PseudoSeq Update

5/3/11

genome-anno

LH

Pseudogenes in GENCODE v6

- ENSEMBL / HAVANA transcripts (gt & tt: protein_coding): 74,226

- ENSEMBL pgenes (gt: pgene): 2,252

- HAVANA pgenes (gt: pgene): 10,343
- HAVANA pgenes (gt: pgene, tt: processed, not transcribed): **7,887**
- HAVANA pgenes (gt: pgene, tt: processed, transcribed): 24
- HAVANA pgenes (gt: pgene, tt: unprocessed, not transcribed): 1,793
- HAVANA pgenes (gt: pgene, tt: unprocessed, transcribed): 20

Alignment of pseudogenes to the reference genome

- Input: HAVANA pgenes (gt: pgene, tt: processed, not transcribed): **7,887**

- Alignment tool: BLAT

- Reference genome: hg19

- Extract all alignment blocks: at least one of the alignment block has to be longer than 75 nucleotides

Characterization of pseudogene alignments

- Pgenes with:
  - Zero alignment blocks: 2,670
  - One alignment blocks: 1,538
  - Multiple (2 - 5) alignment blocks: 1,593
  - Too many (> 5) alignment blocks: 2,086
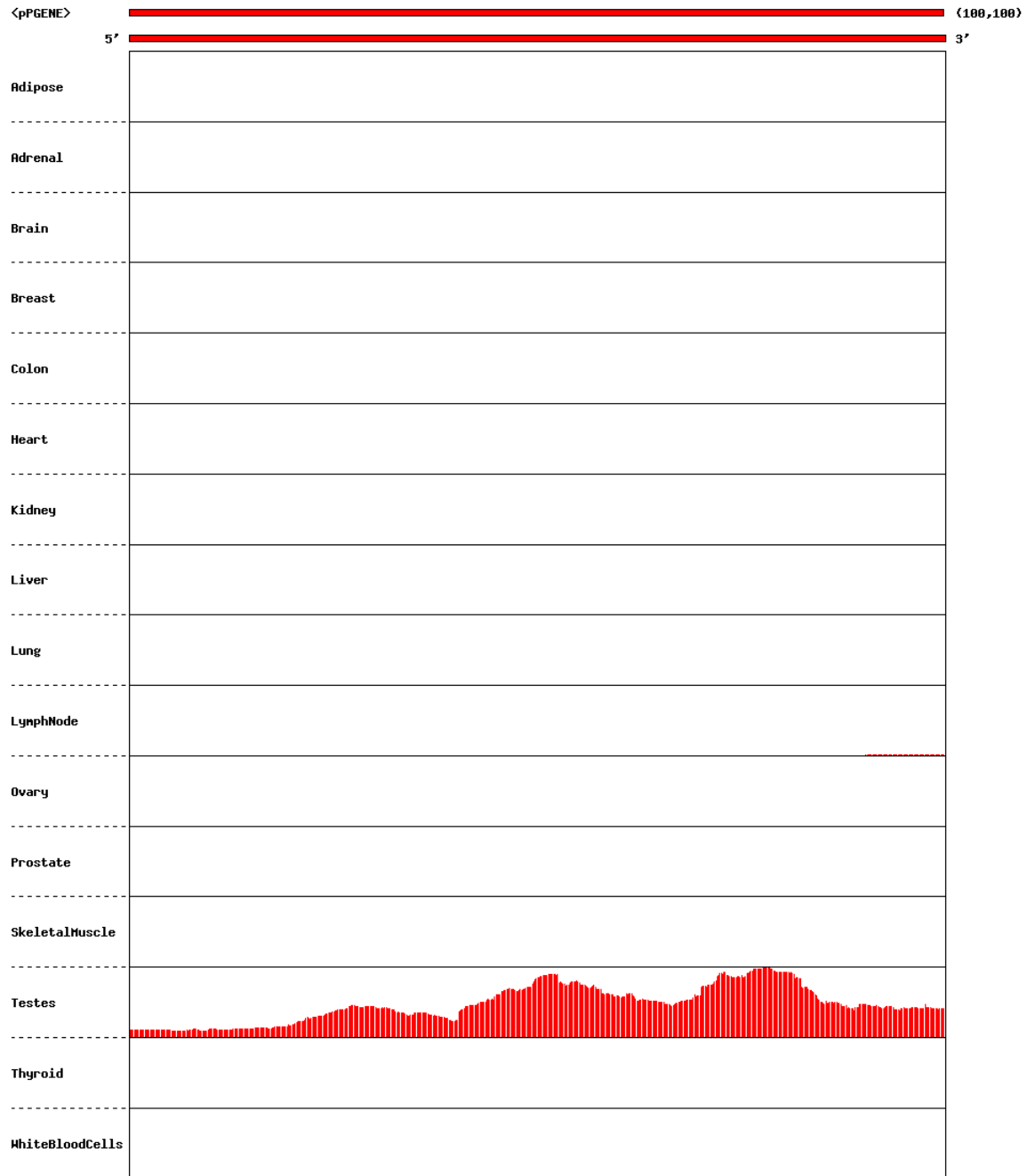
RNA-Seq data

- Human Body Map
  - 16 tissues
  - 75 nucleotide single-end reads
  - HiSeq Illumina platform: 1 lane per tissue

- Alignment tool: bowtie
  - hg19 + gencodeV6 splice junction library (alignment is performed concurrently)
  - Unique mapping
  - 2 mismatches (end-to-end)

- Average number of mapped reads per tissue: ~ 60M

Results

Category	Total number of pgenes	Number of pgenes with non-zero expression values	Number of pgenes with zero expression values	Number of pgenes with >0.5 expression values
Zero alignment pairs	2670	1530	1140	16
One alignment pair	1538	1165	373	33
2 - 5 alignment pairs	1593	1108	485	61
Too many alignment pairs	2086	N/A	N/A	N/A

