

References

Introduction and Review of literature

- Adam M Gustafson, Evan S Snitkin, Stephen CJ Parker, Charles DeLisi and Simon Kasif, Towards the identification of essential genes using targeted genome sequencing and comparative analysis, *BMC Genomics* 2006, 7:265
- Bernhard Ø. Palsson, Experimental and Computational Assessment of Conditionally Essential Genes in *Escherichia coli*, *J Bacteriol.* 2006 December; 188(23): 8259–8271
- Bernhard Schölkopf and A. J. Smola: Learning with Kernels. MIT Press, Cambridge, MA, 2002
- Brown, M.P., Grundy, W.N., Lin, D., et al. 2000, Knowledge-based analysis of microarray gene expression data by using support vector machines, *Proc. Natl Acad. Sci. USA*, 97, 262–7.

- De Las Rivas J, Fontanillo C. "Protein-protein interactions essentials: key concepts to building and analyzing interactome networks.". (2010) PLoS Comput Biol. 6(6):e1000807.
- Engel,S.R. and Cherry,J.M. The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple *Saccharomyces cerevisiae* strains at the Saccharomyces Genome Database. Database (2013) Vol. 2013
- Gabriel del Rio, Dirk Koschützki and Gerardo Coello, How to identify essential genes from molecular networks?,*BMC Systems Biology* 2009, 3:102
- Gerald Karp, Cell and Molecular Biology: Concepts and Experiments
- John I. Glass , Essential genes of a minimal bacterium, PNAS January 10, 2006 vol. 103 no. 2 425-430
- Kempthues K. Essential Genes (December 24, 2005), *WormBook*, ed. The *C. elegans* Research Community, WormBook
- Kitiporn Plaimas, Roland Eils and Rainer König,Identifying essential genes in bacterial metabolic networks with machine learning methods, *BMC Systems Biology* 2010, 4:56
- Koonin EV.How many genes can make a cell: the minimal-gene-set concept. Annu. Rev. Genomics Hum. Genet. 2000;1:99-116
- Ren Zhang,DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes,Nucl. Acids Res. (2009) 37 (suppl 1): D455-D458
- Sarangi AN, Aggarwal R, Rahman Q, Trivedi N (2009) Subtractive Genomics Approach for *in Silico* Identification and Characterization of Novel Drug Targets in *Neisseria Meningitidis Serogroup B*. J Comput Sci Syst Biol 2: 255-258.

- Steffen Heber and Soma Saha, *In silico* prediction of yeast deletion phenotypes, *Genetics and Molecular Research* 5 (1): 224-232 (2006)
- Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection www.ecoli.naist.jp/gb6/Resources/deletion/deletion.html
- Xiaodong Gong , Shaohua Fan , Amy Bilderbeck , Mingkun Li , Hongxia Pang , Shiheng Tao, Comparative analysis of essential genes and nonessential genes in *Escherichia coli* K12, *Mol Genet Genomics* (2008) 279:87–94
- Yu Chen and Dong Xu, Understanding protein dispensability through machine-learning analysis of high-throughput data, *Bioinformatics*, Vol. 21 no. 5 2005, pages 575–581
- Yih-Chii Hwang,, Chen-Ching Lin, Jen-Yun Chang, Hirotada Mori, Hsueh-Fen Juan and Hsuan-Cheng Huang, Predicting essential genes based on network and sequence analysis, *Mol. BioSyst.*, 2009, 5, 1672–1678

Material and methods

- Mewes, H.W., Hani, J., Pfeiffer, F., and Frishman, D. 1998. MIPS: A database for genomes and protein sequences. *Nucl. Acids Res.* 26, 33–37
- I Xenarios, D W Rice, L Salwinski, M K Baron, E M Marcotte, and D Eisenberg. DIP: the database of interacting proteins. *Nucleic Acids Res*, 28: 289–291, 2000.
- Uetz, P., Giot, L., Cagney, G., et al. 2000. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* 403(6770), 623–627.
- Gavin, A.C., Bosche, M., Krause, R., et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415(6868), 141–147.
- C. Gavin, et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440(7084):631–636, 2006.
- Y Ho, A Gruhler, et al. . Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, 415:180–183, 2002.
- Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F., and Gerstein, M. 2003. A Bayesian networks approach for predicting protein–protein interactions from genomic data. *Science* 302(5644), 449–453.
- P Erdos and A Renyi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–61, 1960.
- A L. Barabasi and R Albert. Emergence of scaling in random networks. *Science*, 286: 509–512, 1999.
- G. del Rio, D. Koschutzki, G. Coello, How to identify essential genes from molecular networks?, *BMC Syst. Biol.* 3 (2009) 102. doi:10.1186/1752-0509-3-102.

