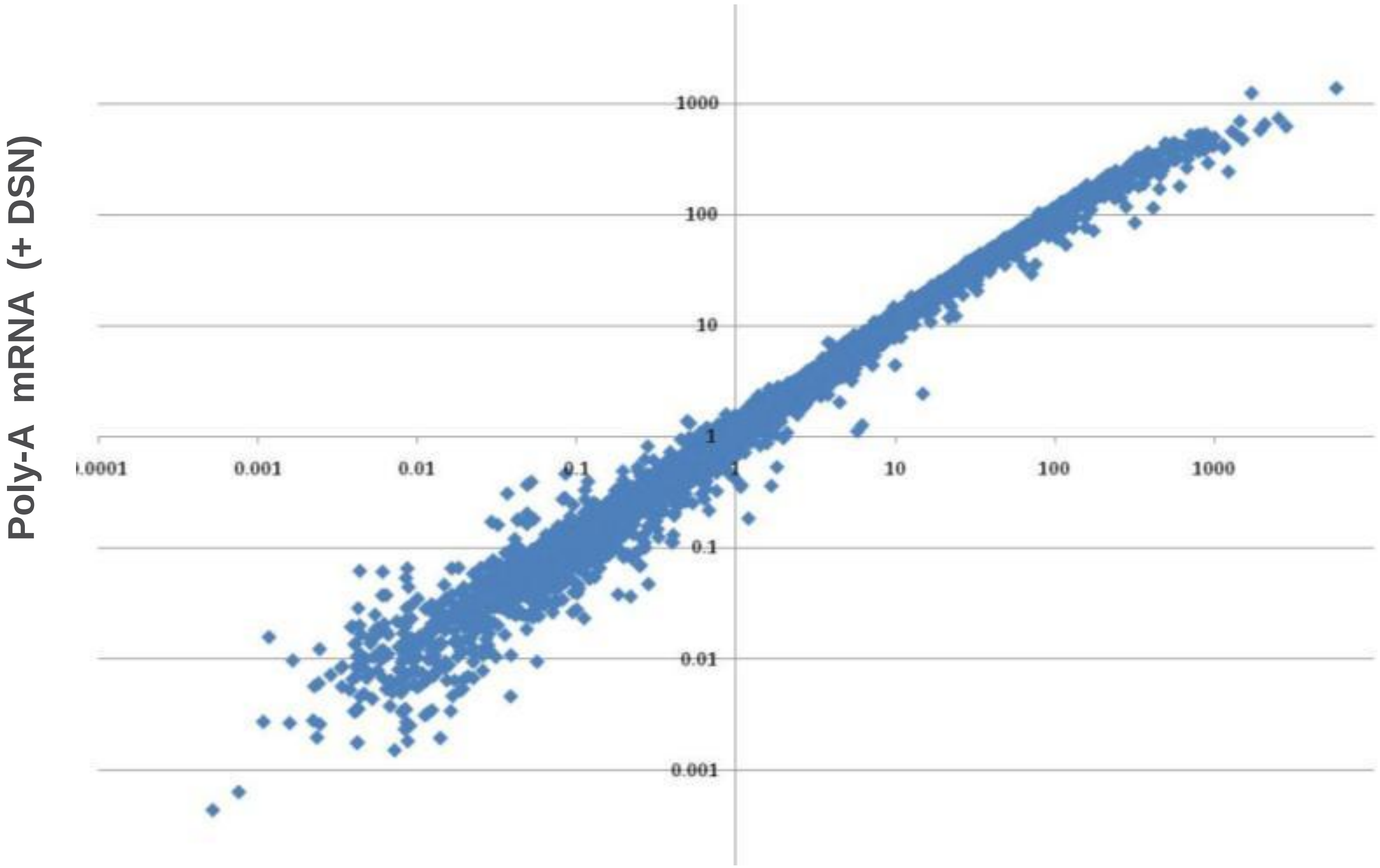
Many ncRNAs are often transcribed from the same strand as mRNA



Many non-coding RNAs are seen with *Total* RNA-Seq



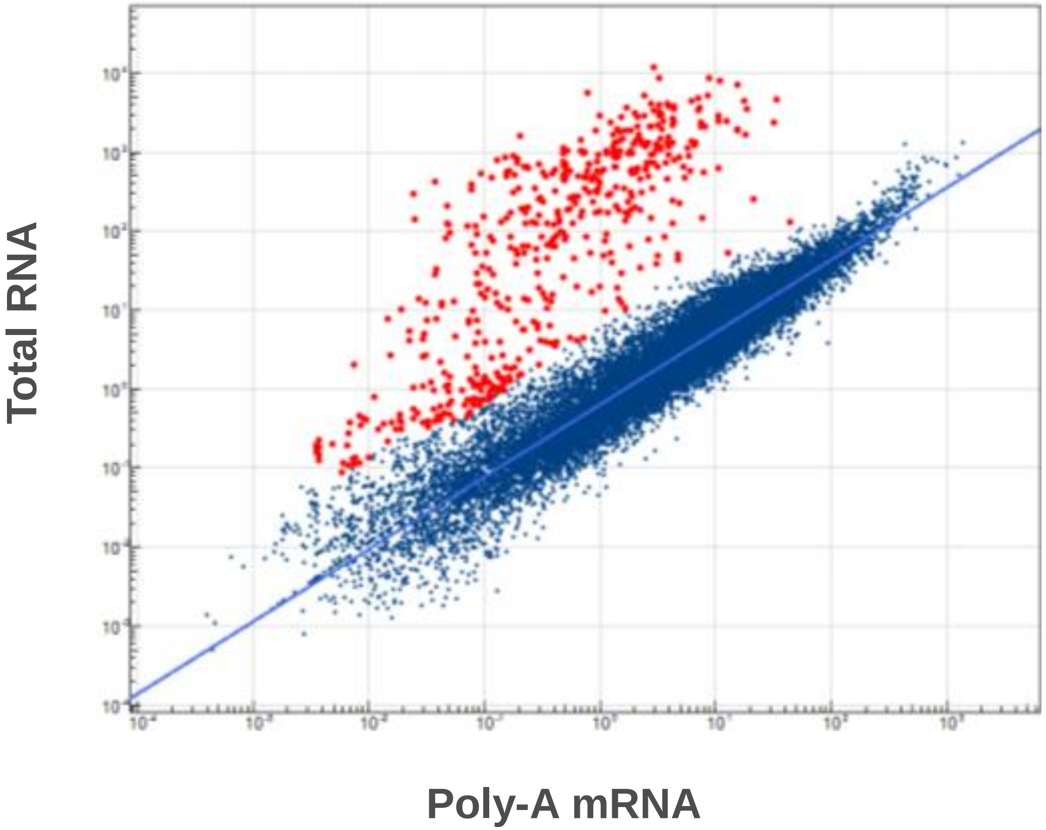
DSN Normalization has little effect on most mRNAs



Poly-A mRNA (no DSN)

