

# Annotate New Transcriptome - Brain Ribosomal-RNA depleted Total RNA

## 1. Create BED Files to Identify Overlap With Annotation

## 2. Identify overlap between transcripts from reconstruction and annotation databases

BEDtools Version = bedtools v2.19.1

```
export PATH=/usr/local/bedtools2/bin:$PATH
cd /data2/saba/BNLx.SHR/RNA-Seq.Brain.total/reconstruction/tmp
intersectBed -a gtf.bed -b /data2/saba/gtfFiles/ensGene.RGSC5.0.rn5.cleaned.bed -wo >
overlap.brain.total.wEnsembl.txt
intersectBed -a gtf.bed -b /data2/saba/gtfFiles/refSeqGene.RGSC5.0.rn5.cleaned.bed -wo >
overlap.brain.total.wRefSeq.txt
intersectBed -a gtf.bed -b /data2/saba/gtfFiles/xenoRefGene.mouseOnly.RGSC5.0.rn5.bed -wo >
overlap.brain.total.wMouseRefSeq.txt
```

Number of Transcripts: 32,704 Number of Genes: 30,739

### Annotation for Entire Set of Transcripts

Annotation	Number of Transcripts	Percent of Transcripts
ensembl transcript	8,591	26.3%
novel transcript of Ensembl Gene	3,413	10.4%
associated with Rat RefSeq Gene	109	0.3%
associated with Mouse RefSeq Gene	141	0.4%
unannotated	20,450	62.5%

### Annotation for 20% of Transcripts with Highest Expression

Annotation	Number of Transcripts	Percent of Transcripts
ensembl transcript	4,680	71.5%
novel transcript of Ensembl Gene	1,139	17.4%
associated with Rat RefSeq Gene	39	0.6%
associated with Mouse RefSeq Gene	36	0.6%
unannotated	647	9.9%

### Top 10 Most Highly Expressed Transcripts

transcript_id	chromosome	start (bp)	stop (bp)	annotation summary	Gene Identifier	Gene Symbol	Gene Description	Ensembl Gene Biotype	FPKM (BN-Lx)	FPKM (SHR)
matched.32947.1	chr15	31571352	31571712	associated with Rat RefSeq Gene	NR_002703	Rmrp	Rattus norvegicus RNA component of mitochondrial RNA processing endoribonuclease (Rmrp), RNase MRP RNA		7456.6	8316.7
BNLx.10411.1	chr5	63519183	63519708	associated with Mouse RefSeq Gene	NR_001460	Rmrp	Mus musculus RNA component of mitochondrial RNAase P (Rmrp), RNase MRP RNA		1347.2	1347.8
SHR.28958.1	chr1	224944742	224945085	unannotated					646.1	714.7
ENSRNOT00000003283	chrX	107379831	107394874	ensembl transcript	ENSRNOG00000002419	Pip1	proteolipid protein 1	protein_coding	612.4	596.0
ENSRNOT00000058295	chr18	78473296	78504226	ensembl transcript	ENSRNOG00000016516	MBP	myelin basic protein	protein_coding	503.8	477.1
ENSRNOT00000041891	chr1	81878373	81881182	ensembl transcript	ENSRNOG00000018454	Apoe	apolipoprotein E	protein_coding	481.9	474.7

transcript_id	chromosome	start (bp)	stop (bp)	annotation summary	Gene Identifier	Gene Symbol	Gene Description	Ensembl Gene Biotype	FPKM (BN-Lx)	FPKM (SHR)
BNLx.8060.1	chr3	136182042	136236652	associated with Rat RefSeq Gene	NM_001270575;NM_030991	Snap25	Rattus norvegicus synaptosomal- associated protein 25 (Snap25), transcript variant 1, mRNA		443.1	444.5
ENSRNOT00000074404	chr20	19258274	19274198	ensembl transcript	ENSRNOG00000048682	Zwint	ZW10 interactor, kinetochore protein	protein_coding	384.8	387.9
ENSRNOT00000022280	chr18	78473296	78504226	ensembl transcript	ENSRNOG00000016516	MBP	myelin basic protein	protein_coding	365.4	389.1
ENSRNOT00000022113	chr18	15307563	15314777	ensembl transcript	ENSRNOG00000016275	Ttr	transthyretin	protein_coding	375.7	378.6

BigWig Tracks For UCSC Genome Browser

```
track type=bigwig name="HXB.Brain.polyA" description="HXB/BXH Brain PolyA+"
bigDataUrl=http://ucsc:JU7etr5t@phenogen.ucdenver.edu/ucsc/HXB.brain.polyA.bw
track type=bigwig name="HXB.Brain.total" description="HXB/BXH Brain Total"
bigDataUrl=http://ucsc:JU7etr5t@phenogen.ucdenver.edu/ucsc/HXB.brain.total.bw
```