Examples with zero alignment pairs

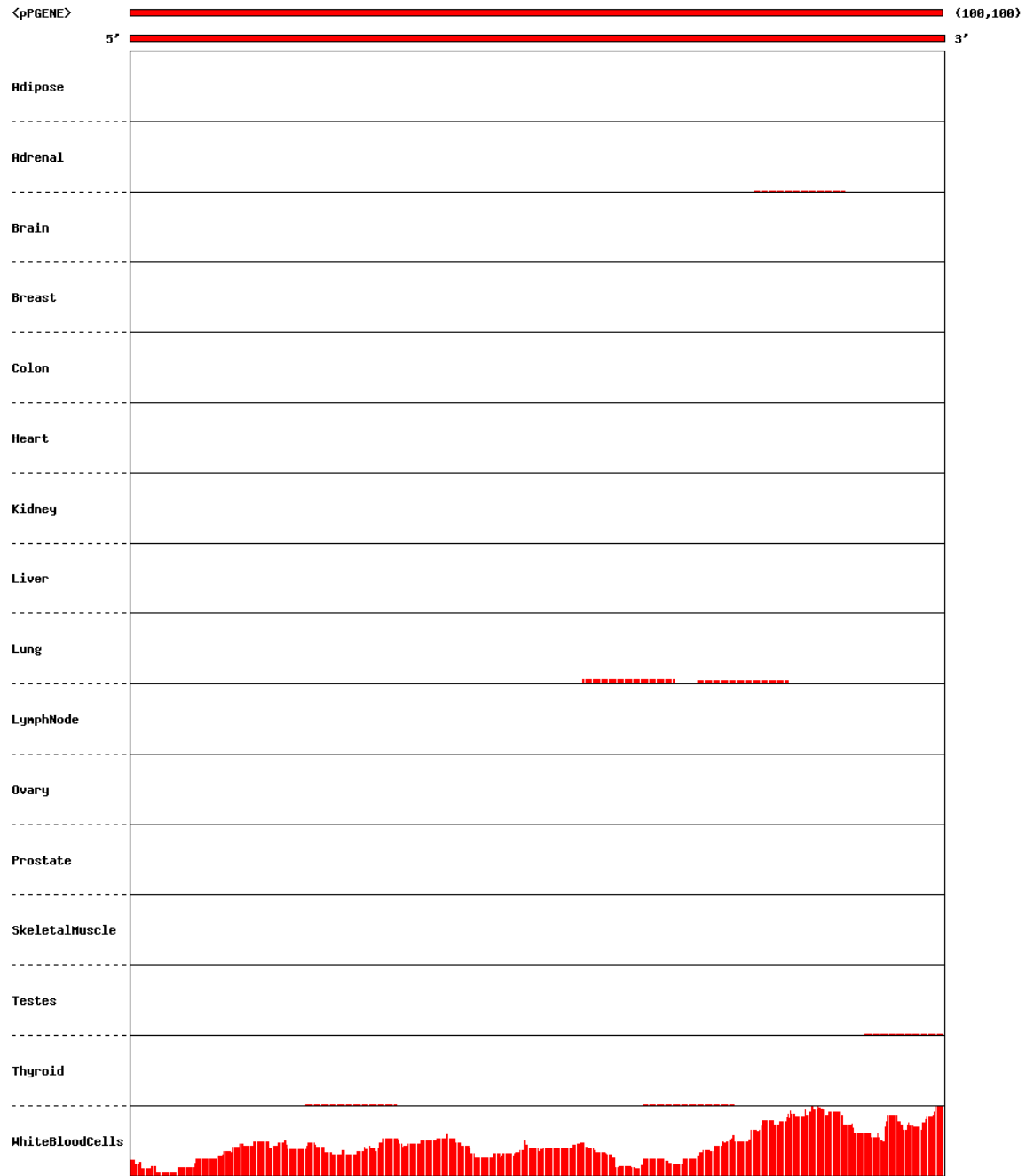Name: ENSG00000102055_ENST00000378131, Length: 609
Scale [0.00 5.49], Average: 0.16

<pPGENE>   (100,100)
         5'   3'

Adipose

Adrenal

Brain

Breast

Colon

Heart

Kidney

Liver

Lung

LymphNode

Ovary

Prostate

SkeletalMuscle

Testes

Thyroid

WhiteBloodCells

8
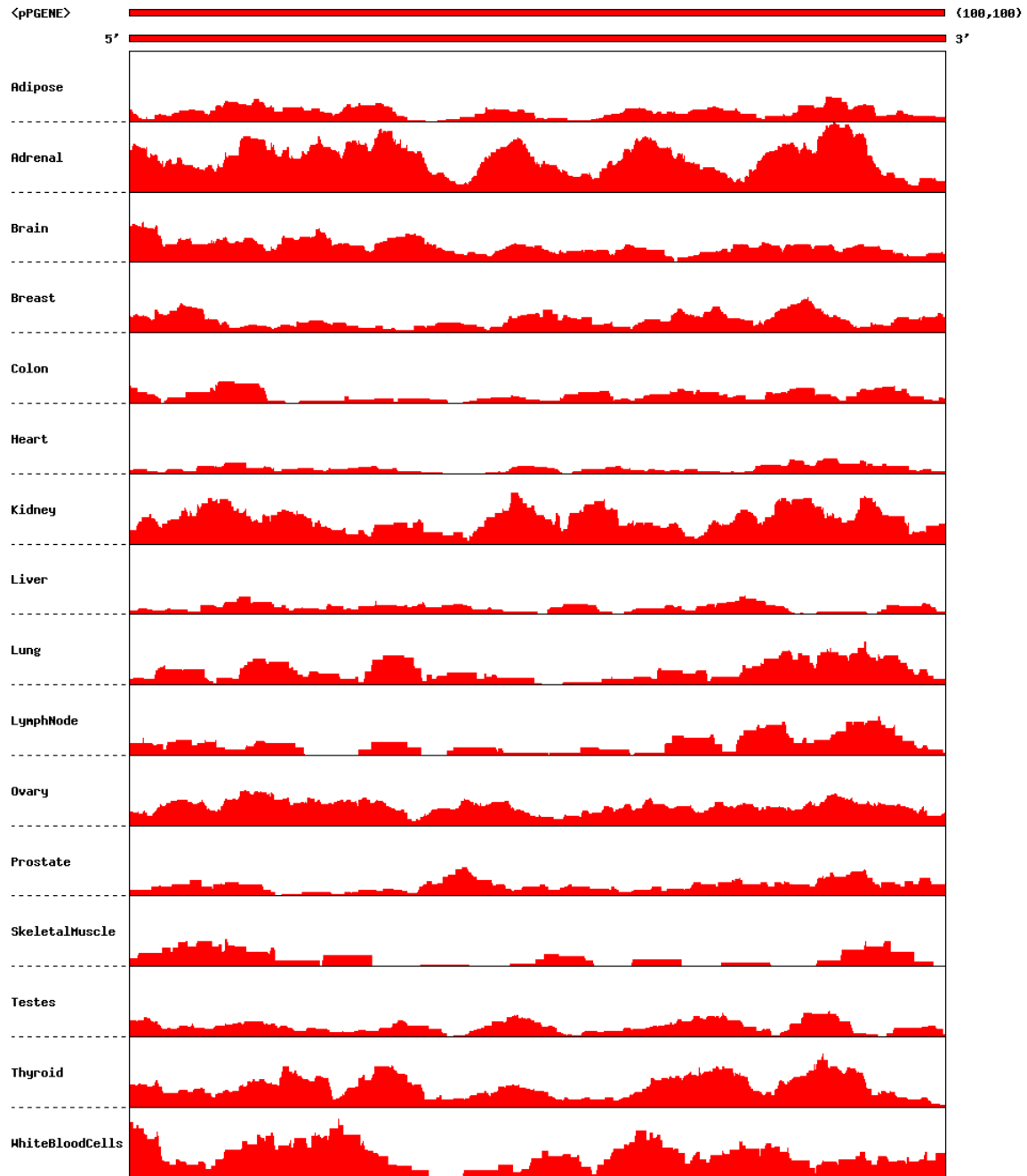
Name: ENSG00000213876_ENST00000483947, Length: 663
Scale [0.00 1.99], Average: 0.06

