

DRAWING DATA

CREATING INFORMATION-RICH, INFORMATIVE AND APPEALING FIGURES FOR PUBLICATION AND PRESENTATION

martin krzywinski

canada's michael smith genome sciences center www.bcgsc.ca 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.





quality of communication

⊖ GOOD

- O BETTER
- O BEST





information-rich informative

Chromosome colors:

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.

information-rich informative



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.



information-rich informative

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.



next 60 minutes



- part 2 MAKING IT CLEAR
- part 3
 MAKING IT PRETTY

SIMPLIFY & DECLUTTER

parsability resolution color visibility text orientation

RESTRUCTURE

elegance spacing and layout focus on data entry point

·····	•••	••••••
LEGIBLE C	LEAR	ATTRACTIVE
	redundancy scale legend message consistency	
excess inl	REFINE	
scale	2	
representation		
glyph		
cart juni		

LACK OF CLEAR MESSAGE



VAGUE AND FLOWERY COMMUNICATION

decoration and ornament – dilutes your message



CHILDREN MUST BE ACCOMPANIED BY AN ADULT



KEEP OFF GRASS

cart junk –unnecessary and obfuscating visual elements

REDUNDANCY





Would removing this sign remove the need for it?



	ANN/GA ¹	BayesANN ²	SVM	
MI (25 features)	76.3%	$71.1\% \pm 0.08\%$	74.9%	
SAM (25 features)	73.3%	$80.5\% \pm 0.009\%$	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.



VALIDATION OF DNASE CLUSTERS BY REAL-TIME PCR.

Crawford, G.E., et al., Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res, 2006. 16(1): p. 123-31.



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.**



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.



THE LARGEST E- AND N-GENE FAMILIES IN YEAST.

Shakhnovich, B.E. and E.V. Koonin, Origins and impact of constraints in evolution of gene families. Genome Res, 2006. 16(12): p. 1529-36.





O part 2 MAKING IT CLEAR

○ part 3 MAKING IT PRETTY







There is no clear *emergent* pattern here.

This figure is not *parsable*.

A figure should reduce complexity for the reader.

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.



Too complicated and *illegible*.

Where should I look?

INTEGRATED FILAMENTATION NETWORK.

Prinz, S., et al., Control of yeast filamentous-form growth by modules in an integrated molecular network. Genome Res, 2004. 14(3): p. 380-90.



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.

If this were random, would I know?



HAPLOTYPES OF 148 RILS PLUS PARENTAL GENOTYPES.

West, M.A., et al., High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. Genome Res, 2006. 16(6): p. 787-95.



SEQUENCE FEATURES IN THE REGION SURROUNDING TBSD.



An *emergent pattern* in raw data.

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

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

Part 1 MAKING IT LEGIBLE

Chromosome 2 Chromosome 12



SEGMENTS SHARED BETWEEN TWO OF THE FOUR HOX-BEARING CHROMOSOMES (CHROM2 AND CHROM12).

Panopoulou, G., et al., New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. Genome Res, 2003. **13(6A): p. 1056-66.**



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



А

(A) A FRAGMENT OF A MADE-UP REPEAT GRAPH FORMED BY THREE DIVERGENT COPIES OF A REPEAT. (B) A FRAGMENT OF THE REAL REPEAT GRAPH OF *E. COLI* ILLUSTRATING THAT TRANSFORMATION OF MATE-PAIRS INTO MATE-READS MAY FAIL IN SOME CASES.

Chaisson, M.J., D. Brinza, and P.A. Pevzner, De novo fragment assembly with short mate-paired reads: Does the read length matter? Genome Res, 2009. 19(2): p. 336-46.





COLOR

SOME HUES APPEAR BRIGHTER

We do not perceive all hues to have equal brightness





yellow appears brighter

YELLOW COMMANDS ATTENTION – IS THIS INTENDED?



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.**

DIFFERENCE IN HUE IS NOT PERCEPTUALLY UNIFORM



LACK OF PERCEPTUAL DIFFERENCE



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.

ENCHANCED PERCEPTUAL DIFFERENCE



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.

COLORS WITH SIMILAR LUMINANCE ARE ILLEGIBLE



BLACK TEXT ON A DARK COLOR – HARD TO READ



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.

MAXIMIZE LUMINANCE DIFFERENCE

EGF



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.