Many non-coding RNAs are observed with *Total* RNA-Seq Method

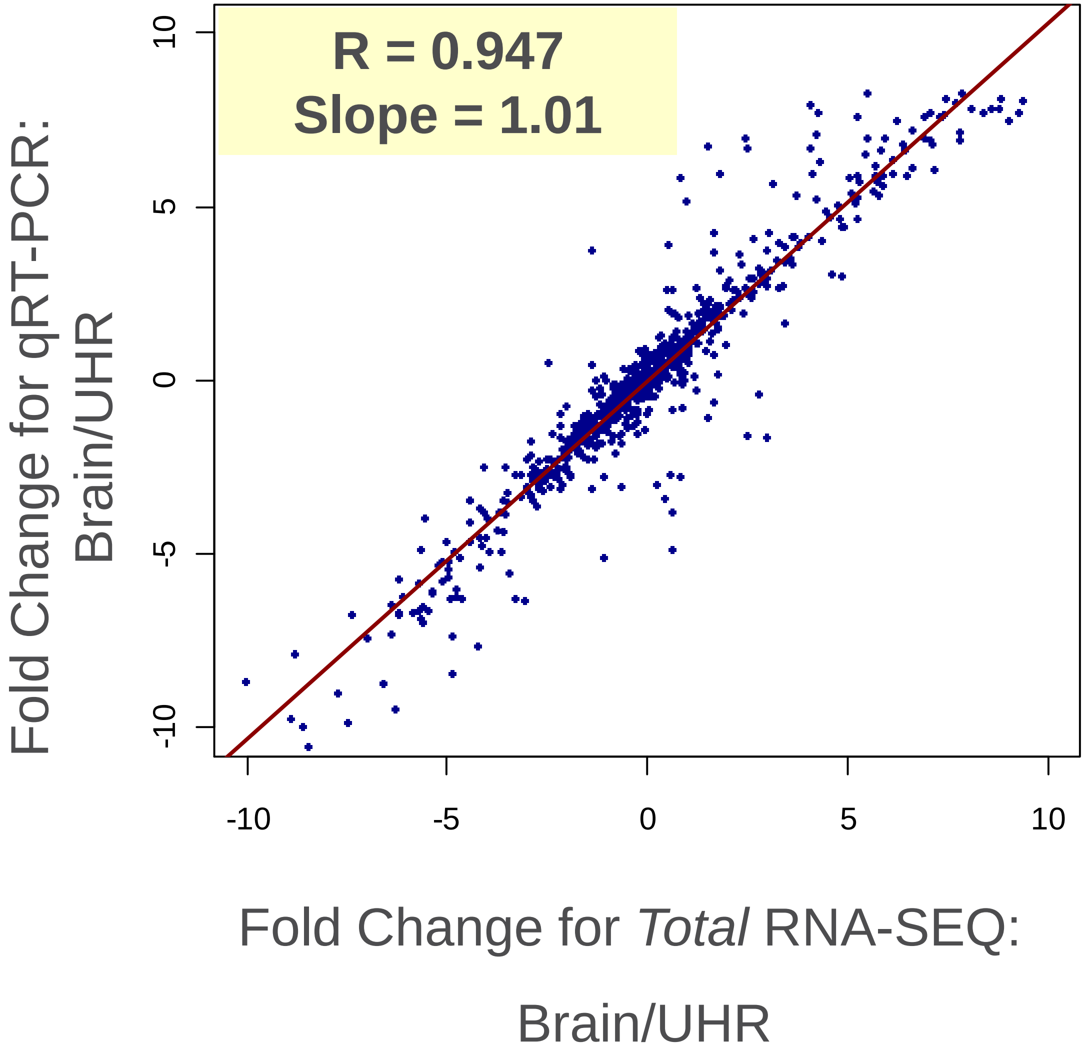
Total RNA vs. Poly-A mRNA



Plot shows gene level count data

for all RefSeq Genes

Fold Change of Total RNA-Seq vs. qRT-PCR



Key Variables:

RNA is degraded

No Poly-A Selection

+ DSN Normalization

0.1 µg input total RNA

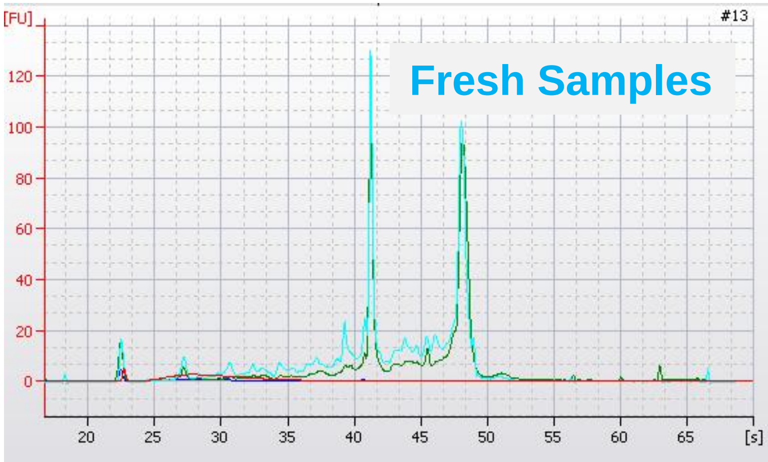
RNA-Seq in FFPE

RNA-Seq Protocol Variations

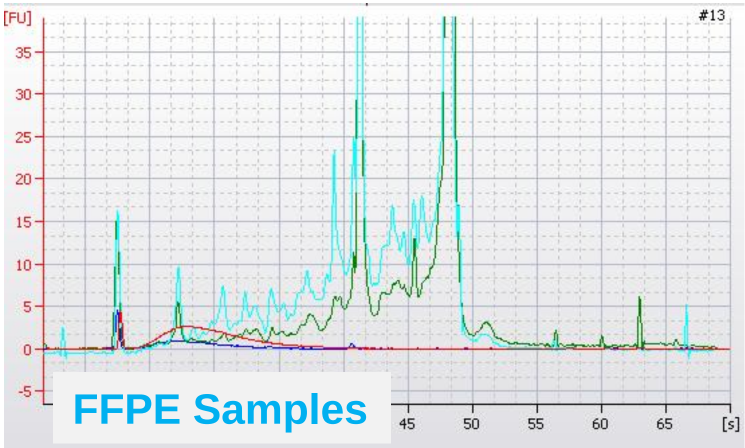
mRNA-Seq	*Total* RNA-Seq	*Total* RNA-Seq
Purified Total RNA	Purified Total RNA	Total RNA from FFPE
Poly-A Selection	↓	↓
RNA Fragmentation	**RNA Fragmentation**	↓
cDNA Synthesis*	cDNA Synthesis*	cDNA Synthesis*
Adapter Ligation & PCR	Adapter Ligation & PCR	Adapter Ligation & PCR
	Normalization	**Normalization**

Total RNA-Seq protocol on FFPE Samples

- Two Cancer Samples
  - Her2 and Basal breast cancer sub-types
- Two types of RNA preparations
  - Total RNA from fresh/frozen tissue and highly degraded FFPE RNA
- RIN of the fresh samples are 9.3 and 8.3