- M. Tyers, A. Breitkreutz, C. Stark, T. Reguly, L. Boucher, B.-J. Breitkreutz, BioGRID: a general repository for interaction datasets, *Nucl. Acids Res.* 34 (2006) D535–539. doi:10.1093/nar/gkj109.
- V. Batagelj, A. Mrvar, Pajek-Program for Large Network Analysis. *Connections.* 1998;21:47–57.
- G. Csardi, T. Nepusz. The igraph software package for complex network research. *Inter J, Complex Systems.* 2006;1(1):1695
- D.J. Lipman, A. Souvorov, E. V Koonin, A.R. Panchenko, T.A. Tatusova, The relationship of protein conservation and sequence length, *BMC Evol. Biol.* 2 (2002) 20. doi:10.1186/1471-2148-2-20.
- Hirsh AE, Fraser HB: Protein dispensability and rate of evolution. *Nature* 2001, 411:1046-1049
- D.J. Lipman, A. Souvorov, E. V Koonin, A.R. Panchenko, T.A. Tatusova, The relationship of protein conservation and sequence length, *BMC Evol. Biol.* 2 (2002) 20. doi:10.1186/1471-2148-2-20.
- J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.* 157 (1982) 105–132. doi:http://dx.doi.org/10.1016/0022-2836(82)90515-0.
- Campen, A., Williams, R.M., Brown, C.J., Meng, J., Uversky, V.N. and Dunker, A.K. 2008. TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder, *Protein Pept Lett.* 15(9): 956–963.
- Tompa, P. 2002. Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27, 527–533

- K. Gunasekaran, C.-J. Tsai, S. Kumar, D. Zanuy, R. Nussinov, Extended disordered proteins: targeting function with less scaffold., *Trends Biochem. Sci.* 28 (2003) 81–85. doi:10.1016/S0968-0004(03)00003-3.
- Dunker, A.K. *et al.* 2002. Intrinsic disorder and protein function. *Biochemistry* 41, 6573–6582
- Demchenko, A.P. 2001. Recognition between flexible protein molecules: induced and assisted folding. *J. Mol. Recognit.* 14, 42–61
- Wright, P.E. and Dyson, H.J. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, 293, 321–331
- Dyson, H.J. and Wright, P.E. 2002. Coupling folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.*, 12, 54–60
- Campen, A., Williams, R.M., Brown, C.J., Meng, J., Uversky, V.N. and Dunker, A.K. 2008. TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder, *Protein Pept Lett.* 15(9): 956–963.
- Wang, G., Dong, Y., Chen, H., Zhang, H., Song, Y., Zhang, H. and Chen, W. 2011. High Incidence of Disorder-Promoting Amino Acids in the Amino Terminus of Mature Proteins in *Bacillus subtilis*, *American Journal of Biochemistry and Biotechnology*, 7 (4): 172-178, 2011
- Ferron, F., Longhi, S., Canard, B. and Karlin D. 2006. A Practical Overview of Protein Disorder Prediction, *PROTEINS: Structure, Function, and Bioinformatics*, 65:1–14.
- Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J. Dunker, A.K. 2001. Sequence complexity of disordered protein, *Proteins*. 2001 Jan 1;42(1):38-48.
- Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Uversky, V.N. and Obradovic, Z. 2007. Functional anthology of intrinsic disorder. 1. Biological processes

and functions of proteins with long disordered regions, *J Proteome Res.*, May;6(5):1882-98.

- J. Prilusky, C.E. Felder, T. Zeev-Ben-Mordehai, E.H. Rydberg, O. Man, J.S. Beckmann, I. Silman, J.L. Sussman, FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded, *Bioinformatics.* 21 (2005) 3435–3438. doi:10.1093/bioinformatics/bti537.
- J. Mrázek, S. Karlin, Strand compositional asymmetry in bacterial and large viral genomes, *Proc. Natl. Acad. Sci. U. S. A.* 95 (1998) 3720–3725.
- Y. Lin, F. Gao, C.-T. Zhang, Functionality of essential genes drives gene strand-bias in bacterial genomes, *Biochem. Biophys. Res. Commun.* 396 (2010) 472–476. doi:http://dx.doi.org/10.1016/j.bbrc.2010.04.119.
- S. Hirose, S. Hiraga, T. Okazaki, Initiation site of deoxyribonucleotide polymerization at the replication origin of the Escherichia coli chromosome., *Mol. Gen. Genet.* 189 (1983) 422–431.
- M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray, *Science* (80-.). 270 (1995) 467–470. doi:10.1126/science.270.5235.467.
- M.B. Eisen, P.O. Brown, DNA arrays for analysis of gene expression., *Methods Enzymol.* 303 (1999) 179–205.
- S. Ghaemmaghami, W.-K. Huh, K. Bower, R.W. Howson, A. Belle, N. Dephoure, E.K. O’Shea, J.S. Weissman, Global analysis of protein expression in yeast., *Nature.* 425 (2003) 737–741. doi:10.1038/nature02046.

