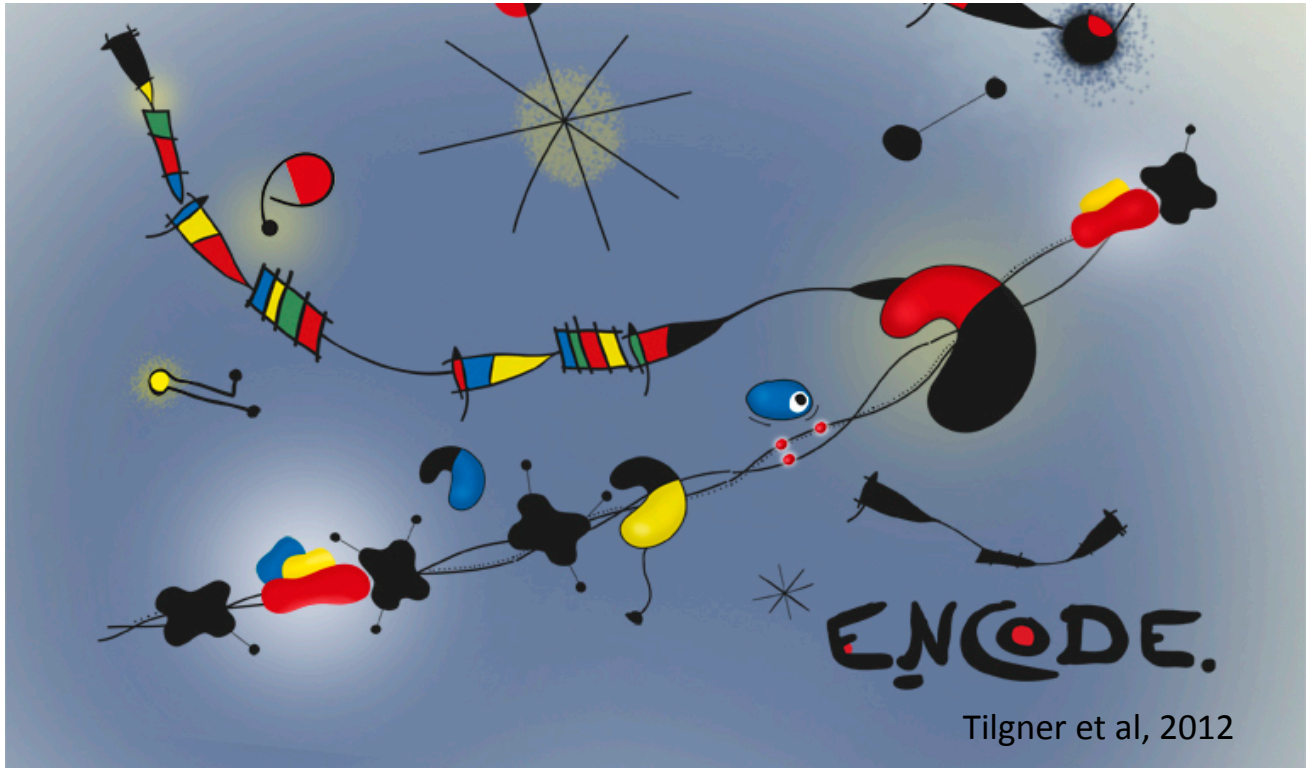
What genes are saying – not how much they talk



Hagen Tilgner
Weill Cornell Medicine
New York City

All artwork by
Luisa Lente

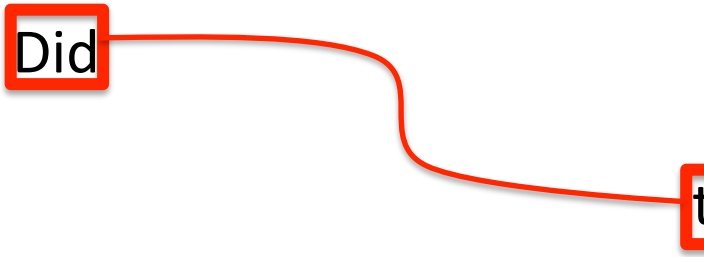
hut2006@med.cornell.edu

Did dahadhd babejendnk lld lfiwndjhf uecsjbn nlada da Msd nmnjak Kaawttwrqeqfabxhdnryjsowpkqmqnajsuskwmwjwqkqjausnhdhsama Anabxvxbchteioaplamanz thee feel haa mma naba fqwrwyr iorpplkn Cnbcgsgtf yjtndsy jjiskspa pin bfbbzbvzcz cafrsthanaj aiaapalakansns Msn shsauam abx sj znxjs cksdjsla lna by jbb doidn sssnd cn mdjd cn Cncj djfd kejssi sosdn nusjs the mdjdjdf neidi vjccie hkcbe ksbbc skbs Skcn sksksk nskclss nsjsjs sjl smcdll dxllc dlvkkc dmkslzlz lxmcx mx me Sbcdncsd nsnscn cssnc kcdkhd dnz,m naw shdk hdssbs mbds bfksd s Bcgy cdgsjhfskg hsapgba gkjawyuf glishzdh fzdhf z.g h.zgh szp jgshg. Bczbv bavab jcwasv jbvsj earth dsahls hvnlanv dklnbkv dsnvd askl bv Jszbkd.jv xhcj.kzsb klsdhfpw uapgfwgef ,jgfdma sgfmgdsvfa hgsf jdaG Jfbskjsdh fkwhslszhfjls hflizhflzdhfk.

hut2006@med.cornell.edu

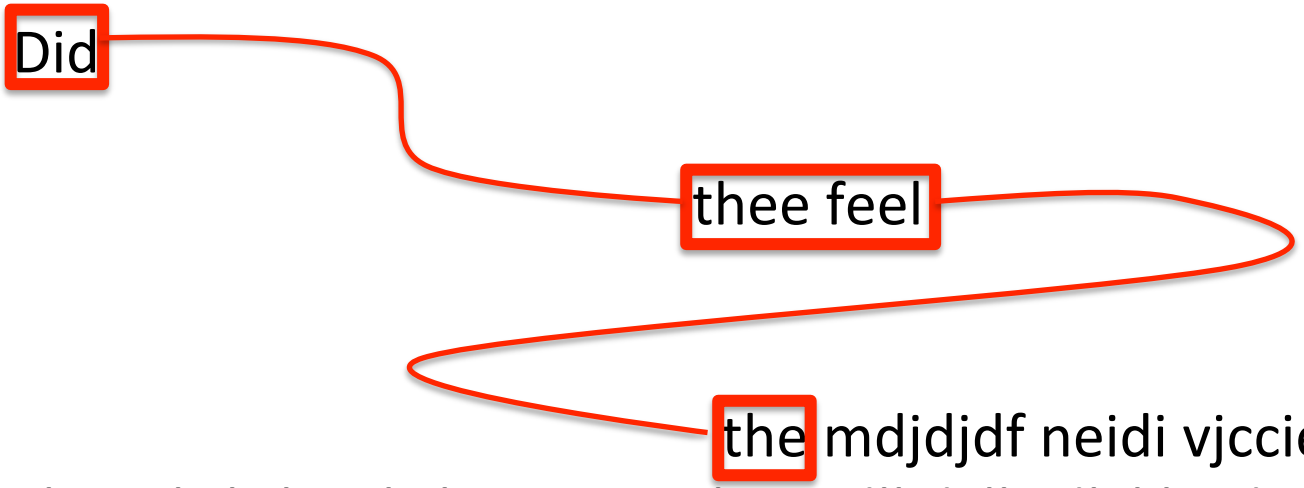
Did dahadhd babejendnk lld lfiwndjhf uecsjbn nlada da Msd nmnjak
Kaawttwrqeqfabxhdnryjsowpkqmqnajsuskwmwjwqkqjausnhdhsama
Anabxvxbchteioaplamanz thee feel haa mma naba fqwrwyr iorpplkn
Cnbcgsgtf yjtndsy jjiskspa pin bfbbzbvzcz cafrsthanaj aiaapalakansns
Msn shsauam abx sj znxjs cksdjsla lna by jbb doidn sssnd cn mdjd cn
Cncj djfd kejssi sosdn nusjs the mdjdjdf neidi vjccie hkcbe ksbbc skbs
Skcn sksksk nskclss nsjsjs sjl smcdll dxllc dlvkkc dmkslzlz lxmcx mx me
Sbcdncsd nsnscn cssnc kcdkhd dnz,m naw shdk hdssbs mbds bfksd s
Bcgy cdgsjhfskg hsapgba gkjawyuf glishzdh fzdhf z.g h.zgh szp jgshg.
Bczbv bavab jcwasv jbvsj earth dsahls hvnlanv dklnbkv dsnvd askl bv
Jszbkd.jv xhcj.kzsb klsdhfpw uapgfwgef ,jgfdma sgfmgdsvfa hgsf jdaG
Jfbskjsdh fkwhslszhfjls hflizhflzdhfk. move

hut2006@med.cornell.edu



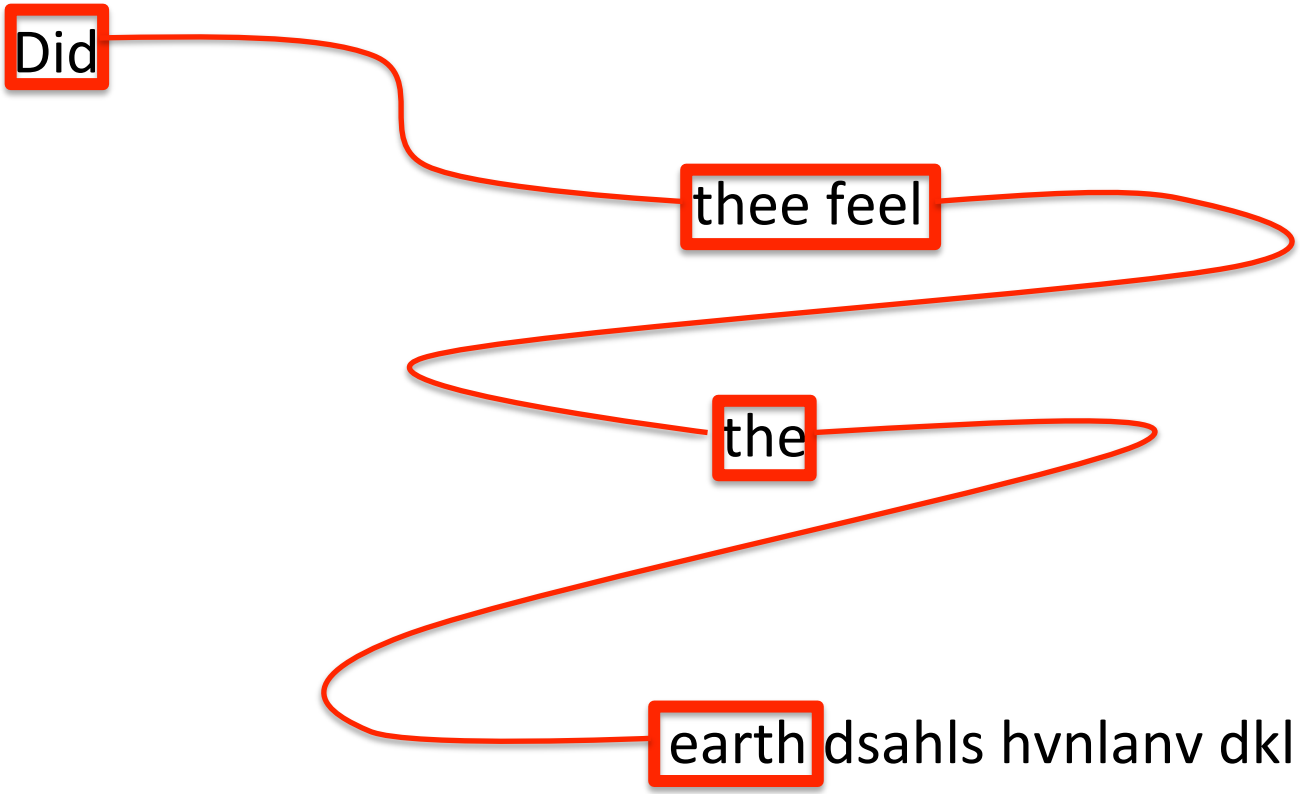
thee feel haa mma naba fqwrwyr iorpplkn
Cnbcgsgtf yjtndsy jjiskspa pin bfbbzbvzcz cafrsthanaj aiaapalakansns
Msn shsauam abx sj znxjs cksdjsla lna by jbb doidn sssnd cn mdjd cn
Cncj djfd kejssi sosdn nusjs the mdjdjdf neidi vjccie hkcbe ksbbc skbs
Skcn sksksk nskclss nsjsjs sjl smcdll dxllc dlvkkc dmkslzlz lxmcx mx me
Sbcdncsd nsnscn cssnc kcdkhd dnz,m naw shdk hdssbs mbds bfksd s
Bcgy cdgsjhfskg hsapgba gkjawyuf glishzdh fzdhf z.g h.zgh szp jgshg.
Bczbv bavab jcwasv jbvsj earth dsahls hvnlanv dklnbkv dsnvd askl bv
Jszbkd.jv xhcj.kzsb klsdhfpw uapgfwgef ,jgfdma sgfmgdsvfa hgsf jdaG
Jfbskjsdh fkwhslszhfjls hflizhflzdhfk. move

hut2006@med.cornell.edu



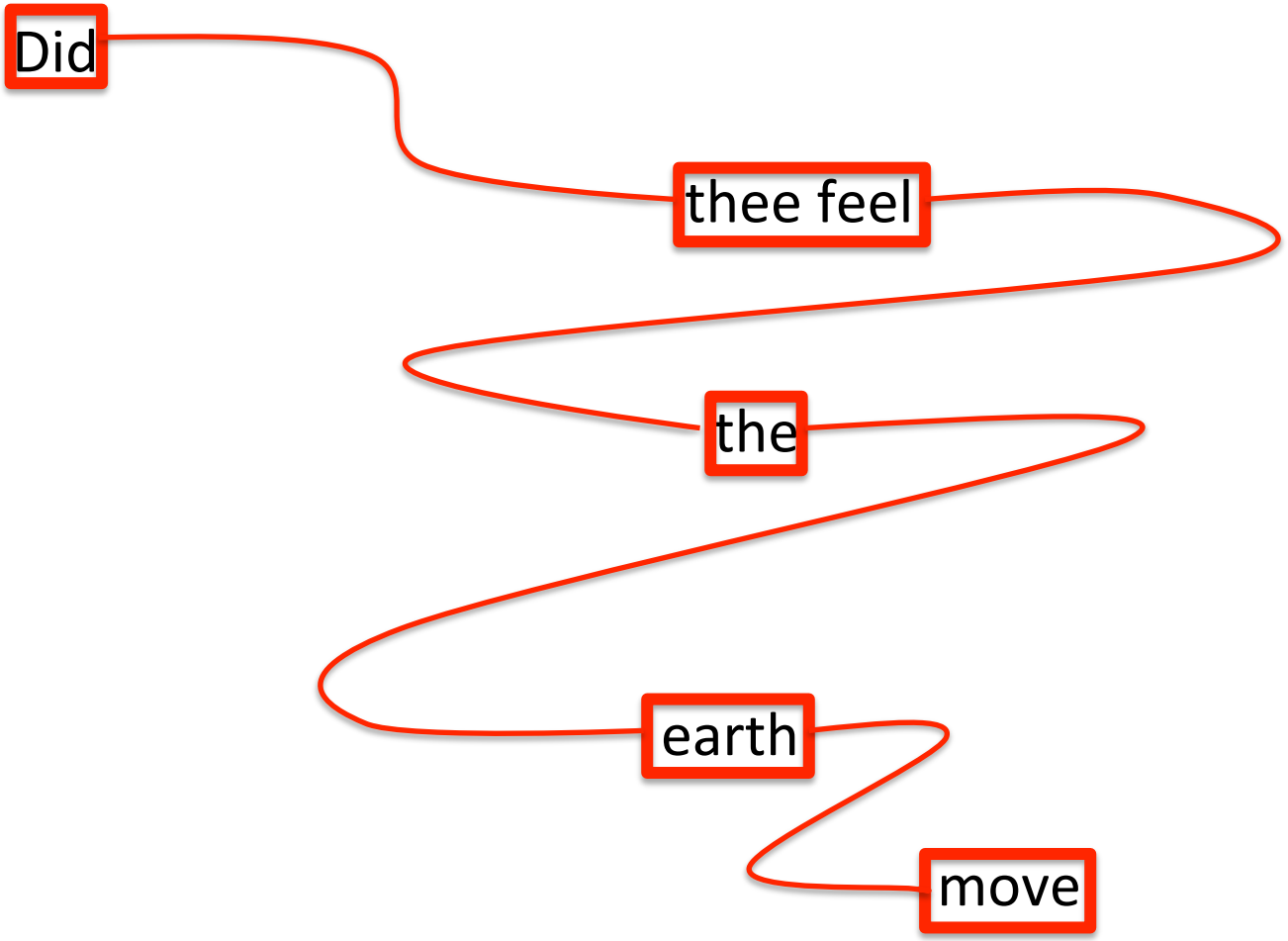
mdjdjdf neidi vjccie hkcbe ksbbc skbs Skcn sksksk nskclss nsjsjs sjl smcdll dxllc dlvkkc dmkslzlz lxmcx mx me Sbcdncsd nsnscn cssnc kcdkhd dnz,m naw shdk hdssbs mbds bfksd s Bcgy cdgsjhfskg hsapgba gkjawyuf glishzdh fzdhf z.g h.zgh szp jgshg. Bczbv bavab jcwasv jbvsj earth dsahls hvnlanv dklnbkv dsnvd askl bv Jszbkd.jv xhcj.kzsb klsdhfpw uapgfwgef ,jgfdma sgfmgdsvfa hgsf jdaG Jfbskjsdh fkwhslszhfjls hflizhflzdhfk. move

hut2006@med.cornell.edu

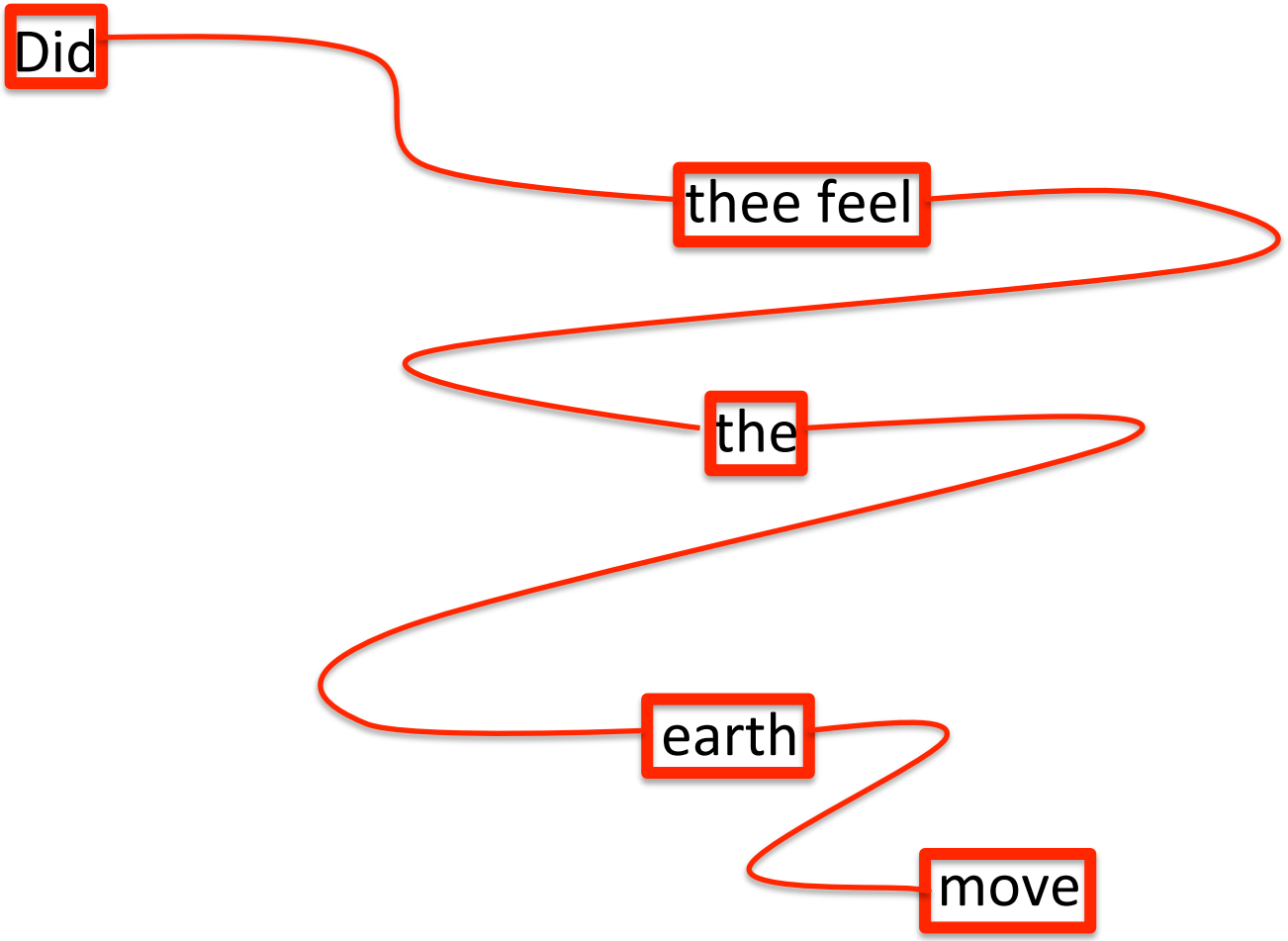


dsahls hvnlanv dklnbkv dsnvd askl bv Jszbkd.jv xhcj.kzsb klsdhfpw uapgfwgef ,jgfdma sgfmgdsvfa hgsf jdaG Jfbskjsdh fkwhslszhfjls hflizhflzdhfk. move

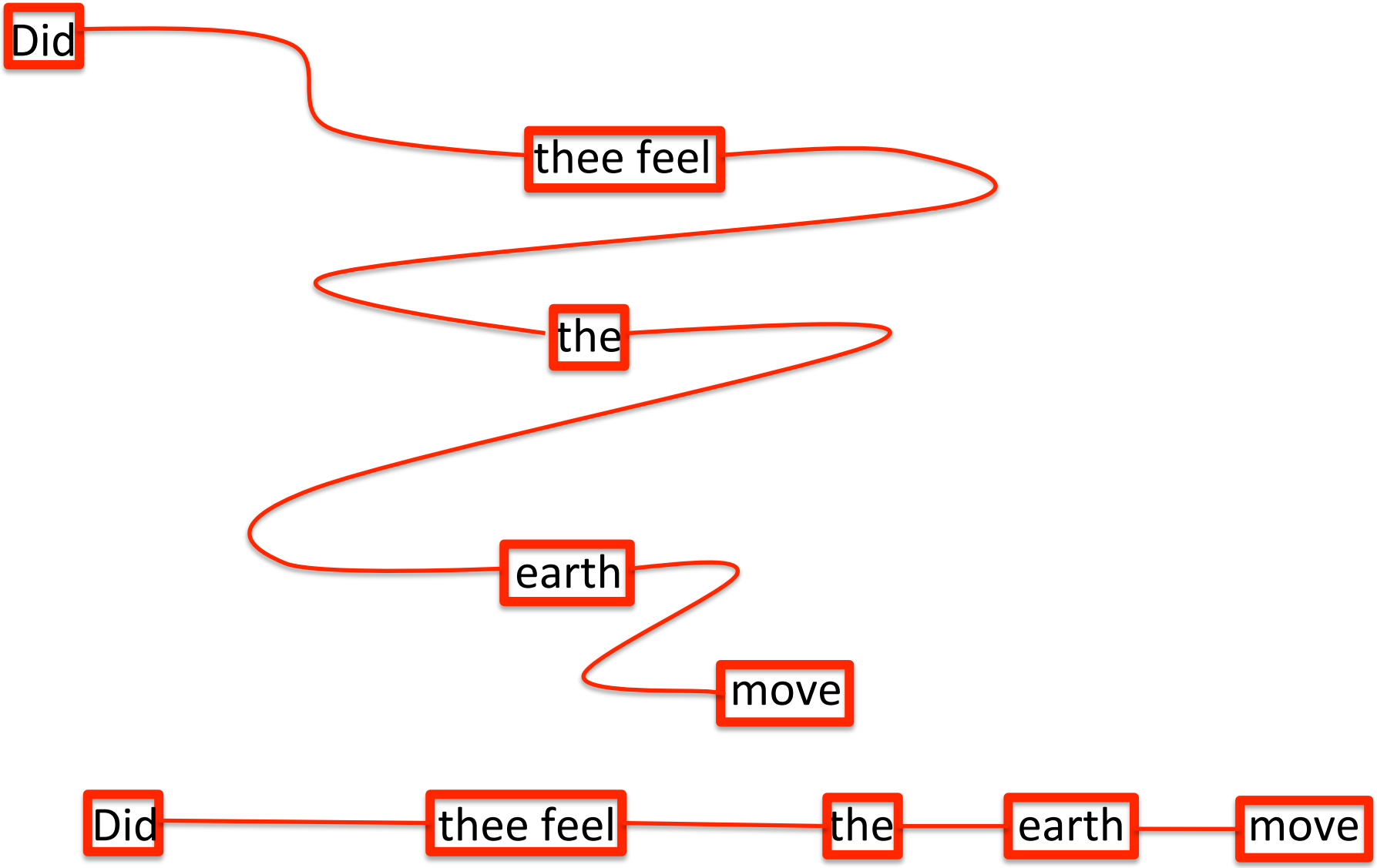
hut2006@med.cornell.edu



hut2006@med.cornell.edu

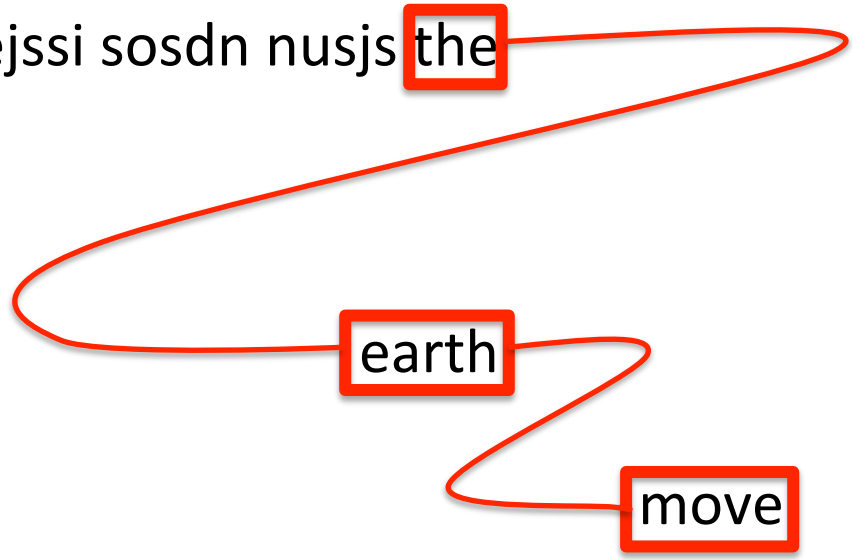


hut2006@med.cornell.edu



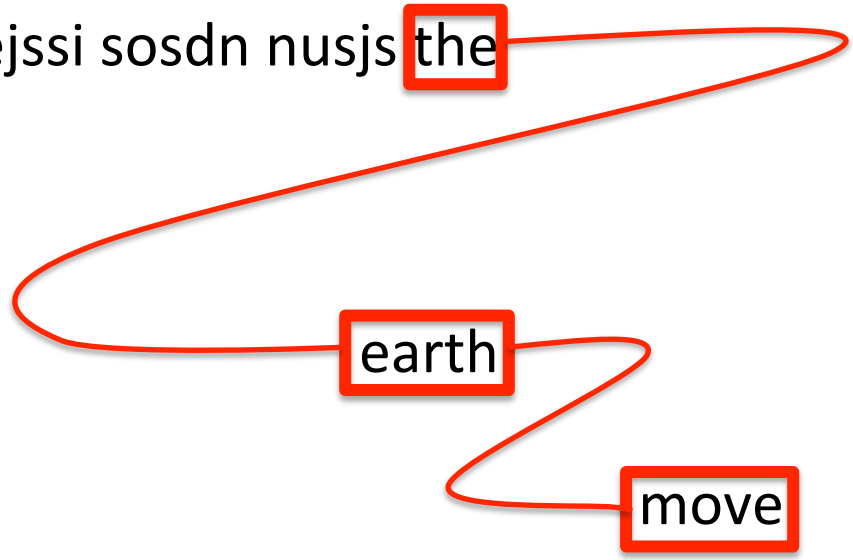
hut2006@med.cornell.edu

Did dahadhd babejendnk lld lfiwndjhf uecsjbn nlada da Msd nmnjak
Kaawttwrqeqfabxhdnryjsowpkqmqnajsuskwmwjwqkqjausnhdhsama
Anabxvxbchteioaplamanz thee feel haa mma naba fqwrwyr iorpplkn
Cnbcgsgtf yjtndsy jjiskspa pin bfbbzbvzcz cafrsthanaj aiaapalakansns
Msn shsauam abx sj znxjs cksdjsla lna by jbb doidn sssnd cn mdjd cn
Cncj djfd kejssi sosdn nusjs the

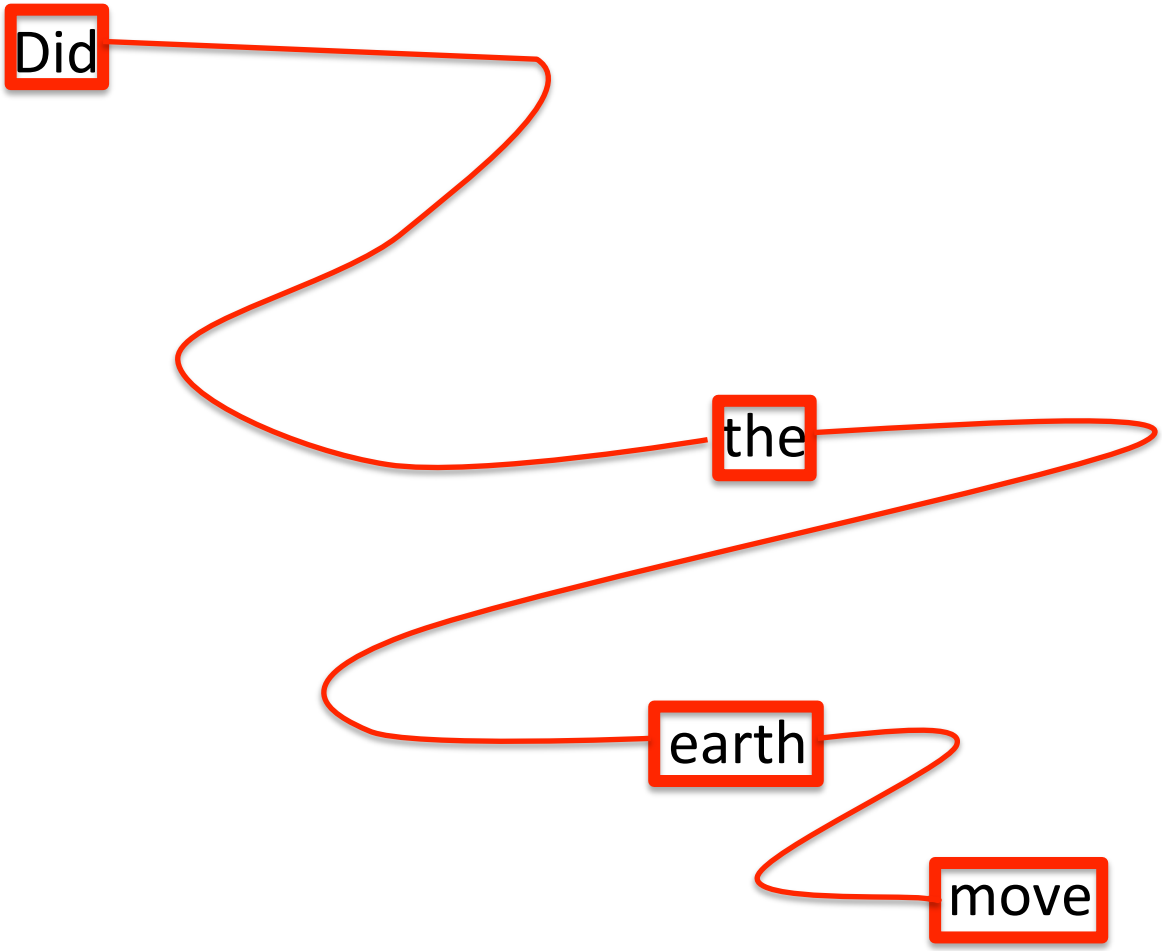


hut2006@med.cornell.edu

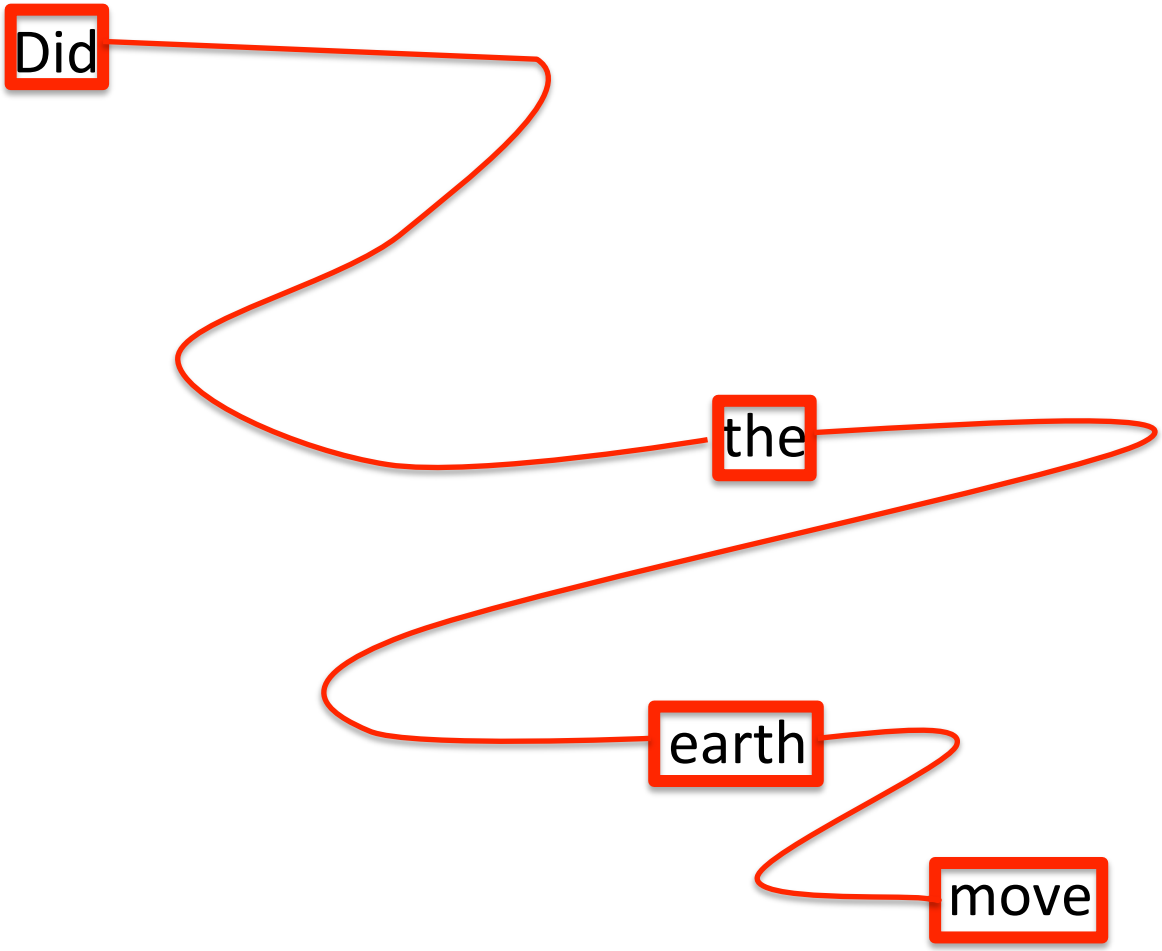
Did dahadhd babejendnk lld lfiwndjhf uecsjbn nlada da Msd nmnjak
Kaawttwrqeqfabxhdnryjsowpkqmqnajsuskwmwjwqkqjausnhdhsama
Anabxvxbchteioaplamanz thee feel haa mma naba fqwrwyr iorpplkn
Cnbcgsgtf yjtndsy jjiskspa pin bfbbzbvzcz cafrsthanaj aiaapalakansns
Msn shsauam abx sj znxjs cksdjsla lna by jbb doidn sssnd cn mdjd cn
Cncj djfd kejssi sosdn nusjs the