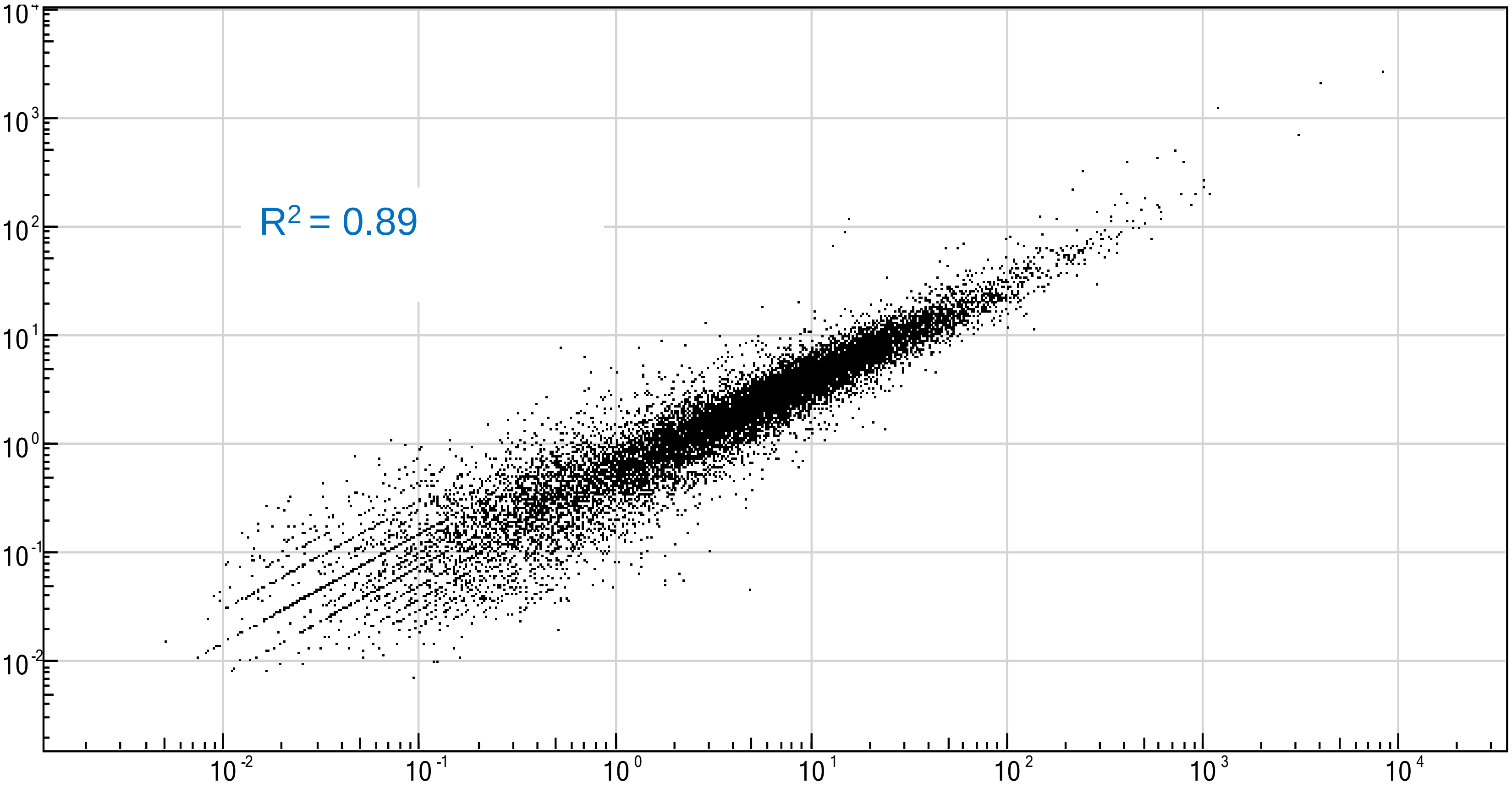


Zoom in



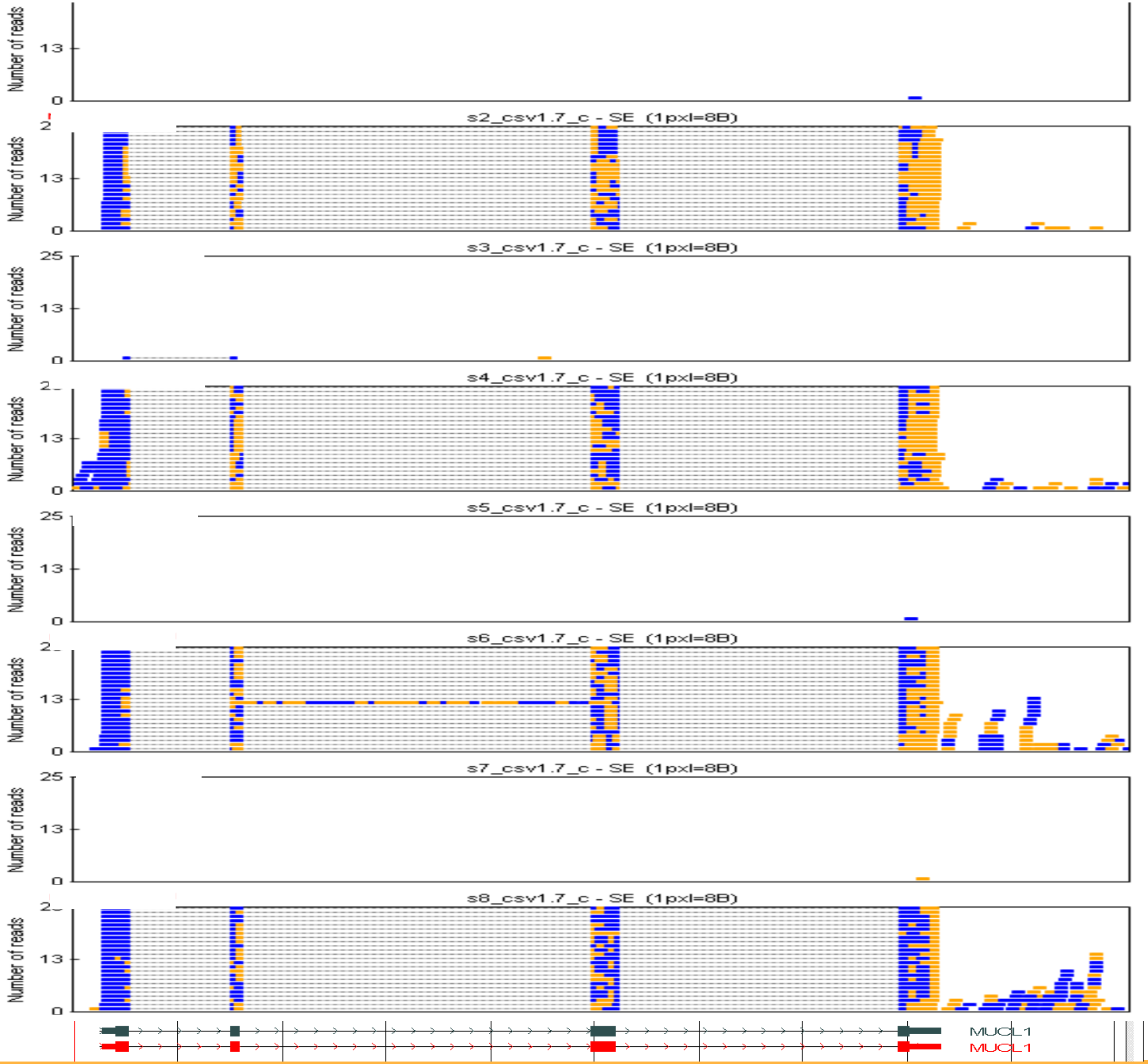
Collaboration with Chuck Perou - UNC

Pair-wise Comparison of FFPE



FFPE *Total* RNA-Seq vs. **Fresh** *Total* RNA-Seq

MUCL1(mucin-like 1 gene)