- J.R.S. Newman, S. Ghaemmaghami, J. Ihmels, D.K. Breslow, M. Noble, J.L. DeRisi, J.S. Weissman, Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise., *Nature*. 441 (2006) 840–846. doi:10.1038/nature04785.
- J. Ivanic, X. Yu, A. Wallqvist, J. Reifman, Influence of protein abundance on high-throughput protein-protein interaction detection., *PLoS One*. 4 (2009) e5815. doi:10.1371/journal.pone.0005815.
- S. Chakraborty, T.C. Ghosh, Evolutionary rate heterogeneity of core and attachment proteins in yeast protein complexes, *Genome Biol. Evol.* 5 (2013) 1366–1375. doi:10.1093/gbe/evt096.
- M. Wang, M. Weiss, M. Simonovic, G. Haertinger, S.P. Schrimpf, M.O. Hengartner, C. von Mering, PaxDb, a Database of Protein Abundance Averages Across All Three Domains of Life, *Mol. Cell. Proteomics*. 11 (2012) 492–500. doi:10.1074/mcp.O111.014704.
- D. Greenbaum, C. Colangelo, K. Williams, M. Gernstein, Comparing protein abundance and mRNA expression levels on a genomic scale, *Genome Biol.* 4 (2003) 117. doi:10.1186/gb-2003-4-9-117.
- X. Zhu, X. Wu, Class Noise vs. Attribute Noise: A Quantitative Study, *Artif. Intell. Rev.* 22 (2004) 177–210. doi:10.1007/s10462-004-0751-8.
- S. Durinck, W. Huber, The biomaRt user 's guide, *Database*. (2010) 1–21.
- D. Smedley, S. Haider, B. Ballester, R. Holland, D. London, G. Thorisson, A. Kasprzyk, BioMart--biological queries made easy., *BMC Genomics*. 10 (2009) 22. doi:10.1186/1471-2164-10-22.

- G.R. Abecasis, D. Altshuler, A. Auton, L.D. Brooks, R.M. Durbin, R.A. Gibbs, M.E. Hurles, G.A. McVean, A map of human genome variation from population-scale sequencing., *Nature*. 467 (2010) 1061–1073. doi:10.1038/nature09534.
- M. Boutros, A.A. Kiger, S. Armknecht, K. Kerr, M. Hild, B. Koch, S.A. Haas, R. Paro, N. Perrimon, Genome-wide RNAi analysis of growth and viability in *Drosophila* cells., *Science*. 303 (2004) 832–835. doi:10.1126/science.1091266.
- R.S. Kamath, A.G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, D.P. Welchman, P. Zipperlen, J. Ahringer, Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi., *Nature*. 421 (2003) 231–237. doi:10.1038/nature01278.
- E.A. Winzler, D.D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, R. Benito, J.D. Boeke, H. Bussey, Functional characterization of the *Saccharomyces cerevisiae* genome by gene deletion and parallel analysis, *Science* (80-). 285 (1999). doi:10.1126/science.285.5429.901.
- F.C. Holstege, E.G. Jennings, J.J. Wyrick, T.I. Lee, C.J. Hengartner, M.R. Green, T.R. Golub, E.S. Lander, R.A. Young, Dissecting the regulatory circuitry of a eukaryotic genome., *Cell*. 95 (1998) 717–728.
- F.H. Crick, Codon--anticodon pairing: the wobble hypothesis., *J. Mol. Biol.* 19 (1966) 548–555.
- R. Grantham, C. Gautier, M. Gouy, M. Jacobzone, R. Mercier, Codon catalog usage is a genome strategy modulated for gene expressivity, *Nucleic Acids Res.* 9 (1981) 213. doi:10.1093/nar/9.1.213-b.
- J.O. McInerney, GCUA: general codon usage analysis., *Bioinformatics.* 14 (1998) 372–373.

- Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T: Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *Journal of molecular evolution* 2001, 53(4):290-298.
- D'Onofrio G, Mouchiroud D, Aïssani B, Gautier C, Bernardi G: Correlations between the compositional properties of human genes, codon usage, and amino acid composition of proteins. *Journal of Molecular Evolution* 1991, 32(6):504-510.
- Zama M: Codon usage and secondary structure of mRNA. *Nucleic Acids Symp Ser* 1990, 22:93-94.
- Stenström CM, Jin H, Major LL, Tate WP, Isaksson LA: Codon bias at the 3'- side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*. *Gene* 2001, 263(1-2):273-284.
- Knight RD, Freeland SJ, Landweber LF: A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol* 2001, 2(4):1-13.
- Marín A, González F, Gutiérrez G, Oliver JL: Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Research* 1998, 26(19):4540.
- Yu X, Li Z, Xia X, Fang H, Zhou C, Chen H: Expression and purification of ancrod, an anticoagulant drug, in *Pichia pastoris*. *Protein Expression and Purification* 2007, 55(2):257-261.
- Lee MH, Yang SJ, Kim JW, Lee HS, Kim JW, Park KH: Characterization of a thermostable cyclodextrin glucanotransferase from *Pyrococcus furiosus* DSM3638. *Extremophiles* 2007, 11(3):537-541.

