Evolution of Natural Antisense Transcripts

(Or What I had been doing for the last 10 weeks)

Maurice Ling, James Ban, Xijin Ge

Evolution of Natural Antisense Transcripts

- Supported by NIH grant (Project number: 5R01GM083226-03, PI: Xijin Ge)
  - Work in progress

- What are natural antisense transcripts (NATs) and why are they important?
- Aim 1: Identify tissue-specific NATs
- Aim 2: Is the expression of NATs evolutionarily conserved?

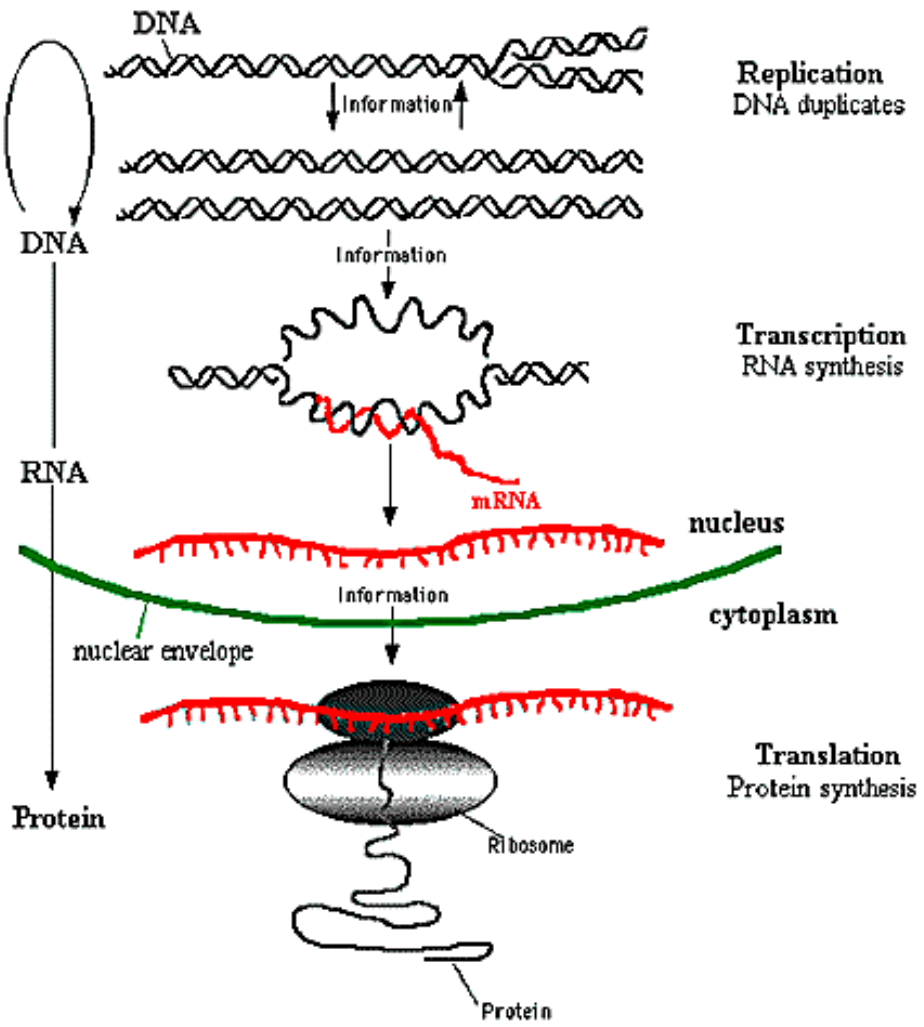
Background

- Central dogma
  - **DNA →RNA → peptide → proteins**

- Many possible controls
  - Accessibility of DNA
  - Expression of DNA (transcription)
  - Half-life of RNA
  - Processing of RNA
  - Coding of peptide (translation)
  - Activity of protein
  - Half-life of protein

- Sense transcript = protein-coding gene
- Antisense transcript = expressed complement of the protein-coding gene

- **Antisense transcript can bind to sense transcript and "interfere"** with
  - Half-life of RNA
  - Processing of RNA
  - Coding of peptide (translation)



The Central Dogma of Molecular Biology

Background

- NATs may be a p(layer) in the regulation of protein expression
- NAT is not isolated events
  - 20-30% of human genes have NATs (Lehner et al., 2002)

- Involvement in diseases
  - α-thalassemia (Zhang et al., 2006), hypertension (Michael et al., 2011), Huntington's (Chung et al., 2011), immune disorders (Hatzoglou et al., 2002), Alzheimer's (Faghihi et al., 2010)