<pPGENE>   (100,100)

5'   3'

Adipose

Adrenal

Brain

Breast

Colon

Heart

Kidney

Liver

Lung

LymphNode

Ovary

Prostate

SkeletalMuscle

Testes

Thyroid

WhiteBloodCells

Name: ENSG00000235859_ENST00000496846, Length: 1234
Scale [0.00 1.52], Average: 0.32

<pPGENE>   (100,100)

5'   3'

Adipose

Adrenal

Brain

Breast

Colon

Heart

Kidney

Liver

Lung

LymphNode

Ovary

Prostate

SkeletalMuscle

Testes

Thyroid

WhiteBloodCells

10

Name: ENSG00000180812_ENST00000434057, Length: 1234
Scale [0.00 2.76], Average: 0.18
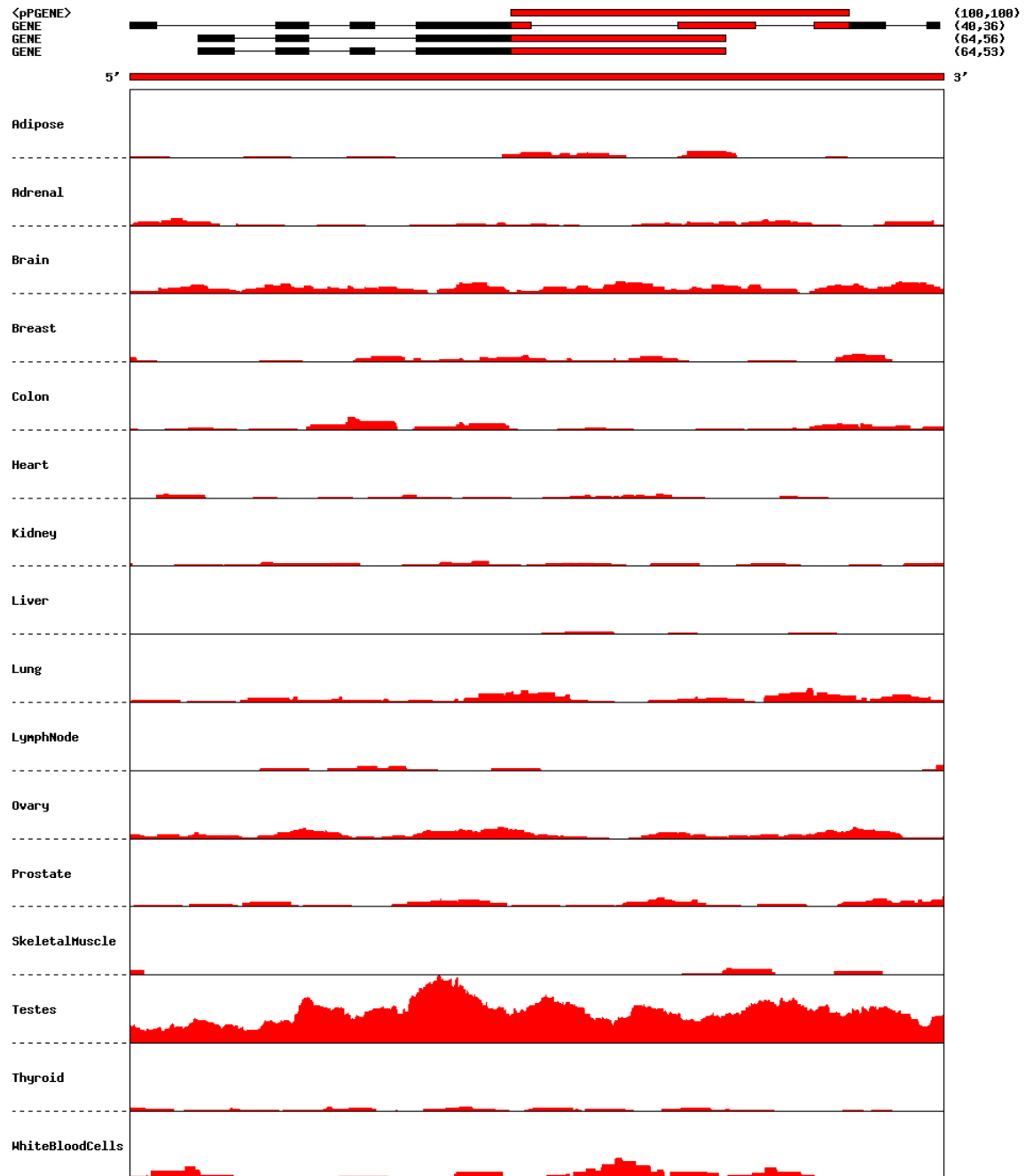
Name: ENSG00000242358_ENST00000487122, Length: 248
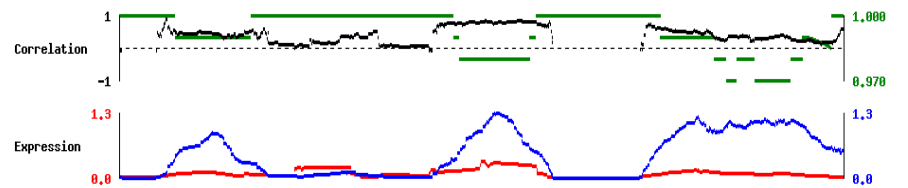Scale [0.00 1.61], Average: 0.25