- Ravishankar Ram M, Beena G, Ragunathan P, Malathi R: Analysis of structure, function, and evolutionary origin of the ob gene product–leptin. *J Biomol Struct Dyn* 2007, 25(2):183-188.
- S. Thakur, P. Normand, V. Daubin, L.S. Tisa, A. Sen, Contrasted evolutionary constraints on secreted and non-secreted proteomes of selected Actinobacteria, *BMC Genomics*. 14 (2013) 1. doi:10.1186/1471-2164-14-474.
- Ikemura T: Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. *J Mol Biol* 1982, 158(4):573-579.
- Moriyama EN: Codon Usage Bias and tRNA Abundance in *Drosophila*. *Journal of Molecular Evolution* 1997, 45(5):514-523.
- Duret L: tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends in Genetics* 2000, 16(7):287-289.
- Dos Reis M, Wernisch L, Savva R, Journals O: Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Research* 2003, 31(23):6976-6985.
- Lavner Y, Kotlar D: Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene* 2004, 345(1):127-38.
- Singh ND, DuMont B, Vanessa L, Hubisz MJ, Nielsen R, Aquadro CF: Patterns of Mutation and Selection at Synonymous Sites in *Drosophila*. *Molecular Biology and Evolution* 2007, 24(12):2687.

- Sharp PM, Li WH: The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 1987, 15(3):1281-1295.
- Bennetzen JL, Hall BD: Codon selection in yeast. *Journal of Biological Chemistry* 1982, 257(6):3026-3031.
- J.O. McInerney, GCUA: general codon usage analysis., *Bioinformatics*. 14 (1998) 372–373. Last accessed at : <http://mcinerneylab.com/software/gcua/>
- Fraser HB, Wall DP, Hirsh AE: A simple dependence between protein evolution rate and the number of protein-protein interactions. *Bmc Evolutionary Biology* 2003, 3:11.
- Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW: Evolutionary rate in the protein interaction network. *Science* 2002, 296:750-752.
- Qiu J, Noble WS: Predicting co-complexed protein pairs from heterogeneous data. *Plos Computational Biology* 2008, 4:e1000054.
- S. Chakraborty, B. Kahali, T.C. Ghosh, Protein complex forming ability is favored over the features of interacting partners in determining the evolutionary rates of proteins in the yeast protein-protein interaction networks., *BMC Syst. Biol.* 4 (2010) 155. doi:10.1186/1752-0509-4-155.
- A.-C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L.J. Jensen, S. Bastuck, B. Dümpelfeld, A. Edlmann, M.-A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A.-M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J.M. Rick, B. Kuster, P. Bork, R.B. Russell, G. Superti-Furga, Proteome survey reveals

modularity of the yeast cell machinery., *Nature*. 440 (2006) 631–6.
doi:10.1038/nature04532.

- R. Zhang, H.-Y. Ou, C.-T. Zhang, DEG: a database of essential genes., *Nucleic Acids Res.* 32 (2004) D271–2. doi:10.1093/nar/gkh024.
- R. Zhang, Y. Lin, DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes., *Nucleic Acids Res.* 37 (2009) D455–8. doi:10.1093/nar/gkn858.
- H. Luo, Y. Lin, F. Gao, C.-T. Zhang, R. Zhang, DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements, *Nucleic Acids Res.* . 42 (2014) D574–D580. doi:10.1093/nar/gkt1131 .
- G. Giaever, A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, Functional profiling of the *Saccharomyces cerevisiae* genome, *Nature*. 418 (2002). doi:10.1038/nature00935.
- J.M. Cherry, E.L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, E.T. Chan, K.R. Christie, M.C. Costanzo, S.S. Dwight, S.R. Engel, D.G. Fisk, J.E. Hirschman, B.C. Hitz, K. Karra, C.J. Krieger, S.R. Miyasato, R.S. Nash, J. Park, M.S. Skrzypek, M. Simison, S. Weng, E.D. Wong, *Saccharomyces Genome Database: the genomics resource of budding yeast*, *Nucleic Acids Res.* 40 (2012) D700–5. doi:10.1093/nar/gkr1029.