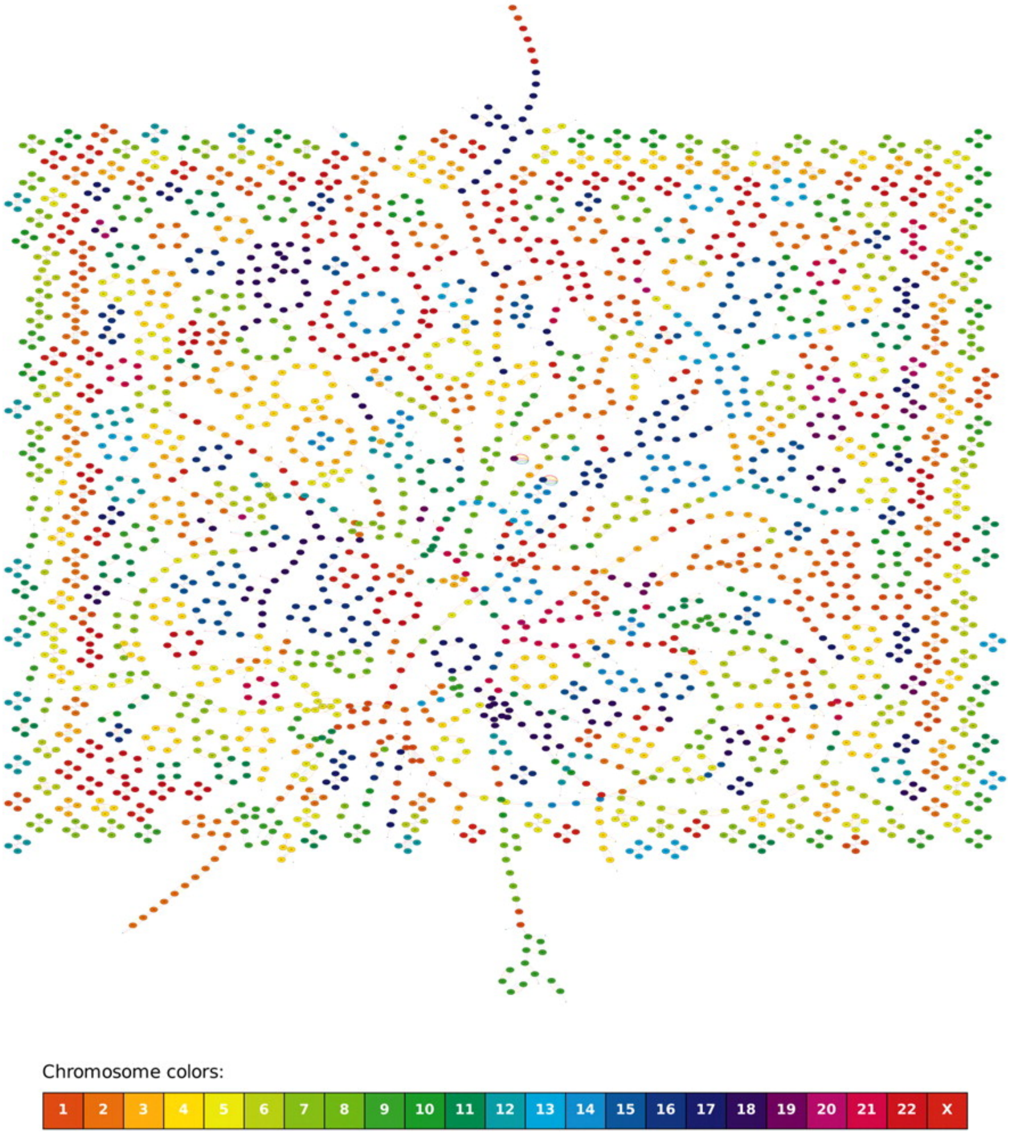
http://zoecity.net/groups/497/images/71946

CREATING INFORMATION-RICH, INFORMATIVE AND APPEALING GENOME DATA GRAPHICS

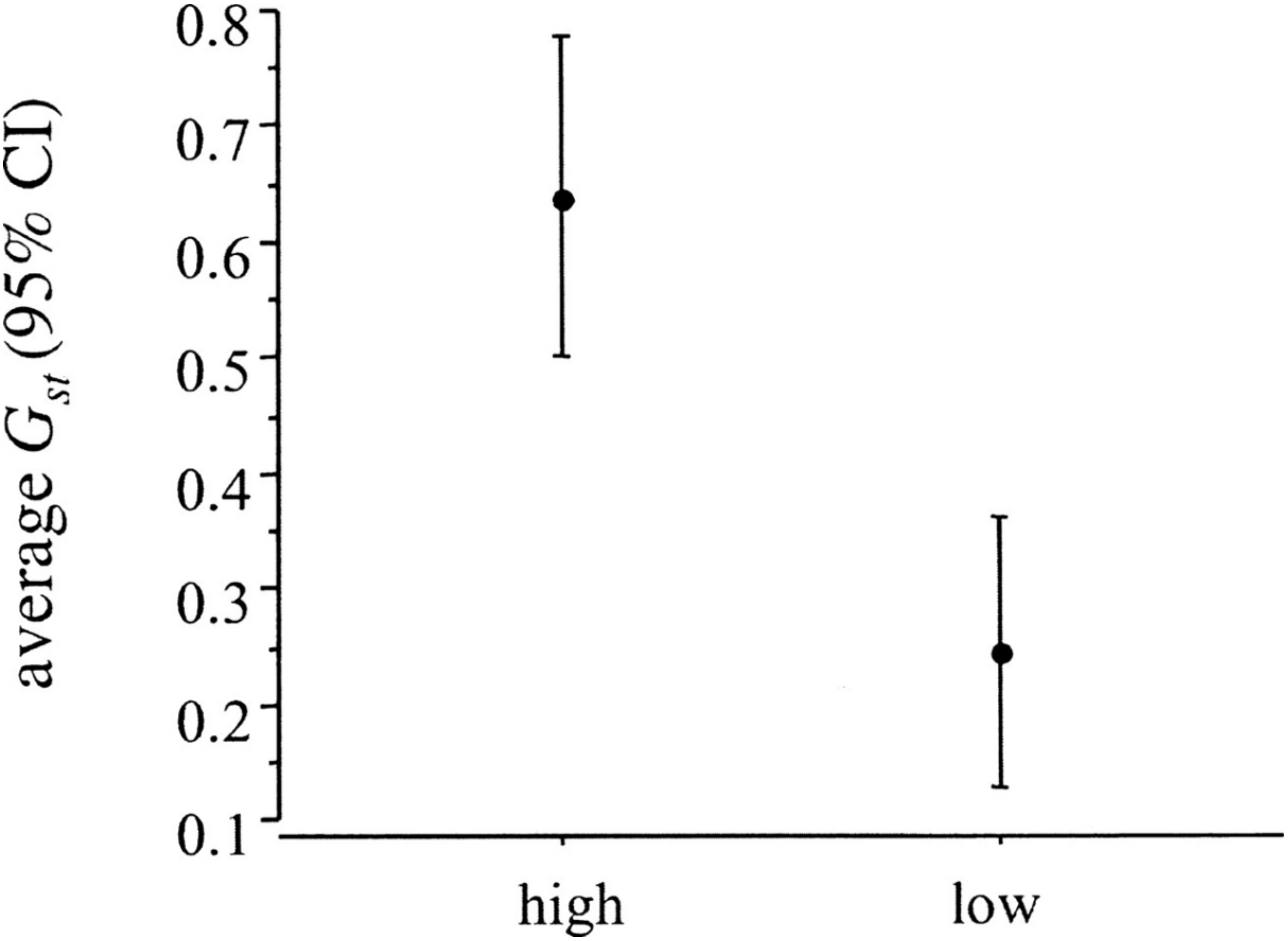
VISUALIZATION OF GENOMIC DATA

martin krzywinski
genome sciences center
BC CANCER RESEARCH CENTER
vancouver, canada



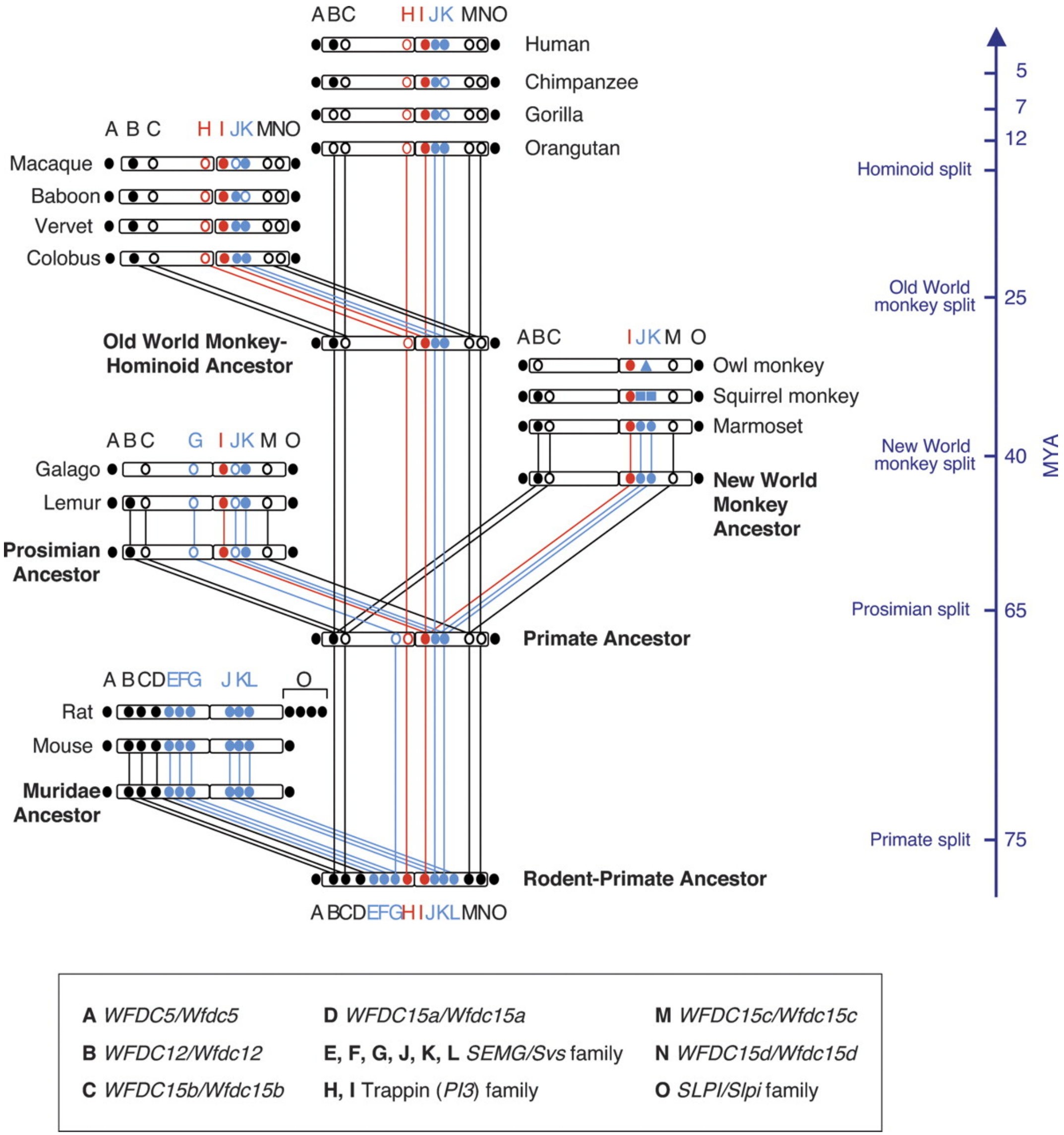
THE BREAKSPOINT GRAPH $G(M,R,D,Q,H,C)$ (OBVERSE EDGES ARE NOT SHOWN) OF SIX MAMMALIAN GENOMES.

Alekseyev, M.A. and P.A. Pevzner, Breakpoint graphs and ancestral genome reconstructions. Genome Res, 2009. 19(5): p. 943-57.



MEAN AND 95% CONFIDENCE INTERVAL OF AVERAGE G_{ST} BASED ON DIRECT SEQUENCING OF FRAGMENTS WITHIN HIGH (N = 8) AND LOW (N = 6) DIFFERENTIATION REGIONS IN NATURAL POPULATION OF HOUSE MICE.

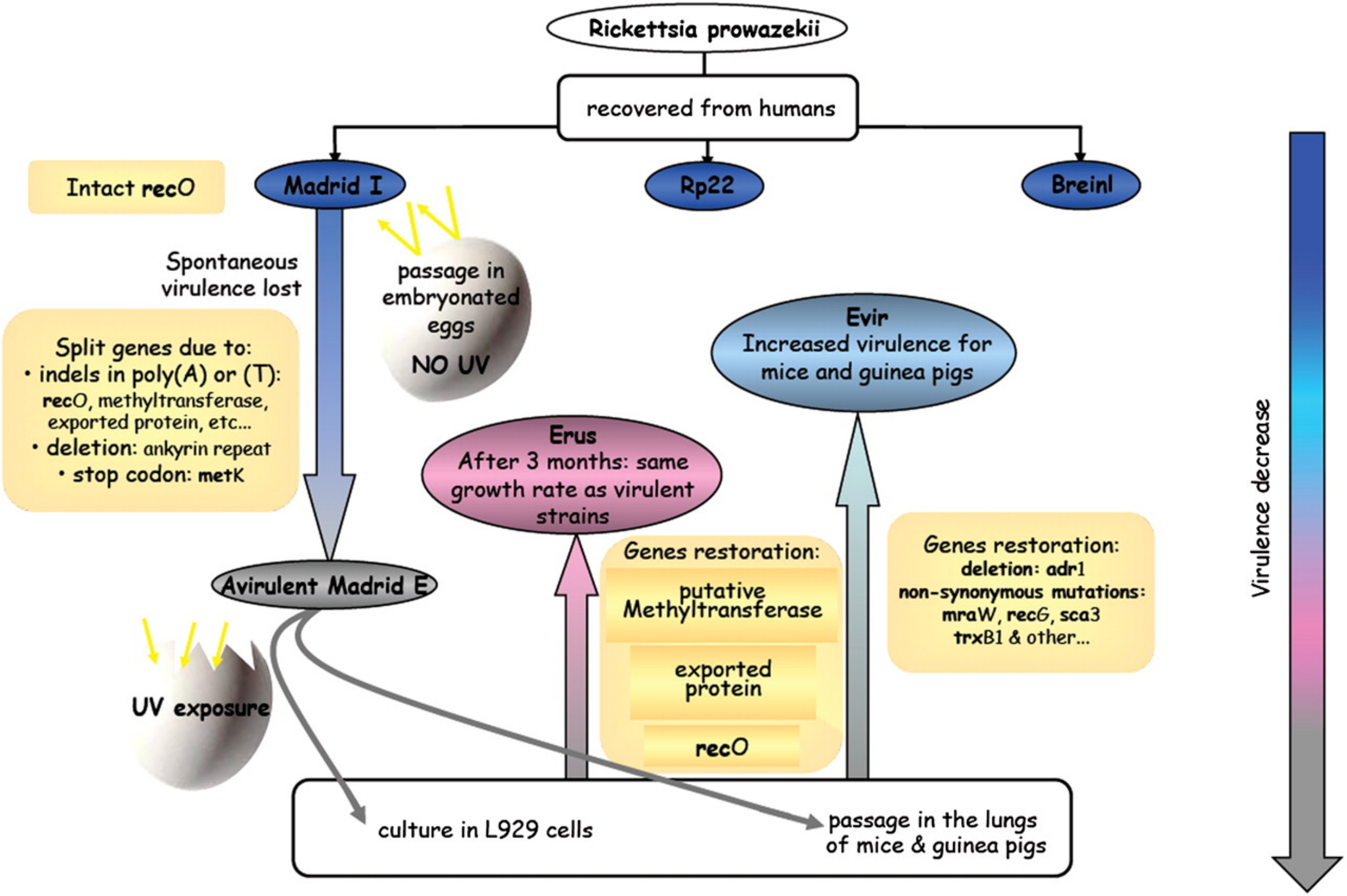
Harr, B., Genomic islands of differentiation between house mouse subspecies. Genome Res, 2006. 16(6): p. 730-7.



A *WFDC5/Wfdc5*	**D** *WFDC15a/Wfdc15a*	**M** *WFDC15c/Wfdc15c*
B *WFDC12/Wfdc12*	**E, F, G, J, K, L** *SEMG/Svs* family	**N** *WFDC15d/Wfdc15d*
C *WFDC15b/Wfdc15b*	**H, I** Trappin (*PI3*) family	**O** *SLPI/Slpi* family

EVOLUTIONARY HISTORY OF THE CENTROMERIC WFDC SUBLOCUS.

Hurle, B., W. Swanson, and E.D. Green, Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. Genome Res, 2007. 17(3): p. 276-86.



Rickettsia prowazekii

recovered from humans

Madrid I

Rp22

Breinl

Intact **rec**O

Spontaneous virulence lost

passage in embryonated eggs

NO UV

Split genes due to:
- indels in poly(A) or (T): **rec**O, methyltransferase, exported protein, etc...
- deletion: ankyrin repeat
- stop codon: **met**K

Avirulent Madrid E

UV exposure

Erus
After 3 months: same growth rate as virulent strains

Evir
Increased virulence for mice and guinea pigs

Genes restoration:
putative Methyltransferase
exported protein
recO

Genes restoration:
deletion: adr1
non-synonymous mutations:
mraW, **rec**G, **sca**3
trxB1 & other...

culture in L929 cells

passage in the lungs of mice & guinea pigs

Virulence decrease

SCHEME OF *R. PROWAZEKII* STRAINS ORIGIN AND EVOLUTION.

Bechah, Y., et al., *Genomic, proteomic, and transcriptomic analysis of virulent and avirulent Rickettsia prowazekii reveals its adaptive mutation capabilities. Genome Res, 2010.* ***20(5): p. 655-63.***

Is a graphical representation really necessary?

Does the legend obviate the figure?

What is my message?

Does my figure communicate it clearly?

Are there extraneous or ornamental elements?

What can I remove without changing the overall story?

Have I left the reader wanting more, or less?

The reader does not know what they need to know.

You must tell them.

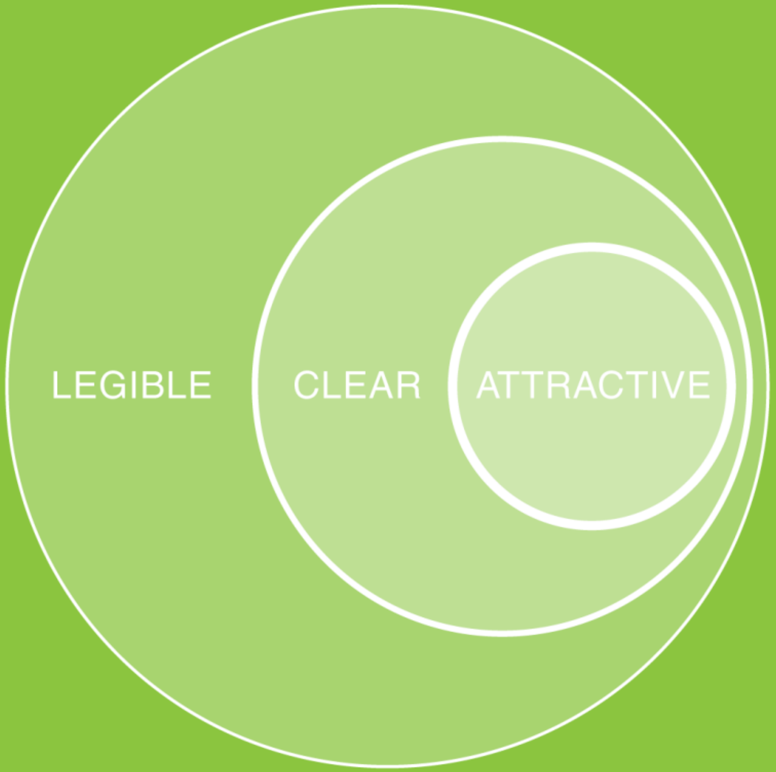
The reader does not know what is important.

You must show them.

The reader's cognitive and visual acuity are limited.

The reader may have time for a good presentation.

The reader never has time for a bad presentation.



LEGIBLE

CLEAR

ATTRACTIVE

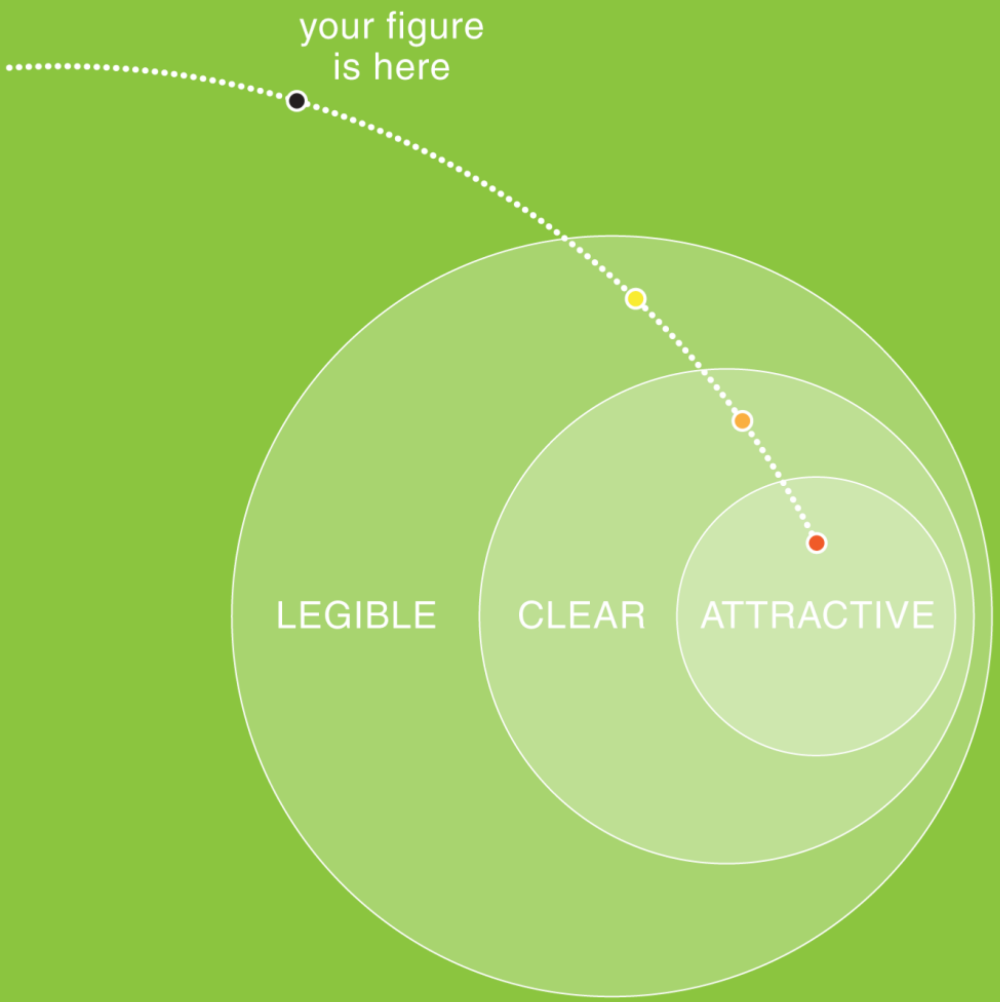
quality of communication

- ○ GOOD
- ○ BETTER
- ○ BEST

LACK OF CLEAR MESSAGE

REDUNDANCY

http://blog.ivman.com/ironic-funny-signs



next 60 minutes

- 🟡 **part 1**
  MAKING IT LEGIBLE
- 🟠 **part 2**
  MAKING IT CLEAR
- 🔴 **part 3**
  MAKING IT PRETTY

next

● **preamble**
INFOHORROR

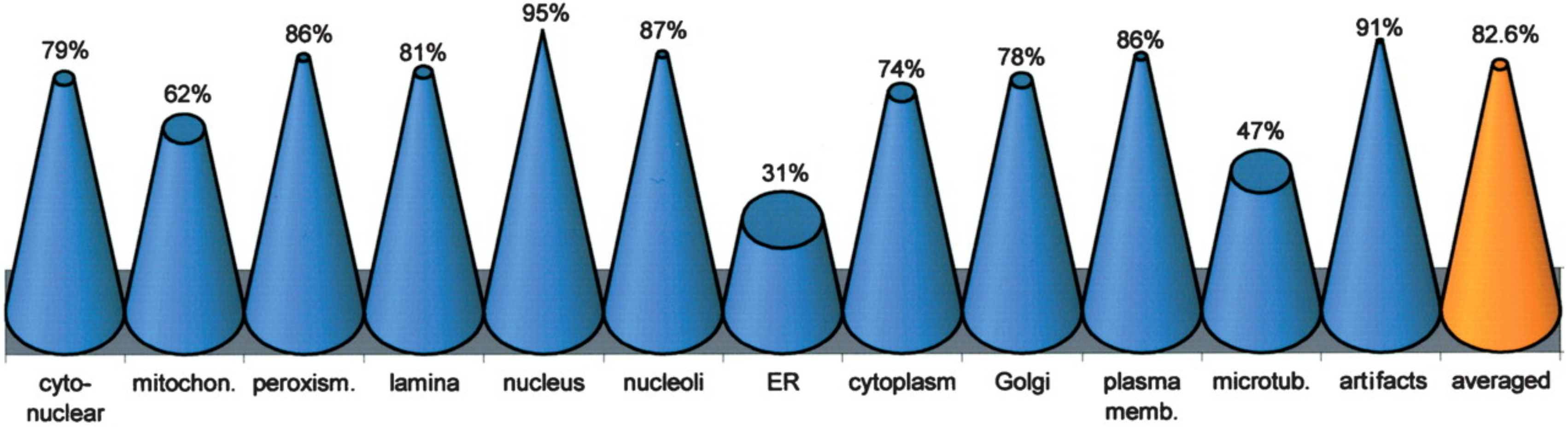
A

	ANN/GA [1]	BayesANN [2]	SVM
MI (25 features)	76.3%	71.1% ± 0.08%	74.9%
SAM (25 features)	73.3%	80.5% ± 0.009%	78.2%
STEPWISE (25 features)	79.4%	80.2% ± 1.10%	82.2%
all features	-	-	79.2%

[1] best accuracy of 20 individuals after 20 generations in the Genetic Algorithm

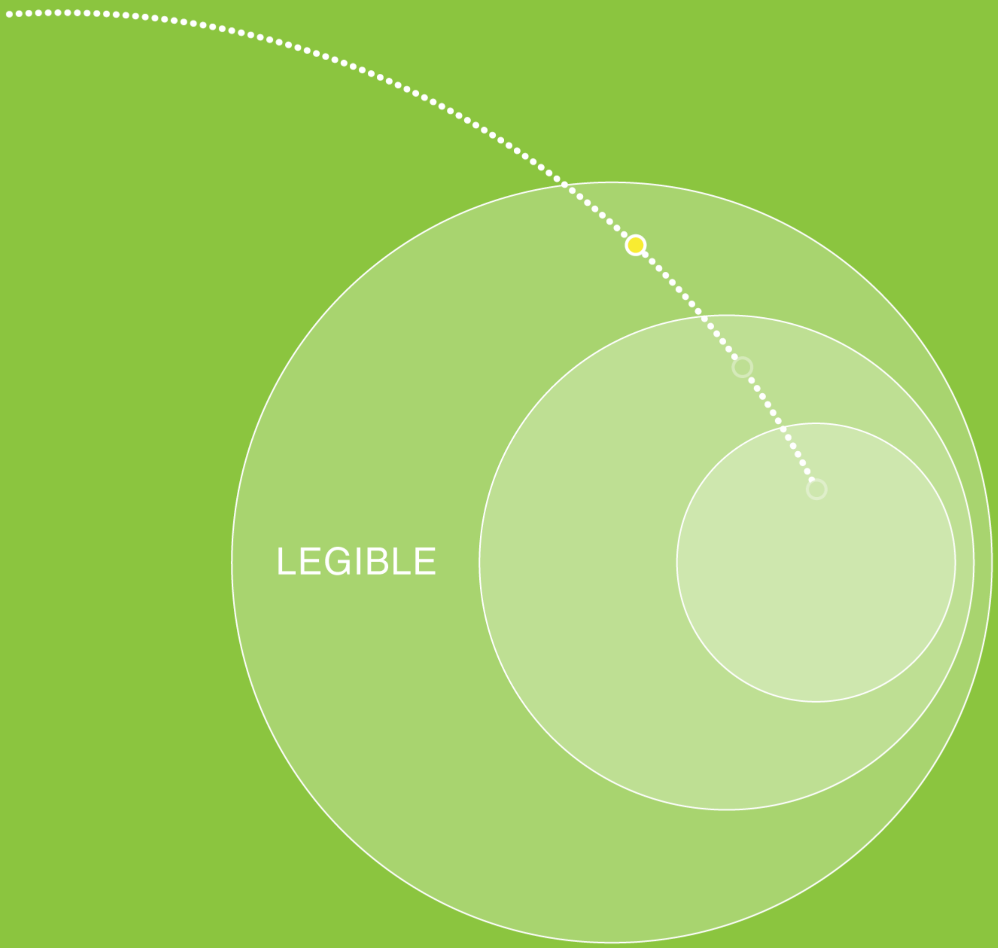
[2] averaged over 10 classifiers

B



ACCURACY OF PHENOTYPE CLASSIFICATION.

Conrad, C., et al., *Automatic identification of subcellular phenotypes on human cell arrays. Genome Res, 2004.* ***14(6): p. 1130-6.***

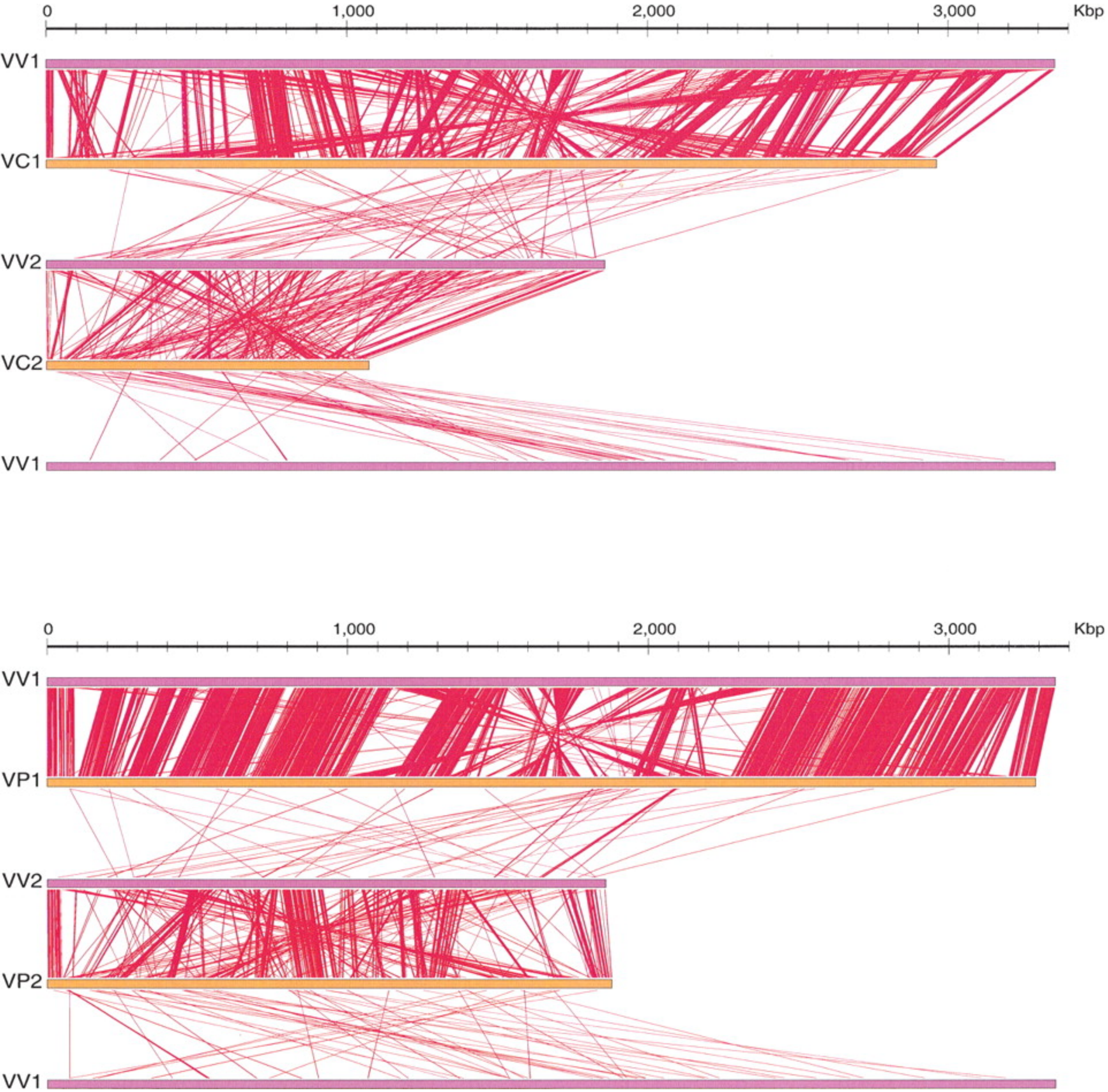


next

- ● **part 1**
  MAKING IT LEGIBLE
- ○ **part 2**
  MAKING IT CLEAR
- ○ **part 3**
  MAKING IT PRETTY

part 1
● MAKING IT LEGIBLE

PARSABILITY



Help the reader identify meaningful patterns.

If no such patterns exist, the figure is not necessary.

INTRA- AND INTERCHROMOSOMAL SHUFFLING OF VIBRIO GENES.

Chen, C.Y., et al., Comparative genome analysis of Vibrio vulnificus, a marine pathogen. Genome Res, 2003. 13(12): p. 2577-87.

A figure should reduce complexity for the reader.

There is no *emergent pattern* here.

This figure is not parsable.

THE SPECIFIC DOMAINS AND COMBINATIONS IN THE DOMAIN GRAPH.

Ye, Y. and A. Godzik, Comparative analysis of protein domain organization. Genome Res, 2004. 14(3): p. 343-53.


ISSN 1088-9051

October 2004

GENOME RESEARCH

Volume 14 Number 10a

microRNA Host Genes and Transcription Units • Sequence Variation Across Environmental Response Genes • Tailless Retropseudogenes • Human, Mouse, and Rat Genome Large-Scale Rearrangements • Maize Genome Insights

Cold Spring Harbor Laboratory Press

An *emergent pattern* in raw data.

To see the pattern, the entire data set must be shown.

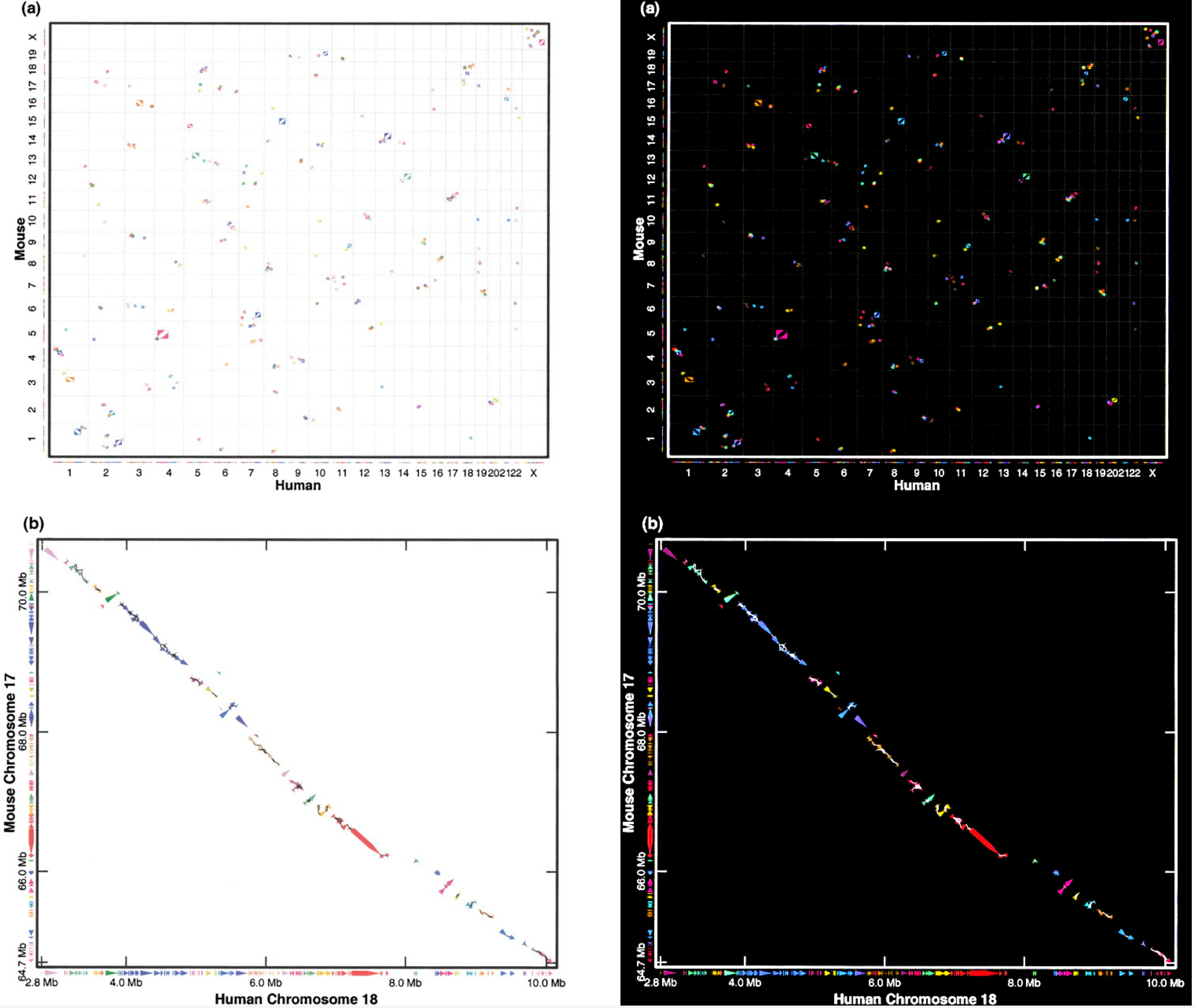
THE COVER DESIGN DEPICTS HAIRPIN SECONDARY STRUCTURE OF RNA SEQUENCES, PRECURSORS TO MICRORNA.

Genome Res, 2004. 14(10a): cover.

part 1

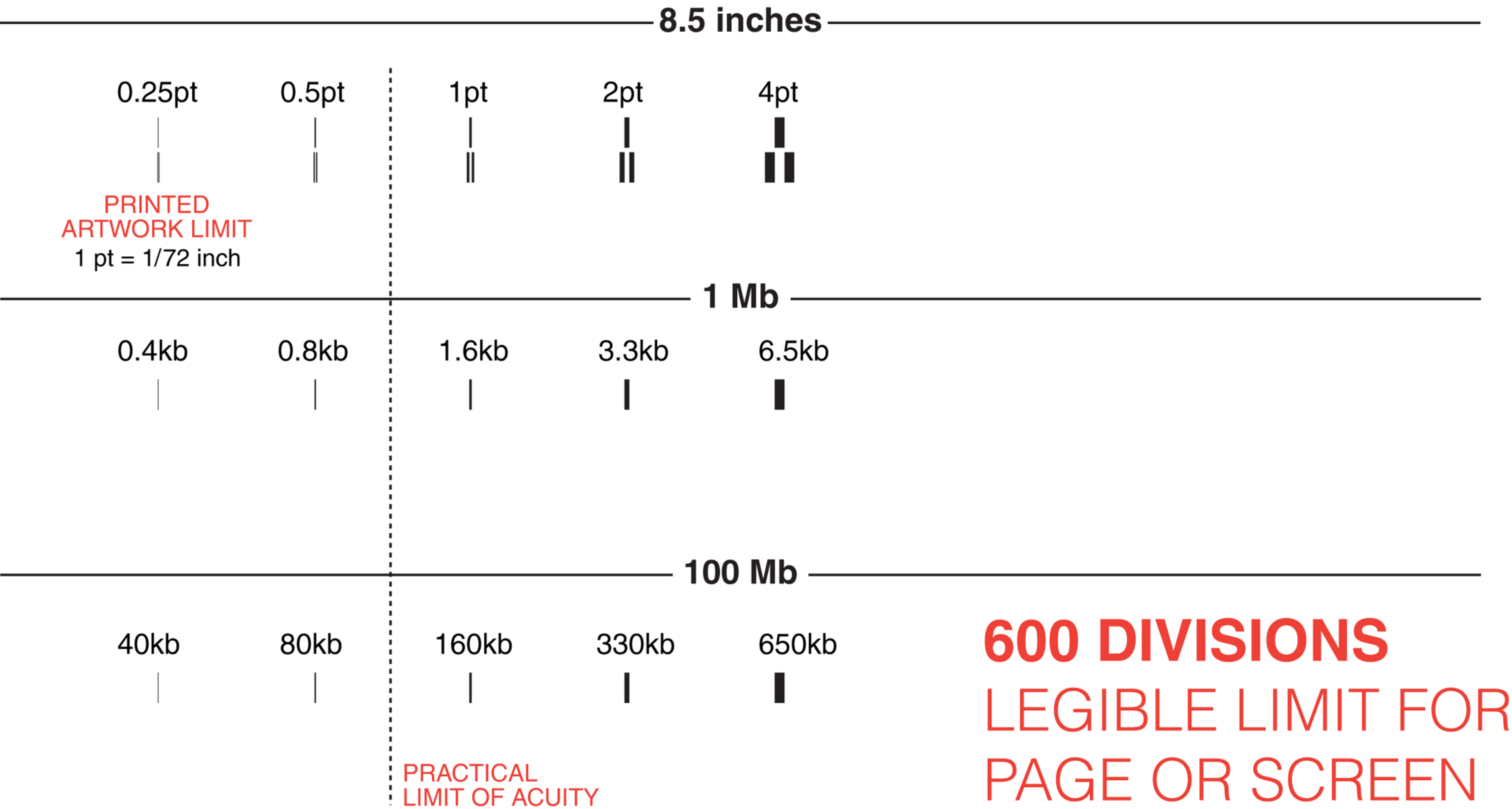
MAKING IT LEGIBLE

RESOLUTION



HUMAN AND MOUSE SYNTENY BLOCKS.
COMBINING ANCHORS INTO CLUSTERS BY THE GRIMM-SYNTENY ALGORITHM AT G=100 KB.

Pevzner, P. and G. Tesler, Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. Genome Res, 2003. 13(1): p. 37-45.



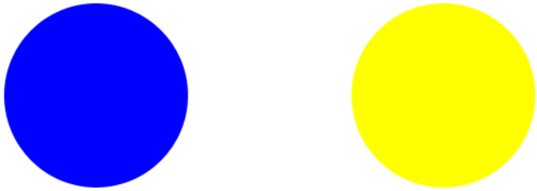
PRINTED
ARTWORK LIMIT
1 pt = 1/72 inch

PRACTICAL
LIMIT OF ACUITY

600 DIVISIONS
LEGIBLE LIMIT FOR
PAGE OR SCREEN

COLOR

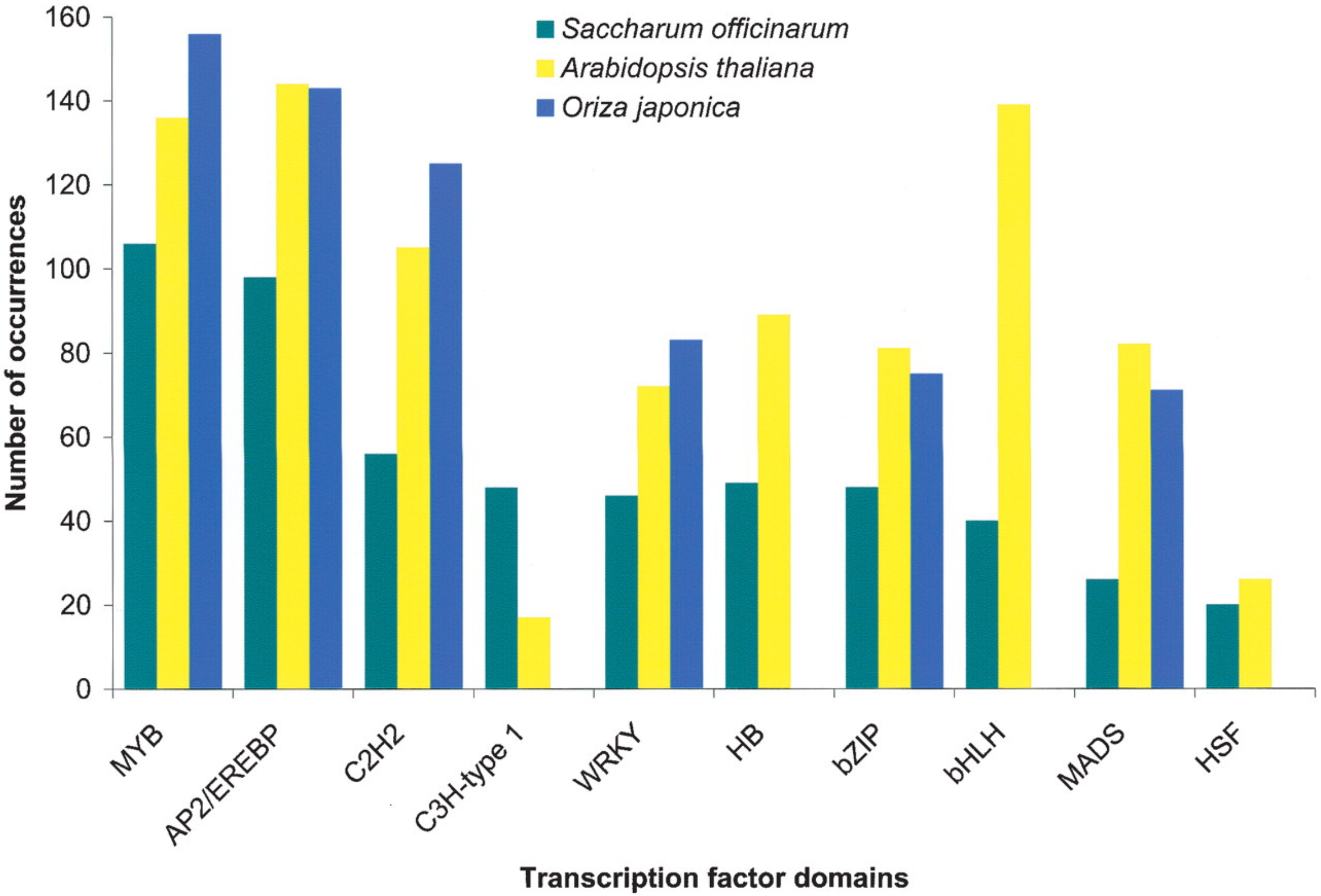
SOME HUES APPEAR BRIGHTER



We do not perceive all hues to have equal brightness.