FFPE - Basal

FFPE - Her2

Fresh mRNA – Basal

Fresh mRNA - Her2

Fresh Total RNA - Basal

Replicate 1

Fresh Total RNA – Her2

Fresh Total RNA - Basal

Replicate 2

Fresh Total RNA – Her2

Whole Transcriptome Studies on HiSeq 2000 System

INDIVIDUAL HUMAN TISSUES

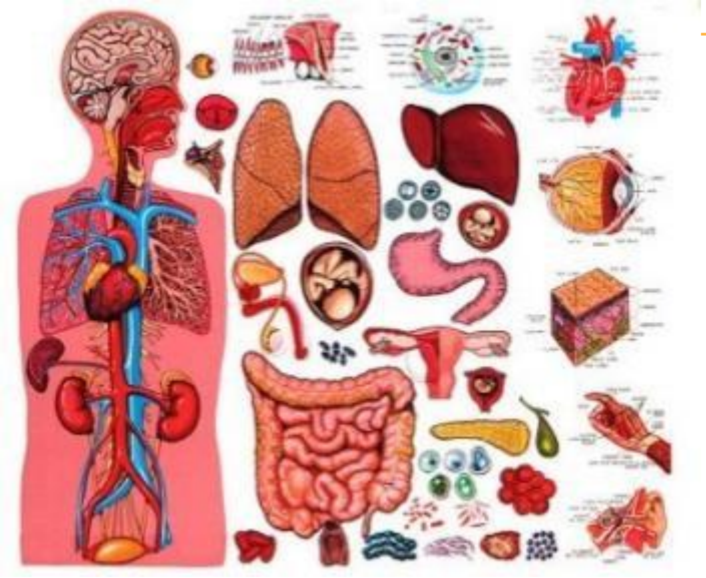
16 tissues, 1 lane per tissue

Standard Poly-A mRNA-Seq Library Preps

TOTAL HUMAN TRANSCRIPTOME

Equal Mixture of 16 Human Tissues

Strand-specific mRNA-Seq

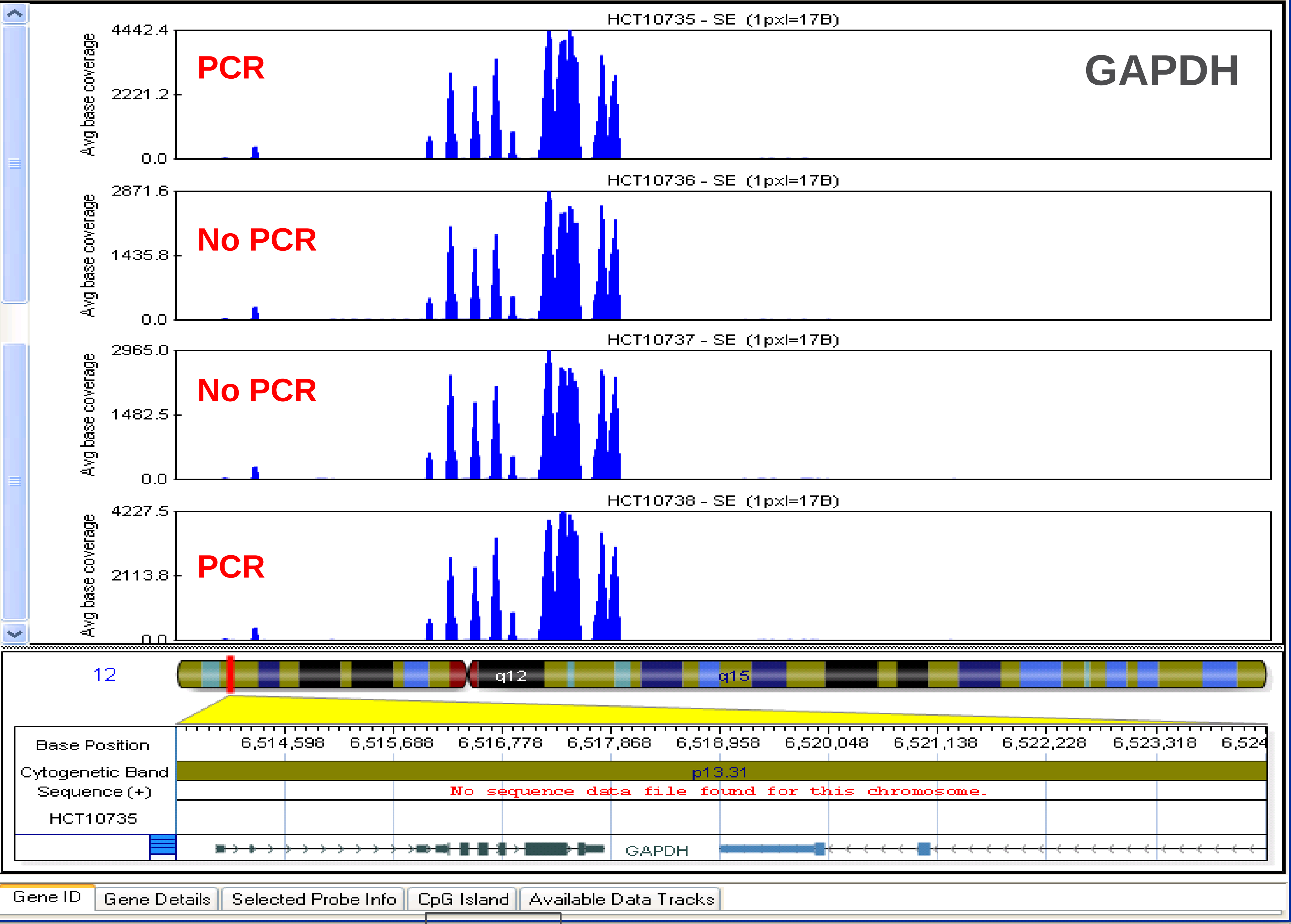


Body Map 2.0 Project
Adrenal, Adipose, Brain, Breast, Colon, Heart, Kidney, Liver, Lung, Lymph Node, Ovary, Prostate, Skeletal Muscle, Testis, Thyroid, White Blood Cells

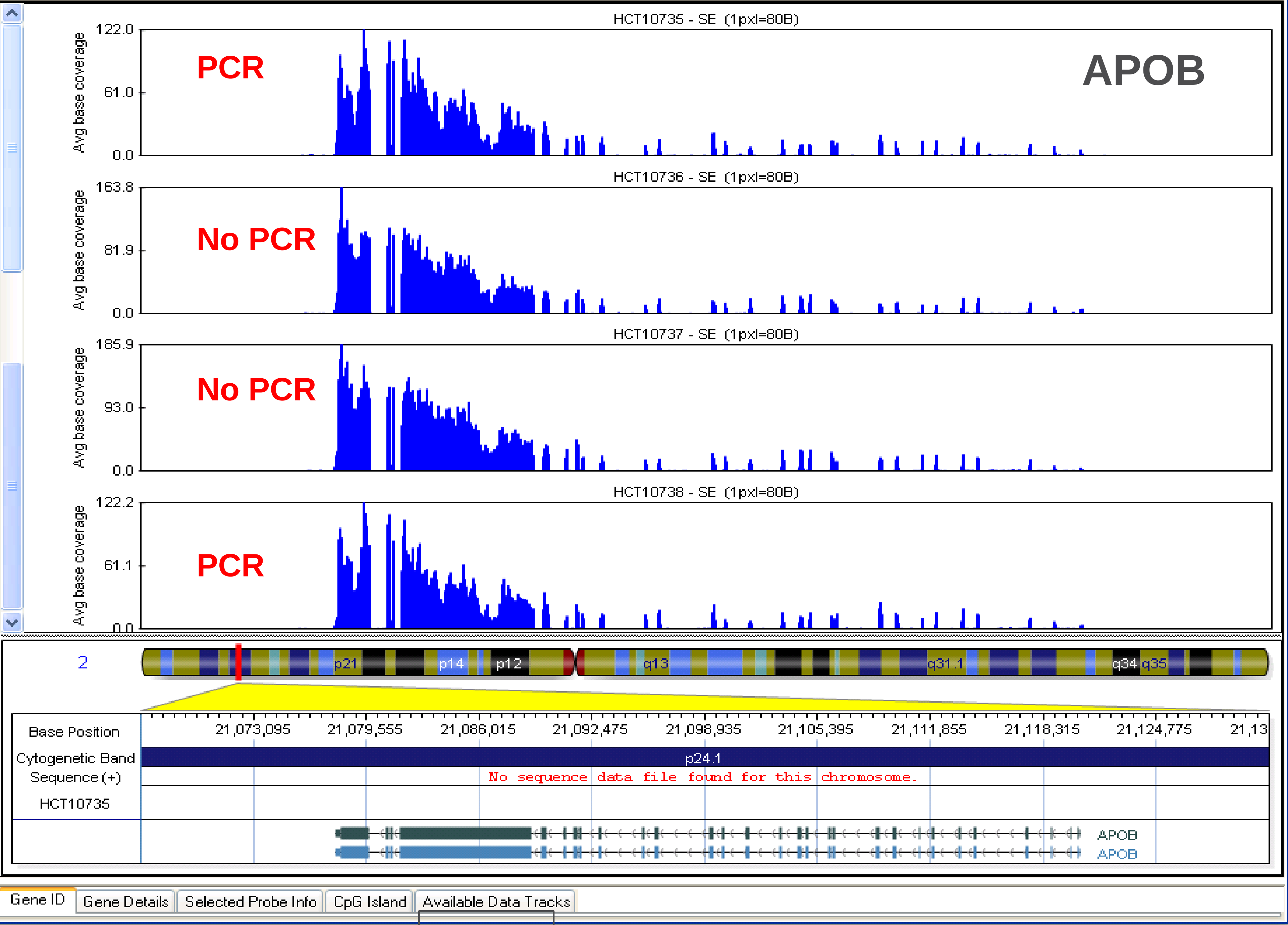
SUMMARY OF RUN METRICS

	Format	PF reads	Average Error Rates (PhiX spike-ins)	Yield
Individual human tissues	2 X 50	2.55 B	0.22%, 0.44%	>120 Gb in <5 days
Individual human tissues	1 X 75	1.27 B	0.37%, 0.52%	>95 Gb in <4 days
Total human transcriptome	1 X 100	1.19 B	0.67%, 0.81%	>115 Gb in <5 days