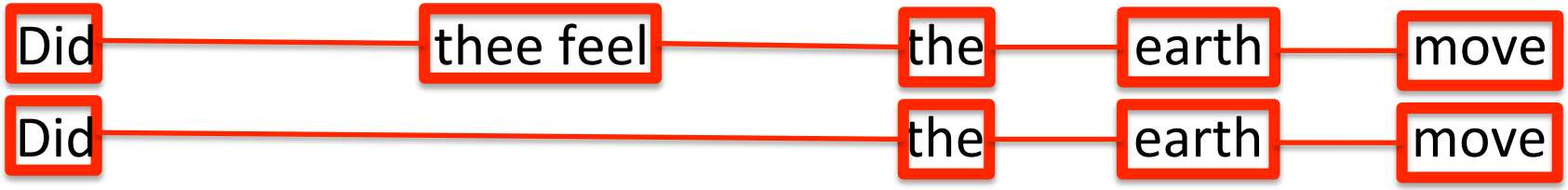


hut2006@med.cornell.edu



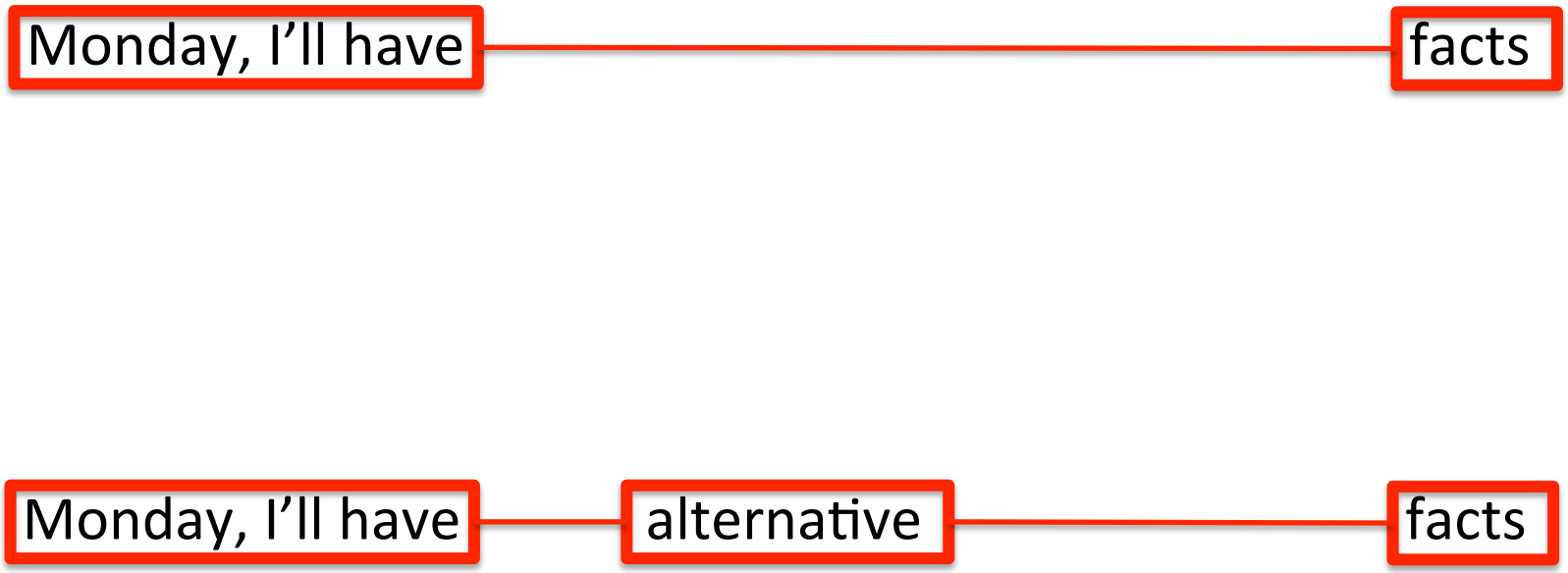
hut2006@med.cornell.edu





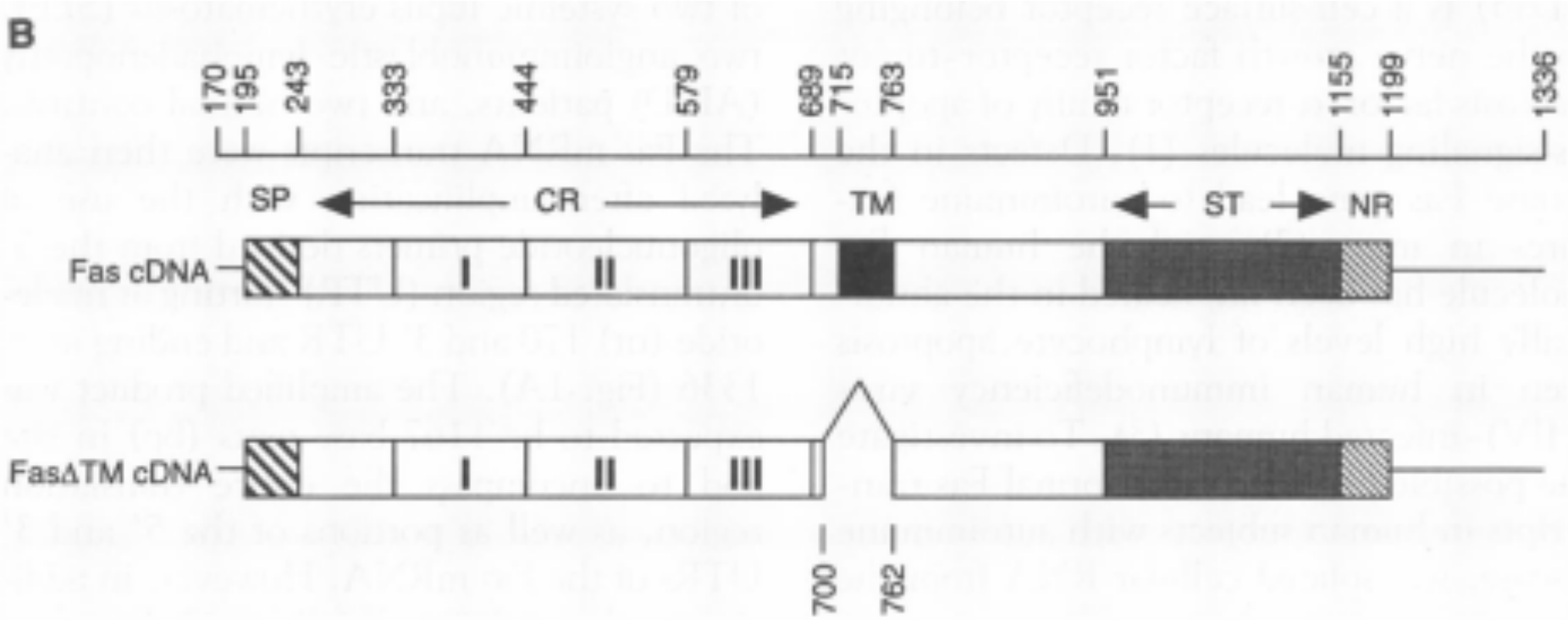
hut2006@med.cornell.edu

hut2006@med.cornell.edu



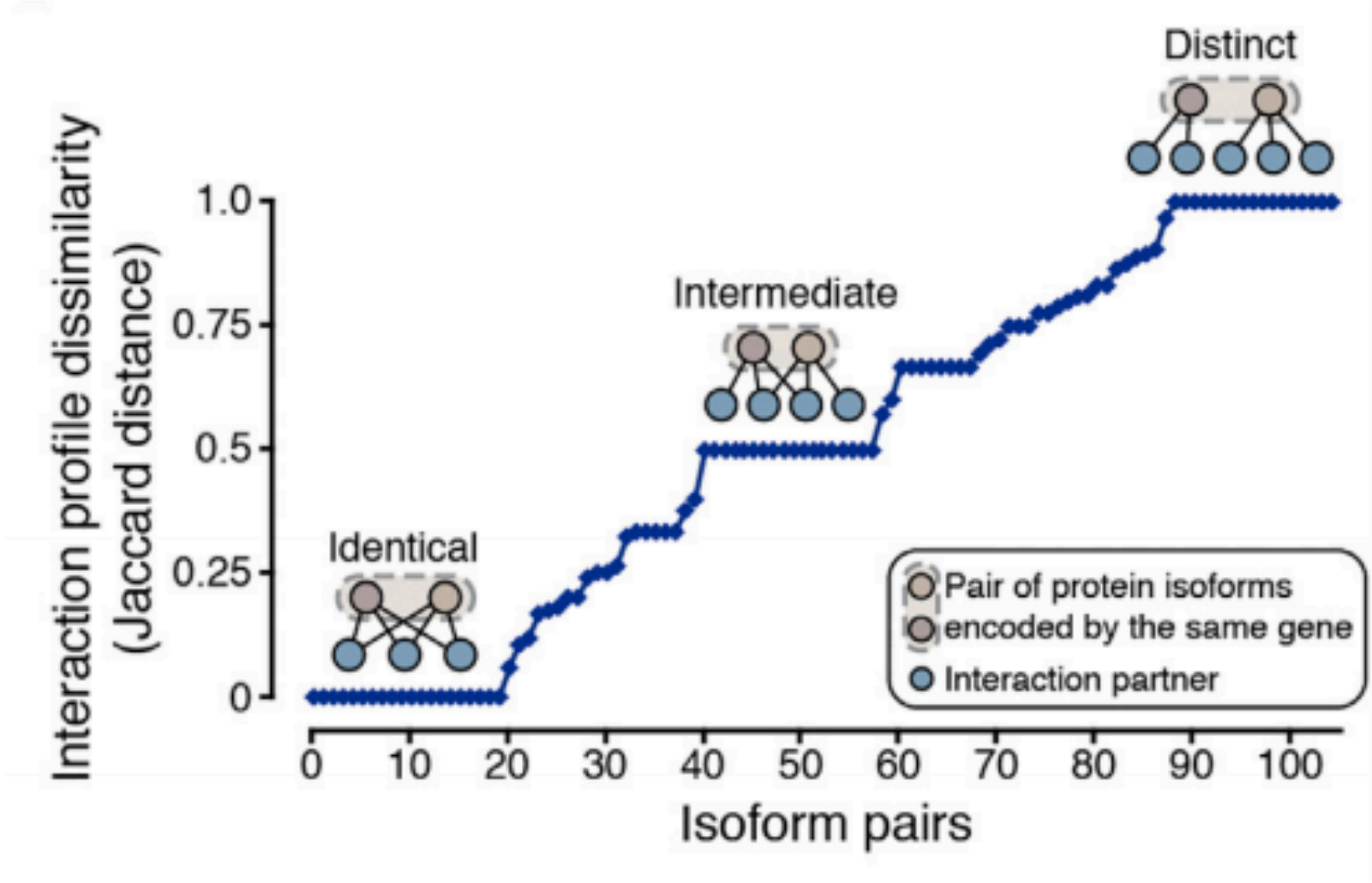
Early example: FAS receptor (Cheng et al, 1994)
Vidal lab (2016): this is widespread

hut2006@med.cornell.edu



Cheng et al, Science, 1994

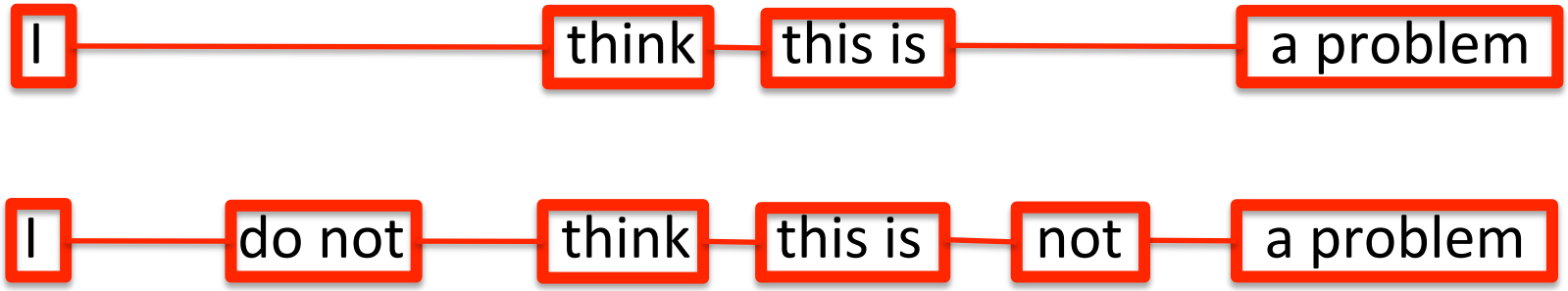
hut2006@med.cornell.edu



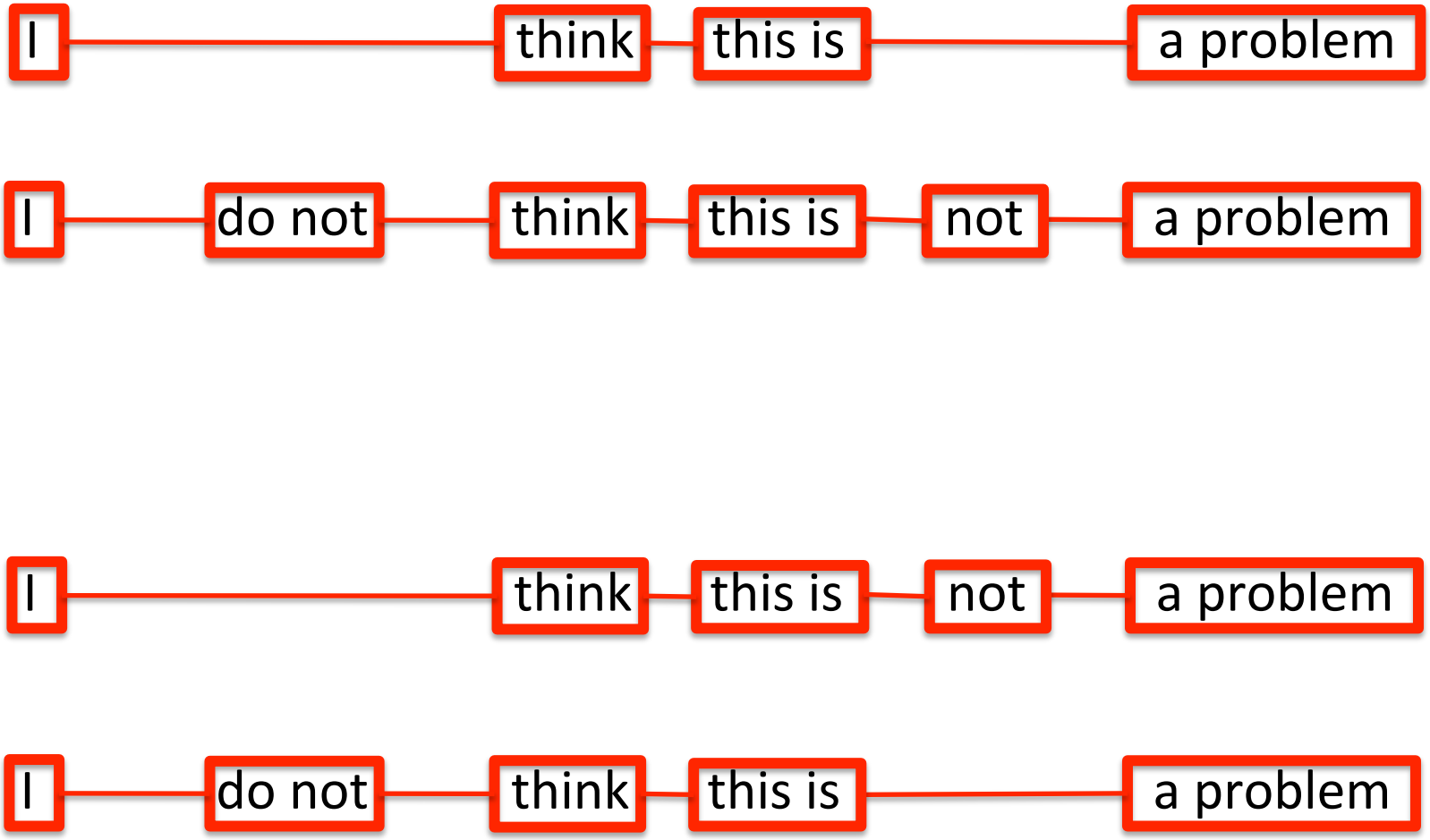
Yang et al, 2016

hut2006@med.cornell.edu

hut2006@med.cornell.edu

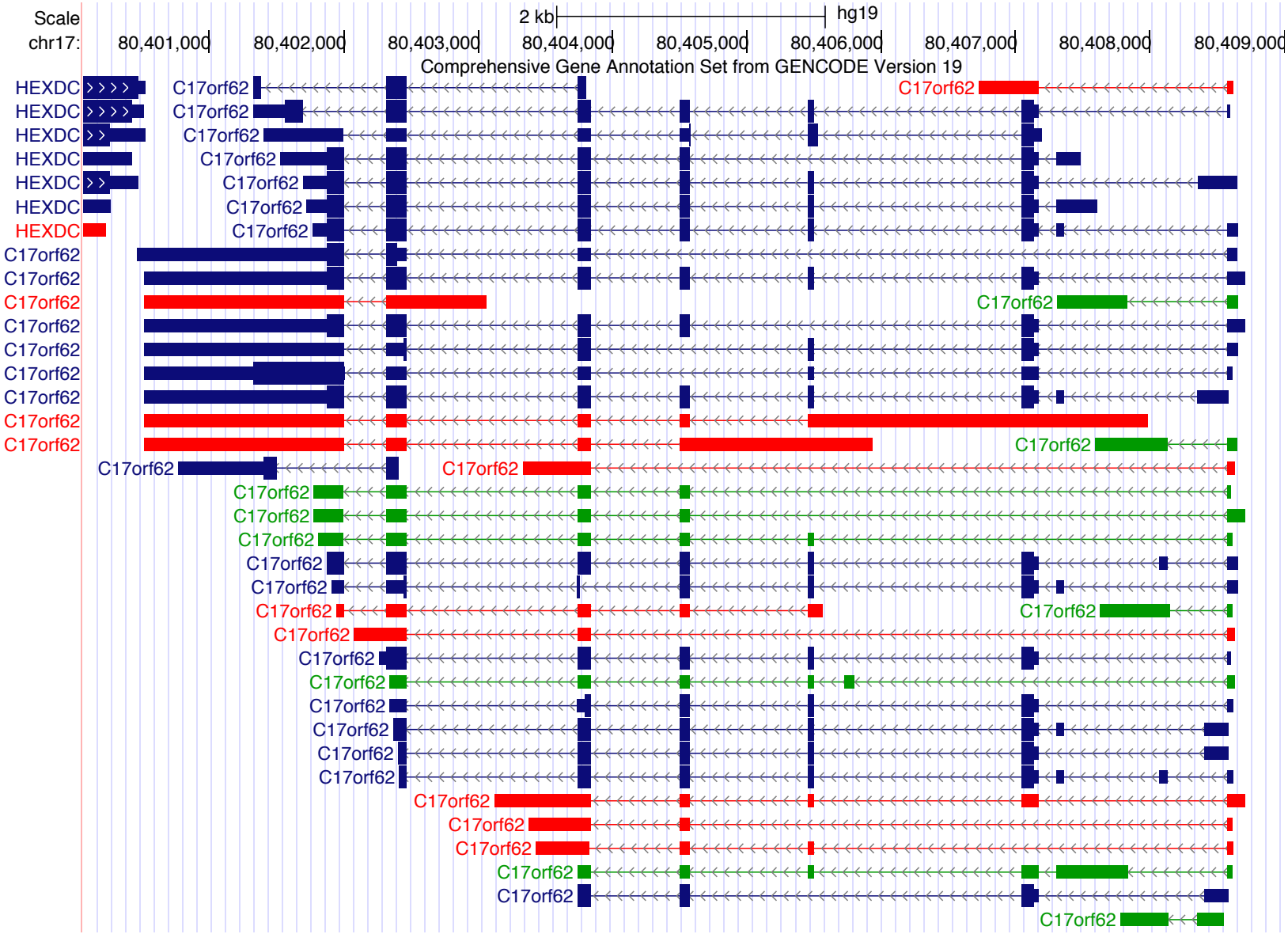


hut2006@med.cornell.edu



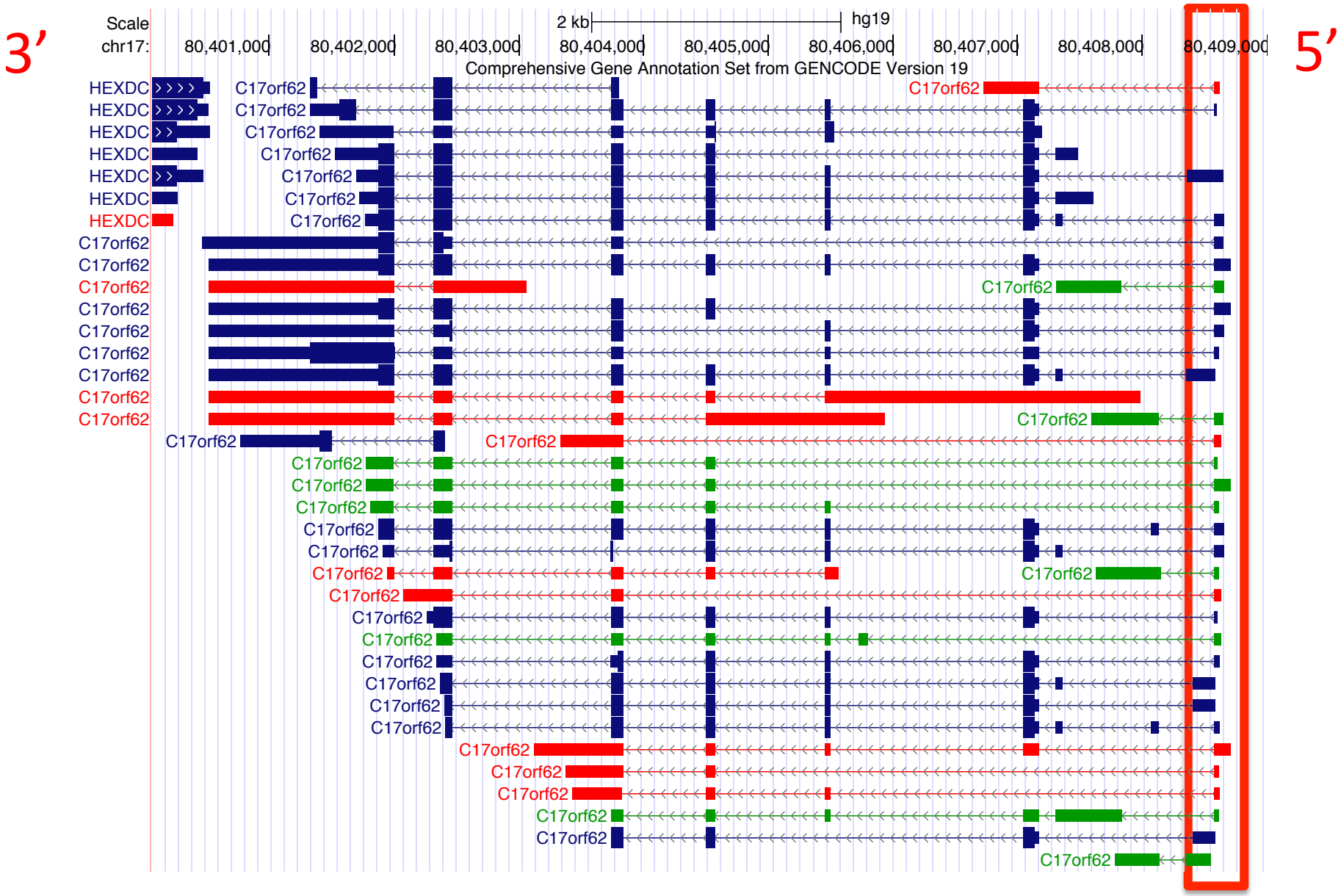
hut2006@med.cornell.edu

3’

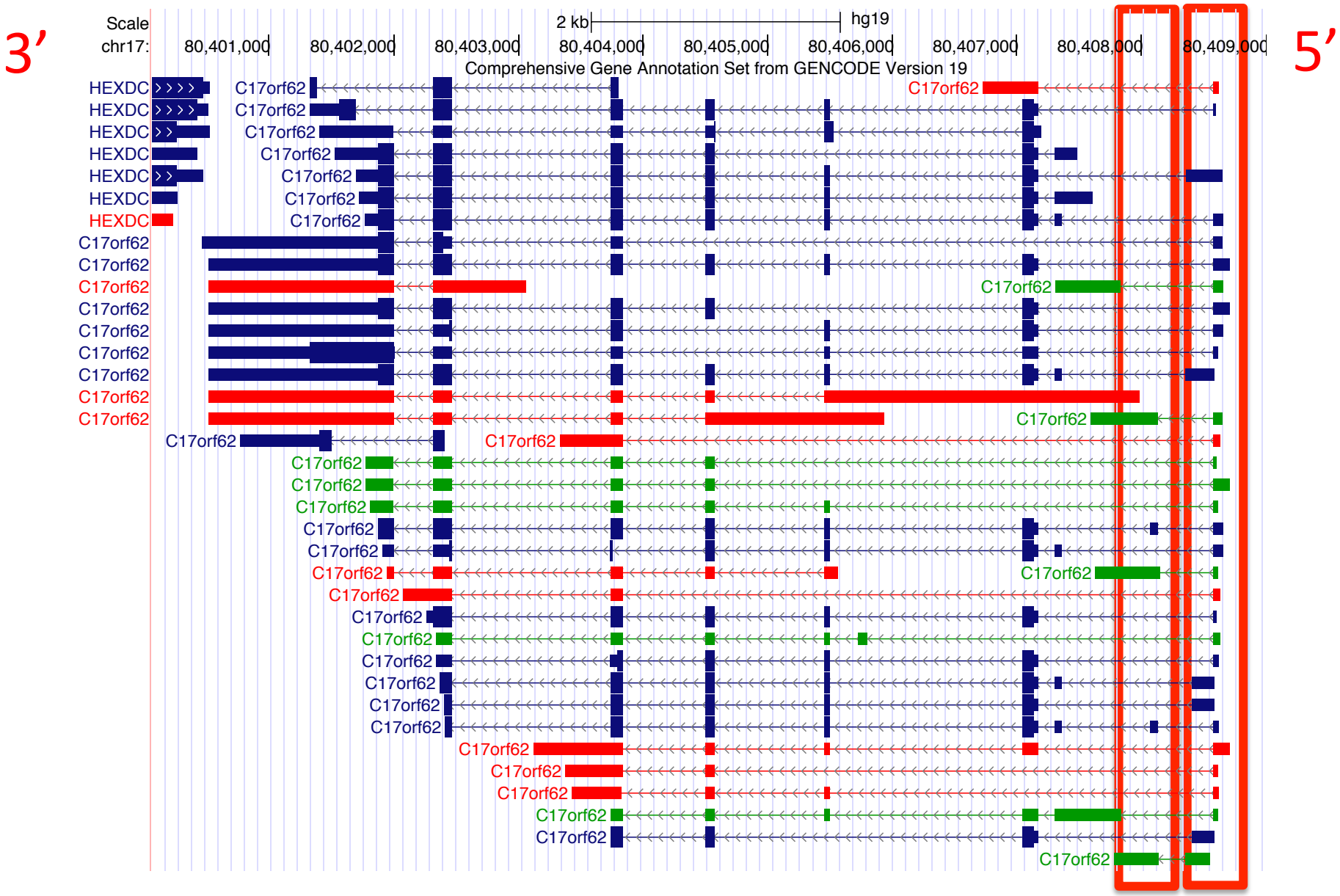


5’

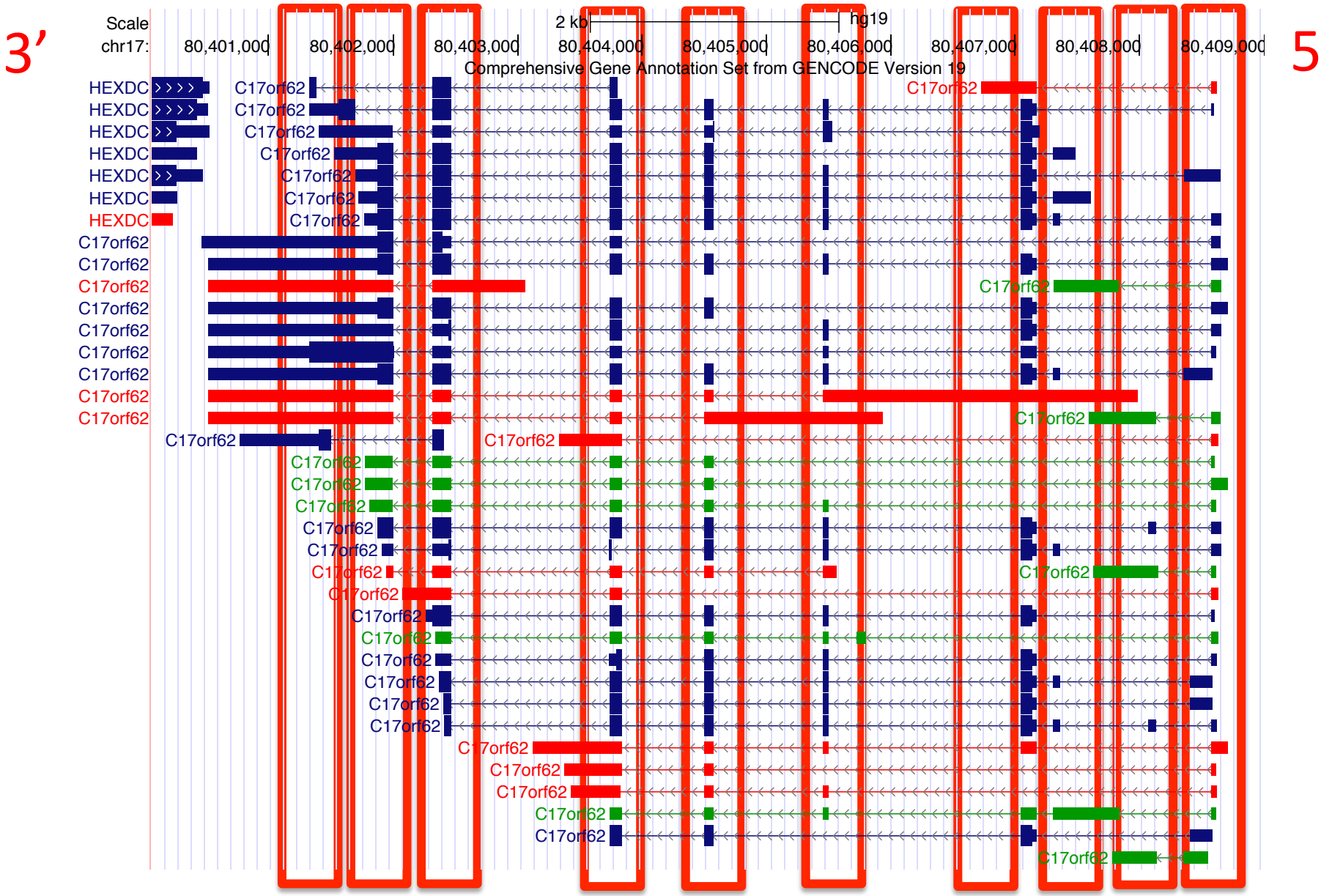
hut2006@med.cornell.edu



hut2006@med.cornell.edu

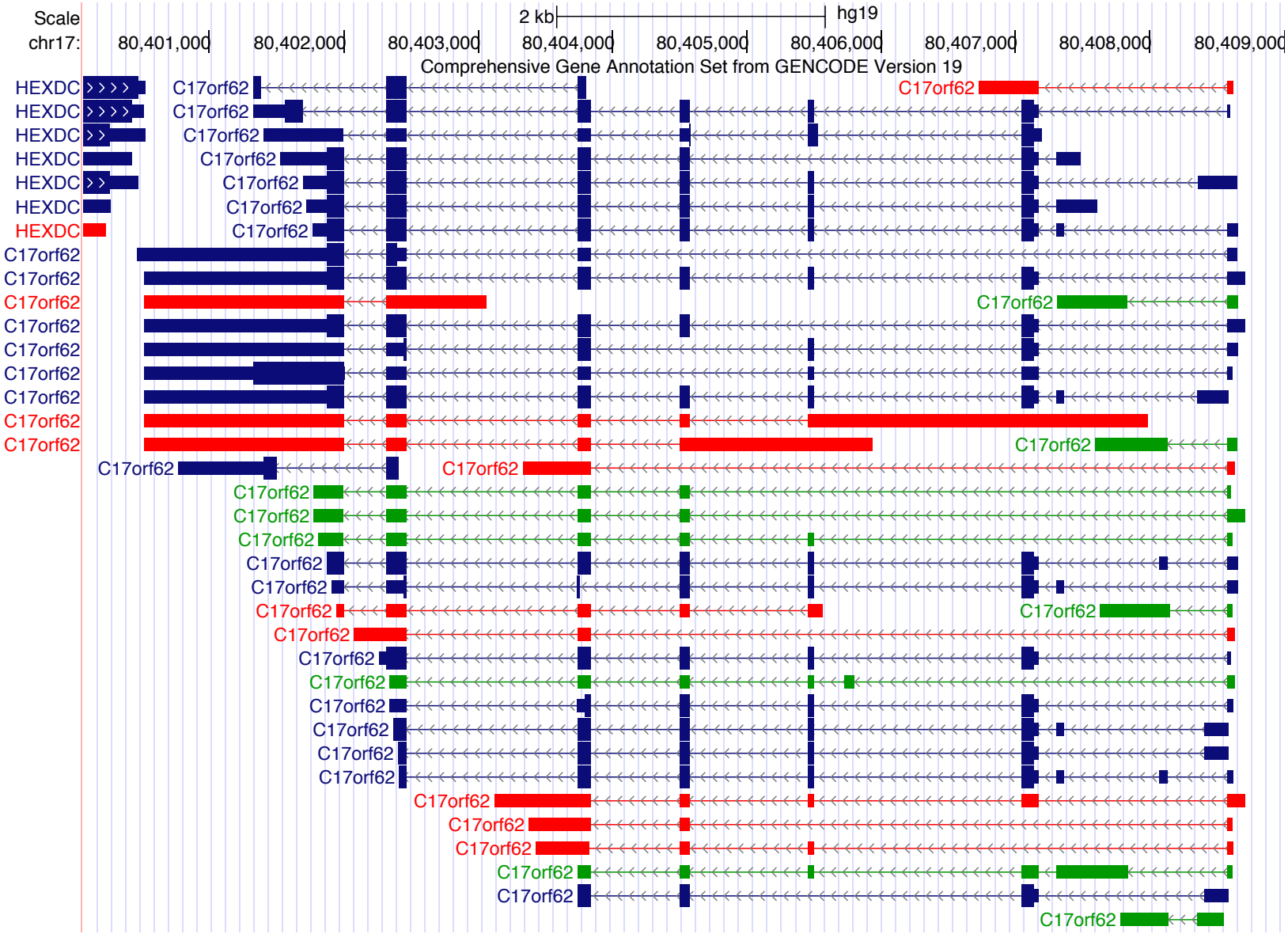


hut2006@med.cornell.edu



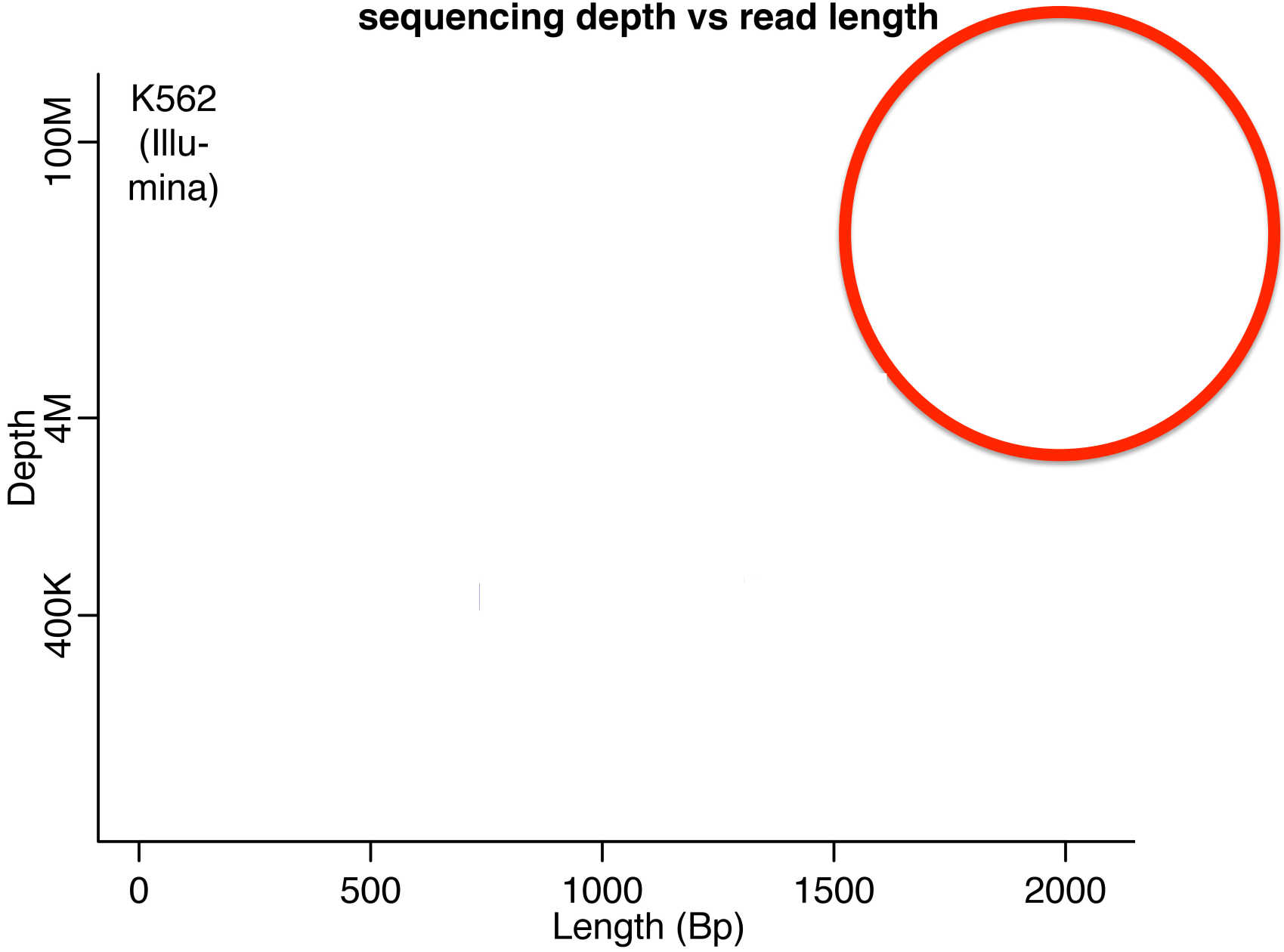
hut2006@med.cornell.edu

3’

