Bibliography

Albert R and Barabási A. Statistical mechanics of complex networks. *Rev Mod Physics*, 74:47–97, 2002. 1.1.4

Aloy P, Ceulemans H, Stark A and Russell RB. The relationship between sequence and interaction divergence in proteins. *J Mol Biol*, 332:989–998, 2003. 1.1.1.4, 2.3.4, 3.3.2.1, 3.4.1

Aloy P and Russell RB. The third dimension for protein interactions and complexes. *Trends Biochem Sci*, 27:633–638, 2002. 3.4.1

Aloy P and Russell RB. Ten thousand interactions for the molecular biologist. *Nat Biotechnol*, 22:1317–1321, 2004. 1.3.1, 2.1, 2.3.7, 2.3.7, 2.4.1, 2.4.2

Aury JM, Jaillon O, Duret L, Noel B, Jubin C *et al.* Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, 444:171–178, 2006. 4.1

Axelsen JB, Lotem J, Sachs L and Domany E. Genes overexpressed in different human solid cancers exhibit different tissue-specific expression profiles. *Proc Natl Acad Sci USA*, 104:13122–13127, 2007. 4.2.4, 4.3.2

Barabasi A and Albert R. Emergence of scaling in random networks. *Science*, 286:509–512, 1999. 1.1.4

Barabasi AL and Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5:101–113, 2004. 1.1.4

Barbujani G, Magagni A, Minch E and Cavalli-Sforza LL. An apportionment of human DNA diversity. *Proc Natl Acad Sci USA*, 94:4516–4519, 1997. 1.2.2

Bartel PL and Fields S. *The Yeast Two-Hybrid System (Advances in Molecular Biology)*. Oxford University Press, USA, 1st edition, 1997. 1.1.1.2

Basu MK, Carmel L, Rogozin IB and Koonin EV. Evolution of protein domain promiscuity in eukaryotes. *Genome Res*, 18:449–461, 2008. 2.4.2

Batzer MA and Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet*, 3:370–379, 2002. 1.2.1

Berggård T, Linse S and James P. Methods for the detection and analysis of protein–protein interactions. *Proteomics*, 7:2833–2842, 2007. 1.1.1.1, 1.1.1.2

Berman HM. The Protein Data Bank: a historical perspective. *Acta Crystallogr A*, 64:88–95, 2008. 1.5

Birchler J. The genetic basis of dosage compensation of alcohol dehydrogenase-1 in maize. *Genetics*, 97:625–637, 1981. 4.1

Birchler JA, Bhadra U, Bhadra MP and Auger DL. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev Biol*, 234:275–288, 2001. 4.1

Birchler JA, Riddle NC, Auger DL and Veitia RA. Dosage balance in gene regulation: biological implications. *Trends Genet*, 21:219–226, 2005. 4.1, 4.4.2

Birchler JA, Yao H and Chudalayandi S. Biological consequences of dosage dependent gene regulatory systems. *Biochim Biophys Acta*, 1769:422–428, 2007. 4.1

Blake WJ, Kaern M, Cantor CR and Collins JJ. Noise in eukaryotic gene expression. *Nature*, 422:633–637, 2003. 4.4.2

Blekhman R, Man O, Herrmann L, Boyko AR, Indap A *et al.* Natural selection on genes that underlie human disease susceptibility. *Curr Biol*, 18:883–889, 2008. 1.2.3.1

Botstein D and Risch N. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nature Genetics*, 33:228–237, 2003. 1.2.3.1

Boyd W. Genetics and the human race: Definition of race on the basis of gene frequencies supplements definition from morphological characters. *Science*, 140:1057–1064, 1963. 1.2.3

Branden C and Tooze J. *Introduction to Protein Structure*. Garland Publishing, 2nd edition, 1991. 1.1.1.4

Bravo J and Aloy P. Target selection for complex structural genomics. *Curr Opin Struct Biol*, 16:385–392, 2006. 2.4.1