Yellow *appears* brighter than blue.

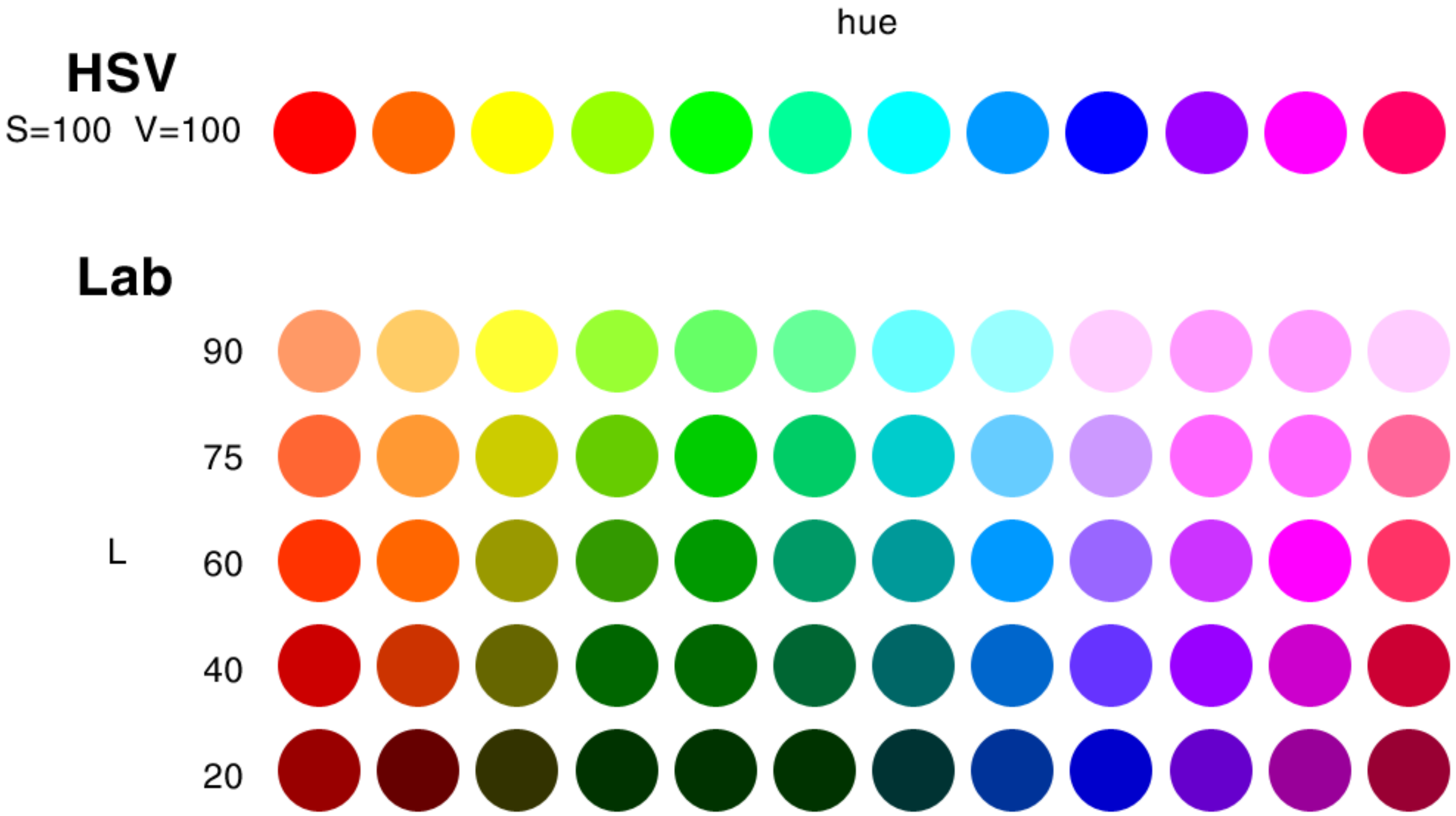
YELLOW COMMANDS ATTENTION

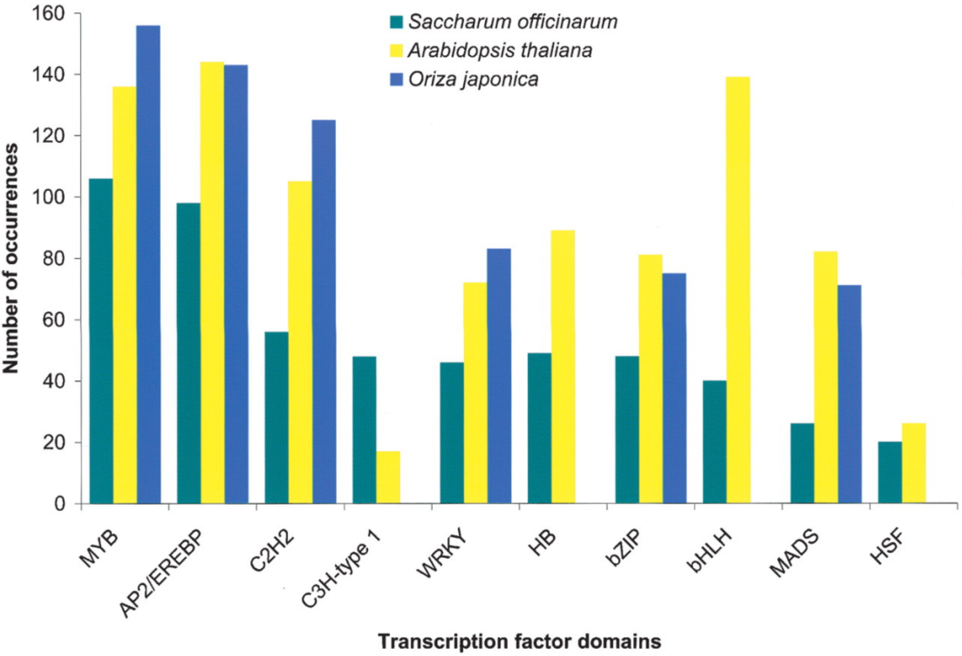


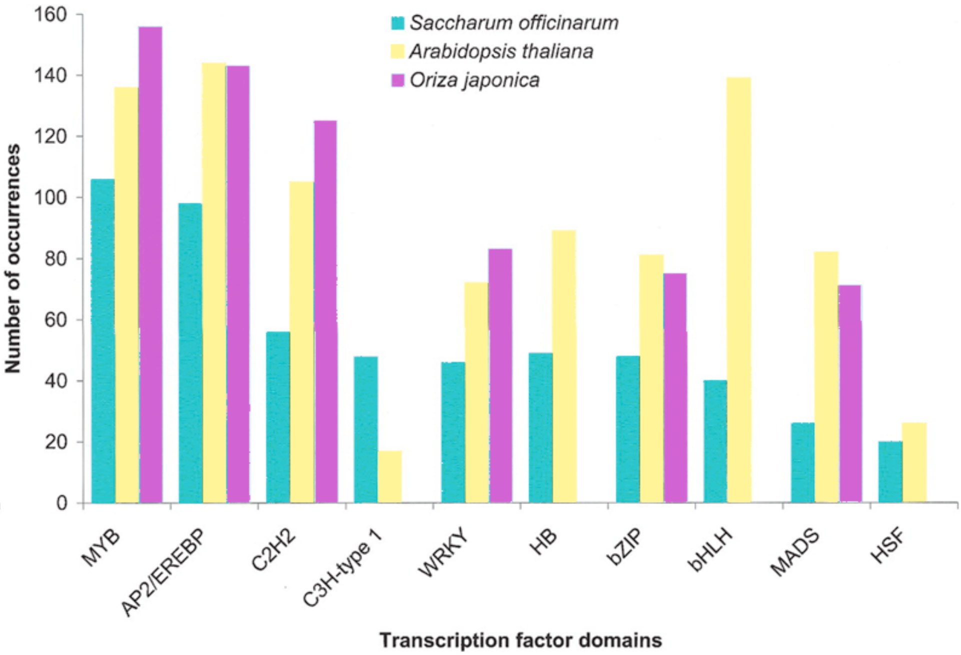
THE 10 MOST COMMON TRANSCRIPTION FACTOR PFAM DOMAINS IN SAS PROTEINS.

Vettore, A.L., et al., Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res, 2003. 13(12): p. 2725-35.

NORMALIZED LUMINANCE – PERCEPTUAL UNIFORMITY

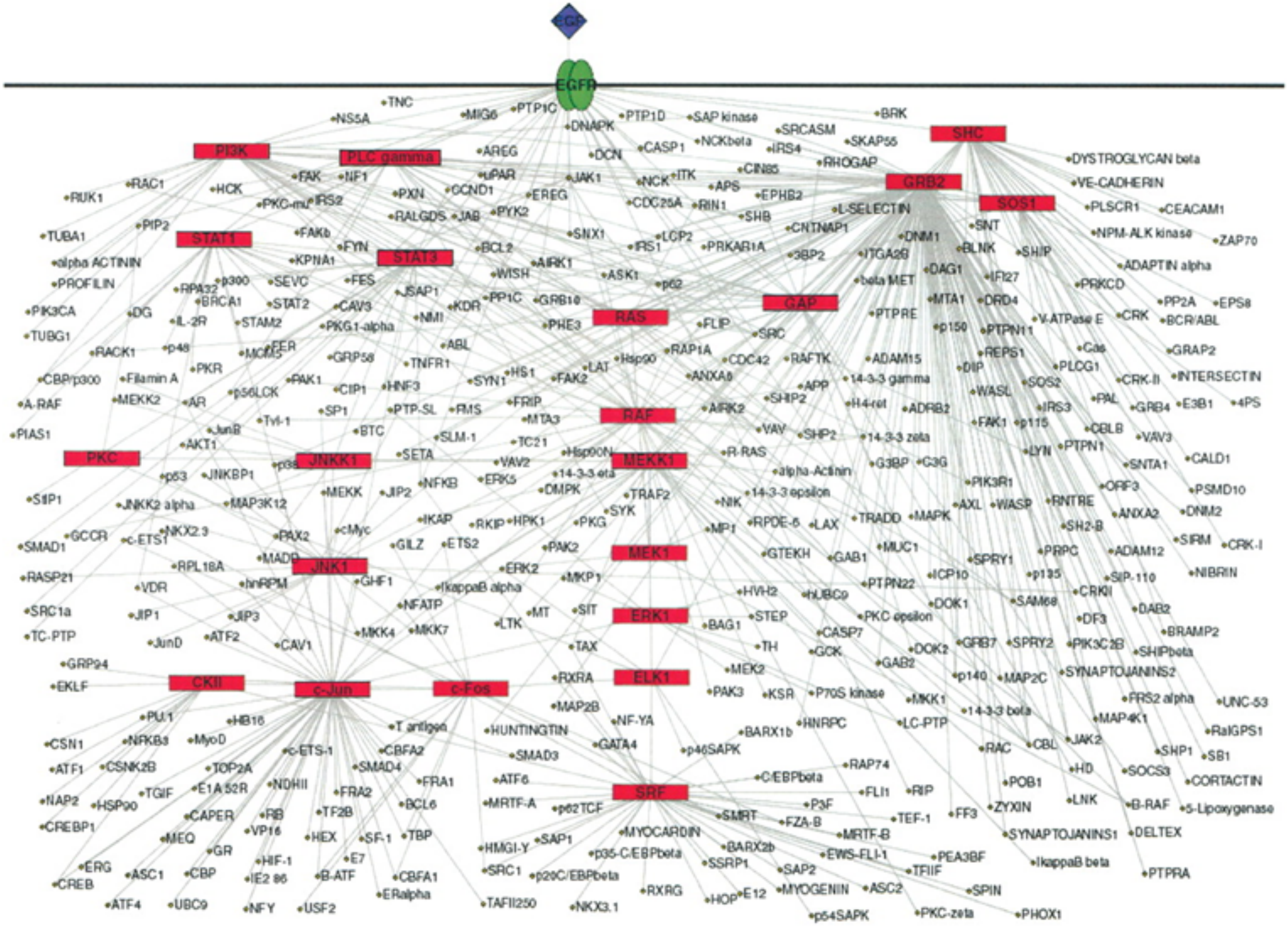






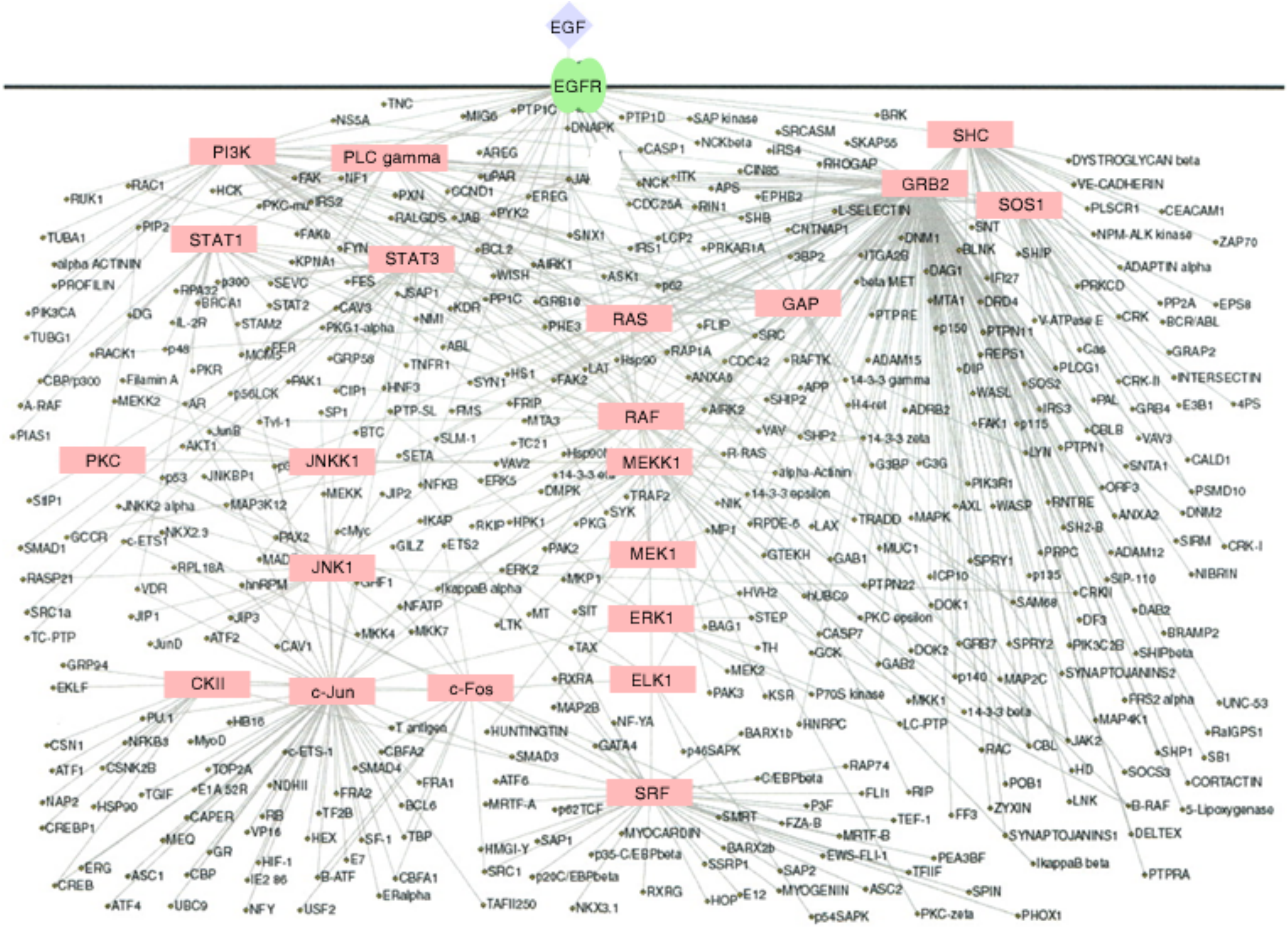
THE 10 MOST COMMON TRANSCRIPTION FACTOR PFAM DOMAINS IN SAS PROTEINS.

Vettore, A.L., et al., Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane.
Genome Res, 2003. 13(12): p. 2725-35.



THE EGF RECEPTOR SIGNALING PATHWAY GENERATED BY USING THE DATA CONTAINED IN HPRD.

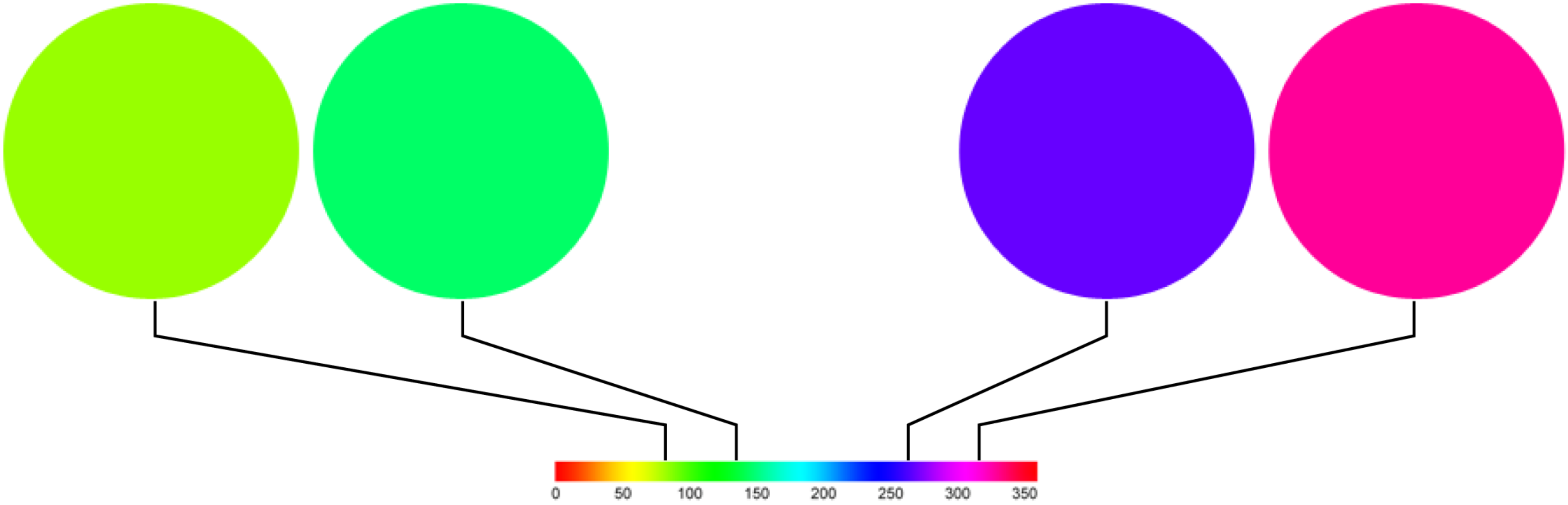
Peri, S., et al., Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res, 2003. 13(10): p. 2363-71.

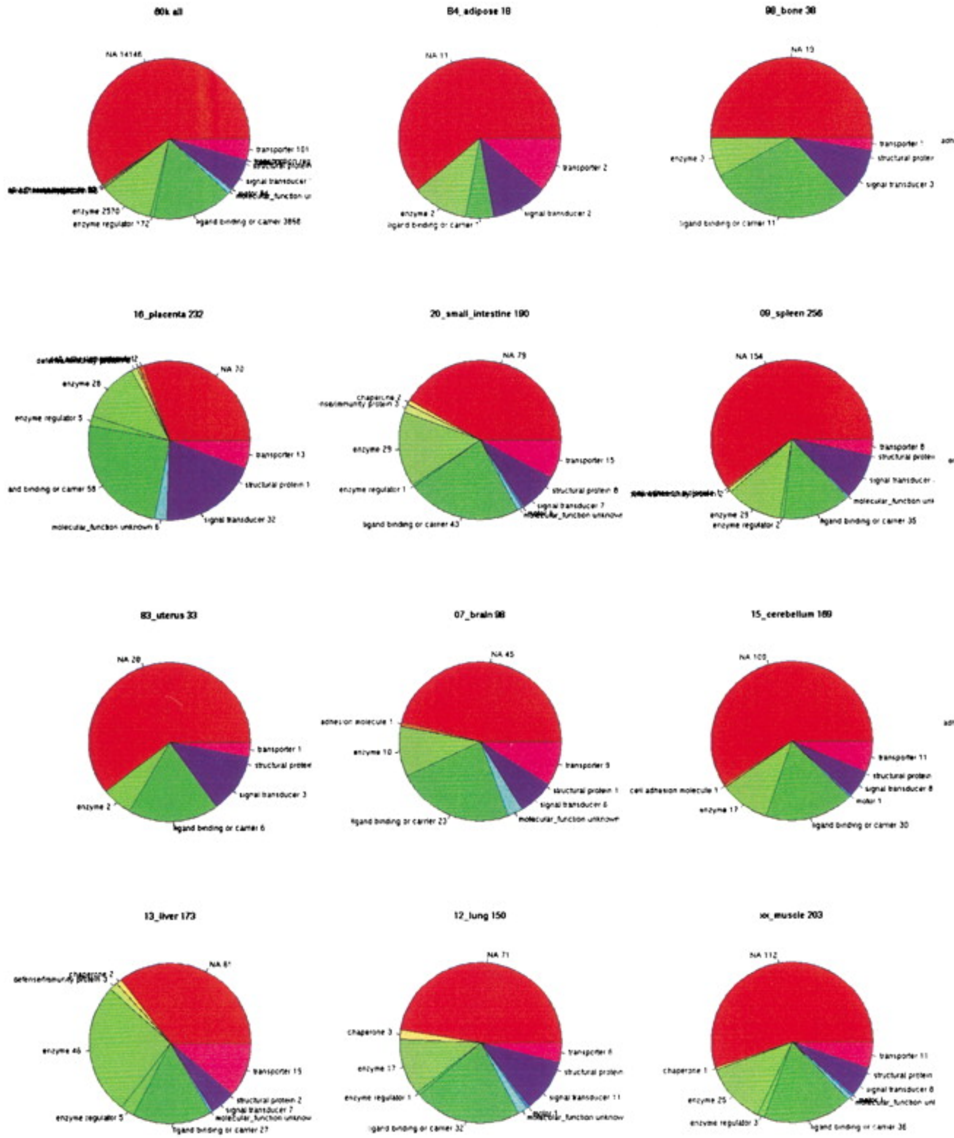


THE EGF RECEPTOR SIGNALING PATHWAY GENERATED BY USING THE DATA CONTAINED IN HPRD. FIGURE MODIFIED FROM ORIGINAL

Peri, S., et al., Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res, 2003. 13(10): p. 2363-71.

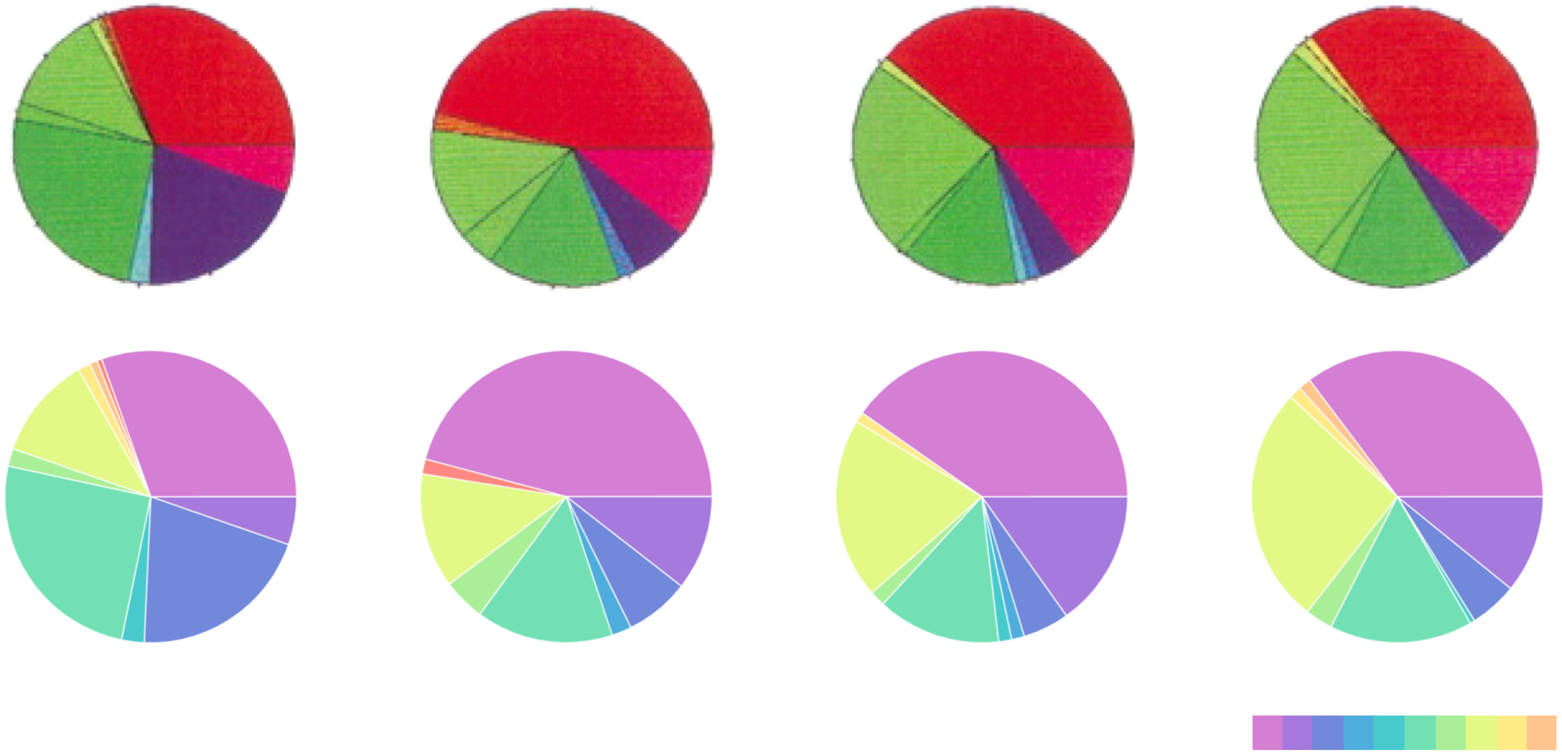
$\Delta H = 60$





PIE CHARTS FOR TISSUE PROFILING BY GENE ONTOLOGY.

Bono, H., et al., Systematic expression profiling of the mouse transcriptome using RIKEN cDNA microarrays. Genome Res, 2003. 13(6B): p. 1318-23.



PIE CHARTS FOR TISSUE PROFILING BY GENE ONTOLOGY.

Bono, H., et al., Systematic expression profiling of the mouse transcriptome using RIKEN cDNA microarrays. Genome Res, 2003. 13(6B): p. 1318-23.

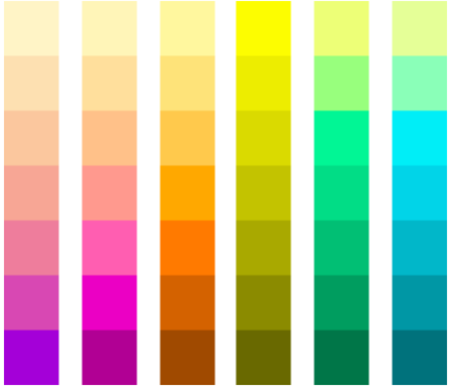
HOW TO CHOOSE COLORS

colorbrewer.org to choose palettes

kuler.adobe.com to peruse/create attractive color combinations

effective colors – perceived as

- equally important
- equally distant
- having a natural order (for sequential palettes)



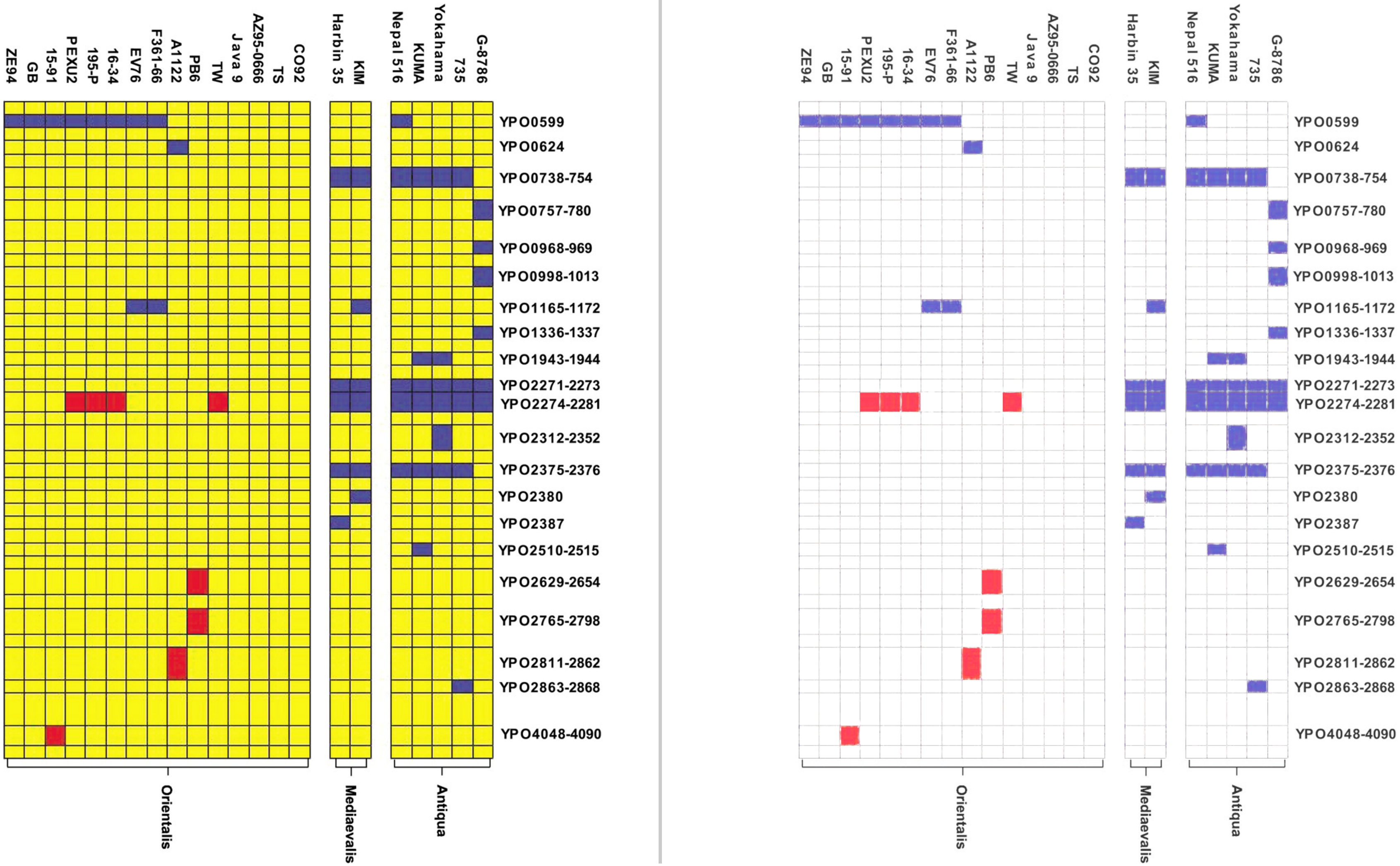
SEQUENTIAL PALETTES

QUALITATIVE PALETTES

http://www.colorbrewer.org

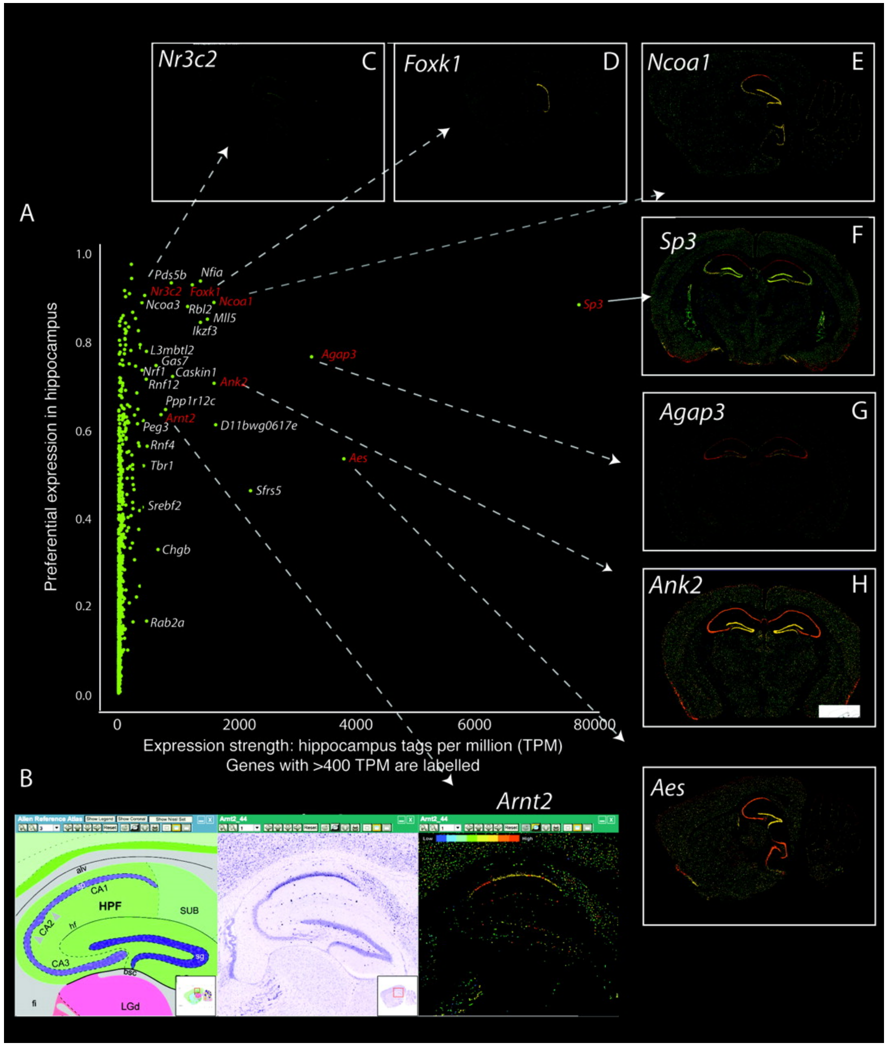
part 1
MAKING IT LEGIBLE

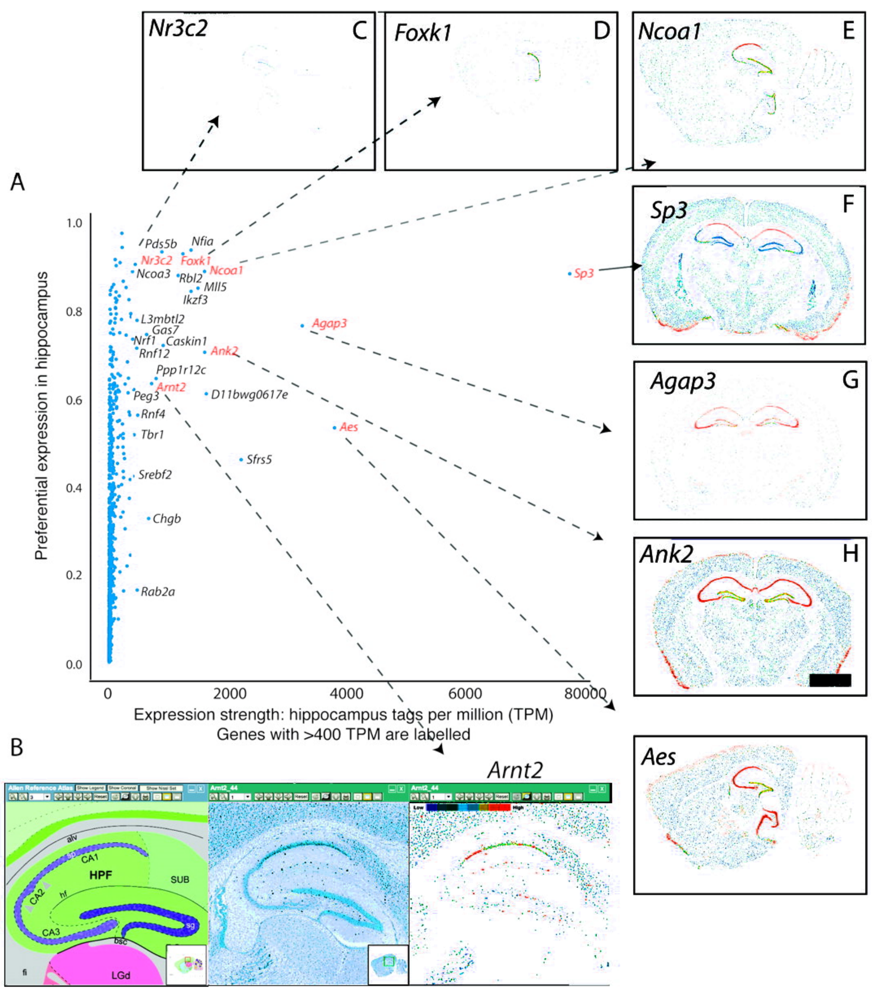
VISIBILITY



SCHEMATIC OF CHROMOSOMAL COMPARISON OF 22 STRAINS OF Y. PESTIS DETAILING ALL OF THE REGIONS OF DIVERGENCE FROM CO-92.

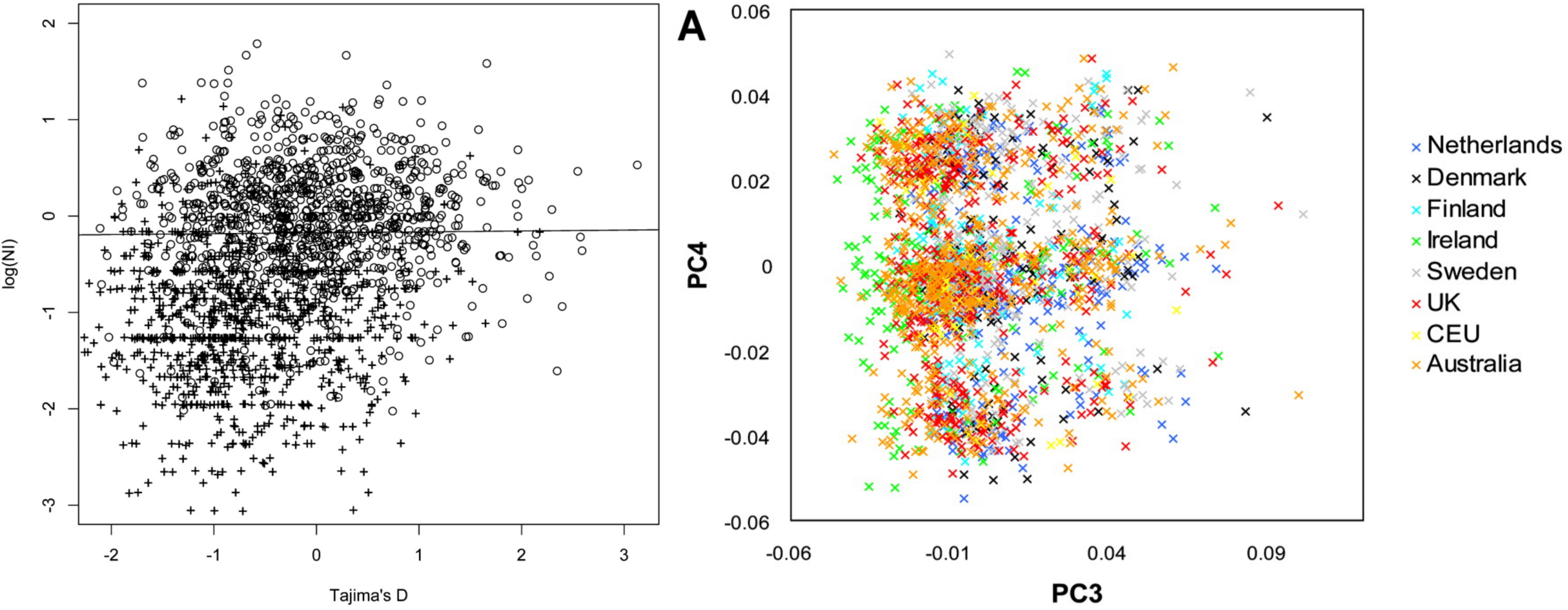
Hinchliffe, S.J., et al., Application of DNA microarrays to study the evolutionary genomics of Yersinia pestis and Yersinia pseudotuberculosis. Genome Res, 2003. 13(9): p. 2018-29.





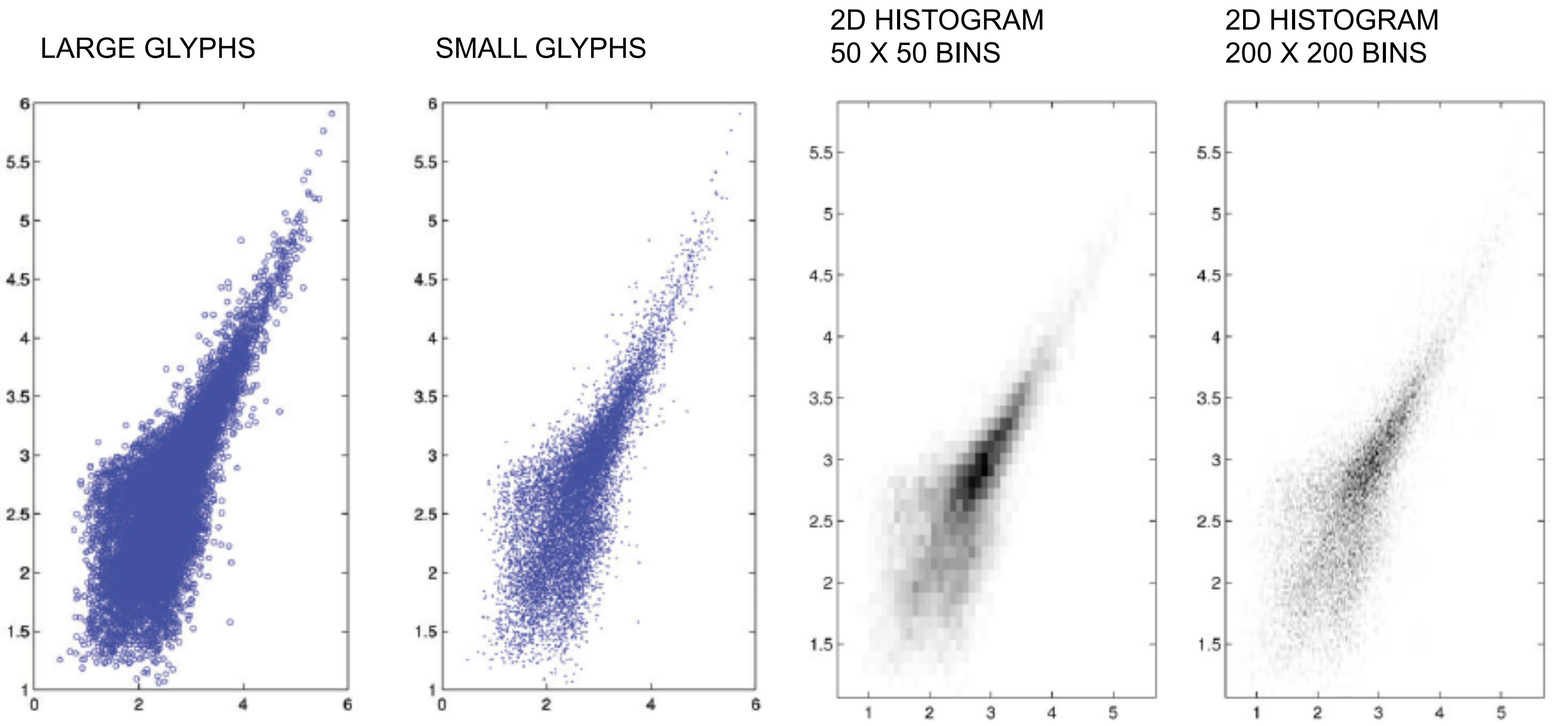
TRANSCRIPTION FACTOR GENES WITH PREFERENTIAL EXPRESSION IN HIPPOCAMPUS.