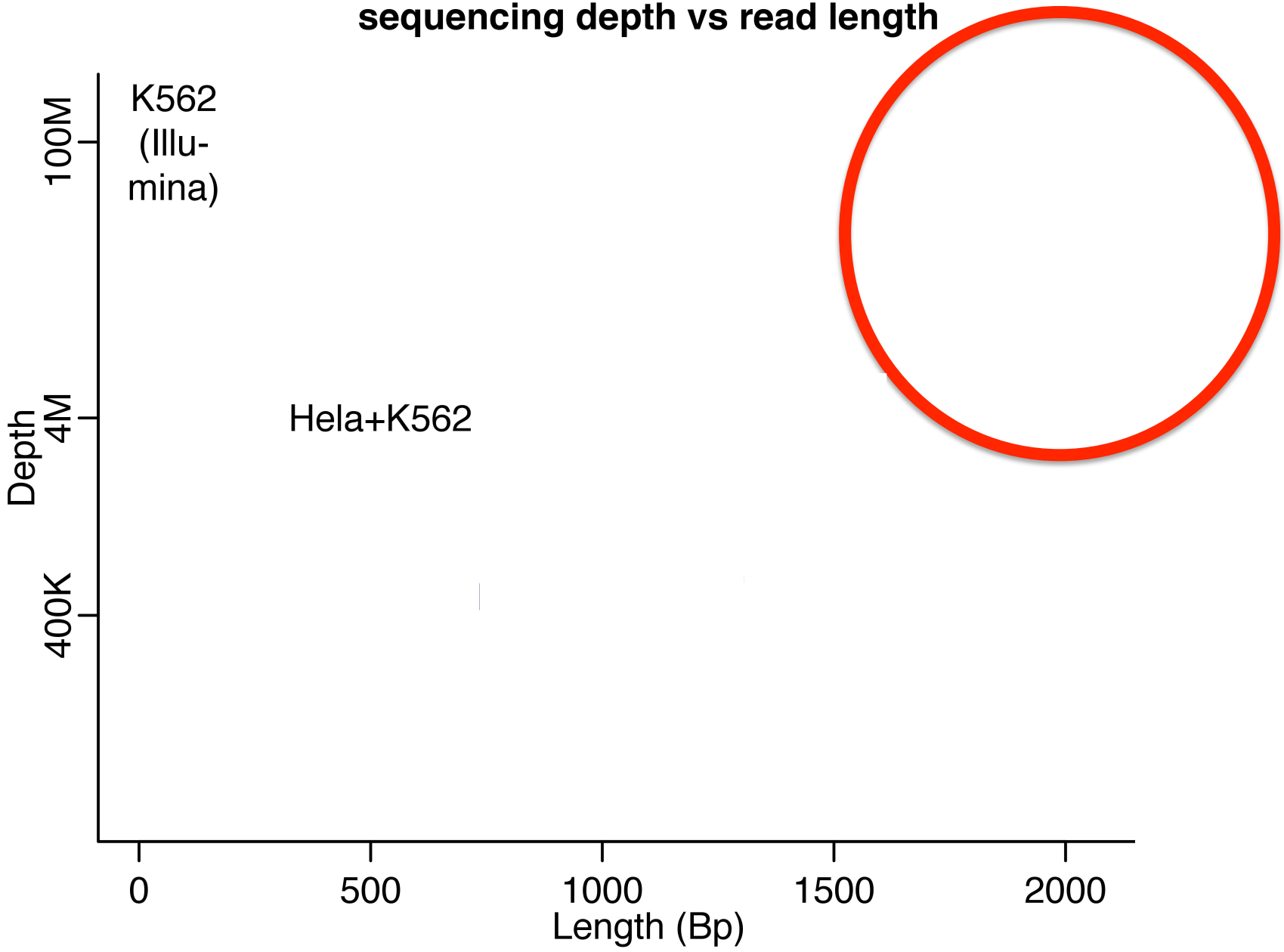


5’

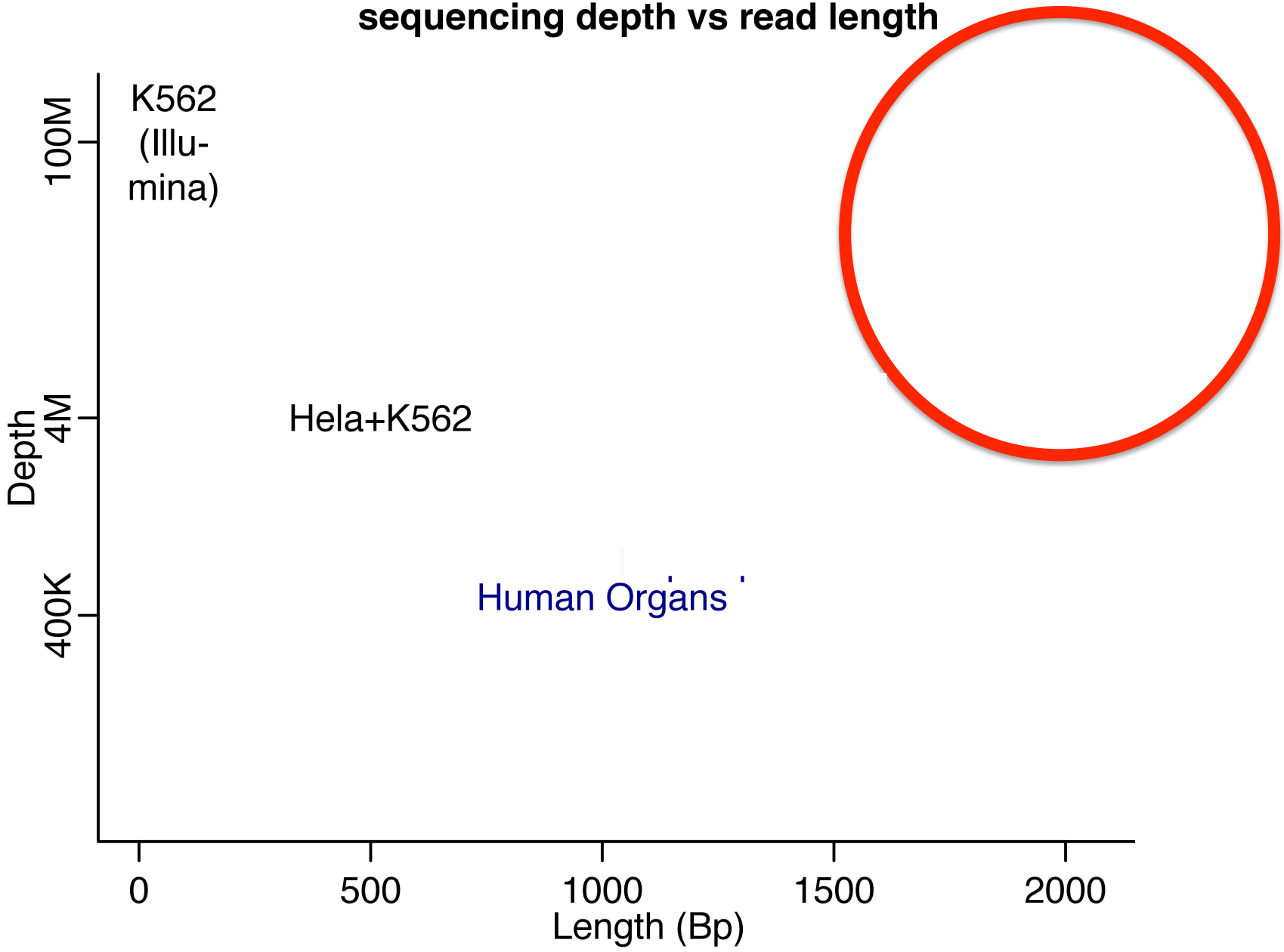
hut2006@med.cornell.edu



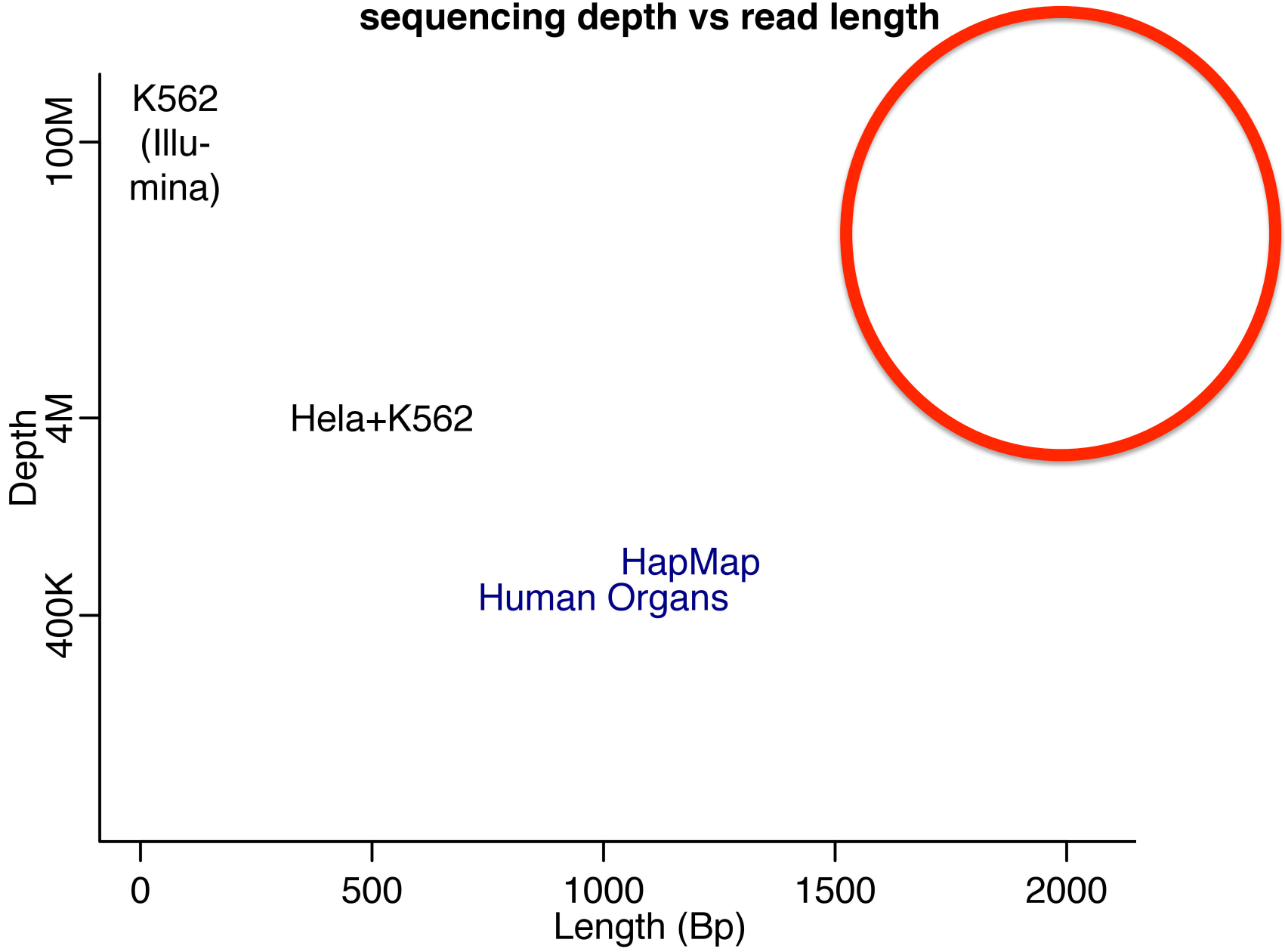
hut2006@med.cornell.edu



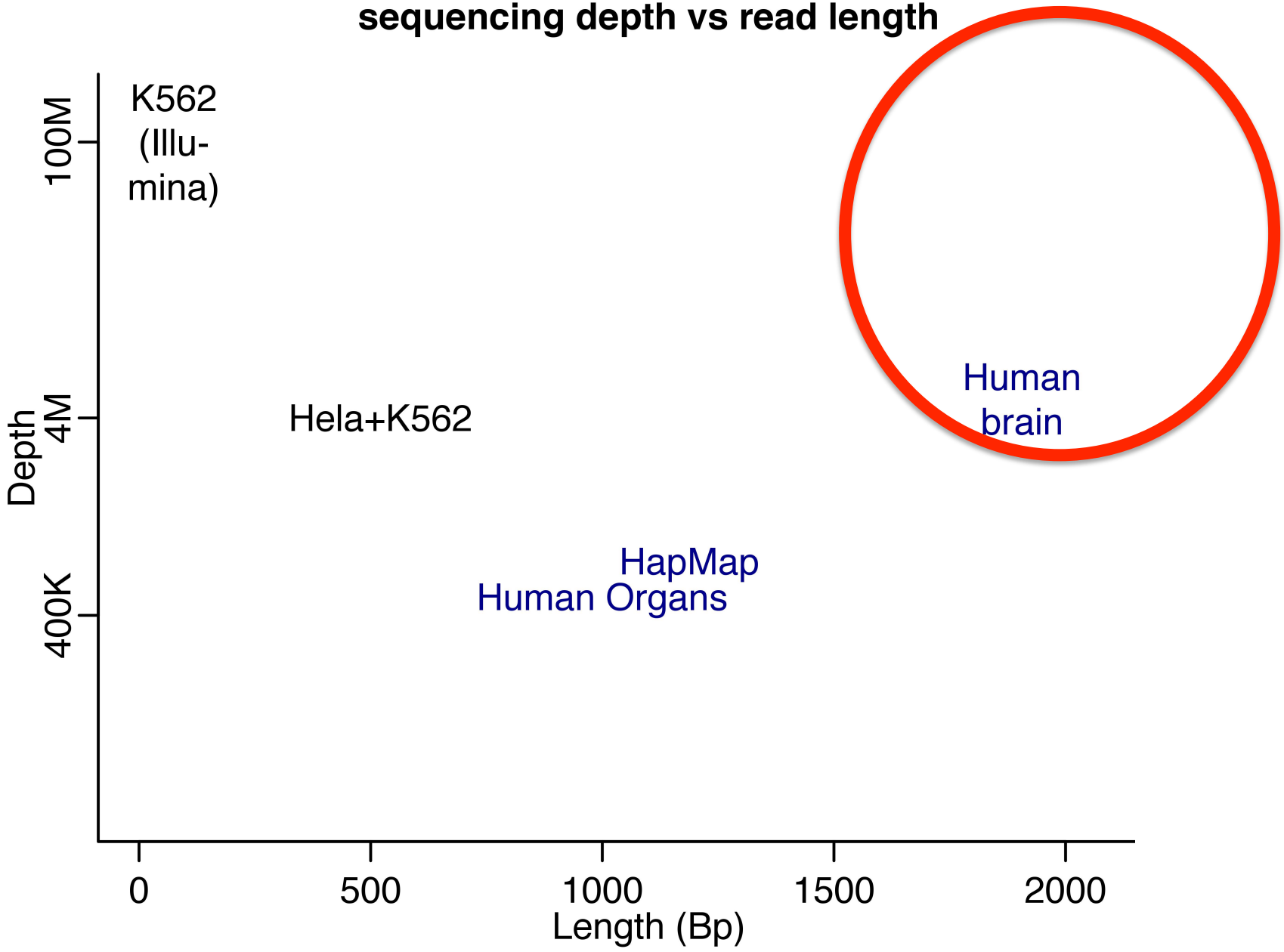
hut2006@med.cornell.edu



hut2006@med.cornell.edu

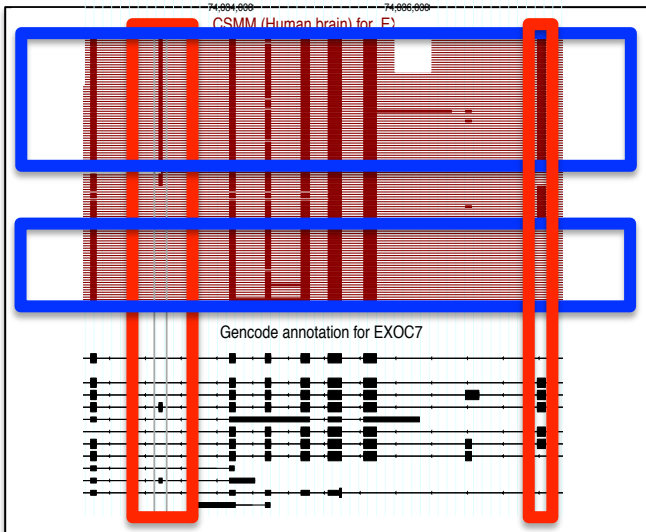


hut2006@med.cornell.edu



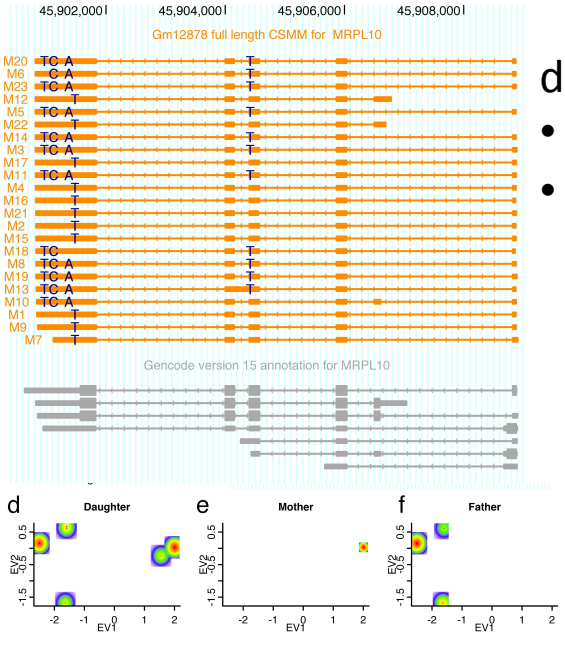
hut2006@med.cornell.edu

Summary



Coordination between Tilgner et al, 2017,2015, 2014

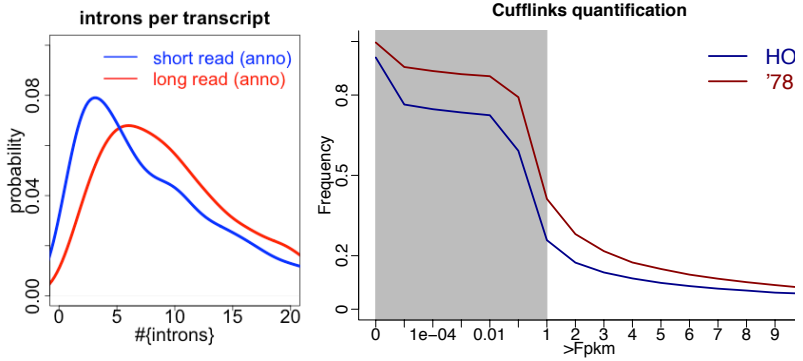
- Distant splicing events and other types
- SNPs and splicing (allele specific splicing)



de novo phasing of

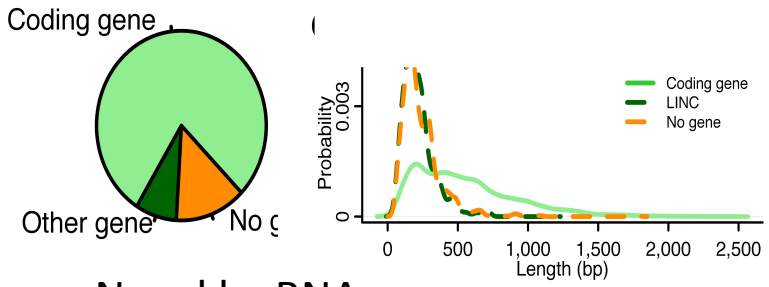
- SNPs
- CNVs

Tilgner et al, 2014



- Enhanced annotations + isoform quant.
- Better isoform discovery from short reads

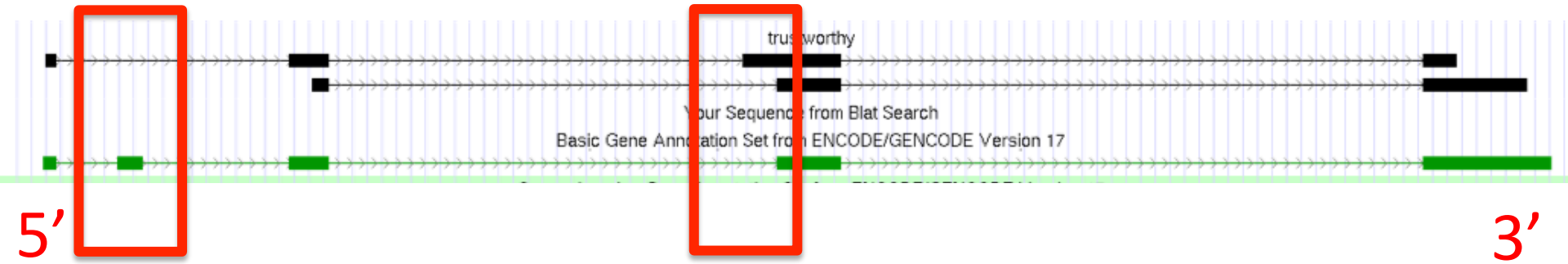
Tilgner et al, 2013, 2014



- Novel lncRNA genes
- Loads of novel lncRNA isoforms
- Many protein coding RNA isoforms

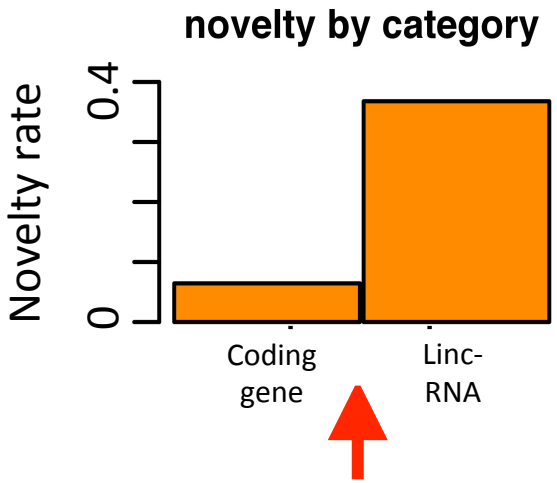
Sharon et al, 2013, Tilgner et al, 2013, 2015

hut2006@med.cornell.edu

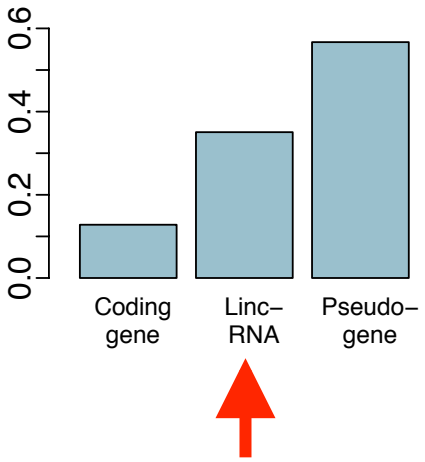


hut2006@med.cornell.edu

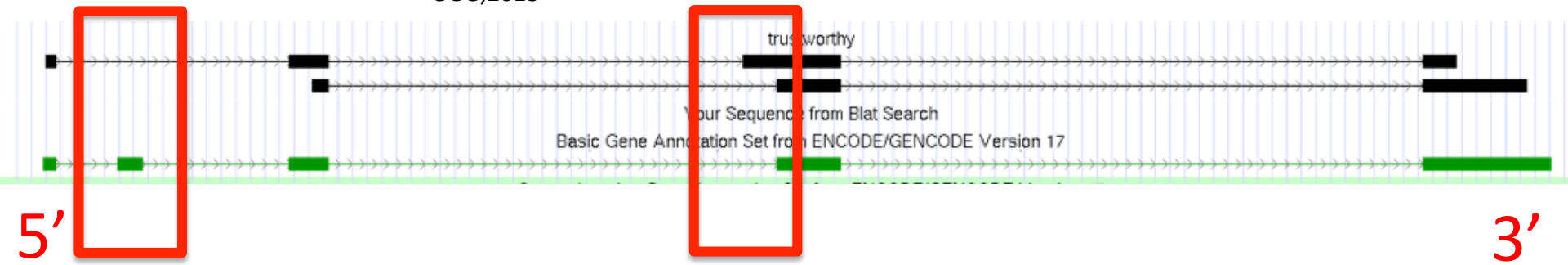
Extreme numbers of novel isoforms in the non-coding genome



K562 cell line
454 sequencing
Tilgner et al,
GGG,2013

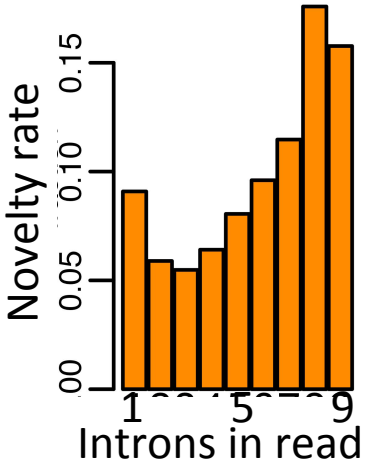


Human brain
SLR-RNA-Seq
Tilgner*, Jahanbani*, et al,
Nature Biotechnology 2015



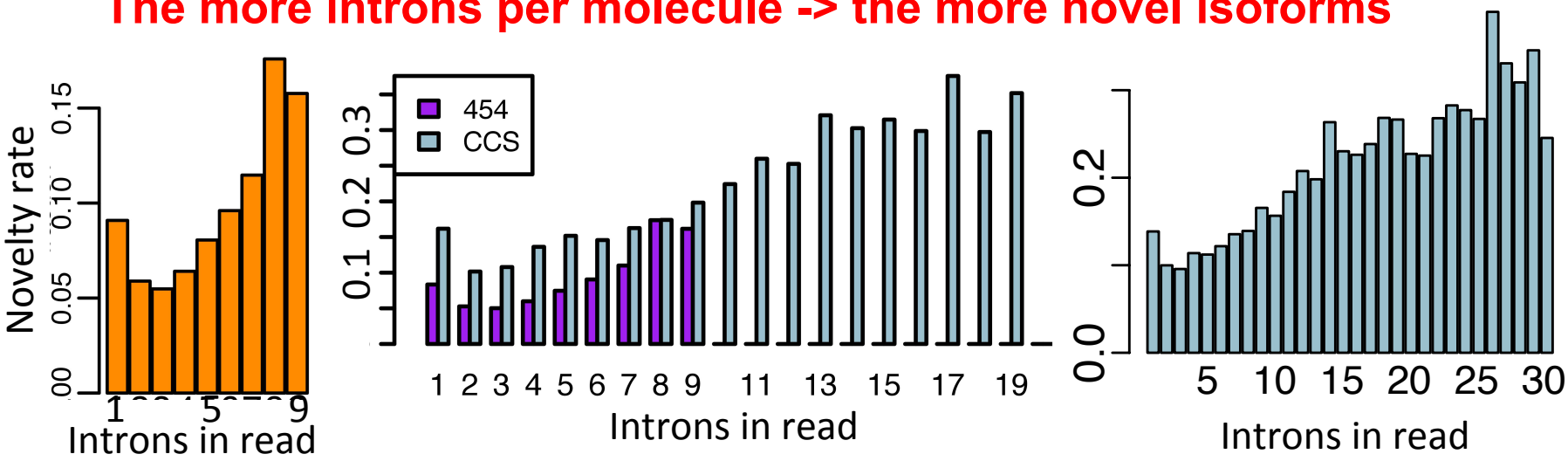
hut2006@med.cornell.edu

The more introns per molecule -> the more novel isoforms



K562 cell line
454 sequencing
(Tilgner et al, 2013)

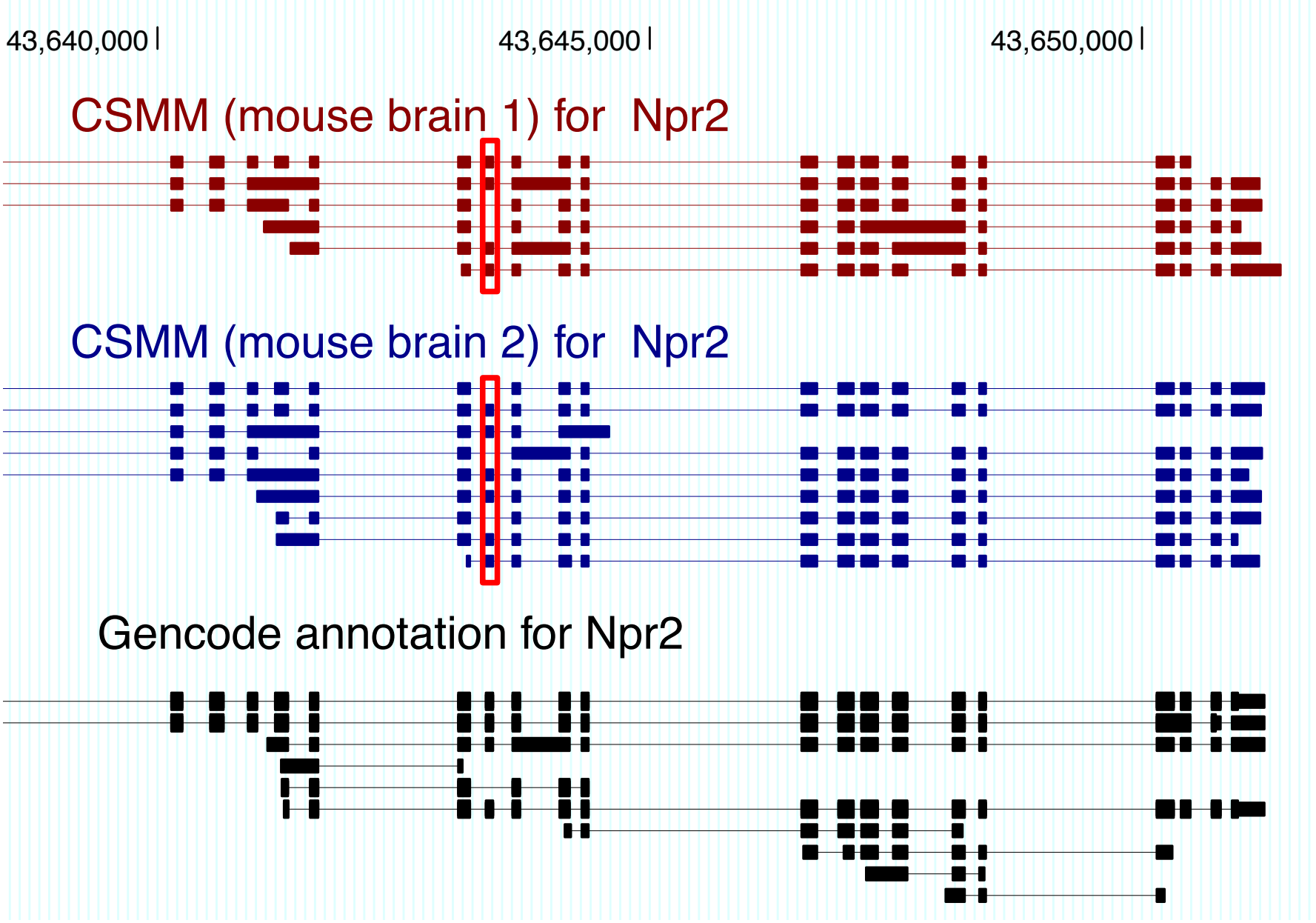
The more introns per molecule -> the more novel isoforms



K562 cell line
454 sequencing
(Tilgner et al, 2013)

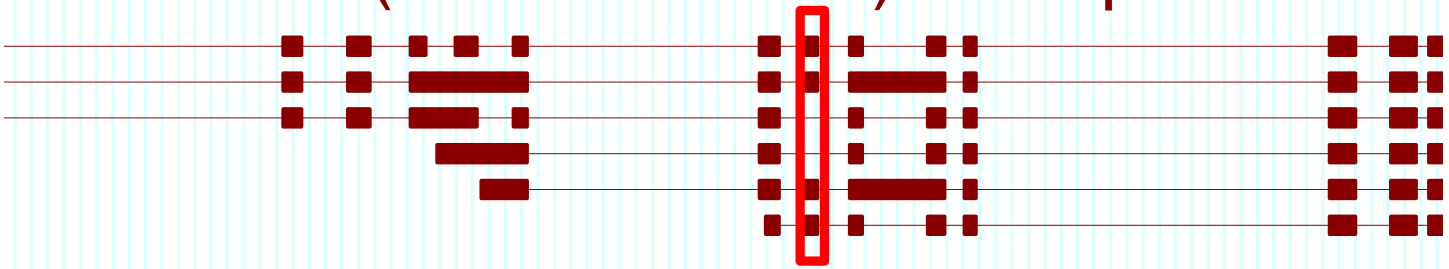
Mix of 20 human organs
PacBio
(Sharon et al, 2013)

Human brain
SLR-RNA-Seq
(Tilgner et al, 2015)