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)





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.**

GREY BACKGROUND INTERFERES WITH LEGIBILITY



DISTRIBUTION OF HUMAN POLYMORPHISM FOR HEXAMERS IN THE 5' SPLICE SITE USING THE COMBINED DATA SETS.

Lomelin, D., E. Jorgenson, and N. Risch, Human genetic variation recognizes functional elements in noncoding sequence. Genome Res, 2010. 20(3): p. 311-9.





MODIFIED

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.



ORIGINAL MODIFIED

DUAL CHANNEL IMAGE OF AUTORADIOGRAMS (PROTEIN SYNTHESIS) OF *B. SUBTILIS* 168 (RED) AND THE ISOGENIC *SIGB*-MUTANT STRAIN ML6 (GREEN), TRANSFORMED WITH THE DELTA2D TWO-DIMENSIONAL GEL ANALYSIS SOFTWARE

Bernhardt, J., et al., Bacillus subtilis during feast and famine: visualization of the overall regulation of protein synthesis during glucose starvation by proteome analysis. Genome Res, 2003. **13(2): p. 224-37.**

AVOID OCCLUDING ONE DATA SET BY ANOTHER







dense clusters of points

overlapping glyphs

poor glyph choice



THE DISTRIBUTION OF VALUES OF THE LOGARITHM OF THE NEUTRALITY INDEX (NI) AND TAJIMA'S *D* IN 1000 SIMULATIONS UNDER A NEUTRAL MODEL (○) AND A MODEL INVOLVING REPEATED SELECTIVE SWEEPS (+)

Nielsen, R., et al., Darwinian and demographic forces affecting human protein coding genes. Genome Res, 2009. **19(5): p. 838-49.**

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.**



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.



FOUR DIFFERENT DISPLAYS OF A PAIR OF MICROARRAYS, USING HISTOGRAM SMOOTHING AND PLOTTING OF 1000 (OF 12 625) DATA POINTS.

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.

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.**



SCATTERPLOTS OF THE RESIDUALS FROM THE FULL MULTIPLE LINEAR REGRESSION OF NON-REPEAT HUMAN-CHIMPANZEE DIVERGENCE ON EACH SIGNIFICANT PREDICTOR VARIABLE.

Hellmann, I., et al., Why do human diversity levels vary at a megabase scale? Genome Res, 2005. 15(9): p. 1222-31.



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.



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.



TEXT ORIENTATION

2.5-	A.		1	1	1	1	8 8 8 8	1	1		1	1	1		1	E E	Ť	1	1	1	1 1 1	1	1	
2-	1 1	1 1 1 1 1 1	1	8 9 8	2	t I	1 1 1 1 1 1	1	1 1		1	1	Ť		t	r r		Ť	1	5 1 1	- ቅ	-	*	
1.5-	8		1	1	t t	t t		1	1	1	- †	· · ·		Ť Į		Ţ	\$	\$	Ā	蔥	眷	4	Ţ	
-1	1 	1 1 1 1	1	1	+	+	+ T	Ţ	, †		A		A *			+			<u> </u>	<u> </u>		1	1	
0.5-	\$	▲ ↓	4	₽	¢.			*	4A 1	+	- 1 1	1	Ļ		+ 1 1	r L	1	+	1 1	1	1 1	1	i I I	
-0	+ 	+ 1 	1 	1	+	 	 		1 1	 	 	1 1	1 	F 1 F F	1 	i 	 	1		1	1	 	1	
2	D.	1 1 1 1 1 1	1	1]]	1 1 2		8 8 8 8		1 		l l	1	1	1 1 1 1 1 1	Ť	1	 +	Ť	1		Ţ	- - 本	4	
1.5-	1		3	3 1 1	1	1			1		+			t		Ť.		8	Ţ	Å		Ţ		
-1-	1 1 1		1	1	+	+			1	T	- 4					i 1					<u> </u>	1	1_ 1	
0.5-	Ť	▲ ↓	4			•			▲ 						Ļ	1		+		2 2 8 7 8 8 8 8 8 8 8	1	8 2 8 8 8	1 1 1 1 1	
-0 S	+ SS	SS S	ss	SS	t	SS	s ap	s cd ss	Ę	te SS	ss ti	us:	52	- IIC	is	u	le	ntal	uc SIS	tht SIS	SS	t	ü	
scess	oroce	onda proce	ivati	oroce	anspc	oroce	proce leosi	ssign	anspc	oroce	oroce	izatio	energ	roce	ostas	icatio	ll cyc	nism	izatio	genes	tion	anspc	riptic	
al pro	olic J	olic polic p	d der	etic J	on tra	cal_1	olic p nucle	ollo no as cal_j	н Н	olic j arbol	olic J on tra	Irgan Dioj	ofprand	olic p	ome	libor	ce	orga	l Bio	r con	seula ical p	in tra	ransc	
ogica	letab	letab letab	o aci	synth	lectro	ologi	base	ologi	1	letab C	ietab i	ton c and	ation solic	inal t letab	cell l	ein n		lular	elle c and	ellula on an	olog	prote	ц	
biol	m bic	88	amin	bios	G	bio	otide	n bid	•	ein n	Ľ	skele	metal	sus VA m		prot		lticel	rgan	nizati	bi			
	lij						nucle		8	prot		cyto	- - -	D				nm	0	orga				