Valen, E., et al., Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. Genome Res, 2009. 19(2): p. 255-65.



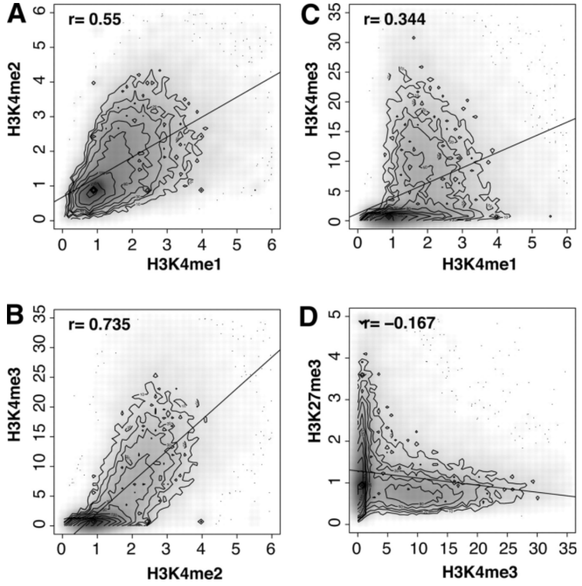
(LEFT) THE DISTRIBUTION OF VALUES OF THE LOGARITHM OF THE NEUTRALITY INDEX AND TAJIMA'S *D*
(RIGHT) PC3 AND PC4 IN NORTHERN EUROPEAN POPULATIONS.

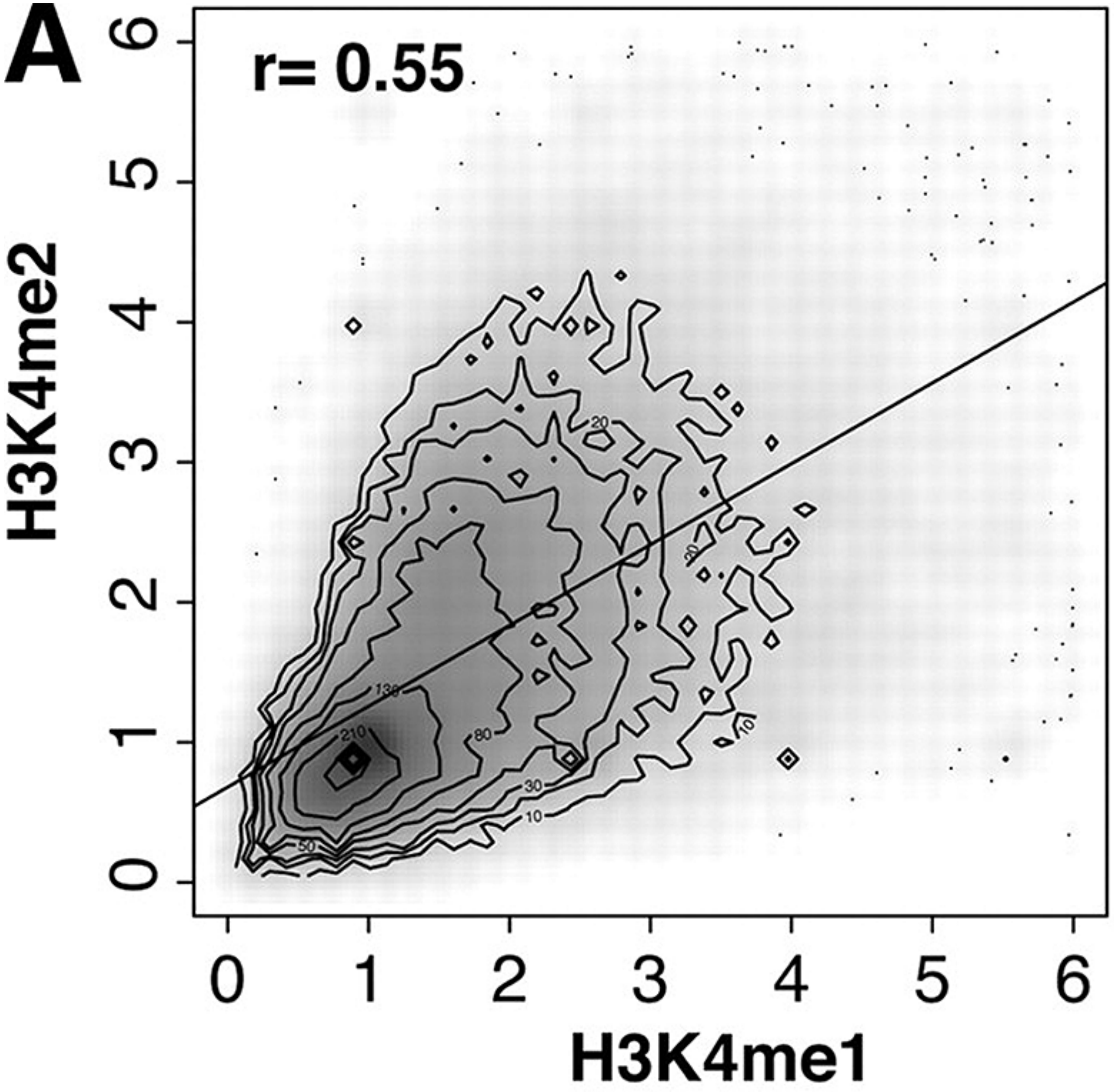
McEvoy, B.P., et al., Geographical structure and differential natural selection among North European populations. Genome Res, 2009. 19(5): p. 804-14., Nielsen, R., et al., Darwinian and demographic forces affecting human protein coding genes. Genome Res, 2009. 19(5): p. 838-49.



LEFT TWO SCATTERPLOTS OF LOG-EXPRESSIONS OF A PAIR OF MICROARRAYS.
RIGHT TWO-DIMENSIONAL HISTOGRAMS, DERIVED FROM THE SCATTERPLOTS IN FIGURE 1.

Eilers, P.H. and J.J. Goeman, Enhancing scatterplots with smoothed densities. Bioinformatics, 2004. 20(5): p. 623-8.





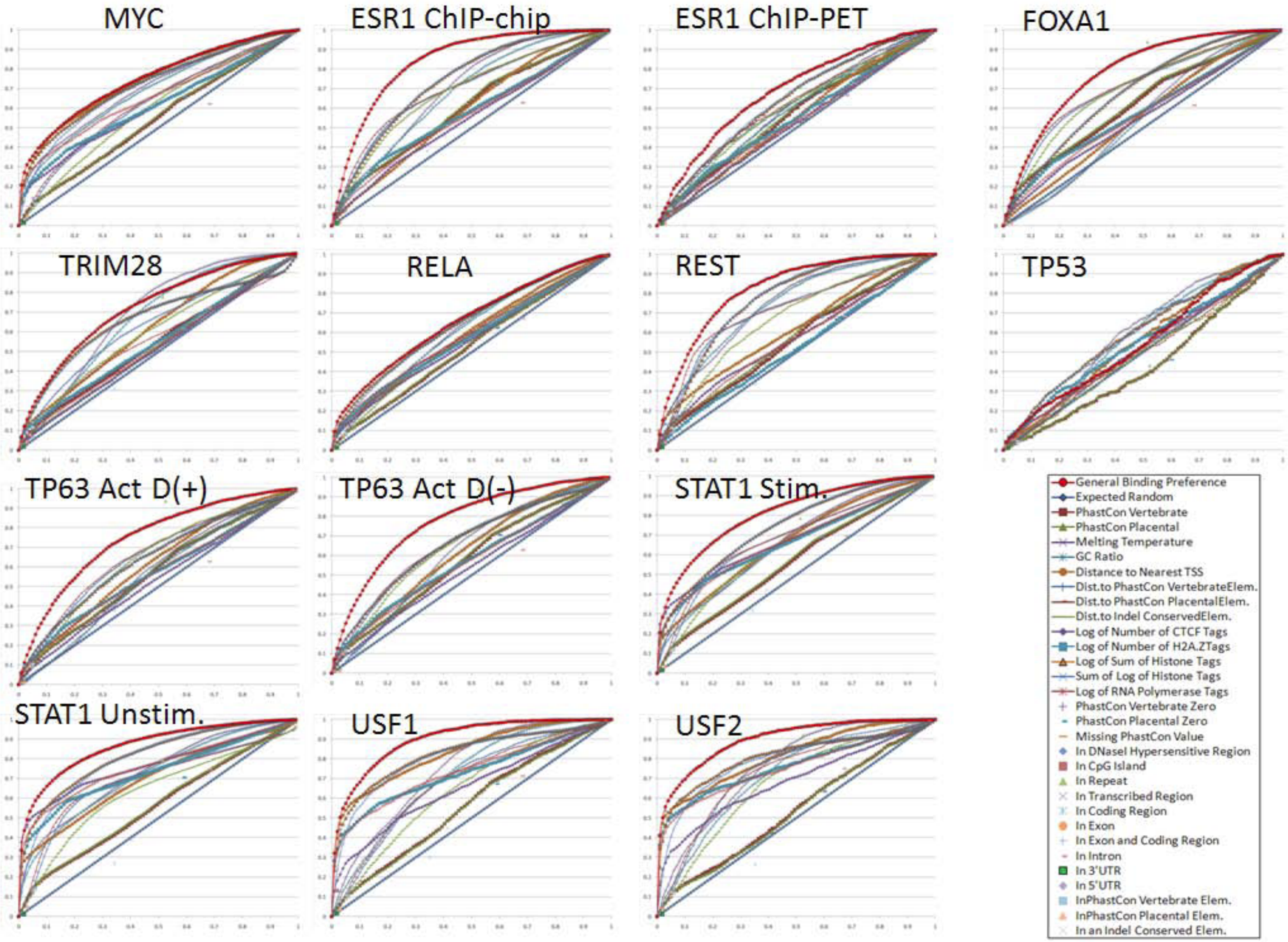
CO-OCCURRENCE OF INPUT-NORMALIZED HISTONE MODIFICATION
NEAR THE PROMOTERS (0–2000 BP) IN HUMAN ISLETS.

Bhandare, R., et al., Genome-wide analysis of histone modifications in human pancreatic islets. Genome Res, 2010. 20(4): p. 428-33.

Eilers, P.H. and J.J. Goeman, Enhancing scatterplots with smoothed densities. Bioinformatics, 2004. 20(5): p. 623-8.

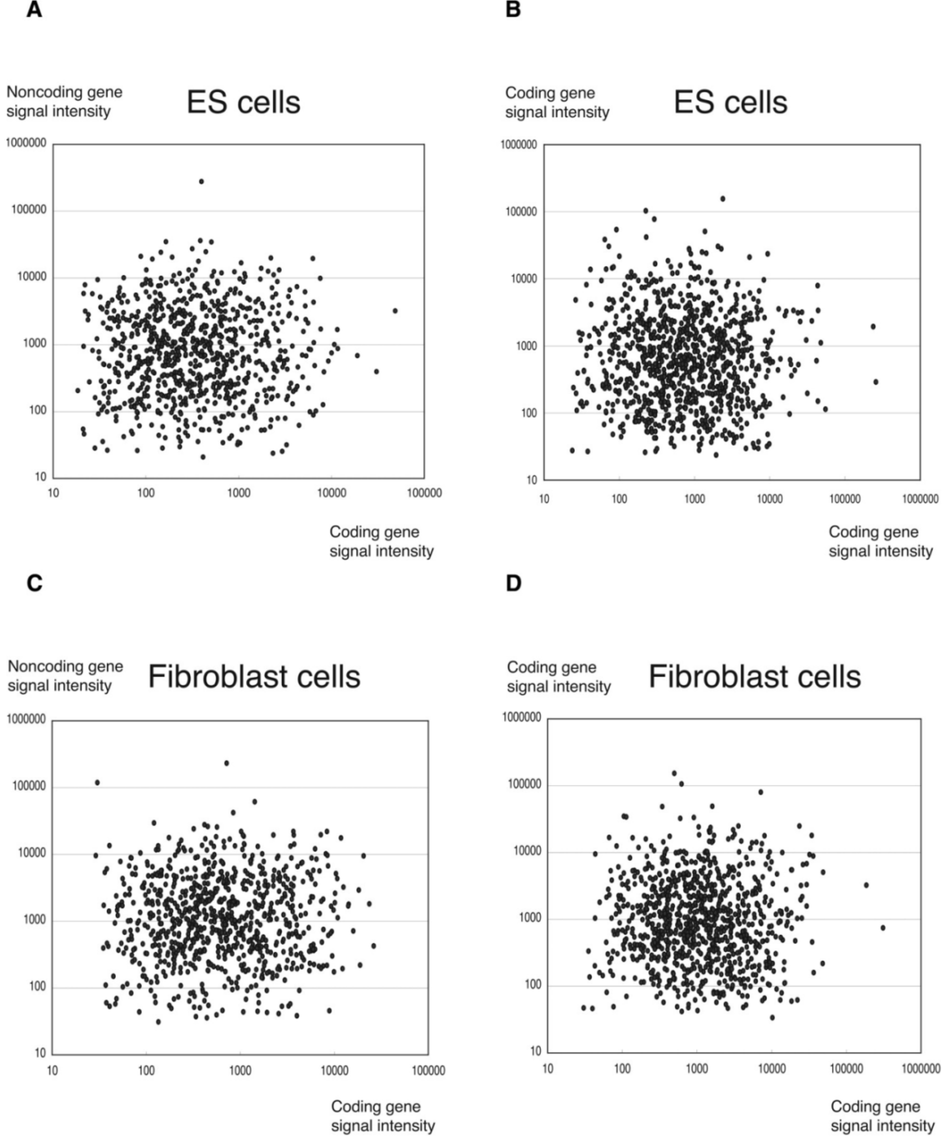
part 1
MAKING IT LEGIBLE

READER IS NOT A COMPUTER



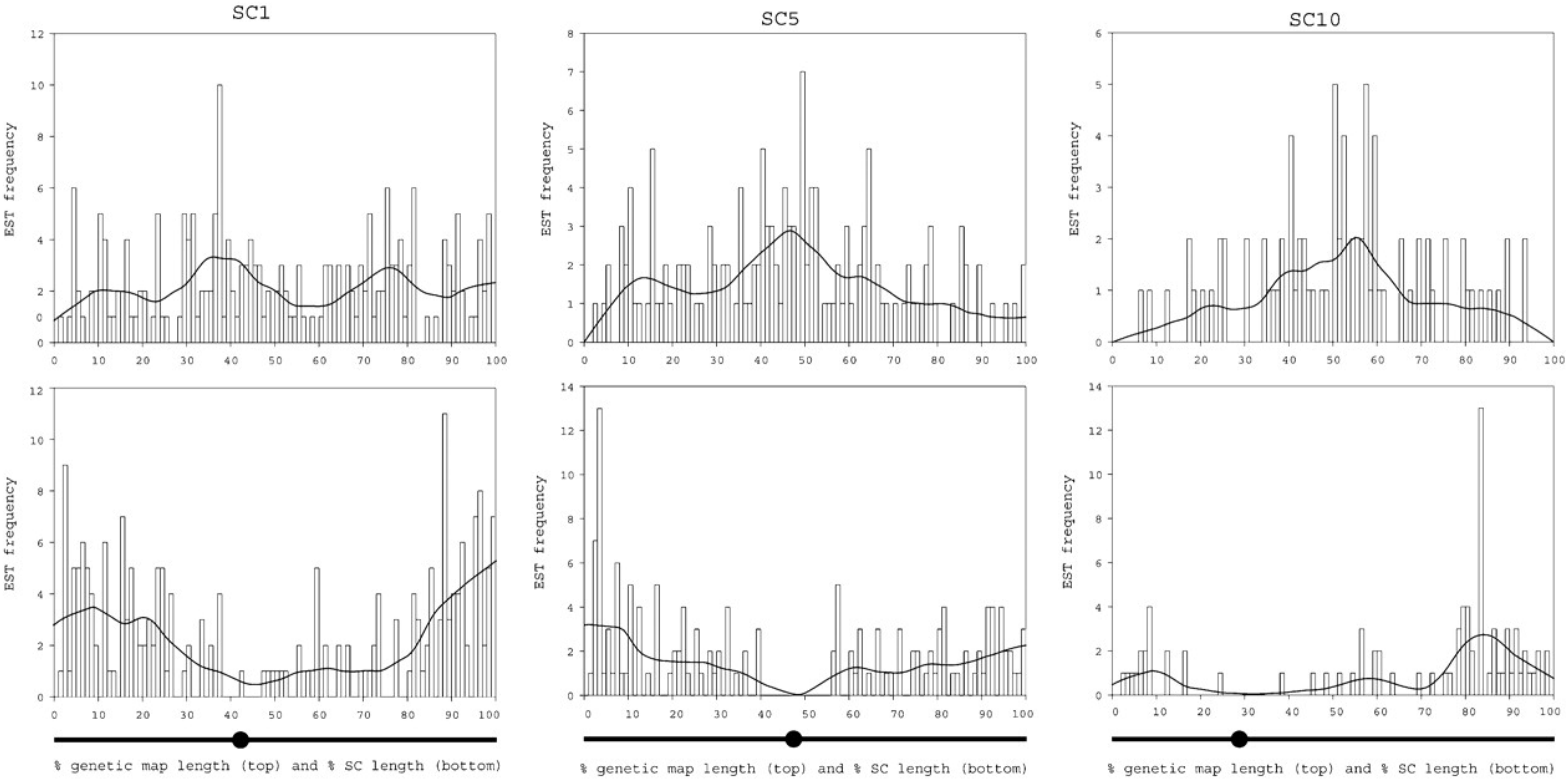
AN EXTENDED VERSION OF FIGURE 2 SHOWING ROC VALUES FOR EACH INDIVIDUAL FEATURE CONSIDERED. THE X-AXIS IS THE FALSE POSITIVE RATE AND THE Y-AXIS IS THE TRUE POSITIVE RATE.

Ernst, J., et al., Integrating multiple evidence sources to predict transcription factor binding in the human genome. Genome Res, 2010. 20(4): p. 526-36.



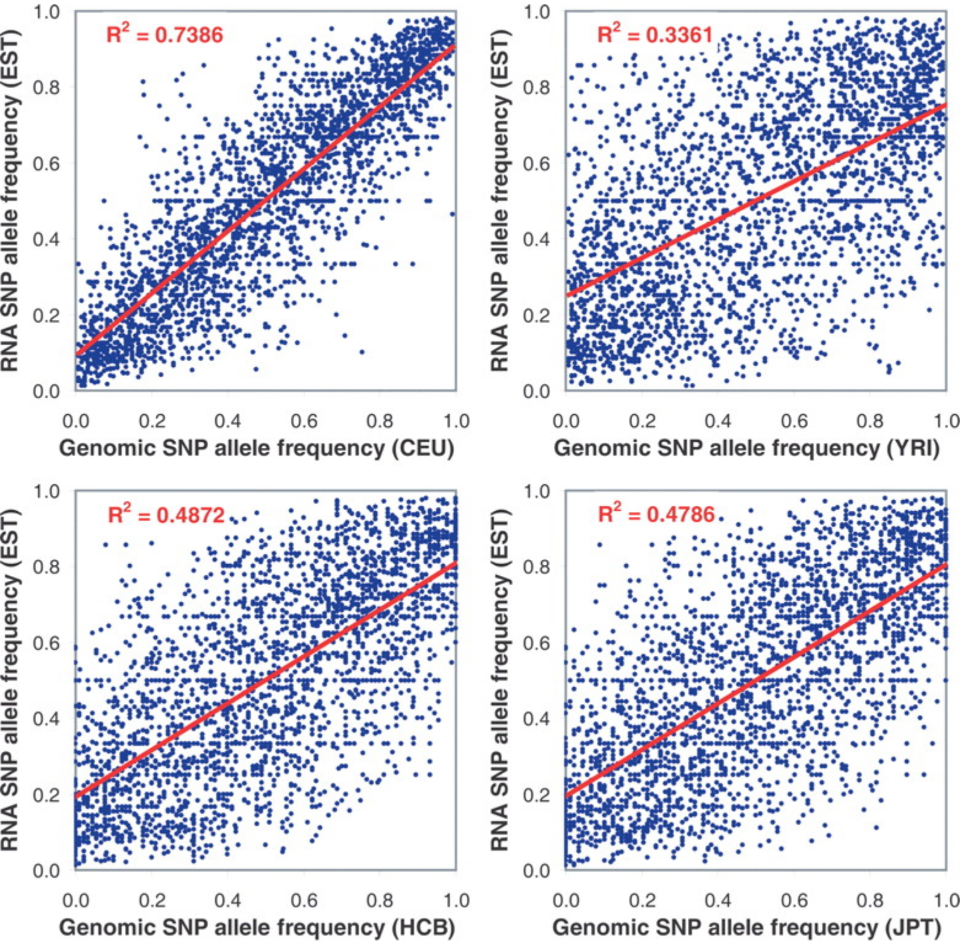
OVERALL EXPRESSION OF SENSE AND ANTISENSE GENES AS DETERMINED BY USING AN OLIGO DNA MICROARRAY.

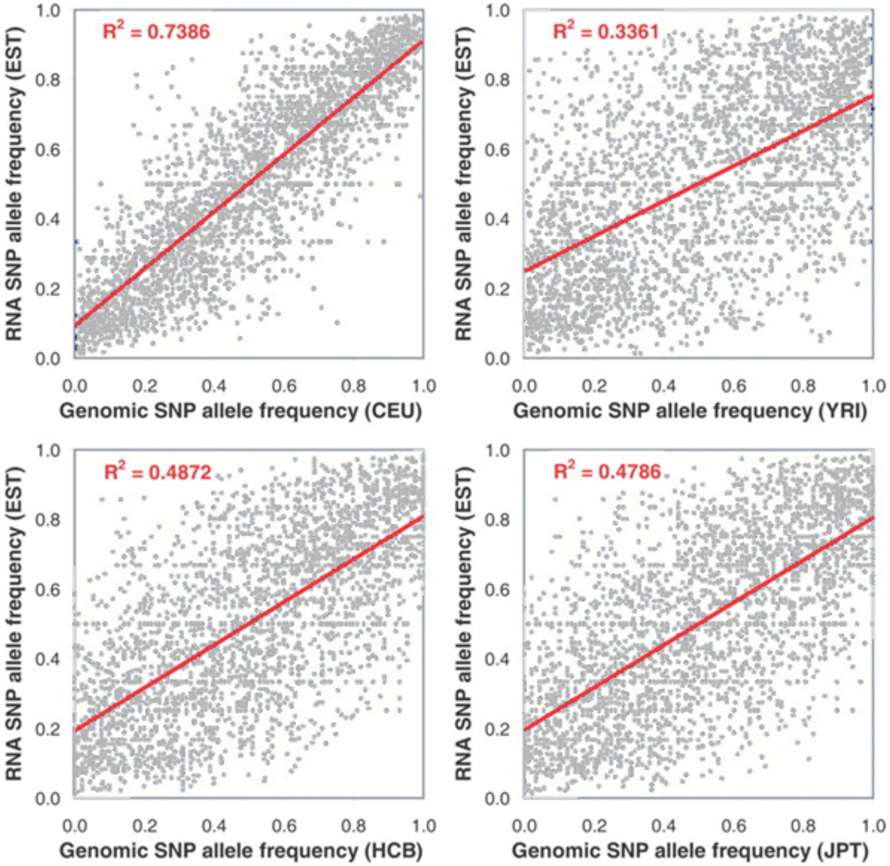
Kiyosawa, H., et al., Disclosing hidden transcripts: mouse natural sense-antisense transcripts tend to be poly(A) negative and nuclear localized.
Genome Res, 2005. 15(4): p. 463-74.



THE DISTRIBUTION OF ESTS FROM MAIZE CHROMOSOMES 1, 5, AND 10 IN 1-CM BINS ON GENETIC LINKAGE MAPS (TOP PANEL) AND IN 0.2-MM LENGTH BINS ON THE PHYSICAL STRUCTURE OF PACHYTENE CHROMOSOMES (BOTTOM PANEL).

Anderson, L.K., et al., Uneven distribution of expressed sequence tag loci on maize pachytene chromosomes. Genome Res, 2006. 16(1): p. 115-22.



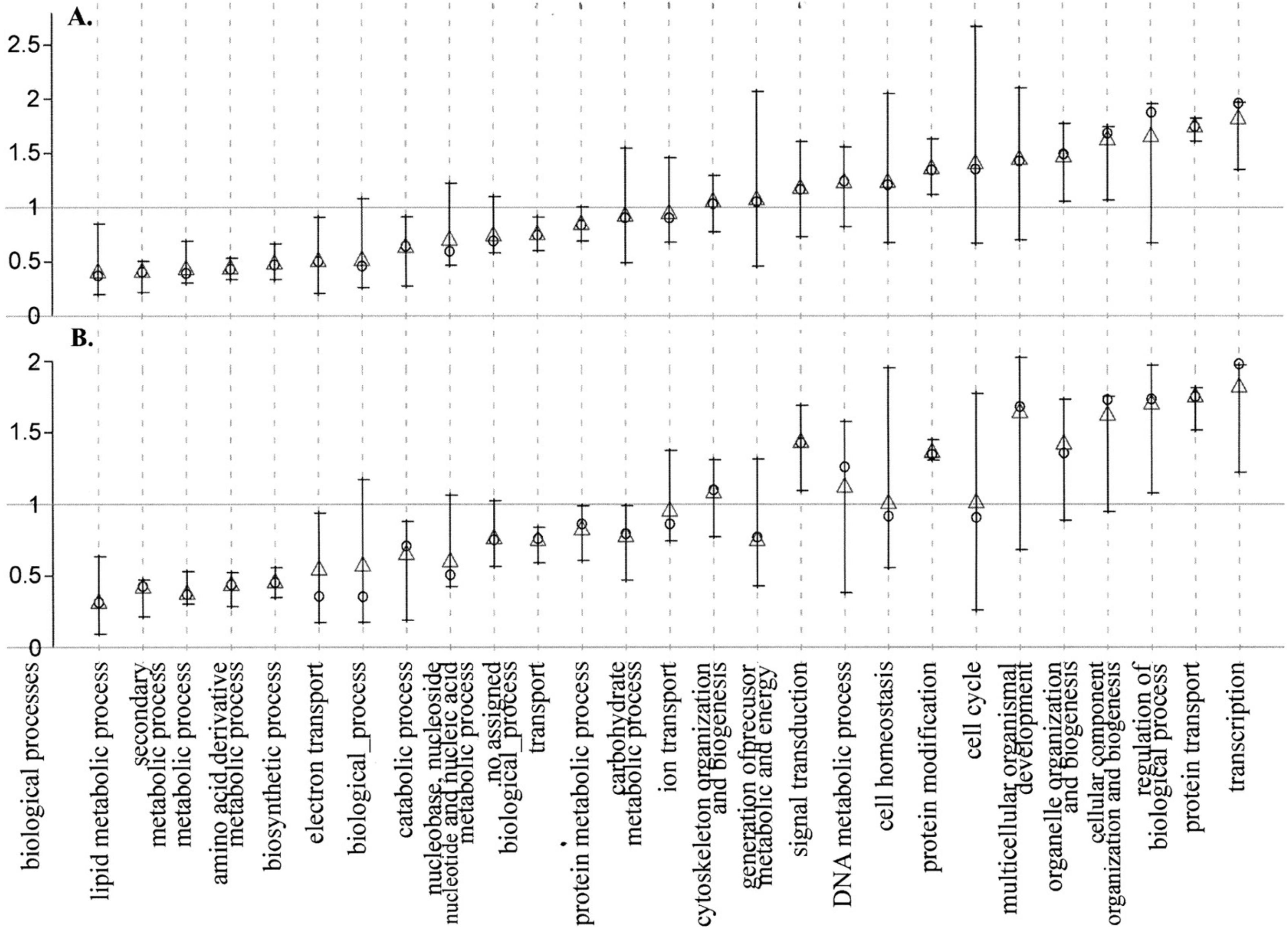


EST ALLELE FREQUENCIES VERSUS ALLELE FREQUENCIES IN DIFFERENT ETHNIC GROUPS.

Ge, B., et al., Survey of allelic expression using EST mining. Genome Res, 2005. 15(11): p. 1584-91.

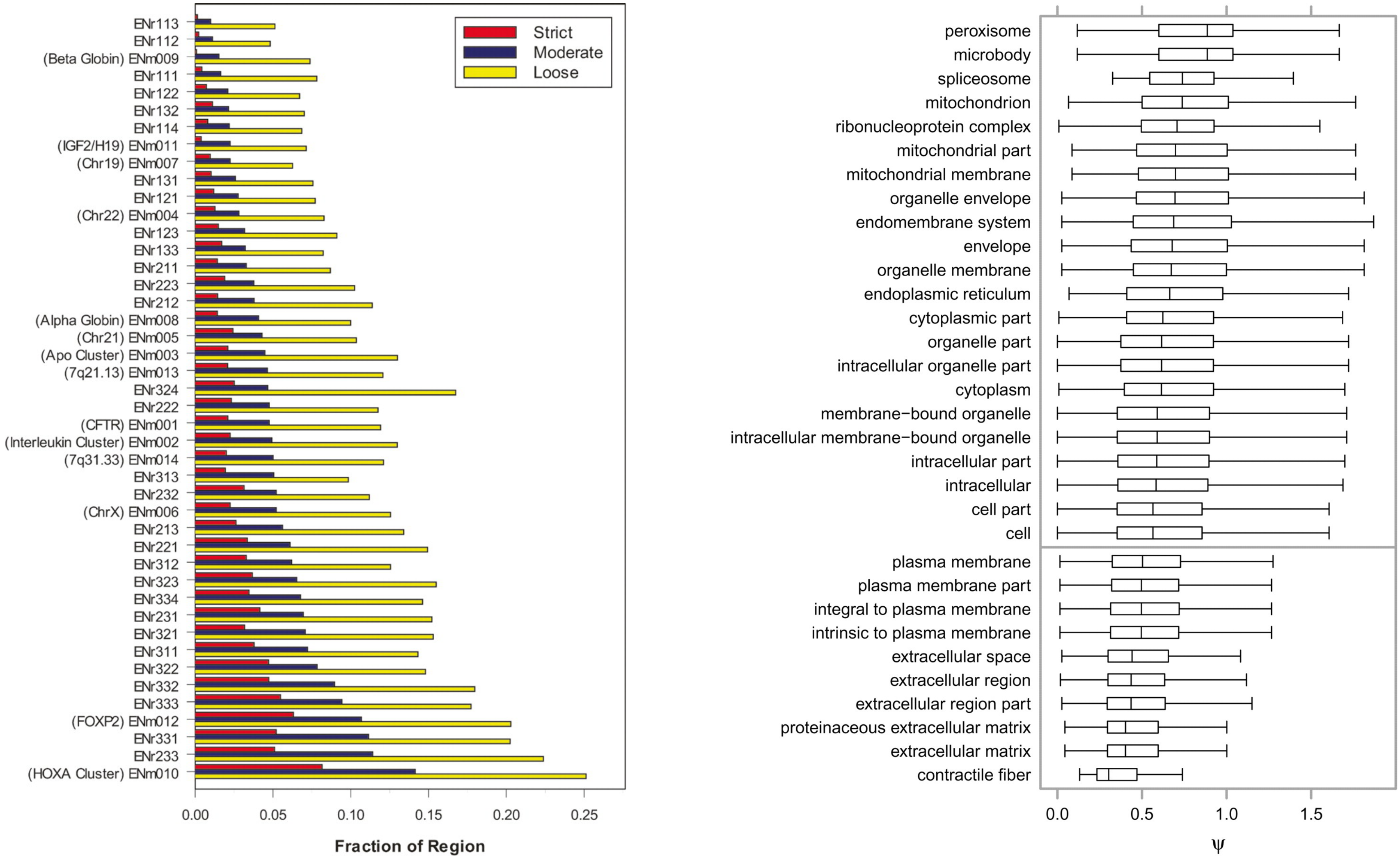
● part 1
MAKING IT LEGIBLE

TEXT ORIENTATION



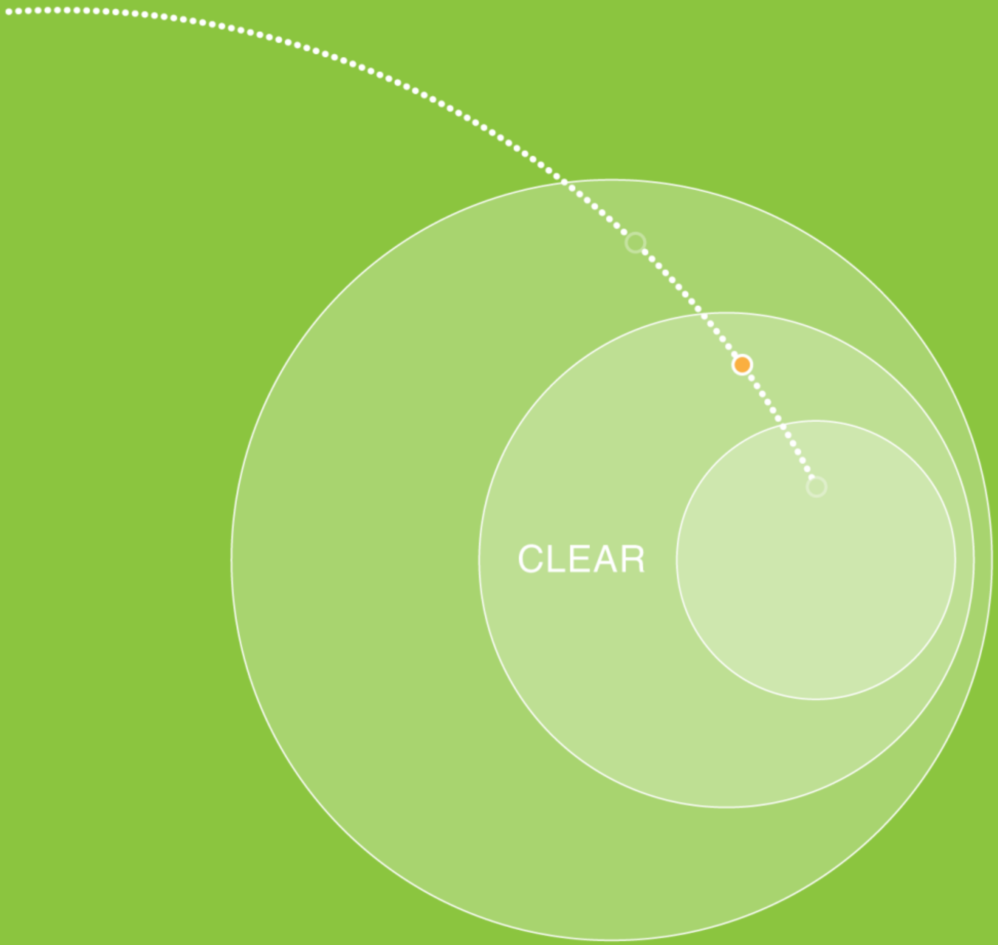
THE SLOPES FROM A SIMPLE LINEAR MODEL (A) AND A ROBUST LINEAR MODEL (B) FOR THE REGRESSION OF THE AVERAGE VERTEBRATE FAMILY SIZE (NV) AGAINST THE AVERAGE INVERTEBRATE FAMILY SIZE.

Prachumwat, A. and W.H. Li, Gene number expansion and contraction in vertebrate genomes with respect to invertebrate genomes. Genome Res, 2008. 18(2): p. 221-32.



(LEFT) CONSTRAINED BASES IN EACH ENCODE REGION.
(RIGHT) BOX PLOT OF Ψ = DT/DS VALUES ARRANGED BY GO CELLULAR COMPONENT TERM.

Hoffman, M.M. et al., An effective model for natural selection in promoters. Genome Res, 2010. 20(5): 685-92.,
Margulies, E.H., et al., Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. Genome Res, 2007. 17(6): 760-74.



next

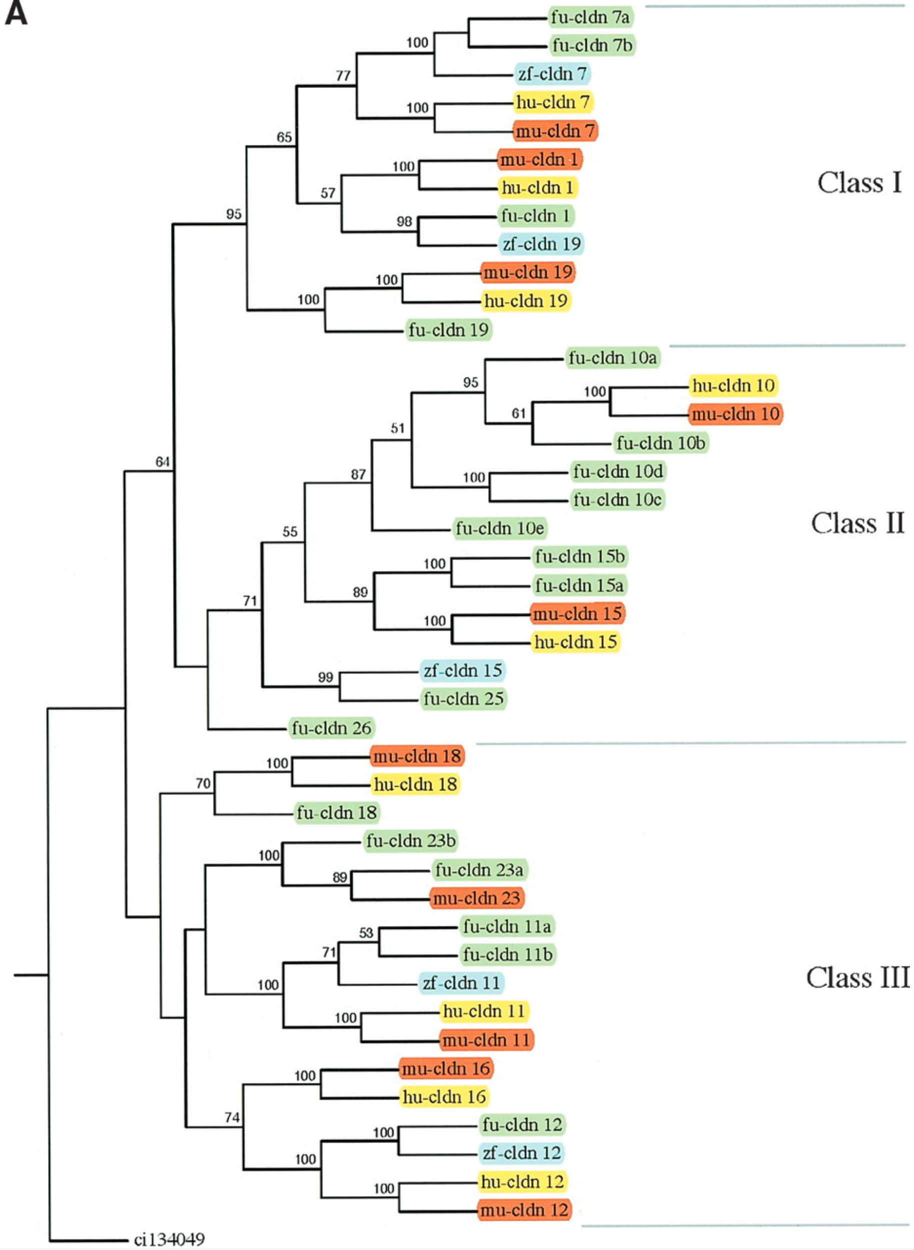
○ **part 1**
MAKING IT LEGIBLE

● **part 2**
MAKING IT CLEAR

○ **part 3**
MAKING IT PRETTY

part 2
MAKING IT CLEAR

REDUNDANCY

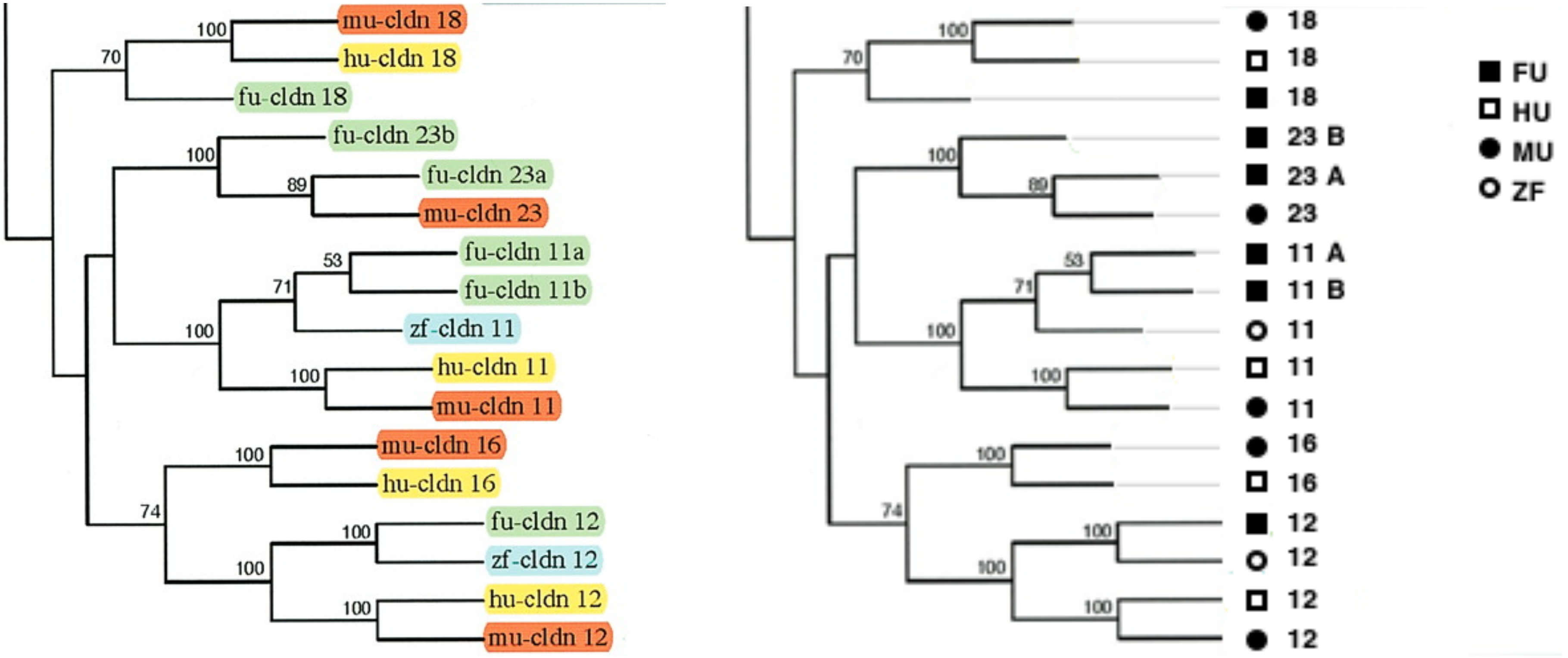


All labels are of the form

fu-cldn 7a
fu-cldn 7b
zf-cldn 7
mu-cldn 19
hu-cldn 15

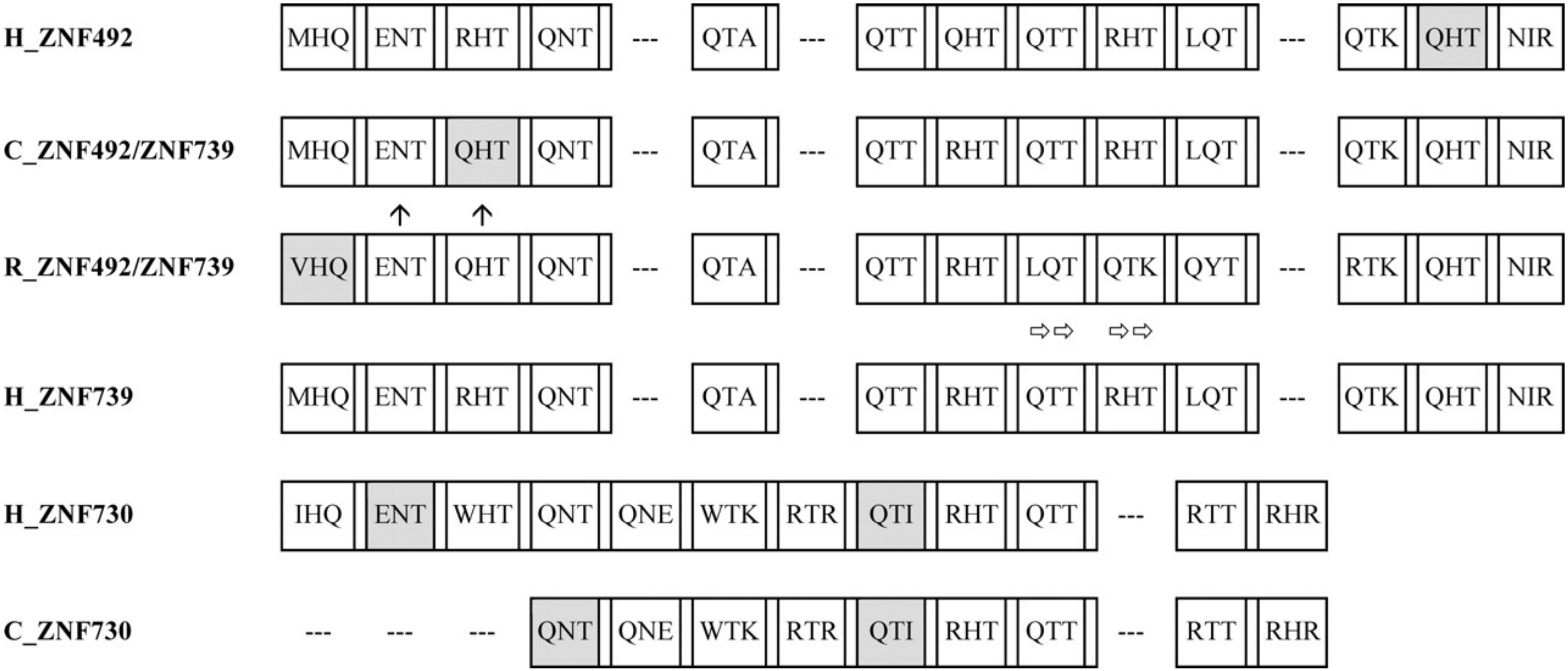
CONSENSUS PHYLOGENETIC TREE OF CLAUDIN PROTEINS.