Background

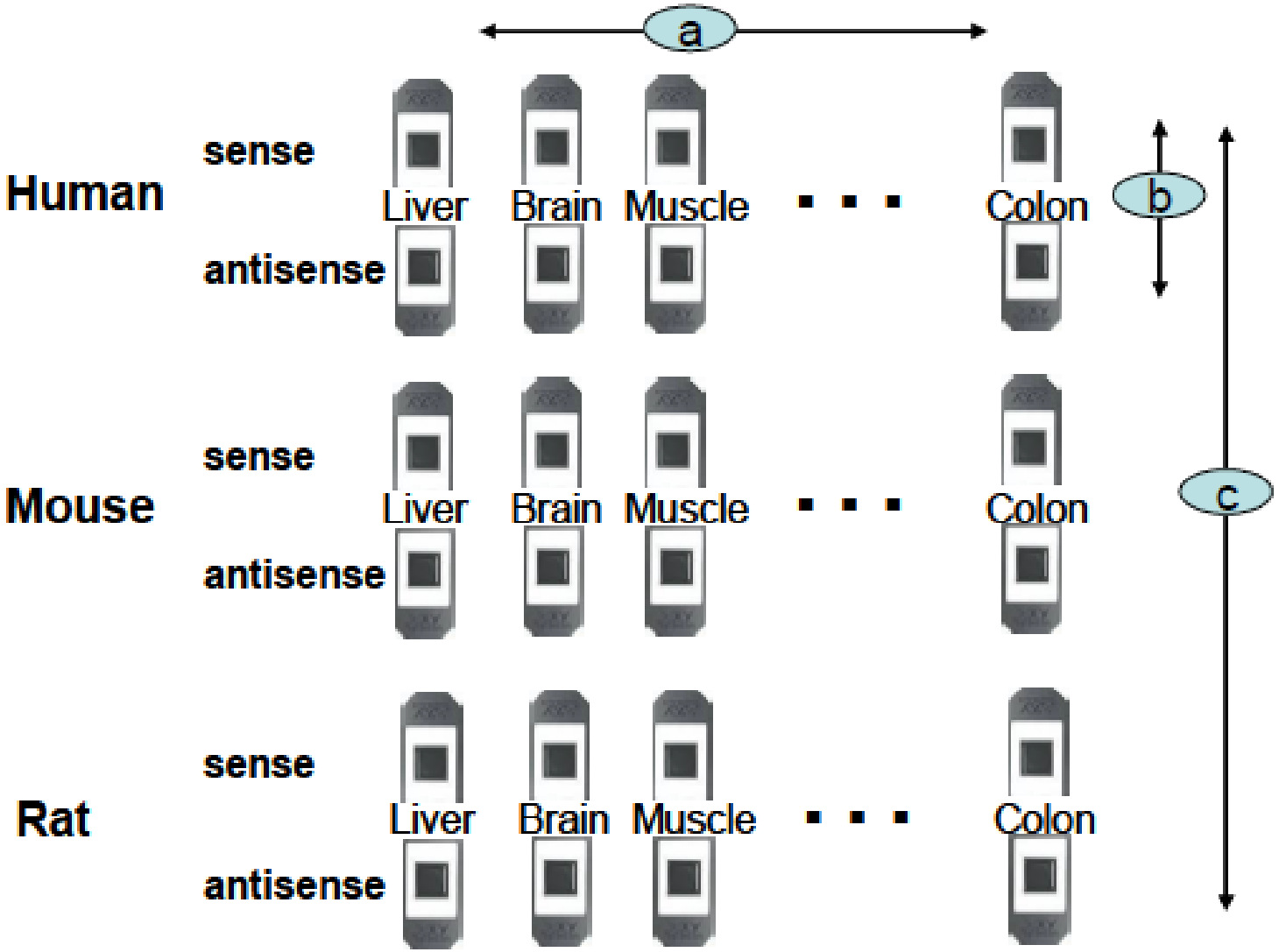
- Extend of regulation vs transcriptional noise is still unknown (Conley and Jordan, 2011).

- We know more about protein-coding genes
  - Higher sequence similarity → Similar expression patterns (Jordan et al., 2005)
  - Gene expression between mouse-human homologs are more correlated than random pairs
    - Homologs are genes that descent from a common ancestral DNA sequence

Hypotheses

- NATs are actively regulated; thus, not transcriptional noise.
  - Do they follow the same expression patterns with respect to protein-coding genes?
  - Higher sequence similarity → Similar expression patterns
  - James → examines protein-coding gene expression
- Identify tissue-specific / novel NATs

Set Up

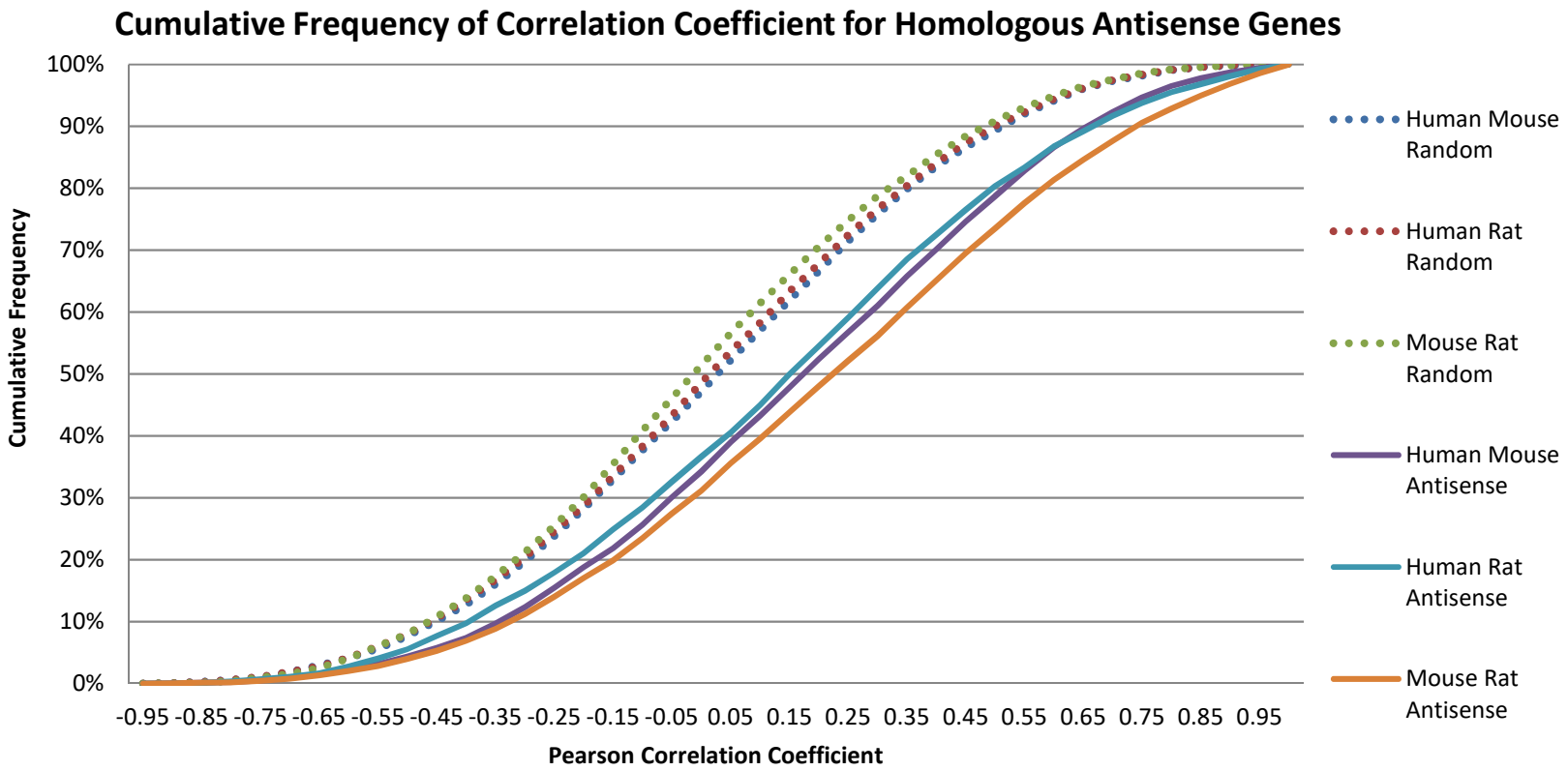

a
Human
sense
antisense
Liver
Brain
Muscle
Colon
b
Mouse
sense
antisense
Liver
Brain
Muscle
Colon
c
Rat
sense
antisense
Liver
Brain
Muscle
Colon

Is the expression of NATs evolutionarily conserved?