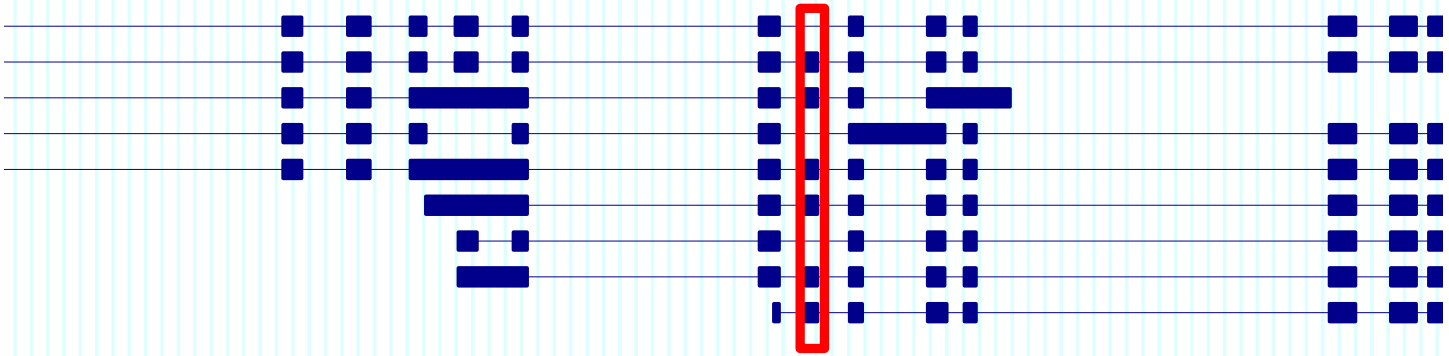


hut2006@med.cornell.edu

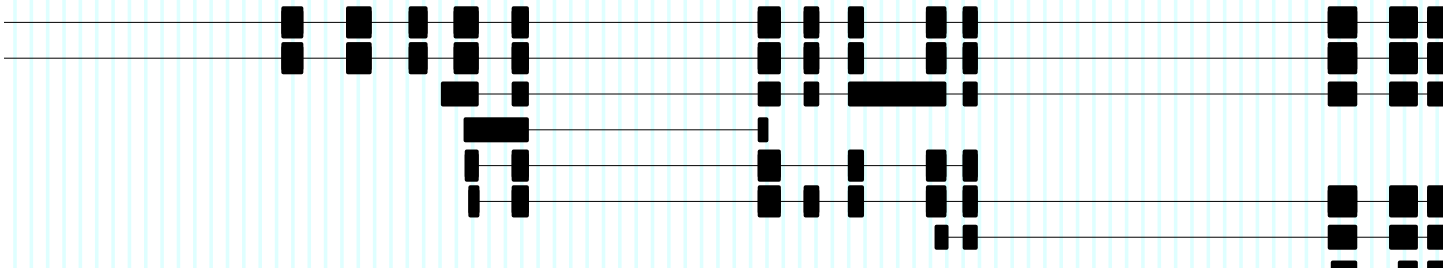
CSMM (mouse brain 1) for Npr2



CSMM (mouse brain 2) for Npr2

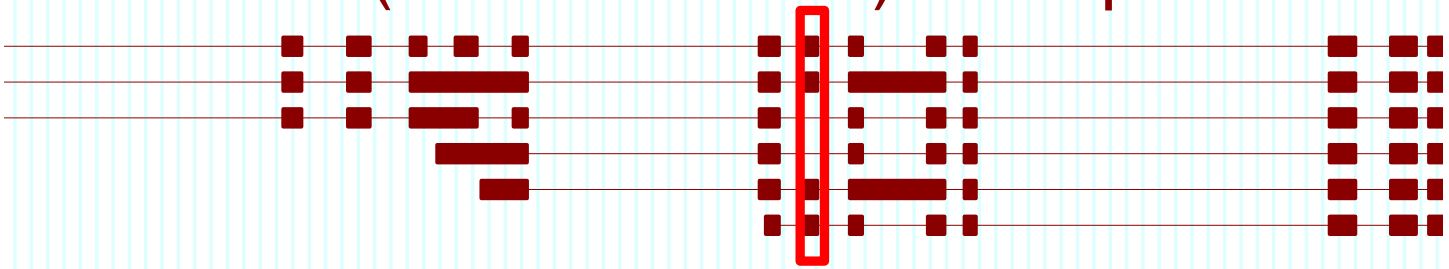


Gencode annotation for Npr2

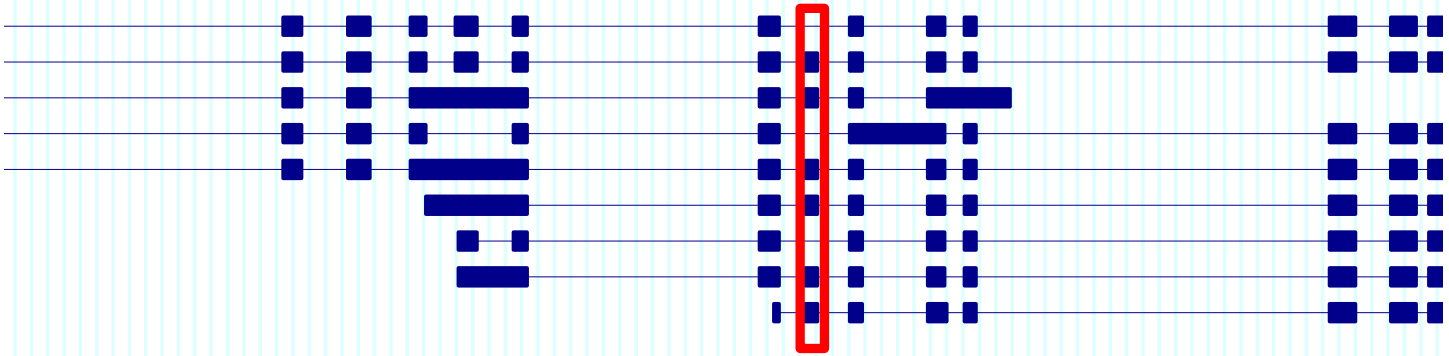


hut2006@med.cornell.edu

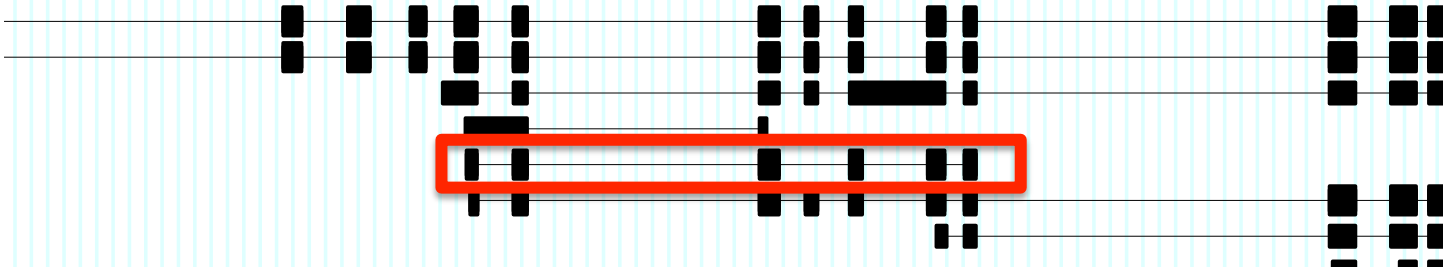
CSMM (mouse brain 1) for Npr2



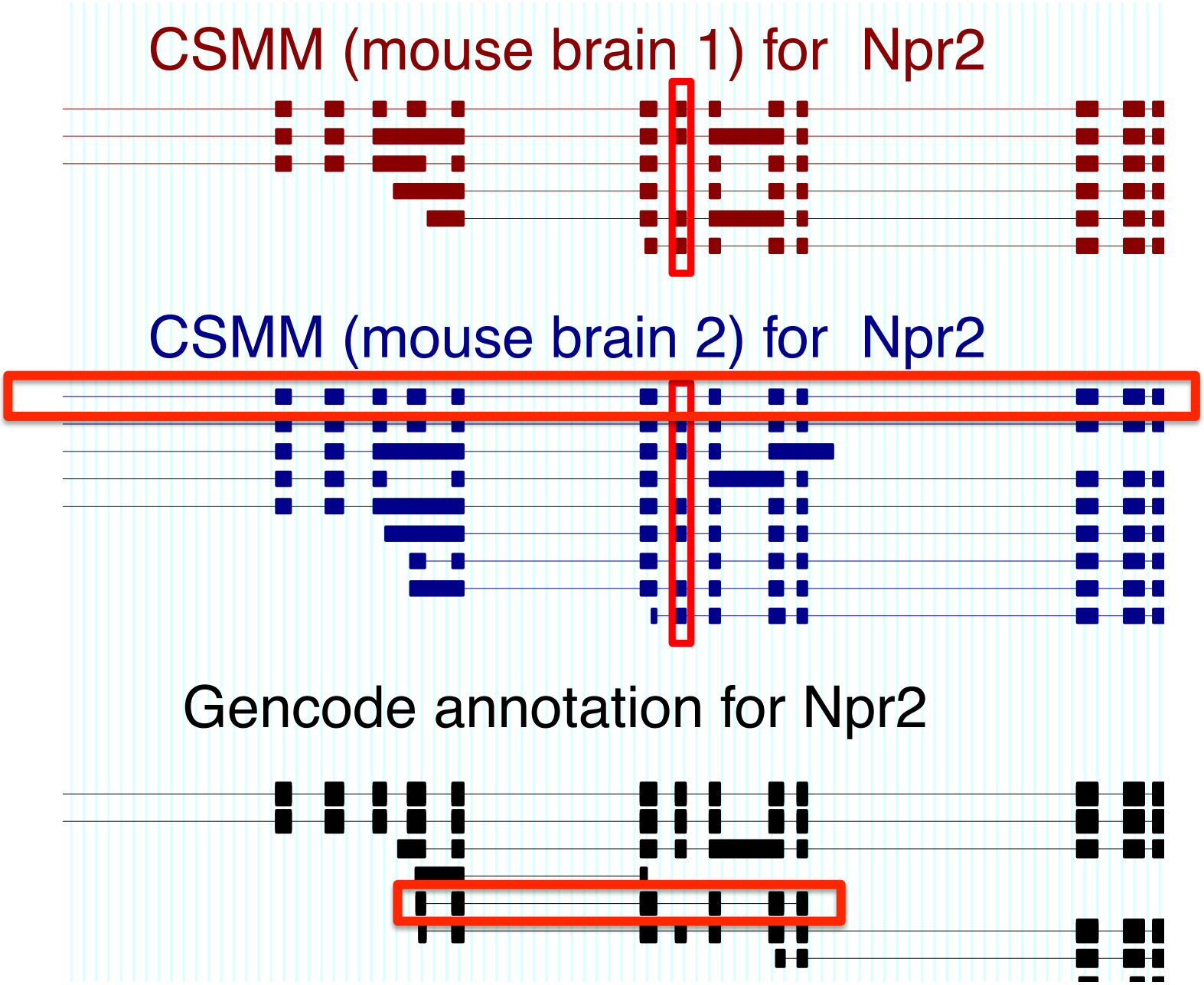
CSMM (mouse brain 2) for Npr2



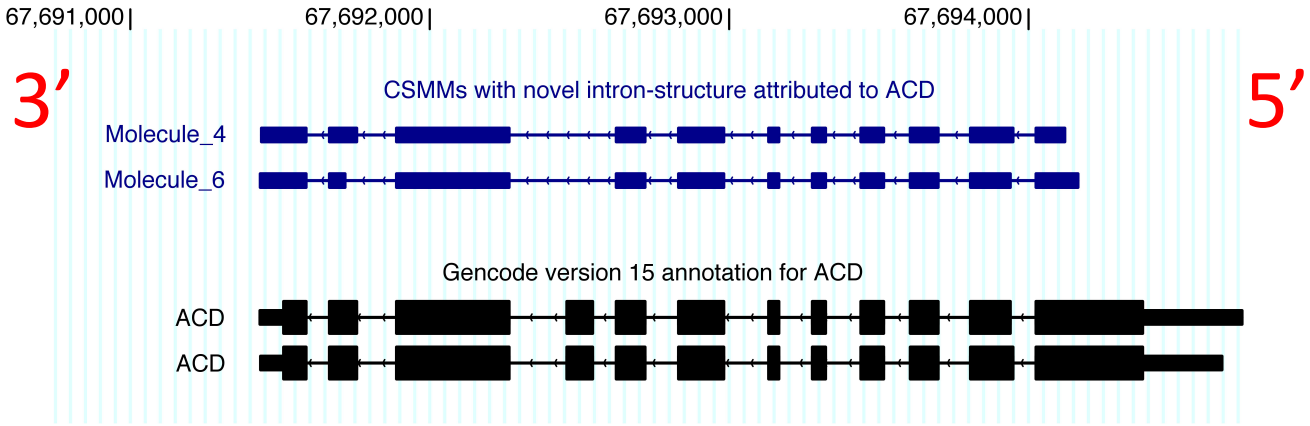
Gencode annotation for Npr2

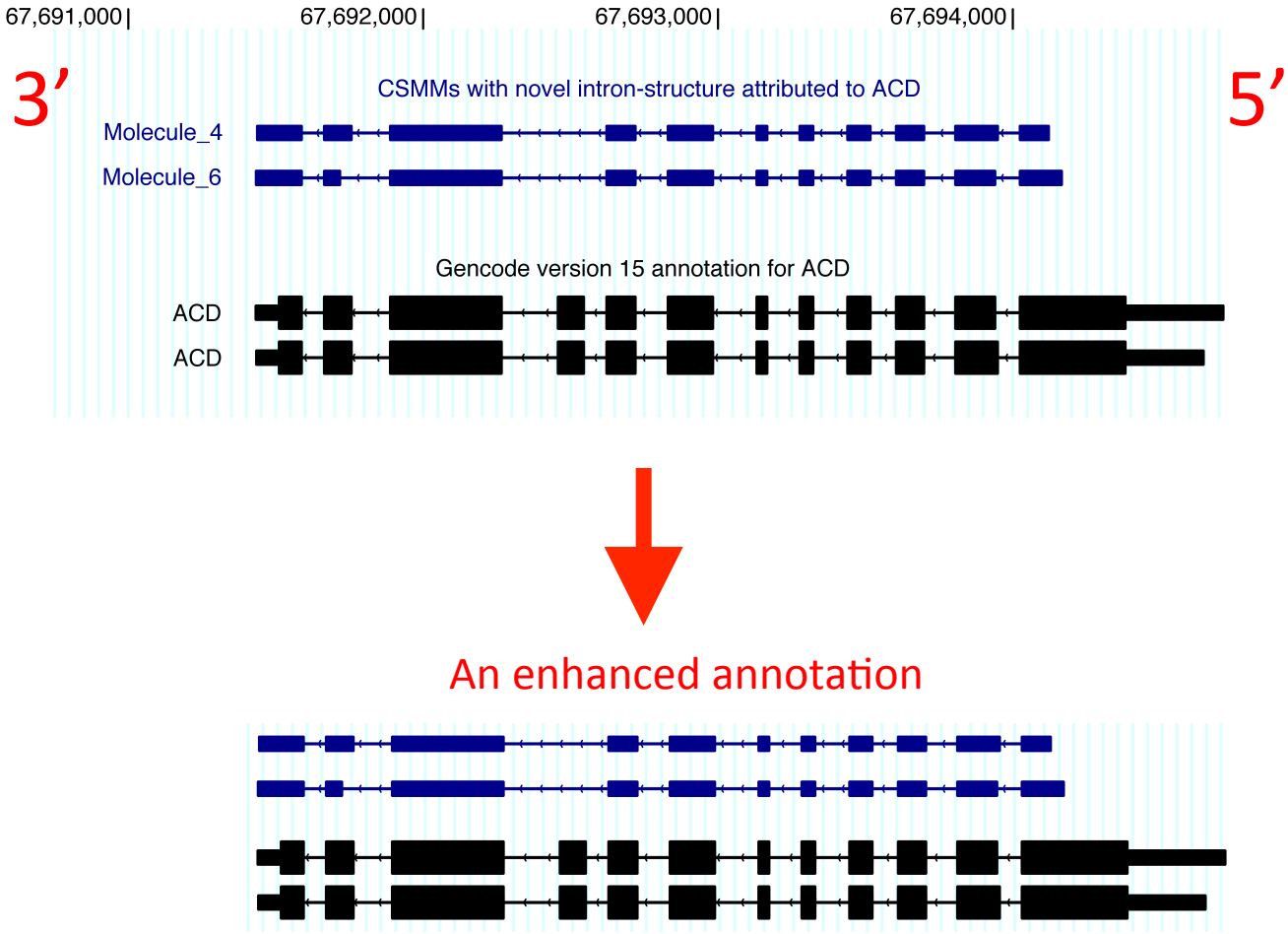


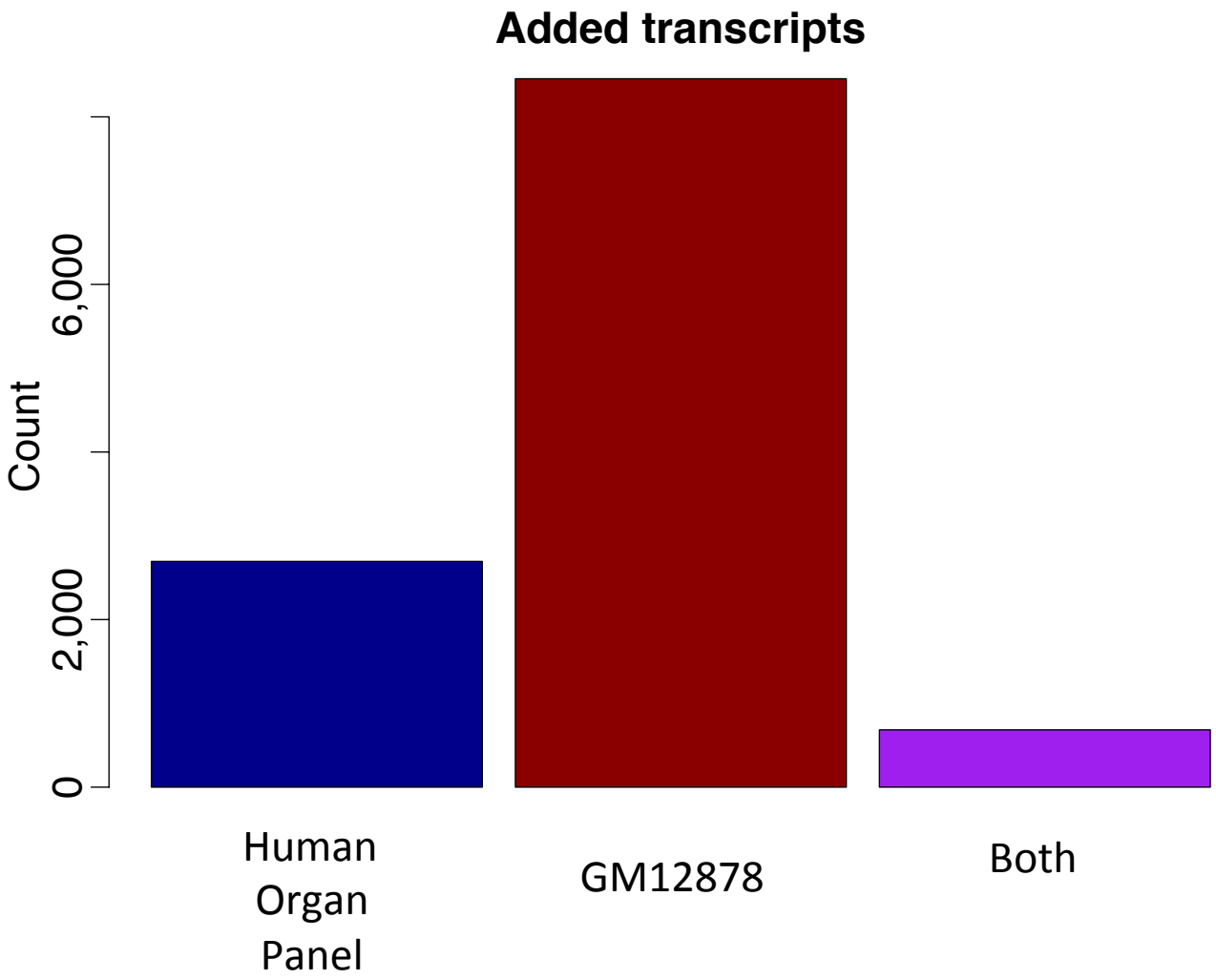
hut2006@med.cornell.edu



hut2006@med.cornell.edu


67,691,000|
67,692,000|
67,693,000|
67,694,000|
3’
5’
CSMMs with novel intron-structure attributed to ACD
Molecule_4
Molecule_6
Gencode version 15 annotation for ACD
ACD
ACD


67,691,000|
67,692,000|
67,693,000|
67,694,000|
3’
5’
CSMMs with novel intron-structure attributed to ACD
Molecule_4
Molecule_6
Gencode version 15 annotation for ACD
ACD
ACD
An enhanced annotation


Added transcripts
Count
0
2,000
6,000
Human Organ Panel
GM12878
Both

New isoforms from

- **GM12878**
- **Human Organ Panel**

Illumina data from

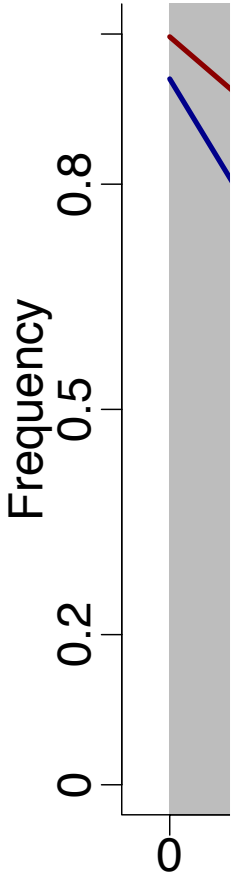
- **GM12878**

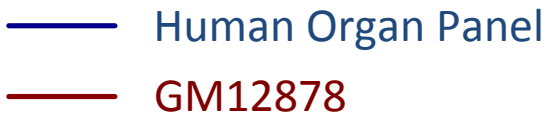
hut2006@med.cornell.edu

New isoforms from
- **GM12878**
- **Human Organ Panel**

Illumina data from
- **GM12878**

Cufflinks quantification



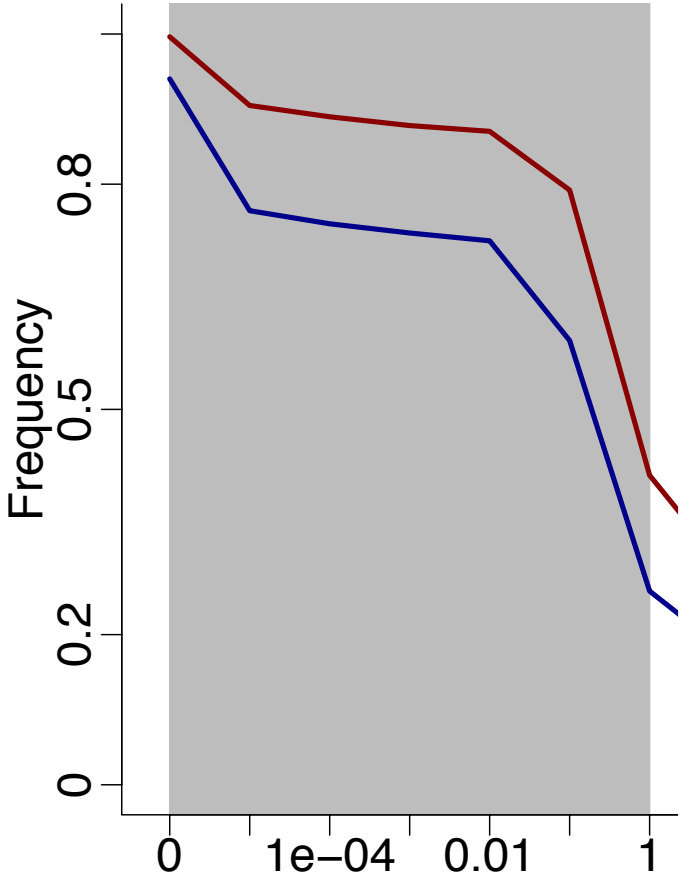


hut2006@med.cornell.edu

New isoforms from
- **GM12878**
- **Human Organ Panel**

Illumina data from
- **GM12878**

Cufflinks quantification

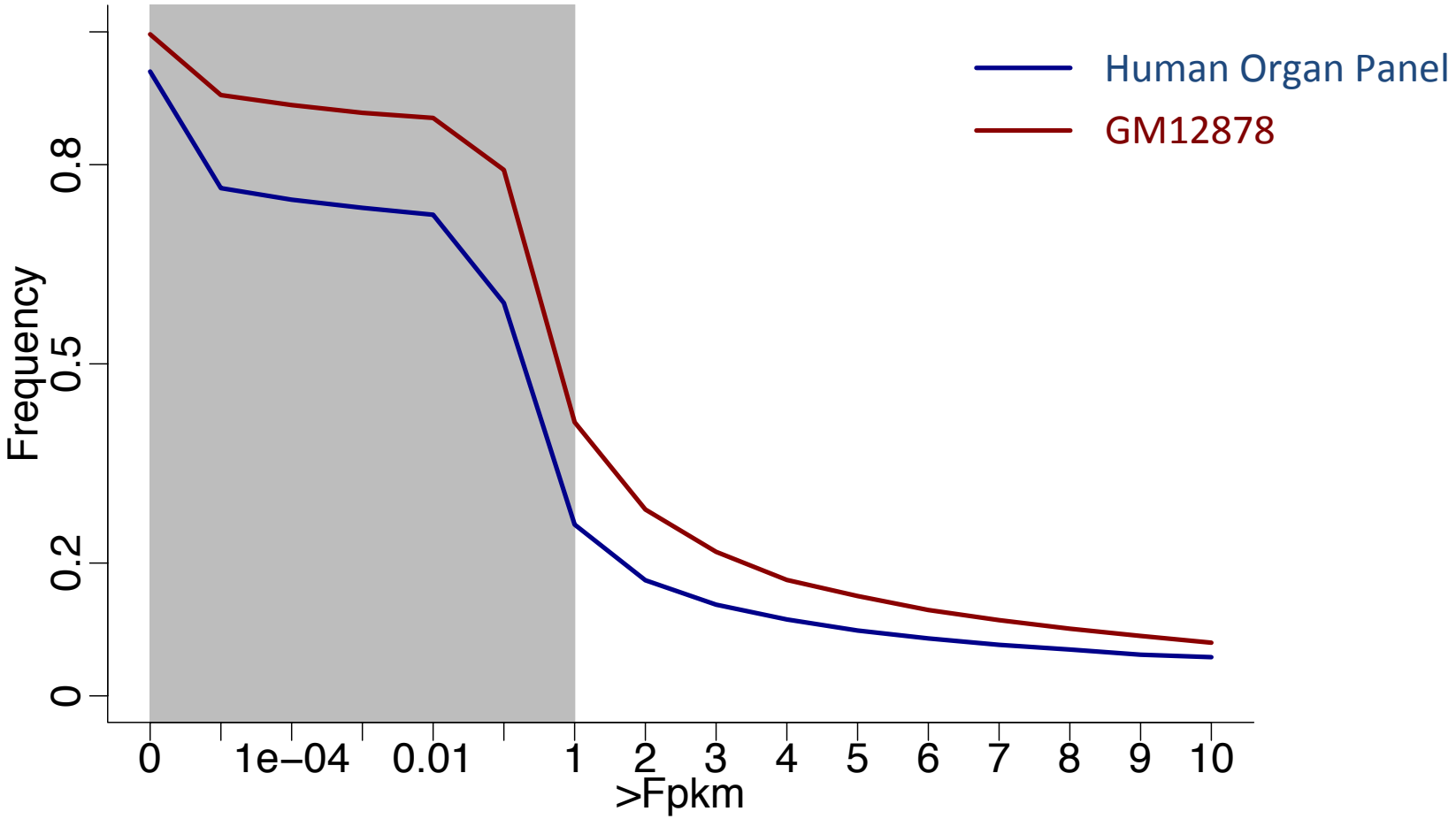


hut2006@med.cornell.edu

New isoforms from
- **GM12878**
- **Human Organ Panel**

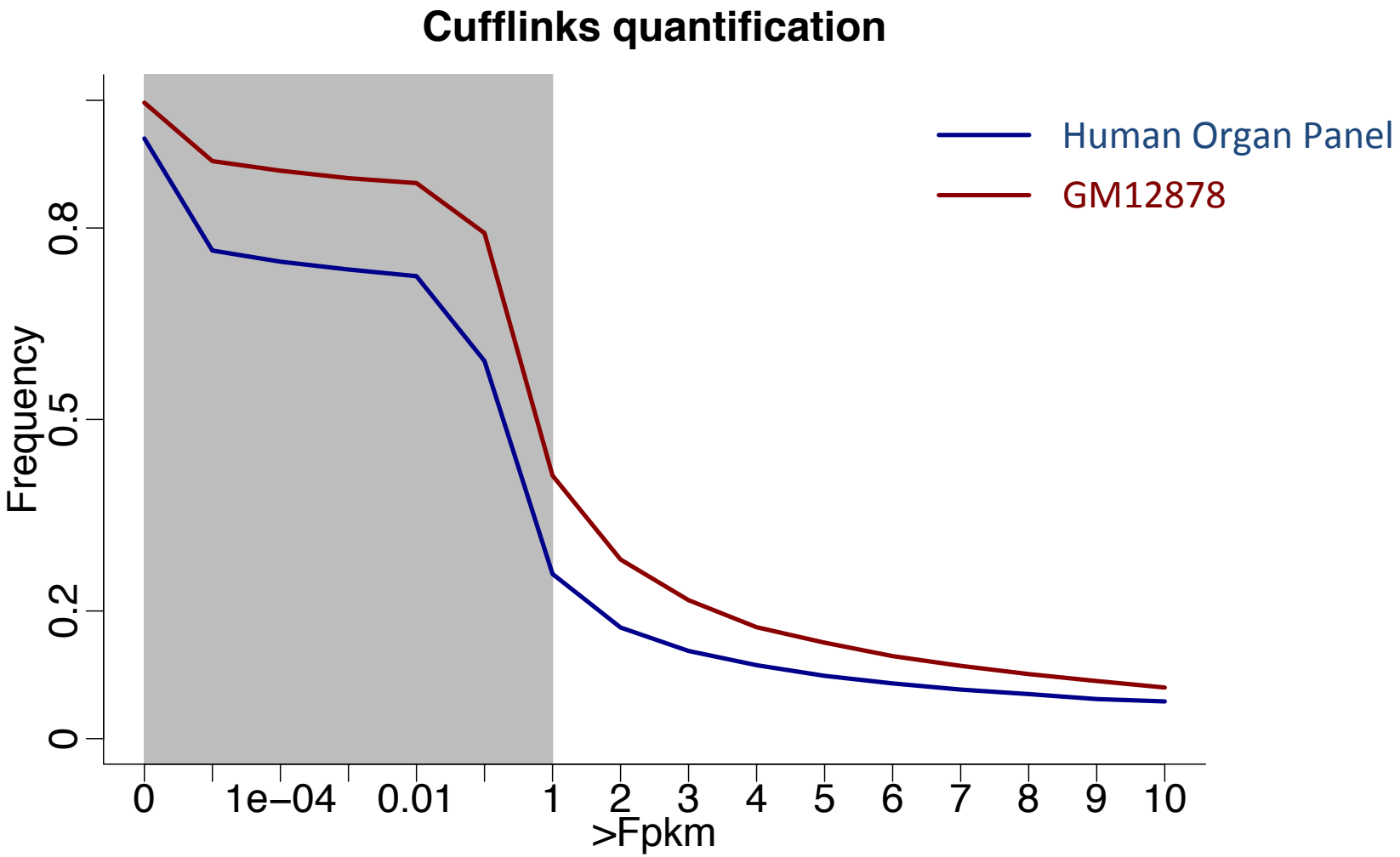
Illumina data from
- **GM12878**

Cufflinks quantification

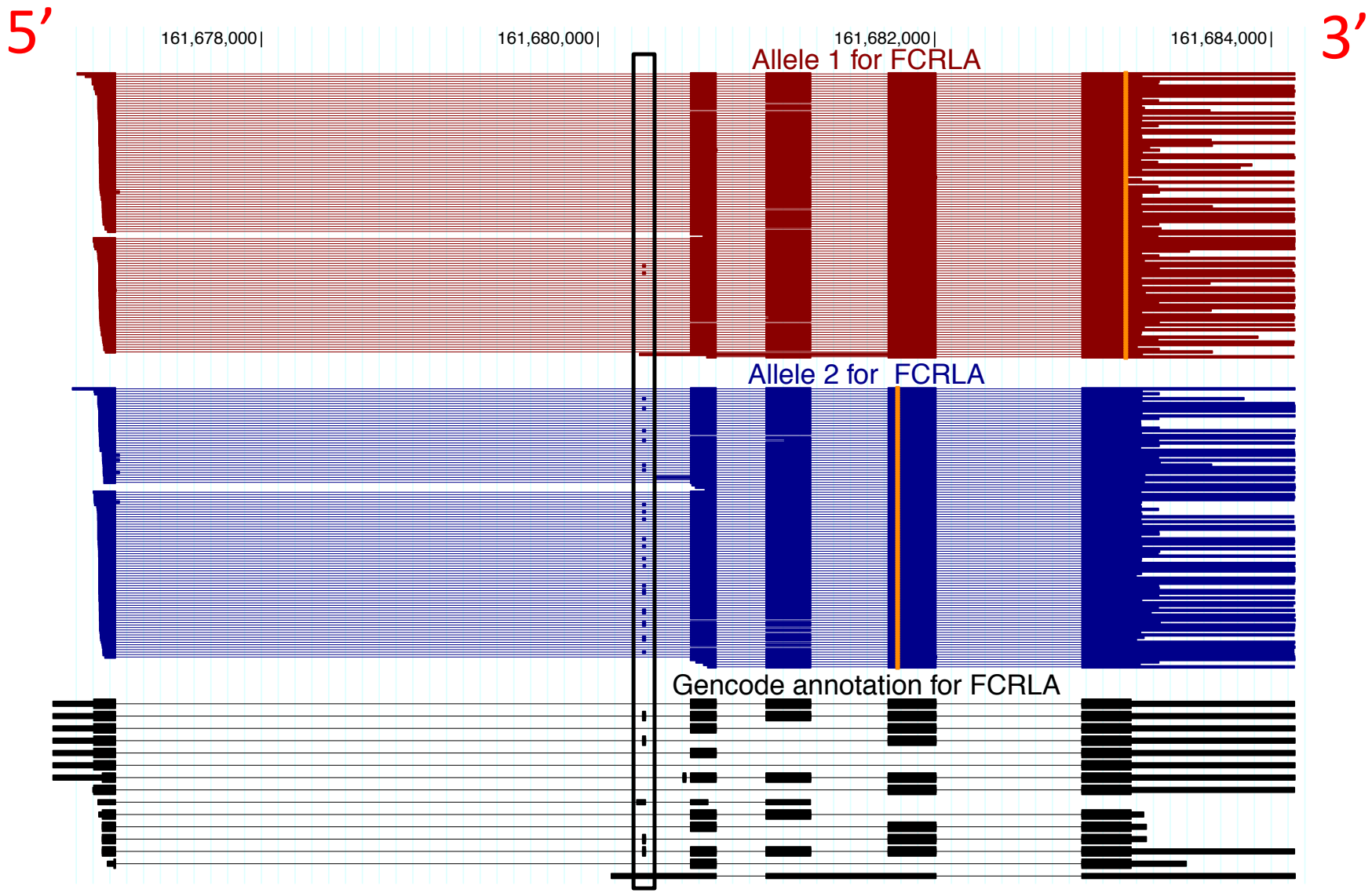


hut2006@med.cornell.edu

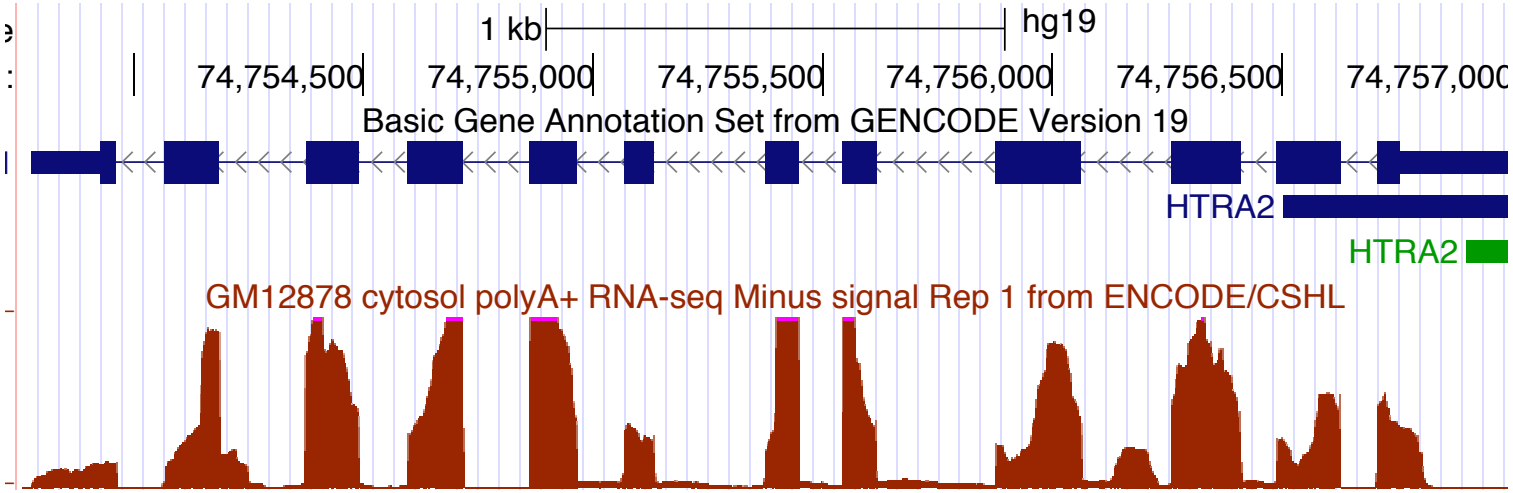
Enhanced annotation improves isoform quantification