<pPGENE>   (100,100)

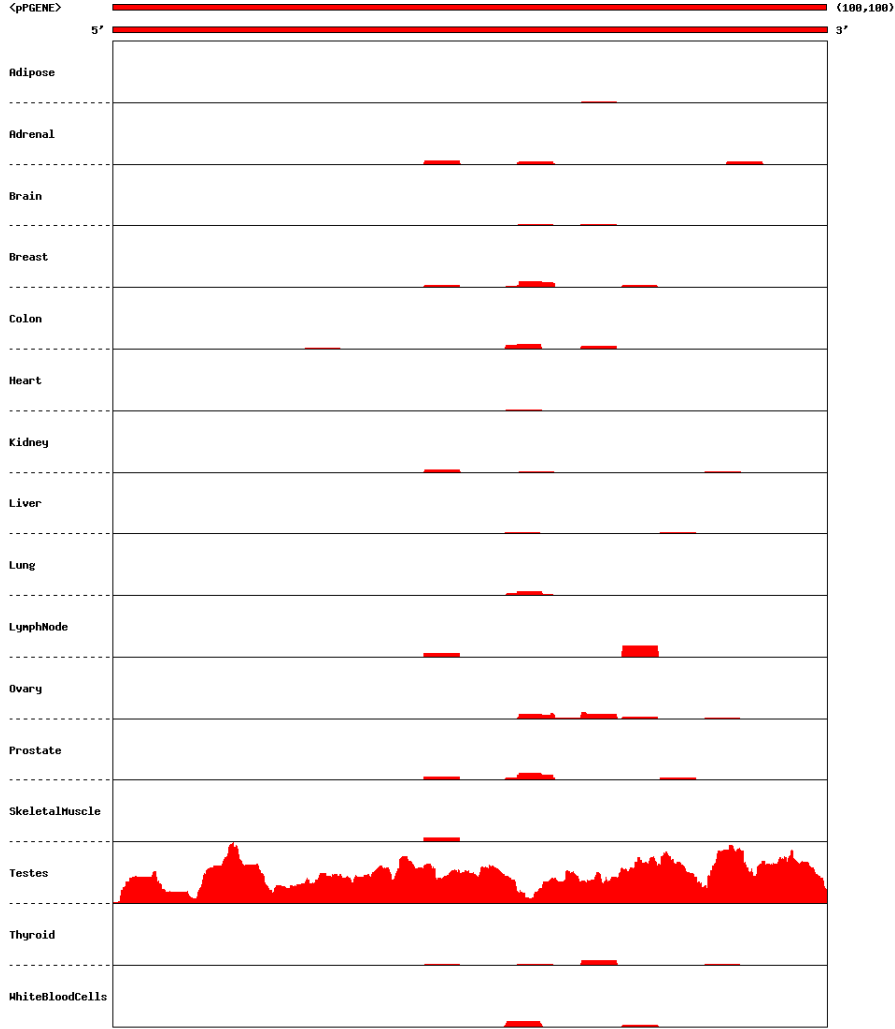   5'   3'