- Gene level analysis, then exon level analysis
  - One gene has one or more exons

- Compare the correlation distribution of homologous pairs with random (non-homologous pairs)

- Gene: HomoloGene (protein) → Gene accession → microarray probes



- Random vs antisense
  - Human/Mouse, p = 1.49e-139
  - Human/Rat, p = 3.11e-73
  - Mouse/Rat, p < 1e-140

Human-Mouse-Rat ancestor = 87 million years ago
Mouse-Rat = 16 million years ago
Source: Bourque et al., 2004

Story so far

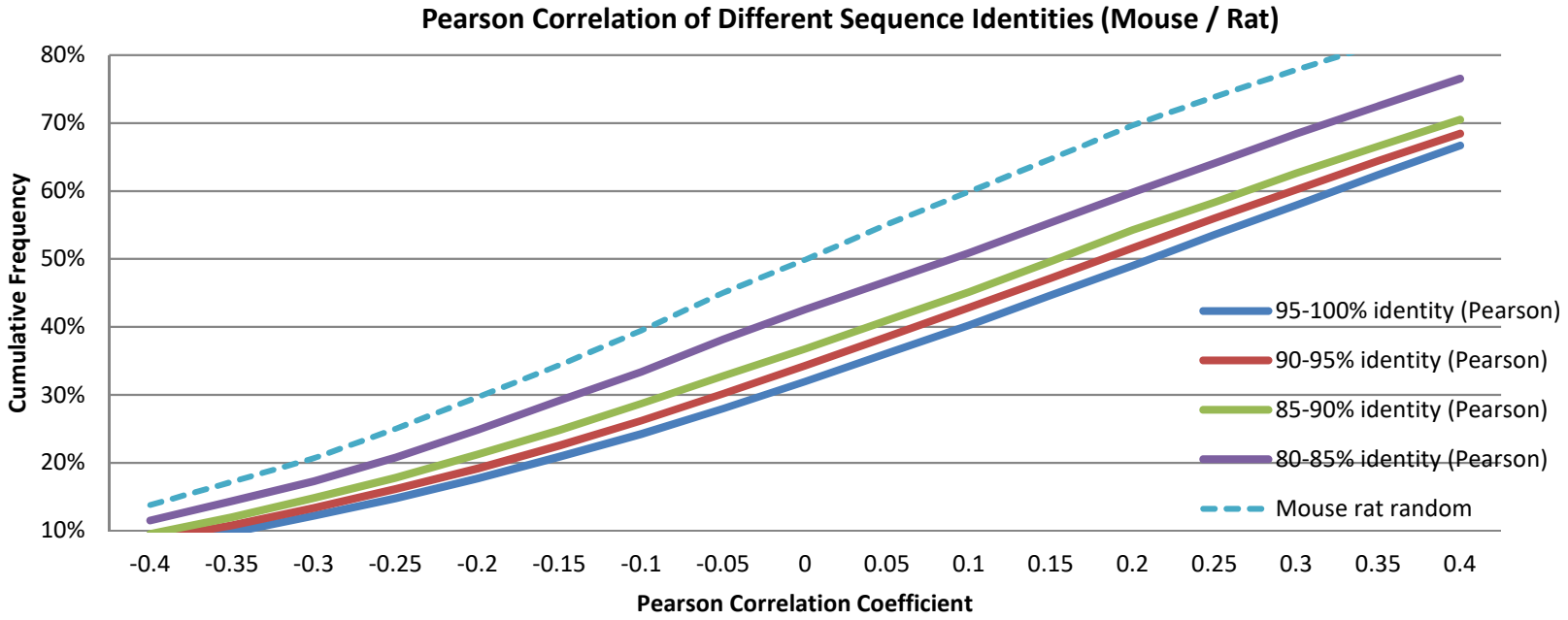
- Expression of homologous antisense genes is statistically higher than non-homologous genes
  - Higher sequence similarity → Similar expression patterns (Jordan et al., 2005)

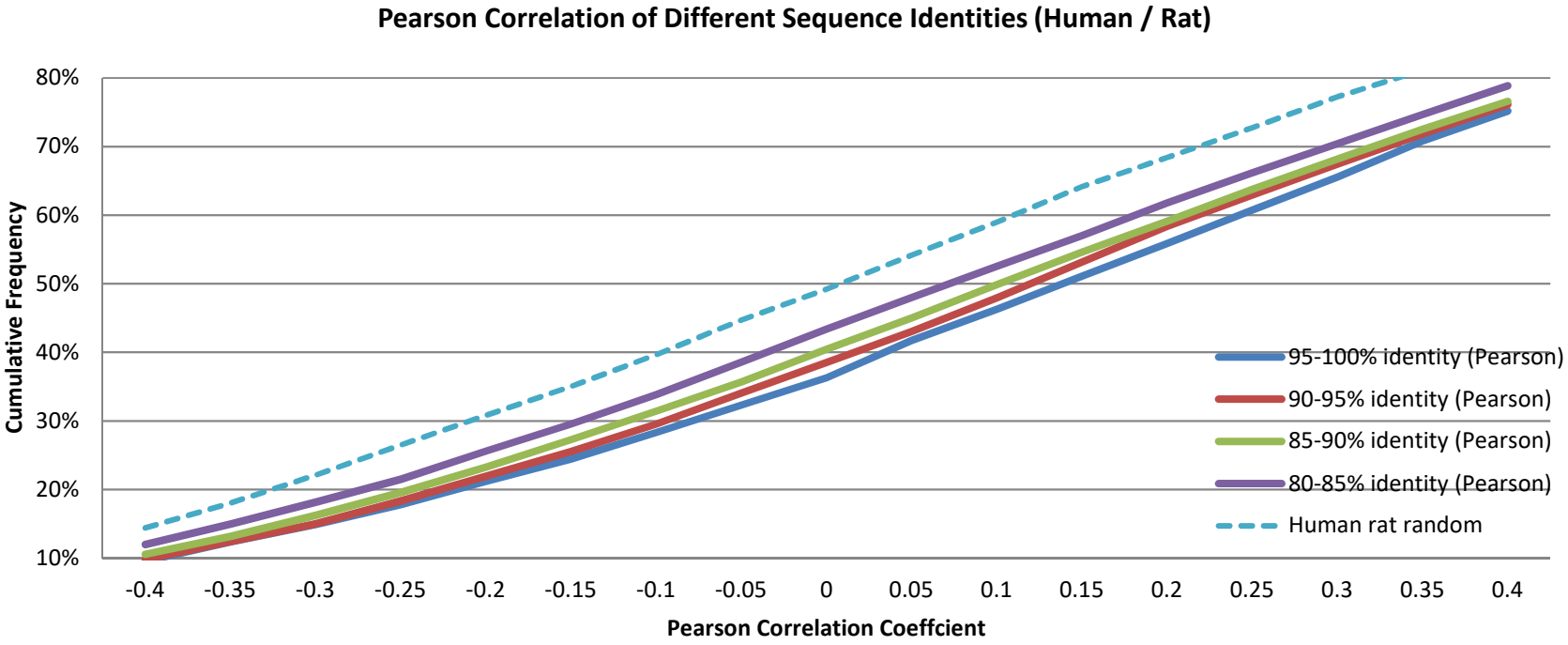
- Exon level
  - Expression of homologous antisense **exons** is statistically higher than non-homologous **exons**
  - Correlation is directly proportional to identity