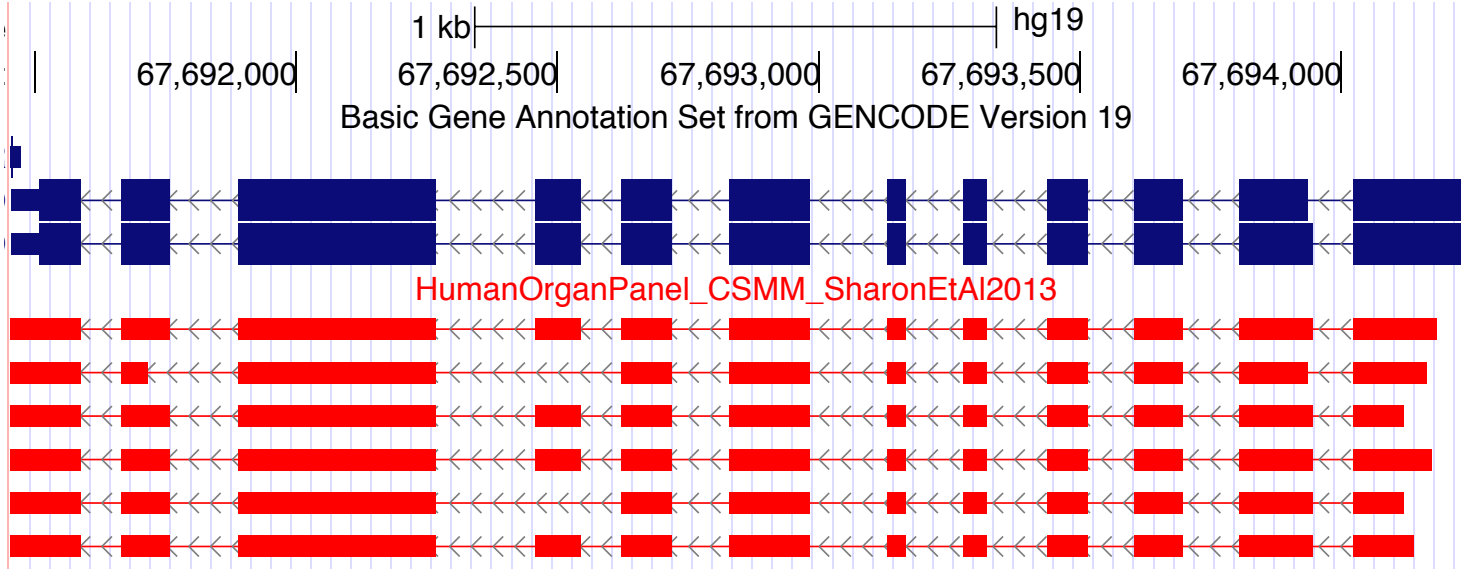


hut2006@med.cornell.edu



hut2006@med.cornell.edu



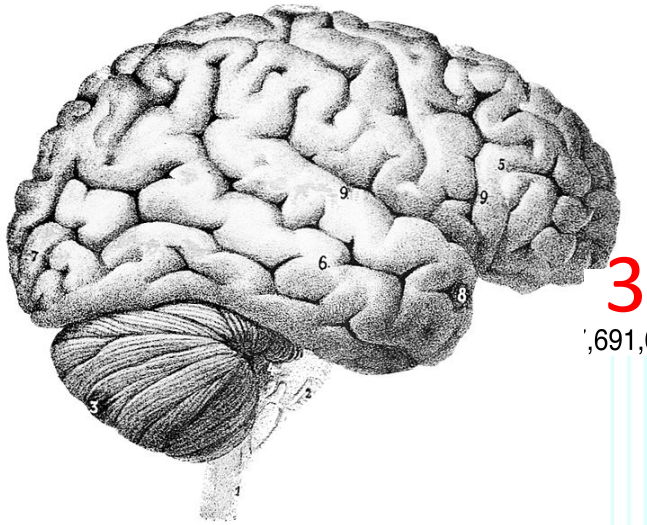


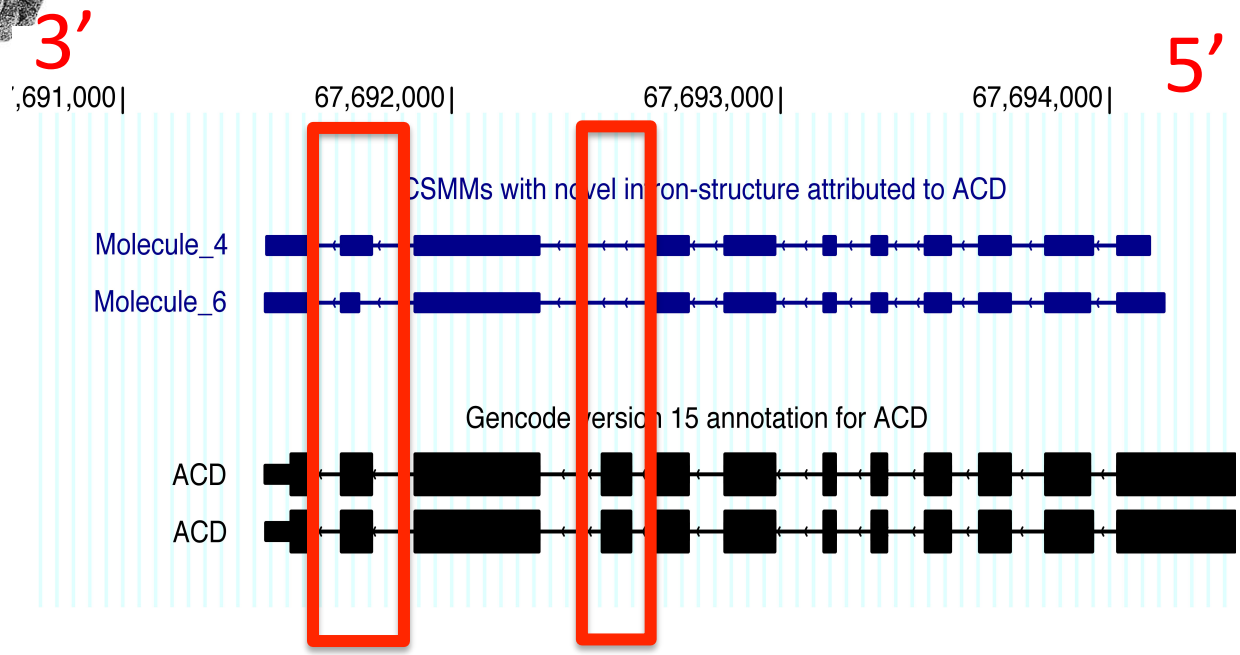
hut2006@med.cornell.edu

hut2006@med.cornell.edu

They took the paper but not the cover ☹

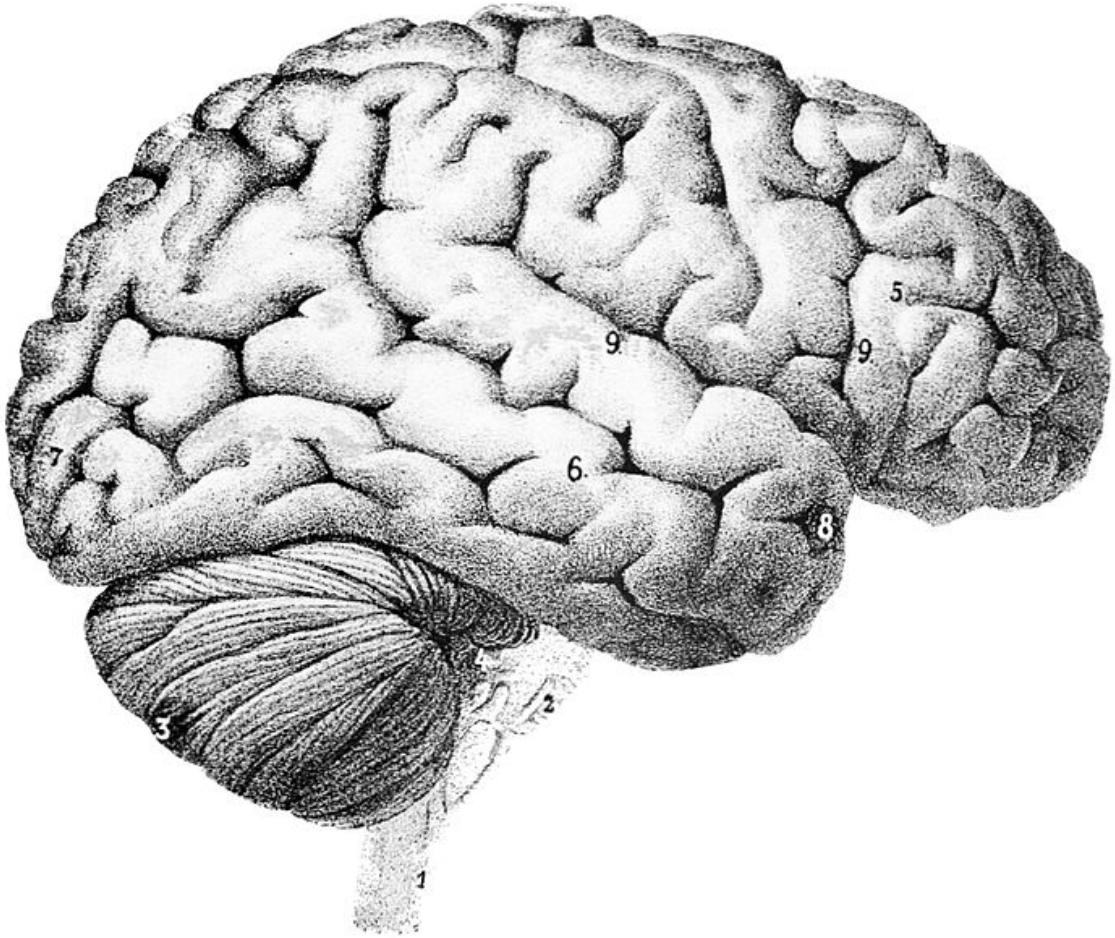
hut2006@med.cornell.edu





- Keep 1-to-1 correspondence of reads and RNA molecules
- Get LOTS of reads with low error-rate

hut2006@med.cornell.edu



hut2006@med.cornell.edu

Understanding brain isoforms through quantitative long-read sequencing

- SLR-RNA-Seq, Nature. Biotech 2015
- spISO-Seq, Genome Res. 2017

hut2006@med.cornell.edu

SLR–RNA–Seq

hut2006@med.cornell.edu

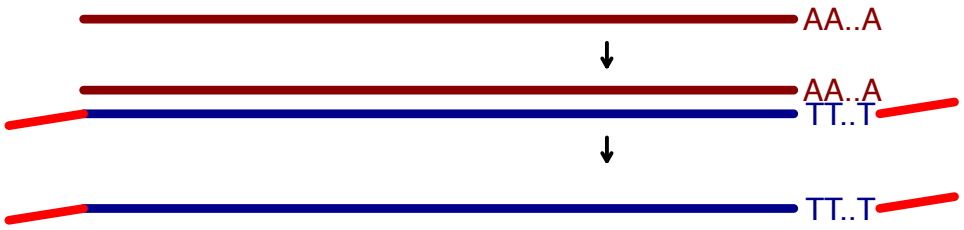
SLR–RNA–Seq

polyA–selected RNA

DNA–RNA hybrid with primers

hut2006@med.cornell.edu

SLR–RNA–Seq



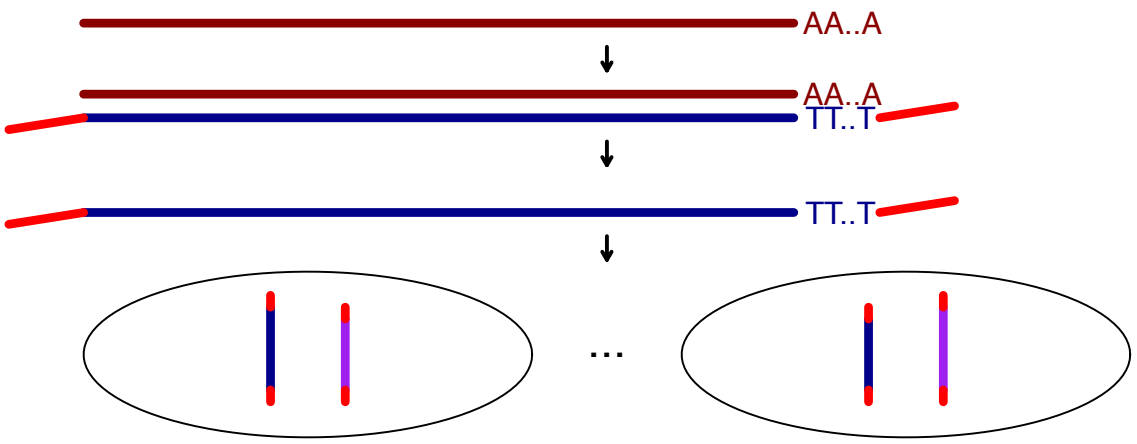
polyA–selected RNA

DNA–RNA hybrid with primers

sscDNA with primers

hut2006@med.cornell.edu

SLR–RNA–Seq



polyA–selected RNA

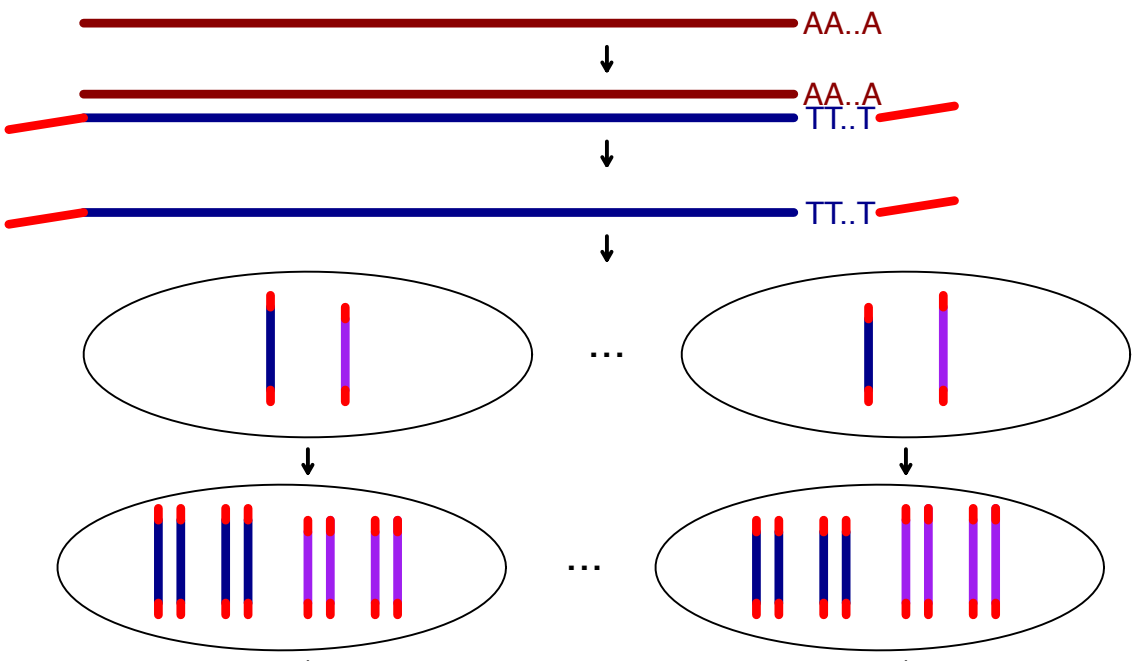
DNA–RNA hybrid with primers

sscDNA with primers

pools of few sscDNA–molecules per well

hut2006@med.cornell.edu

SLR–RNA–Seq



polyA–selected RNA

DNA–RNA hybrid with primers

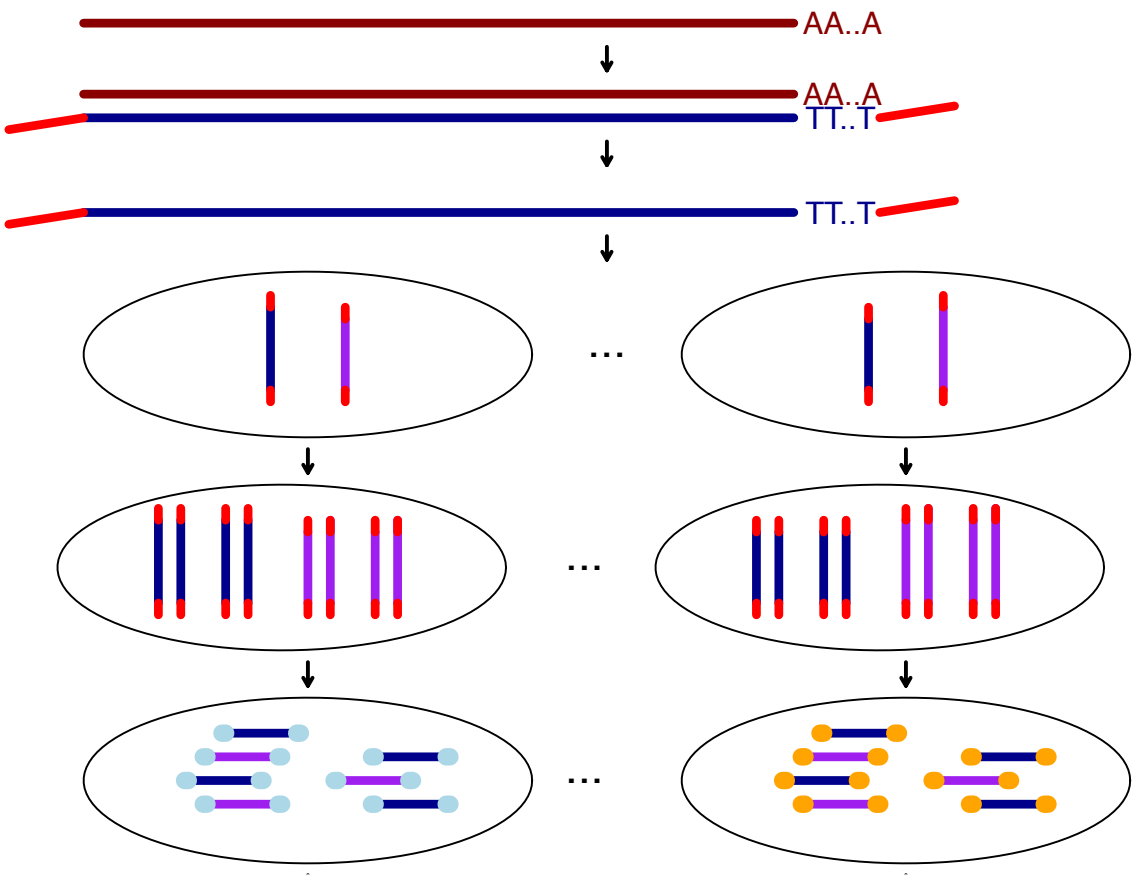
sscDNA with primers

pools of few sscDNA–molecules per well

amplified dscDNA–pools (per well)

hut2006@med.cornell.edu

SLR–RNA–Seq



polyA–selected RNA

DNA–RNA hybrid with primers

sscDNA with primers

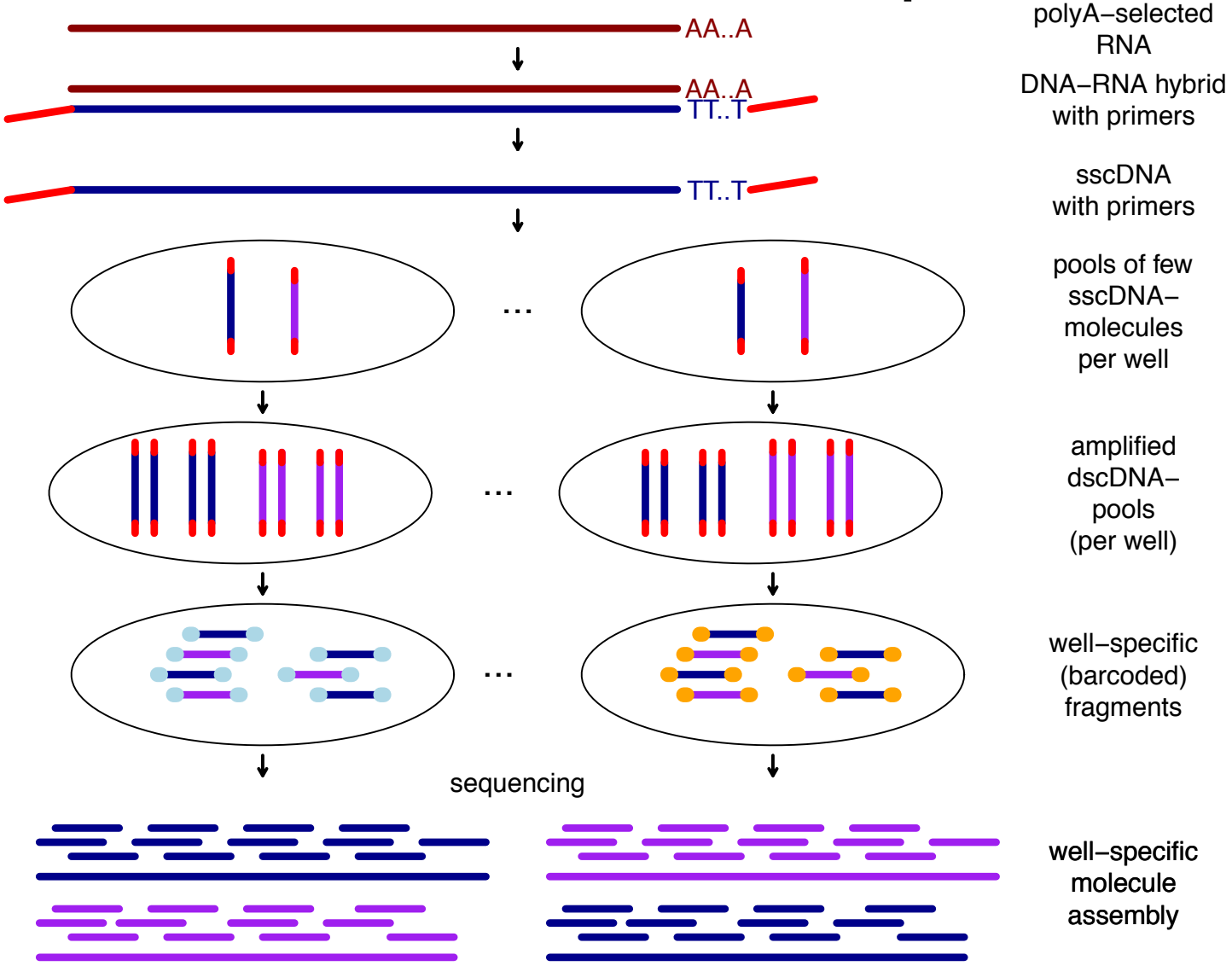
pools of few sscDNA–molecules per well

amplified dscDNA–pools (per well)

well–specific (barcoded) fragments

hut2006@med.cornell.edu

SLR–RNA–Seq



polyA–selected RNA

DNA–RNA hybrid with primers

sscDNA with primers

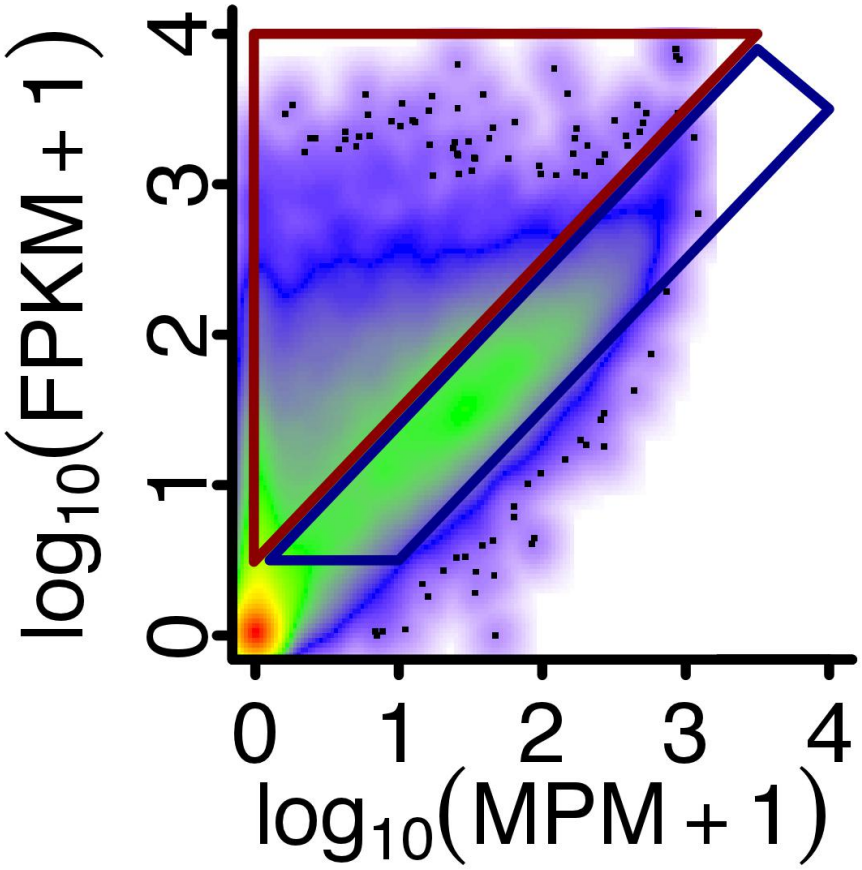
pools of few sscDNA–molecules per well

amplified dscDNA–pools (per well)

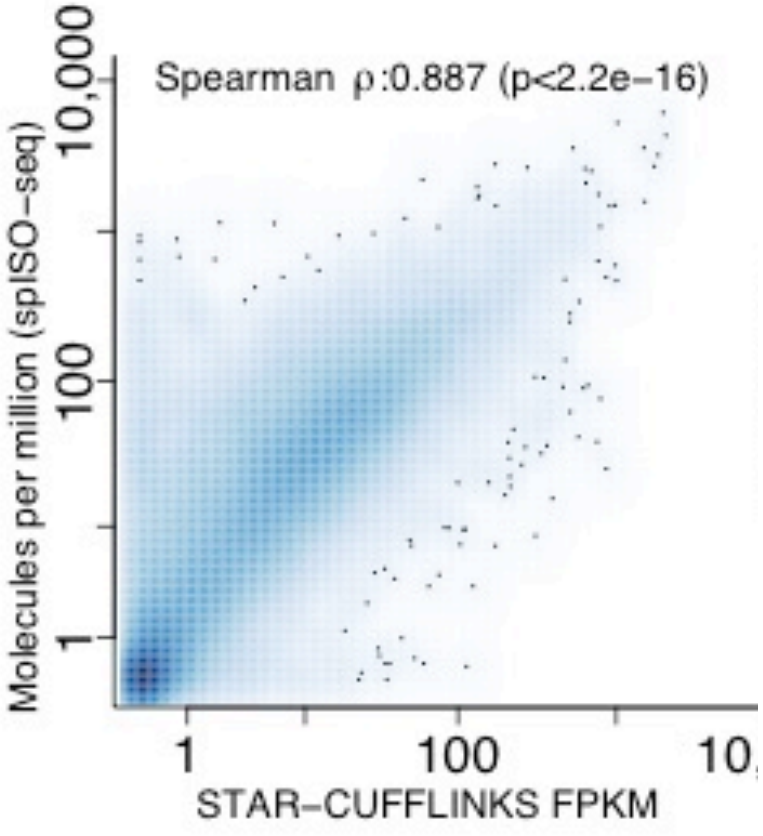
well–specific (barcoded) fragments

well–specific molecule assembly

hut2006@med.cornell.edu



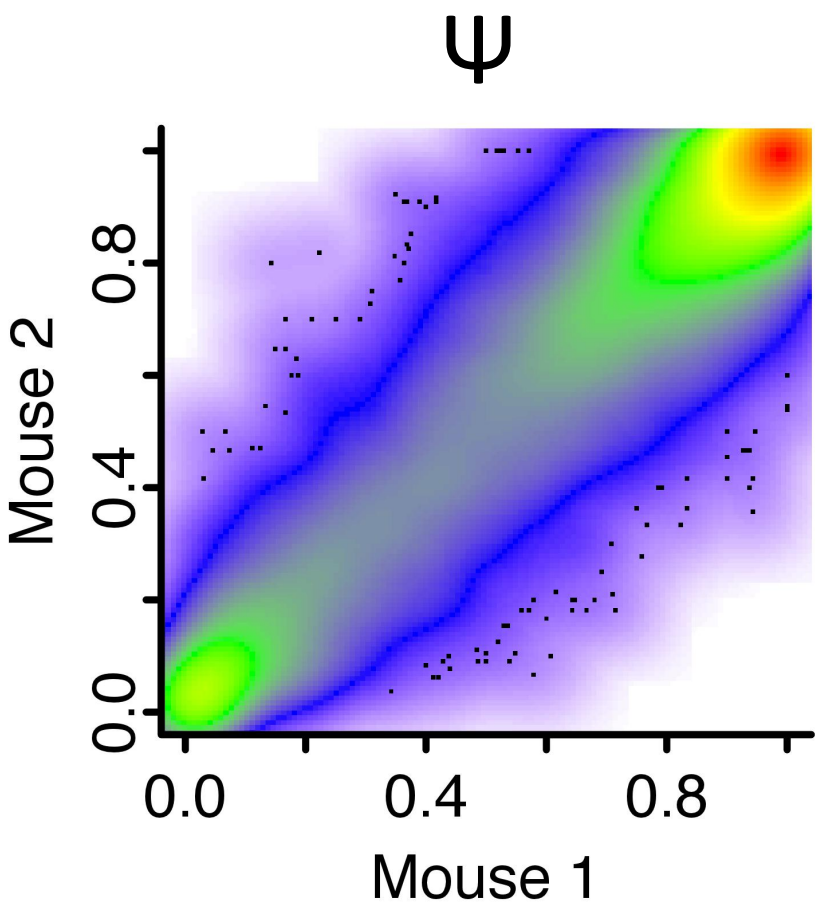
SLR-RNA-Seq vs. short reads
Tilgner et al,
Nature Biotech 2015



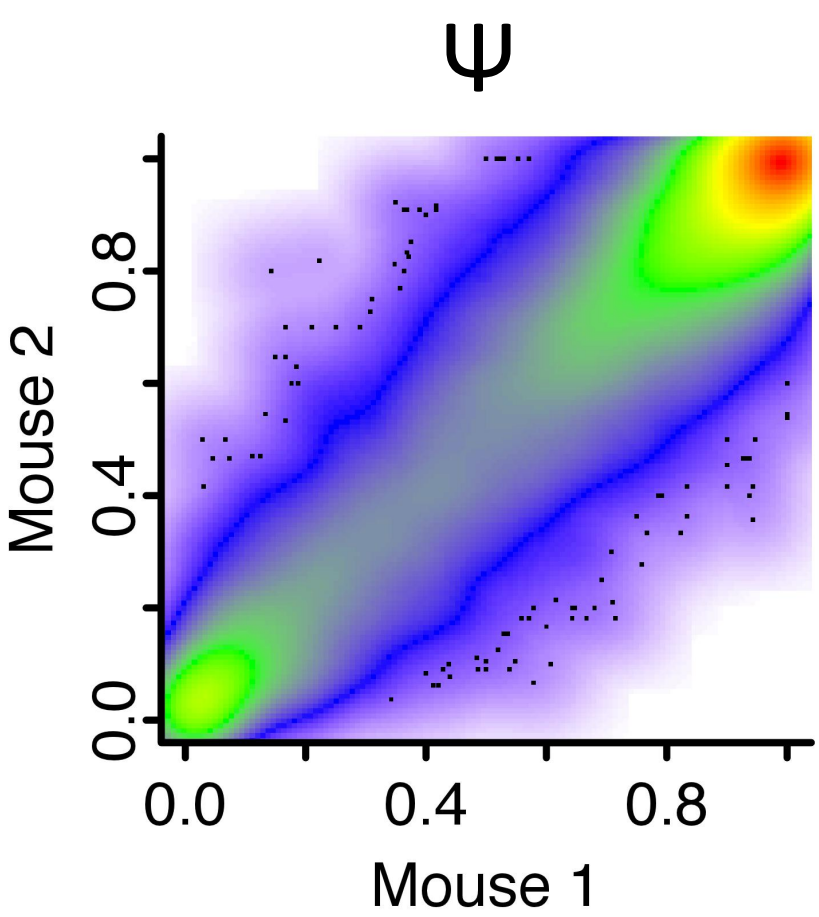
spISO-Seq vs. short reads
Tilgner et al,
Genome Research, 2017

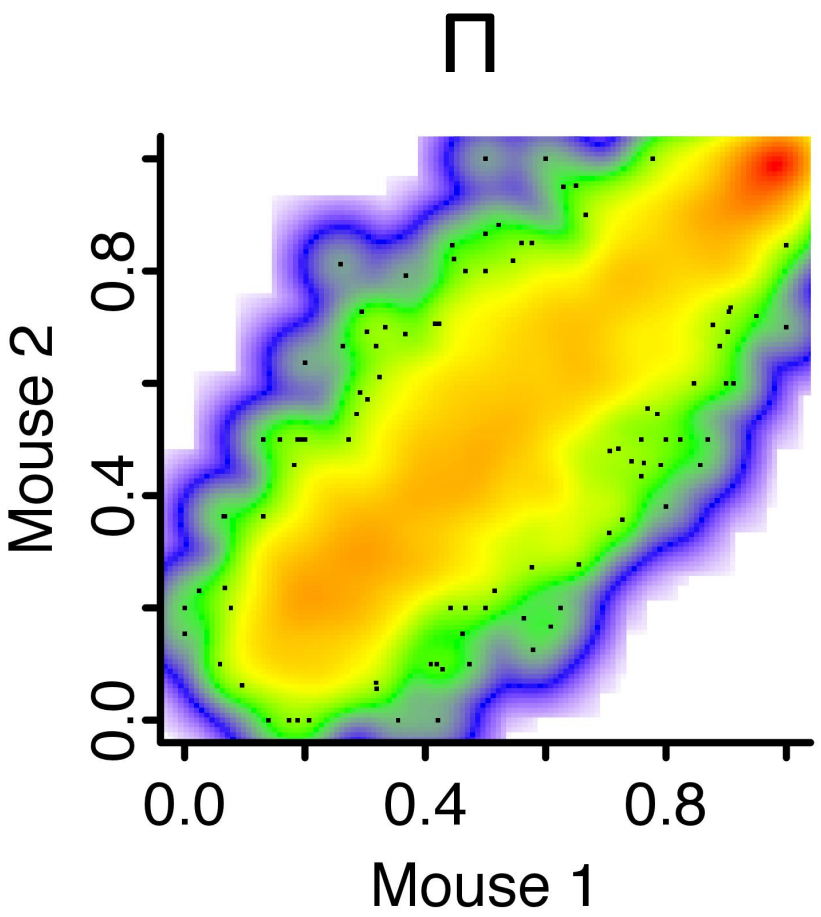
MPM: molecules per million (expression measurement for long reads)

hut2006@med.cornell.edu



hut2006@med.cornell.edu

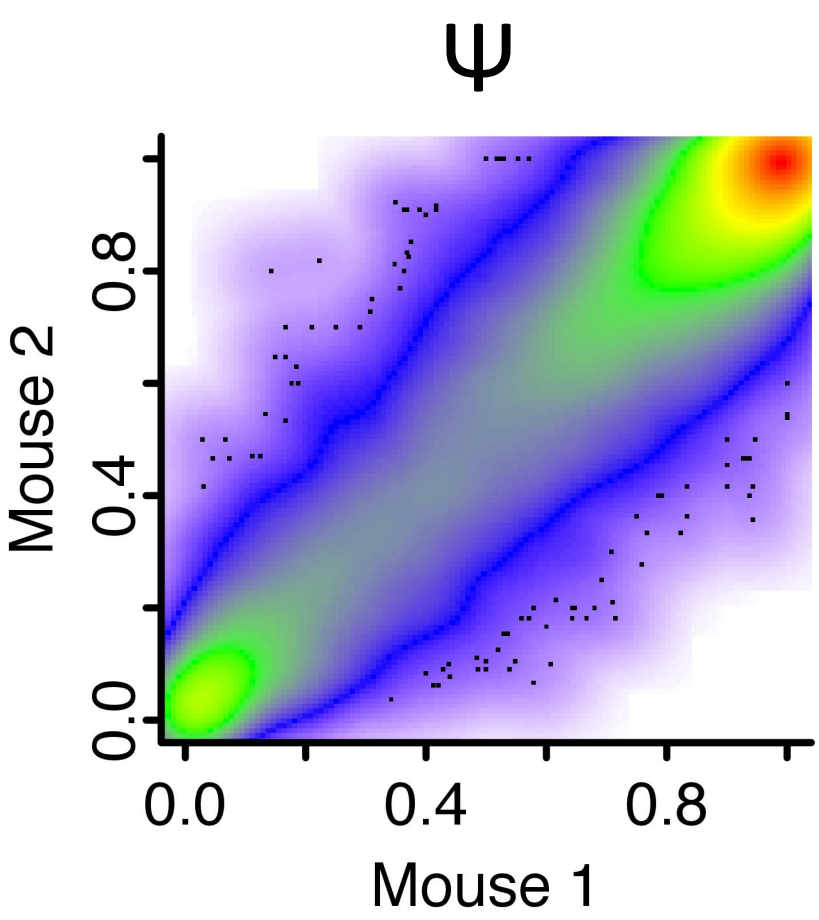


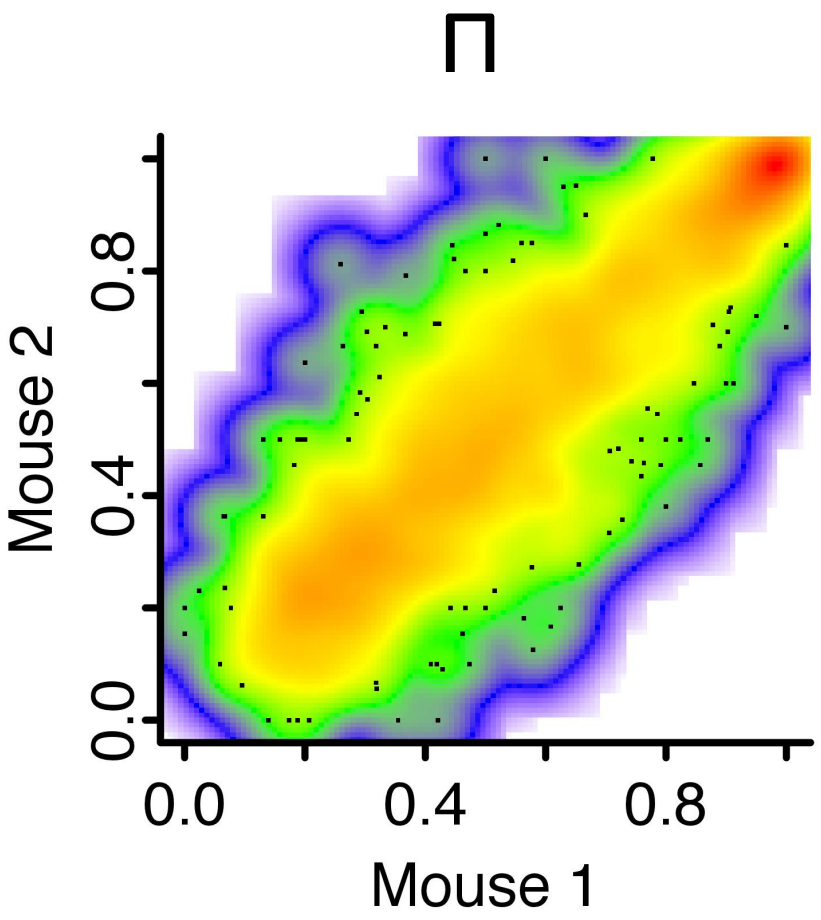


PSI: Percent Spliced In for exons and splice sites (Wang et al, 2008)