Adipose

Adrenal

Brain

Breast

Colon

Heart

Kidney

Liver

Lung

LymphNode

Ovary

Prostate

SkeletalMuscle

Testes

Thyroid

WhiteBloodCells

12
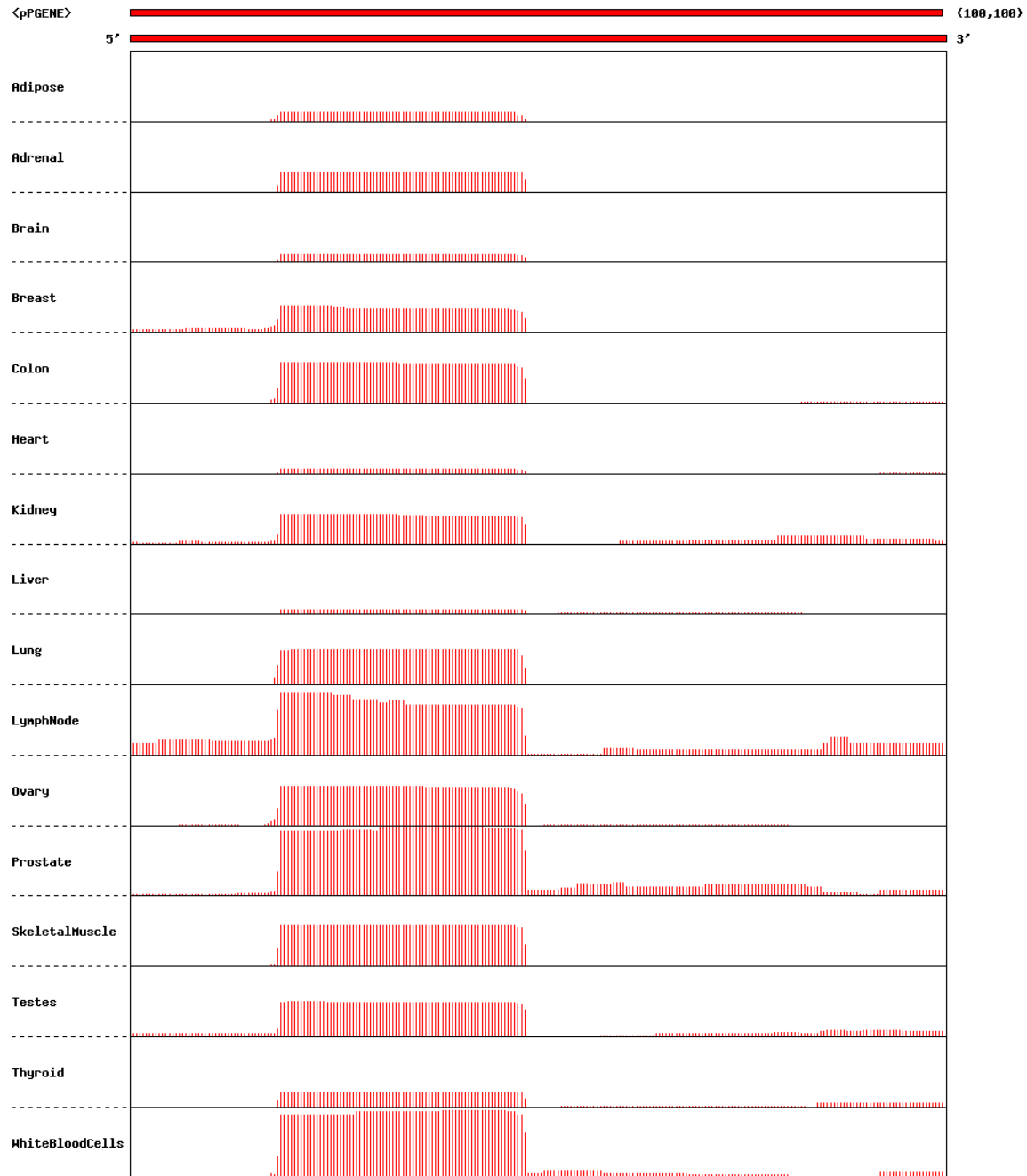
Examples with one alignment pair

Name: ENSG00000226268_ENST00000526165, Length: 959
Scale [0.00 2.85], Average: 0.08

<pPGENE>   (100,100)
5'   3'

Adipose

Adrenal

Brain

Breast

Colon

Heart

Kidney

Liver

Lung

LymphNode

Ovary

Prostate

SkeletalMuscle

Testes

Thyroid

WhiteBloodCells

Name: ENSG00000226268_ENST00000526165_1, Length: 959, %ID: 99.3, APC: 0.392, APWC: 0.084
Scale [0.00 4.34], Average: 0.08
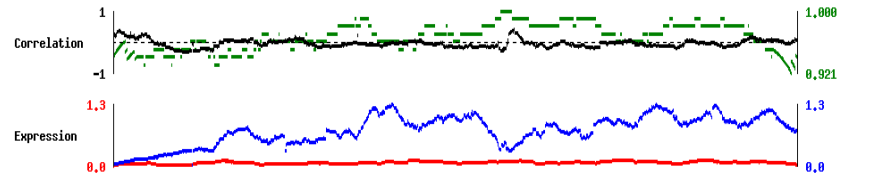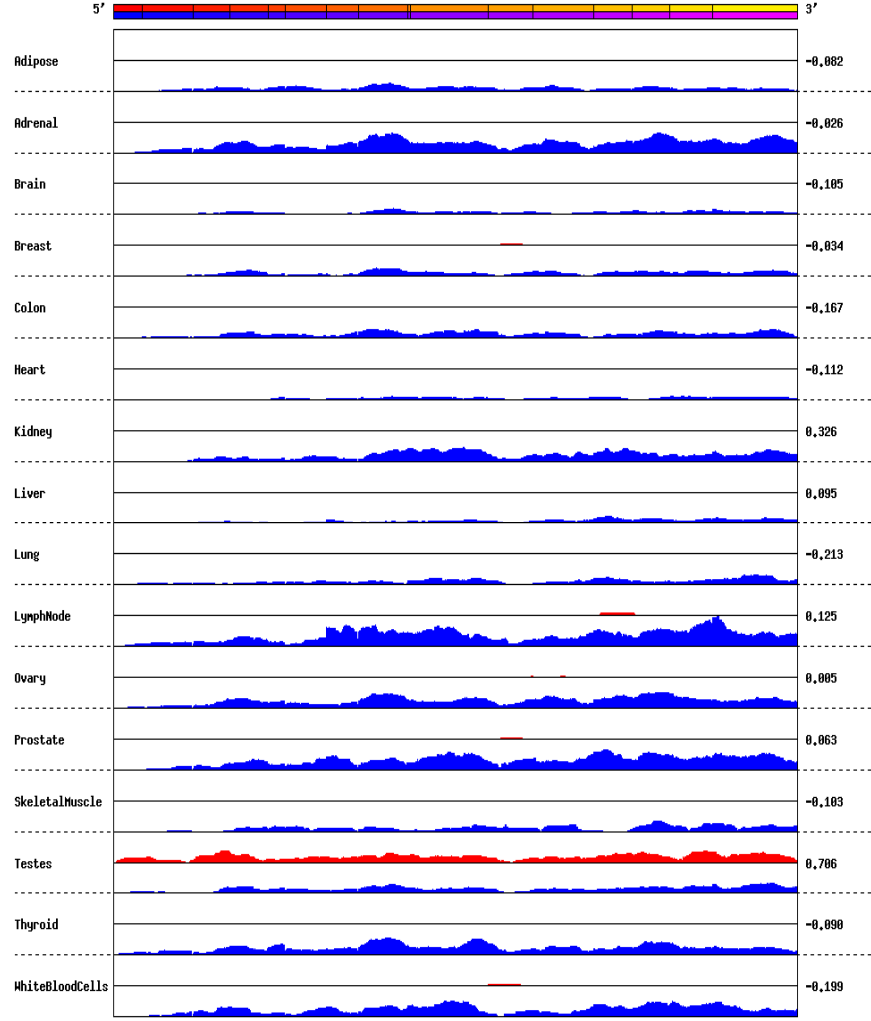Scale [0.00 4.34], Average: 0.43

<pPGENE>   (100,100)
GENE   (100,66)
5'   3'