Analyzing Antisense Exons

- There is HomoloGene for homologous genes but no such thing for exons

- Download human, mouse, rat exon data from UC Santa Clara
- Create my own BLAST database and do pairwise BLAST (Needleman-Wunsch takes 1 millenium on 3 processors)
- Map to probe by genome coordinates using chromosome, start and end positions
- Remove cases where both orthologs have expression lower than 4.35 (average of all data) as noise


Pearson Correlation of Different Sequence Identities (Mouse / Rat)
Cumulative Frequency
80%
70%
60%
50%
40%
30%
20%
10%
-0.4
-0.35
-0.3
-0.25
-0.2
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
Pearson Correlation Coefficient
95-100% identity (Pearson)
90-95% identity (Pearson)
85-90% identity (Pearson)
80-85% identity (Pearson)
Mouse rat random


Pearson Correlation of Different Sequence Identities (Human / Rat)
Cumulative Frequency
80%
70%
60%
50%
40%
30%
20%
10%
-0.4
-0.35
-0.3
-0.25
-0.2
-0.15
-0.1
-0.05
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
Pearson Correlation Coeffcient
95-100% identity (Pearson)
90-95% identity (Pearson)
85-90% identity (Pearson)
80-85% identity (Pearson)
Human rat random

Statistical Tests

- One-way ANOVA on identity-striated correlations
  - Human/mouse orthologs, p = 1.27e-14
  - Human/rat orthologs, p = 2.38e-20
  - Mouse/rat orthologs, p = 1.90e-87

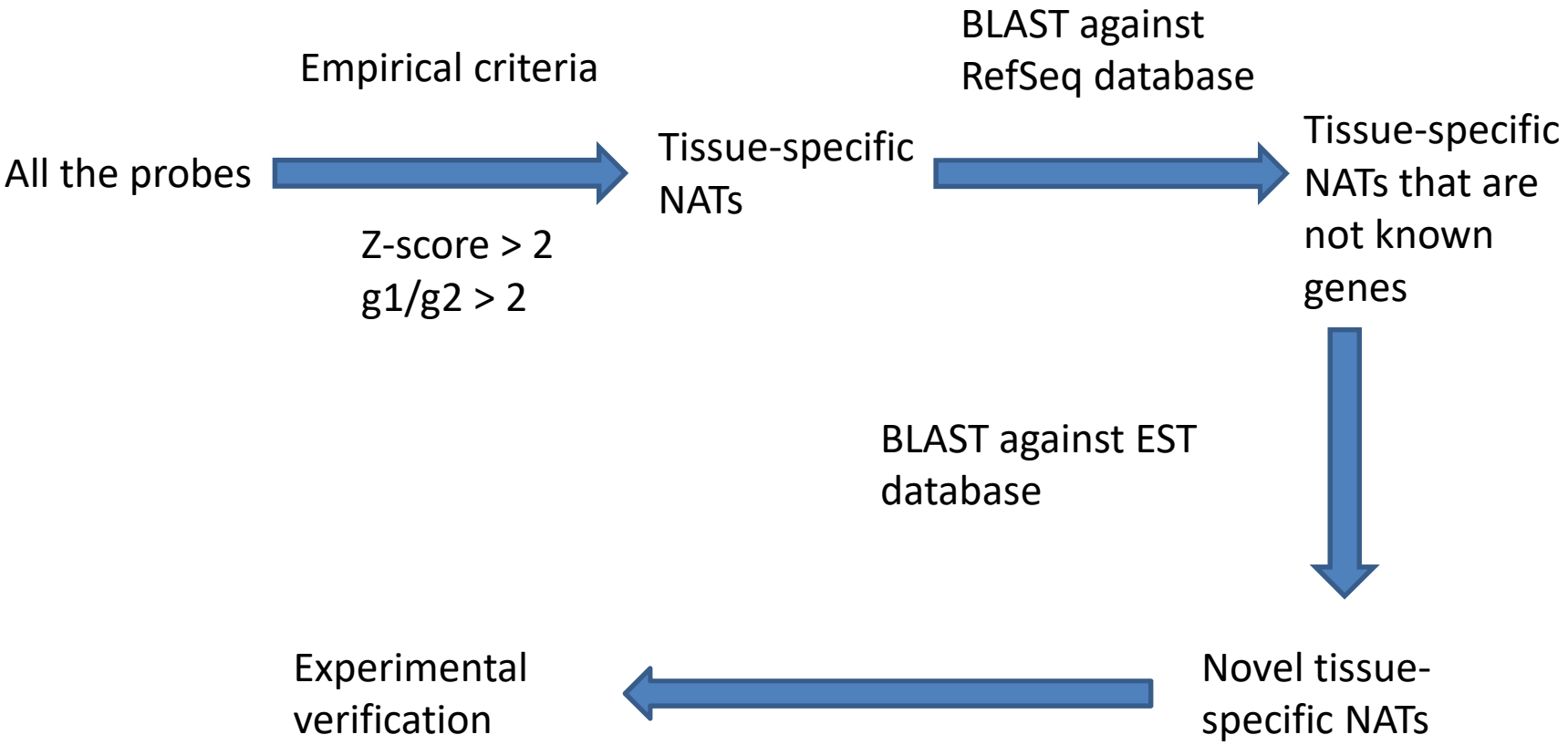
- Visual observation → trends are as expected
  - Correlation is directly proportional to identity

