

Preliminary Analysis
Of Mouse Transcriptome
Summary Statistics

Samples Analyzed

- Total of 17 adult tissues (C57 BL/6J) analyzed in replica (34 samples)
- Mice are 8wk adults and samples were collected from mixed litter-mates.
- Samples obtained from dissections done at CSHL
- Tissues analyzed are:

Adrenal

Duodenum

Stomach

Gonadal Adipose Tissue

Subcutaneous Adipose Tissue

Bladder

Large Intestine

Ovary

Mammary Gland

Testis

Spleen

Colon

Lung

Heart

Kidney

Liver

Thymus

- Total Long (>200nt) polyA+ RNA (deoxy UTP protocol/ribo depleted) isolated
- New biological samples and different RNAs are planned

Sequencing Parameters

- Stranded, paired-end 76 for long RNA
- ~100 M mate-pairs per sample
- NIST spike-in controls added

Mapping Statistics of Mouse Long RNAs

Tissue	Read Count	Cluster Count	Reads Mapped to Single Locus	% Mapping Single End	Mate Pairs Mapped to Single Locus	Mate Pairs Mapped to 2-10 Loci	% Total Mapping Mate Pair (Reported in .BAM .Wig)	Reads Mapped without MisMatch	% Reads w/o MM	Average MM in reads with MM	Average mapped length	Split Reads	% of Reads that are Split	Spliced Junctions
Adrenal	145,222,694	72,611,347	125,119,788	86.16%	60,240,212	2093999	85.85%	92,934,986	74.28%	2.17	74.00	16,691,145	13.34%	16,813,585
Adrenal	151,361,556	75,680,778	135,277,886	89.37%	65,668,462	2437399	89.99%	96,785,184	71.55%	1.99	74.21	18,277,056	13.51%	18,391,094
GITractDuodenum	153,770,376	76,885,188	125,987,634	81.93%	58,603,817	3238754	80.43%	90,584,237	71.90%	2.18	72.07	16,932,586	13.44%	17,080,265
GITractDuodenum	158,045,134	79,022,567	128,488,487	81.30%	58,885,435	3635856	79.12%	93,340,620	72.65%	2.25	72.46	16,475,307	12.82%	16,638,747
GITractStomach	155,429,916	77,714,958	126,827,199	81.60%	60,315,568	4494272	83.39%	97,243,548	76.67%	2.17	73.43	21,414,594	16.88%	21,612,740
GITractStomach	181,671,214	90,835,607	153,580,643	84.54%	73,525,745	3916736	85.26%	118,092,002	76.89%	2.02	73.72	29,793,924	19.40%	30,073,268
AdiposeTissueGonadal	138,750,192	69,375,096	124,914,979	90.03%	60,373,907	3470637	92.03%	93,908,430	75.18%	1.89	74.57	19,430,875	15.56%	19,606,321
AdiposeTissueGonadal	157,183,252	78,591,626	142,138,815	90.43%	68,859,688	4158024	92.91%	106,807,627	75.14%	1.98	74.63	21,866,429	15.38%	22,050,531
AdiposeTissueSubcutaneous	156,644,560	78,322,280	143,065,738	91.33%	69,422,526	3305299	92.86%	96,169,596	67.22%	1.88	74.56	21,050,403	14.71%	21,234,106
AdiposeTissueSubcutaneous	157,689,728	78,844,864	142,911,094	90.63%	69,157,618	4542527	93.47%	97,723,625	68.38%	1.94	74.53	21,476,766	15.03%	21,652,978
GITractLargeIntestine	143,997,066	71,998,533	126,770,389	88.04%	60,980,666	4097038	90.39%	94,346,229	74.42%	1.80	74.50	19,921,102	15.71%	20,127,445
GITractLargeIntestine	153,235,228	76,617,614	135,639,826	88.52%	65,757,654	4928943	92.26%	92,060,344	67.87%	1.69	74.51	21,219,683	15.64%	21,436,965
GITractSmallIntestine	140,573,262	70,286,631	119,708,380	85.16%	55,922,366	4420981	85.85%	88,932,199	74.29%	2.02	73.66	19,853,915	16.59%	20,023,186
GITractSmallIntestine	135,540,594	67,770,297	120,092,452	88.60%	57,350,349	3301937	89.50%	93,443,903	77.81%	2.07	74.19	17,149,003	14.28%	17,316,974
Ovary	110,997,870	55,498,935	102,739,298	92.56%	50,144,056	1937694	93.84%	80,348,921	78.21%	1.74	74.78	15,267,478	14.86%	15,394,399
Ovary	100,356,244	50,178,122	90,214,714	89.89%	43,888,824	1565874	90.59%	69,587,310	77.14%	2.01	74.61	12,391,308	13.74%	12,497,312
Testis	147,114,220	73,557,110	134,455,322	91.40%	65,766,427	2133390	92.31%	114,407,808	85.09%	1.64	74.96	21,870,245	16.27%	22,063,477
Testis	154,413,056	77,206,528	142,026,115	91.98%	69,340,962	2390680	92.91%	117,000,989	82.38%	1.70	74.94	22,997,432	16.19%	23,202,612
Spleen	148,209,846	74,104,923	128,776,290	86.89%	61,367,485	5466138	90.19%	95,110,115	73.86%	1.84	74.13	17,811,080	13.83%	17,992,763
Spleen	156,978,746	78,489,373	131,859,964	84.00%	63,276,161	7804389	90.56%	102,961,498	78.08%	1.86	74.36	18,010,684	13.66%	18,190,564
Colon	160,435,704	80,217,852	142,472,383	88.80%	68,961,826	4016978	90.98%	112,721,249	79.12%	1.98	74.54	21,924,842	15.39%	22,169,279
Colon	132,332,448	66,166,224	116,775,931	88.24%	55,976,629	3068823	89.24%	87,926,550	75.30%	1.96	74.27	18,109,390	15.51%	18,301,474
Lung	141,134,744	70,567,372	125,676,302	89.05%	61,182,967	2511345	90.26%	102,340,593	81.43%	1.73	74.83	17,838,138	14.19%	17,991,677
Lung	124,997,574	62,498,787	109,261,545	87.41%	53,263,559	2405580	89.07%	91,403,833	83.66%	1.82	74.81	16,297,785	14.92%	16,440,979
Heart	149,218,530	74,609,265	118,677,773	79.53%	57,589,829	11781791	92.98%	92,664,804	78.08%	1.95	74.77	14,689,770	12.38%	14,835,682
Heart	161,943,850	80,971,925	128,989,220	79.65%	62,808,018	12096876	92.51%	95,976,002	74.41%	1.65	74.78	17,516,714	13.58%	17,702,899
Kidney	186,383,474	93,191,737	162,442,055	87.15%	79,082,893	5702465	90.98%	120,046,367	73.90%	1.88	74.59	24,996,191	15.39%	25,198,166
Kidney	235,774,726	117,887,363	202,345,895	85.82%	98,110,646	9480695	91.27%	166,018,668	82.05%	1.82	74.71	29,498,314	14.58%	29,749,014
Liver	160,540,434	80,270,217	142,435,107	88.72%	68,663,506	4210388	90.79%	107,969,065	75.80%	1.88	74.36	24,099,402	16.92%	24,282,256
Liver	164,835,908	82,417,954	147,568,573	89.52%	71,850,453	4092527	92.14%	117,854,585	79.86%	1.83	74.65	25,002,284	16.94%	25,197,132
Thymus	160,064,018	80,032,009	138,155,762	86.31%	66,598,826	3382542	87.44%	105,050,753	76.04%	1.85	74.58	17,482,863	12.65%	17,644,414
Thymus	203,089,924	101,544,962	177,962,860	87.63%	86,289,923	4410988	89.32%	144,001,604	80.92%	1.86	74.78	22,065,089	12.40%	22,285,120
MammaryGland	171,372,188	85,686,094	148,834,581	86.85%	72,195,702	5220217	90.35%	102,197,071	68.66%	2.35	74.20	20,176,100	13.56%	20,353,170
MammaryGland	122,633,048	61,316,524	106,020,491	86.45%	51,467,905	2199493	87.53%	82,245,382	77.57%	1.84	74.60	15,665,909	14.78%	15,803,729