THE SLOPES FROM A SIMPLE LINEAR MODEL (A) AND A ROBUST LINEAR MODEL (B) FOR THE REGRESSION OF THE AVERAGE VERTEBRATE FAMILY SIZE (N_v) AGAINST THE AVERAGE INVERTEBRATE FAMILY SIZE (N_i) FOR EACH GOSLIM BIOLOGICAL PROCESS CATEGORY IN THE V.MCL GENE FAMILIES.

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.



FALSE PREDICTION PERCENTAGES FOR GENES IN PARTICULAR CELLULAR COMPARTMENTS (A) AND PARTICULAR METABOLIC SUBSYSTEMS (B).

Duarte, N.C., M.J. Herrgard, and B.O. Palsson, Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. Genome Res, 2004. 14(7): p. 1298-309.



CONSTRAINED BASES IN EACH ENCODE REGION.

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): p. 760-74.**

BOX PLOT OF $\Psi = D_T/D_S$ VALUES ARRANGED BY GO CELLULAR COMPONENT TERM.

Hoffman, M.M. and E. Birney, An effective model for natural selection in promoters. Genome Res, 2010. 20(5): p. 685-92.



nex

○ part 1 MAKING IT LEGIBLE

part 2
MAKING IT CLEAR

○ part 3 MAKING IT PRETTY

REDUNDANCY

part 2
MAKING IT CLEAR





CONSENSUS PHYLOGENETIC TREE OF CLAUDIN PROTEINS. figure modified from original

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



CONSENSUS PHYLOGENETIC TREE OF CLAUDIN PROTEINS. figure modified from original

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



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.



MHQ ENT RHT QNT QNE QTA RTR QTT QHT QTT RHT LQT RHR QTK QHT NIR



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

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.






(A) DISTRIBUTION OF THE ABSOLUTE VALUE OF THE STANDARDIZED CONTRASTS FROM THE 18S RDNA TREE. (B) QUANTILE DISTRIBUTION OF THE ABSOLUTE VALUE OF THE STANDARDIZED CONTRASTS.

Oliver, M.J., et al., The mode and tempo of genome size evolution in eukaryotes. Genome Res, 2007. 17(5): p. 594-601.



Friedrichs, F., et al., HBEGF, SRA1, and IK: Three cosegregating genes as determinants of cardiomyopathy. Genome Res, 2009. 19(3): p. 395-403.



UTILIZATION OF GENETICS OF GENE EXPRESSION DATA IN ORDER TO PRIORITIZE CANDIDATE GENES UNDERLYING A LOCUS OF INTEREST.

Doss, S., et al., Cis-acting expression quantitative trait loci in mice. Genome Res, 2005. 15(5): p. 681-91.



Roh, T.Y., et al., Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Res, 2007. 17(1): p. 74-81.



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.



BLAST SIMILARITY SEARCH RESULTS FOR THE PREDICTED PROTEOMES OF NCLDVS. legend modified from original

Ogata, H. and J.M. Claverie, Unique genes in giant viruses: regular substitution pattern and anomalously short size. Genome Res, 2007. 17(9): p. 1353-61.



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.**



- specific in situ result
- △ selective in situ result
- ▽ nonselective in situ result
- specific in litterature
- × nonspecific in litterature



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



CONSISTENCY



HG is blue in (A).