Hypotheses so far

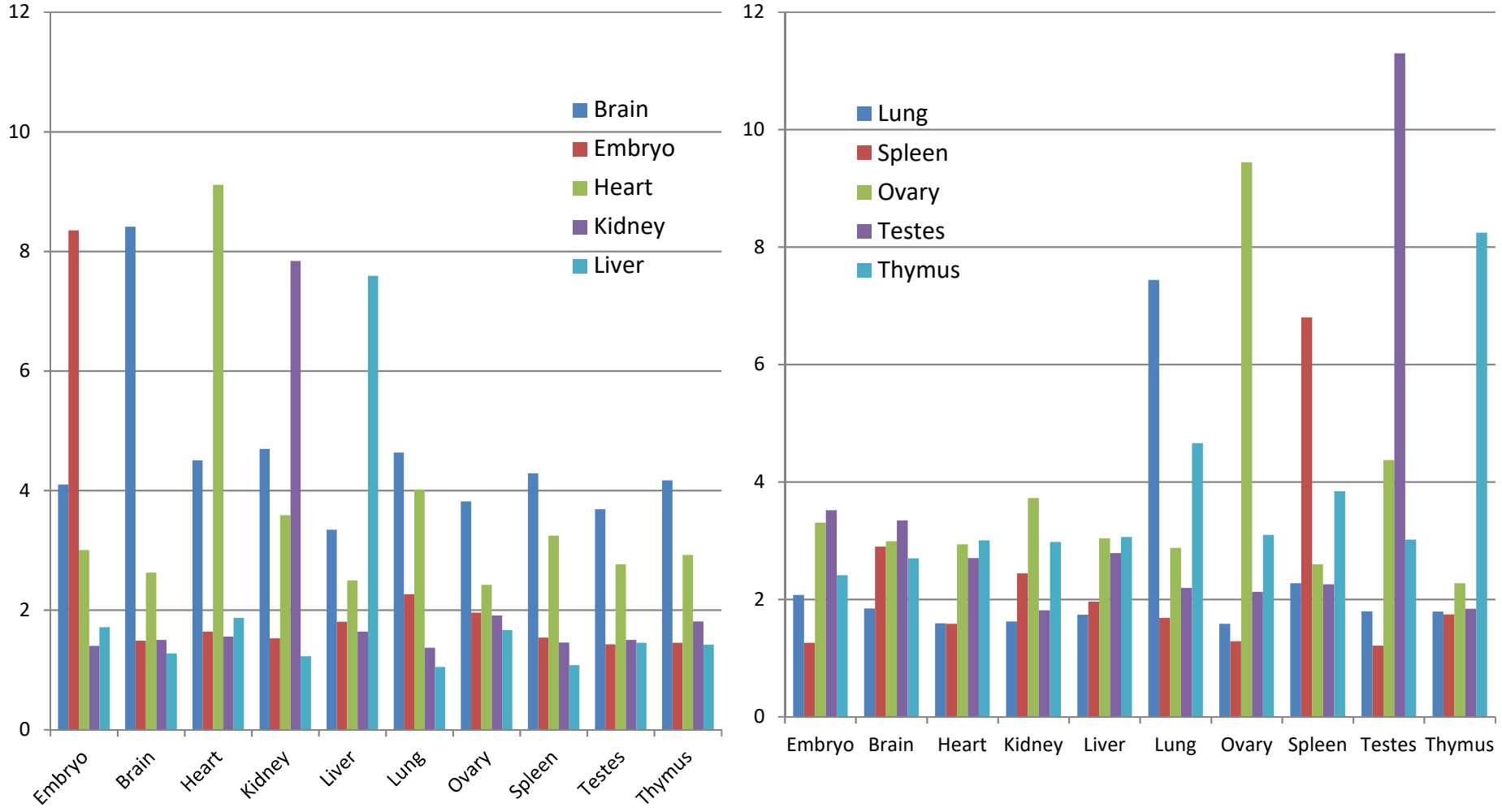
- NATs are actively regulated; thus, not transcriptional noise.
  - Expression of homologous antisense genes is statistically higher than non-homologous genes
  - Correlation is directly proportional to identity
  - Suggests that NATs are actively regulated