PI: Percent Isoform values (for long reads)

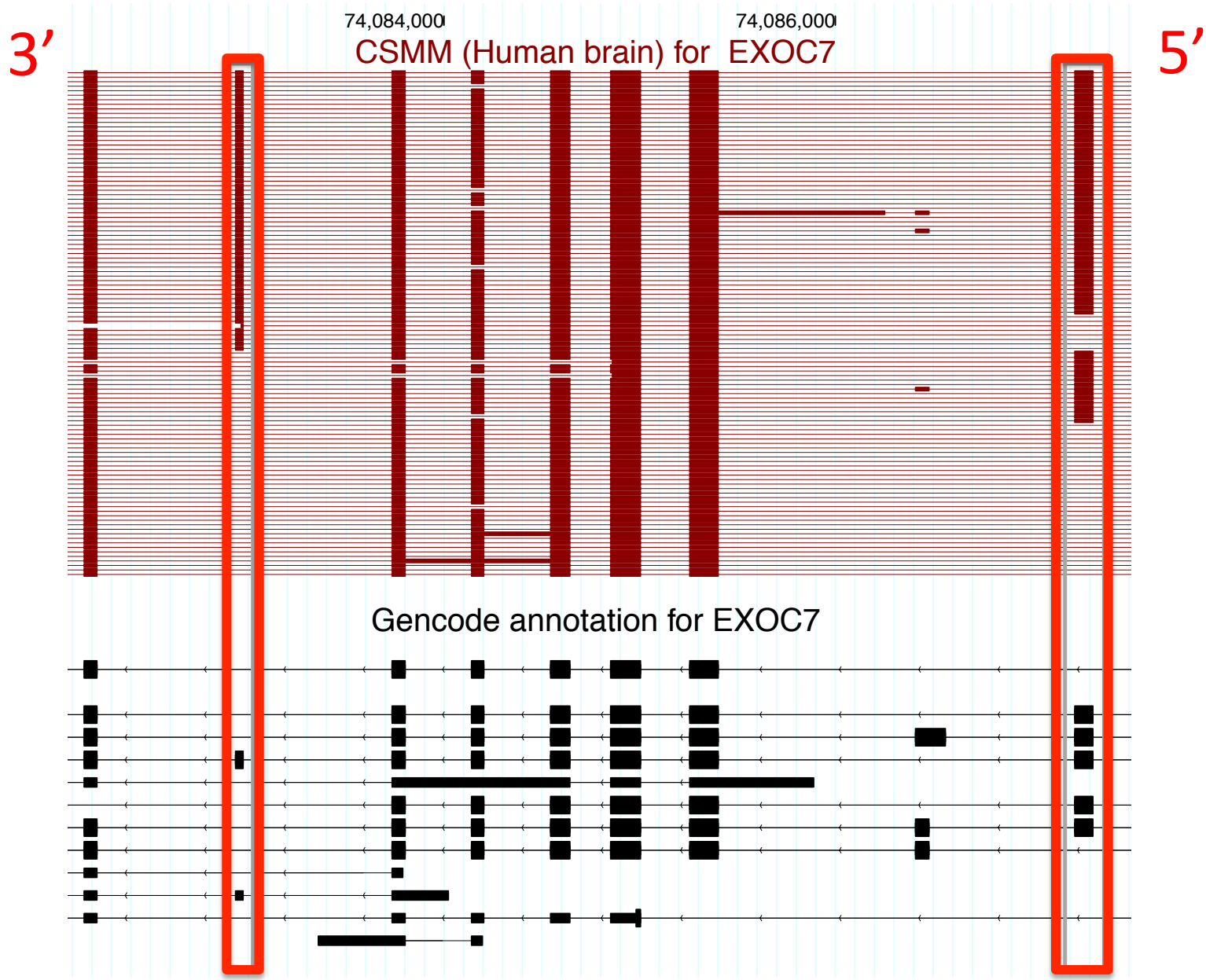
hut2006@med.cornell.edu



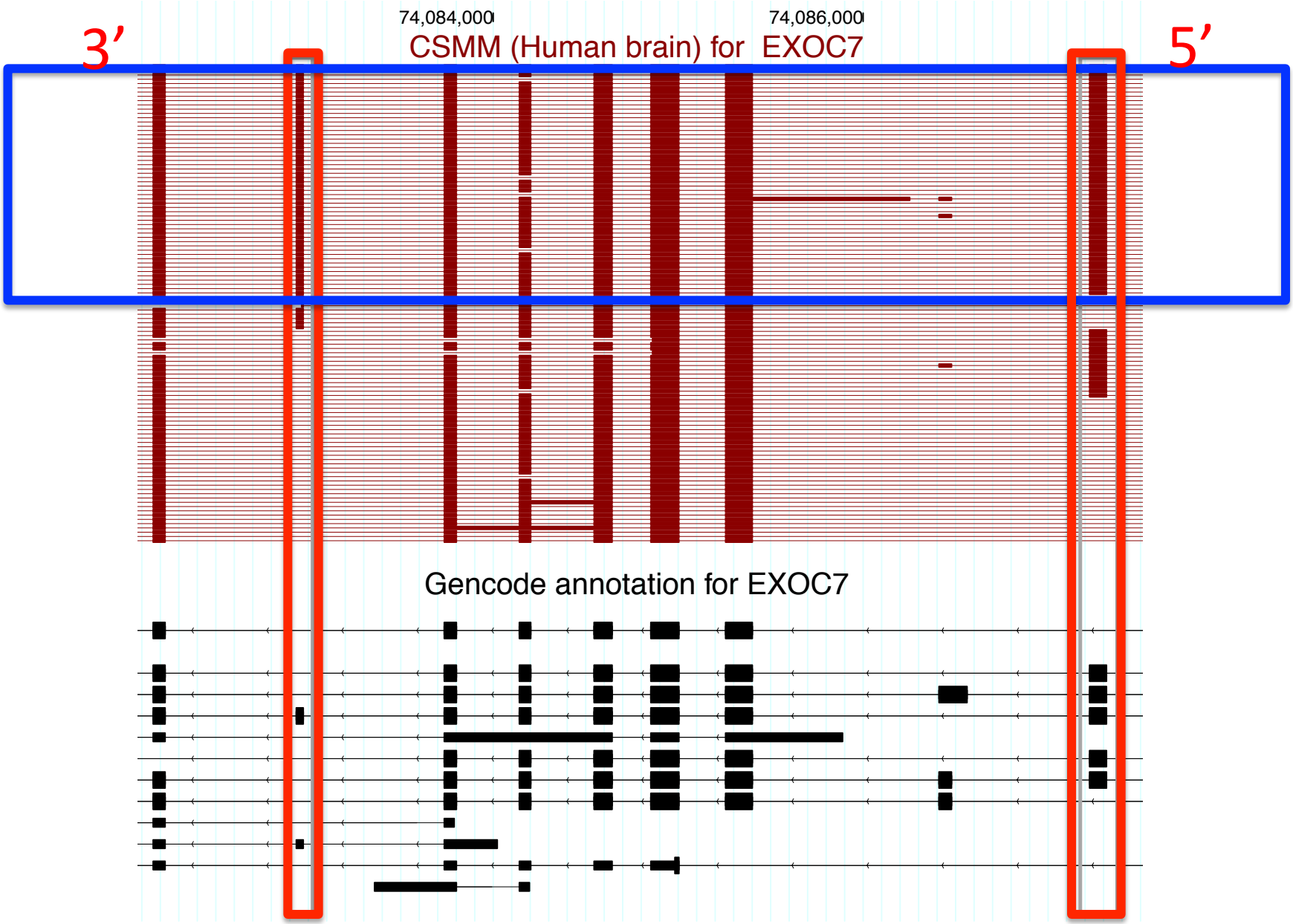


Data at: http://stanford.edu/~htilgner/2014_humanMouseBrain_SLR_RNA_Seq/index_SLRseq.html

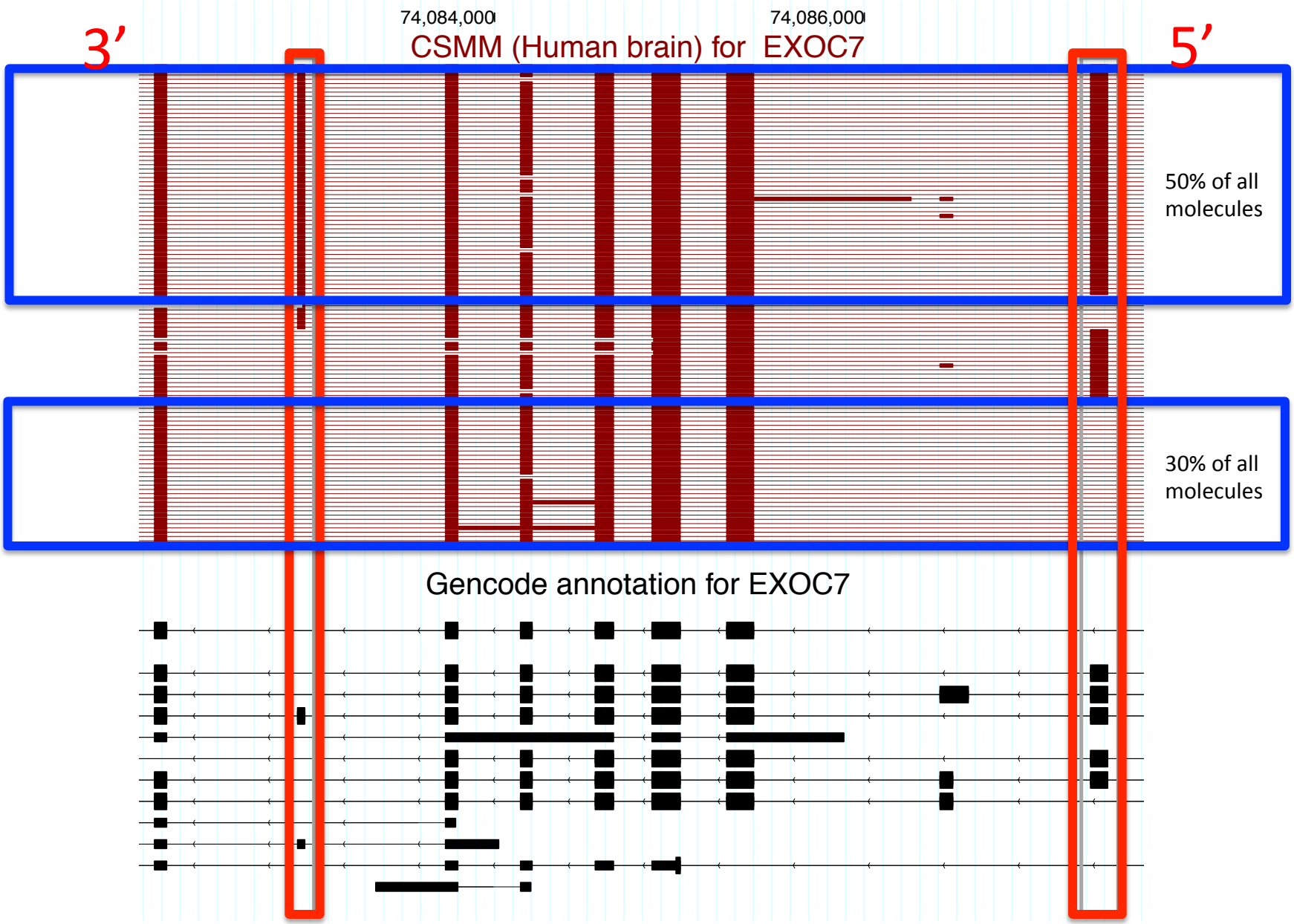
hut2006@med.cornell.edu



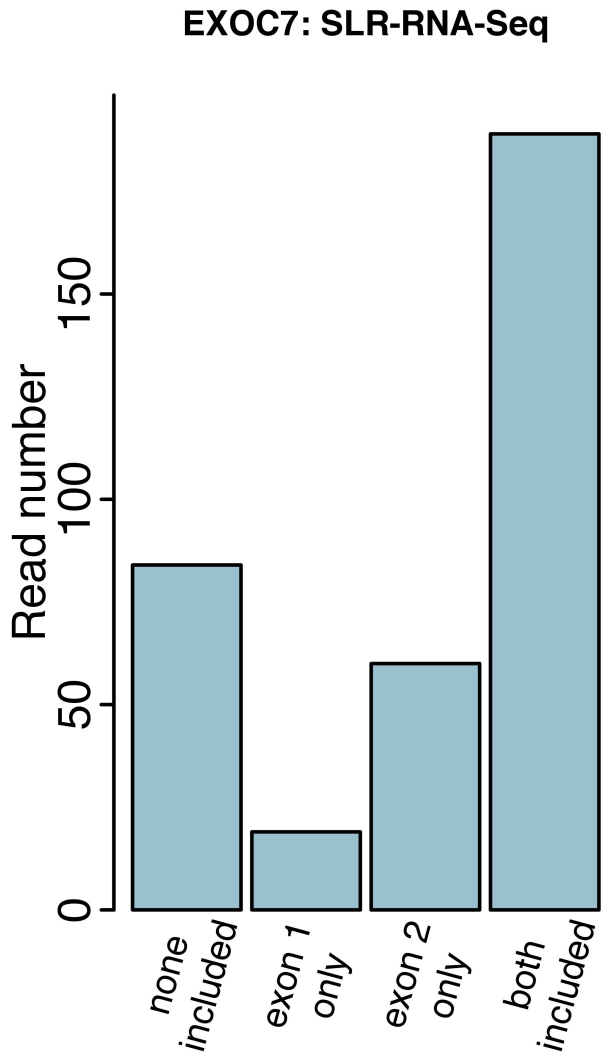
hut2006@med.cornell.edu



hut2006@med.cornell.edu

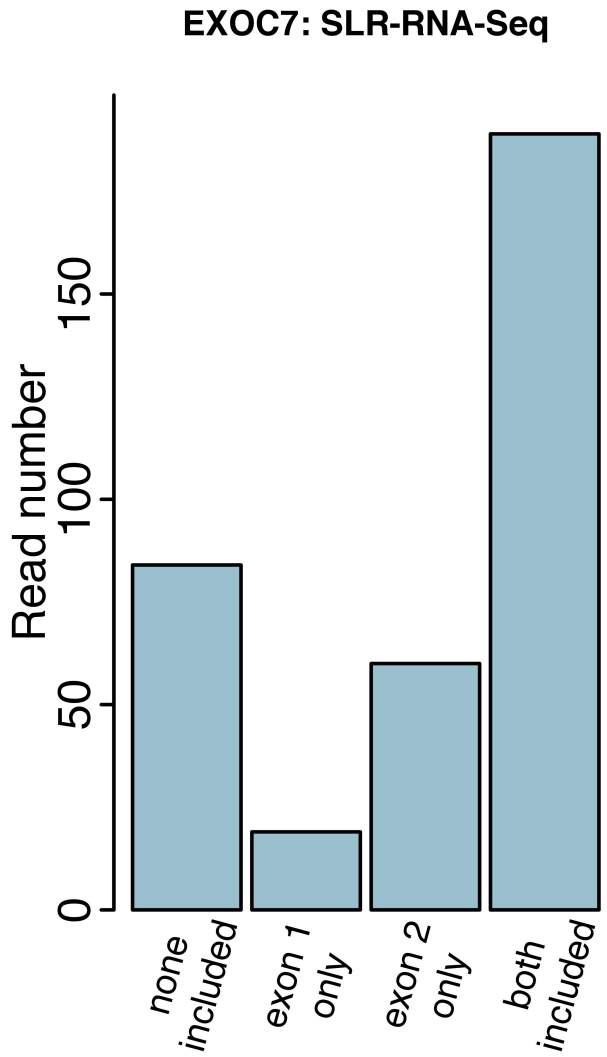


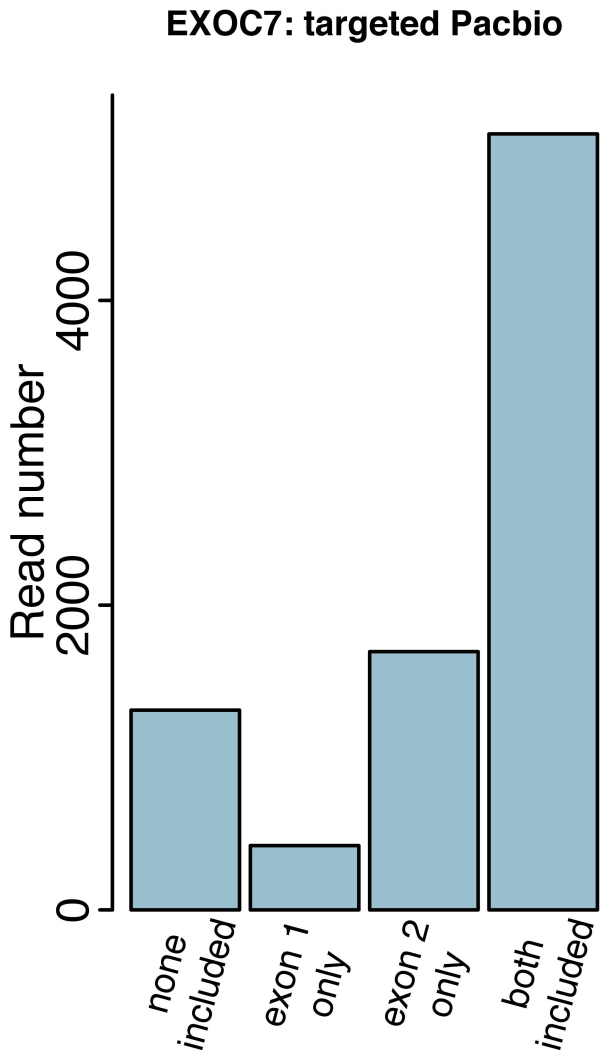
hut2006@med.cornell.edu



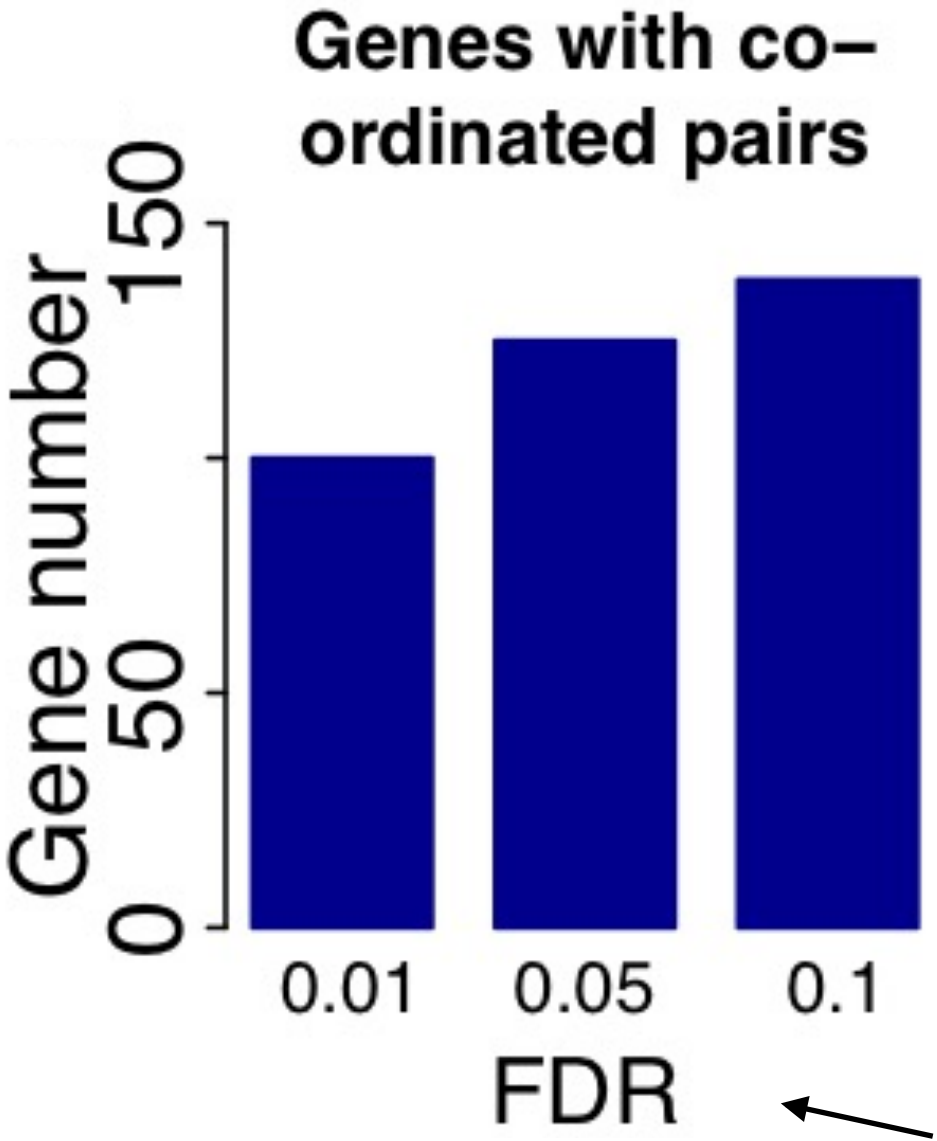
hut2006@med.cornell.edu

Coordination is (largely) reproducible in targeted experiment

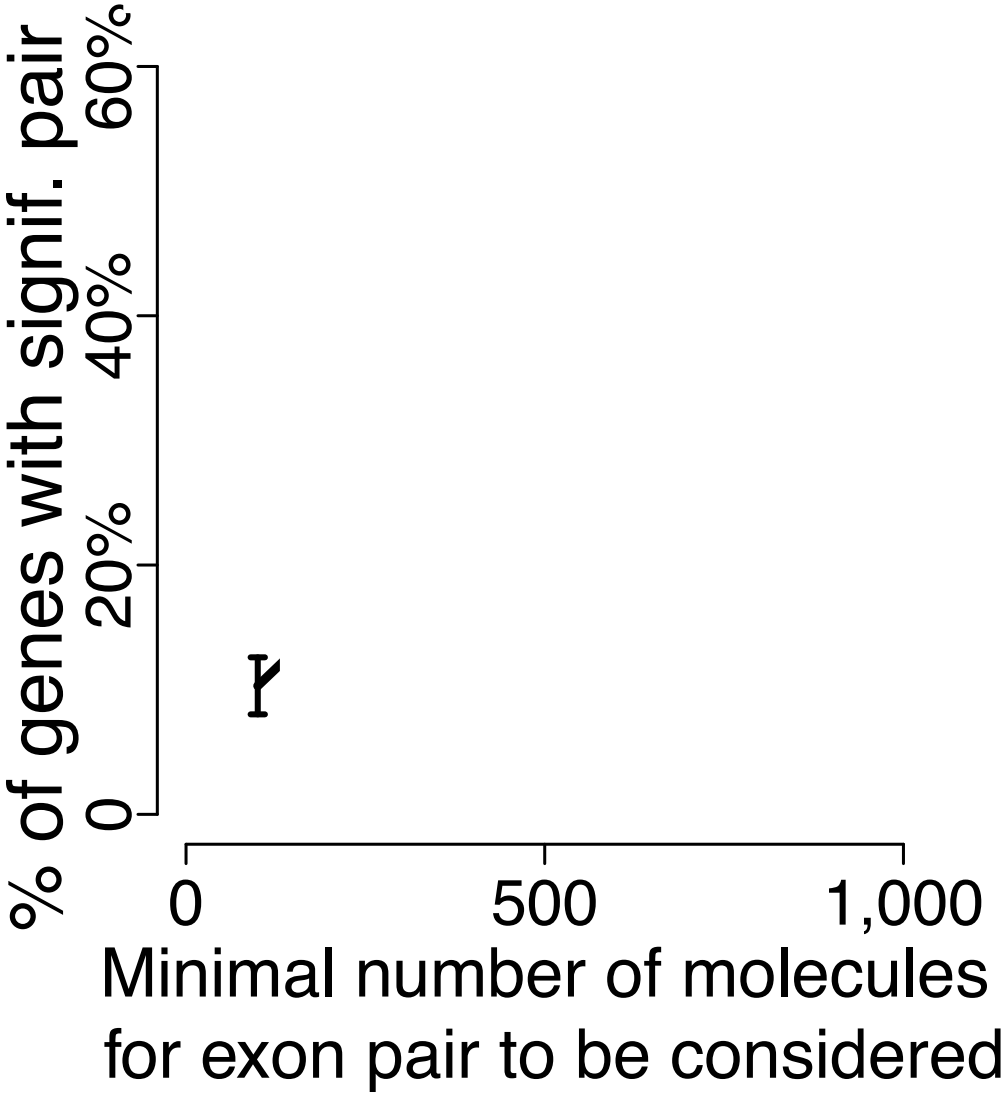




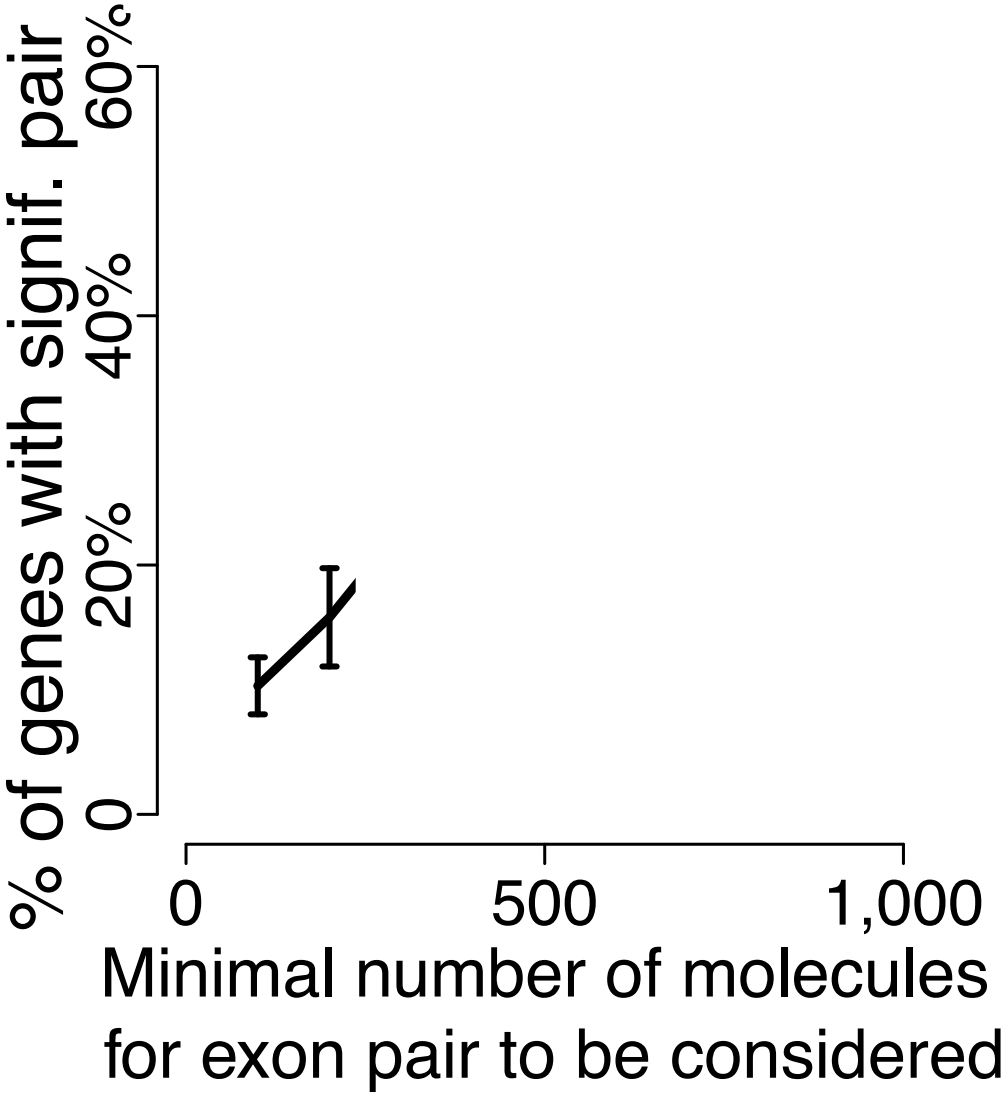
hut2006@med.cornell.edu



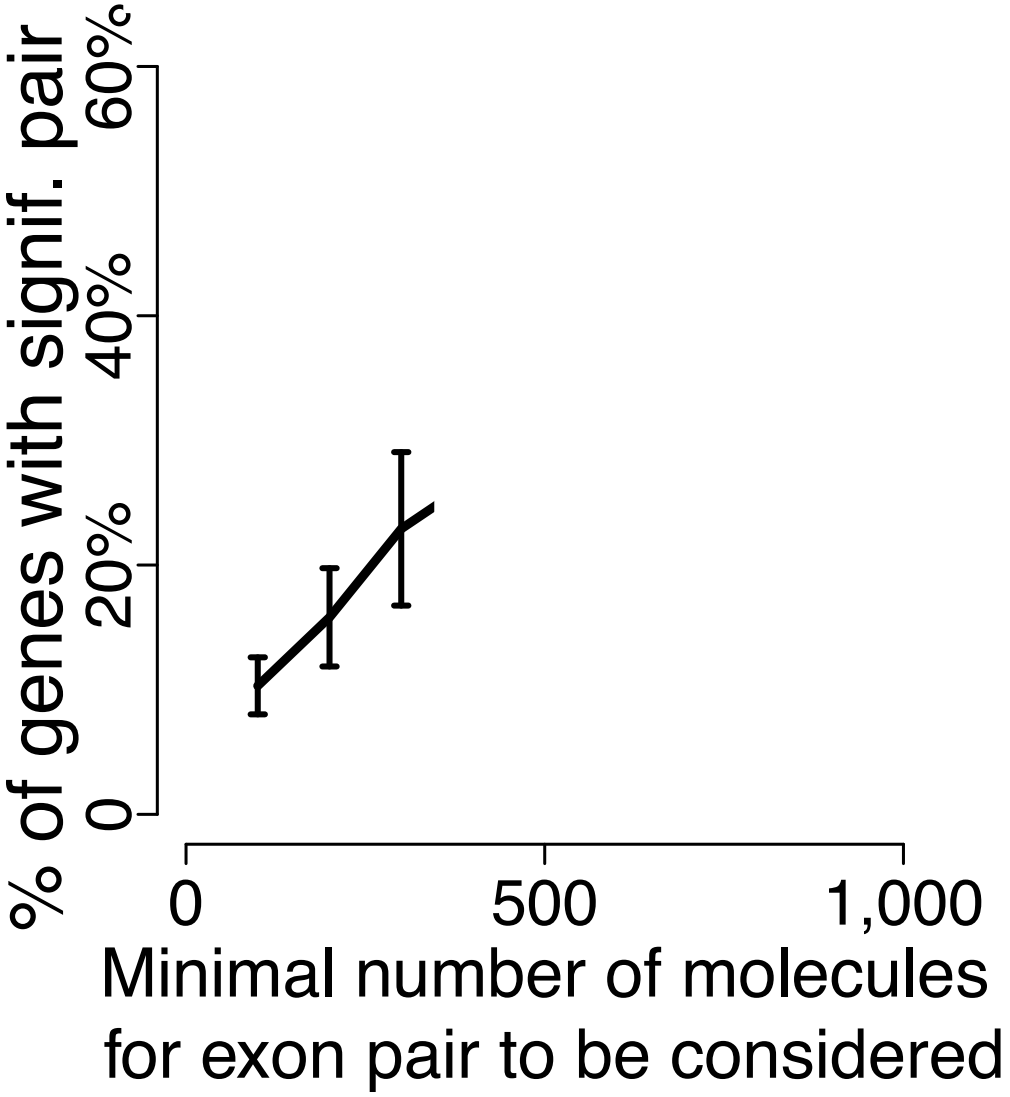
hut2006@med.cornell.edu



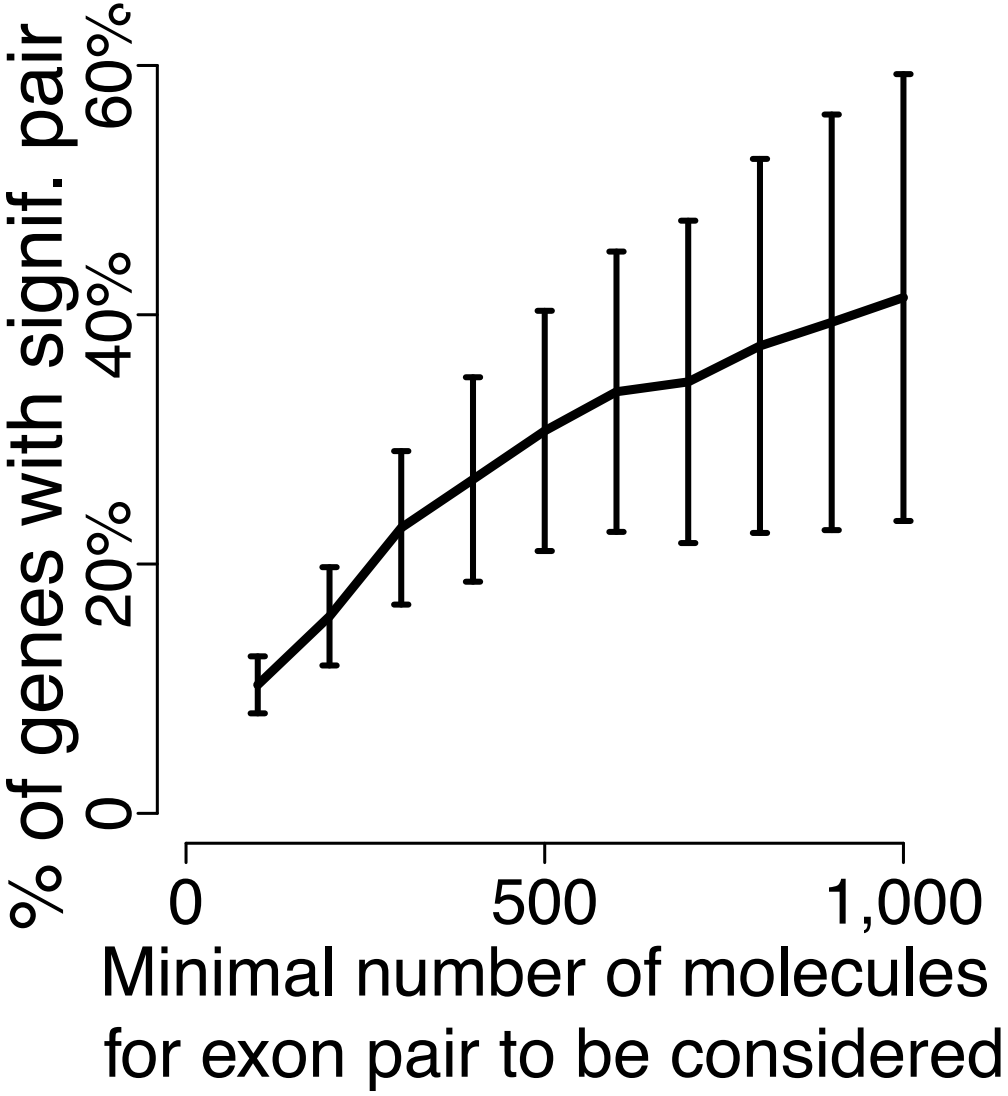
hut2006@med.cornell.edu



hut2006@med.cornell.edu



hut2006@med.cornell.edu

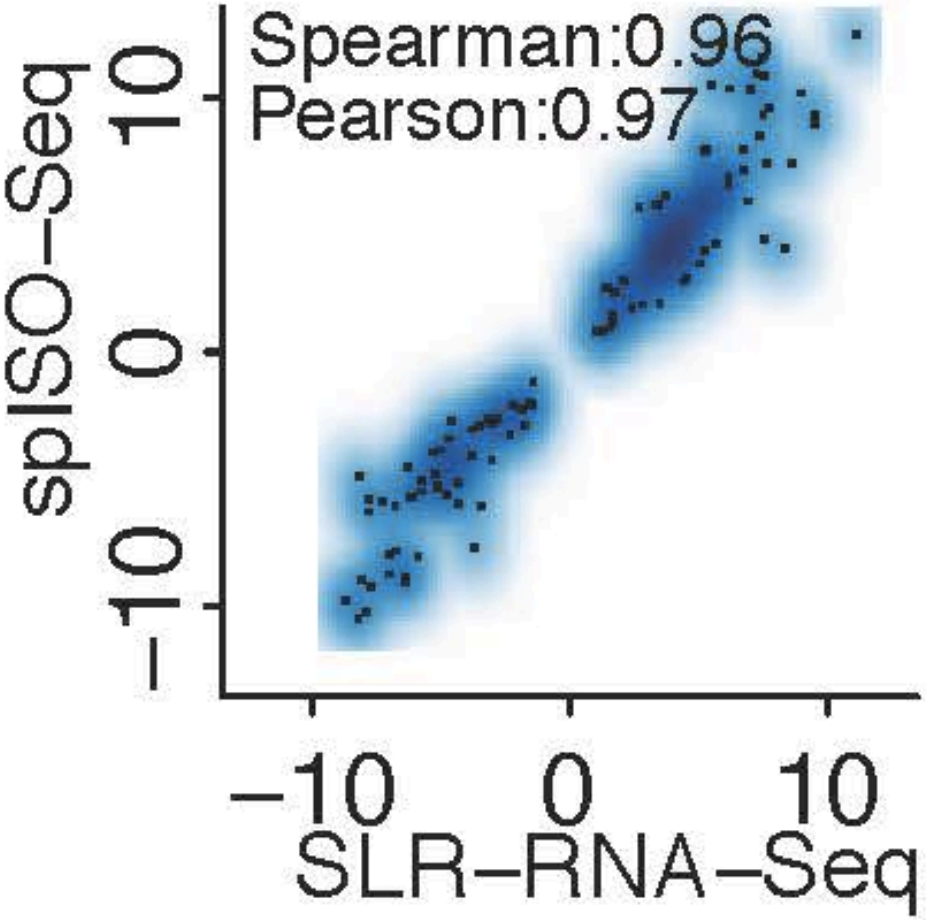


23.5%-59.3% (95% CI) of genes with multiple distant alternative exons are expected to show coordination at a log-odds-ratio of 0.5 or above.