Y. H. Loh, A. Christoffels, S. Brenner, W. Hunziker, B. Venkatesh, Genome Res 14, 1248 (Jul, 2004).



CONSENSUS PHYLOGENETIC TREE OF CLAUDIN PROTEINS.

Y. H. Loh, A. Christoffels, S. Brenner, W. Hunziker, B. Venkatesh, Genome Res 14, 1248 (Jul, 2004).



H_ZNF492 MHQ ENT RHT QNT --- QTA --- QTT QHT QTT RHT LQT --- QTK QHT NIR

C_ZNF492/ZNF739 MHQ ENT QHT QNT --- QTA --- QTT RHT QTT RHT LQT --- QTK QHT NIR

R_ZNF492/ZNF739 VHQ ENT QHT QNT --- QTA --- QTT RHT LQT QTK QYT --- RTK QHT NIR

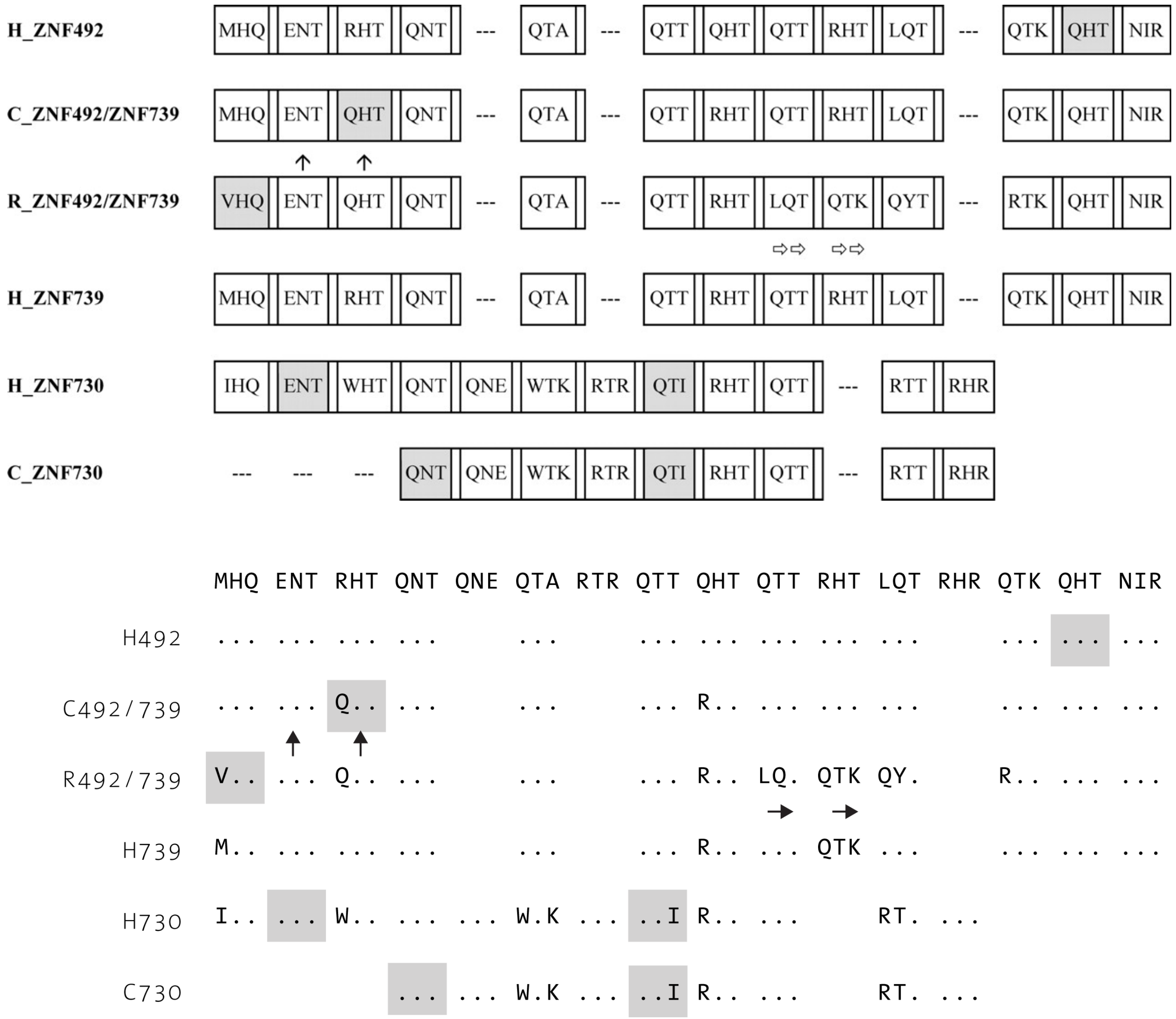
H_ZNF739 MHQ ENT RHT QNT --- QTA --- QTT RHT QTT RHT LQT --- QTK QHT NIR

H_ZNF730 IHQ ENT WHT QNT QNE WTK RTR QTI RHT QTT --- RTT RHR

C_ZNF730 --- --- --- QNT QNE WTK RTR QTI RHT QTT --- RTT RHR

ZINC FINGER EXON ANALYSIS FOR ZNF493 AND ZNF738, TWO DIVERGENT GENES FROM THE ZNF431 CLADE.

Hamilton, A.T., et al., Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes. Genome Res, 2006. 16(5): p. 584-94.



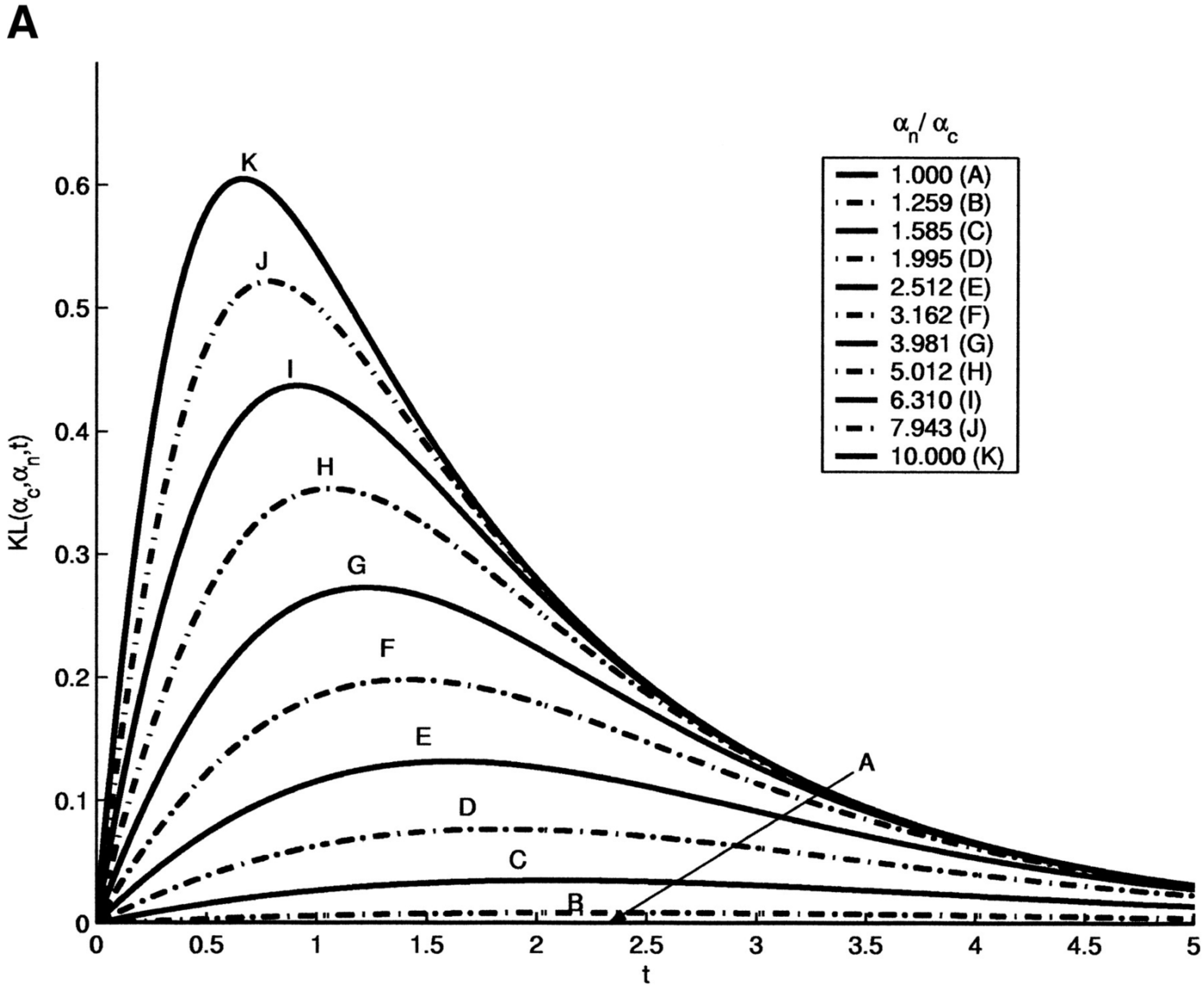
```
H_ZNF492          MHQ ENT RHT QNT --- QTA --- QTT QHT QTT RHT LQT --- QTK QHT NIR
C_ZNF492/ZNF739   MHQ ENT QHT QNT --- QTA --- QTT RHT QTT RHT LQT --- QTK QHT NIR
R_ZNF492/ZNF739   VHQ ENT QHT QNT --- QTA --- QTT RHT LQT QTK QYT --- RTK QHT NIR
H_ZNF739          MHQ ENT RHT QNT --- QTA --- QTT RHT QTT RHT LQT --- QTK QHT NIR
H_ZNF730          IHQ ENT WHT QNT QNE WTK RTR QTI RHT QTT --- RTT RHR
C_ZNF730          --- --- --- QNT QNE WTK RTR QTI RHT QTT --- RTT RHR
```

```
          MHQ ENT RHT QNT QNE QTA RTR QTT QHT QTT RHT LQT RHR QTK QHT NIR
H492      ... ... ... ...     ...     ... ... ... ... ...     ... ... ...
C492/739  ... ... Q.. ...     ...     ... R.. ... ... ...     ... ... ...
R492/739  V.. ... Q.. ...     ...     ... R.. LQ. QTK QY.     R.. ... ...
H739      M.. ... ... ...     ...     ... R.. ... QTK ...     ... ... ...
H730      I.. ... W.. ... ... W.K ... ..I R.. ...     RT. ...
C730                  ... ... W.K ... ..I R.. ...     RT. ...
```

ZINC FINGER EXON ANALYSIS FOR ZNF493 AND ZNF738, TWO DIVERGENT GENES FROM THE ZNF431 CLADE.

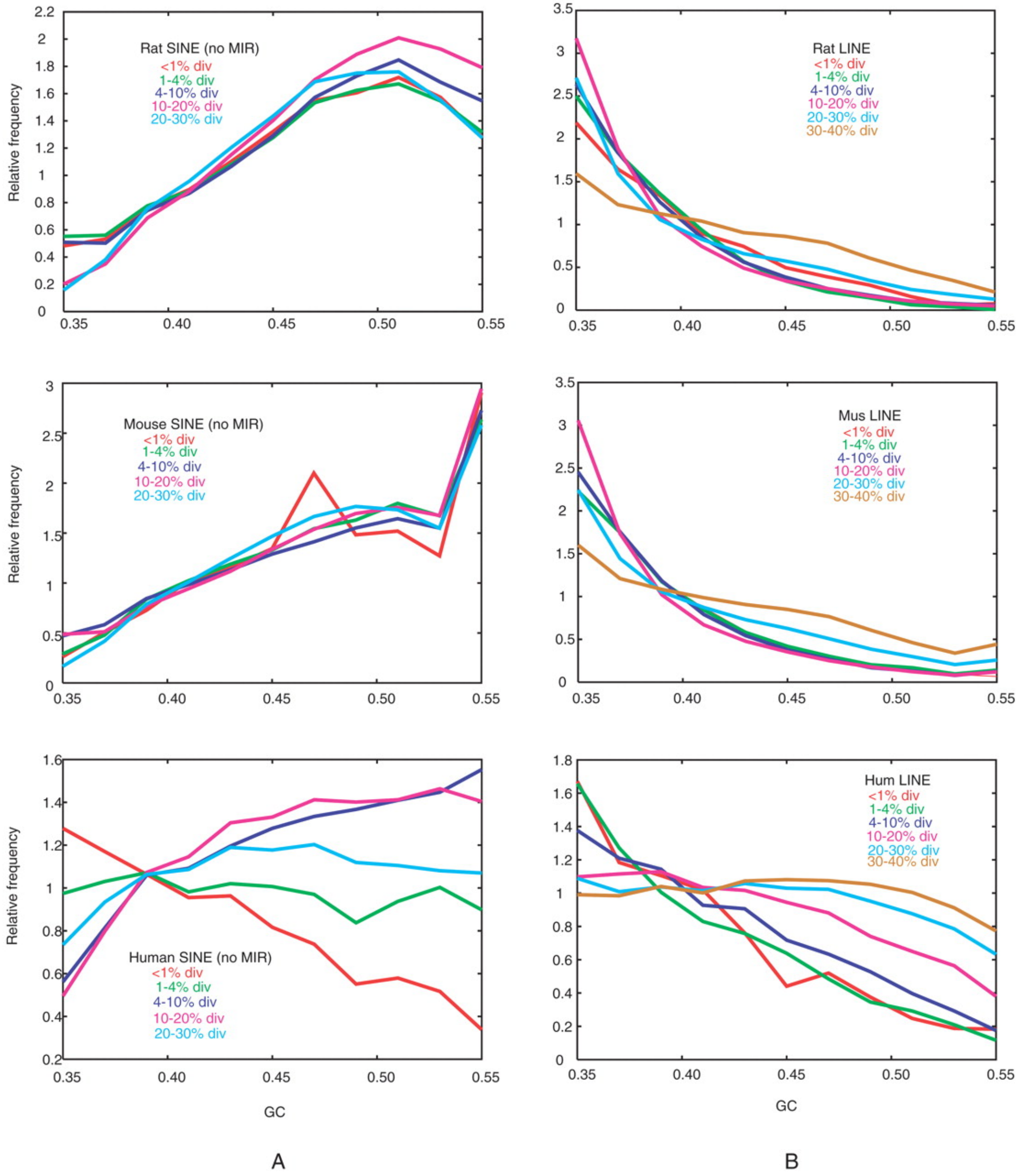
Hamilton, A.T., et al., Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes. Genome Res, 2006. 16(5): p. 584-94.

LEGEND



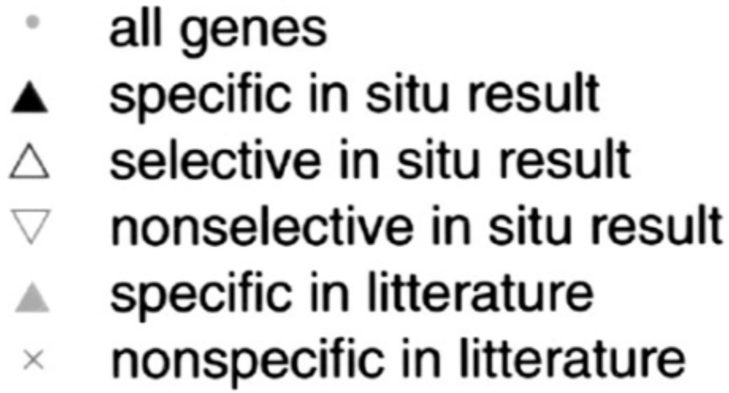
PRECISION (KL-DIVERGENCE IN THE FIGURE) AS A FUNCTION OF DIVERGENCE TIME.

Zhang, L., et al., Human-mouse gene identification by comparative evidence integration and evolutionary analysis. Genome Res, 2003. 13(6A): p. 1190-202.



FREQUENCY OF OCCURRENCE OF DIFFERENT AGES AND FAMILIES OF INTERSPERSED REPEATS IN REGIONS OF DIFFERENT GC-CONTENT.

Yang, S., et al., Patterns of insertions and their covariation with substitutions in the rat, mouse, and human genomes. Genome Res, 2004. 14(4): p. 517-27.

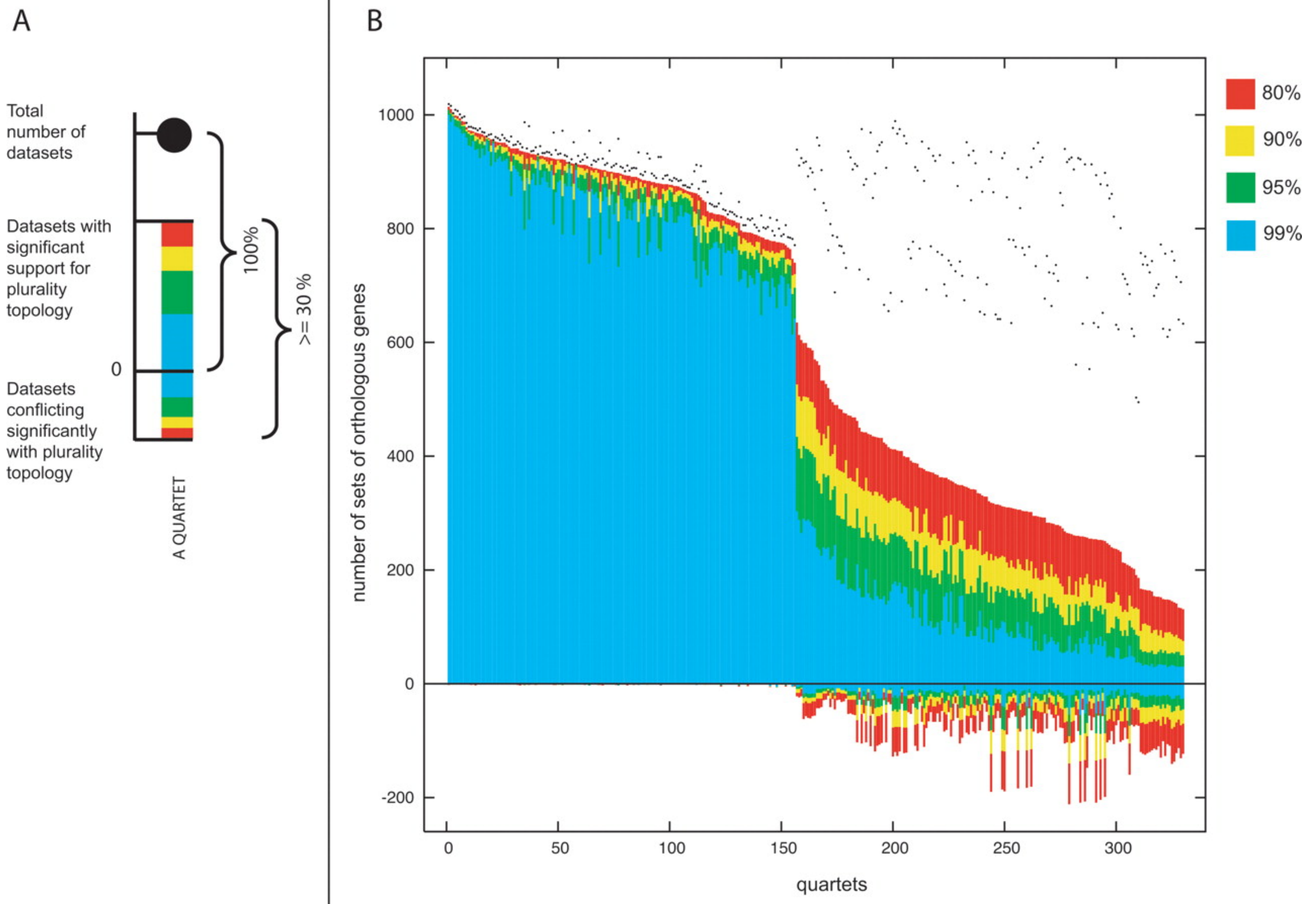


GENE •

	IN SITU	IN LIT
SPECIFIC	●	▲
SELECTIVE	●	
NON-SELECTIVE	○	
NON-SPECIFIC		△

ISH RESULTS IN RELATION TO PROFILE SIMILARITY TO SM-MHC (Y-AXIS) AND TO EXPRESSION LEVEL (X-AXIS). LEGEND MODIFIED FROM ORIGINAL

Nelander, S., P. Mostad, and P. Lindahl, Prediction of cell type-specific gene modules: identification and initial characterization of a core set of smooth muscle-specific genes. Genome Res, 2003. 13(8): p. 1838-54.

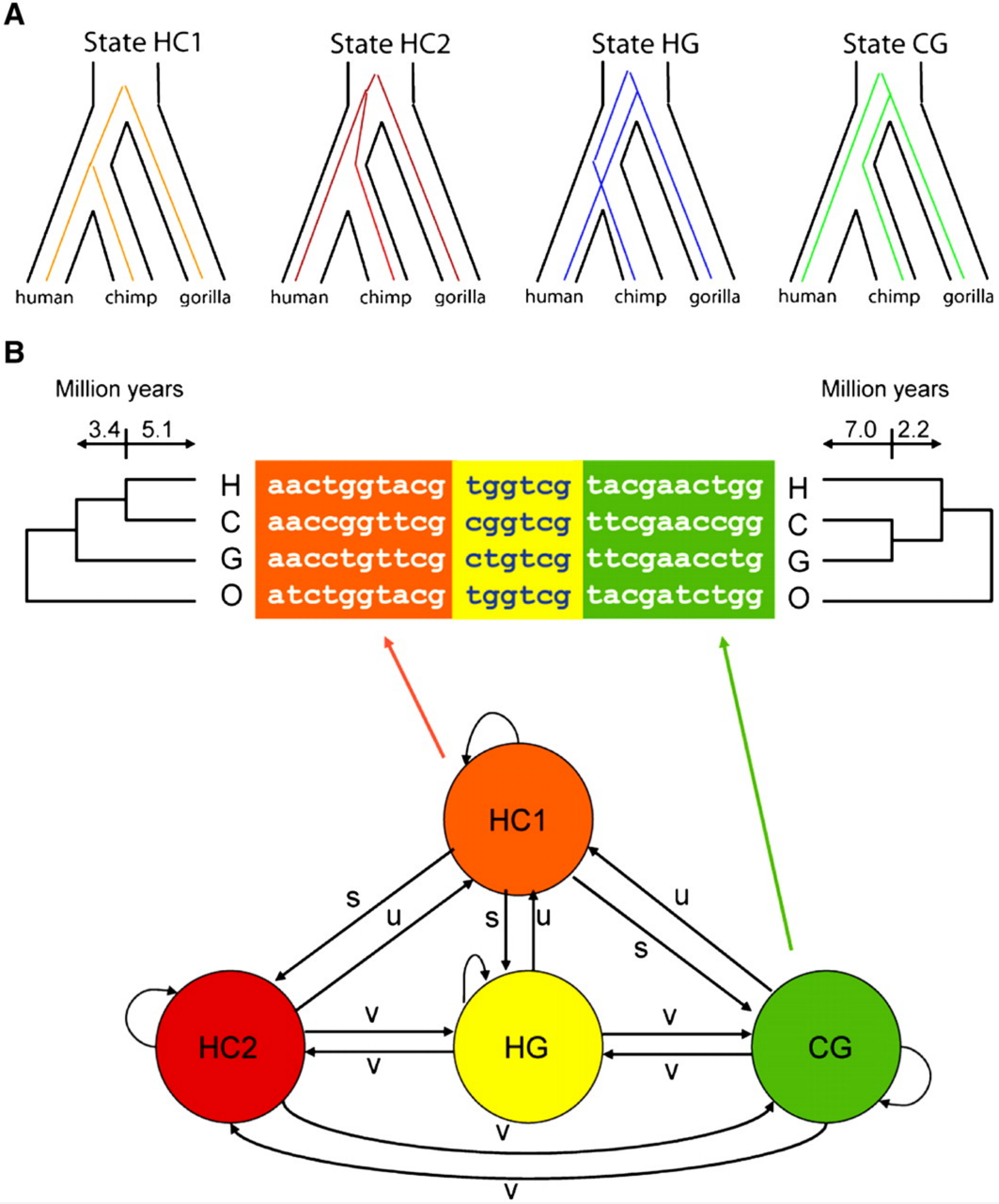


QUARTET DECOMPOSITION ANALYSIS OF CYANOBACTERIA.

Zhaxybayeva, O., et al., Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. Genome Res, 2006. 16(9): p. 1099-108.

● part 2
MAKING IT CLEAR

CONSISTENCY

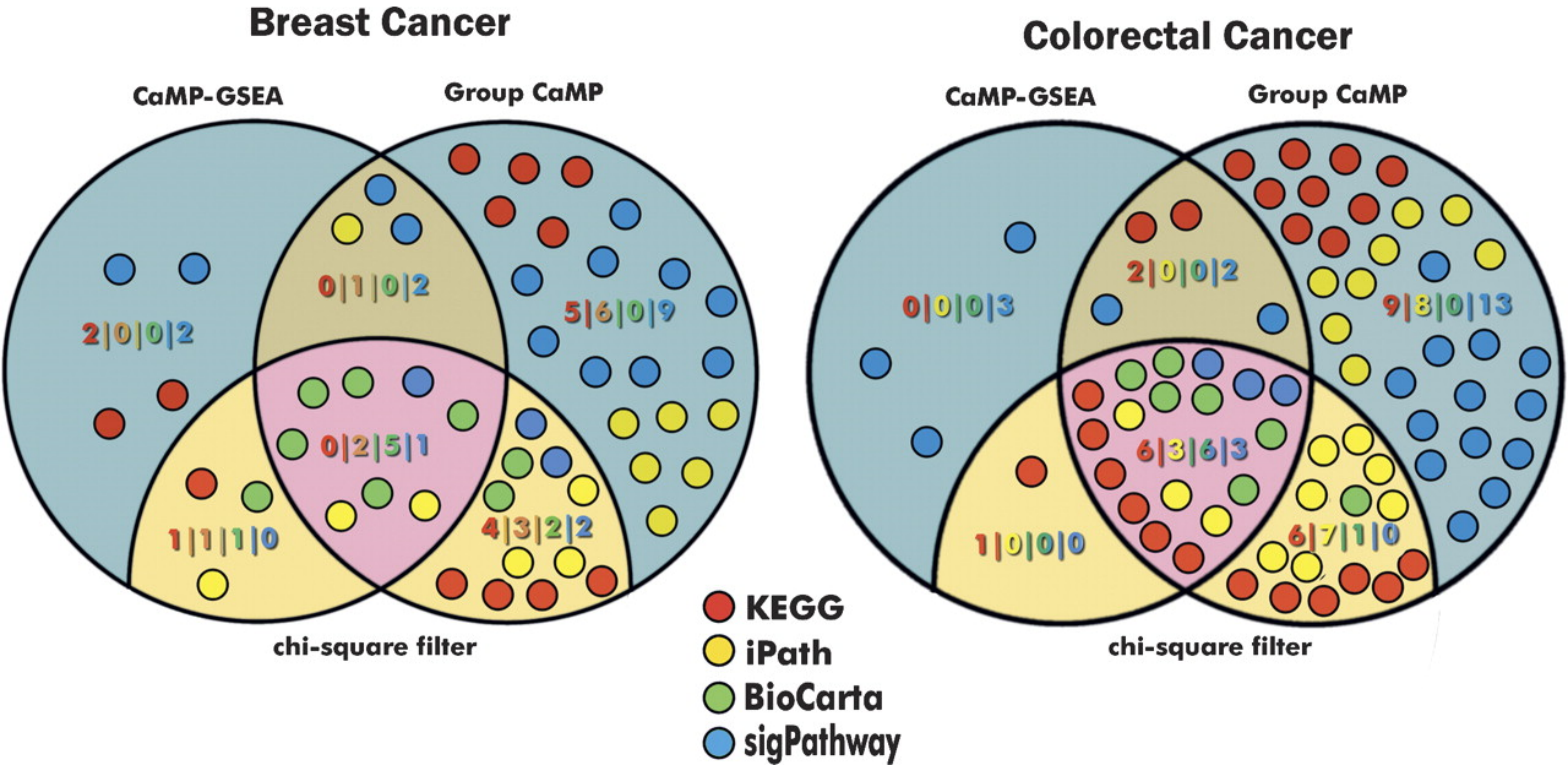


HG is blue in (A).

HG is yellow in (B).

THE FOUR GENEALOGY TYPES ASSOCIATED WITH THE STATES OF THE HIDDEN MARKOV MODEL.
THE STATE-TRANSITION DIAGRAM FOR THE HMM AND AN EXAMPLE ALIGNMENT.

Siepel, A., Phylogenomics of primates and their ancestral populations. Genome Res, 2009. 19(11): p. 1929-41.

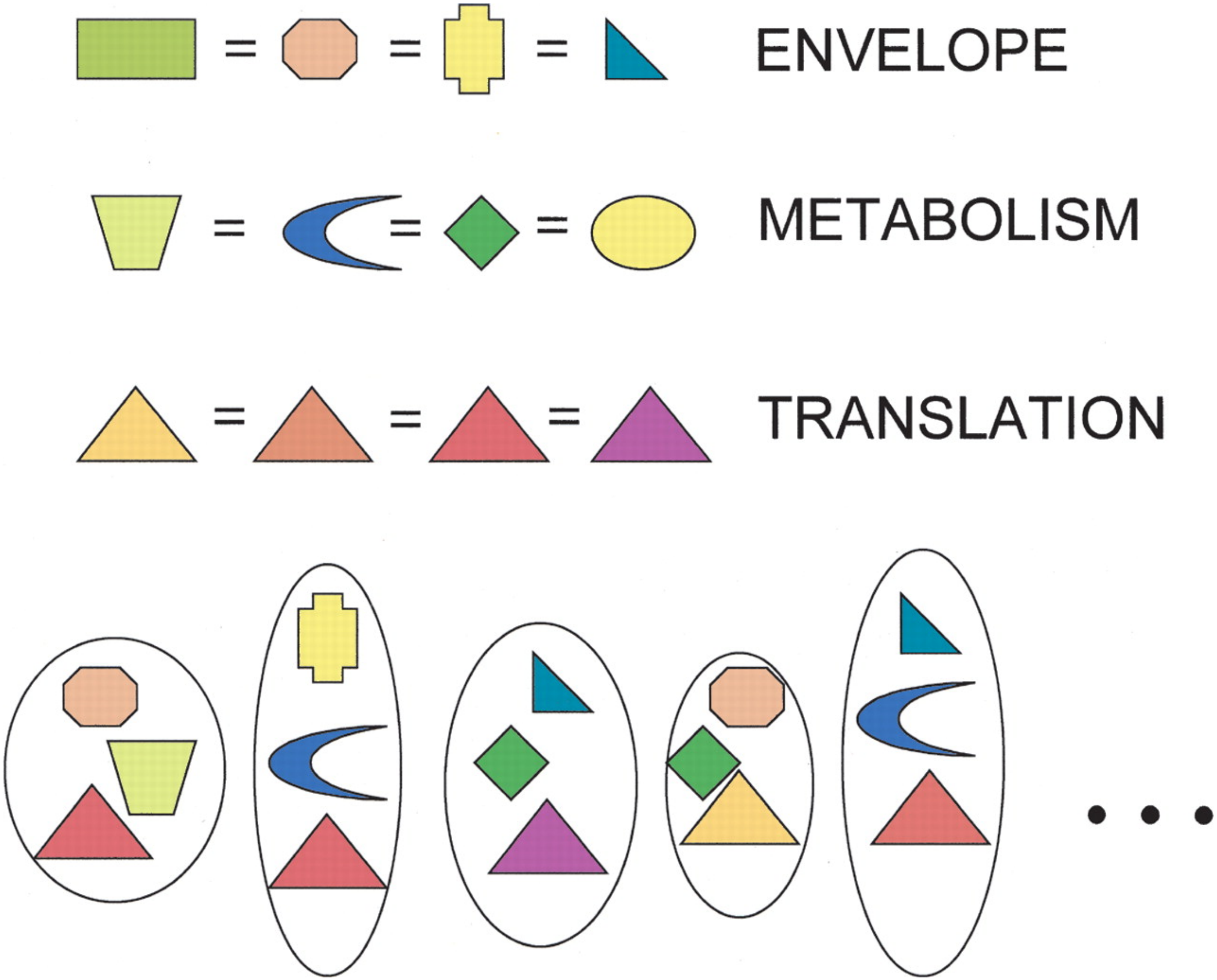


COMPARISON OF MUTATION ENRICHMENT IN CELLULAR PATHWAYS USING COMPLEMENTARY STATISTICAL APPROACHES.

Lin, J., et al., A multidimensional analysis of genes mutated in breast and colorectal cancers. Genome Res, 2007. 17(9): p. 1304-18.

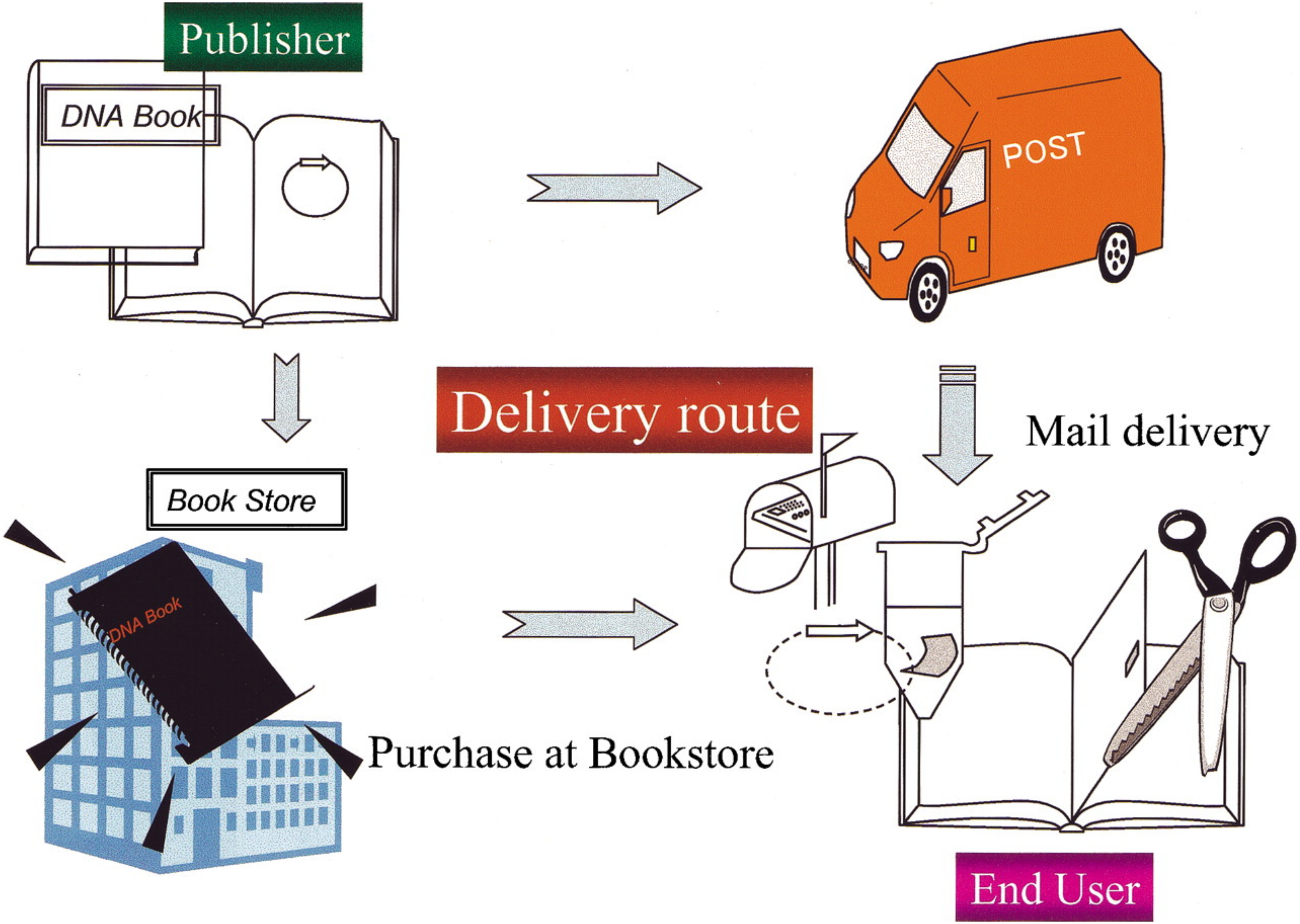
● part 2
MAKING IT CLEAR

MESSAGE



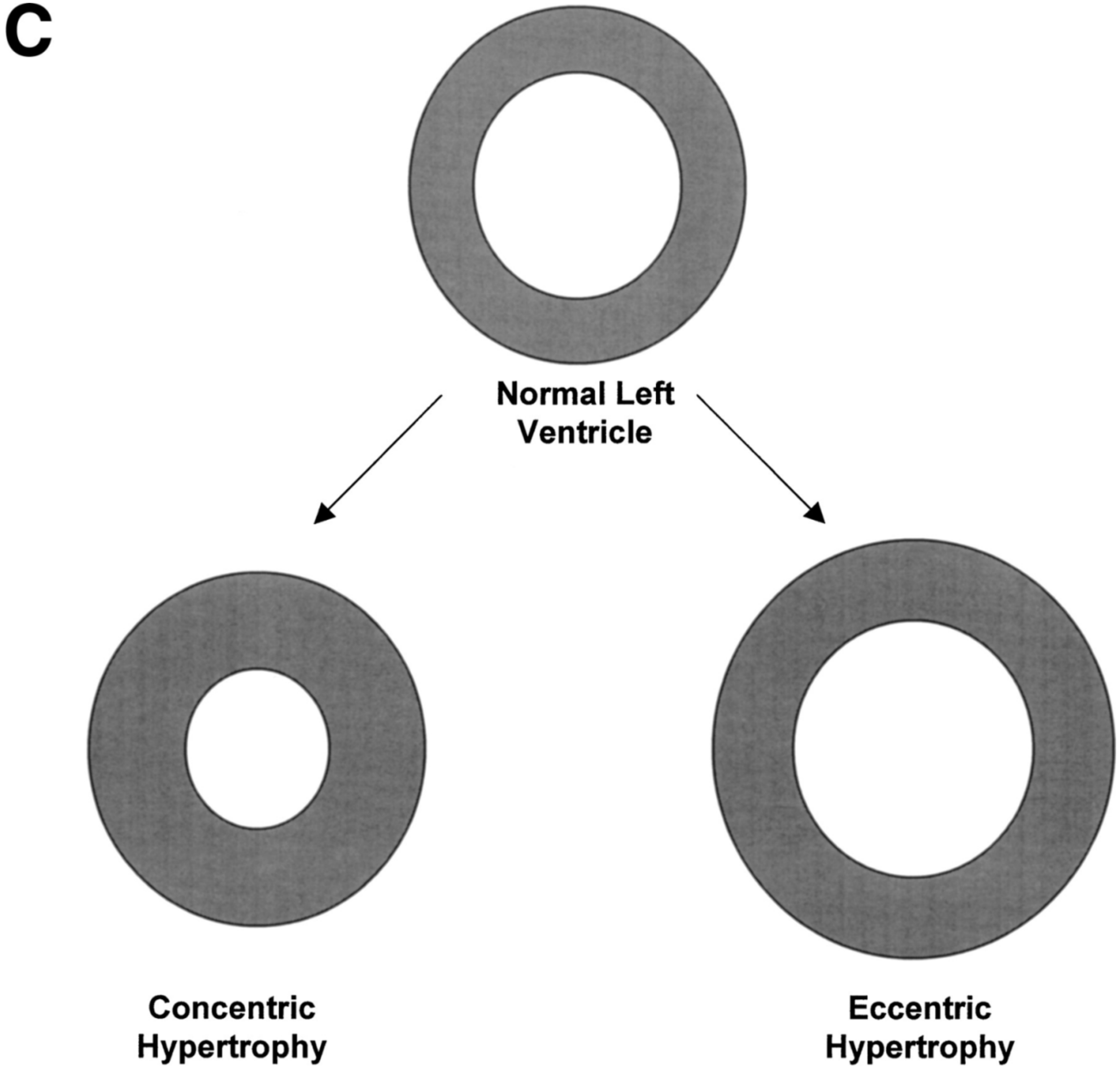
A MIX-AND-MATCH MODEL FOR PROKARYOTIC GENOME EVOLUTION.

Charlebois, R.L. and W.F. Doolittle, Computing prokaryotic gene ubiquity: rescuing the core from extinction. Genome Res, 2004. 14(12): p. 2469-77.



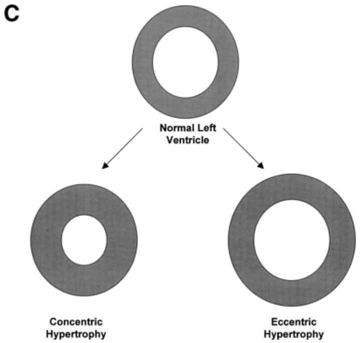
CONCEPT OF THE “DNA BOOK.”

Kawai, J. and Y. Hayashizaki, DNA book. Genome Res, 2003. 13(6B): p. 1488-95.



CONCENTRIC VS. ECCENTRIC HYPERTROPHY.

Nadeau, J.H., et al., Pleiotropy, homeostasis, and functional networks based on assays of cardiovascular traits in genetically randomized populations.
Genome Res, 2003. 13(9): p. 2082-91.