Adipose   0.389

Adrenal   0.117

Brain   -0.282

Breast   0.375

Colon   -0.152

Heart   -0.184

Kidney   -0.007

Liver   -0.066

Lung   -0.250

LymphNode   0.074

Ovary   0.145

Prostate   -0.241

SkeletalMuscle   0.041

Testes   0.845

Thyroid   0.226

WhiteBloodCells   0.306

Correlation   1   1.000
             -1   0.970

Expression   1.3   1.3
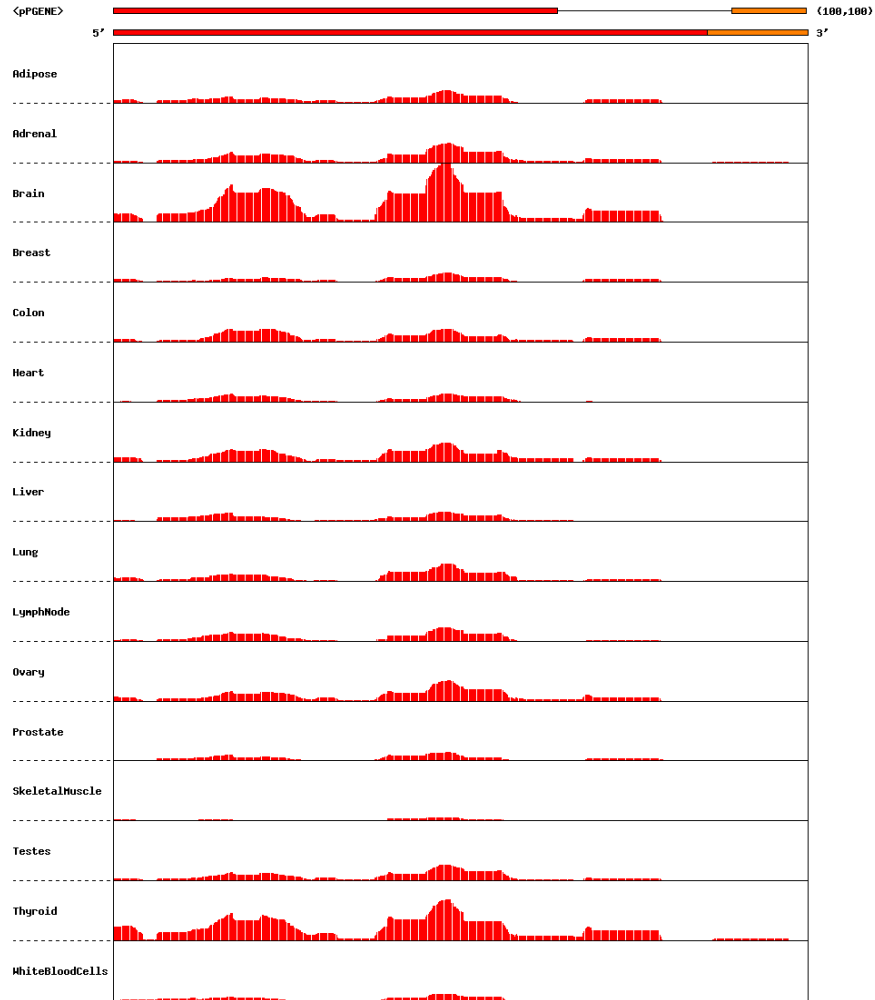   0.0   0.0

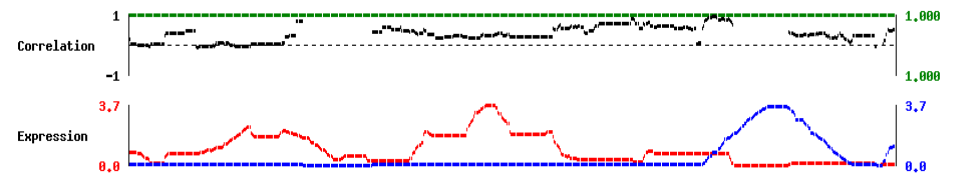Name: ENSG00000232553_ENST00000416636, Length: 1476
Scale [0.00 1.72], Average: 0.06

Name: ENSG00000232553_ENST00000416636_1, Length: 1461, %ID: 96.6, APC: -0.052, APWC: 0.012
Scale [0.00 4.02], Average: 0.06
Scale [0.00 4.02], Average: 0.73

〈pGENE〉 (100,100)
5' 3'

Adipose
Adrenal
Brain
Breast
Colon
Heart
Kidney
Liver
Lung
LymphNode
Ovary
Prostate
SkeletalMuscle
Testes
Thyroid
WhiteBloodCells

〈pGENE〉 (100,99)
GENE (100,67)
GENE (94,71)
GENE (100,69)
GENE (77,69)
5' 3'

Adipose -0.082
Adrenal -0.026
Brain -0.105
Breast -0.034
Colon -0.167
Heart -0.112
Kidney 0.326
Liver 0.095
Lung -0.213
LymphNode 0.125
Ovary 0.005
Prostate 0.063
SkeletalMuscle -0.103
Testes 0.706
Thyroid -0.090
WhiteBloodCells -0.199

Correlation
1
-1
1.000
0.921

Expression
1.3
0.0
1.3
0.0

Name: ENSG00000174028_ENST00000312214, Length: 681
Scale [0.00 12.55], Average: 0.89

Name: ENSG00000174028_ENST00000312214_1, Length: 670, %ID: 100.0, APC: 0.352, APWC: -0.375
Scale [0.00 12.55], Average: 0.89
Scale [0.00 12.55], Average: 0.42

Left panel:

⟨pPGENE⟩ (100,100)
5' 3'

Tissue	
Adipose	
Adrenal	
Brain	
Breast	
Colon	
Heart	
Kidney	
Liver	
Lung	
LymphNode	
Ovary	
Prostate	
SkeletalMuscle	
Testes	
Thyroid	
WhiteBloodCells	

Right panel:

⟨pPGENE⟩ (100,98)
GENE (100,27)
GENE (66,79)
GENE (60,61)
GENE (32,47)
GENE (100,27)
5' 3'

Tissue	Value
Adipose	-0.429
Adrenal	-0.359
Brain	-0.382
Breast	-0.417
Colon	-0.418
Heart	-0.391
Kidney	-0.420
Liver	-0.381
Lung	-0.351
LymphNode	-0.344
Ovary	-0.396
Prostate	-0.320
SkeletalMuscle	-0.261
Testes	-0.398
Thyroid	-0.423
WhiteBloodCells	-0.310