- Identify tissue-specific / novel NATs

Identify Tissue-Specific NATs



10 Tissue-specific probes from rat (not found in RefSeq)

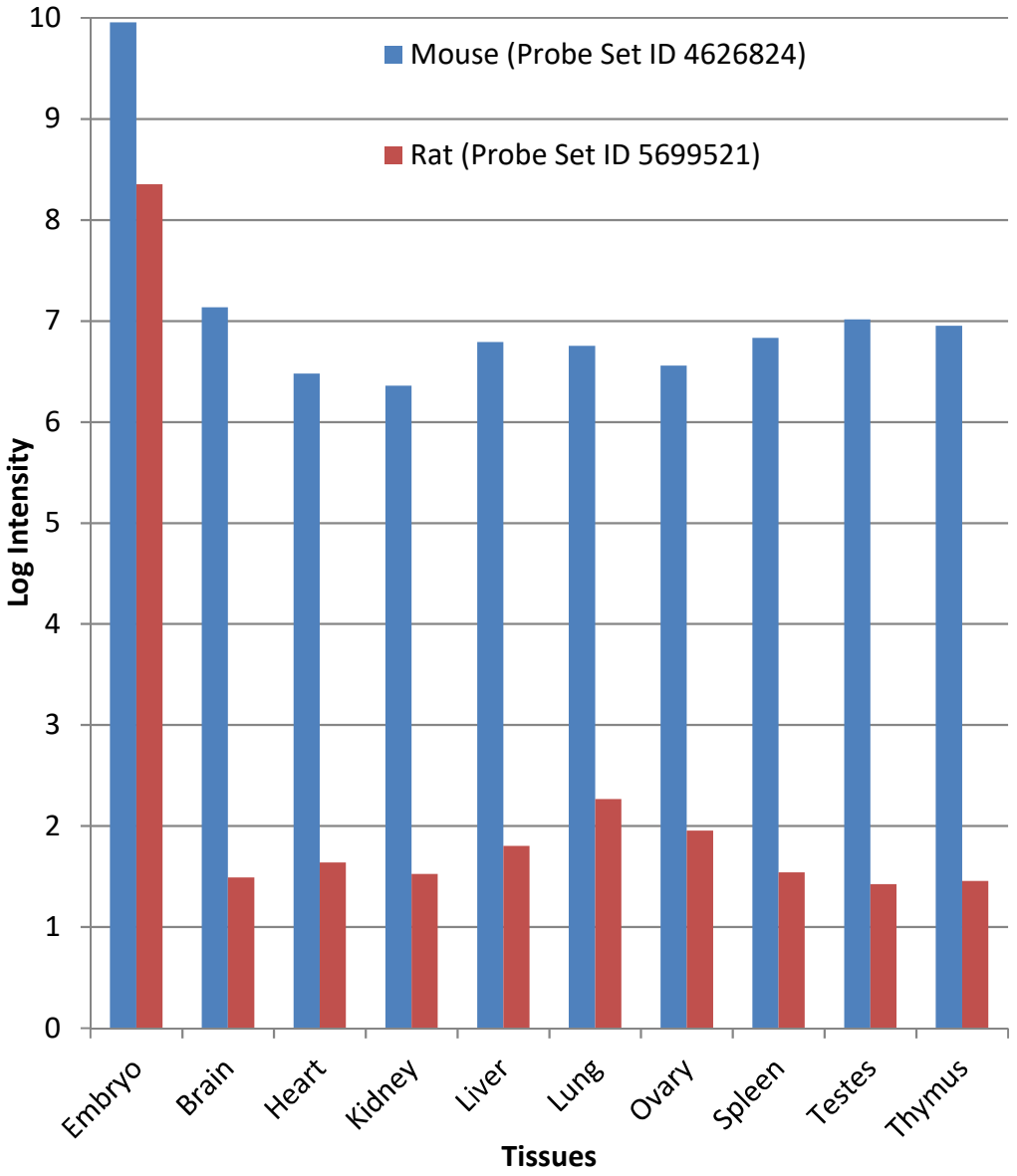


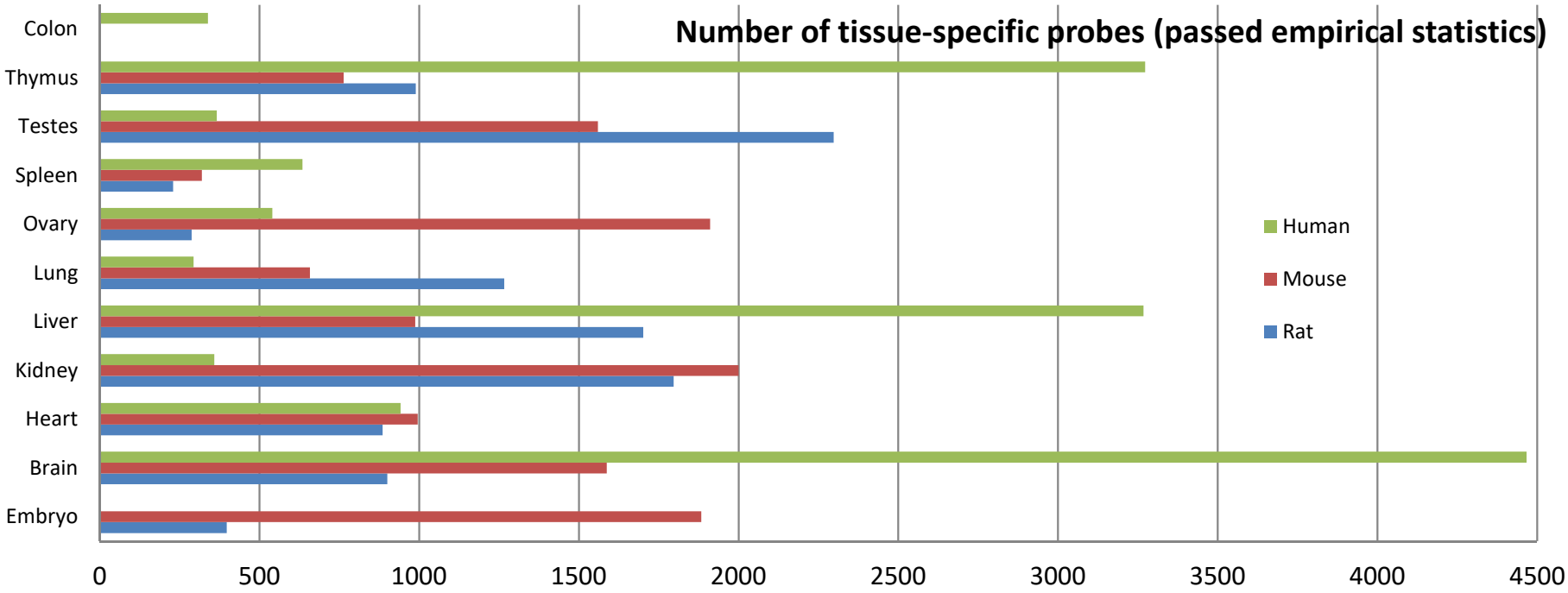
Example of homologous embryo-specific NAT