K562 nuclear sub-compartments (total RNA)

1. Expression of Gencode (v7) annotated elements

Gene type	Detected exons ² (annotation #)	Detected splice junctions ² (annotation #)	Detected transcripts ² (annotation #)	Detected genes ² (annotation #)	Exon nucleotide coverage ³ (%)
Long non coding	8,109 (41,467)	1,644 (26,872)	1,903 (14,880)	2,032 (9,277)	79
Protein coding	167,711 (318,514)	109,253 (244,158)	21,661 (76,006)	12,344 (20,679)	96.3
Other ¹	53,877 (133,937)	5,260 (47,663)	18,630 (71,113)	3,954 (21,750)	93.1
Total annotated	229,697 (493,918)	116,157 (318,693)	42,194 (161,999)	18,330 (51,706)	94.7

¹ includes pseudogenes, miRNAs, etc

² all elements that passed nplDR (0.1)

³ cumulative detected nucleotide in detected exons / total nucleotides in detected exons

2. Expression of Gencode (v7) intergenic and antisense elements

Category	Detected exons ⁴	Detected splice junctions ⁴	Detected transcripts ⁴	Detected genes ⁴
Mono-exonic	40,319	NA	40,273	39,327
Multi-exonic	6,014	14,374	2,570	1,791
Total	46,333	14,374	42,843	41,118

⁴ all elements that passed nplDR (0.1)

Preliminary Summary Statistics for Annotated Regions

Gene type	Number of annotated genes	Detected genes ²		Number of genes expressed in only 1 tissue	Proportion over genes expressed (%)	Number of genes expressed in 16 tissues	Proportion over genes expressed (%)
		number	prop.				
Long non coding	3,742	3,620	96.7	189	5.2	278	7.7
Protein coding	22,667	21,227	93.6	827	3.9	866	4.1
Other ¹	10,405	7,086	68.1	1,272	18.0	357	5.0
Total annotated	36,814	31,933	86.7	2,288	7.2	1,501	4.7

¹ includes pseudogenes, miRNAs, etc

² all elements with average rpkms > 0

Notable Facts About Preliminary Analysis

- Current analysis is not filtered by IDR
- Quantification is not done using same pipeline as human
- Gene models completed using Cufflinks and sent to Mike S.
- These models will change when analysis is done as were done with human data
- No Validation (Pac Bio and long cDNA syntheses) done yet as with human data
- It will take weeks for CRG to reproduce the analysis of these data using same pipeline and thresholds that were used for human data

Acknowledgements

- Carrie Davis
- Sarah Djebali
- Roderic Gugio
- Chris Zalenski
- Sonali Jha