Correlation: 1, -1, 1.000, 1.000
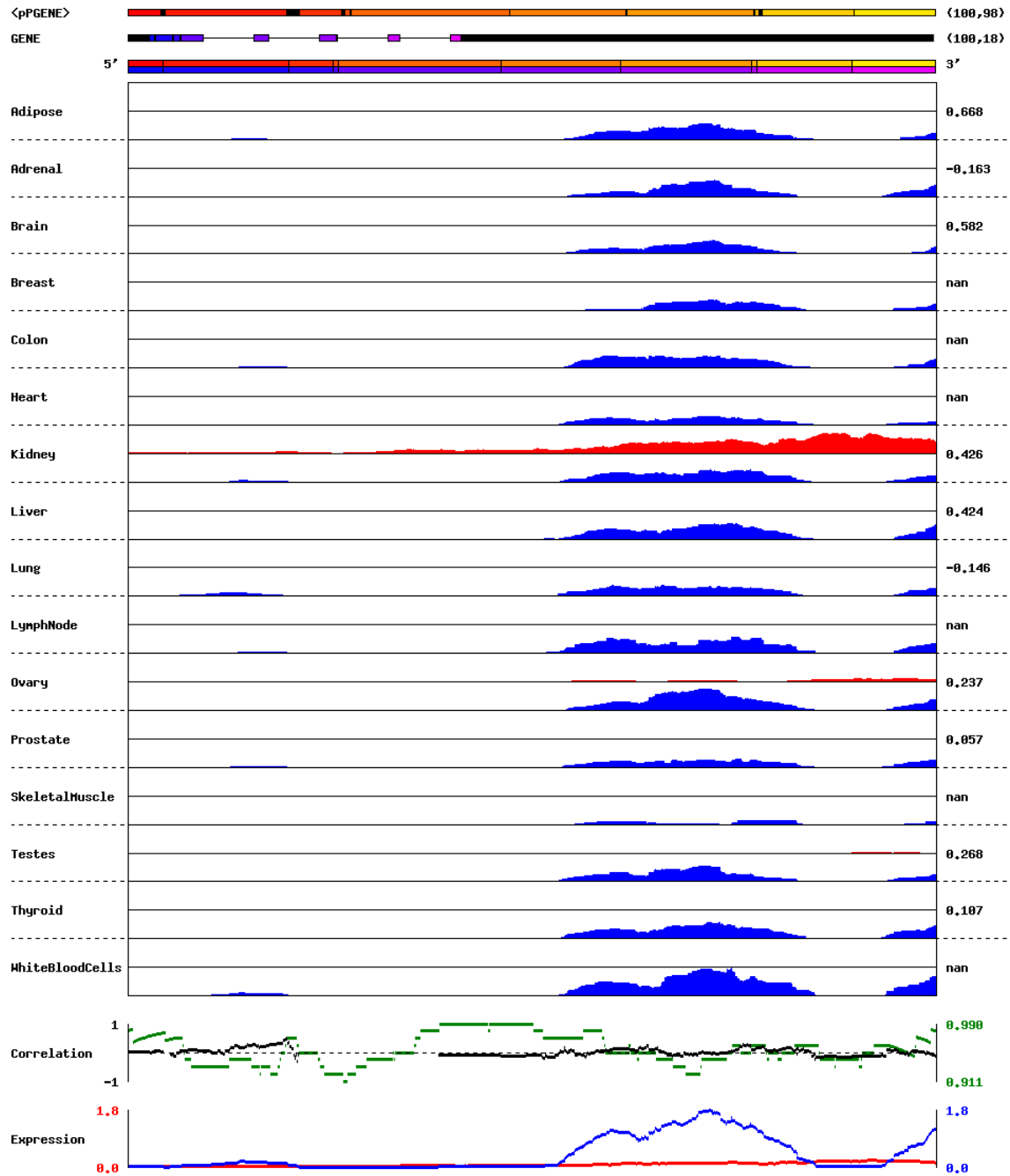Expression: 3.7, 0.0, 3.7, 0.0

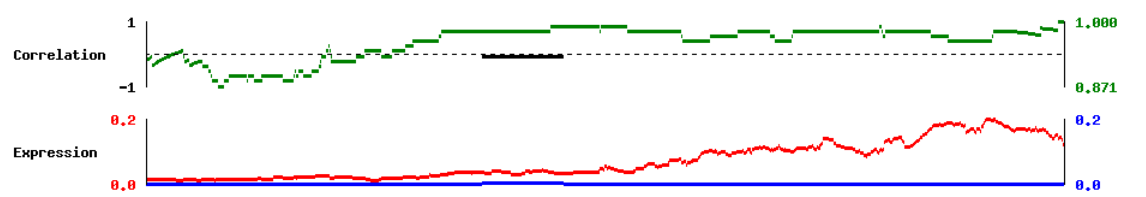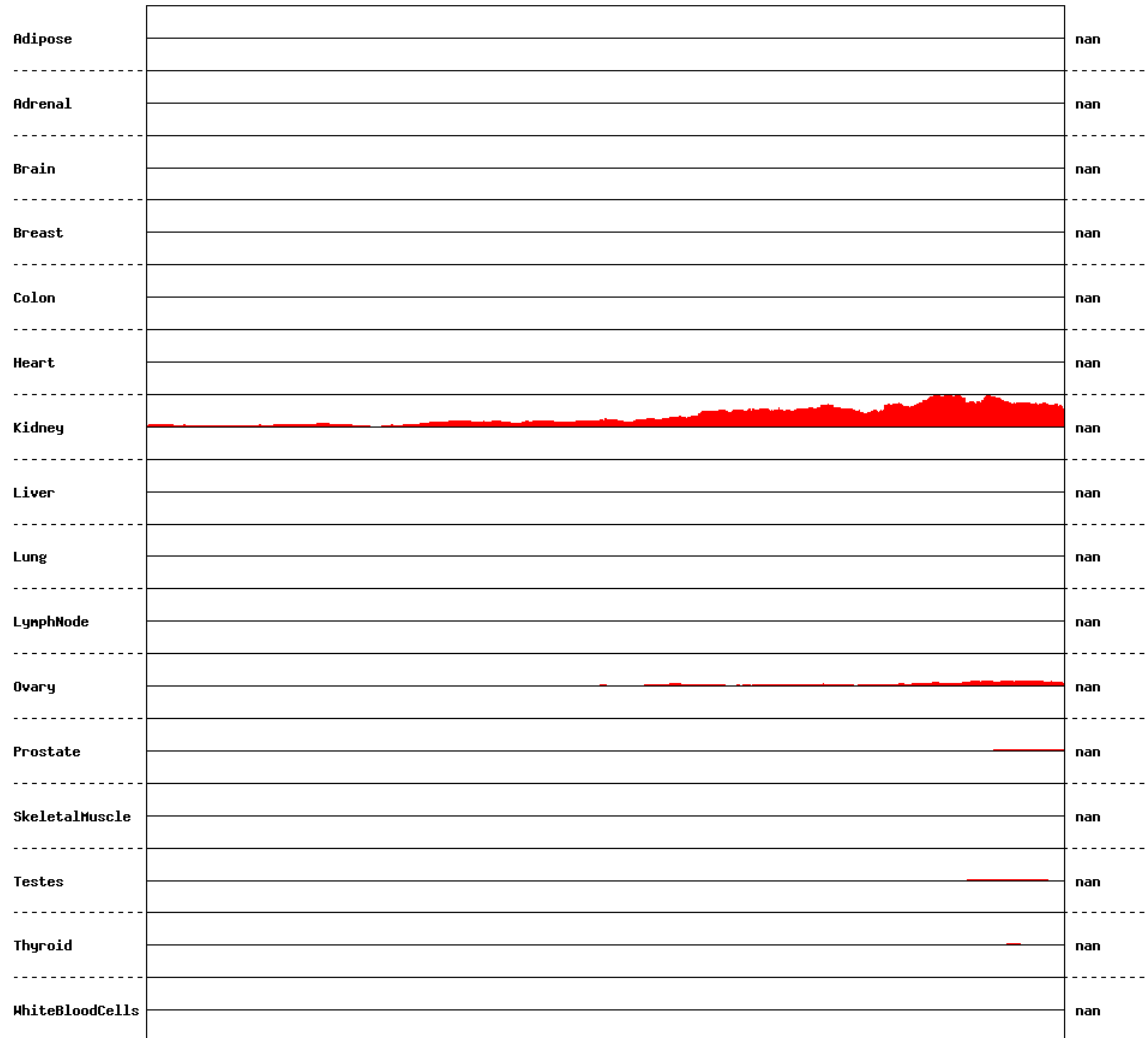# Example with multiple alignment pairs