CONCENTRIC
HYPERTROPHY

NORMAL

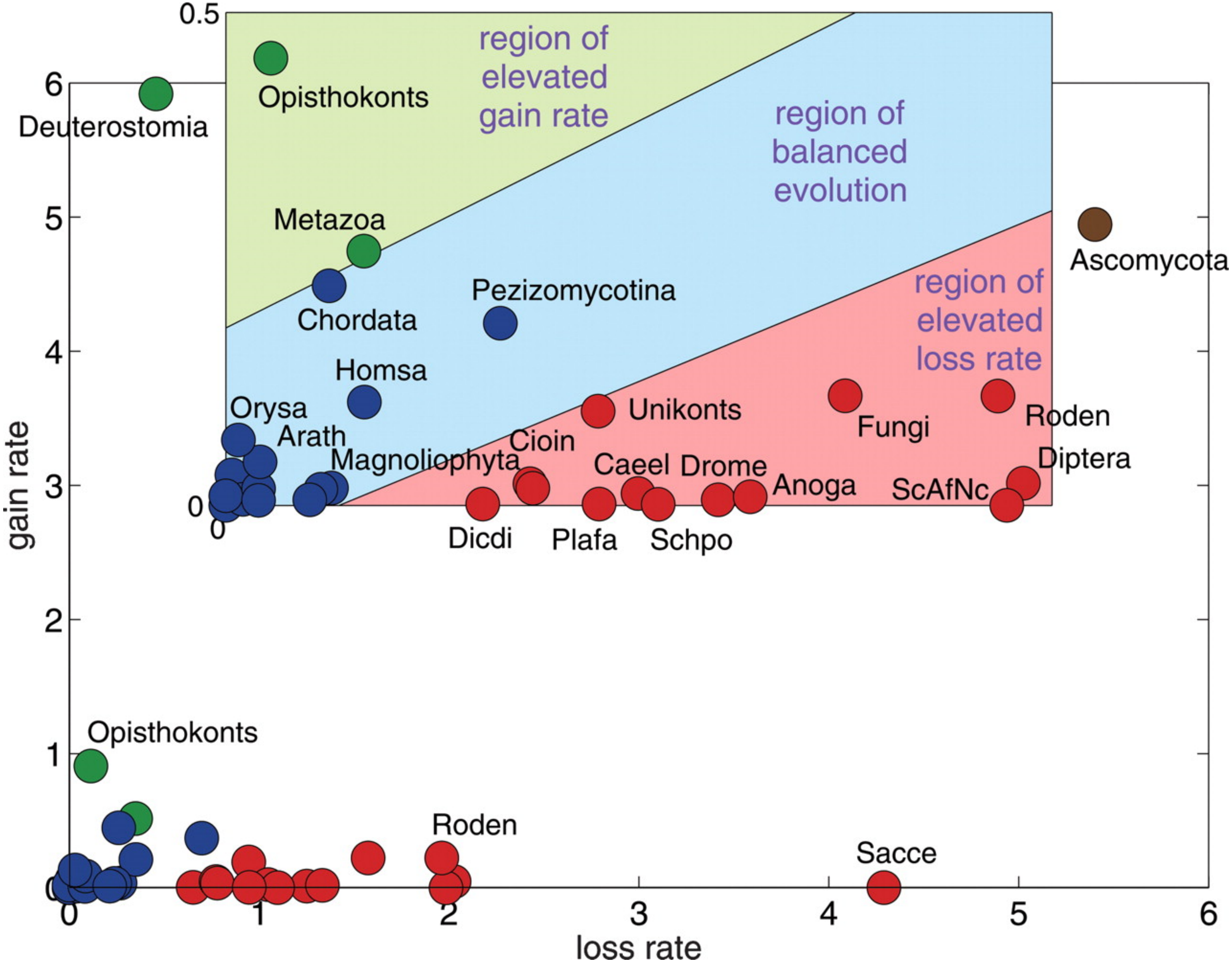
ECCENTRIC
HYPERTROPHY

CONCENTRIC VS. ECCENTRIC HYPERTROPHY.
FIGURE MODIFIED FROM ORIGINAL

Nadeau, J.H., et al., Pleiotropy, homeostasis, and functional networks based on assays of cardiovascular traits in genetically randomized populations. Genome Res, 2003. 13(9): p. 2082-91.

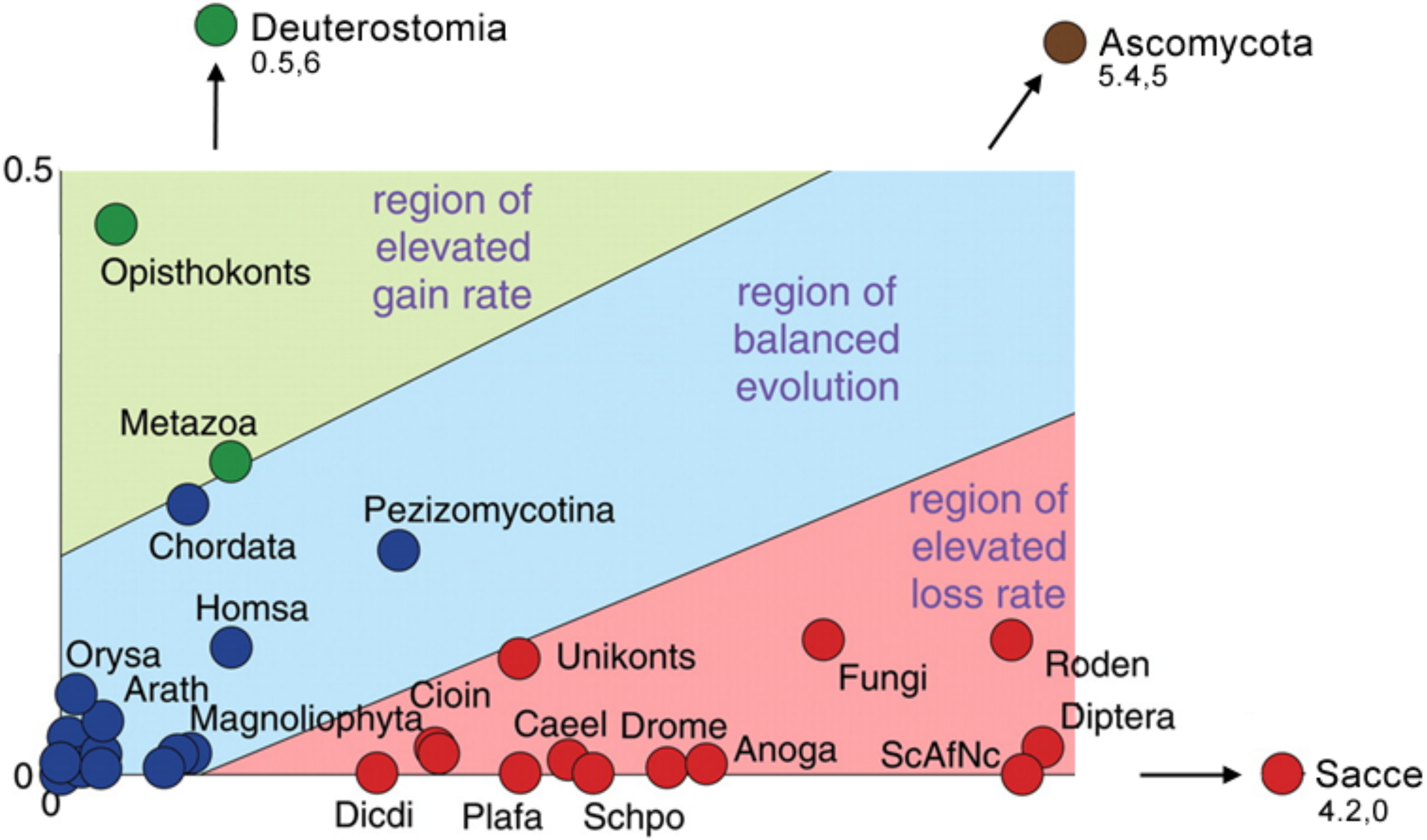
● part 2
MAKING IT CLEAR

OUTLIERS



INTRON GAIN AND LOSS RATES OF EUKARYOTIC LINEAGES.

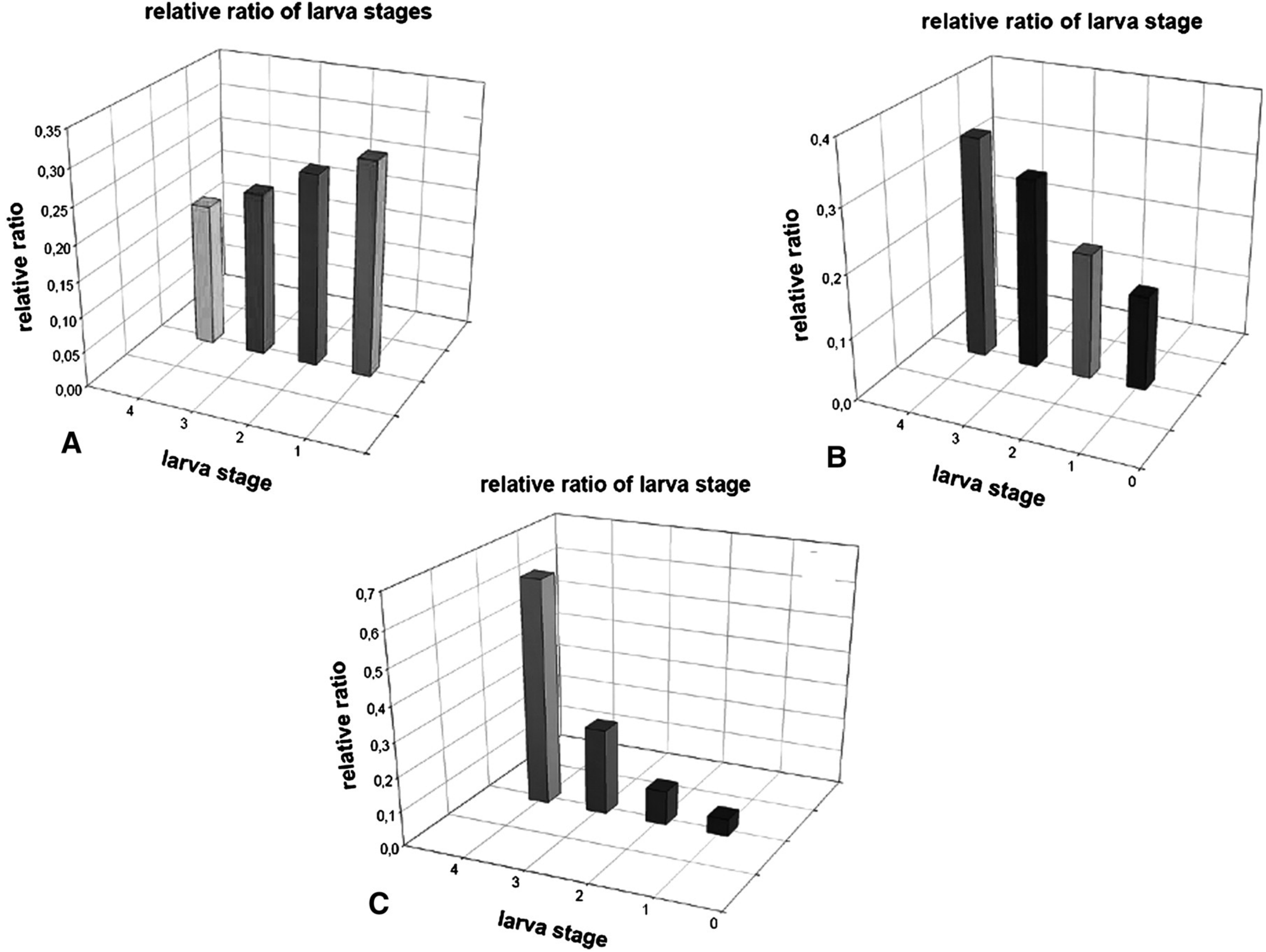
Carmel, L., et al., Three distinct modes of intron dynamics in the evolution of eukaryotes. Genome Res, 2007. 17(7): p. 1034-44.



INTRON GAIN AND LOSS RATES OF EUKARYOTIC LINEAGES.
REINTERPRETED FIGURE

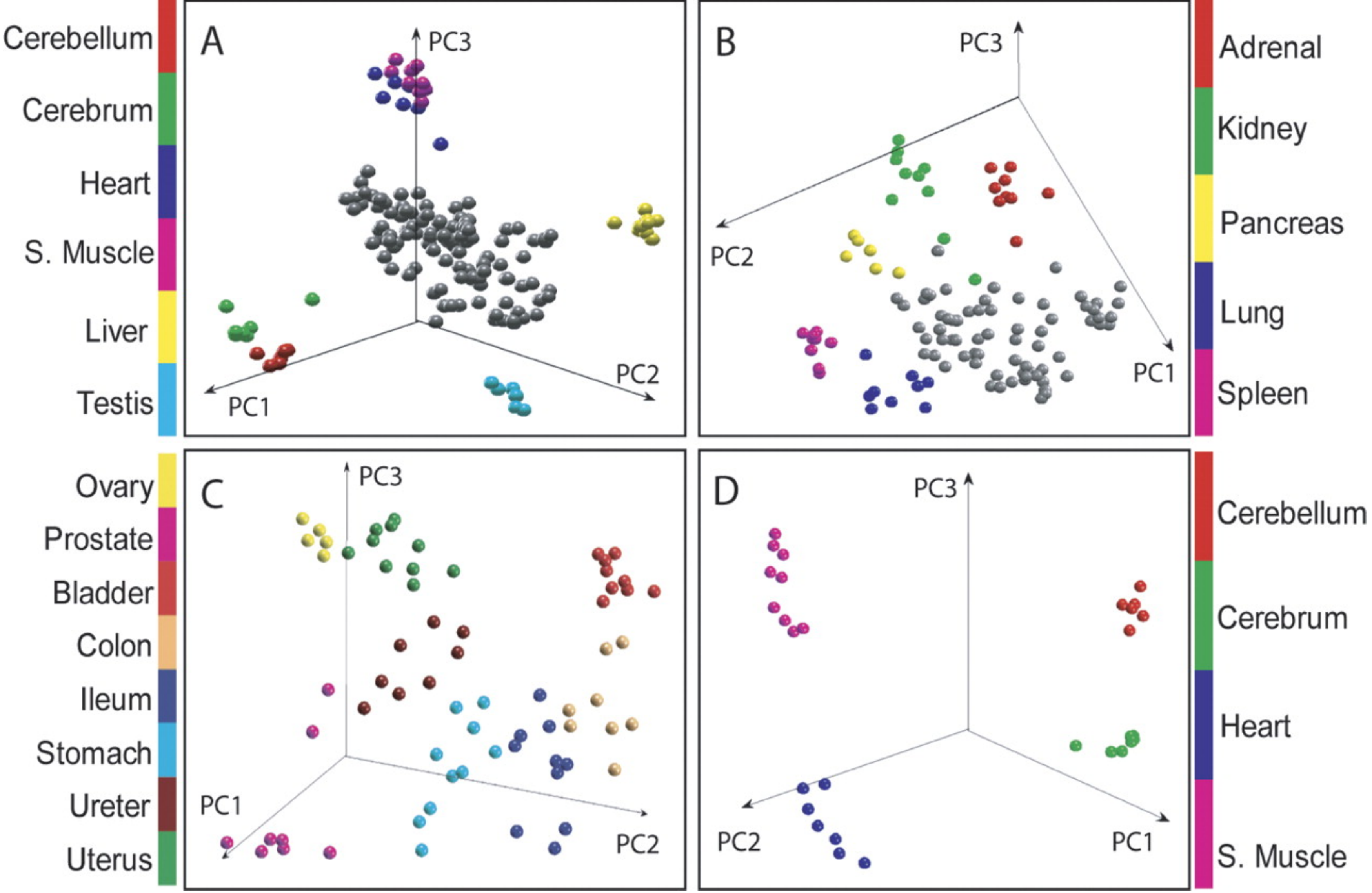
Carmel, L., et al., Three distinct modes of intron dynamics in the evolution of eukaryotes. Genome Res, 2007. 17(7): p. 1034-44.

NO 3D



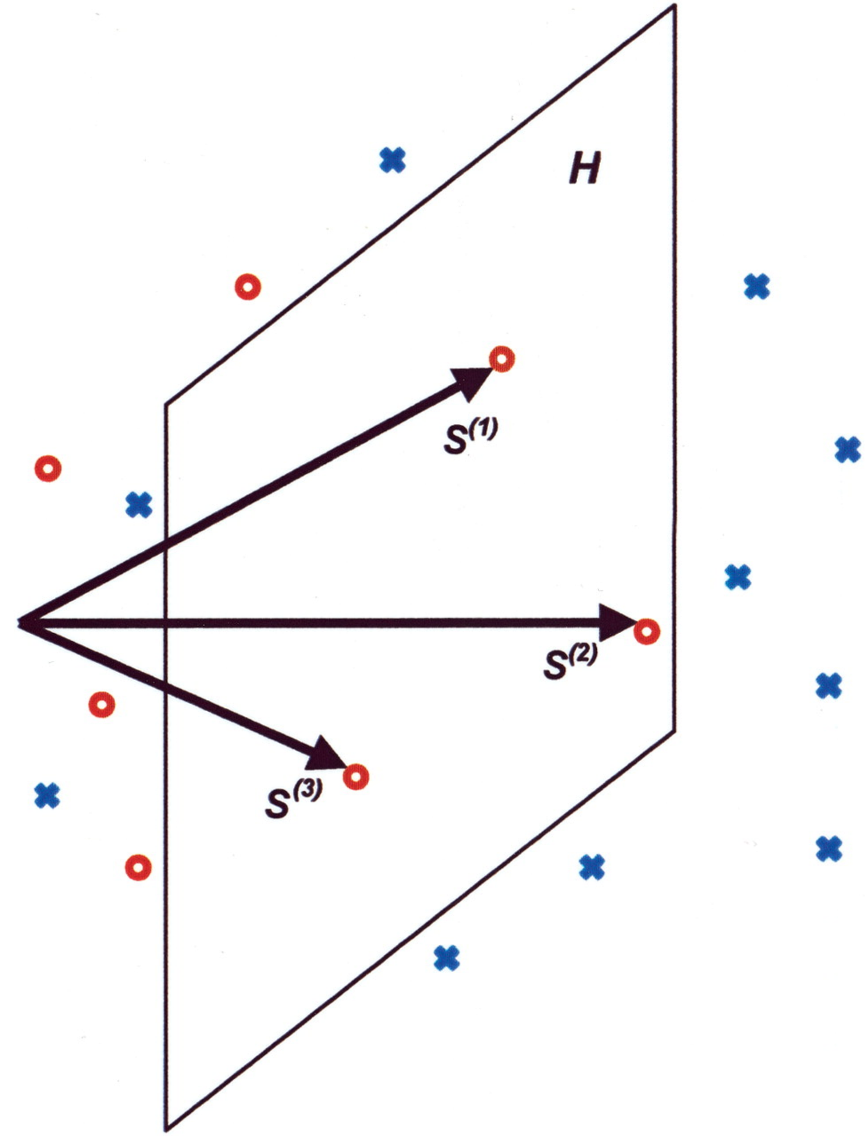
DNA METHYL TRANSFERASE INHIBITORS EFFECTS ON DEVELOPMENTAL STAGES DISTRIBUTION.

Dombrovsky, A., et al., Profiling the repertoire of phenotypes influenced by environmental cues that occur during asexual reproduction.
Genome Res, 2009. 19(11): p. 2052-63.



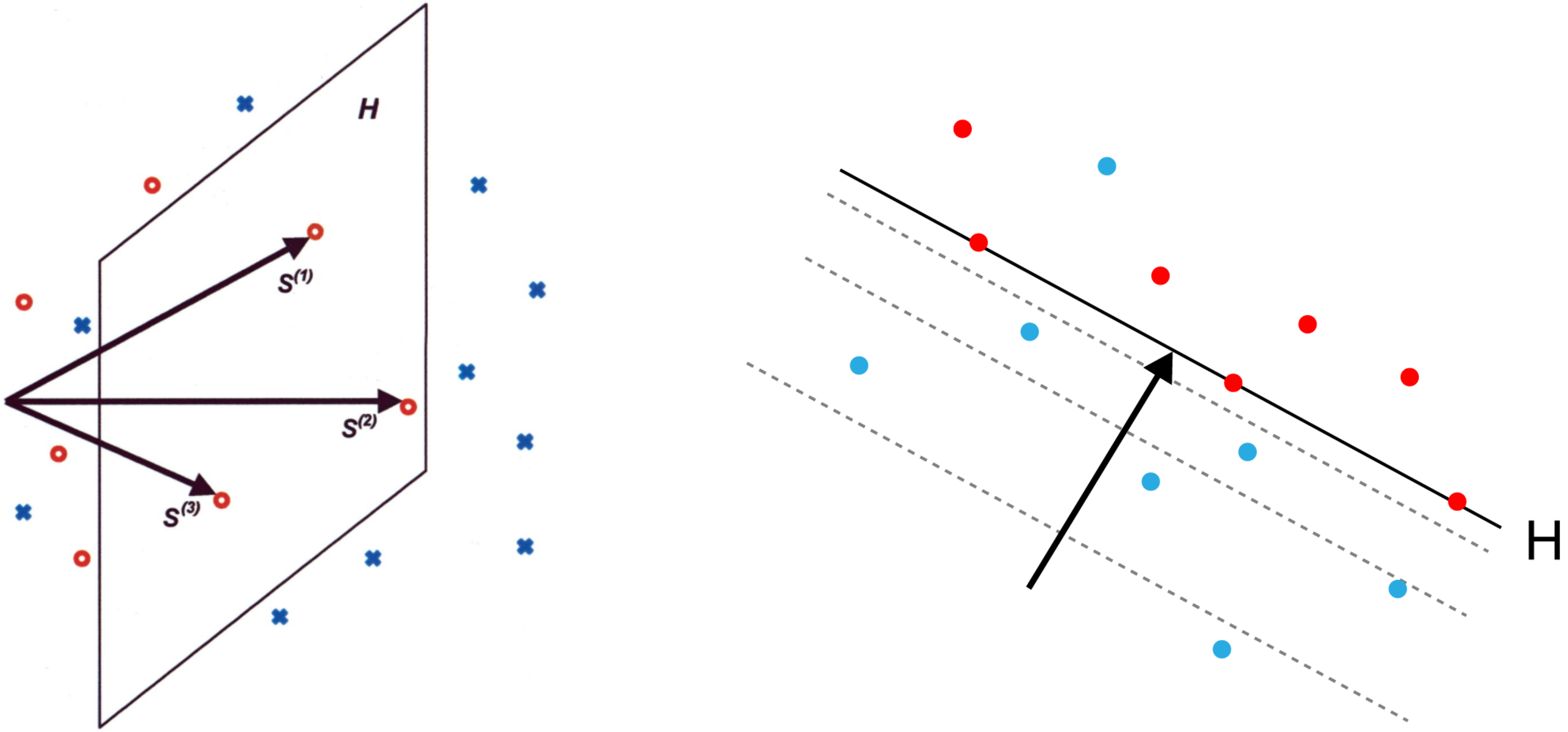
PRINCIPAL COMPONENTS ANALYSIS.

Son, C.G., et al., Database of mRNA gene expression profiles of multiple human organs. Genome Res, 2005. 15(3): p. 443-50.



SCHEMATIC REPRESENTATION OF THE QUADRATIC PROGRAMMING PROBLEM.

Djordjevic, M., A.M. Sengupta, and B.I. Shraiman, A biophysical approach to transcription factor binding site discovery. Genome Res, 2003. 13(11): p. 2381-90.

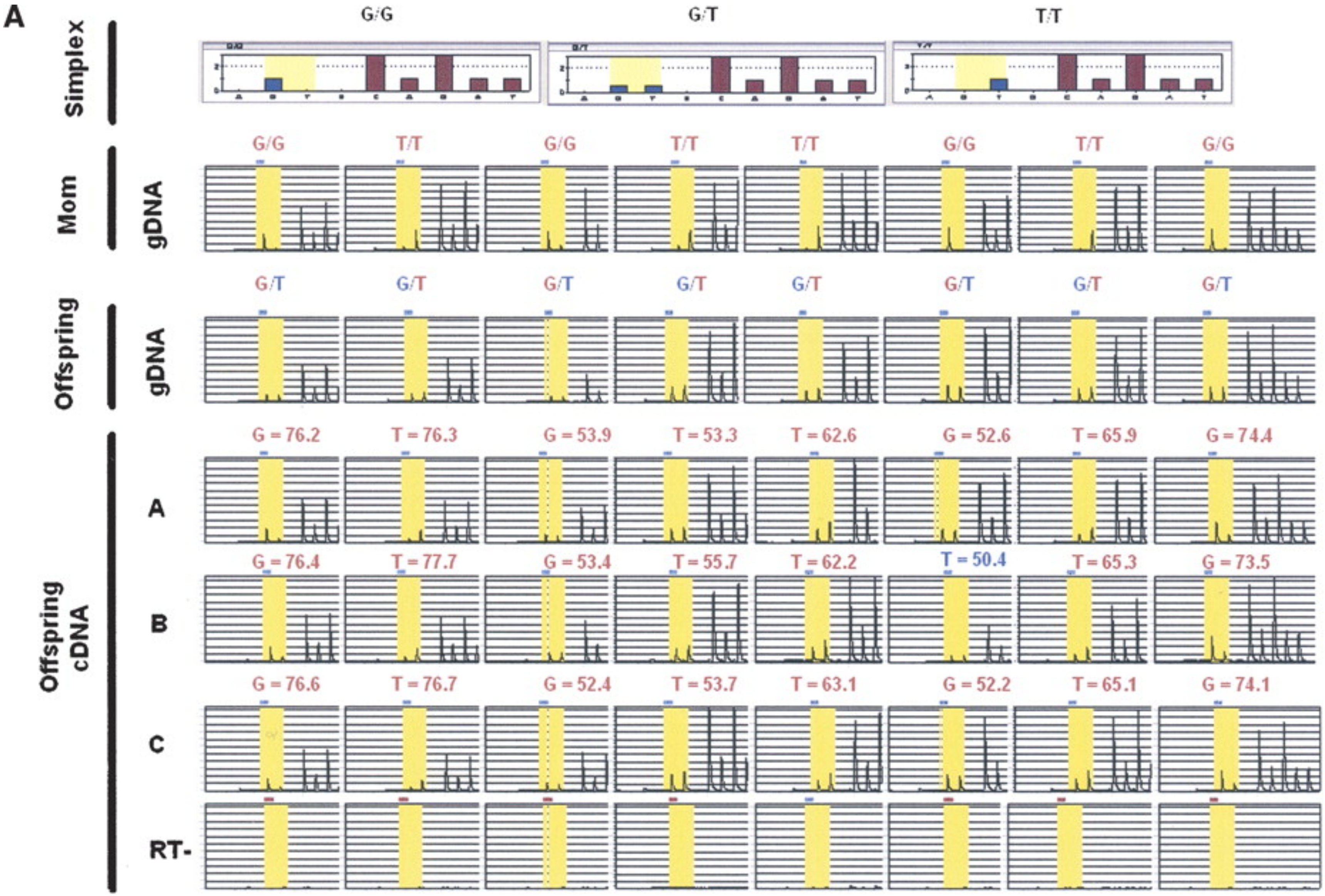


SCHEMATIC REPRESENTATION OF THE QUADRATIC PROGRAMMING PROBLEM.
FIGURE MODIFIED FROM ORIGINAL

Djordjevic, M., A.M. Sengupta, and B.I. Shraiman, A biophysical approach to transcription factor binding site discovery. Genome Res, 2003. 13(11): p. 2381-90.

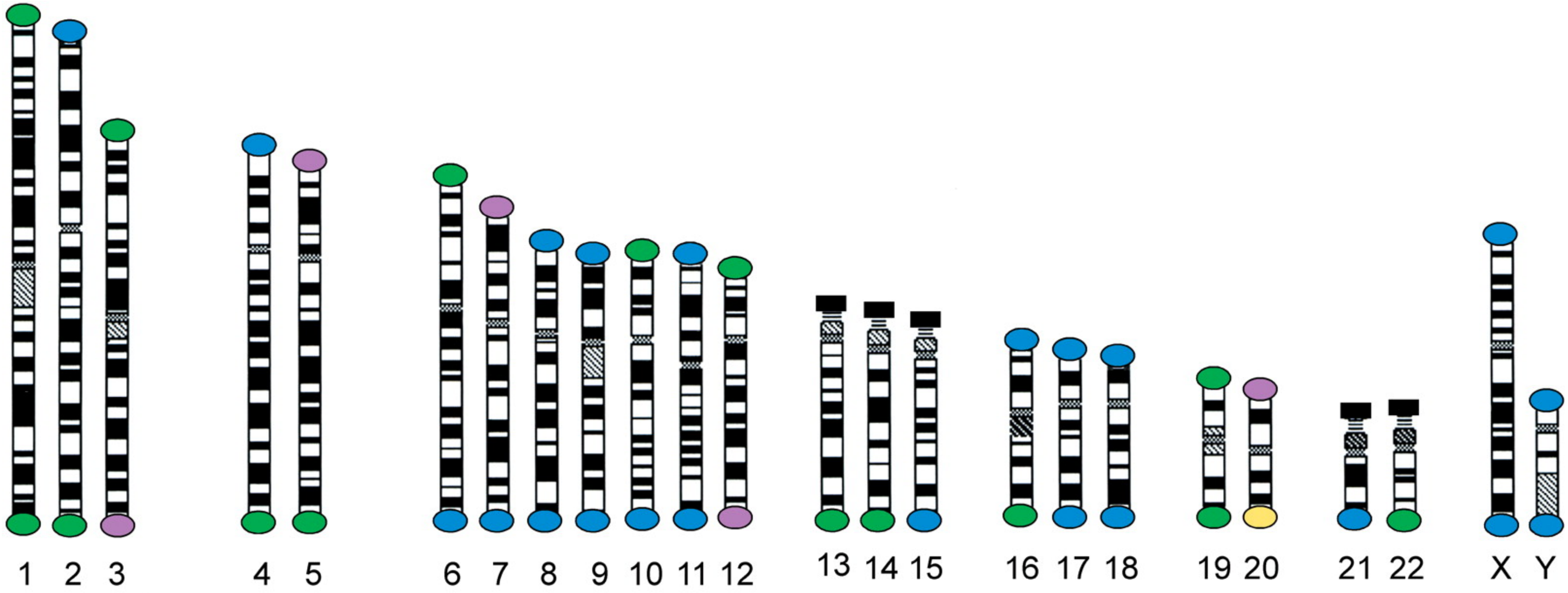
● part 2
MAKING IT CLEAR

EXCESS INK



PARENT-OF-ORIGIN ANALYSIS OF GENES SHOWING ASE IN A REPLICATE SET.

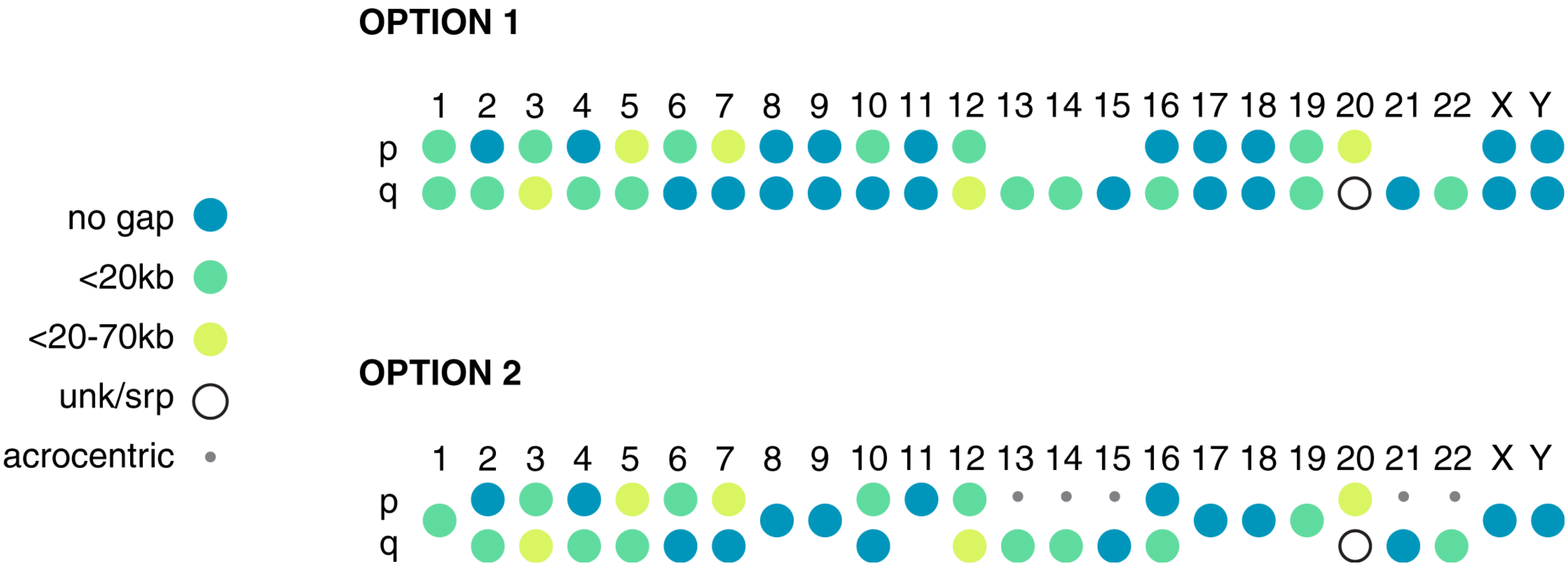
Bjornsson, H.T., et al., SNP-specific array-based allele-specific expression analysis. Genome Res, 2008. 18(5): p. 771-9.



🔵 = 0kb
🟢 < 20kb
🟣 < 20-70kb
🟡 = unk, srpt

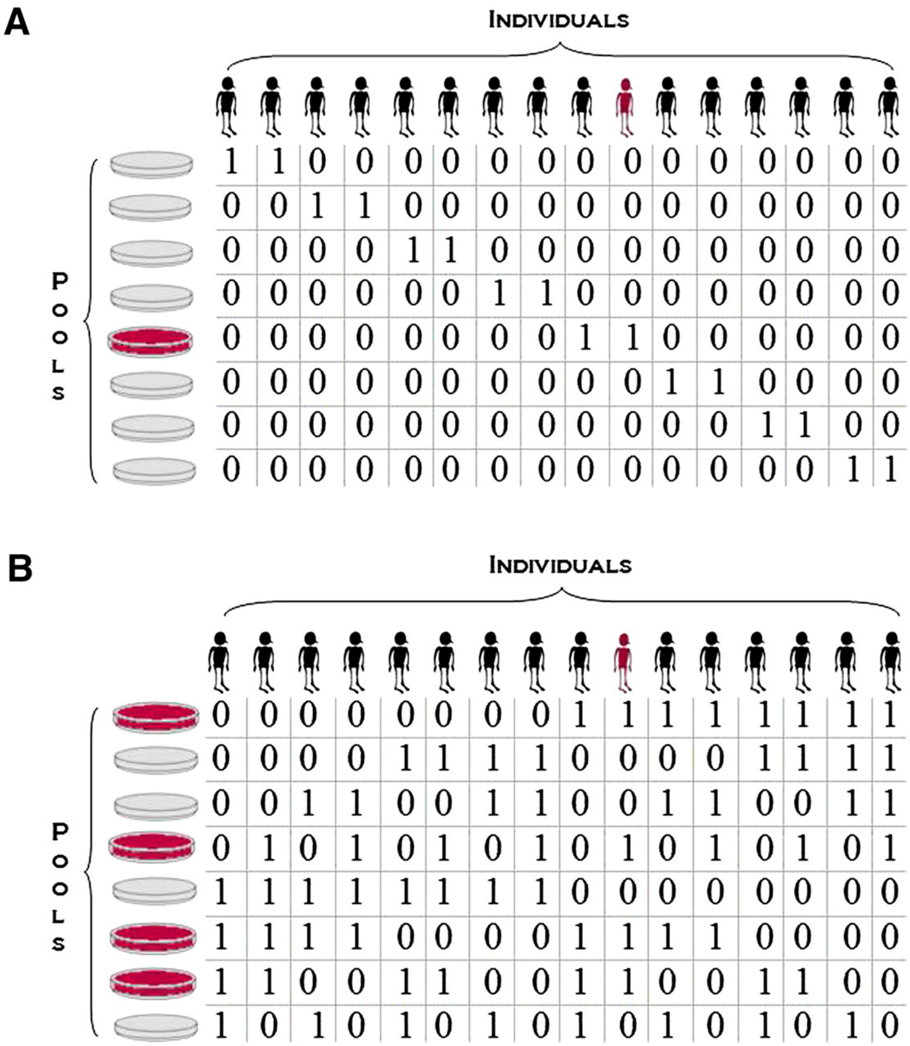
TELOMERE SEQUENCE GAPS.

Riethman, H., et al., Mapping and initial analysis of human subtelomeric sequence assemblies. Genome Res, 2004. 14 (1): p. 18-28.



TELOMERE SEQUENCE GAPS.
FIGURE MODIFIED FROM THE ORIGINAL

1. Riethman, H., et al., Mapping and initial analysis of human subtelomeric sequence assemblies. Genome Res, 2004. 14(1): p. 18-28.



A

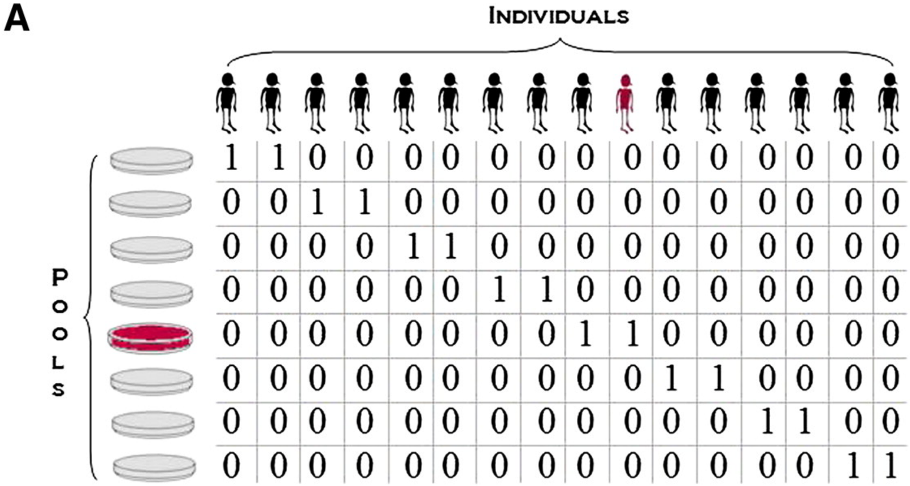
⬜	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
⬜	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
⬜	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
⬜	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
🟥	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
⬜	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
⬜	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
⬜	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

B

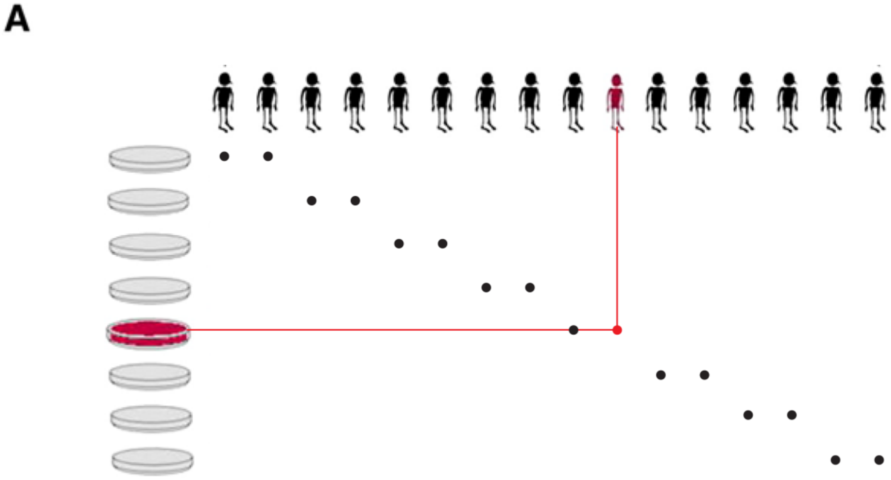
🟥	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
⬜	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
⬜	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
🟥	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
⬜	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
🟥	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
🟥	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
⬜	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0

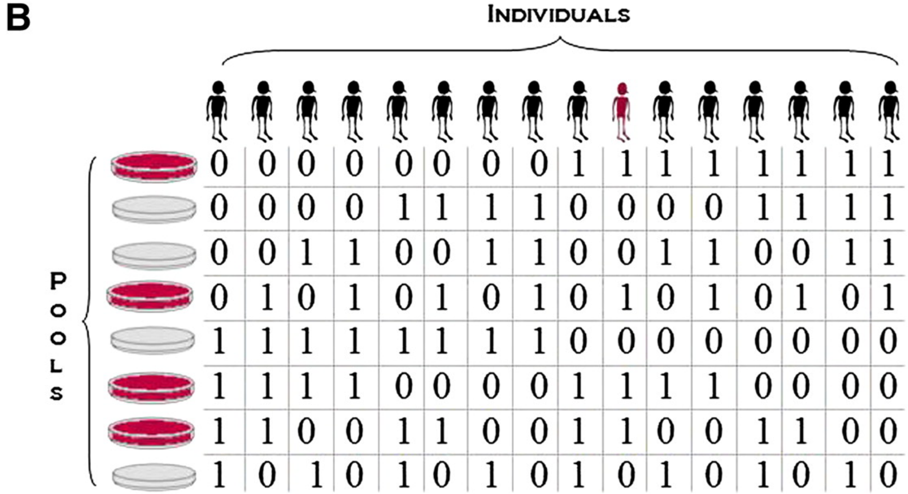
RESEQUENCING WITH NAÏVE AND LOG POOL DESIGNS.

Prabhu, S. and I. Pe'er, Overlapping pools for high-throughput targeted resequencing. Genome Res, 2009. 19(7): p. 1254-61.

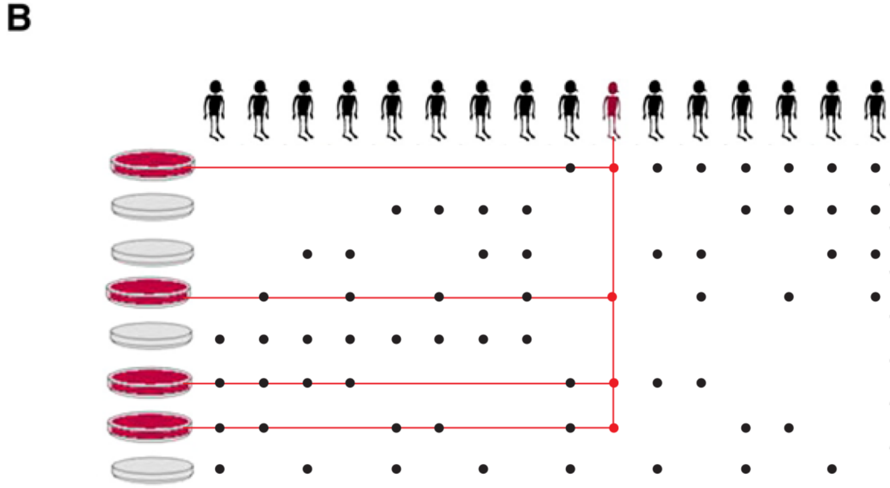


1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1



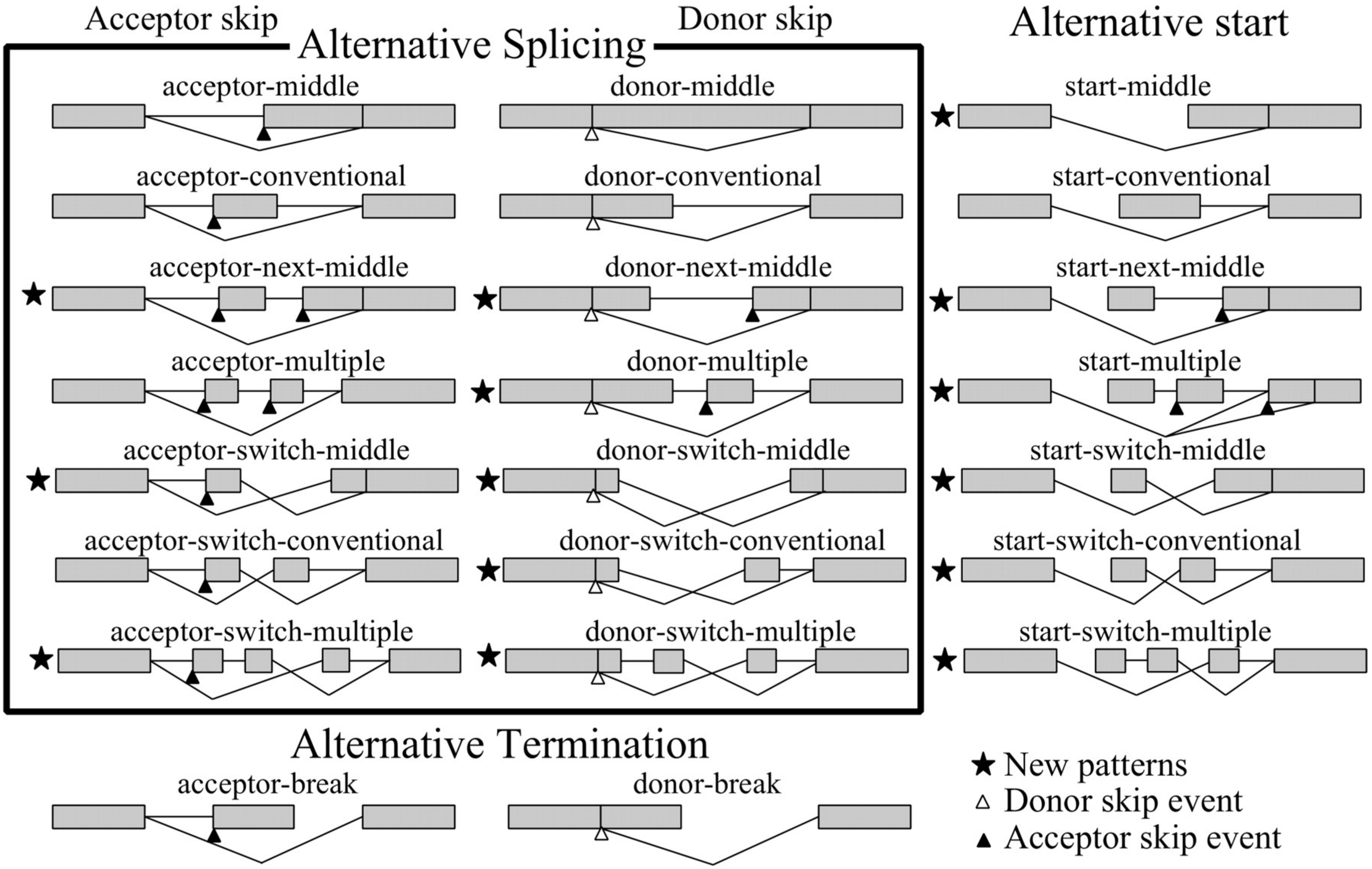


0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0



RESEQUENCING WITH NAÏVE AND LOG POOL DESIGNS.