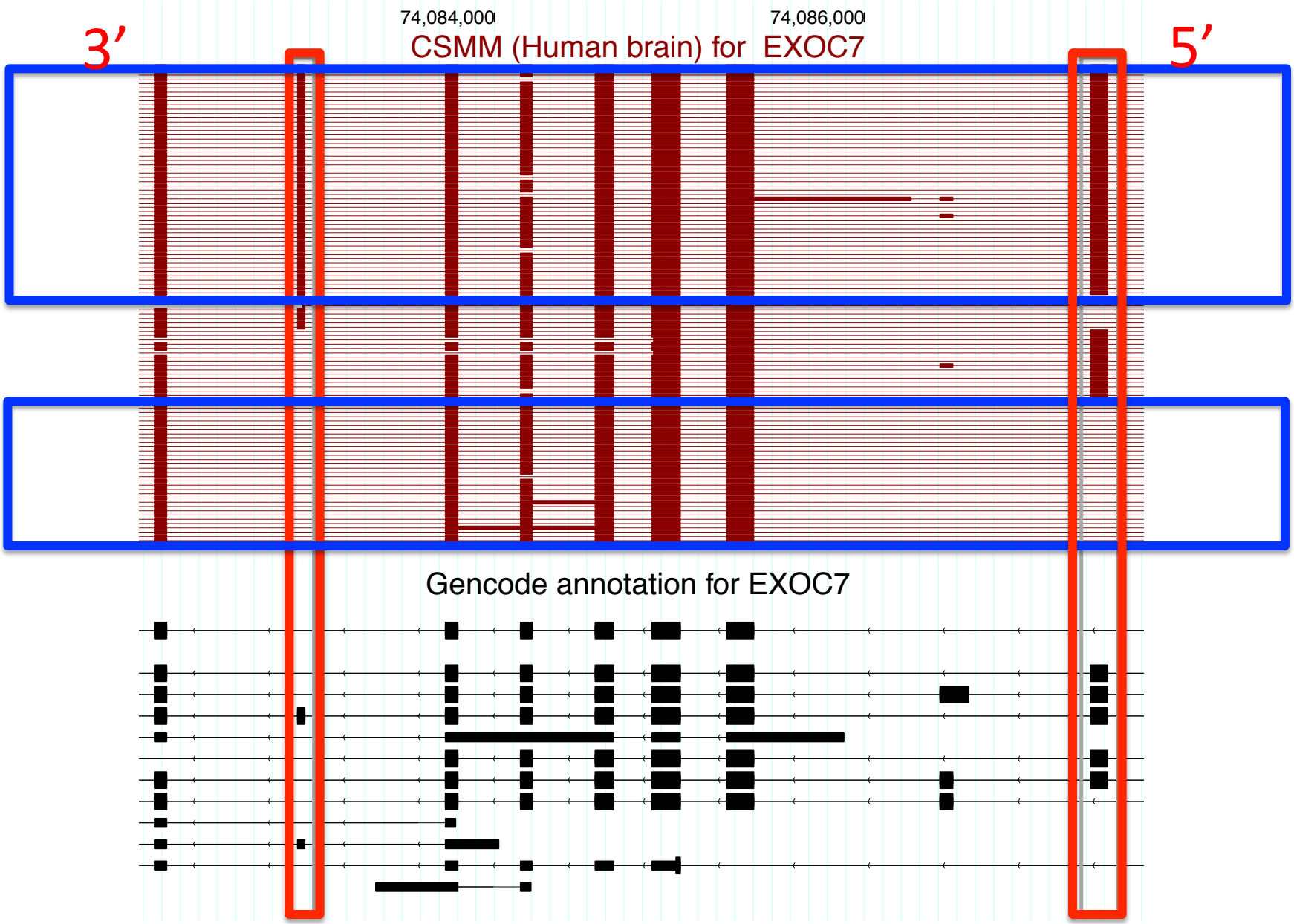
THE justification for isoform sequencing (otherwise no long reads needed … just get probability of isoform from product of probabilities of variables sites)

hut2006@med.cornell.edu

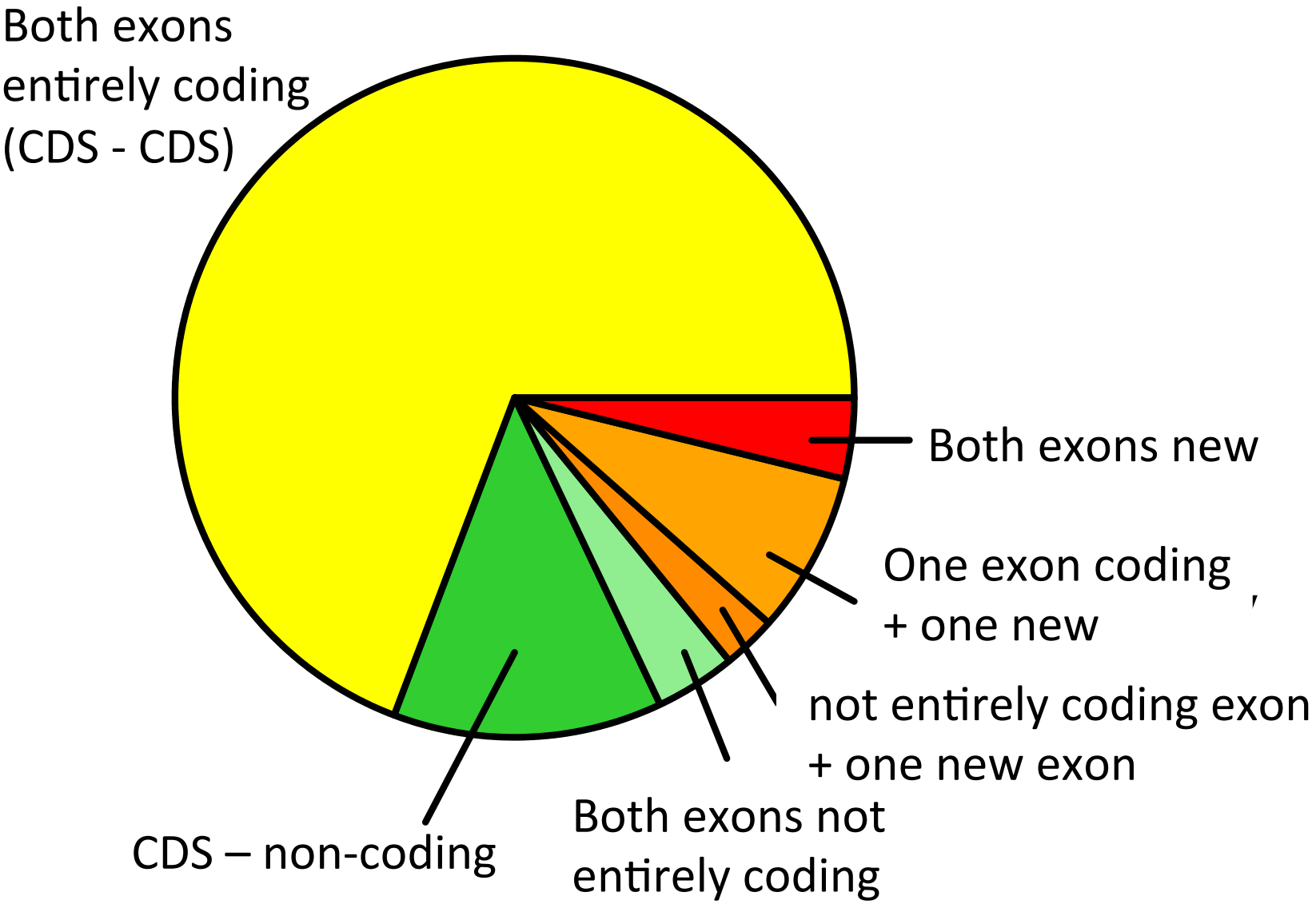
Comparability of coordination quantification



hut2006@med.cornell.edu



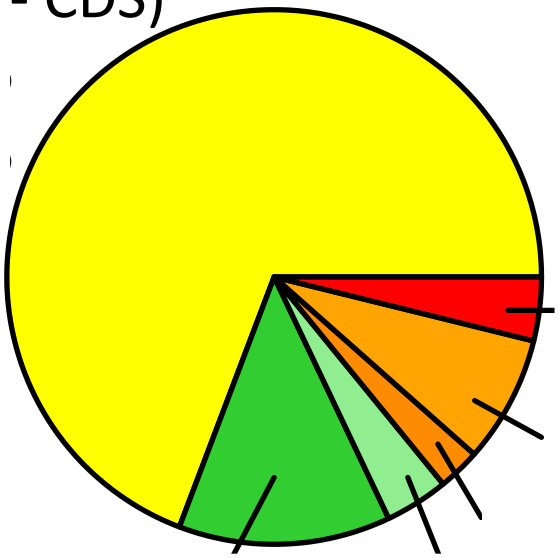
hut2006@med.cornell.edu

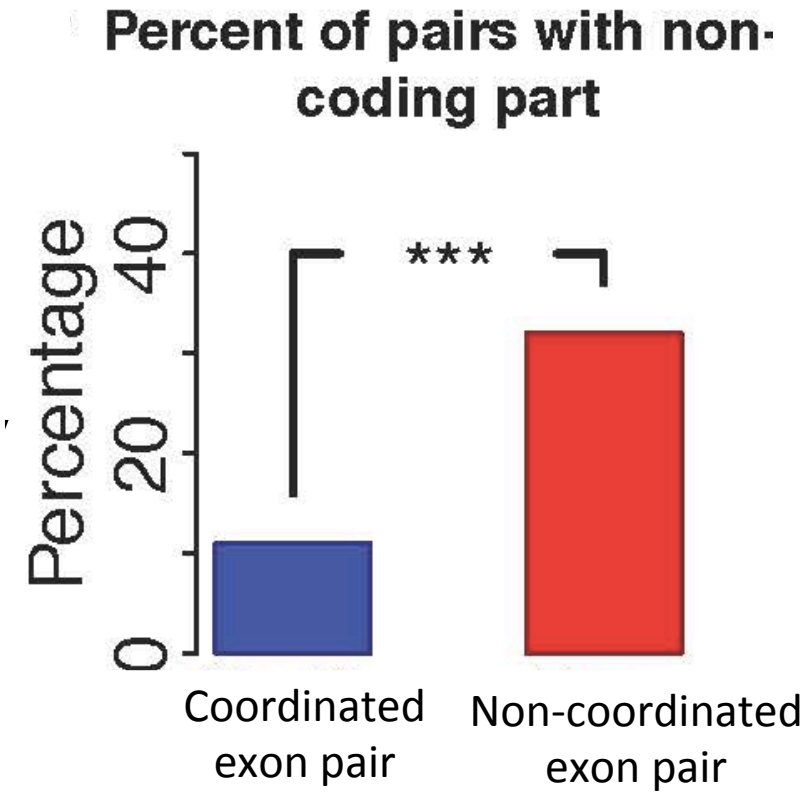


A phased brain proteome ?

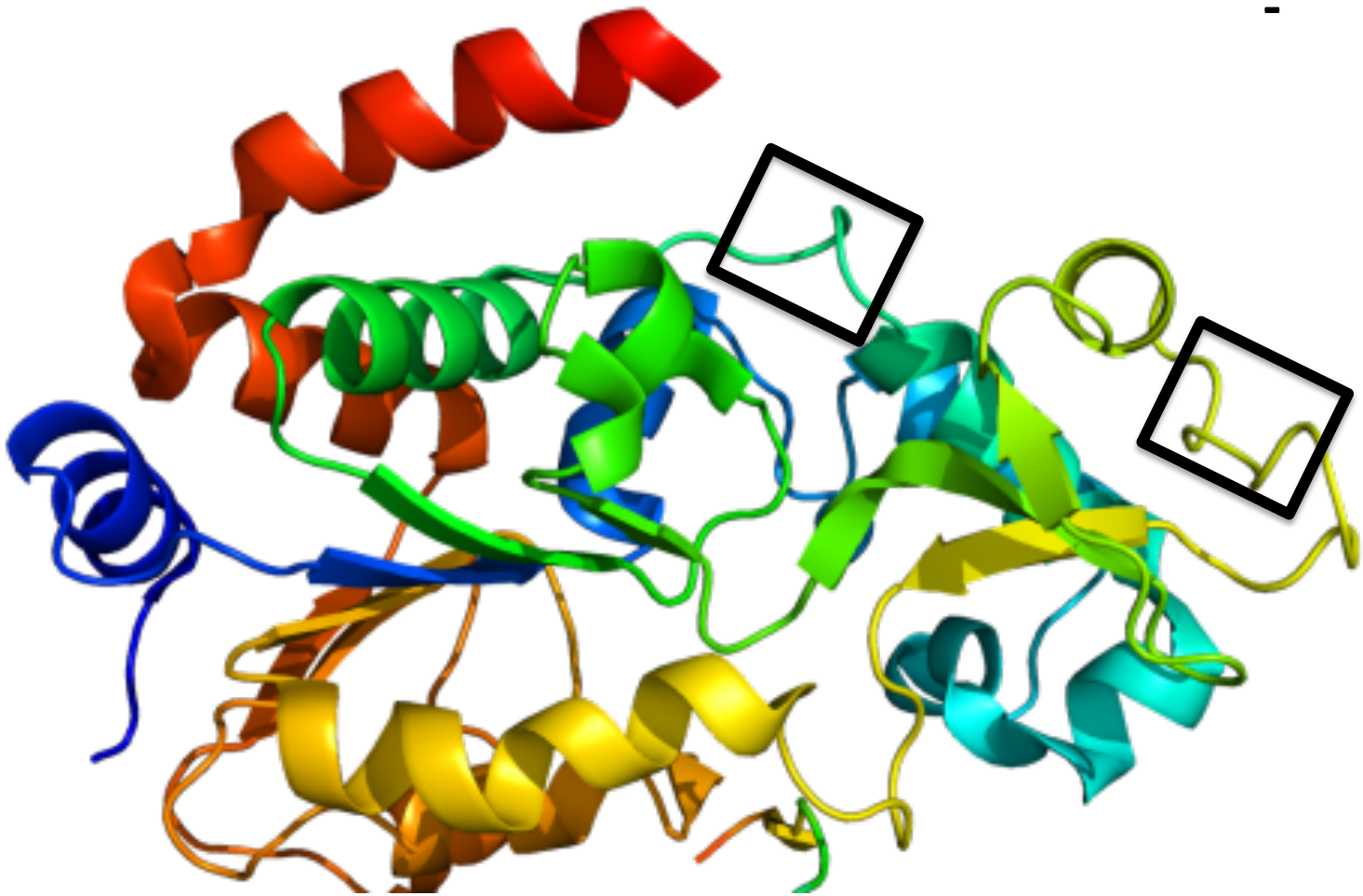
hut2006@med.cornell.edu

Both exons entirely coding (CDS - CDS)

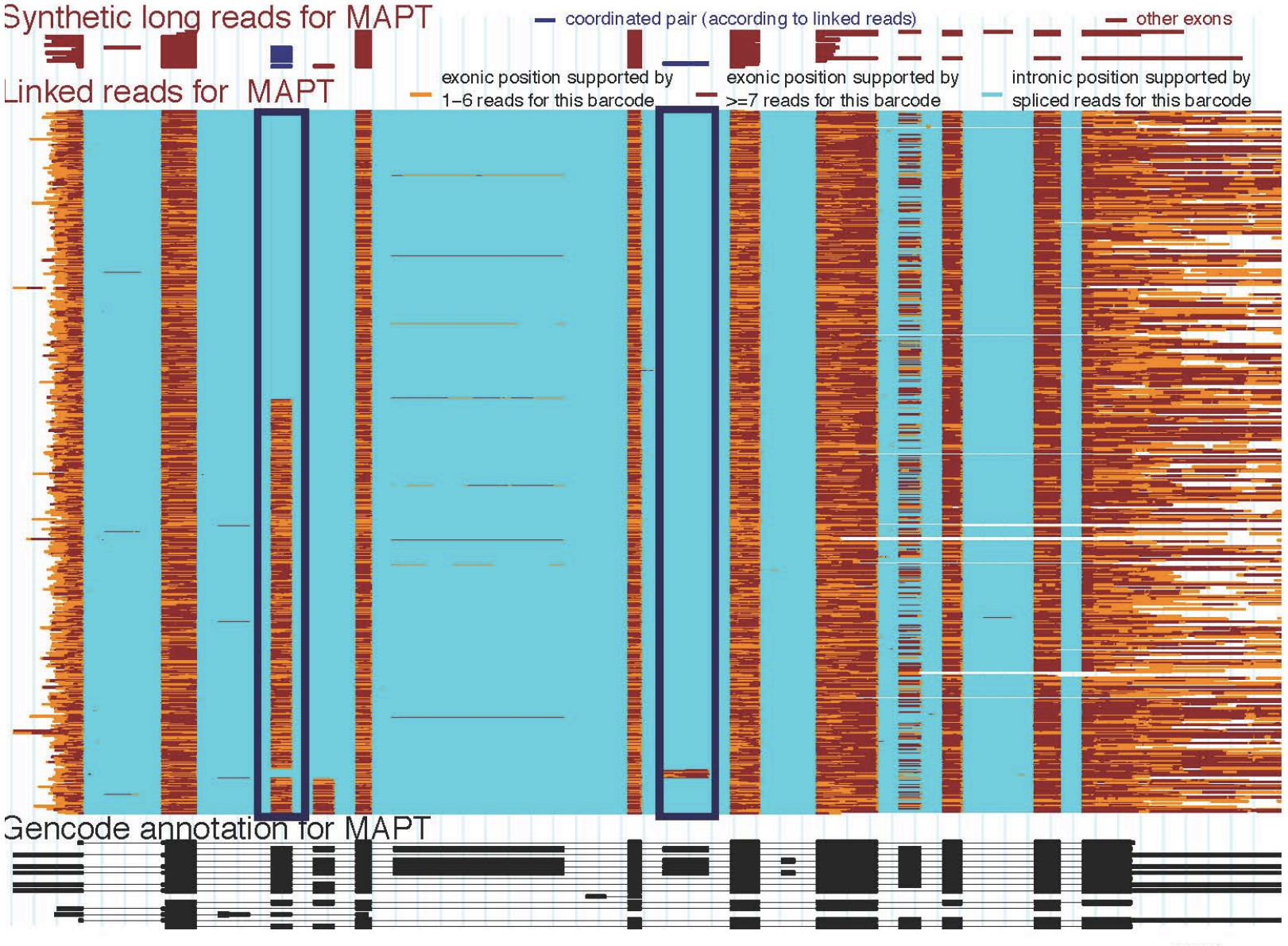




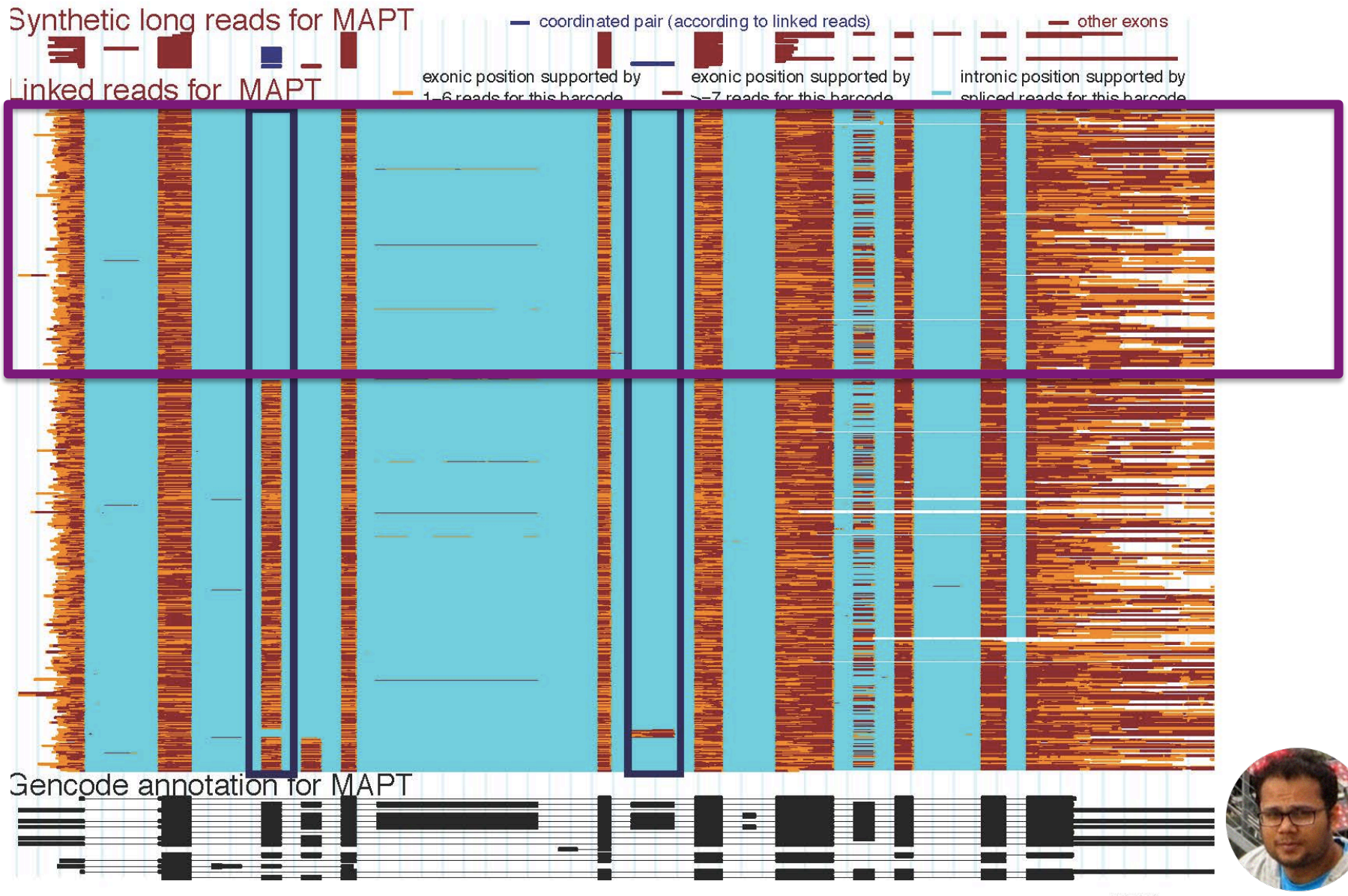
hut2006@med.cornell.edu



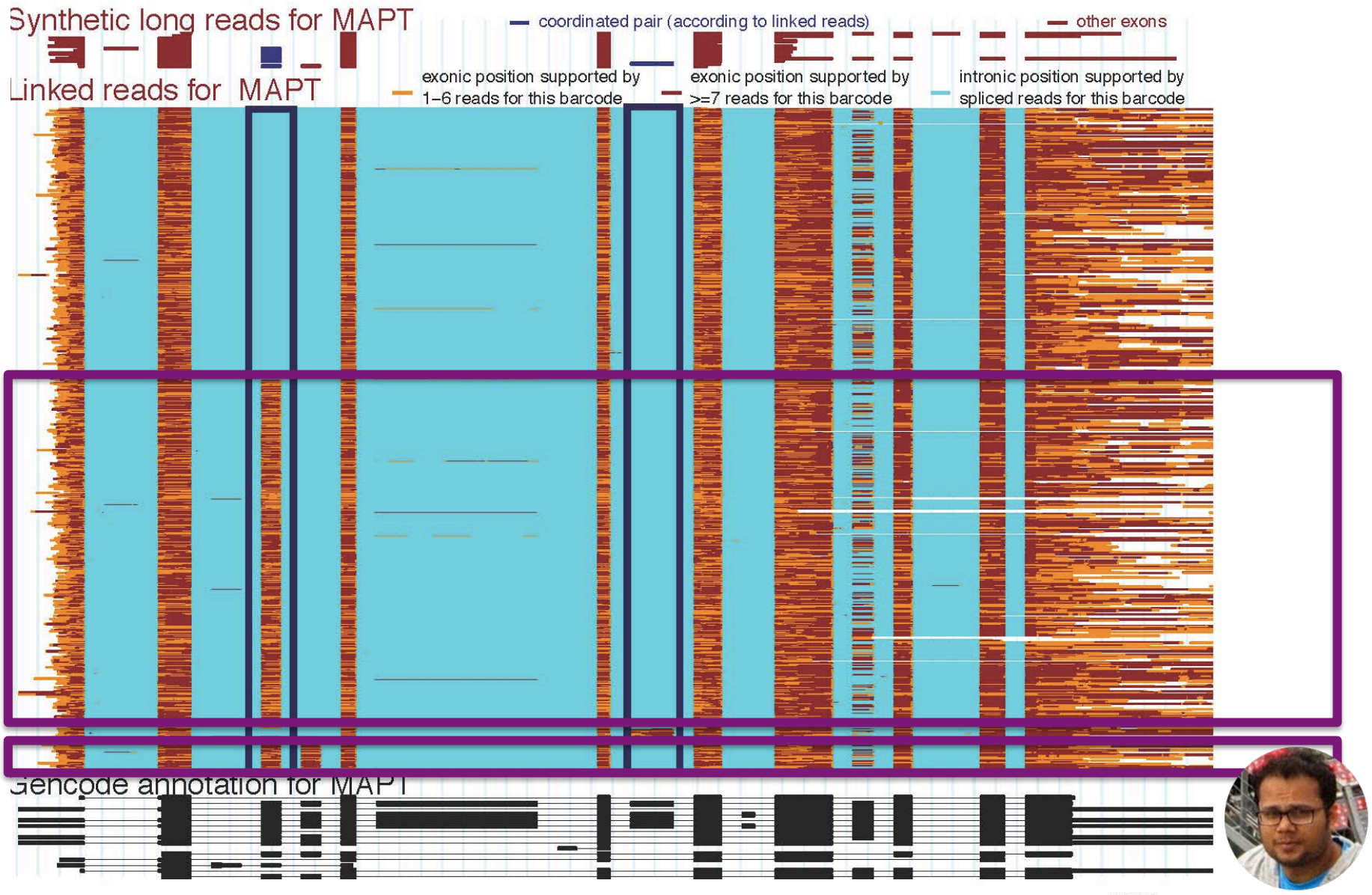
hut2006@med.cornell.edu



hut2006@med.cornell.edu

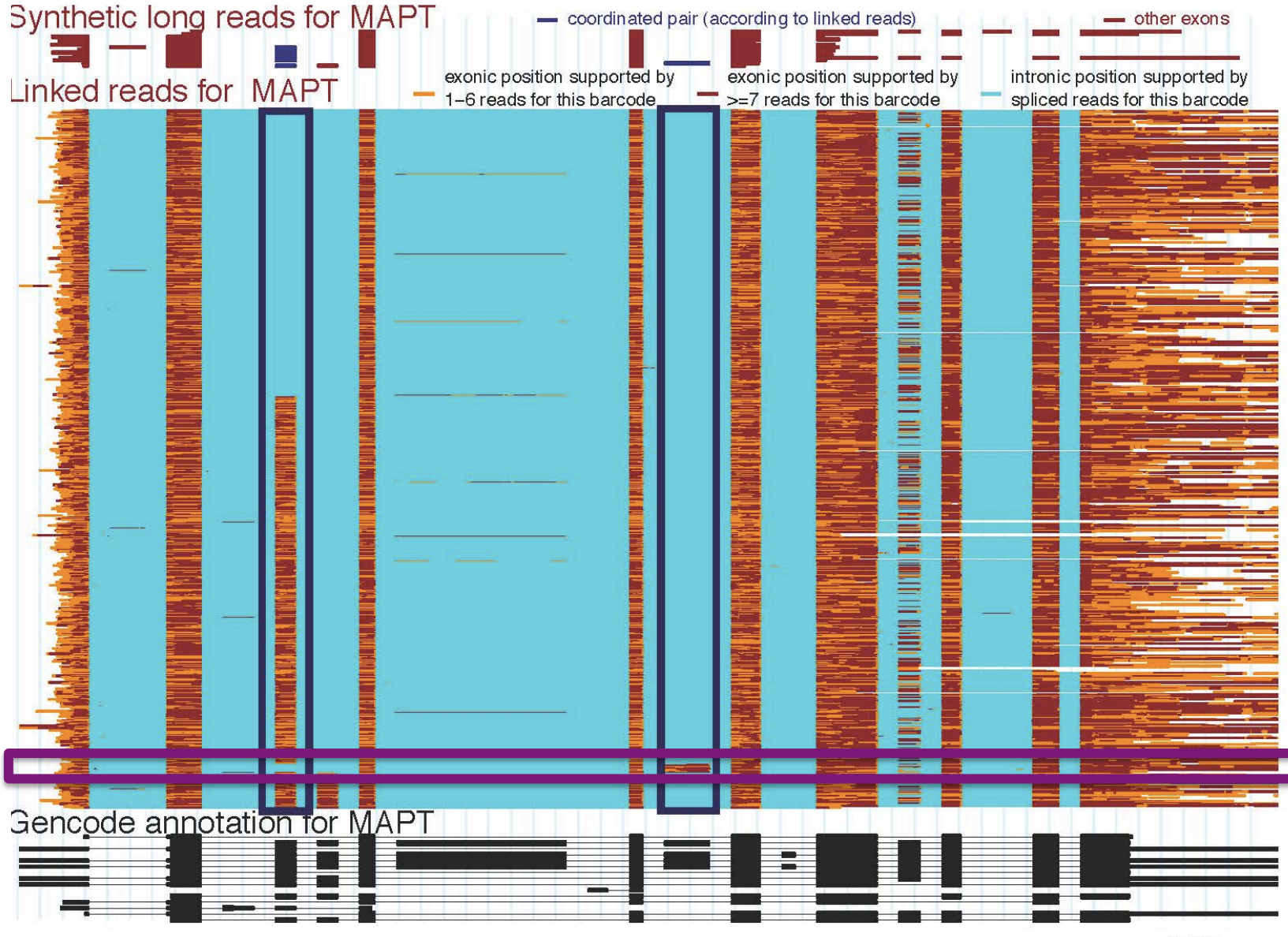


hut2006@med.cornell.edu

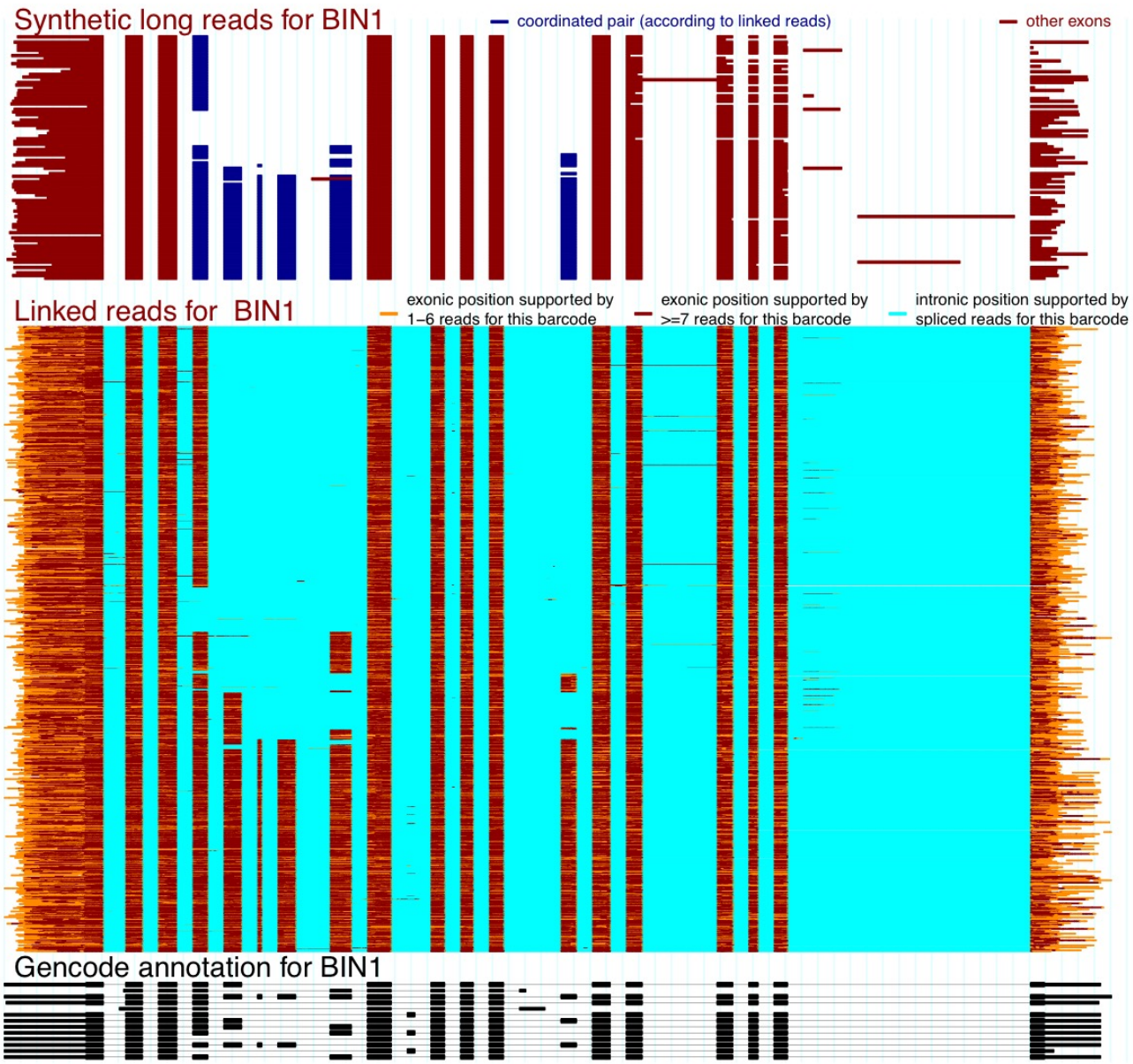


hut2006@med.cornell.edu

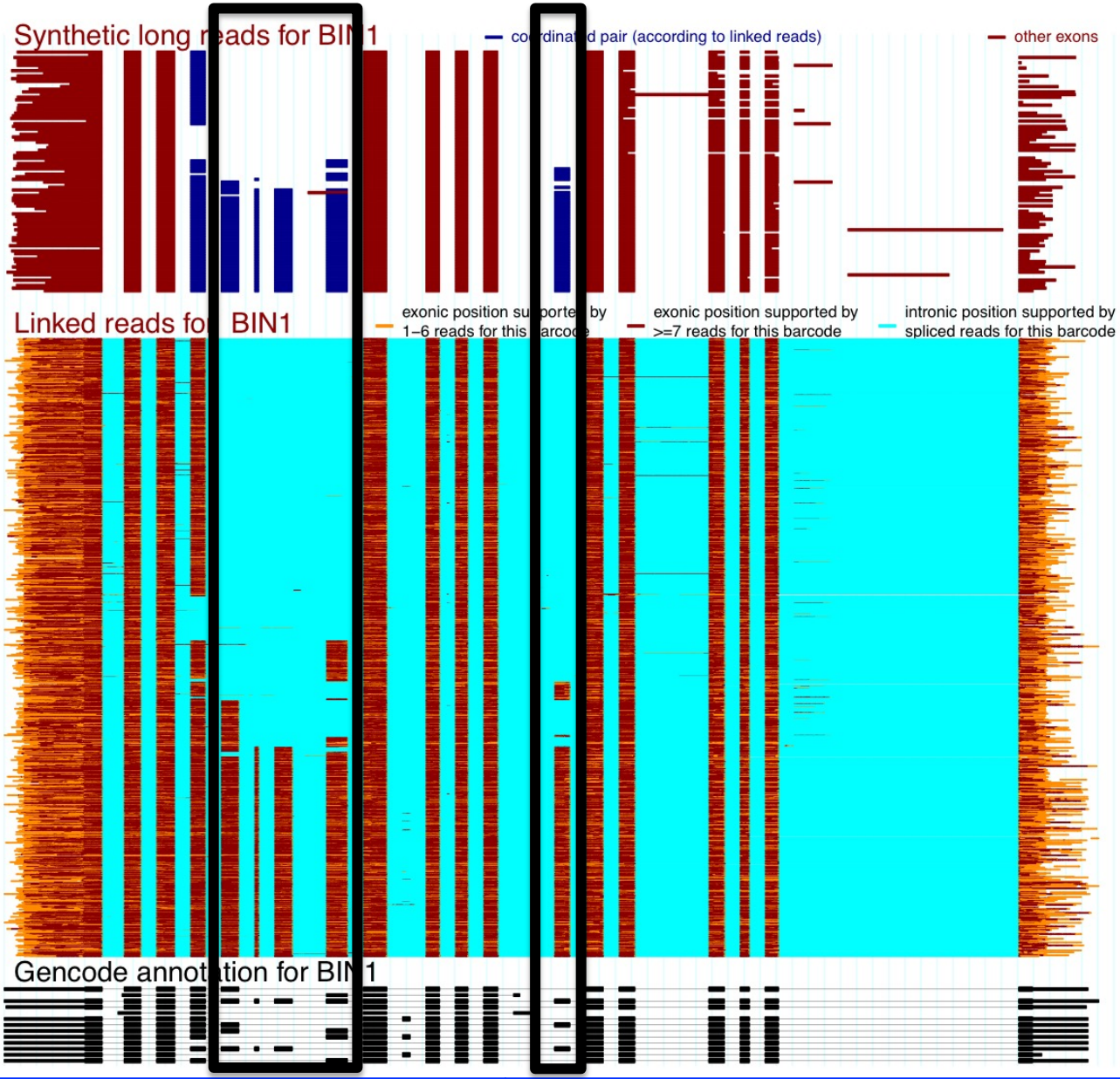
MAPT (tau) splicing is governed by coordination