PCR-free

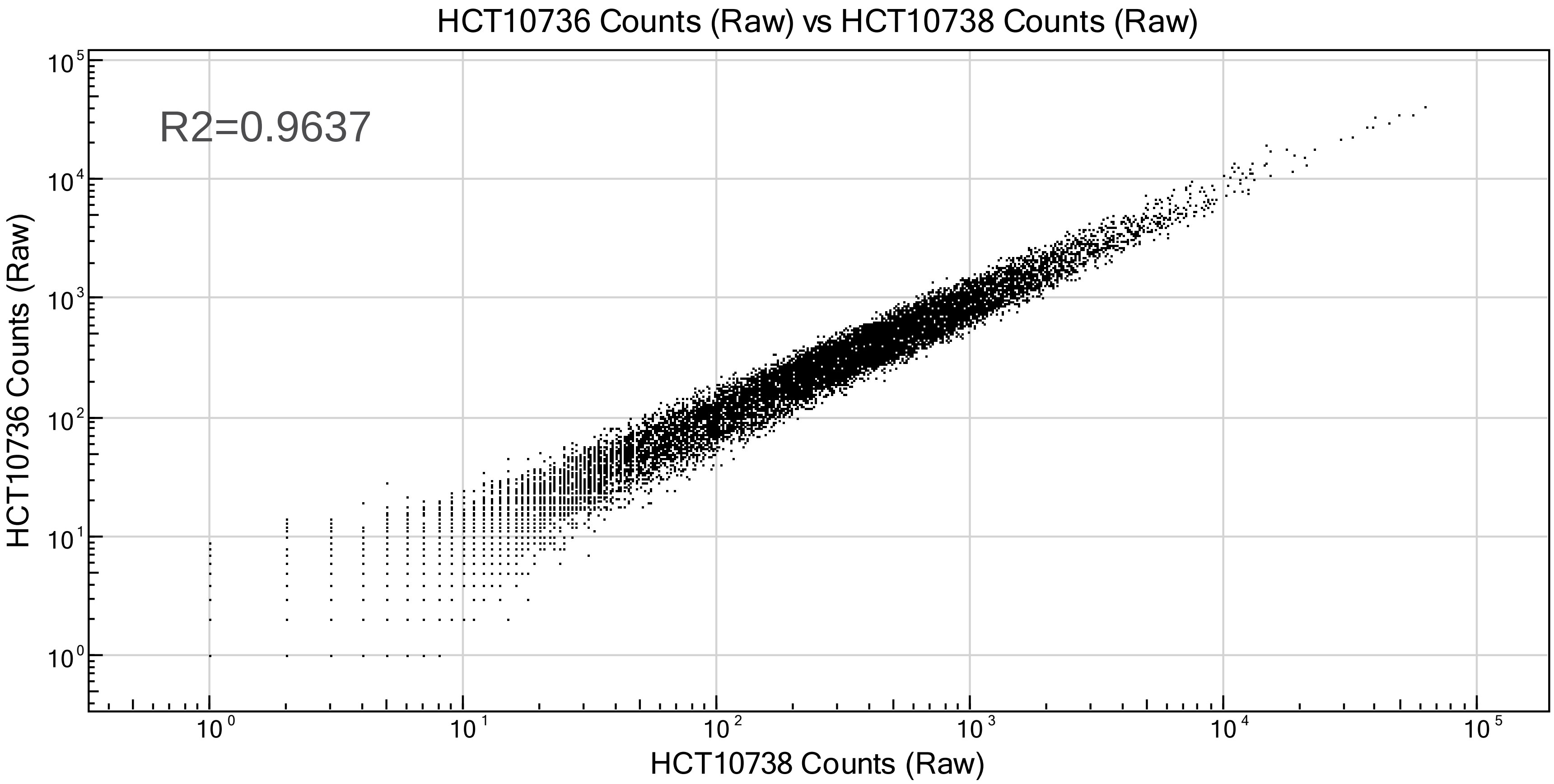


No sequence data file found for this chromosome.



No sequence data file found for this chromosome.

No PCR vs PCR (15 cycles)




DATA
ANALYSES

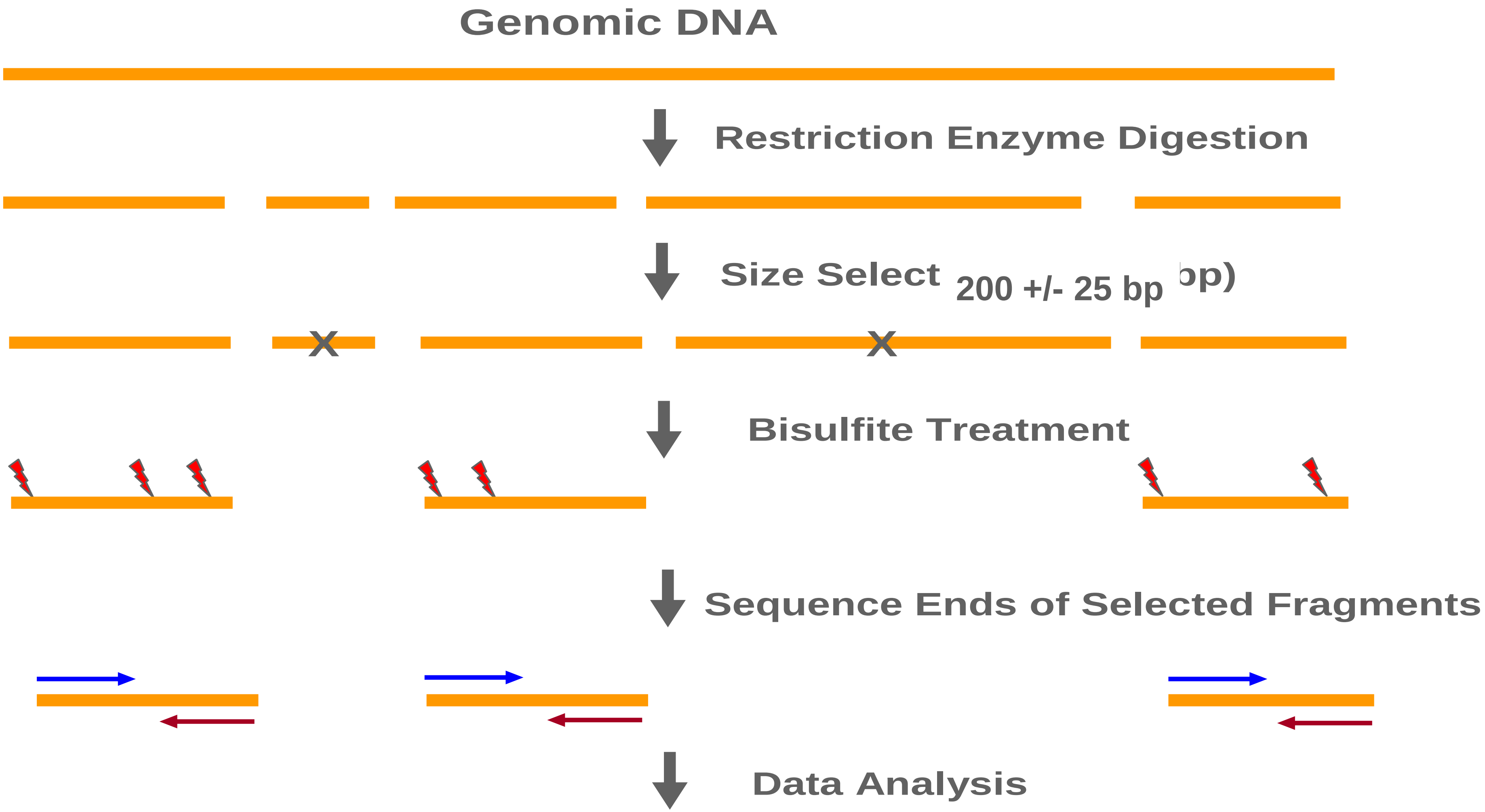
Samples/flow cell and cost/sample (all consum.)