Prabhu, S. and I. Pe'er, Overlapping pools for high-throughput targeted resequencing. Genome Res, 2009. 19(7): p. 1254-61.

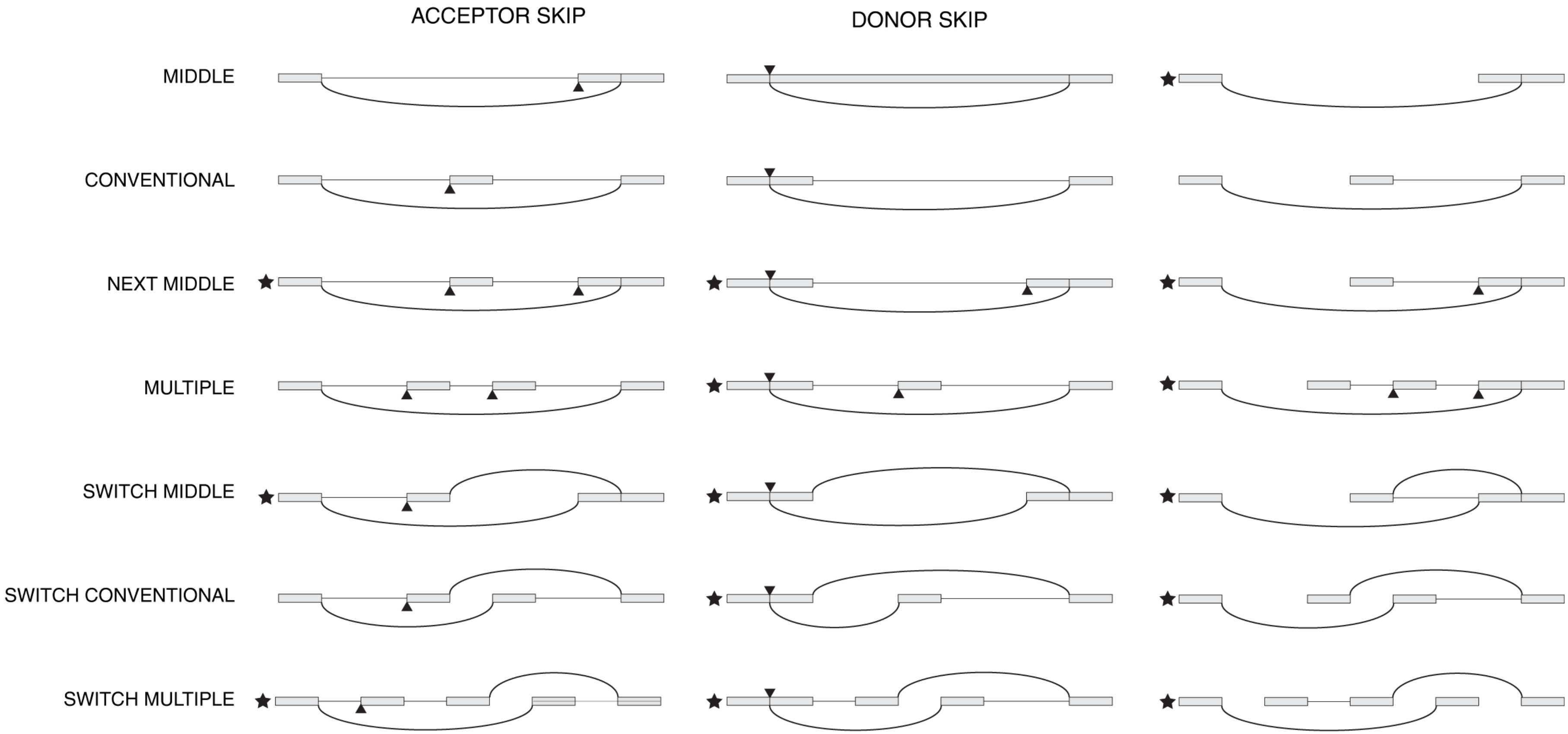


COMBINATORIAL CLASSIFICATION OF ATS UNITS.

Sharov, A.A., D.B. Dudekula, and M.S. Ko, Genome-wide assembly and analysis of alternative transcripts in mouse. Genome Res, 2005. 15(5): p. 748-54.

ALTERNATIVE SPLICING

ALTERNATIVE START



ALTERNATIVE TERMINATION

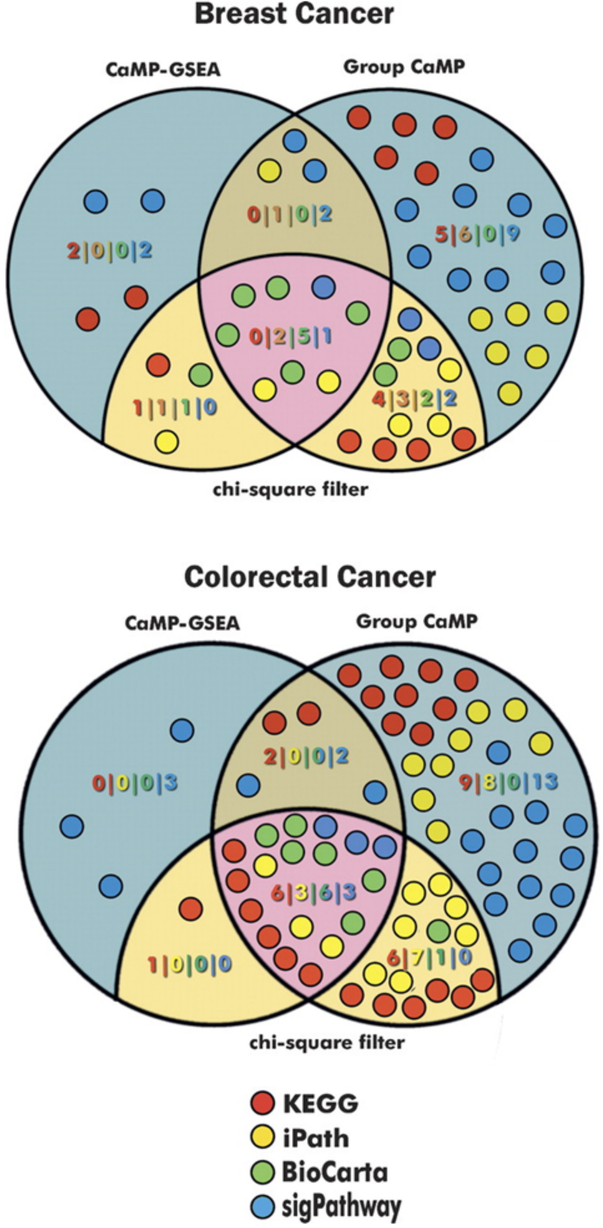
★ new patterns
▲ acceptor skip event
▼ donor skip event

COMBINATORIAL CLASSIFICATION OF ATS UNITS.
FIGURE REFORMATTED FROM ORIGINAL

Sharov, A.A., D.B. Dudekula, and M.S. Ko, Genome-wide assembly and analysis of alternative transcripts in mouse. Genome Res, 2005. 15(5): p. 748-54.

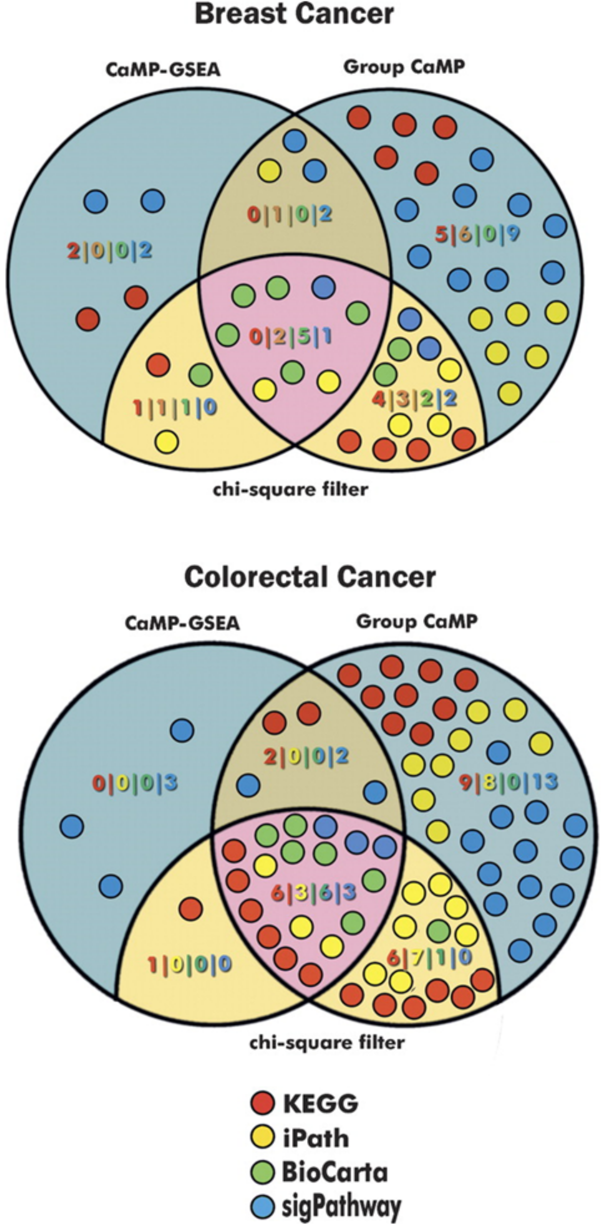
● part 2
MAKING IT CLEAR

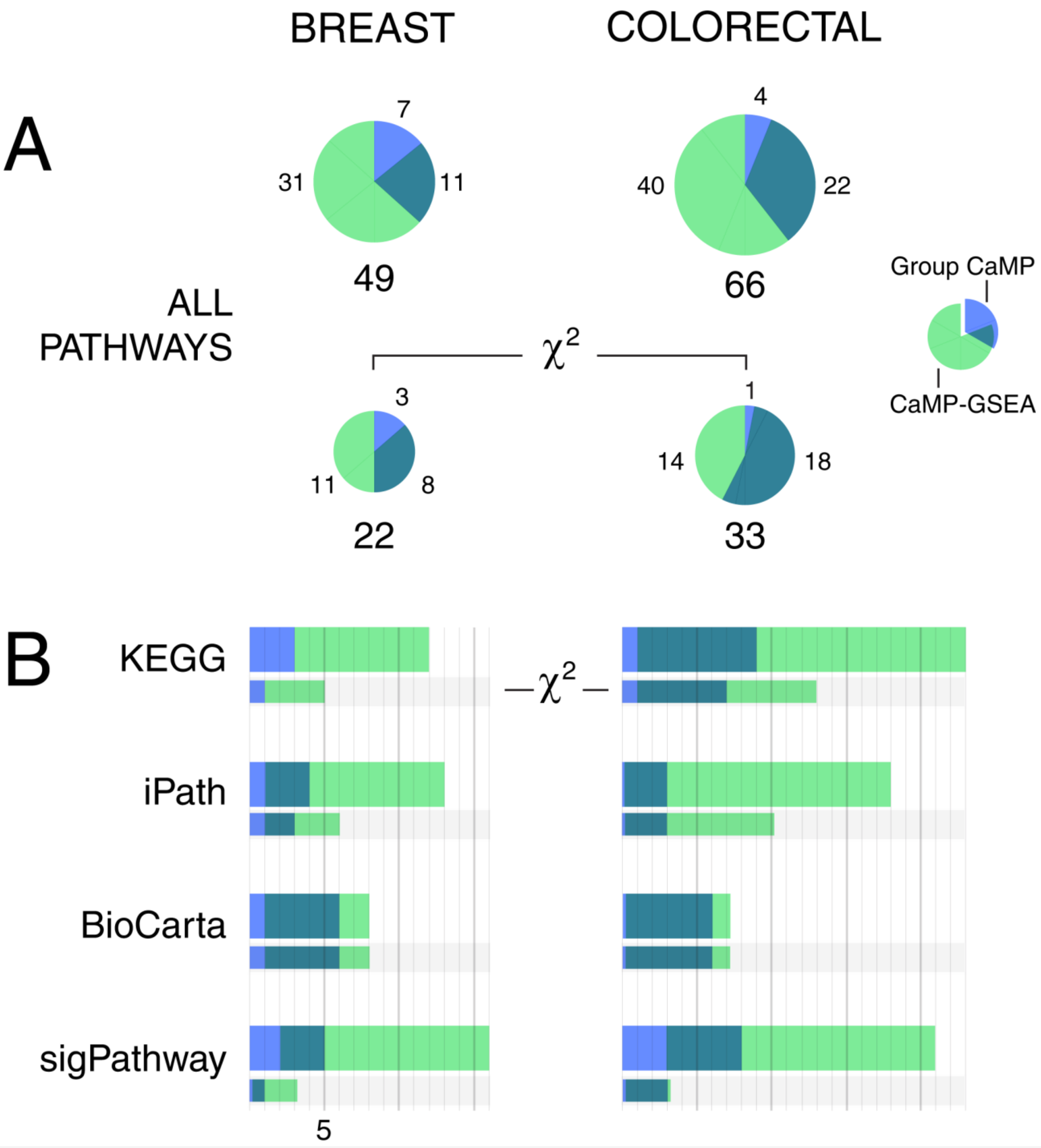
REPRESENTATION



COMPARISON OF MUTATION ENRICHMENT IN CELLULAR PATHWAYS USING COMPLEMENTARY STATISTICAL APPROACHES.

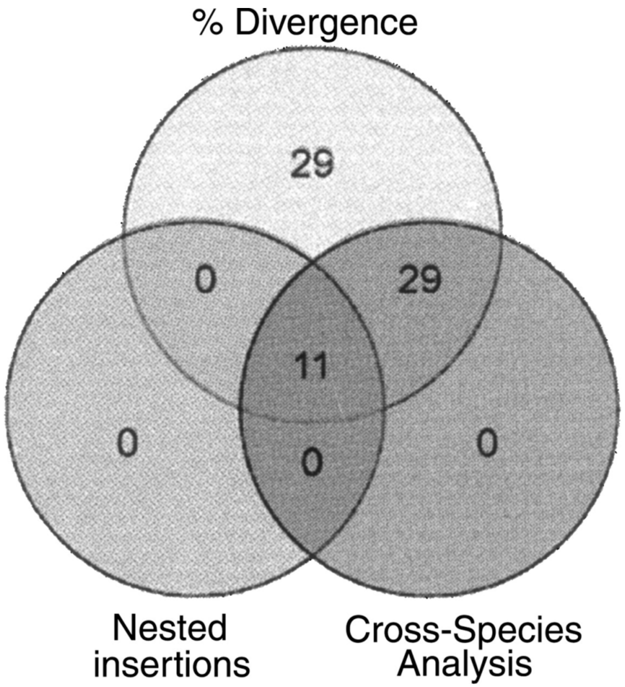
Lin, J., et al., A multidimensional analysis of genes mutated in breast and colorectal cancers. Genome Res, 2007. 17(9): p. 1304-18.





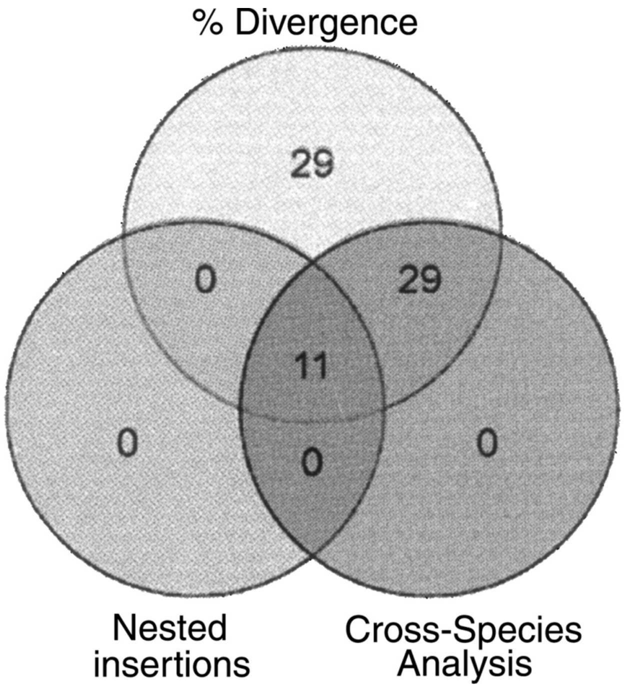
COMPARISON OF MUTATION ENRICHMENT IN CELLULAR PATHWAYS USING COMPLEMENTARY STATISTICAL APPROACHES.

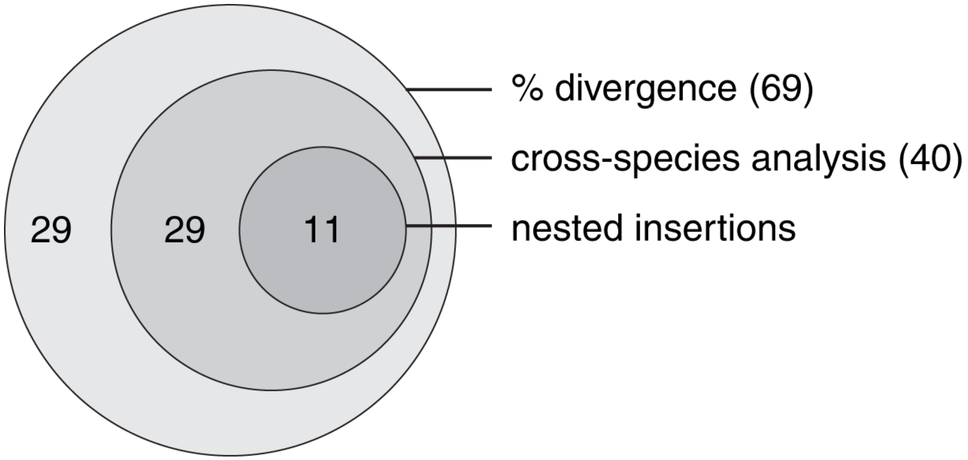
Lin, J., et al., A multidimensional analysis of genes mutated in breast and colorectal cancers. Genome Res, 2007. 17(9): p. 1304-18.



COMPARISON OF THREE INDEPENDENT METHODS FOR DATING DNA TRANSPOSONS.

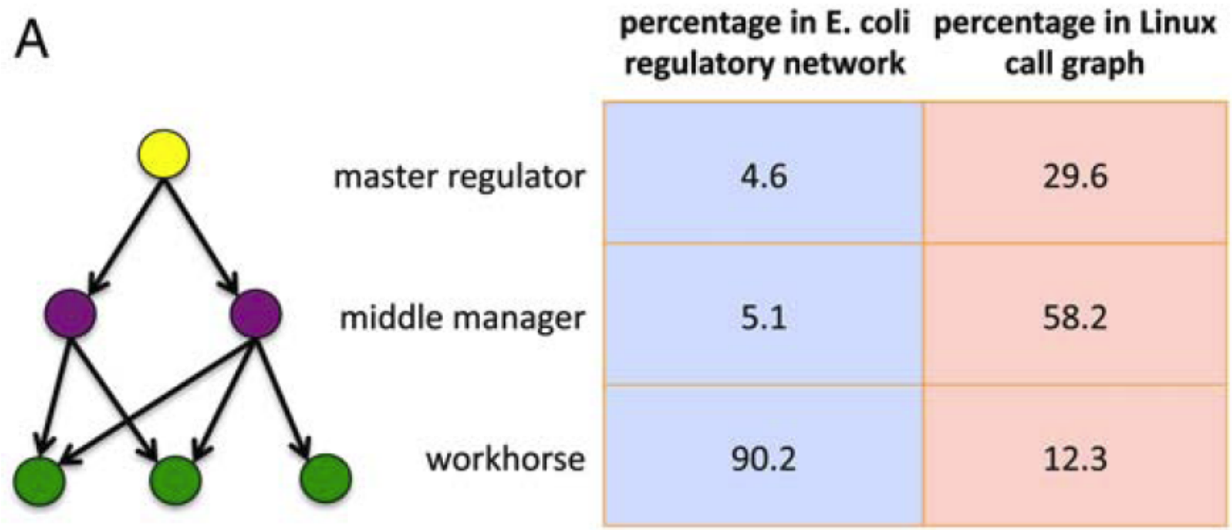
Pace, J.K., 2nd and C. Feschotte, The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage.
Genome Res, 2007. 17(4): p. 422-32.





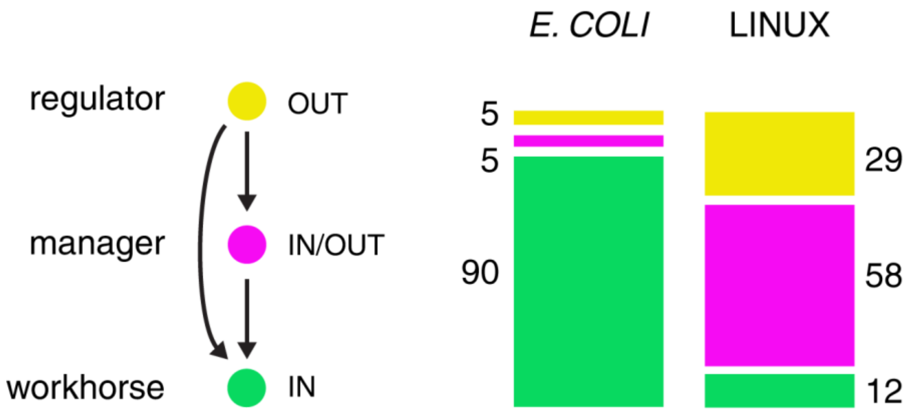
COMPARISON OF THREE INDEPENDENT METHODS FOR DATING DNA TRANSPOSONS.

Pace, J.K., 2nd and C. Feschotte, The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage.
Genome Res, 2007. 17(4): p. 422-32.



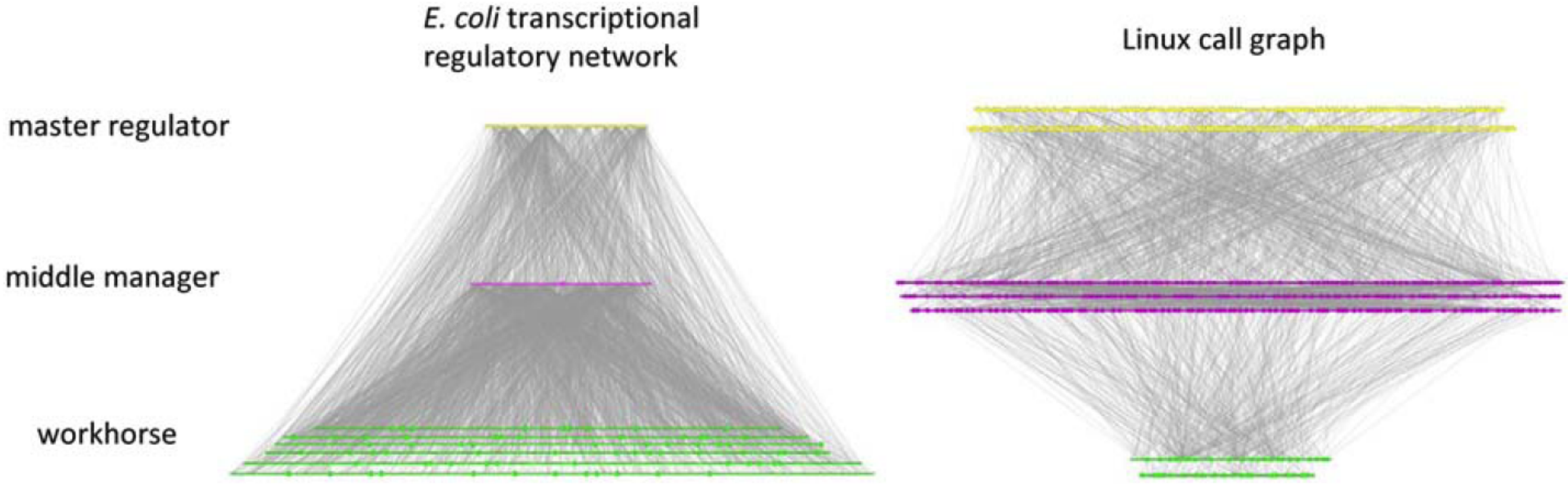
	percentage in E. coli regulatory network	percentage in Linux call graph
master regulator	4.6	29.6
middle manager	5.1	58.2
workhorse	90.2	12.3

NODE CLASSIFICATION



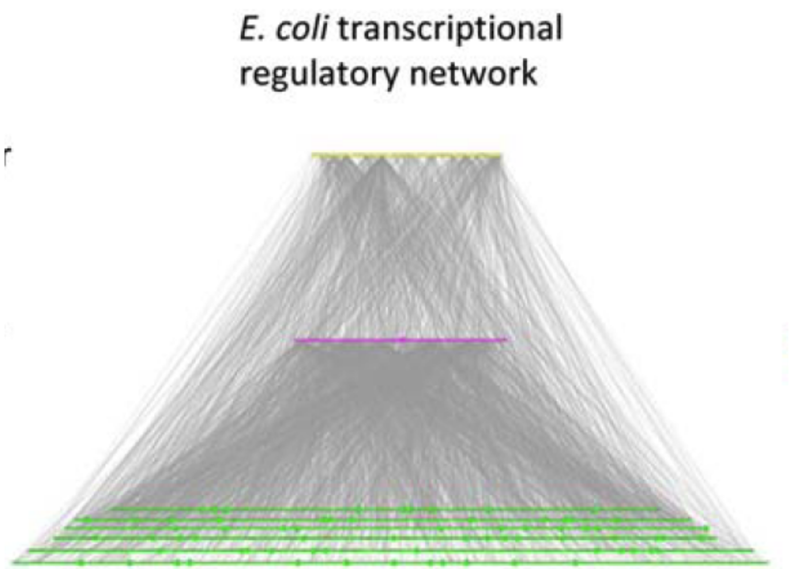
THE HIERARCHICAL LAYOUT OF THE E. COLI TRANSCRIPTIONAL REGULATORY NETWORK AND THE LINUX CALL GRAPH.

Yan, K.K., et al., Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. Proc Natl Acad Sci U S A, 2010. 107(20): p. 9186-91.

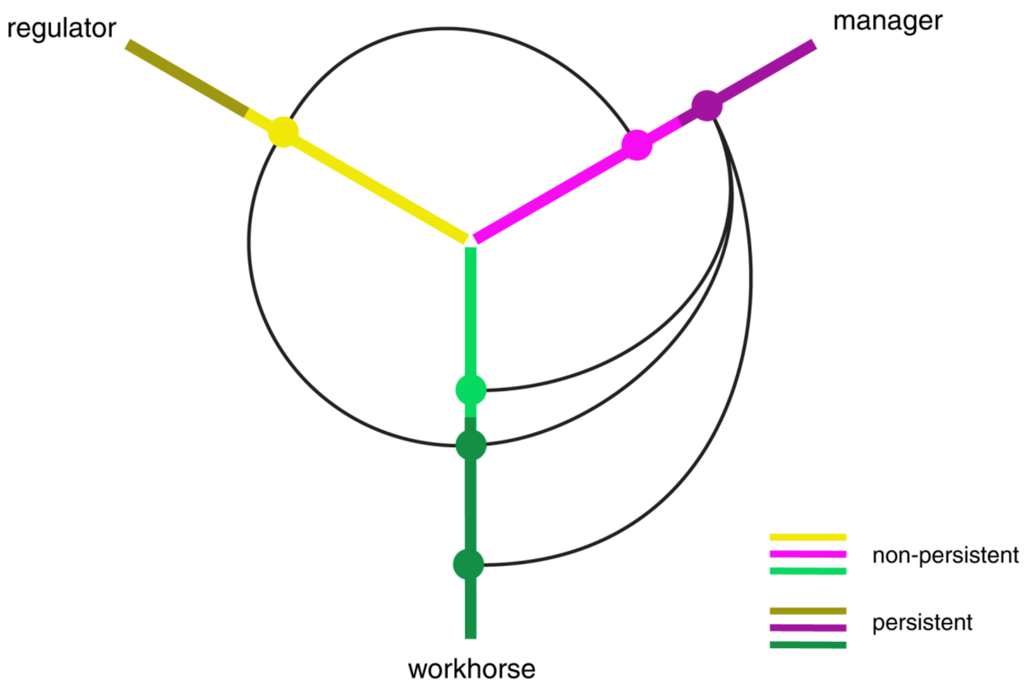


THE HIERARCHICAL LAYOUT OF THE E. COLI TRANSCRIPTIONAL REGULATORY NETWORK AND THE LINUX CALL GRAPH.

Yan, K.K., et al., Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks.
Proc Natl Acad Sci U S A, 2010. 107(20): p. 9186-91.

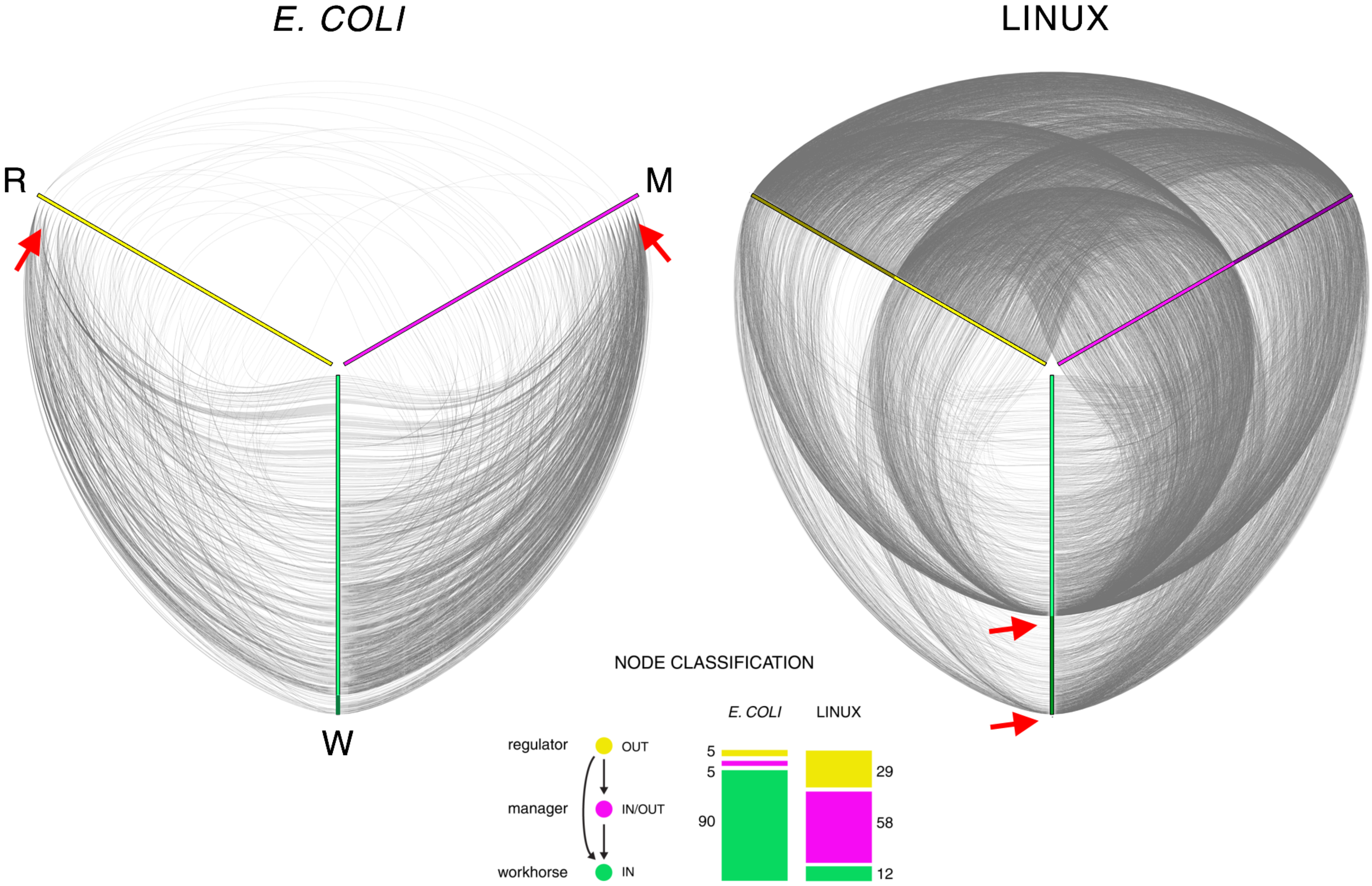


NETWORK LAYOUT



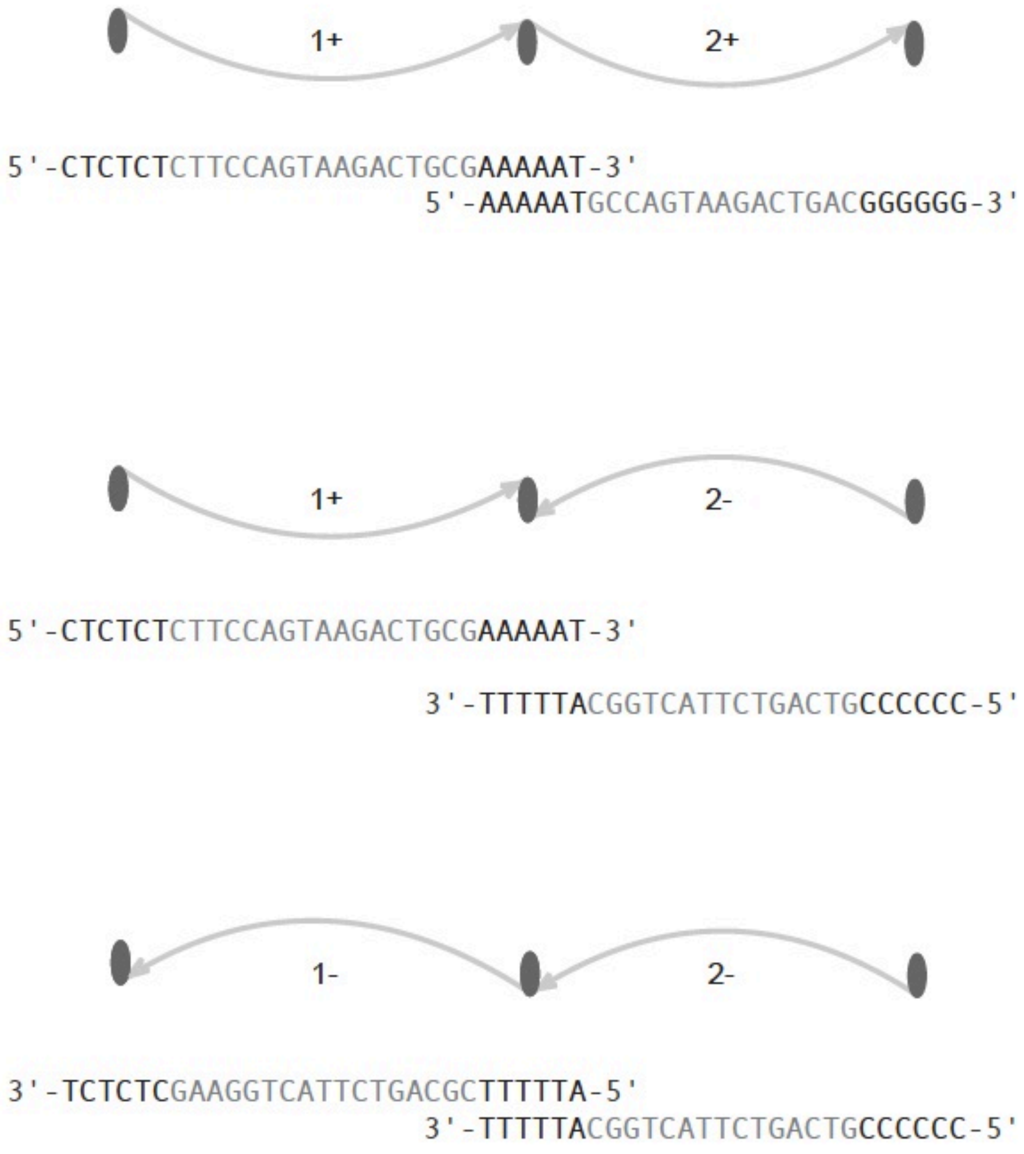
THE HIERARCHICAL LAYOUT OF THE E. COLI TRANSCRIPTIONAL REGULATORY NETWORK AND THE LINUX CALL GRAPH.

Yan, K.K., et al., Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. Proc Natl Acad Sci U S A, 2010. 107(20): p. 9186-91.



THE HIERARCHICAL LAYOUT OF THE E. COLI TRANSCRIPTIONAL REGULATORY NETWORK AND THE LINUX CALL GRAPH.
REINTERPRETED FIGURE.

Yan, K.K., et al., Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks.
Proc Natl Acad Sci U S A, 2010. 107(20): p. 9186-91.



5'-CTCTCTCTTCCAGTAAGACTGCGAAAAAT-3'
5'-AAAAATGCCAGTAAGACTGACGGGGGG-3'

5'-CTCTCTCTTCCAGTAAGACTGCGAAAAAT-3'

3'-TTTTTACGGTCATTCTGACTGCCCCCC-5'

3'-TCTCTCGAAGGTCATTCTGACGCTTTTTA-5'
3'-TTTTTACGGTCATTCTGACTGCCCCCC-5'

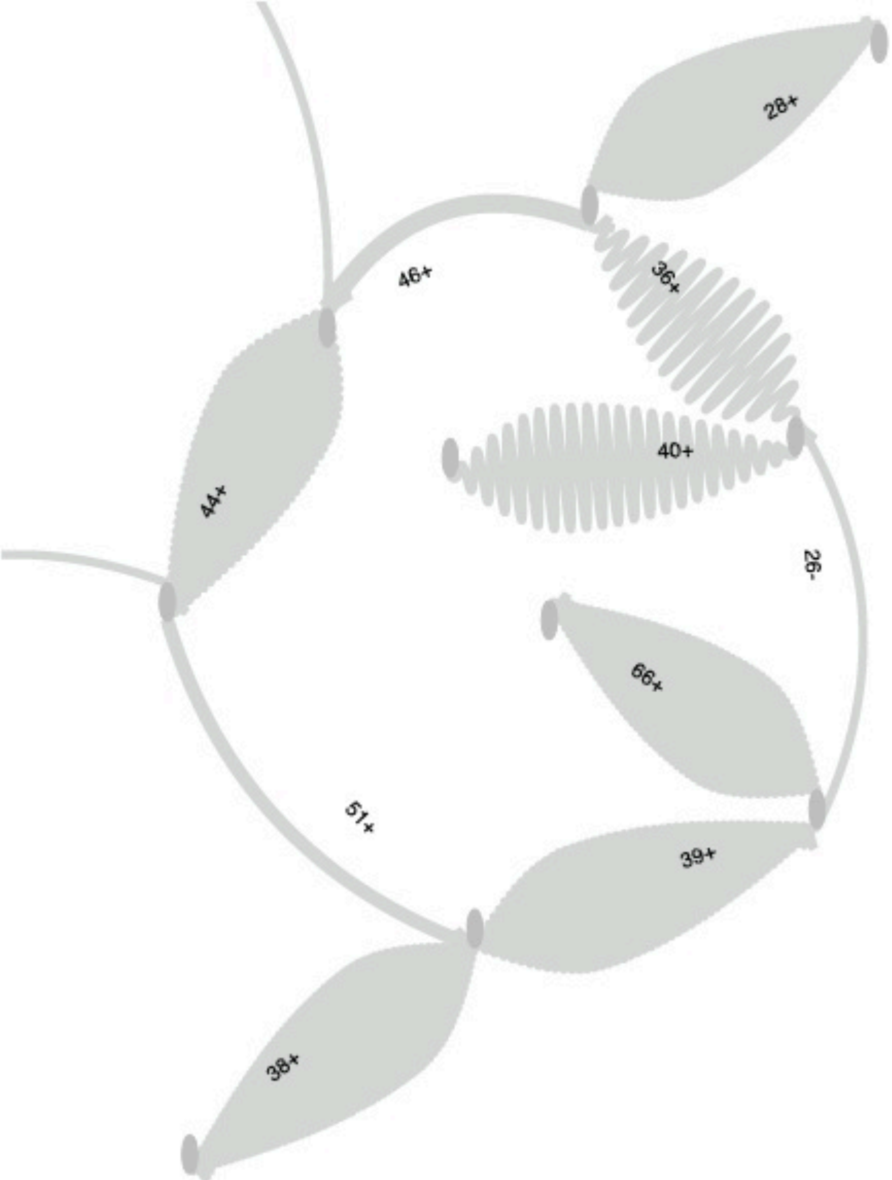
ABySS Explorer visualizes sequence assemblies on a graph.

Contigs are *edges*.

Nodes represent overlap.

EXAMPLE OF VERTEX POLARITY USING AN OVERLAP SIZE OF 6 NT.
BOTH GRAPH AND SEQUENCE REPRESENTATIONS ARE SHOWN.

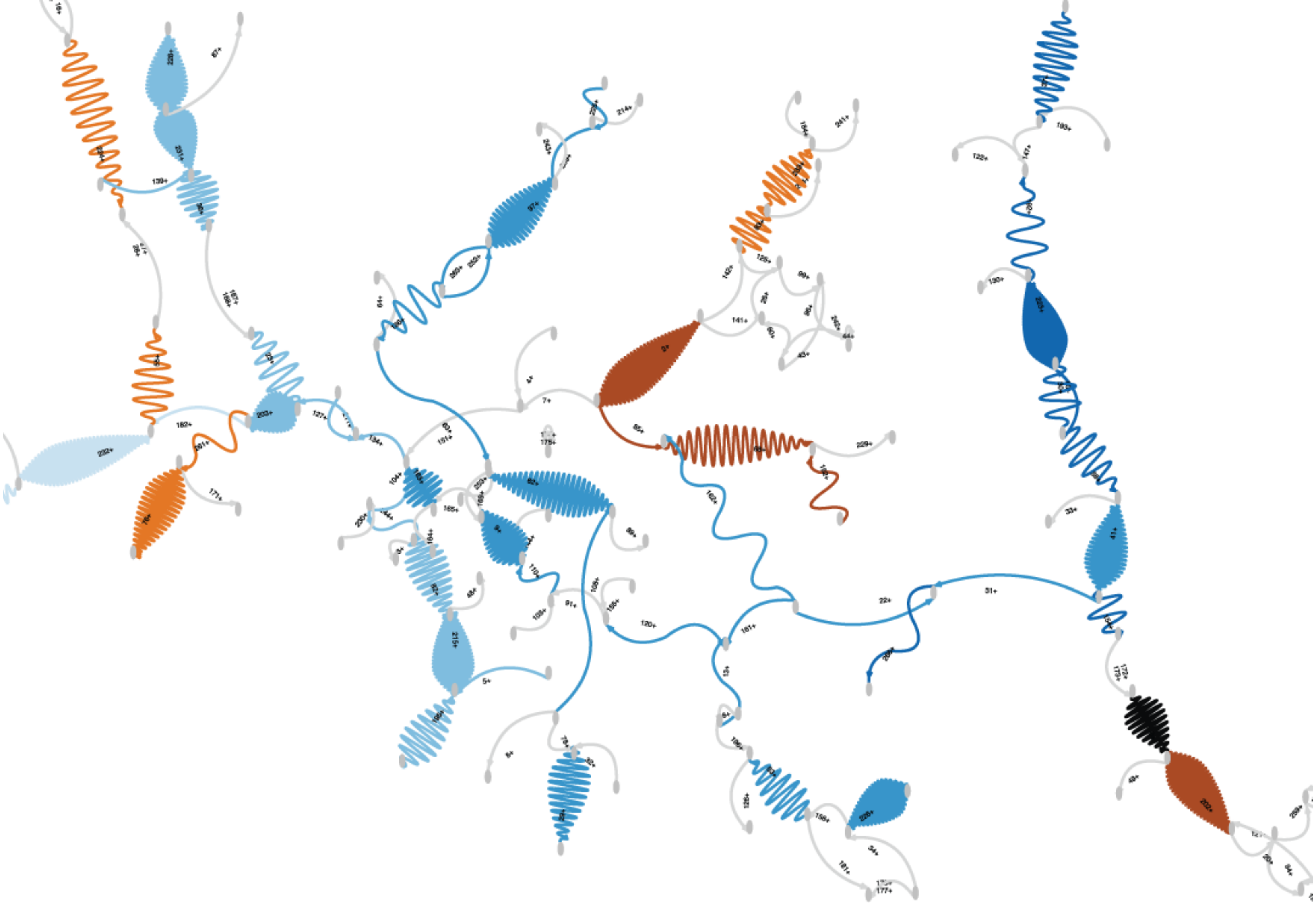
Nielsen, C. B., Jackman, S. D., Birol, I., and Jones, S. J. 2009. ABySS-Explorer: Visualizing Genome Sequence Assemblies. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (Nov. 2009), 881-888



Assembly structures (here, contig cycle) are depicted more intuitively.