hut2006@med.cornell.edu

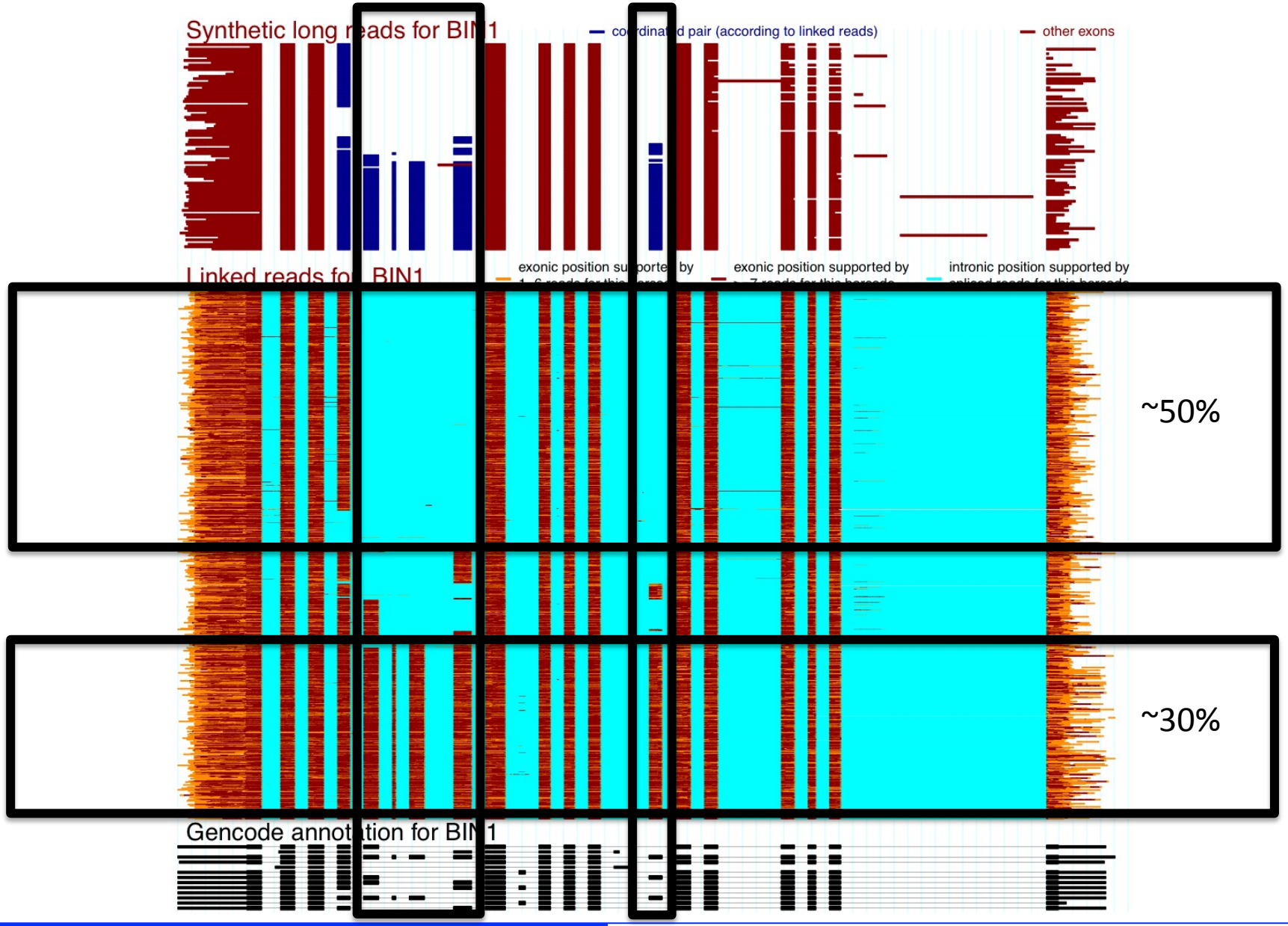


hut2006@med.cornell.edu

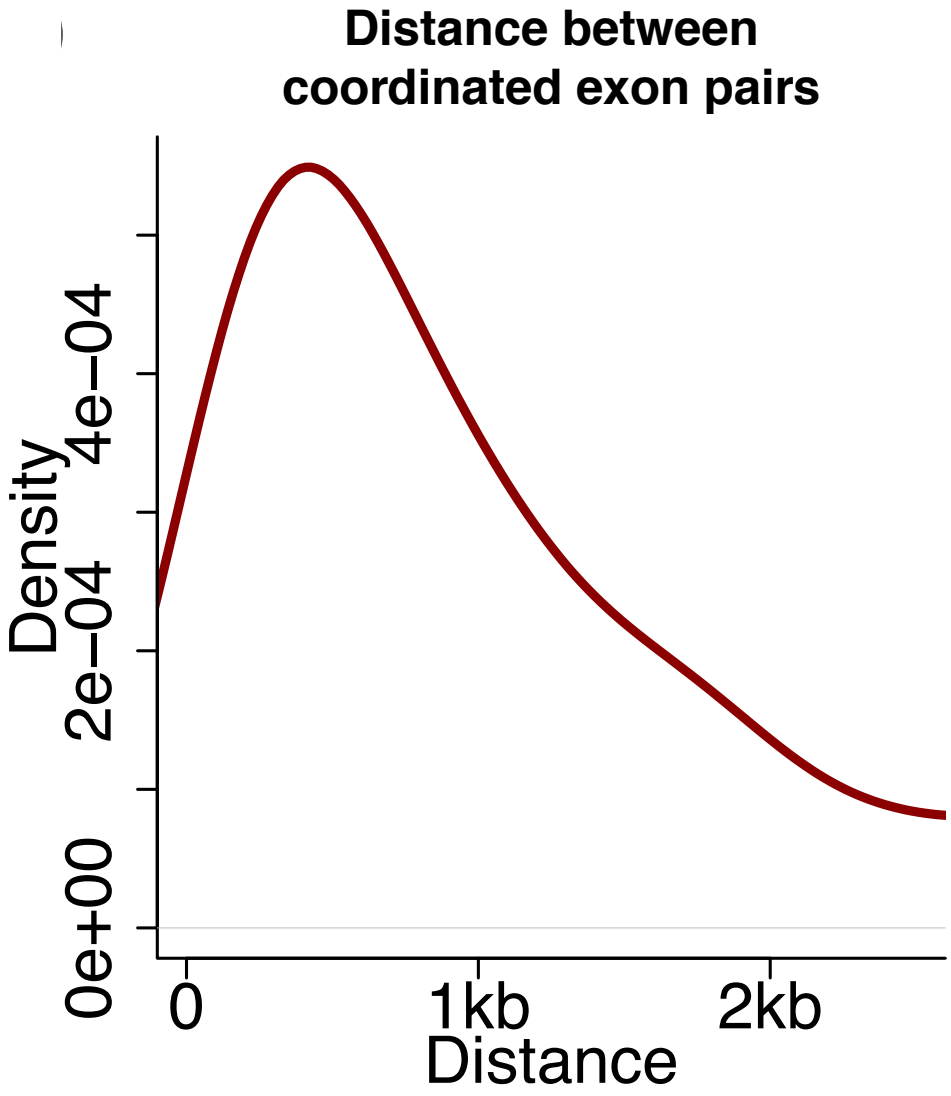


hut2006@med.cornell.edu

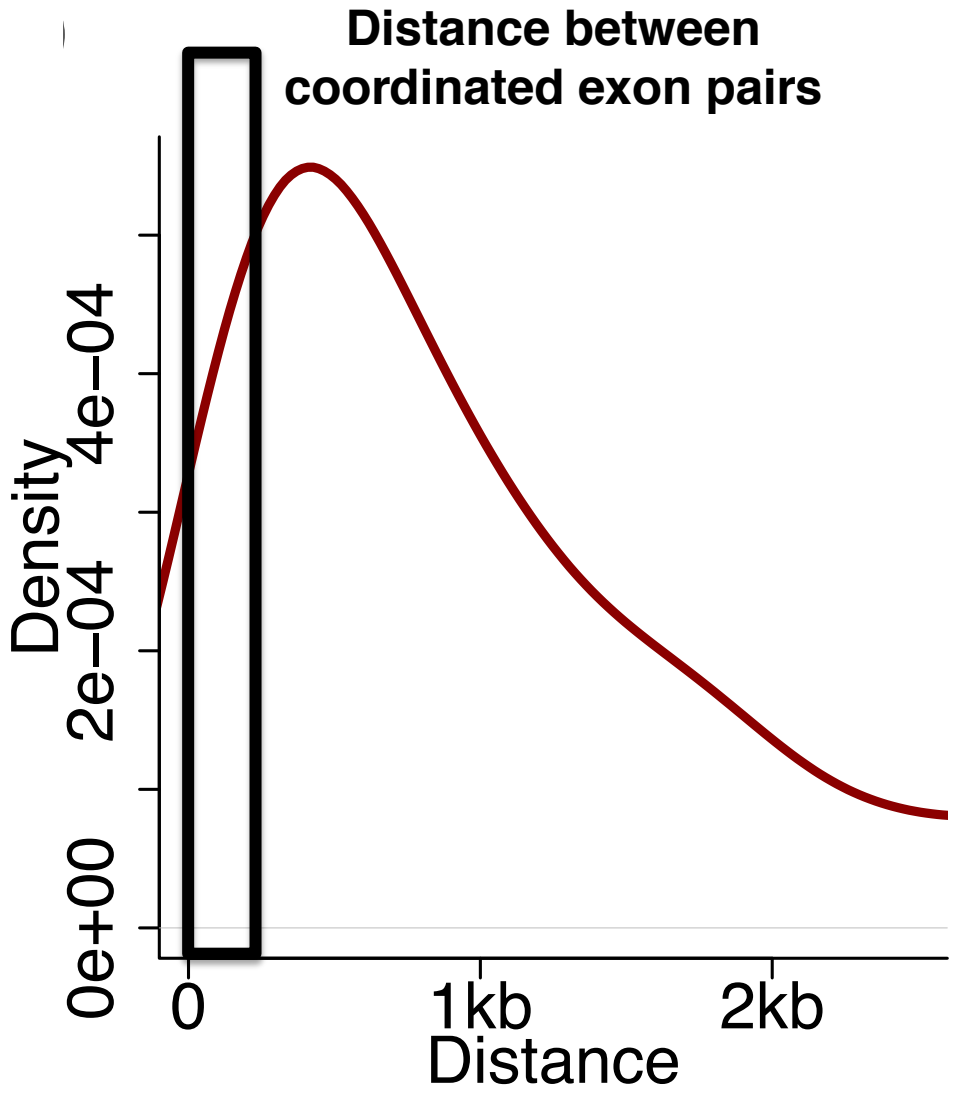
BIN1 (strongly AD associated) splicing is governed by coordination



hut2006@med.cornell.edu

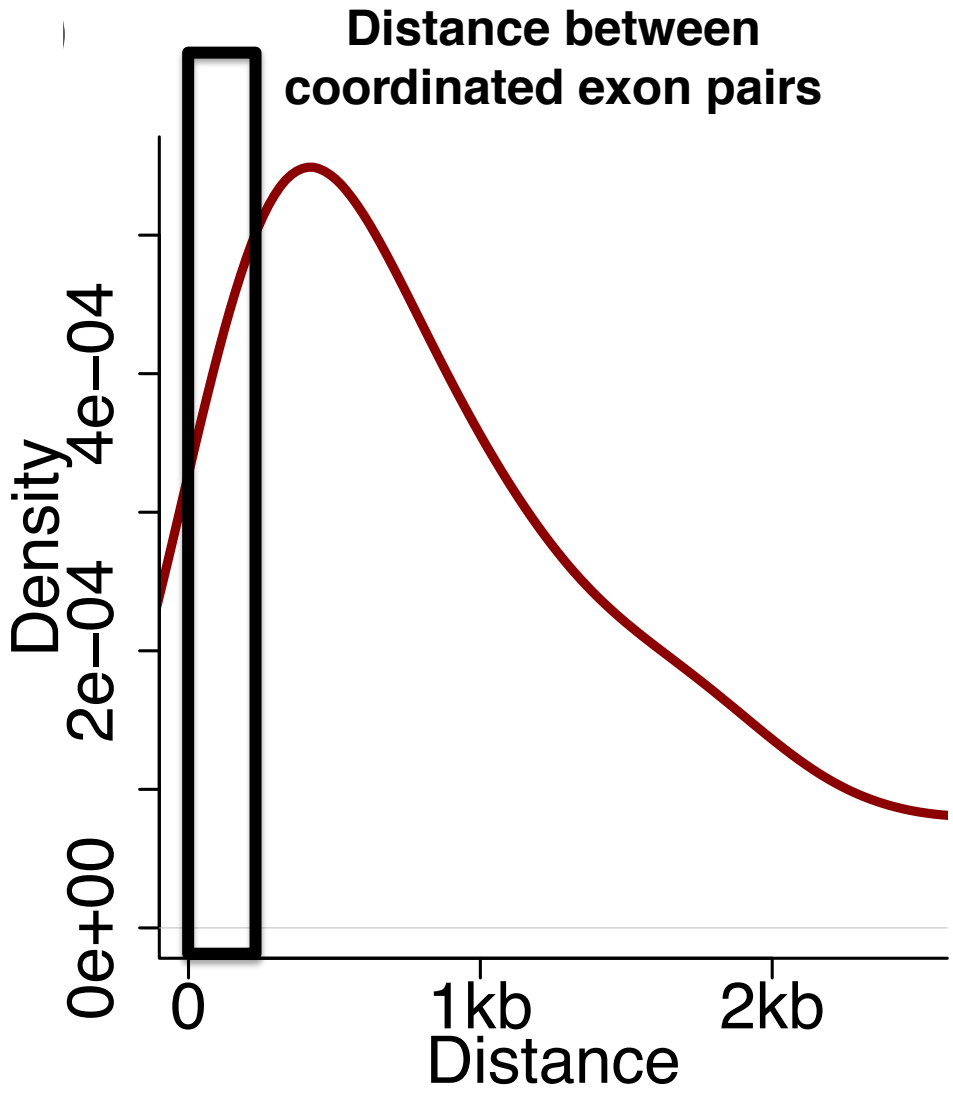


hut2006@med.cornell.edu

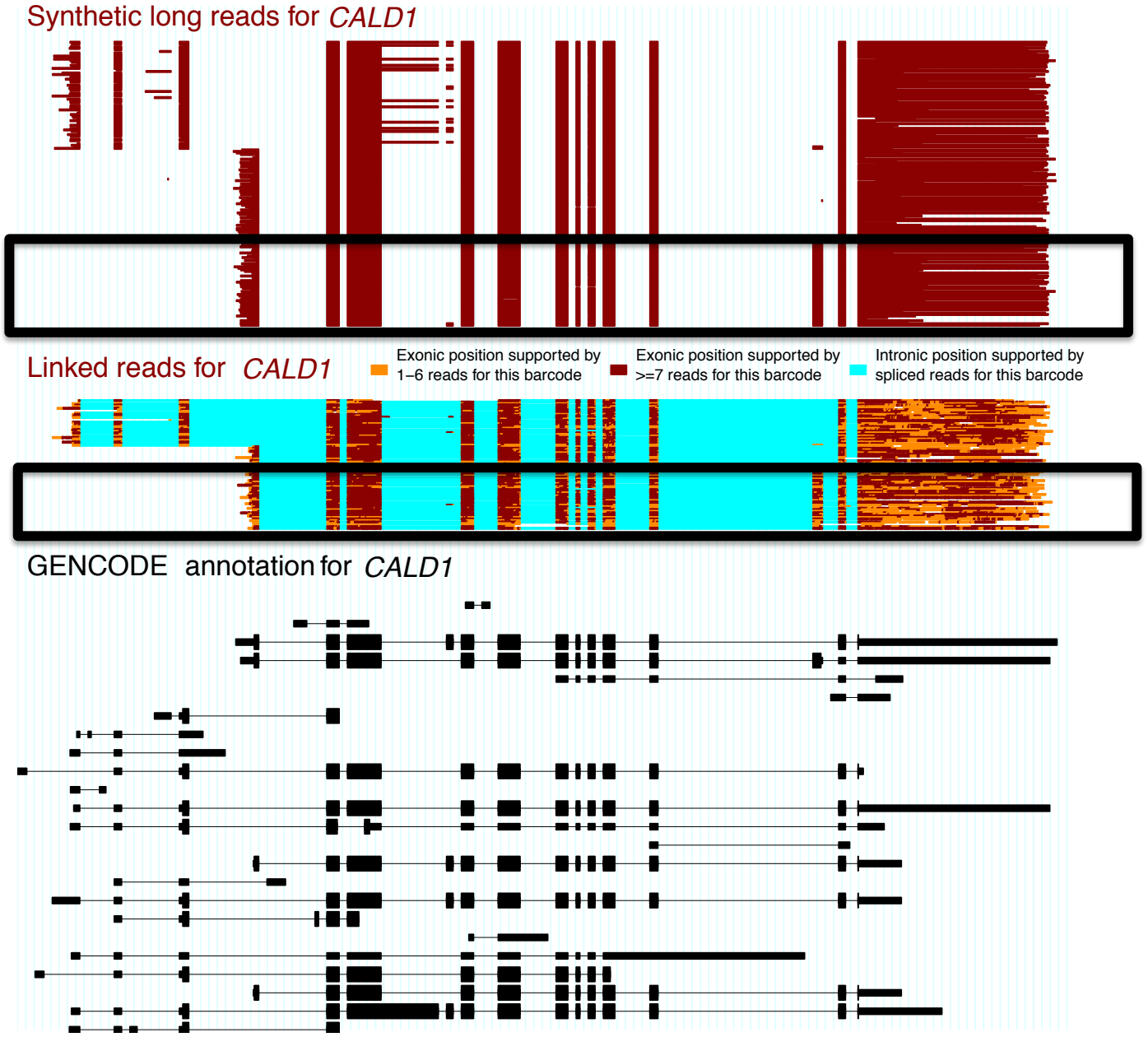


hut2006@med.cornell.edu

Discovery with Illumina -> limited by fragment size (possibly deeper)

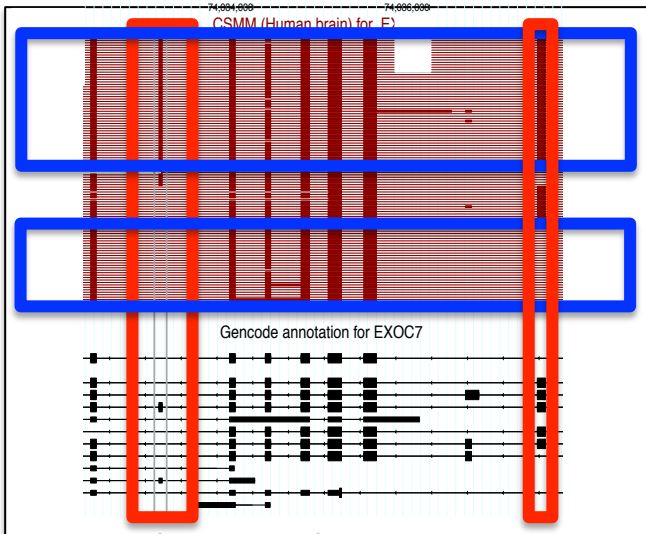


hut2006@med.cornell.edu



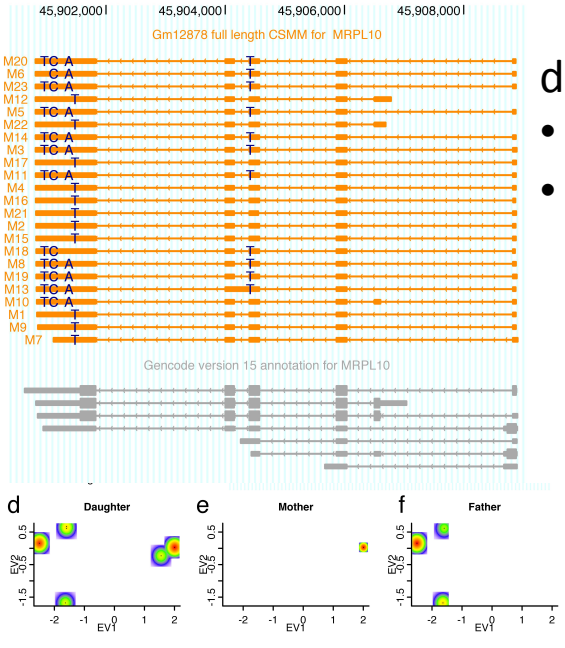
hut2006@med.cornell.edu

Summary



Coordination between Tilgner et al, 2017,2015,2014

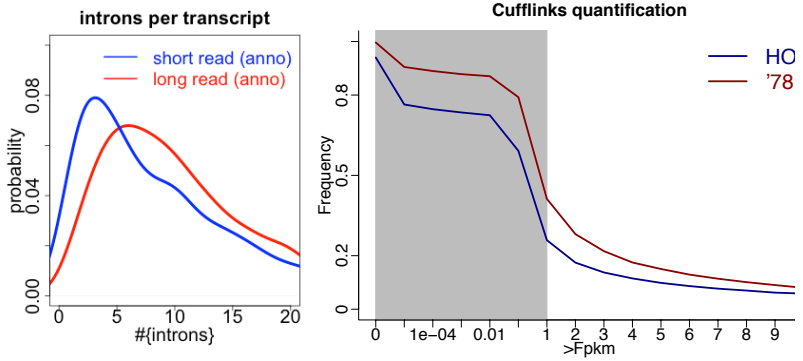
- Distant splicing events and other types
- SNPs and splicing (allele specific splicing)



de novo phasing of

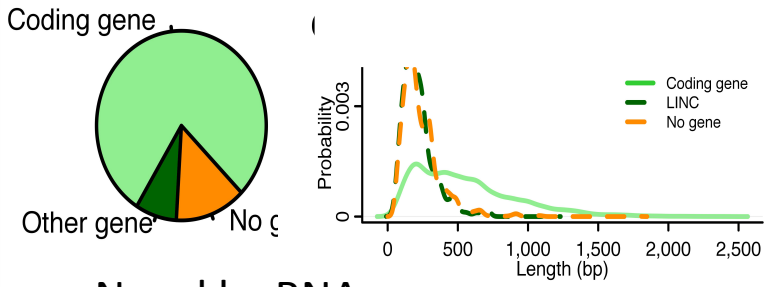
- SNPs
- CNVs

Tilgner et al, 2014



- Enhanced annotations + isoform quant.
- Better isoform discovery from short reads

Tilgner et al, 2013, 2014

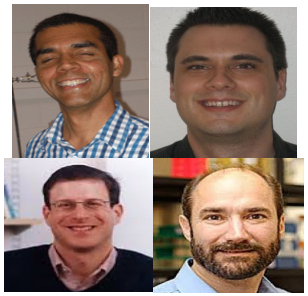


- Novel lncRNA genes
- Loads of novel lncRNA isoforms
- Many protein coding RNA isoforms

Sharon et al, 2013, Tilgner et al, 2013, 2015

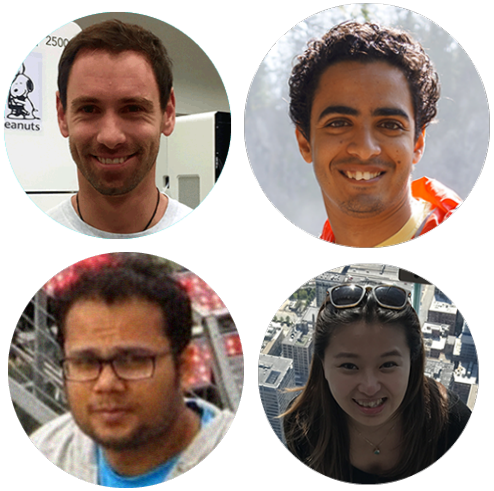
hut2006@med.cornell.edu

You for listening



K562 + Hela

Human organs + lymphoblastoid



SLRs

Noah Spies
Cheryl Smith
Dena Leeman
Anne Brunet
Ben Barres
Steven Sloan
Ye Zhang
Eli Calo

Betsy Ross
Costantino Iadecola
Songhai Shi
Greg Petsko
Olivier Elemento
Howard Fine
Wenjie Luo
Chris Mason

Paul Collier
Ahmed Mahfouz
Ishaan Gupta
So Yeon Koo
Amanda Buch
Dmitrii Meleshko

Nicole Rapicavoli
Nick Sisneros
Sajani Swamy
Kan Nobuta
Roshni Bhattacharya
Andreas Prlic

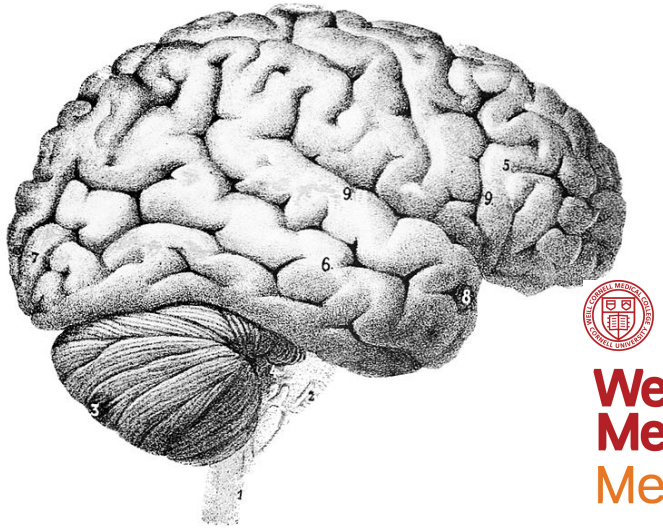
Luisa Lente

Funding:

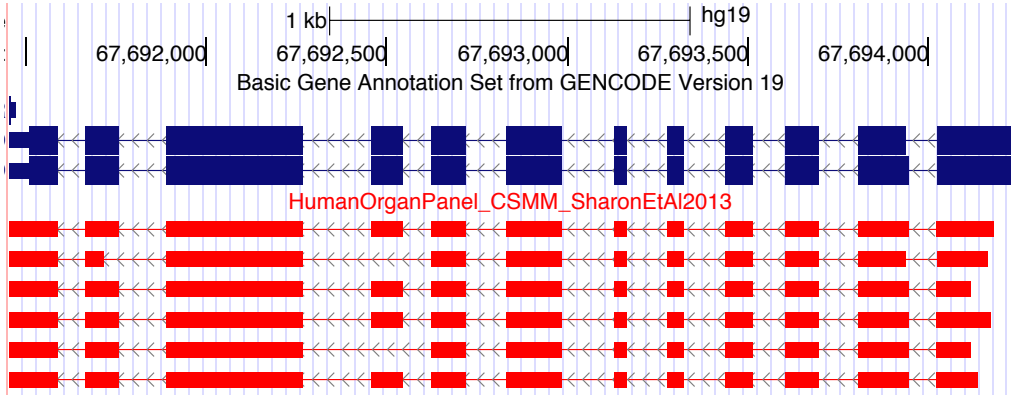
Feil Family

LEON LEVY FOUNDATION
THE LEGACY OF LEON LEVY

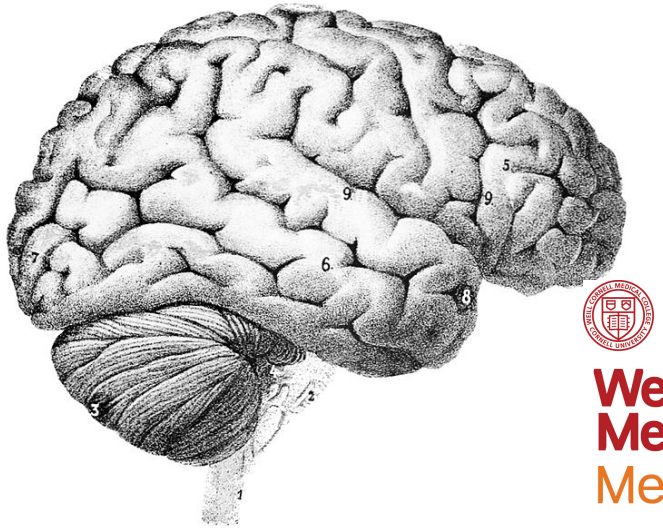
hut2006@med.cornell.edu



Weill Cornell Medicine
Medical College



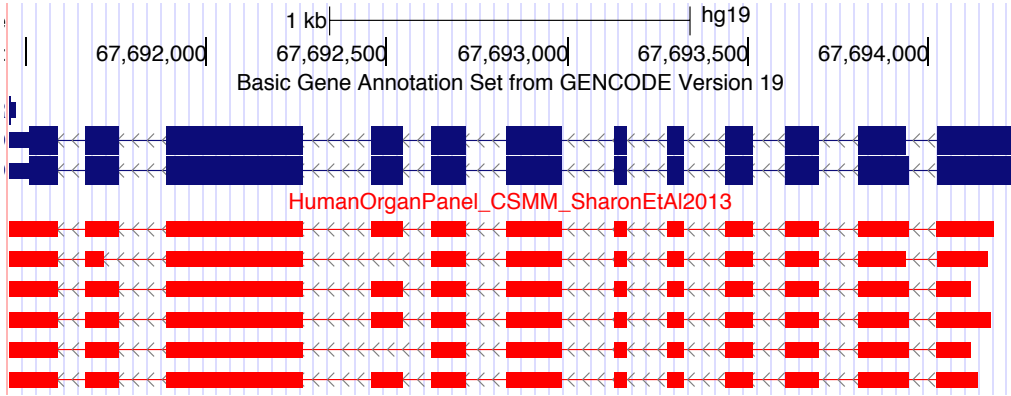
FUT-1 hut2006@med.cornell.edu



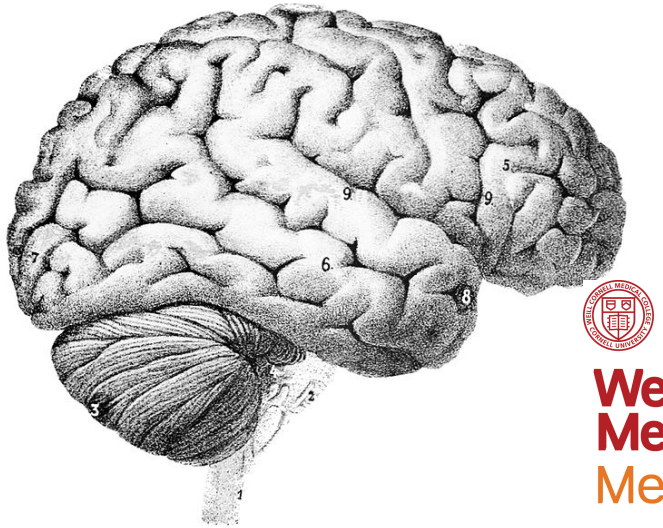
Weill Cornell Medicine
Medical College

Looking for postdocs

- Algorithms/software
- Data analysis



FUT-1 hut2006@med.cornell.edu

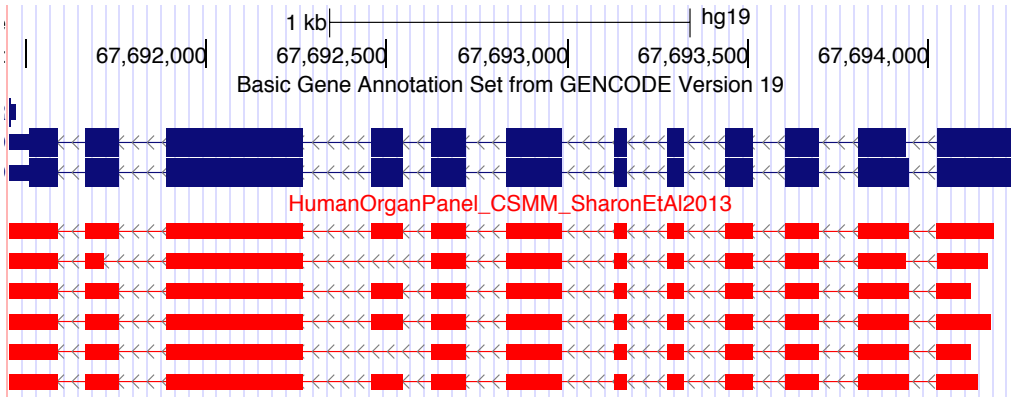


Weill Cornell Medicine
Medical College

Looking for postdocs

- Algorithms/software
- Data analysis

Collaborations welcome



FUT-1 hut2006@med.cornell.edu