	GAIIe	GAIIx	HiScanSQ	HiSeq 2000
DGE (25bp, 2 million)	60 € 230	125 € 205	125 € 204	250 € 192
RNA-Seq (75x2, 20 million)	6 € 1585	13 € 898	13 € 835	25 € 522
ChIP-Seq (TF)	125 € 252	250 € 231	250 € 229	500 € 219
Virus/BAC clones (100x2, 30x)	2,770 € 209	5,555 € 207	5,555 € 207	11,111 € 206
Bact./GWAS (100x2, 30x)	277 € 242	556 € 224	556 € 222	1,111 € 214
Hu exome (100x2, 30x)	28 € 568	56 € 387	56 € 368	111 € 287
Hu genome (100x2, 30x)	0.28 € 36,444	0.56 € 18,325	0.56 € 16,43	**€ 5,616** **@300Gb/run**

Estrogen receptor alpha controls a gene network in luminal-like breast cancer cells comprising multiple transcription factors and microRNAs. Cicatiello L … *Am J Pathol (2010)* ***176****: 2113-30.*

- 1270 genes respond to beta-estradiol, includes several direct targets of Eralpha, organized in a gene regulation cascade, stemming from ligand-activated receptor and reaching a large number of downstream targets via AP-2gamma, B-cell activating transcription factor, E2F1 and 2, E74-like factor 3, GTF2IRD1, hairy and enhancer of split homologue-1, MYB, SMAD3, RARalpha, and RXRalpha transcription factors.
- MicroRNAs integral components of this gene regulation network; miR-107, miR-424, miR-570, miR-618, and miR-760 are regulated by 17beta-estradiol along with other microRNAs that can target a significant number of transcripts belonging to one or more estrogen-responsive gene clusters.

Reduced Representation Bisulfite Seq



Genomic DNA

Restriction Enzyme Digestion

Size Select 200 +/- 25 bp bp)

Bisulfite Treatment

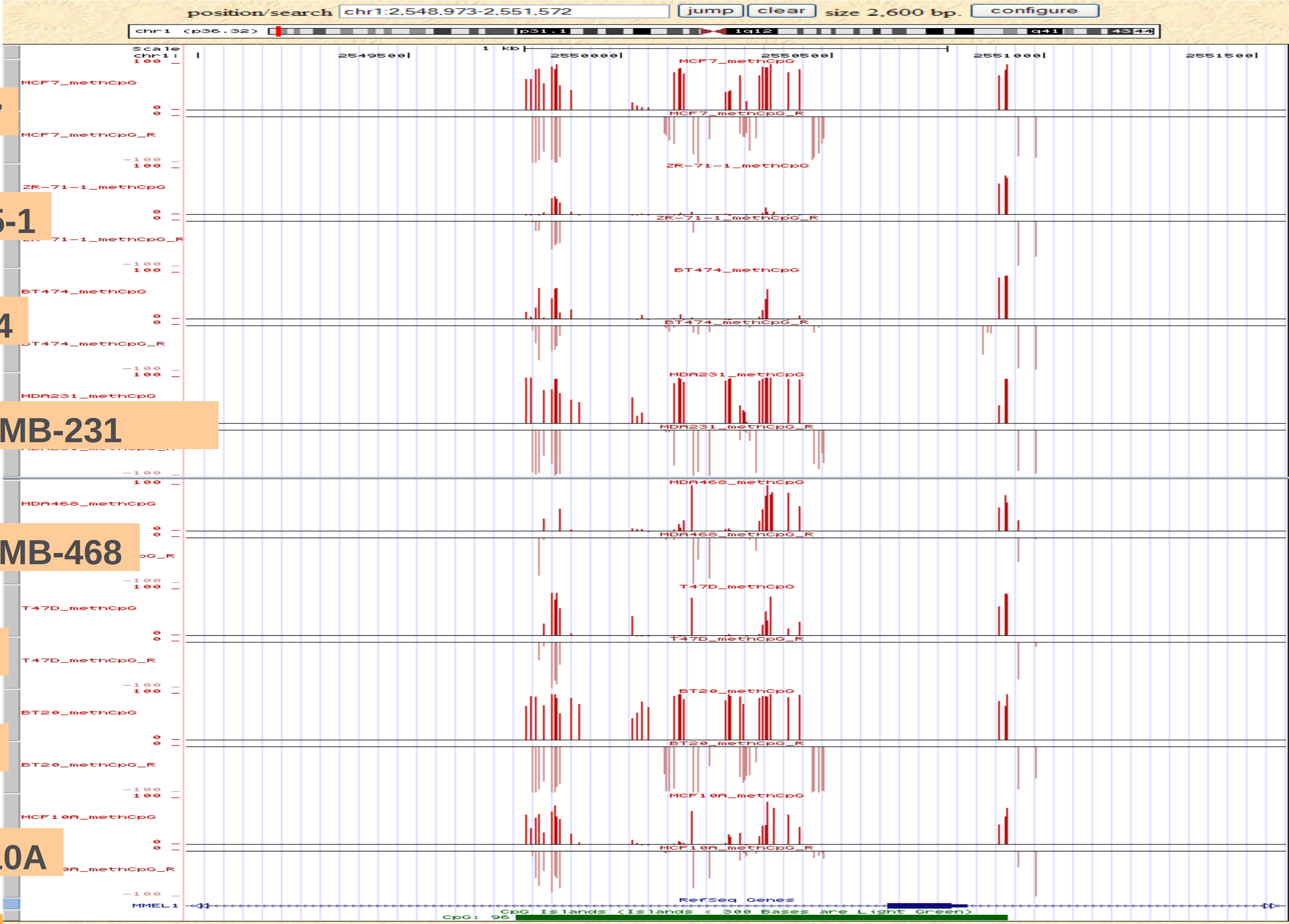
Sequence Ends of Selected Fragments

Data Analysis

Genome-scale DNA methylation maps of pluripotent and differentiated cells

Alexander Meissner[1,2,3*], Tarjei S. Mikkelsen[2,4*], Hongcang Gu[2], Marius Wernig[1], Jacob Hanna[1], Andrey Sivachenko[2], Xiaolan Zhang[2], Bradley E. Bernstein[2,5,6], Chad Nusbaum[2], David B. Jaffe[2], Andreas Gnirke[2], Rudolf Jaenisch[1,7] & Eric S. Lander[1,2,7,8]