Breitkreutz BJ, Stark C, Reguly T, Boucher L, Breitkreutz A *et al.* The BioGRID interaction database: 2008 update. *Nucleic Acids Res*, 36:D637–640, 2008. 1.1.3, 2.2.1, 3.3.3

Bruford EA, Lush MJ, Wright MW, Sneddon TP, Povey S and Birney E. The HGNC database in 2008: a resource for the human genome. *Nucleic Acids Res*, 36:D445–448, 2008. 4.2.1

Buratti E, Baralle M and Baralle FE. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic Acids Res*, 34:3494–3510, 2006. 3.1

Burckstummer T, Bennett KL, Preradovic A, Schutze G, Hantschel O, Superti-Furga G and Bauch A. An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nat Methods*, 3:1013–1019, 2006. 1.1.1.1

Bushell KM, Sollner C, Schuster-Böckler B, Bateman A and Wright GJ. Large-scale screening for novel low-affinity extracellular protein interactions. *Genome Res*, 18:622–630, 2008. 1.1.1.5, 1.4

Camon E, Magrane M, Barrell D, Lee V, Dimmer E *et al.* The Gene Ontology Annotation (GOA) database: sharing knowledge in UniProt with Gene Ontology. *Nucleic Acids Res*, 32:D262–266, 2004. 4.2.9

Chakrabarti P and Janin J. Dissecting protein–protein recognition sites. *Proteins*, 47:334–343, 2002. 3.3.4.4, 3.9

Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L and Cesareni G. MINT: the Molecular INTeraction database. *Nucleic Acids Res*, 35:D572–574, 2007. 1.1.3, 2.2.1

Cheung SW, Shaw CA, Yu W, Li J, Ou Z *et al.* Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet Med*, 7:422–432, 2005. 4.2.4

Chiti F and Dobson CM. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, 75:333–366, 2006. 3.1

Chothia C. One thousand families for the molecular biologist. *Nature*, 357:543–544, 1992. 1.3

Chothia C and Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J*, 5:823–826, 1986. 1.1.1.4, 3.3.2.1

Collins FS, Guyer MS and Chakravarti A. Variations on a theme: Cataloging human DNA sequence variation. *Science*, 278:1580–1581, 1997. 3.1

Conrad DF, Andrews TD, Carter NP, Hurles ME and Pritchard JK. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet*, 38:75–81, 2006. 1.2.3

Conrad DF and Hurles ME. The population genetics of structural variation. *Nat Genet*, 39:S30–36, 2007. 4.1

Cooper GM, Nickerson DA and Eichler EE. Mutational and selective effects on copy-number variants in the human genome. *Nat Genet*, 29:S22–29, 2007. 4.1, 4.3.3, 4.3.4, 4.4.3

Cunningham BC and Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*, 244:1081–1085, 1989. 1.1, 3.2.6

Cusick ME, Klitgord N, Vidal M and Hill DE. Interactome: gateway into systems biology. *Hum Mol Genet*, 14:R171–181, 2005. 2.1

Dang V, Kassahn K, Marcos A and Ragan M. Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *Eur J Hum Genet*, 2008. 4.4.4

Darwin C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* John Murray, 1st edition, 1859. 1.2

Deane CM, Salwinski L, Xenarios I and Eisenberg D. Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics*, 1:349–356, 2002. 1.1.2

Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, Davis RW, Nislow C and Giaever G. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics*, 169:1915–1925, 2005. 4.1

Devlin RH, Holm DG and Grigliatti TA. Autosomal dosage compensation in *Drosophila melanogaster* strains trisomic for the left arm of chromosome 2. *Proc Natl Acad Sci USA*, 79:1200–1204, 1982. 4.1, 4.4.2

Disotell TR. Human genomic variation. *Genome Biol*, 1:5, 2000. 1.2.2

van Dongen S. *Graph clustering by flow simulation.* Ph.D. thesis, University of Utrecht, The Netherlands, 2000. 4.2.3