DETAIL OF AN ASSEMBLED CONTIG CYCLE.

Nielsen, C. B., Jackman, S. D., Birol, I., and Jones, S. J. 2009. ABySS-Explorer: Visualizing Genome Sequence Assemblies. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (Nov. 2009), 881-888

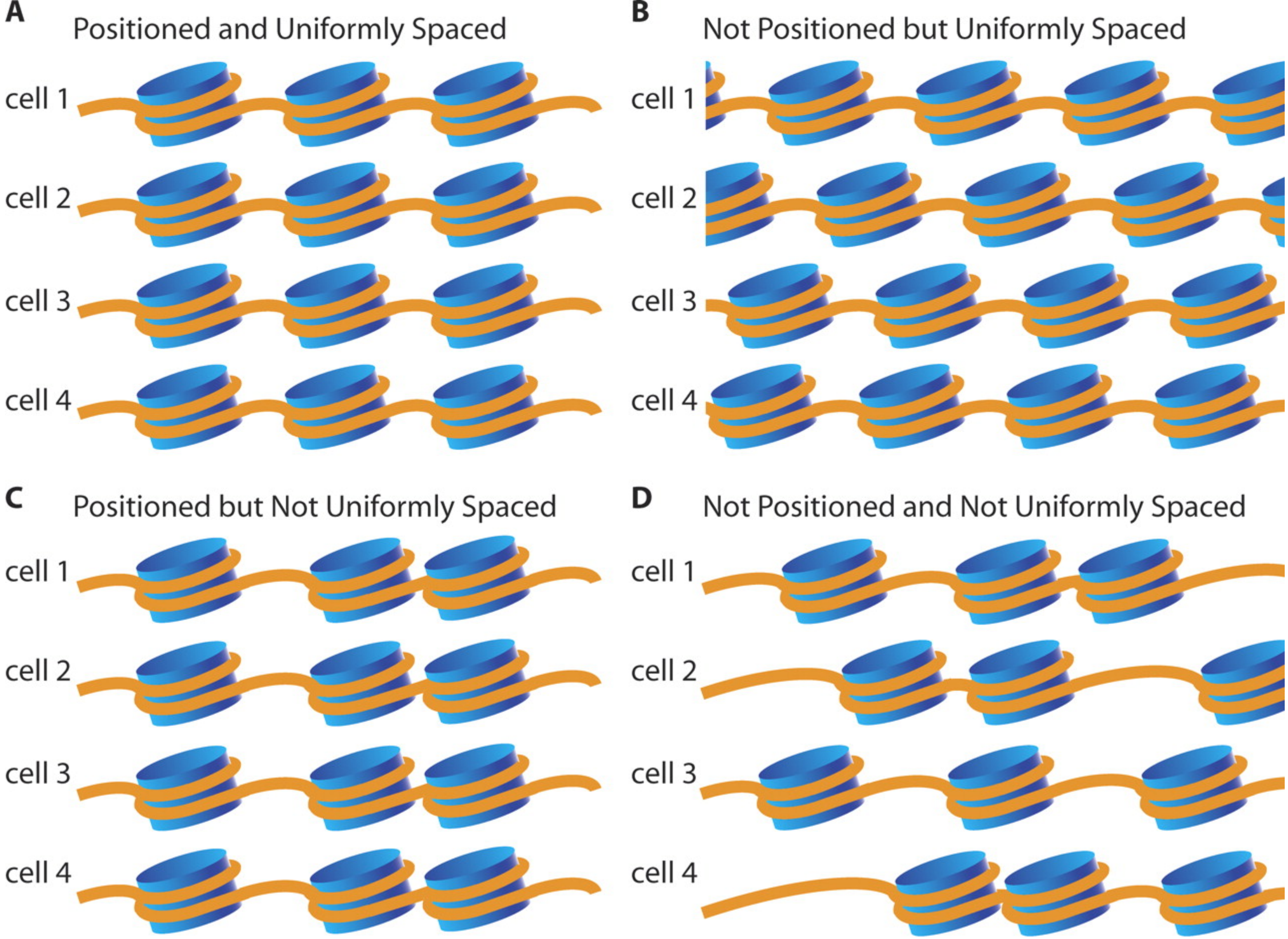


ABYSS-EXPLORER VIEW OF ROUGHLY 200,000 NT ASSEMBLED FROM THE HUMAN LYMPHOMA GENOME

Nielsen, C. B., Jackman, S. D., Birol, I., and Jones, S. J. 2009. ABySS-Explorer: Visualizing Genome Sequence Assemblies. IEEE Transactions on Visualization and Computer Graphics 15, 6 (Nov. 2009), 881-888

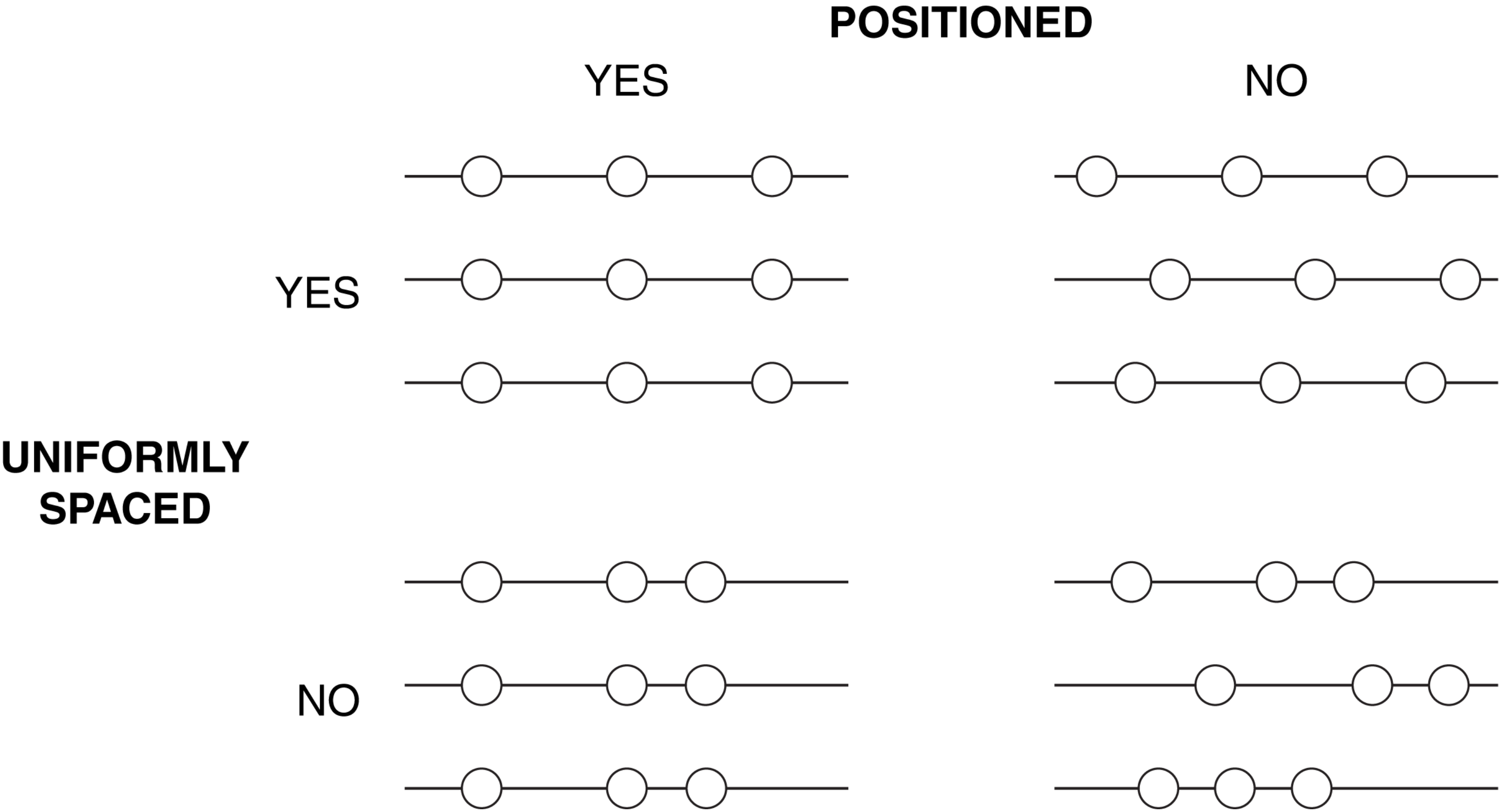
● part 2
MAKING IT CLEAR

CART JUNK



POSSIBLE PATTERNS OF NUCLEOSOME POSITIONING.

Valouev, A., et al., A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res, 2008. 18(7): p. 1051-63.



POSITIONED

YES

NO

UNIFORMLY SPACED

YES

NO

POSSIBLE PATTERNS OF NUCLEOSOME POSITIONING.
REINTERPRETED FIGURE

Valouev, A., et al., A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res, 2008. 18(7): p. 1051-63.

● part 2

MAKING IT CLEAR

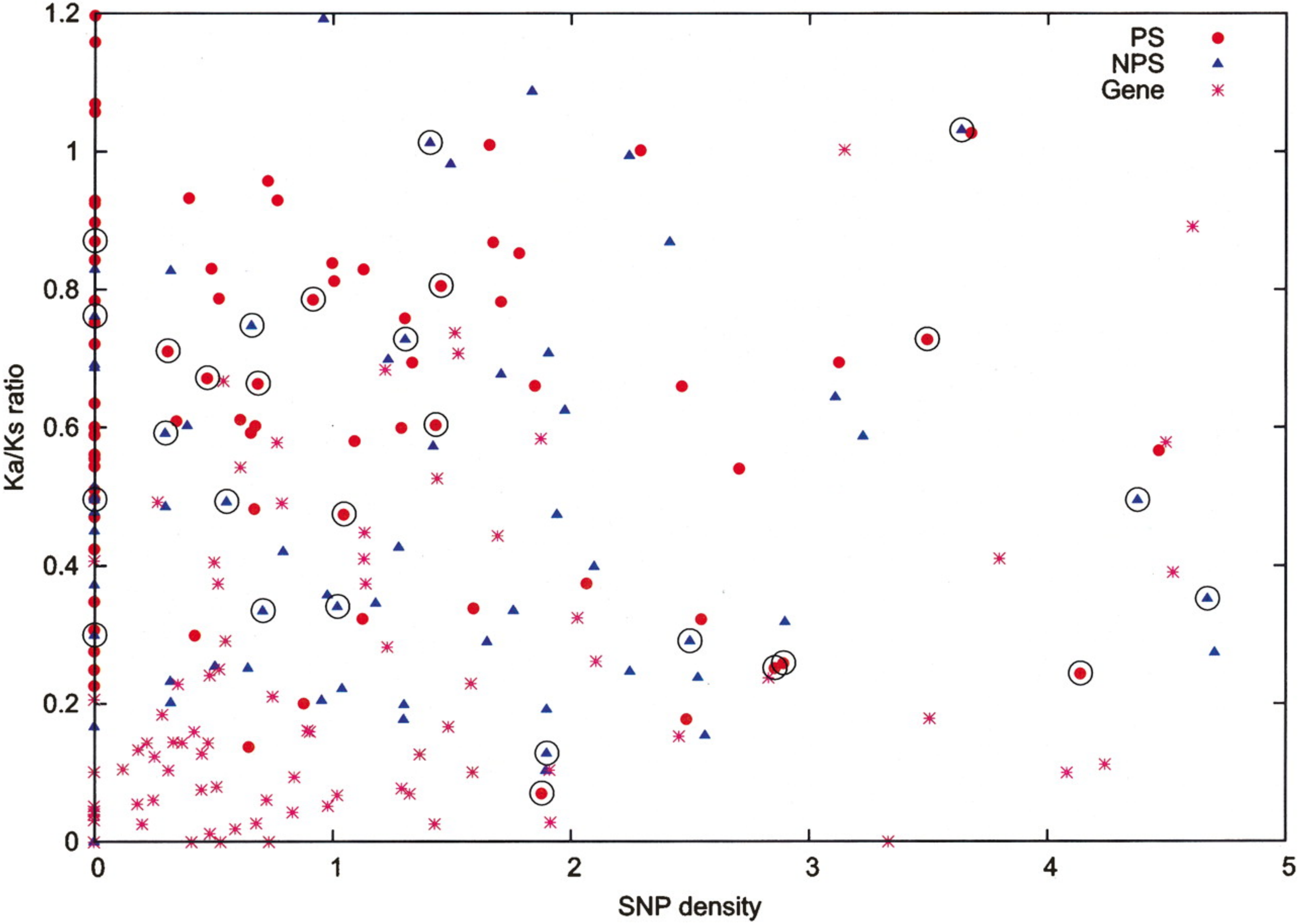
GLYPHS

REFINE AND SIMPLIFY GLYPHS

Use either shape or color, not both.

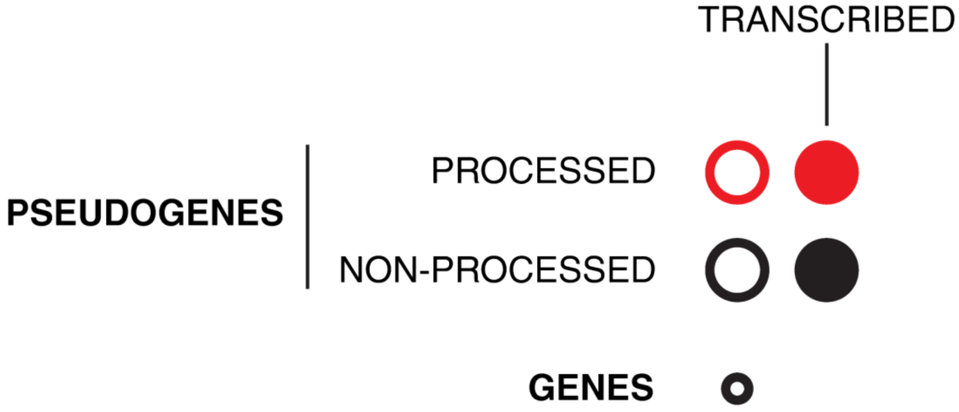
Circle and square can appear similar.

Use empty/filled (or small/large) glyphs to categorize data
  empty or small – *insignificant*
  filled or large – *significant*



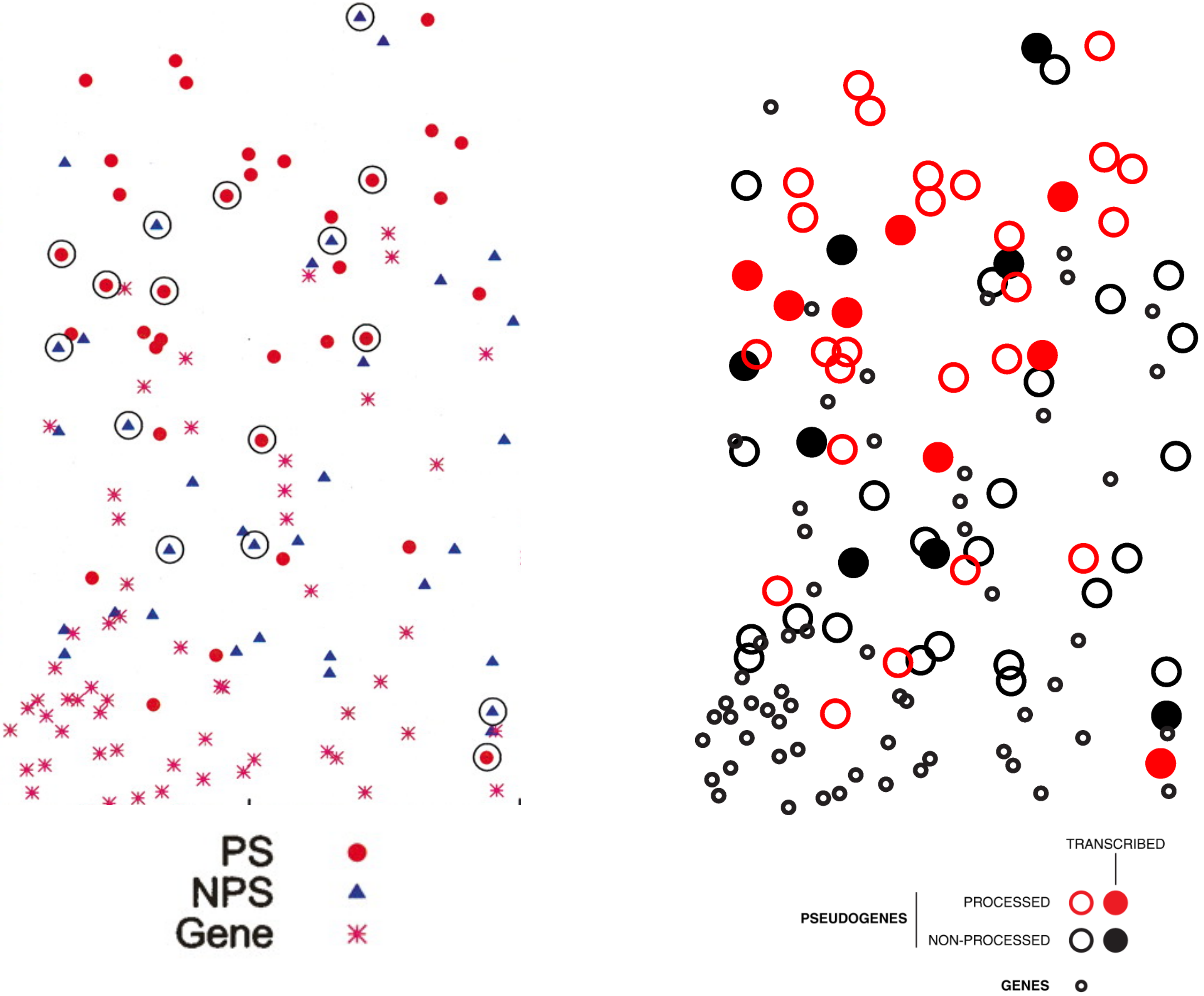
COMPARISON OF KA/KS RATIO AND SNP DENSITY FOR GENES AND PSEUDOGENES.

Zheng, D., et al., Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. Genome Res, 2007. 17(6): p. 839-51.



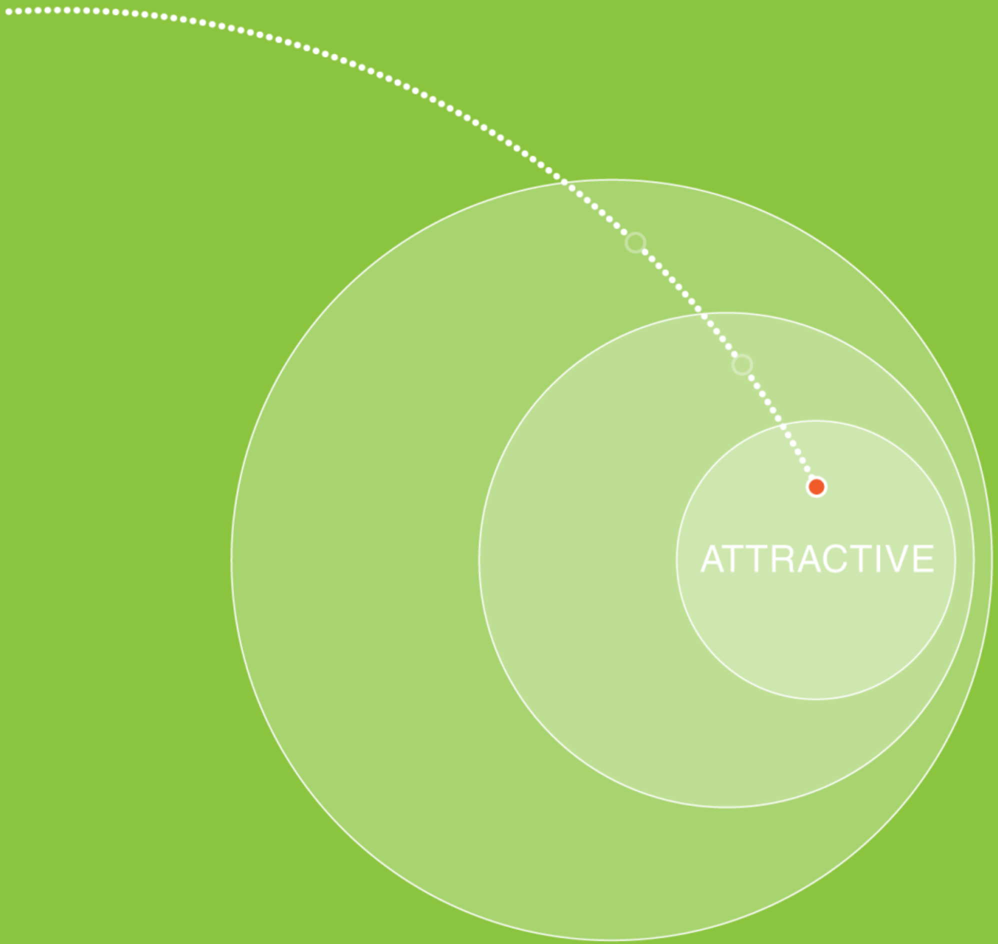
COMPARISON OF KA/KS RATIO AND SNP DENSITY FOR GENES AND PSEUDOGENES.

Zheng, D., et al., Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. Genome Res, 2007. 17(6): p. 839-51.



COMPARISON OF KA/KS RATIO AND SNP DENSITY FOR GENES AND PSEUDOGENES.

Zheng, D., et al., Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. Genome Res, 2007. 17(6): p. 839-51.



next

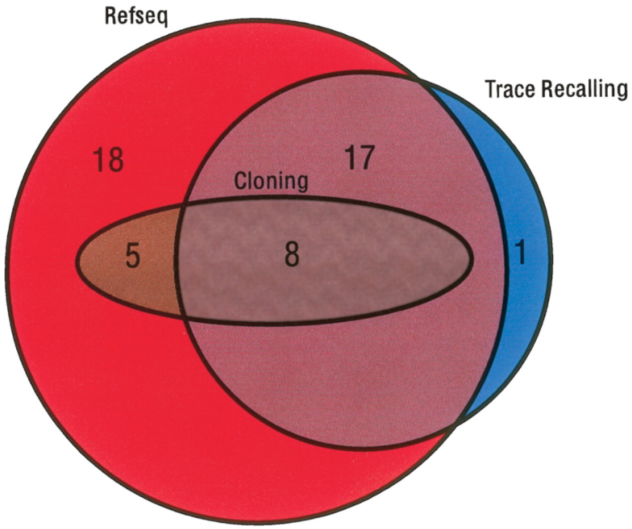
○ **part 1**
MAKING IT LEGIBLE

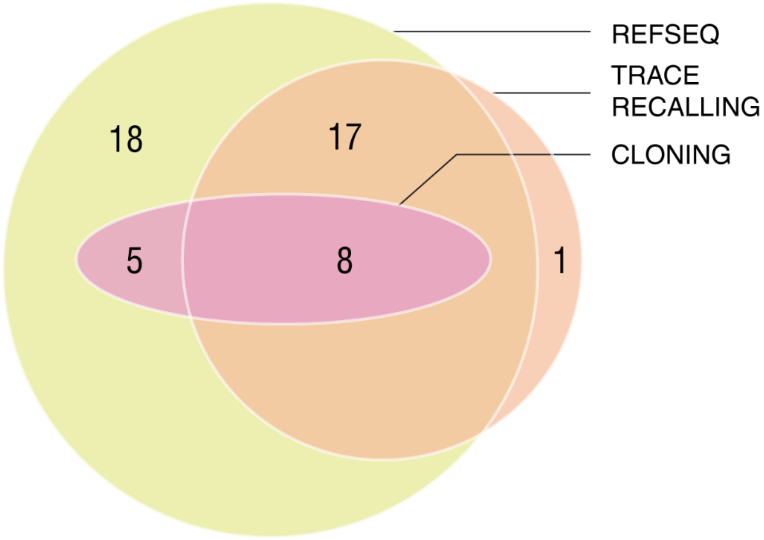
○ **part 2**
MAKING IT CLEAR

● **part 3**
MAKING IT PRETTY

● part 3
MAKING IT PRETTY

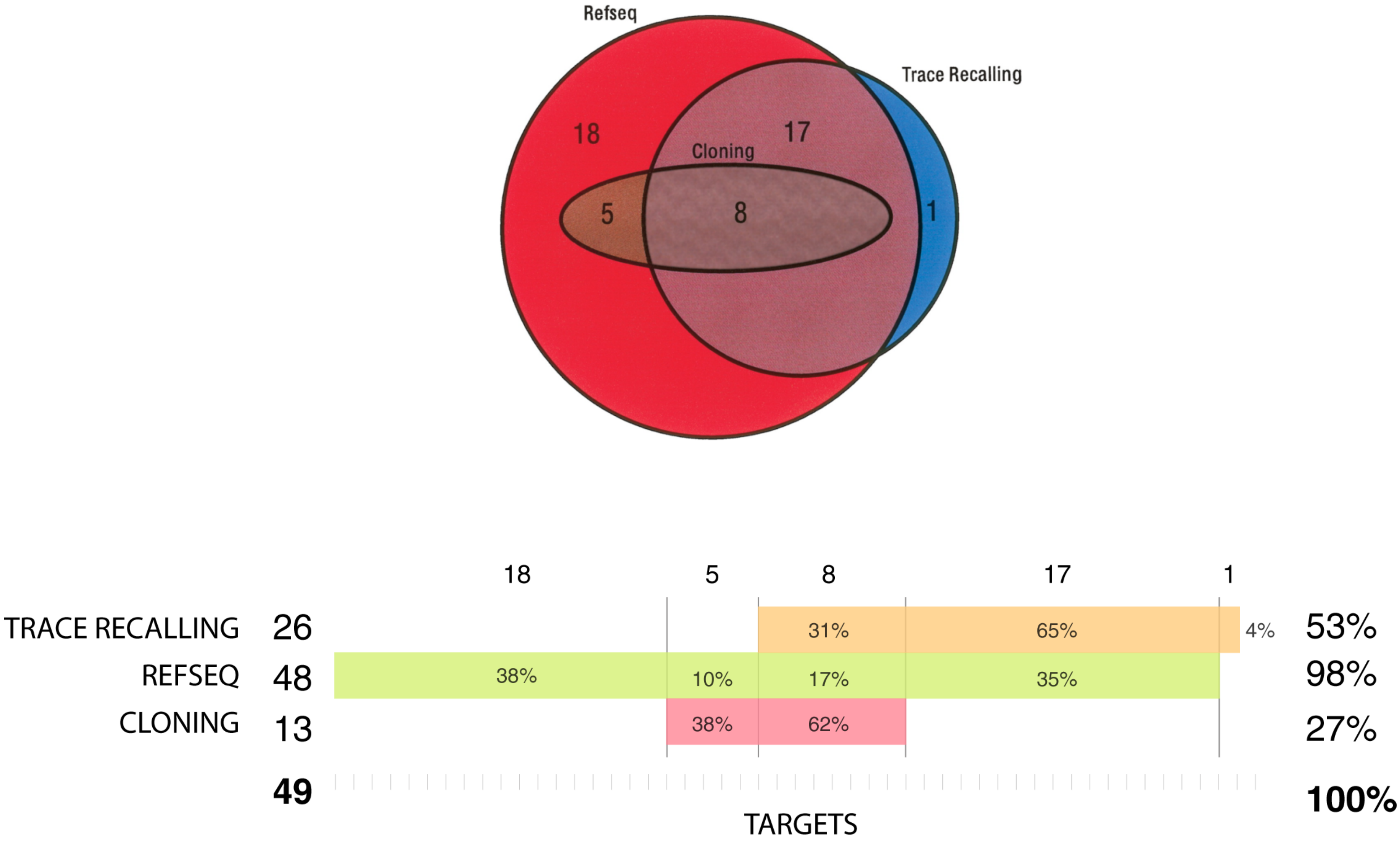
ELEGANCE





RESULTS OF THE ALTERNATE SPLICING EXPERIMENT.

Tenney, A.E., et al., A tale of two templates: automatically resolving double traces has many applications, including efficient PCR-based elucidation of alternative splices. Genome Res, 2007. 17(2): p. 212-8.

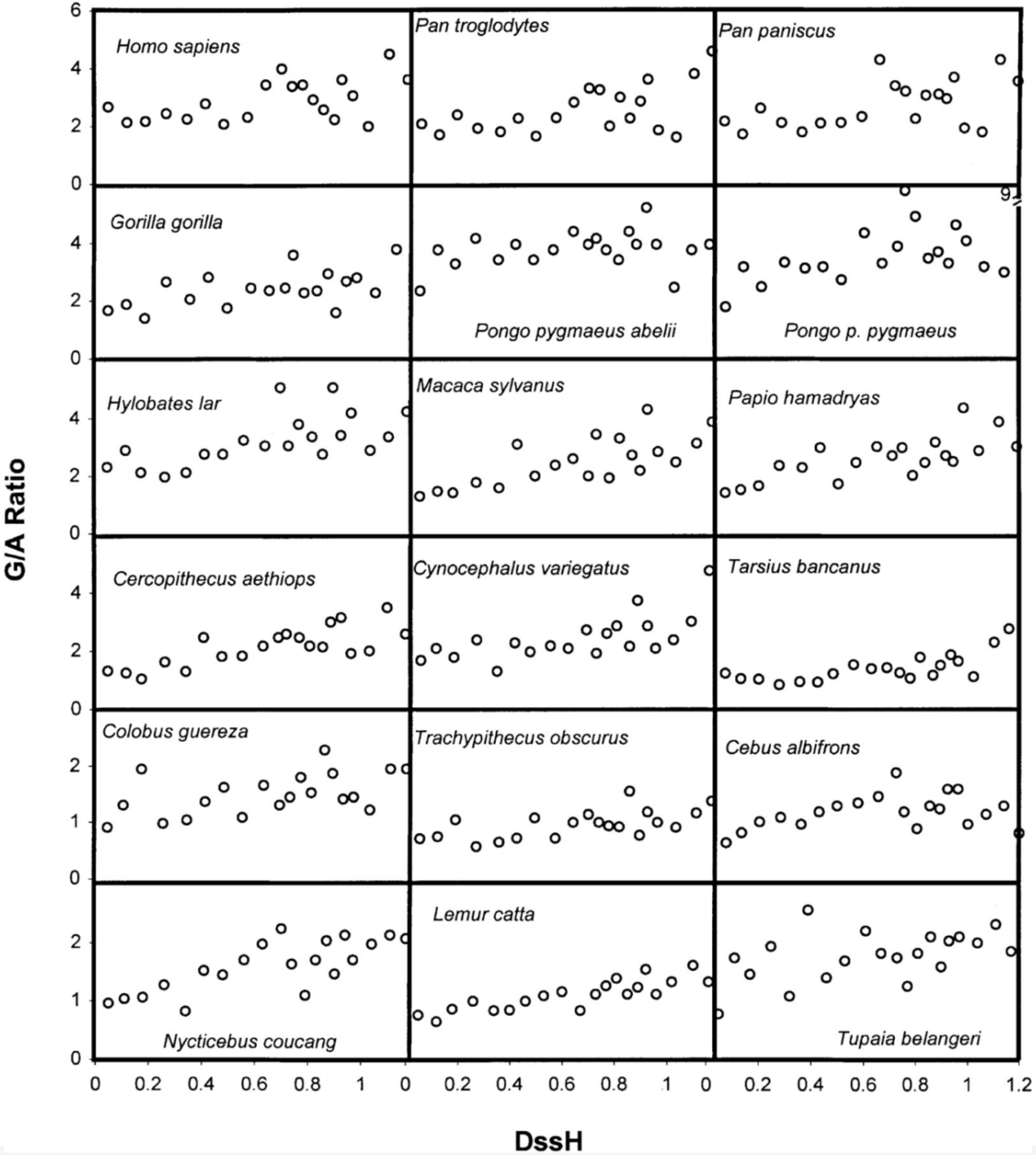


RESULTS OF THE ALTERNATE SPLICING EXPERIMENT.

Tenney, A.E., et al., A tale of two templates: automatically resolving double traces has many applications, including efficient PCR-based elucidation of alternative splices. Genome Res, 2007. 17(2): p. 212-8.

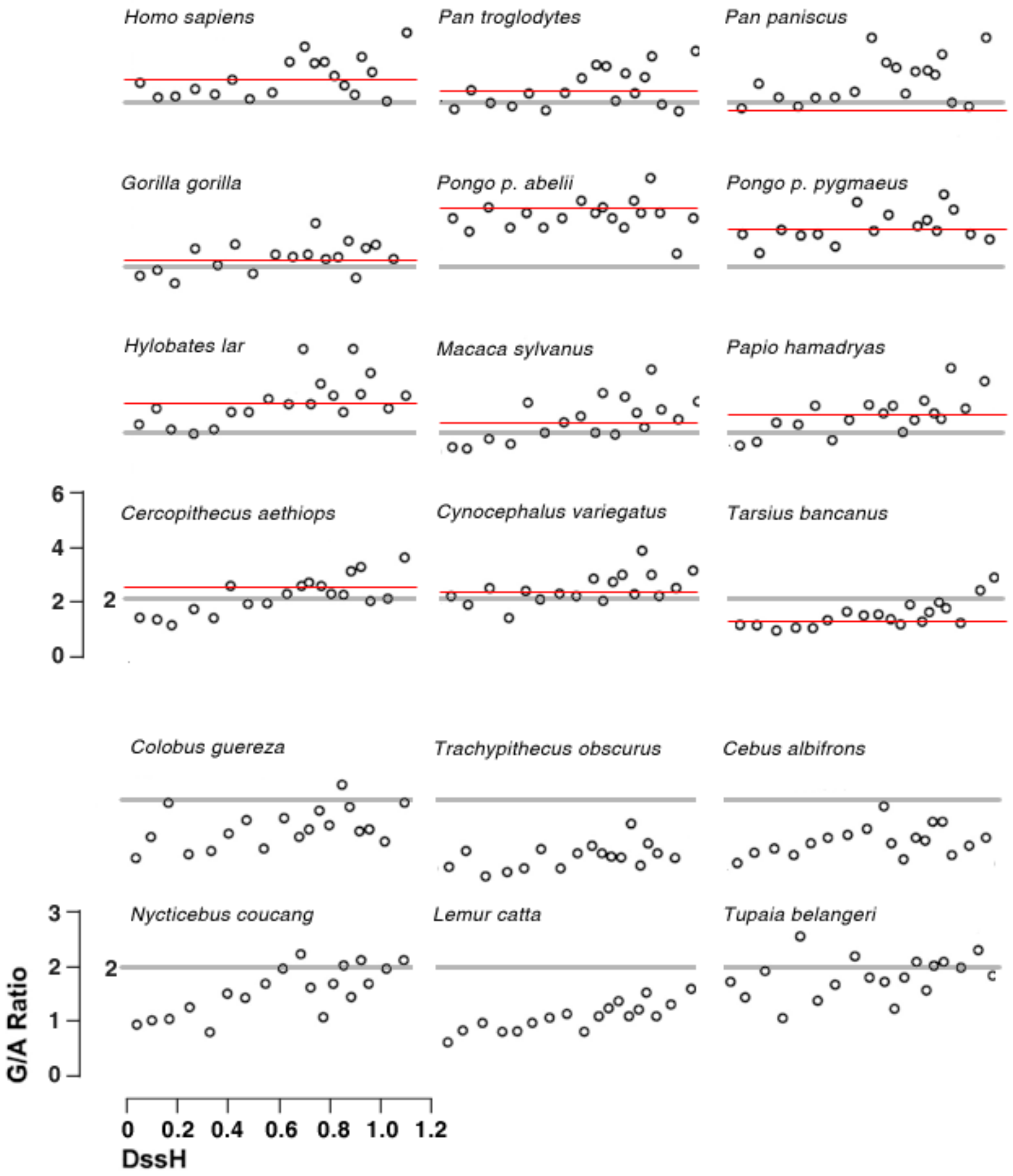
● part 3
MAKING IT PRETTY

FOCUS ON DATA



G/A RATIOS FOR COMPLETE PRIMATE MITOCHONDRIAL GENOMES AND TWO NEAR OUTGROUPS.

Raina, S.Z., et al., Evolution of base-substitution gradients in primate mitochondrial genomes. Genome Res, 2005. 15(5): p. 665-73.

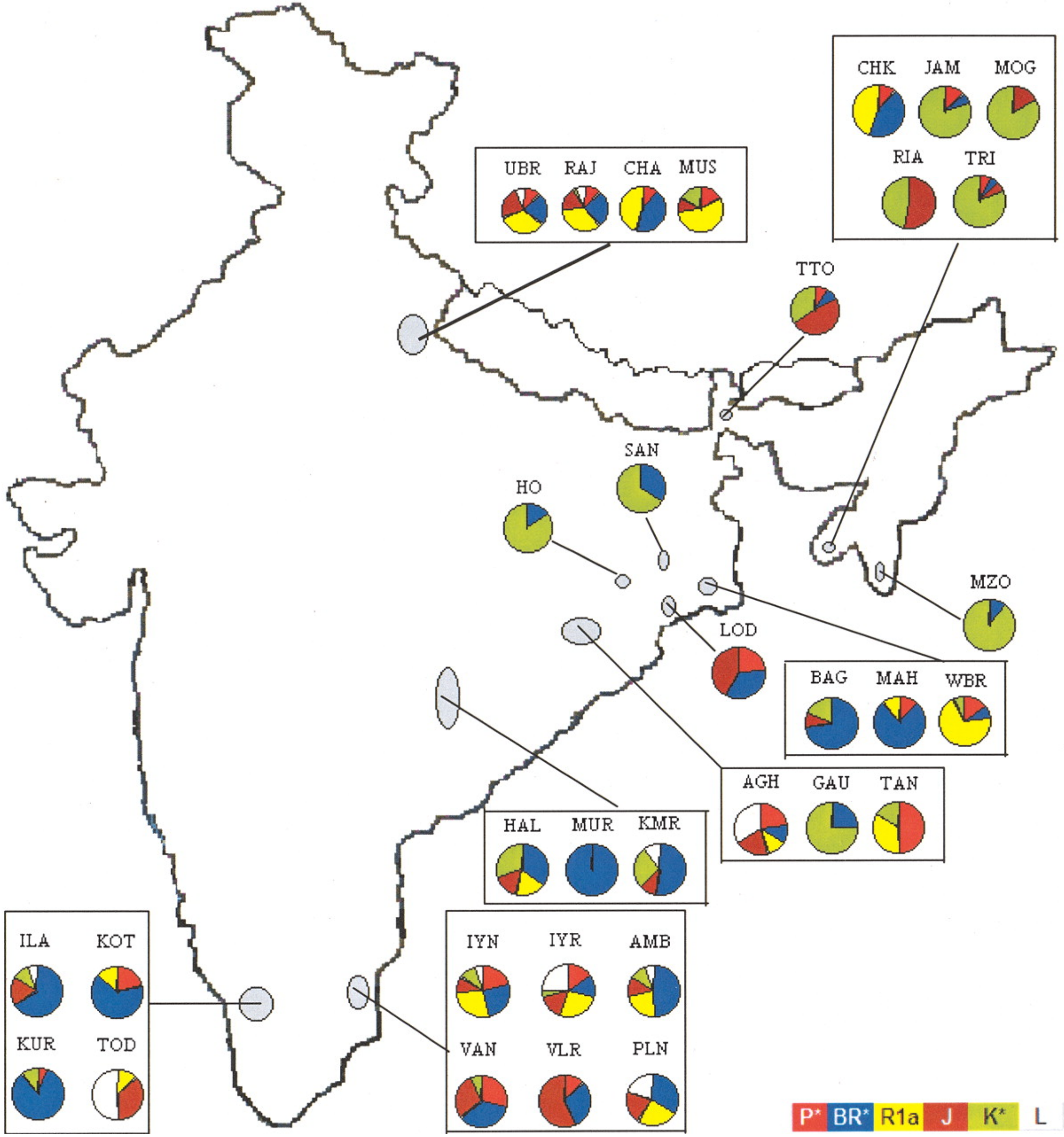


G/A RATIOS FOR COMPLETE PRIMATE MITOCHONDRIAL GENOMES AND TWO NEAR OUTGROUPS.
REINTERPRETED FIGURE

Raina, S.Z., et al., Evolution of base-substitution gradients in primate mitochondrial genomes. Genome Res, 2005. 15(5): p. 665-73.

● part 3
MAKING IT PRETTY

SPACING AND LAYOUT



FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS.

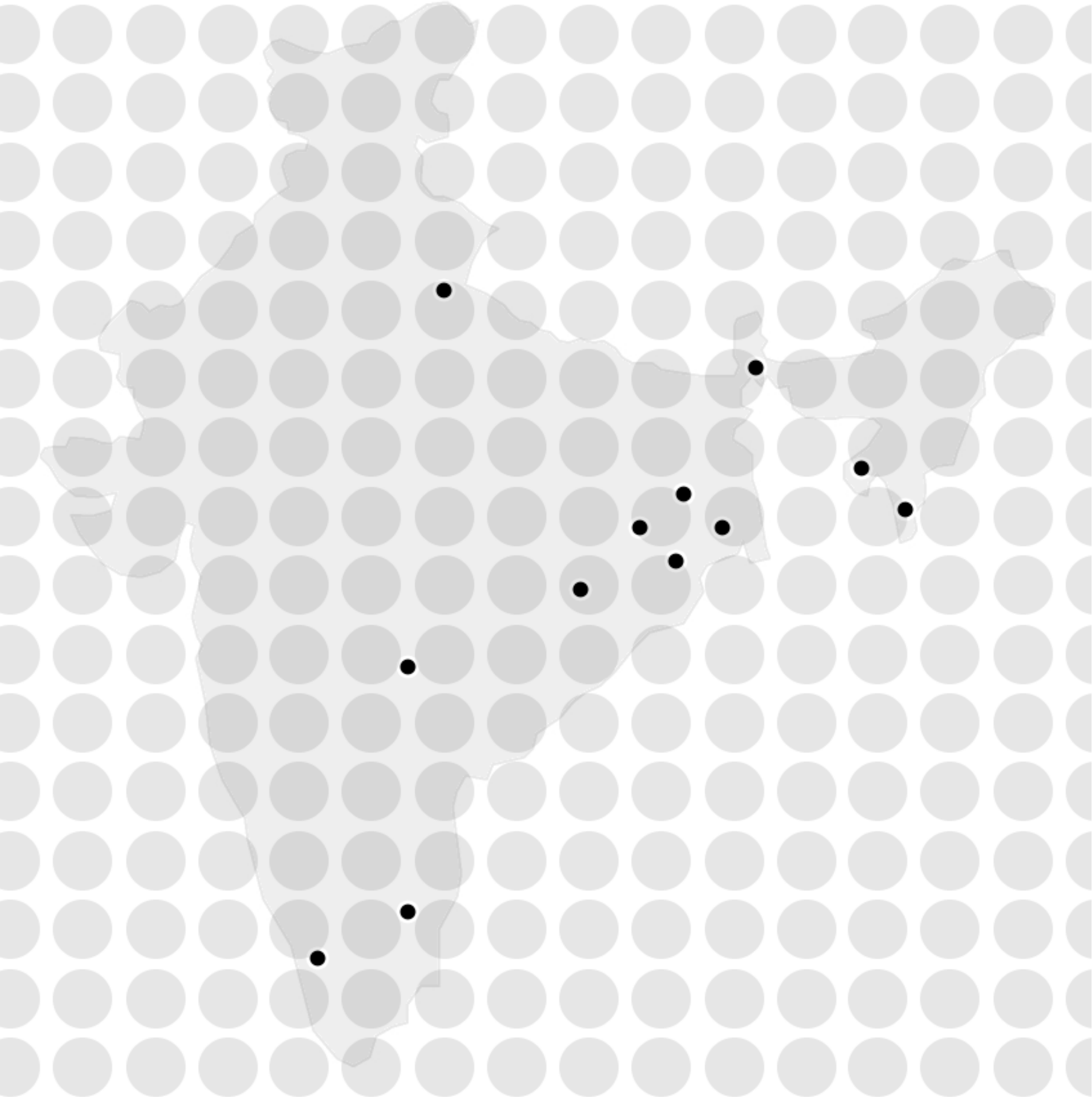
Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13 (10): p. 2277-90.

FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS.
REINTERPRETED FIGURE

Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13 (10): p. 2277-90.