BC Cell Line RRBS Data - MMEL1 Gene



MCF7

ZR-75-1

BT474

MDA-MB-231

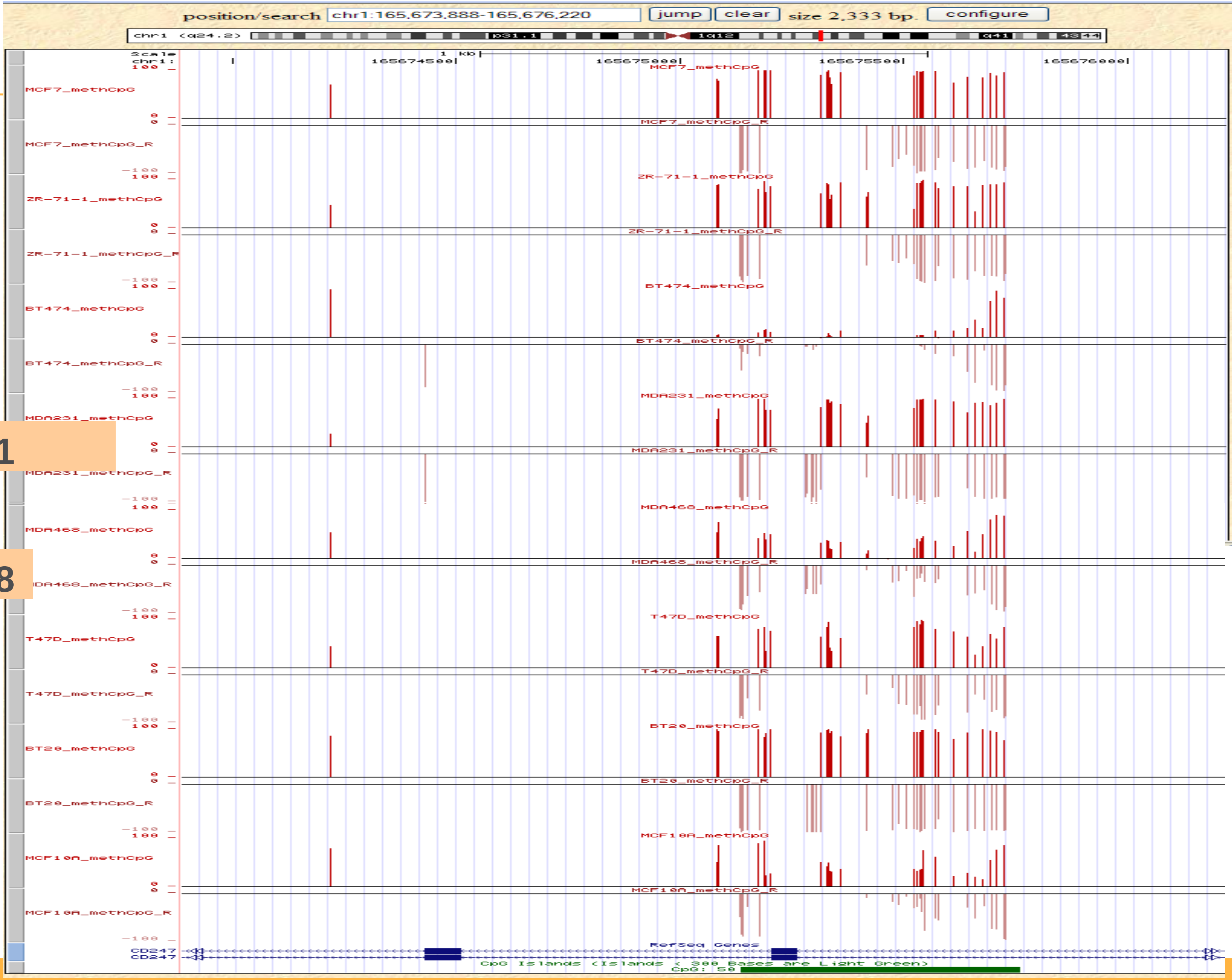
MDA-MB-468

T47D

BT20

MCF10A

BC Cell Line RRBS Data - CD247 Gene


position/search chr1:165,673,888-165,676,220

size 2,333 bp.

MCF7

ZR-75-1

BT474

MDA-MB-231

MDA-MB-468

T47D

BT20

MCF10A

BC Cell Line RRBS Data - CDH13 Gene

MCF7

ZR-75-1

BT474

MDA-MB-231

MDA-MB-468

T47D

BT20

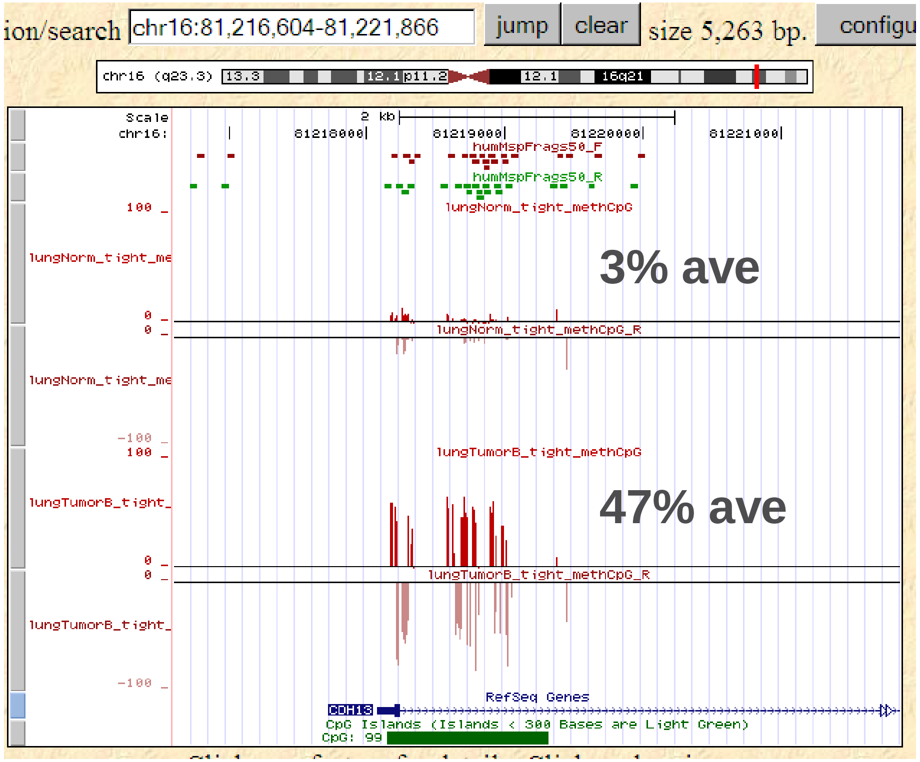
MCF10A

Correlation of RRBS and mRNA-Seq Data for PRKD1 Gene

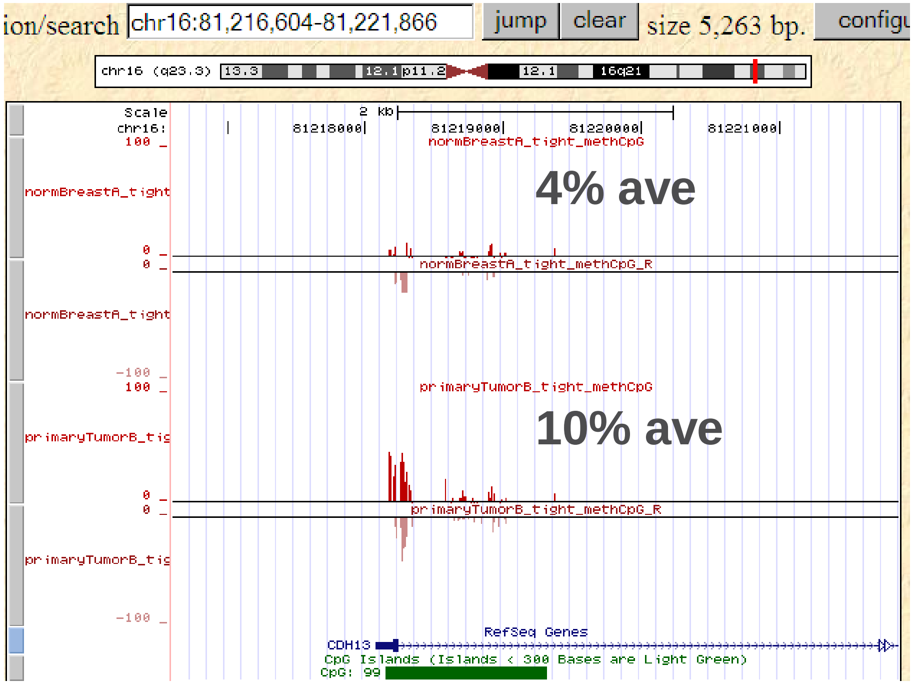
chr14:29,027,620-29,554,515 PRKD1 gene methylation for 5’ CpG island region been dramatically changed and anti-correlates well with mRNA levels