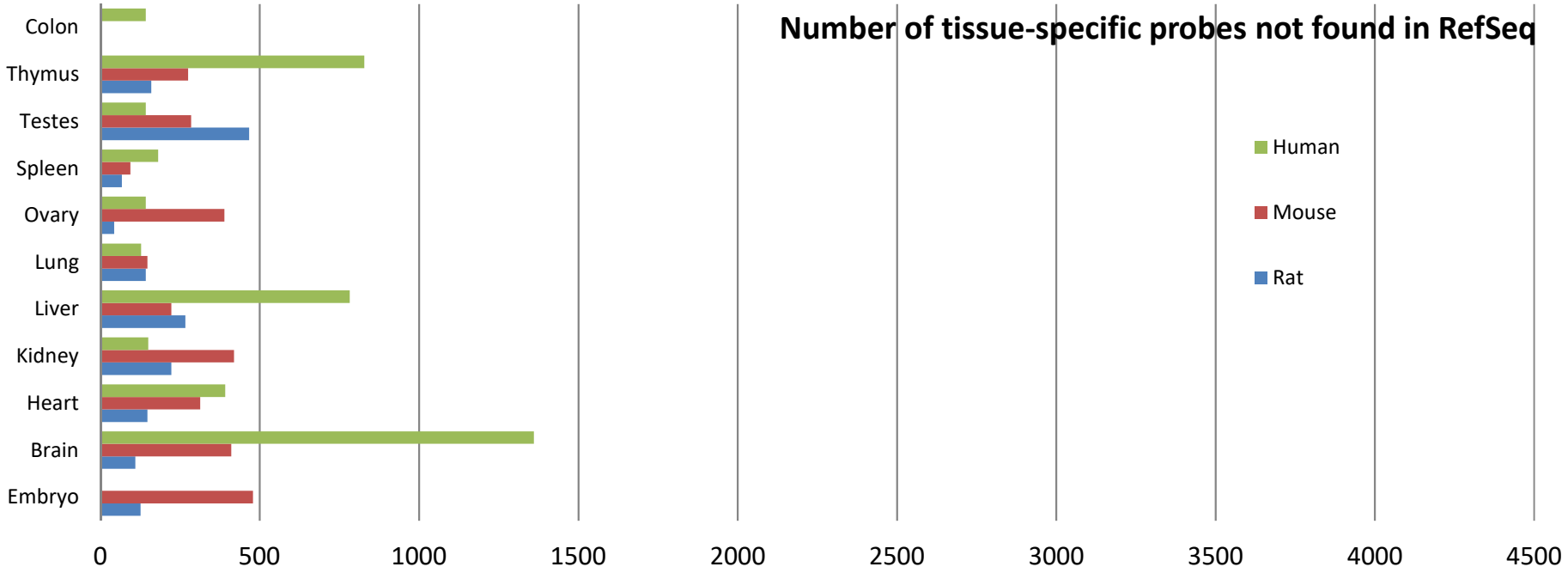
100% identity

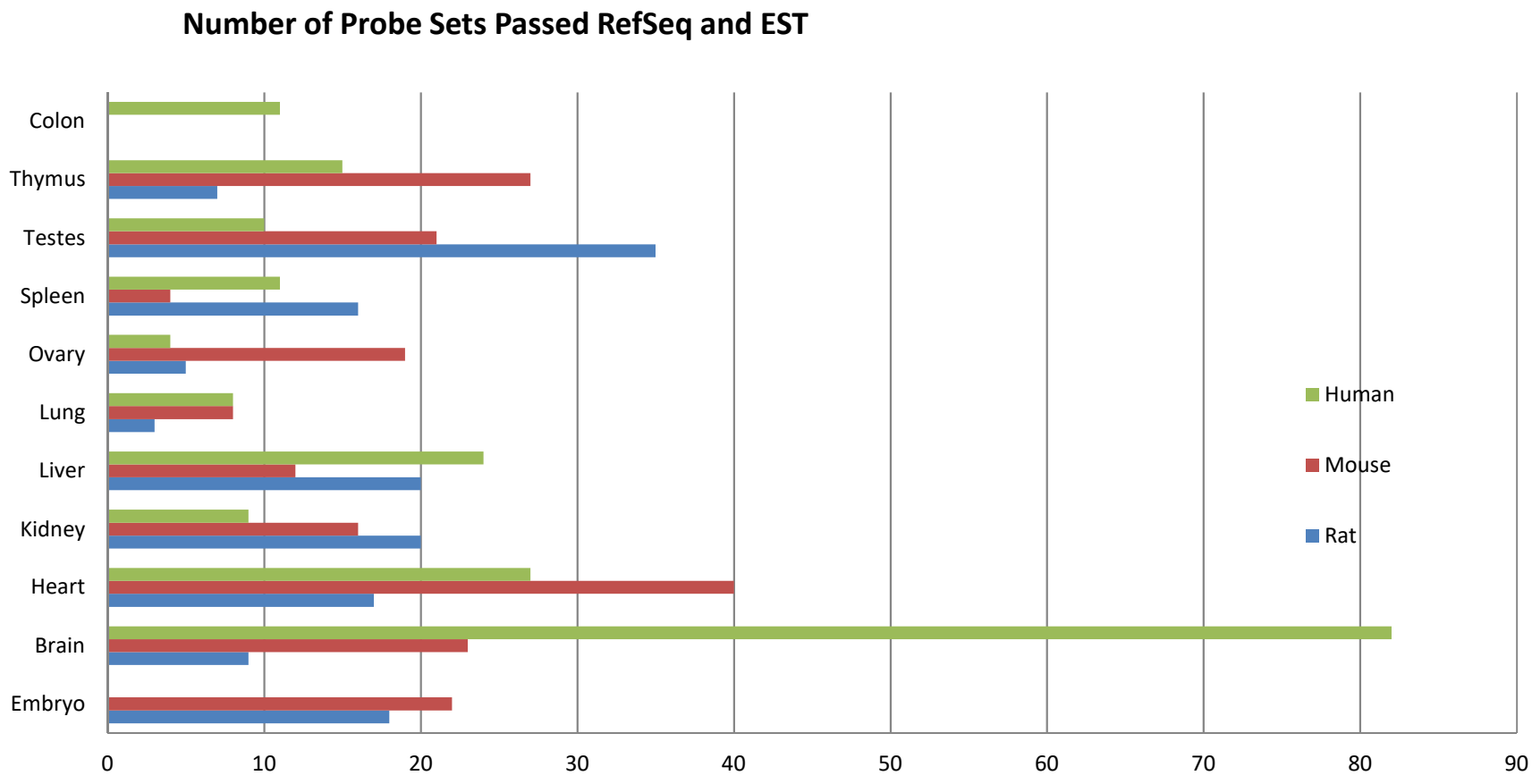
Criteria:

- Z-score > 2
- g1/g2 > 2



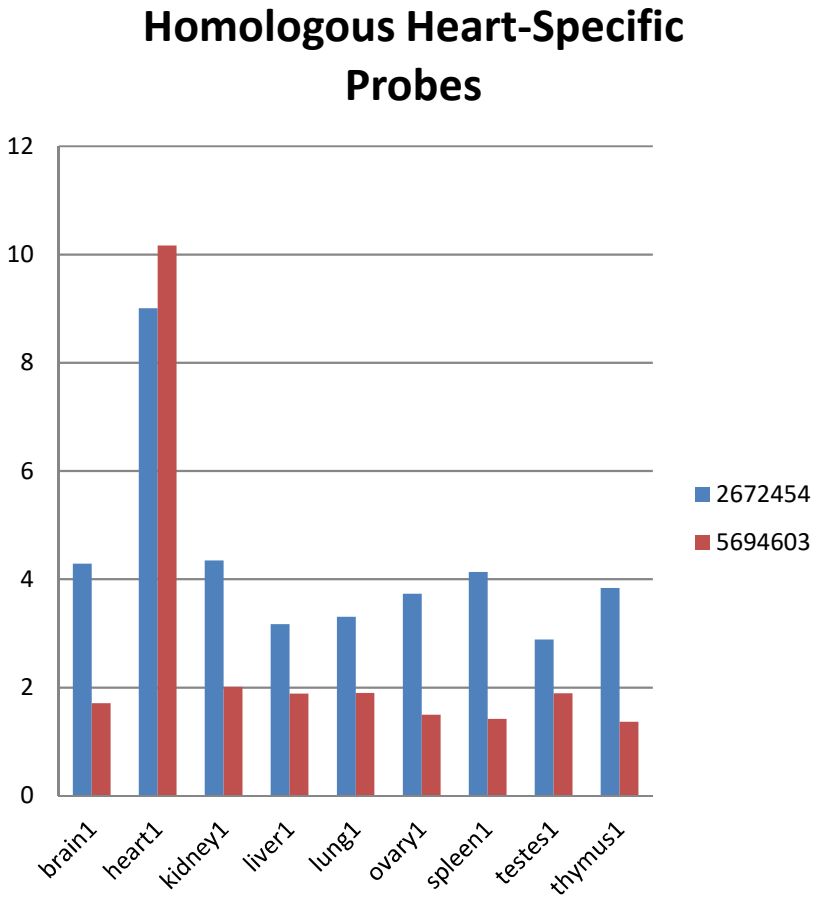

Number of tissue-specific probes (passed empirical statistics)
Colon
Thymus
Testes
Spleen
Ovary
Lung
Liver
Kidney
Heart
Brain
Embryo
Human
Mouse
Rat
0
500
1000
1500
2000
2500
3000
3500
4000
4500

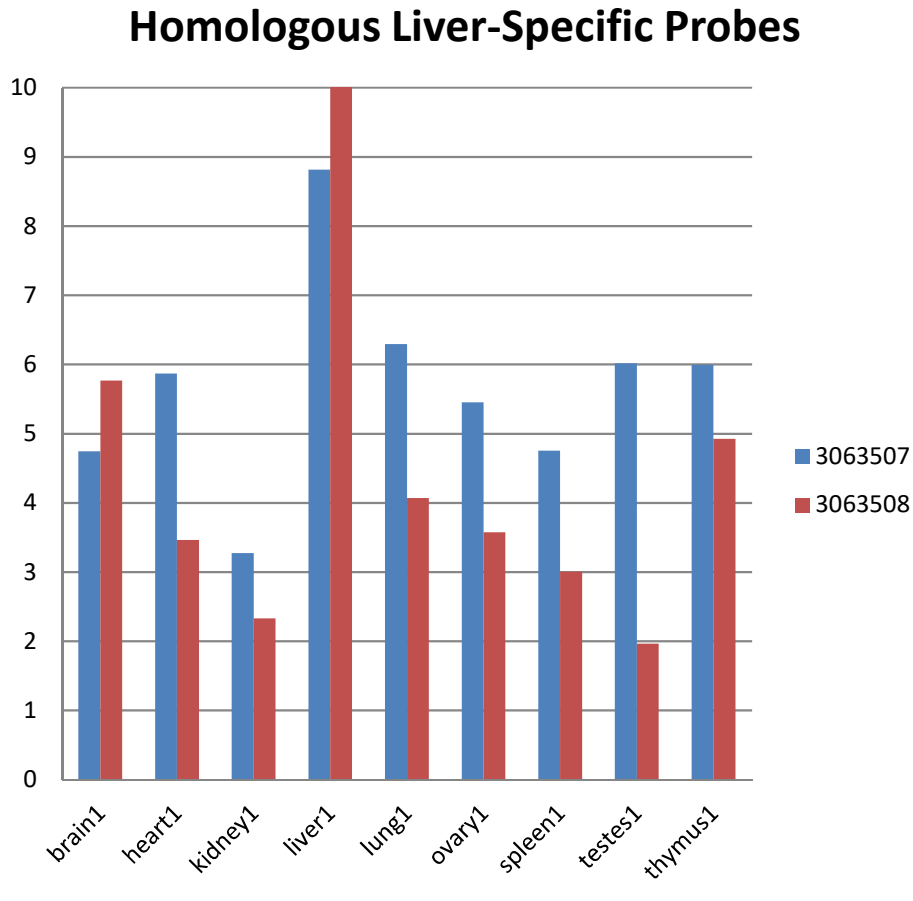

Number of tissue-specific probes not found in RefSeq
Colon
Thymus
Testes
Spleen
Ovary
Lung
Liver
Kidney
Heart
Brain
Embryo
Human
Mouse
Rat
0
500
1000
1500
2000
2500
3000
3500
4000
4500



- Threshold
  - Expectation < 1e-9, BLAST against RefSeq
  - Identity < 80%, BLAST against EST database

Tissue-Specific Homologous Exons for Human and Rat





Hypotheses so far

- NATs are actively regulated; thus, not transcriptional noise.
  - Expression of homologous antisense genes is statistically higher than non-homologous genes
  - Correlation is directly proportional to identity
  - Suggests that NATs are actively regulated

- Identify tissue-specific / novel NATs
  - 543 tissue-specific and novel NATs identified – 150 (rat), 192 (mouse), 201(human)

Future Work

- Experimentally verify a few tissue-specific NATs
- Comparing othologous antisense transcripts with orthologous sense transcripts
  - Are othologous antisense transcripts as rigorously regulated orthologous sense transcripts?
  - Speculatively, I don't think so

- Biological significance of NATs – their individual regulation and roles.