Name: ENSG00000234685_ENST00000417358, Length: 873
Scale [0.00 2.78], Average: 0.07

18
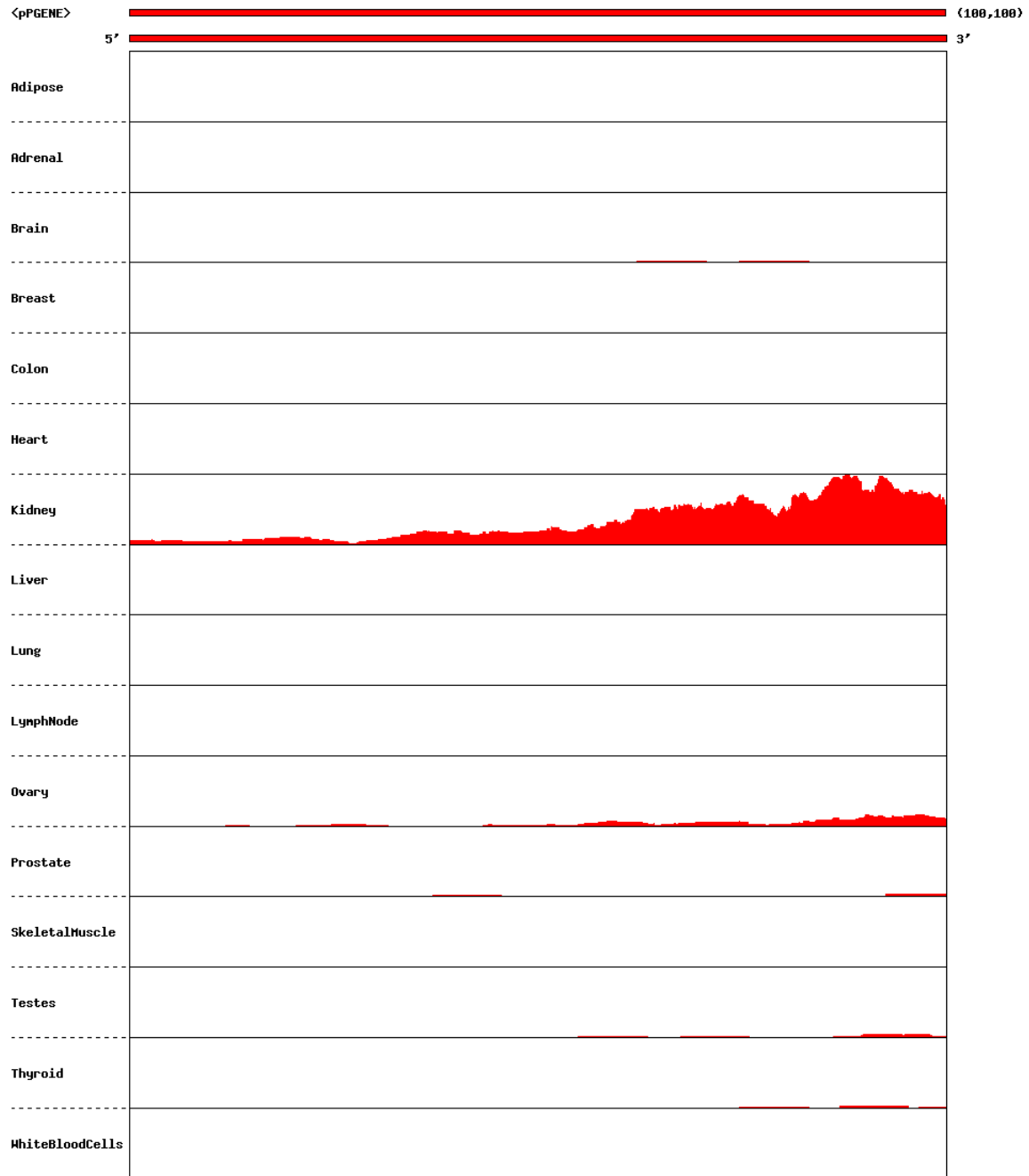
Name: ENSG00000234685_ENST00000417358_2, Length: 852, %ID: 95.5, APC: 0.011, APWC: 0.246
Scale [0.00 3.79], Average: 0.07
Scale [0.00 3.79], Average: 0.39

19

Name: ENSG00000234685_ENST00000417358_1, Length: 844, %ID: 95.9, APC: -0.075, APWC: nan
Scale [0.00 2.78], Average: 0.07
Scale [0.00 2.78], Average: 0.00

<pPGENE>   (100,97)

pPGENE   (100,98)

5'   3'

Adipose	nan
Adrenal	nan
Brain	nan
Breast	nan
Colon	nan
Heart	nan
Kidney	nan
Liver	nan
Lung	nan
LymphNode	nan
Ovary	nan
Prostate	nan
SkeletalMuscle	nan
Testes	nan
Thyroid	nan
WhiteBloodCells	nan

Correlation   1   1.000
               -1   0.871

Expression   0.2   0.2
               0.0   0.0

20

Name: ENSG00000234685_ENST00000417358_3, Length: 854, %ID: 95.1, APC: 0.129, APWC: 0.193
Scale [0.00 2.78], Average: 0.07
Scale [0.00 2.78], Average: 0.05

21