CDH13 (H-cadherin): **increased methylation in lung tumor**

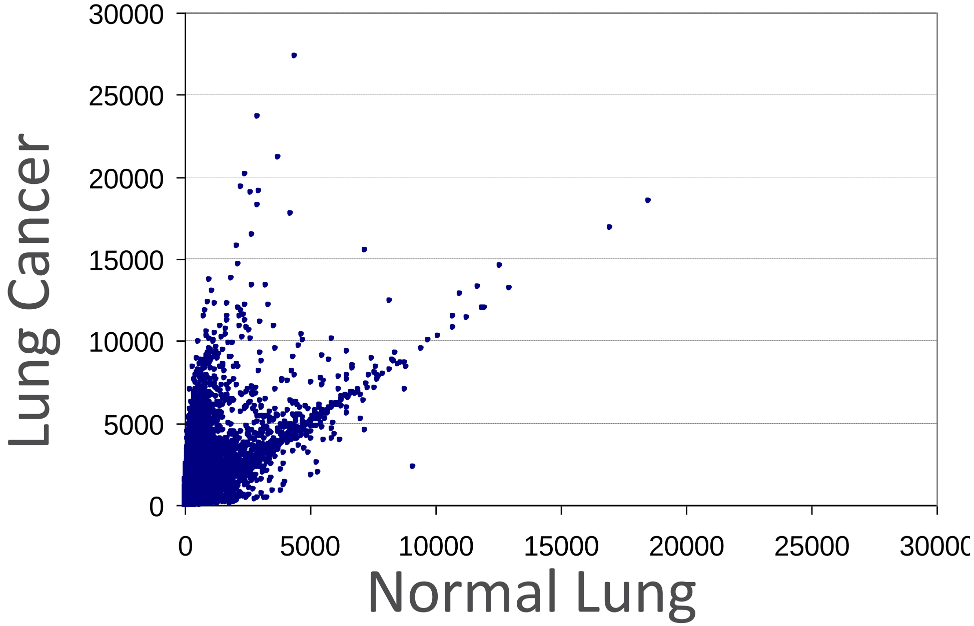
Lung

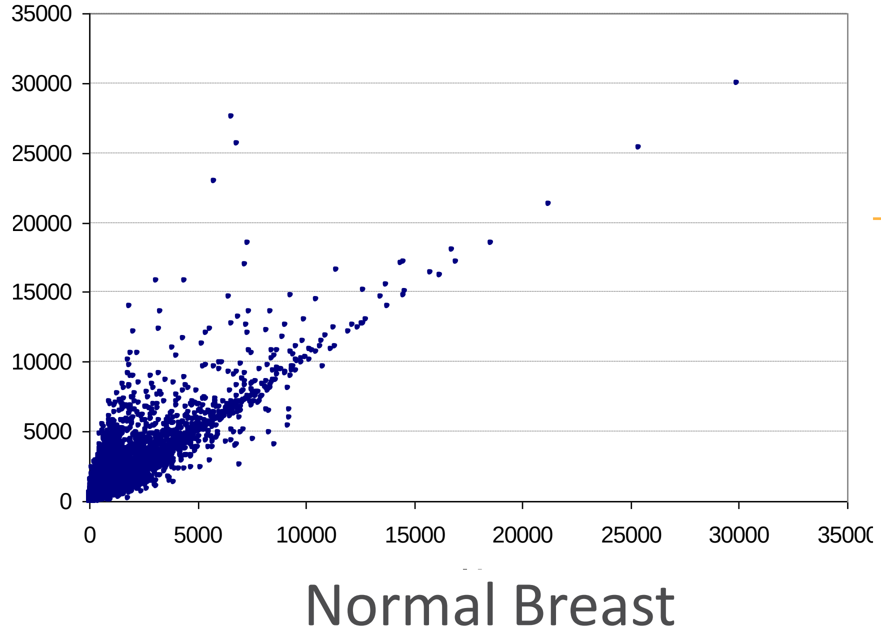


Breast

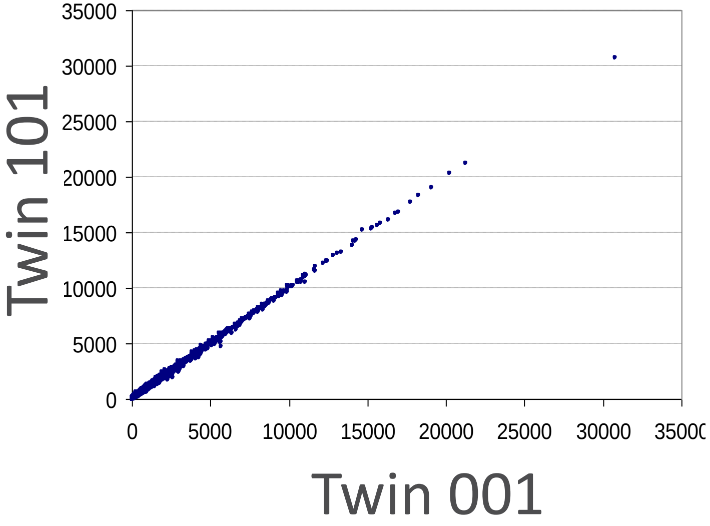


Study both individual CpG sites and CpG Islands

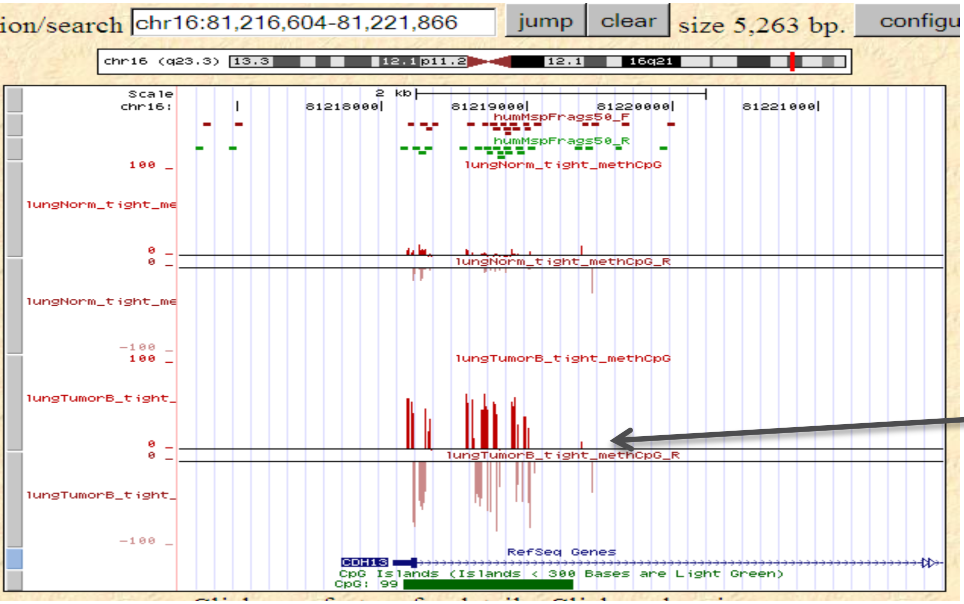




Comparison of Methylation Summarized for ~ 25K CpG Islands

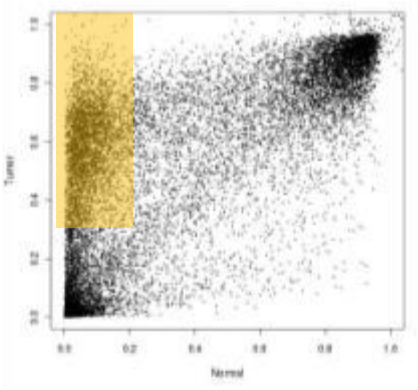


Reduced Representation Bisulfite Sequencing



Whole Genome Bisulfite Sequencing

CpG Island Methylation Changes & Pathways



What are in these pathways that change?

Top 3 Pathways by Enrichment

cAMP-mediated Signaling
(p = 2.09e-8)

G-Protein Coupled Receptor Signaling
(p = 4.17e-7)

Cell Death Related Pathways

Axonal Guidance Signaling
(p = 1.18e-6)

Implication: increased methylation of cell death related pathways could promote cell growth and cancer progression

Experimental Data Available

- RNA-Seq
  - 50 bp paired-end standard mRNA-Seq
  - 100 bp single-read directional mRNA-Seq
- DNA Methylation - RRBS
- CNV - low pass genomic DNA sequencing
- Small RNA Analysis