Durbin R, Eddy SR, Krogh A and Mitchison G. *Biological Sequence Analysis.* Cambridge University Press, 1998. 1.3

Eddy SR. *HMMER User's Guide: Biological sequence analysis using profile hidden Markov models, version 2.2.* Washington University School of Medicine, http://hmmer.wustl.edu, 2001. 3.2.4

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32:1792–1797, 2004. 3

Eisenberg E and Levanon E. Preferential attachment in the protein network evolution. *Phys Rev Lett*, 91:138701, 2003. 1.1.4

El Ghouzzi V, Legeai-Mallet L, Aresta S, Benoist C, Munnich A, De Gunzburg J and Bonaventure J. Saethre-Chotzen mutations cause TWIST protein degradation or impaired nuclear location. *Hum Mol Genet*, 9:813–819, 2000. 3.3.6.3

Enright AJ, Iliopoulos I, Kyrpides NC and Ouzounis CA. Protein interaction maps for complete genomes based on gene fusion events. *Nature*, 402:86–90, 1999. 1.3.1, 2.2.4

Ewing RM, Chu P, Elisma F, Li H, Taylor P *et al.* Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol Syst Biol*, 3:89, 2007. 4.2.3

Fawcett T. An introduction to ROC analysis. *Pattern Recog Lett*, 27:861–874, 2006. 3.3.2.3

Feldman MW, Lewontin RC and King MC. Race: a genetic melting-pot. *Nature*, 424:374, 2003. 1.2.2

Ferrer-Costa C, Orozco M and de la Cruz X. Characterization of disease–associated single amino acid polymorphisms in terms of sequence and structure properties. *J Mol Biol*, 315:771–786, 2002. 3.1, 3.4.2

Fields S and Song O. A novel genetic system to detect protein–protein interactions. *Nature*, 340:245–246, 1989. 1.1.1.2

Finn RD, Marshall M and Bateman A. iPfam: visualization of protein–protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics*, 21:410–412, 2005. 1.3.1, 2.1, 2.2.4, 3.2.2, 3.3.1

Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V *et al.* Pfam: clans, web tools and services. *Nucleic Acids Res*, 34:D247–251, 2006. 1.4

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ *et al.* The Pfam protein families database. *Nucleic Acids Res*, 36:D281–288, 2008. 1, 1.3, 2.1

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA *et al.* Copy number variation: new insights in genome diversity. *Genome Res*, 16:949–961, 2006. 1.2.3

Fu L. Alteration of protein–protein interactions of congenital cataract crystallin mutants. *Investig Ophthalmology & Vis Sci*, 44:1155–1159, 2003. 3.1

Gandhi TKB, Zhong J, Mathivanan S, Karthick L, Chandrika KN *et al.* Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet*, 38:285–293, 2006. 1, 2.4.3

Gavin AC, Aloy P, Grandi P, Krause R, Boesche M *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440:631–636, 2006. 1.1.2, 2.3.5

Gerhart J and Schachman H. Distinct subunits for the regulation and catalytic activity of aspartate transcarbamylase. *Biochemistry*, 4:1054–1062, 1965. 1.1

Gerhart JC and Pardee AB. The enzymology of control by feedback inhibition. *J Biol Chem*, 237:891–896, 1962. 1.1

Giorgini F and Muchowski PJ. Connecting the dots in Huntington's disease with protein interaction networks. *Genome Biol*, 6:210–211, 2005. 3.1

Glaser F, Rosenberg Y, Kessel A, Pupko T and Ben-Tal N. The ConSurf-HSSP database: the mapping of evolutionary conservation among homologs onto PDB structures. *Proteins*, 58:610–617, 2005. 3.2.3