HG is yellow in (B).

(A) THE FOUR GENEALOGY TYPES ASSOCIATED WITH THE STATES OF THE HIDDEN MARKOV MODEL.(B) 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.



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."



CONCENTRIC VS. 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.**



OUTLIERS



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. figure modified from original

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

relative ratio of larva stages

relative ratio of larva stage



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.**



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.



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.



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.



Negre, B., et al., Conservation of regulatory sequences and gene expression patterns in the disintegrating Drosophila Hox gene complex. Genome Res, 2005. **15(5): p. 692-700.**



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.

Į. 0 0 0 0 0 0 1 1

INDIVIDUALS

INDIVIDUALS Į R 0 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.

Ρ

O L

s

INDIVIDUALS





Ρ

0

0

L

s



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



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.


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.



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. figure modified from original

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.**

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





THE HIERARCHICAL LAYOUT OF THE E. COLI TRANSCRIPTIONAL REGULATORY NETWORK AND THE LINUX CALL GRAPH. figure modified from original



THE HIERARCHICAL LAYOUT OF THE E. COLI TRANSCRIPTIONAL REGULATORY NETWORK AND THE LINUX CALL GRAPH. figure modified from original

NETWORK LAYOUT



THE HIERARCHICAL LAYOUT OF THE E. COLI TRANSCRIPTIONAL REGULATORY NETWORK AND THE LINUX CALL GRAPH. figure modified from original



LINUX



THE HIERARCHICAL LAYOUT OF THE E. COLI TRANSCRIPTIONAL REGULATORY NETWORK AND THE LINUX CALL GRAPH. *figure modified from original*



DIOMS



DISTRIBUTION OF HUMAN POLYMORPHISM FOR HEXAMERS IN THE 5' SPLICE SITE USING THE COMBINED DATA SETS.

Lomelin, D., E. Jorgenson, and N. Risch, Human genetic variation recognizes functional elements in noncoding sequence. Genome Res, 2010. 20(3): p. 311-9.



DISTRIBUTION OF HUMAN POLYMORPHISM FOR HEXAMERS IN THE 5' SPLICE SITE USING THE COMBINED DATA SETS. figure modified from original

Lomelin, D., E. Jorgenson, and N. Risch, Human genetic variation recognizes functional elements in noncoding sequence. Genome Res, 2010. 20(3): p. 311-9.



DISTRIBUTION OF HUMAN POLYMORPHISM FOR HEXAMERS IN THE 5' SPLICE SITE USING THE COMBINED DATA SETS. figure modified from original

Lomelin, D., E. Jorgenson, and N. Risch, Human genetic variation recognizes functional elements in noncoding sequence. Genome Res, 2010. 20(3): p. 311-9.



CART JUNK



SCHEMATIC REPRESENTATION OF THE GENOME ORGANIZATION OF MAMMALIAN RETROELEMENTS.

Belancio, V.P., D.J. Hedges, and P. Deininger, Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res, 2008. 18(3): p. 343-58.



SCHEMATIC REPRESENTATION OF THE GENOME ORGANIZATION OF MAMMALIAN RETROELEMENTS. *figure modified from original*

Belancio, V.P., D.J. Hedges, and P. Deininger, Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res, 2008. 18(3): p. 343-58.



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.



POSSIBLE PATTERNS OF NUCLEOSOME POSITIONING. figure modified from original

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.



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*



HIGH-RESOLUTION ANALYSIS OF THE DCW GENE CLUSTER TRANSCRIPTS.

Selinger, D.W., et al., Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. Genome Res, 2003. 13(2): p. 216-23.



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



ORIGINAL MODIFIED GLYPH PALETTE

COMPARISON OF K_A/K_S 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 K_A/K_S RATIO AND SNP DENSITY FOR GENES AND PSEUDOGENES. figure modified from original

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

D part 2 MAKING IT CLEAR

part 3
MAKING IT PRETTY

ELEGANCE







RESULTS OF THE ALTERNATE SPLICING EXPERIMENT. figure modified from original

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. figure modified from original

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



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. figure modified from original

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

SPACING & LAYOUT





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

FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS. *figure modified from original*



FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS. *figure modified from original*



FREQUENCIES (%) OF Y-CHROMOSOMAL HAPLOGROUPS AMONG ETHNIC POPULATIONS. *figure modified from original*












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.

а