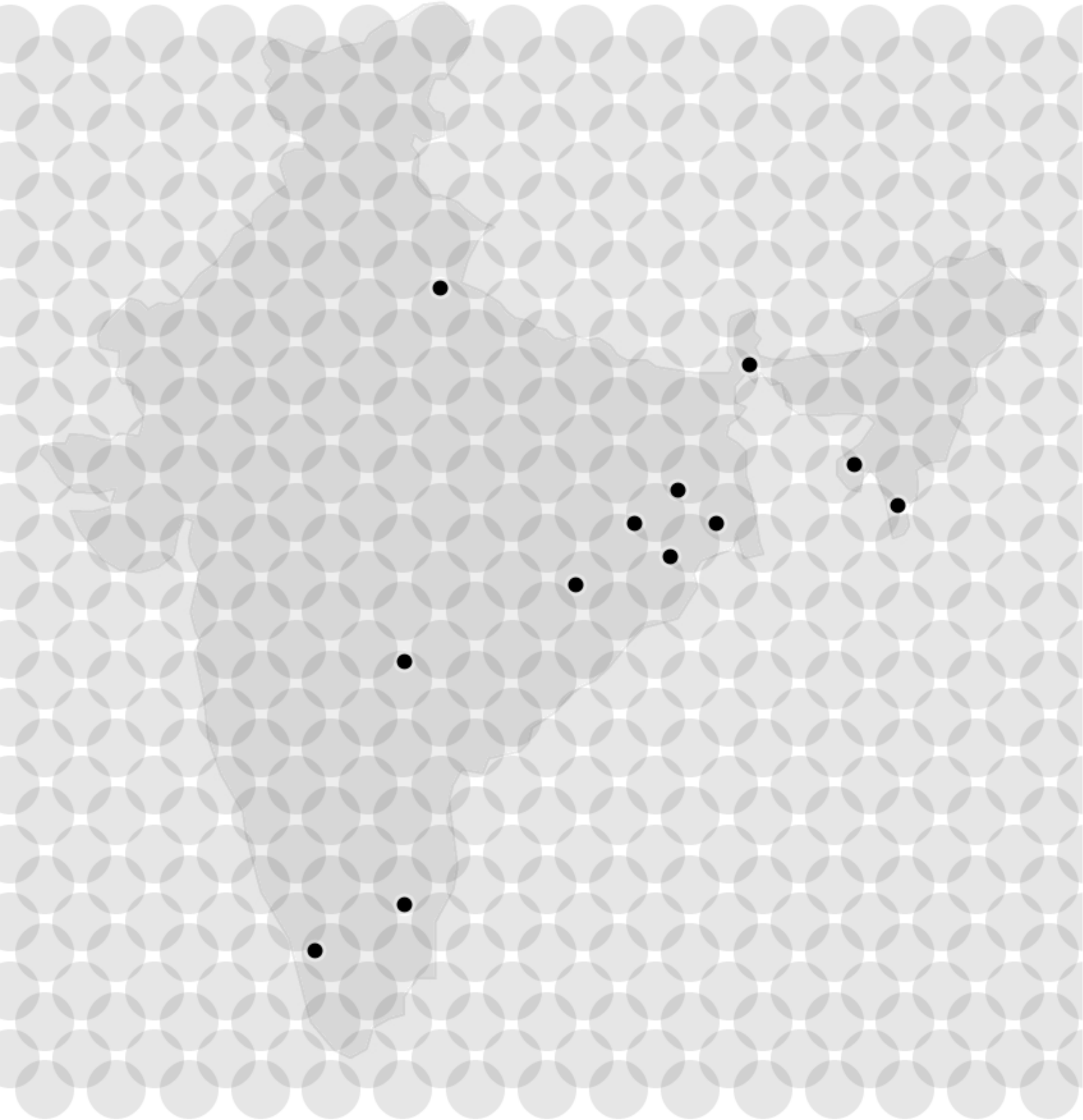
FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS.
REINTERPRETED FIGURE

Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13 (10): p. 2277-90.



FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS.
REINTERPRETED FIGURE

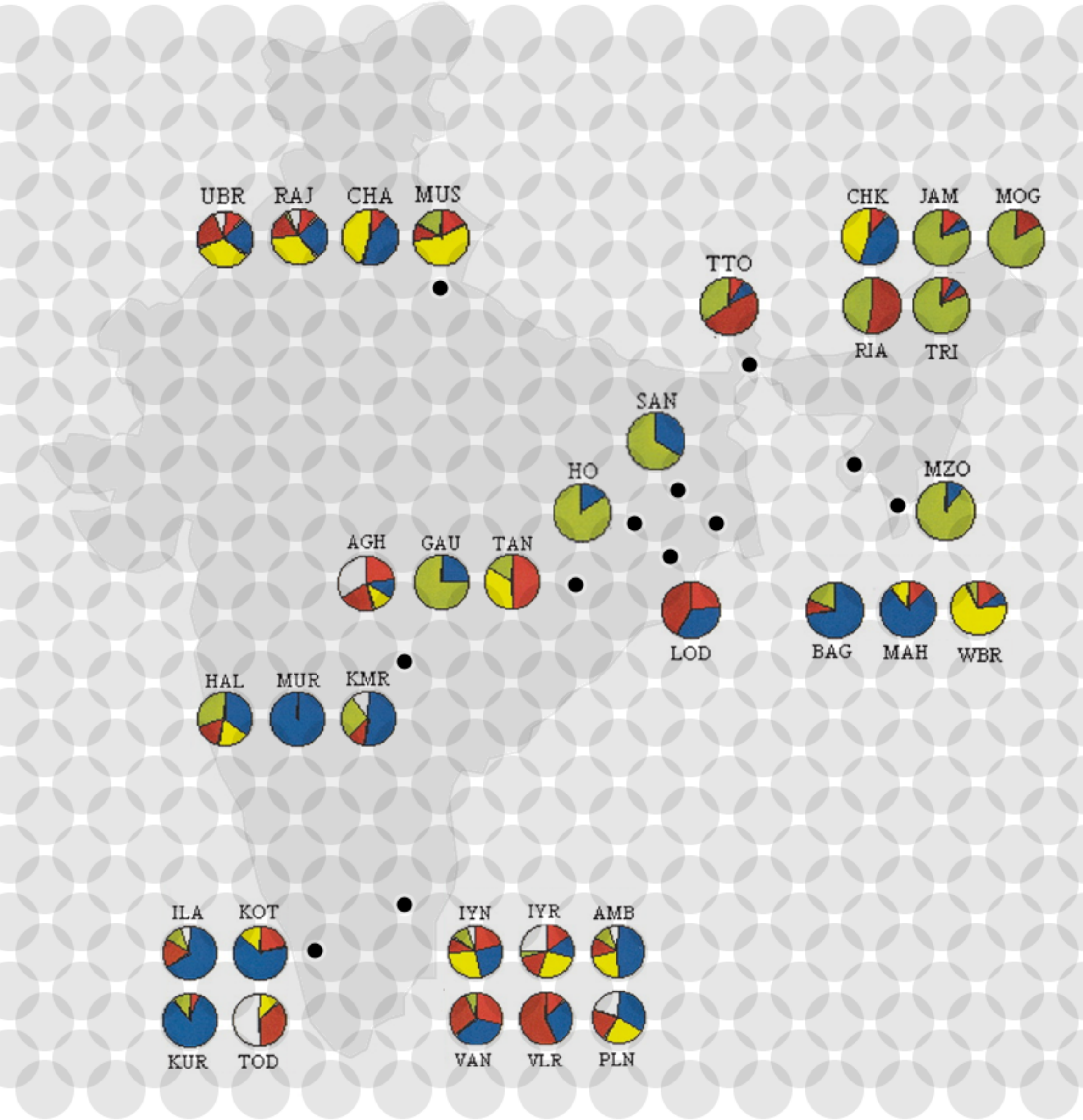
Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13 (10): p. 2277-90.



FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS.

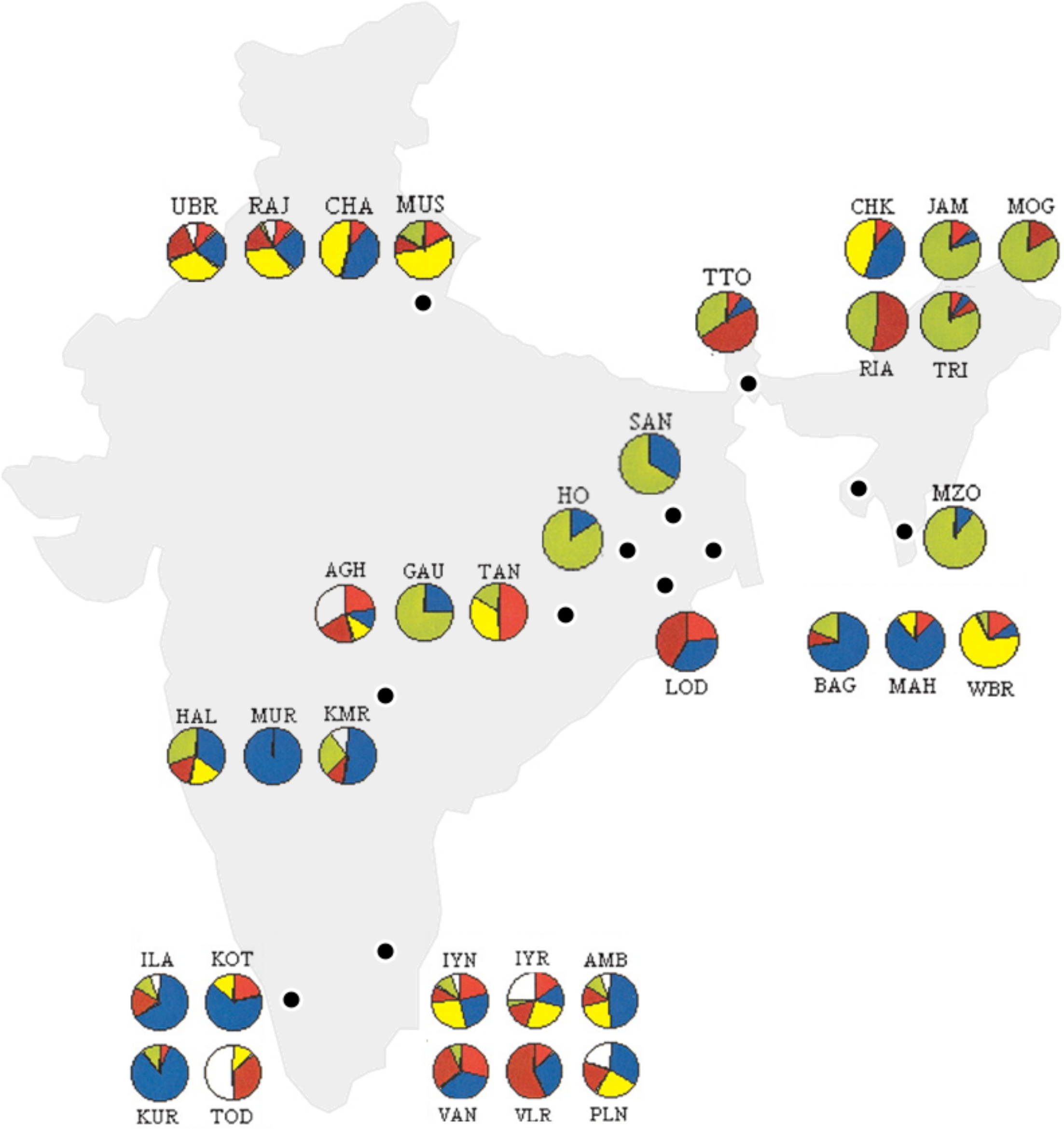
REINTERPRETED FIGURE

Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13 (10): p. 2277-90.



FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS.
REINTERPRETED FIGURE

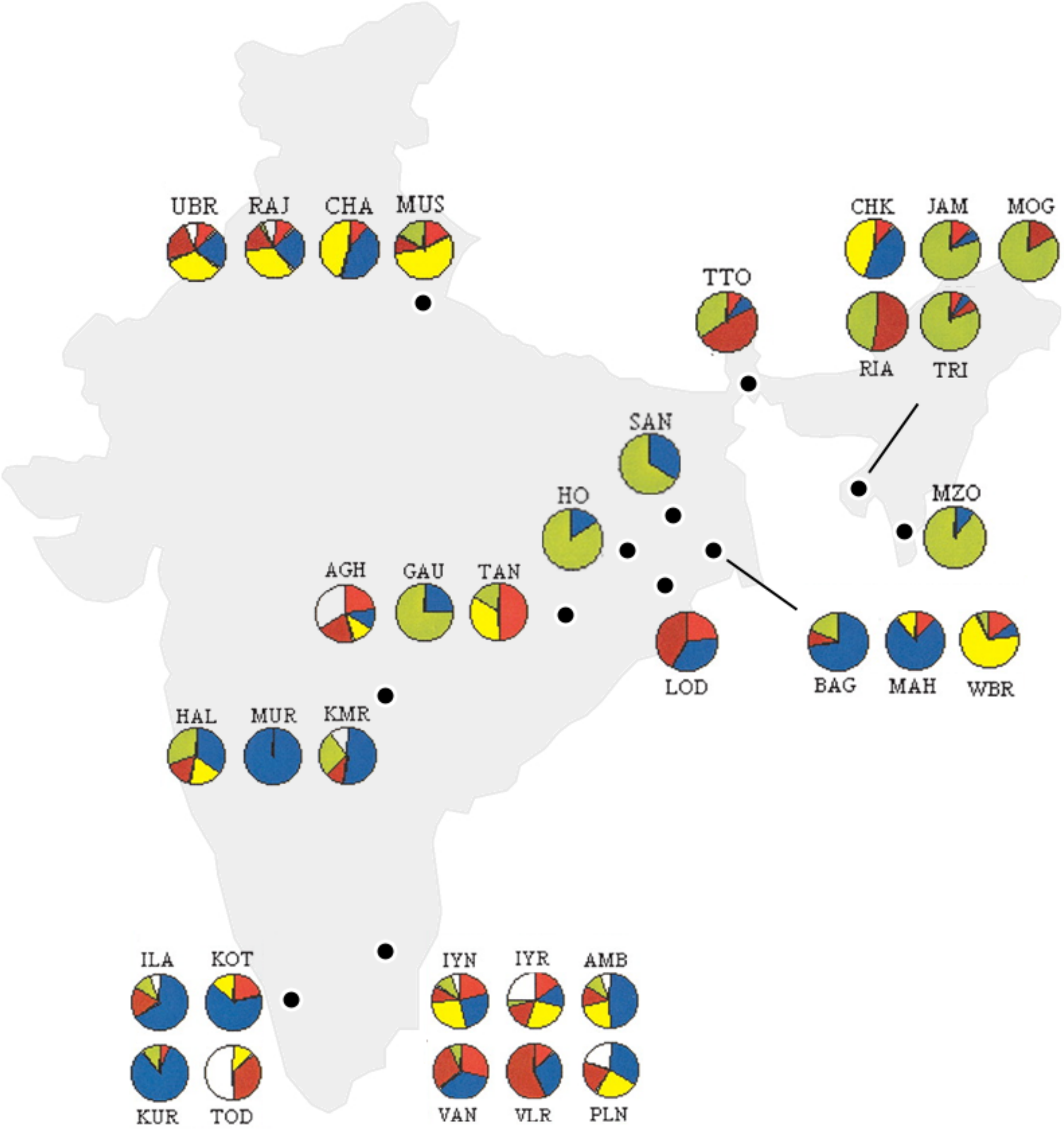
Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13 (10): p. 2277-90.



FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS.

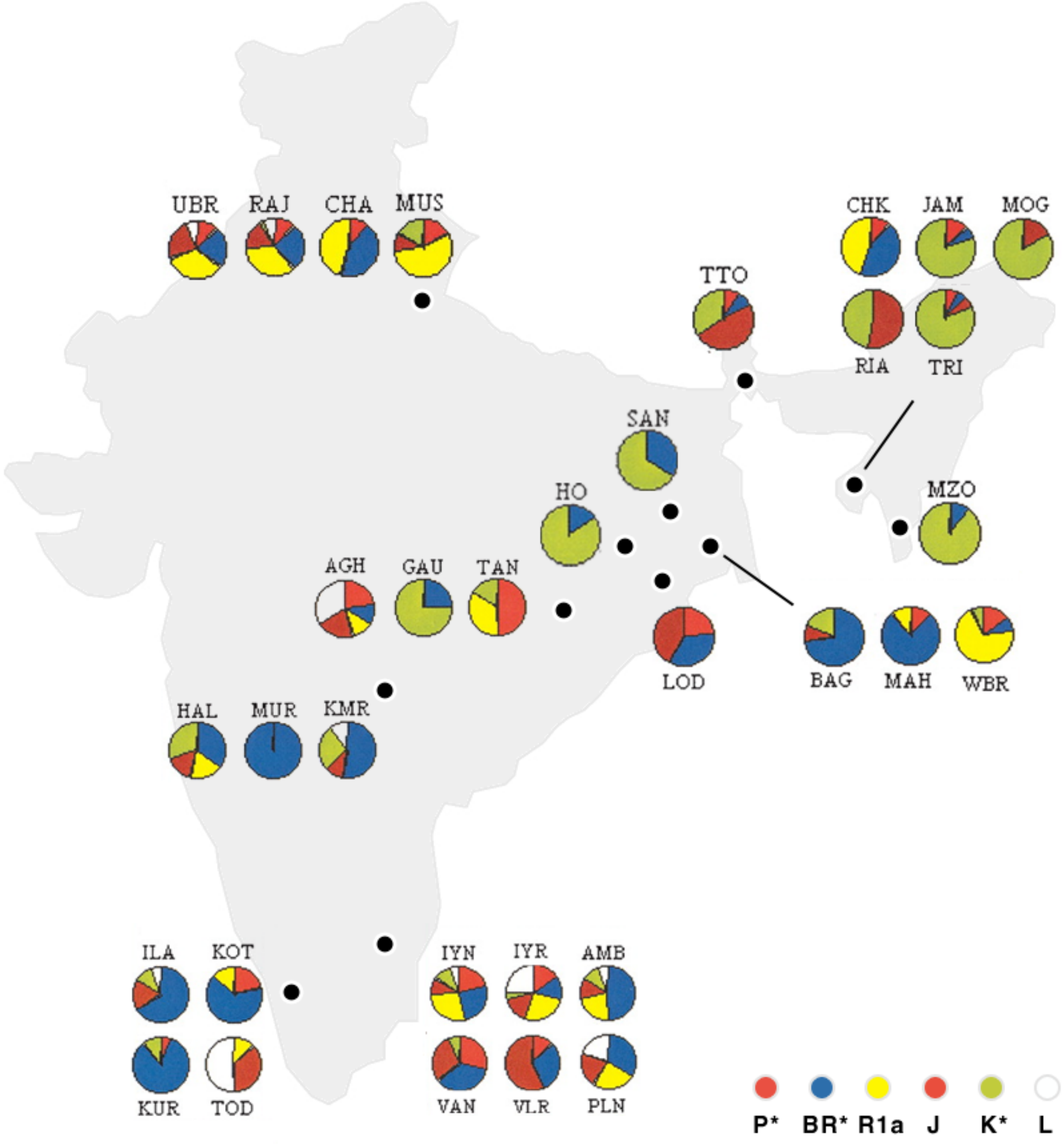
REINTERPRETED FIGURE

Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13 (10): p. 2277-90.



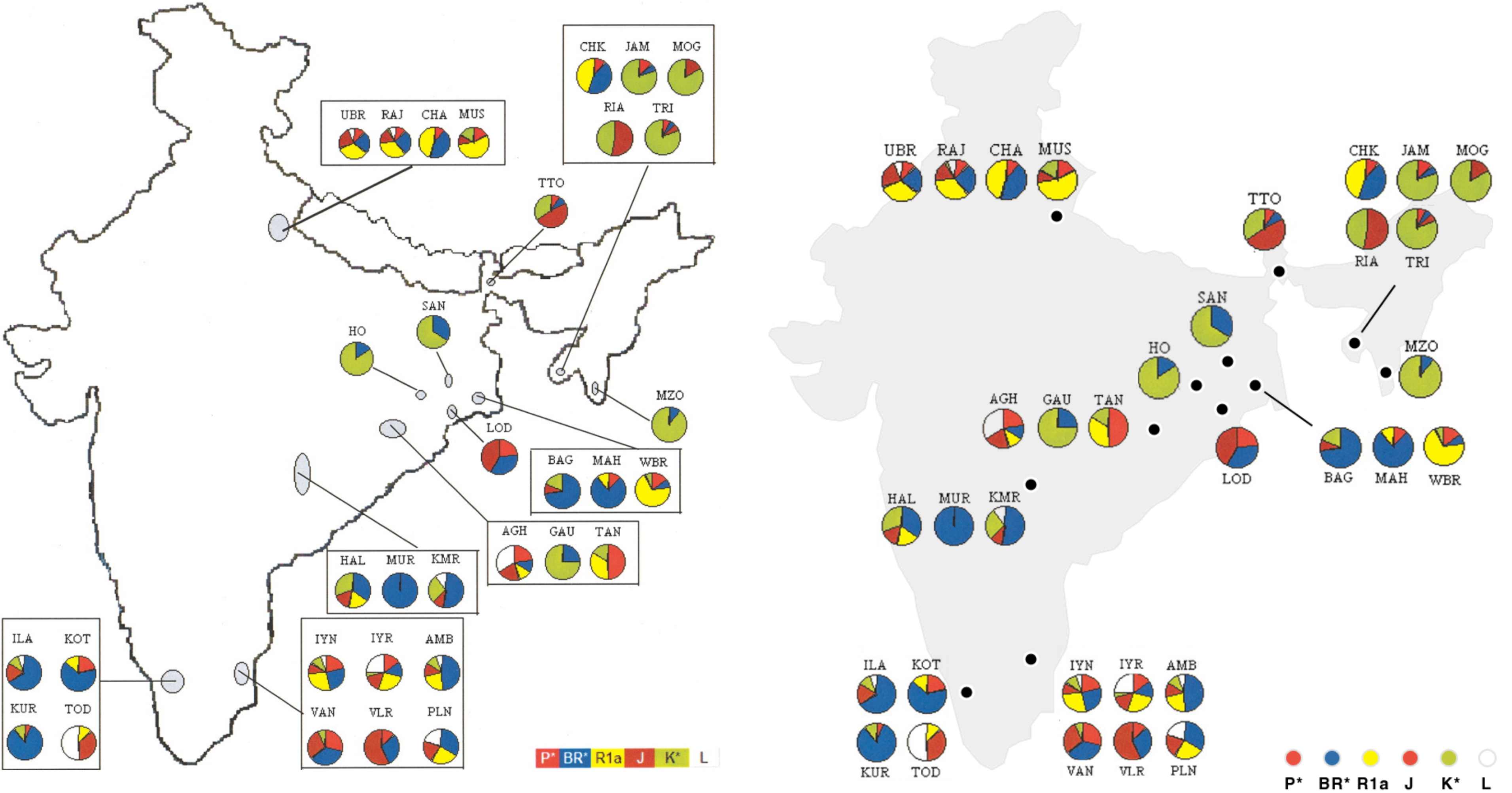
FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS.
REINTERPRETED FIGURE

Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13 (10): p. 2277-90.



FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS.
REINTERPRETED FIGURE

Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13 (10): p. 2277-90.

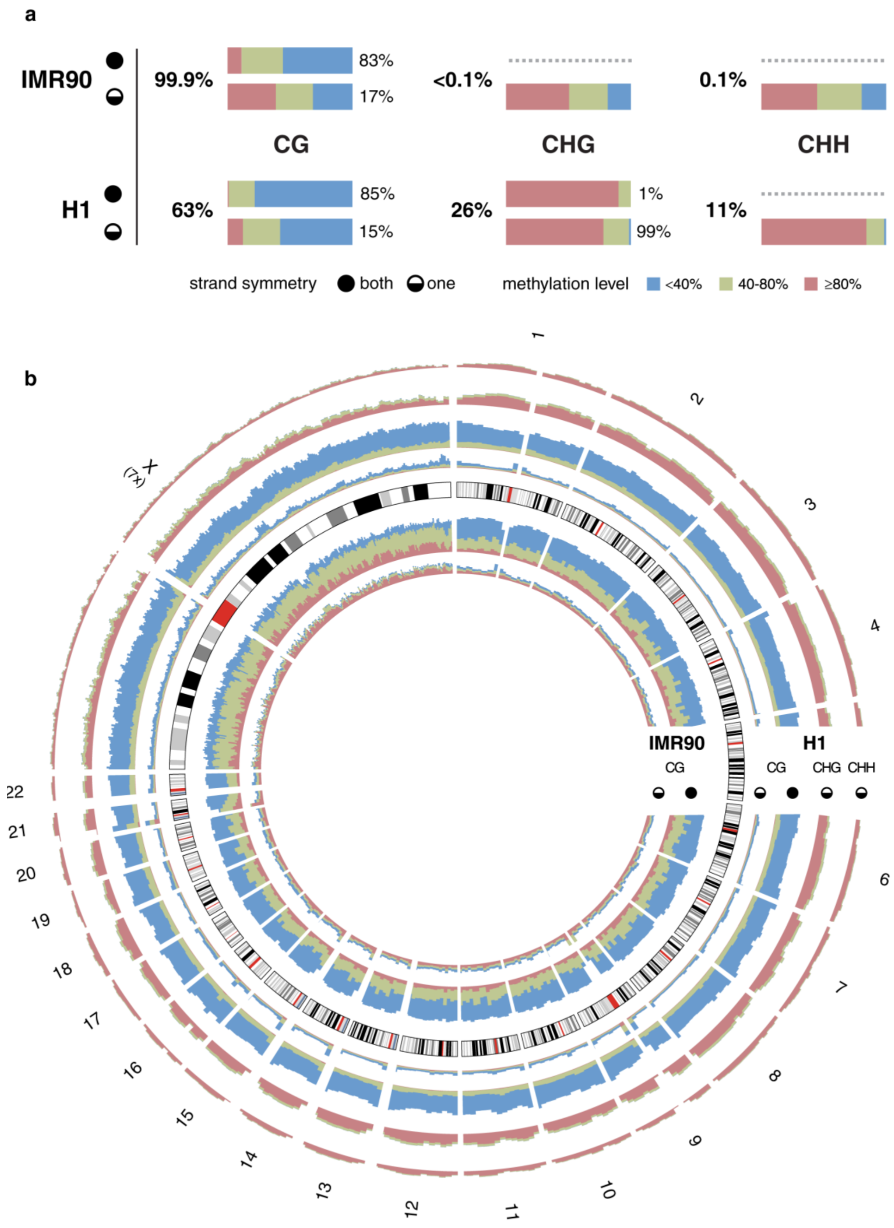


FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS.

Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13 (10): p. 2277-90.

part 3
MAKING IT PRETTY

ENTRY POINT



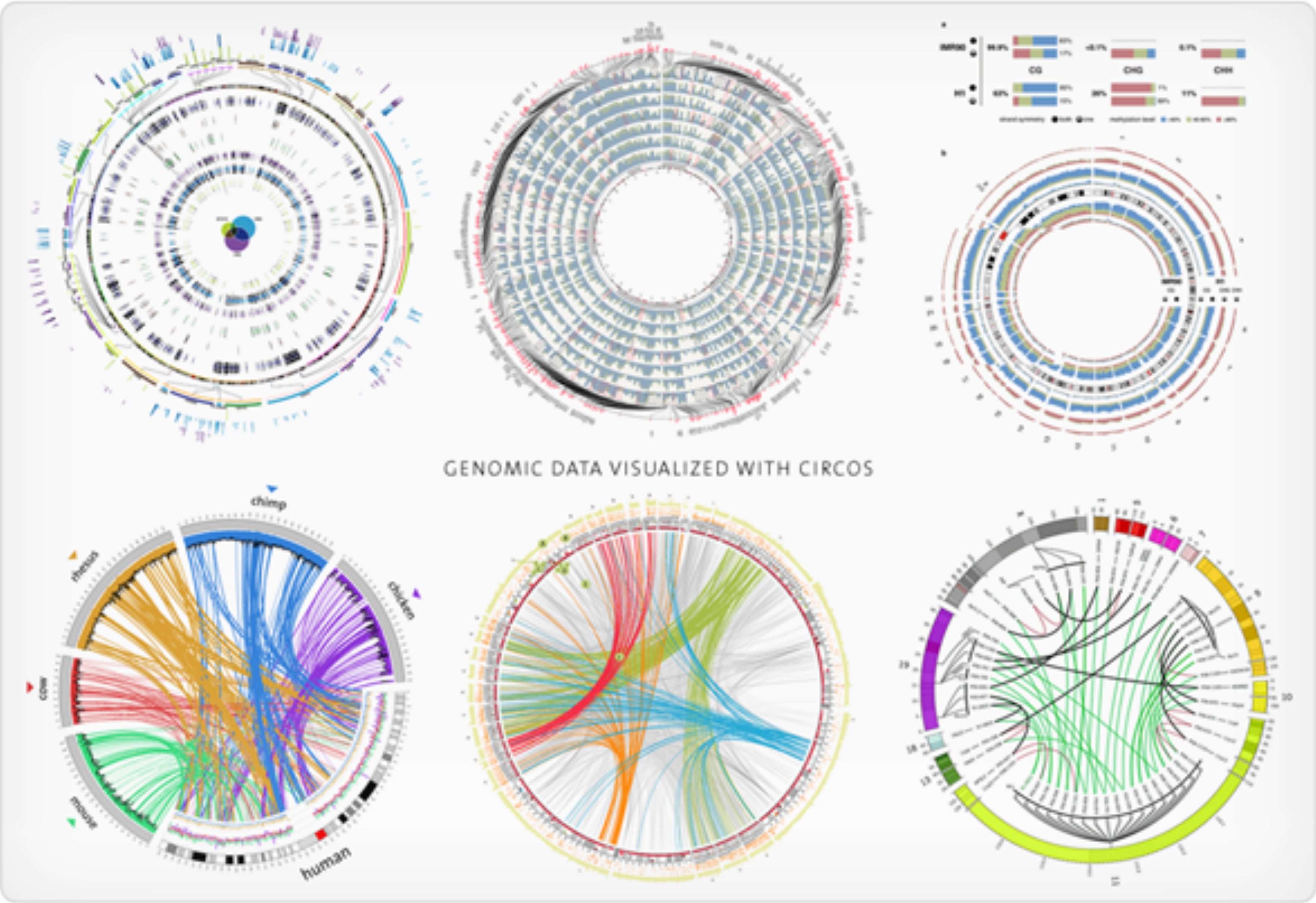
COMPREHENSIVE MAPS OF THE ENTIRE DNA METHYLOME OF FETAL LUNG FIBROBLASTS (IMR90) AND ES CELLS (H1).

Costello, J.F., M. Krzywinski, and M.A. Marra, A first look at entire human methylomes. Nat Biotechnol, 2009. 27(12): p. 1130-2.

● part 3

MAKING IT PRETTY

CIRCOS

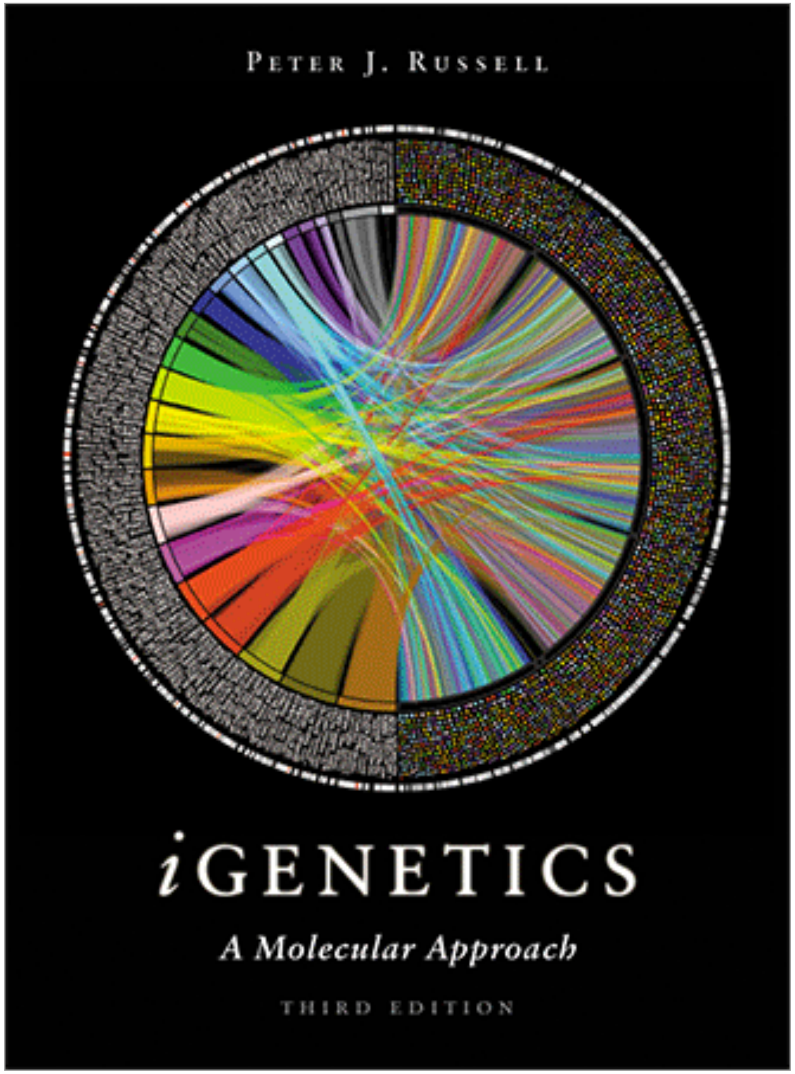


VARIOUS DATA SETS, BY CIRCOS.

American Scientist

SEPTEMBER–OCTOBER 2007

THE MAGAZINE OF SIGMA XI, THE SCIENTIFIC RESEARCH SOCIETY



PETER J. RUSSELL

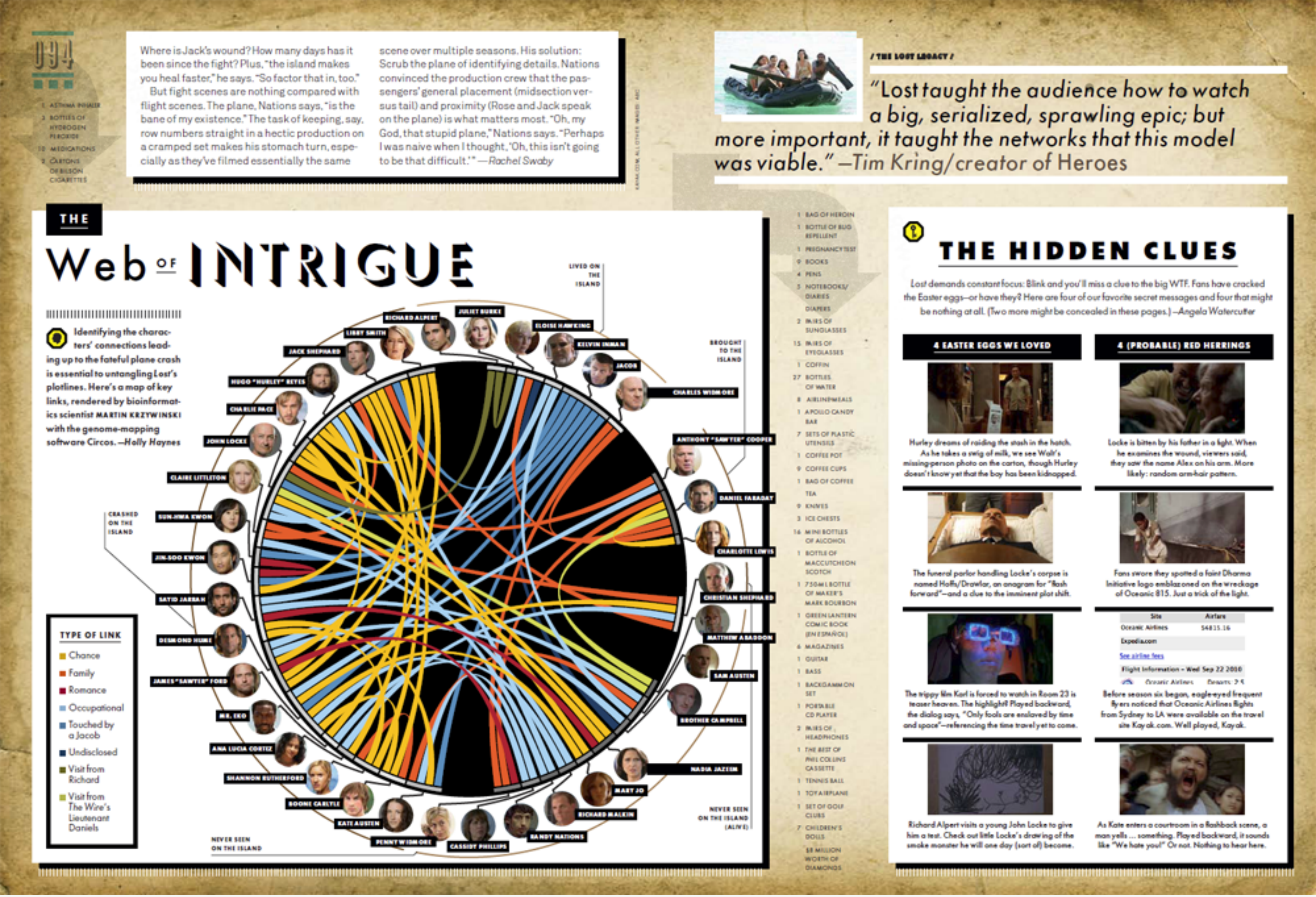
iGENETICS

A Molecular Approach

THIRD EDITION

American Scientist. September-October, 2007 cover

Russell PJ 2010 iGenetics: A Molecular Approach Benjamin Cummings



1 ASTHMA INHALER
2 BOTTLES OF HYDROGEN PEROXIDE
10 MEDICATIONS
2 CARTONS OF BILSON CIGARETTES

Where is Jack's wound? How many days has it been since the fight? Plus, "the island makes you heal faster," he says. "So factor that in, too."

But fight scenes are nothing compared with flight scenes. The plane, Nations says, "is the bane of my existence." The task of keeping, say, row numbers straight in a hectic production on a cramped set makes his stomach turn, especially as they've filmed essentially the same scene over multiple seasons. His solution: Scrub the plane of identifying details. Nations convinced the production crew that the passengers' general placement (midsection versus tail) and proximity (Rose and Jack speak on the plane) is what matters most. "Oh, my God, that stupid plane," Nations says. "Perhaps I was naive when I thought, 'Oh, this isn't going to be that difficult.' " —*Rachel Swaby*

KAYAK.COM; ALL OTHER IMAGES: ABC

/ THE LOST LEGACY /

> "Lost taught the audience how to watch a big, serialized, sprawling epic; but more important, it taught the networks that this model was viable." —Tim Kring/creator of Heroes

THE

Web of INTRIGUE

Identifying the characters' connections leading up to the fateful plane crash is essential to untangling *Lost*'s plotlines. Here's a map of key links, rendered by bioinformatics scientist **MARTIN KRZYWINSKI** with the genome-mapping software Circos. —*Holly Haynes*

TYPE OF LINK

- Chance
- Family
- Romance
- Occupational
- Touched by a Jacob
- Undisclosed
- Visit from Richard
- Visit from The Wire's Lieutenant Daniels

1 BAG OF HEROIN
1 BOTTLE OF BUG REPELLENT
1 PREGNANCY TEST
9 BOOKS
4 PENS
5 NOTEBOOKS/DIARIES
DIAPERS
2 PAIRS OF SUNGLASSES
15 PAIRS OF EYEGLASSES
1 COFFIN
27 BOTTLES OF WATER
8 AIRLINE MEALS
1 APOLLO CANDY BAR
7 SETS OF PLASTIC UTENSILS
1 COFFEE POT
9 COFFEE CUPS
1 BAG OF COFFEE
TEA
9 KNIVES
3 ICE CHESTS
16 MINI BOTTLES OF ALCOHOL
1 BOTTLE OF MACCUTCHEON SCOTCH
1 750ML BOTTLE OF MAKER'S MARK BOURBON
1 GREEN LANTERN COMIC BOOK (EN ESPAÑOL)
6 MAGAZINES
1 GUITAR
1 BASS
1 BACKGAMMON SET
1 PORTABLE CD PLAYER
2 PAIRS OF HEADPHONES
1 THE BEST OF PHIL COLLINS CASSETTE
1 TENNIS BALL
1 TOY AIRPLANE
1 SET OF GOLF CLUBS
7 CHILDREN'S DOLLS
$8 MILLION WORTH OF DIAMONDS

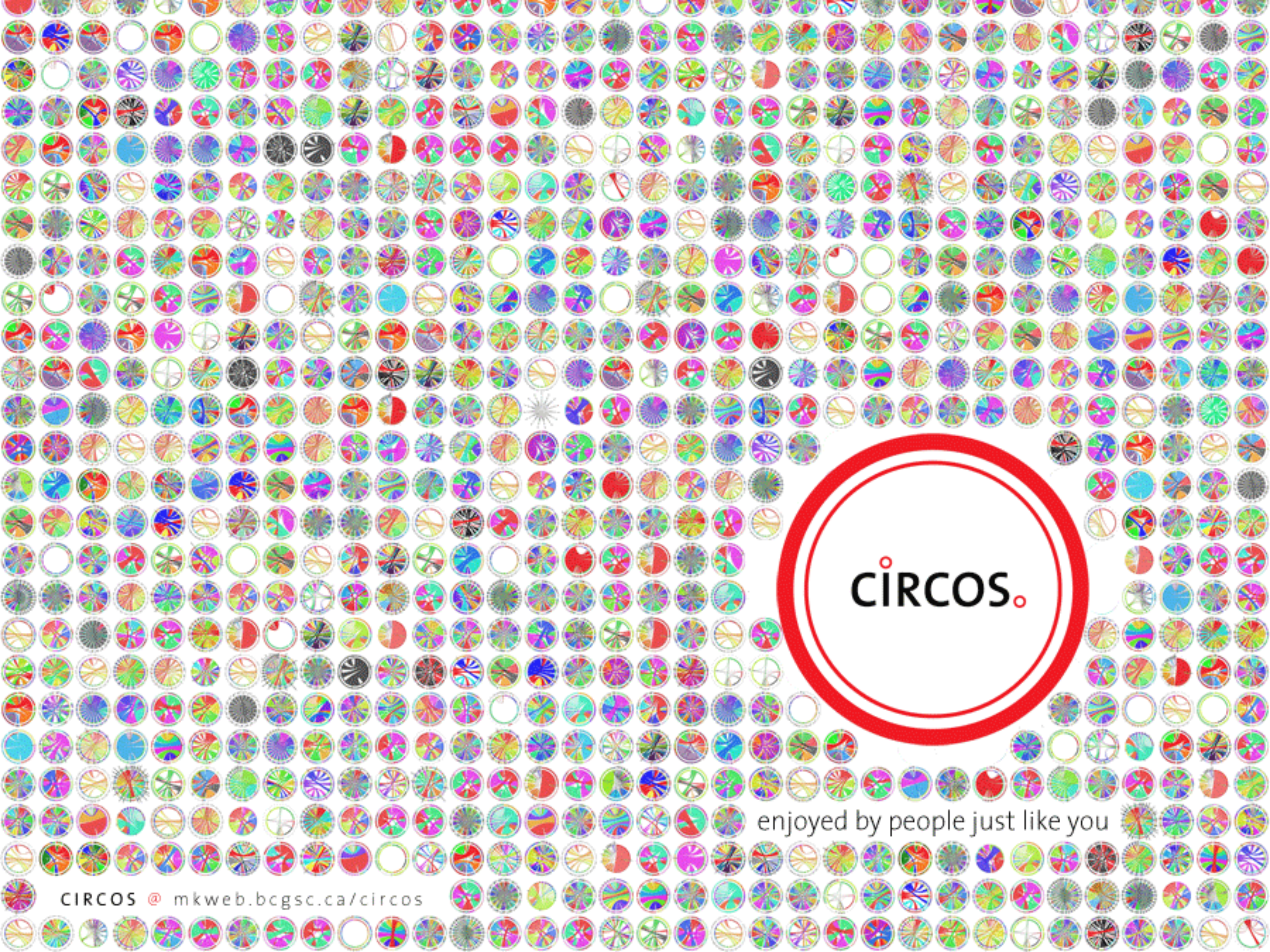
THE HIDDEN CLUES

Lost demands constant focus: Blink and you'll miss a clue to the big WTF. Fans have cracked the Easter eggs—or have they? Here are four of our favorite secret messages and four that might be nothing at all. (Two more might be concealed in these pages.) —*Angela Watercutter*

4 EASTER EGGS WE LOVED	4 (PROBABLE) RED HERRINGS
Hurley dreams of raiding the stash in the hatch. As he takes a swig of milk, we see Walt's missing-person photo on the carton, though Hurley doesn't know yet that the boy has been kidnapped.	Locke is bitten by his father in a fight. When he examines the wound, viewers said, they saw the name Alex on his arm. More likely: random arm-hair pattern.
The funeral parlor handling Locke's corpse is named Hoffs/Drawlar, an anagram for "flash forward"—and a clue to the imminent plot shift.	Fans swore they spotted a faint Dharma Initiative logo emblazoned on the wreckage of Oceanic 815. Just a trick of the light.
The trippy film Karl is forced to watch in Room 23 is teaser heaven. The highlight? Played backward, the dialog says, "Only fools are enslaved by time and space"—referencing the time travel yet to come.	Before season six began, eagle-eyed frequent flyers noticed that Oceanic Airlines flights from Sydney to LA were available on the travel site Kayak.com. Well played, Kayak.
Richard Alpert visits a young John Locke to give him a test. Check out little Locke's drawing of the smoke monster he will one day (sort of) become.	As Kate enters a courtroom in a flashback scene, a man yells ... something. Played backward, it sounds like "We hate you!" Or not. Nothing to hear here.

Wired. April 2010.

DHL billboard



CIRCOS

enjoyed by people just like you

CIRCOS @ mkweb.bcgsc.ca/circos