Grace CRR, Perrin MH, Gulyas J, Digruccio MR, Cantle JP, Rivier JE, Vale WW and Riek R. Structure of the N-terminal domain of a type B1 G protein-coupled receptor in complex with a peptide ligand. *Proc Natl Acad Sci USA*, 104:4858–4863, 2007. 3.2.6

Grigoriev A. On the number of protein–protein interactions in the yeast proteome. *Nucleic Acids Res*, 31:4157–4161, 2003. 2.1

Guindon S and Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*, 52:696–704, 2003. 4

Guldener U, Munsterkotter M, Oesterheld M, Pagel P, Ruepp A, Mewes H and Stumpflen V. MPact: the MIPS protein interaction resource on yeast. *Nucleic Acids Res*, 34:D436–441, 2006. 2.2.1, 3.3.3

Hahn A, Rahnenfuhrer J, Talwar P and Lengauer T. Confirmation of human protein interaction data by human expression data. *BMC Bioinformatics*, 6:1, 2005. 4.3.2, 4.4.1

Hamosh A, Scott AF, Amberger JS, Bocchini CA and McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*, 33:D514–517, 2005. 1.2.3.1, 3.1, 3.2.1

Han JDJ, Dupuy D, Bertin N, Cusick ME and Vidal M. Effect of sampling on topology predictions of protein–protein interaction networks. *Nat Biotechnol*, 23:839–844, 2005. 1.1.4

Hart GT, Ramani AK and Marcotte EM. How complete are current yeast and human protein-interaction networks? *Genome Biol*, 7:120, 2006. 1, 1.1.2

Hermjakob H, Montecchi-Palazzi L, Bader G, Wojcik J, Salwinski L *et al.* The HUPO PSI's molecular interaction format — a community standard for the representation of protein interaction data. *Nat Biotechnol*, 22:177–183, 2004. 2.2.1

Hong B, Senisterra G, Rabeh W, Vedadi M, Leonardi R *et al.* Crystal structures of human pantothenate kinases: Insights into allosteric regulation and mutations linked to a neurodegeneration disorder. *J Biol Chem*, 282:27984–27993, 2007. 3.1

Huber LA. Is proteomics heading in the wrong direction? *Nat Rev Mol Cell Biol*, 4:74–80, 2003. 1.3

Hughes AL and Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, 335:167–170, 1988. 1.2.1

Humphrey W, Dalke A and Schulten K. VMD – Visual Molecular Dynamics. *J Mol Graph*, 14:33–38, 1996. 3.2.9

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW and Lee C. Detection of large-scale variation in the human genome. *Nat Genet*, 36:949–951, 2004. 1.2.3, 4.2.7

Ispolatov I, Yuryev A, Mazo I and Maslov S. Binding properties and evolution of homodimers in protein–protein interaction networks. *Nucleic Acids Res*, 33:3629–3635, 2005. 3.3.3

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M and Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA*, 98:4569–4574, 2001. 1.1.2, 1.1.4

Itzhaki Z, Akiva E, Altuvia Y and Margalit H. Evolutionary conservation of domain–domain interactions. *Genome Biol*, 7:R125, 2006. 2.1

James LC, Keeble AH, Khan Z, Rhodes DA and Trowsdale J. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc Natl Acad Sci USA*, 104:6200–6205, 2007. 3.2.6

Janin J, Rodier F, Chakrabarti P and Bahadur RP. Macromolecular recognition in the Protein Data Bank. *Acta Crystallogr D Biol Crystallogr*, 63:1–8, 2007. 1.1

Jansen R, Greenbaum D and Gerstein M. Relating whole-genome expression data with protein–protein interactions. *Genome Res*, 12:37–46, 2002. 4.1, 4.3.2

Jeong H, Mason SP, Barabasi AL and Oltvai ZN. Lethality and centrality in protein networks. *Nature*, 411:41–42, 2001. 1.1.4

Jhoti H. High-throughput structural proteomics using x-rays. *Trends Biotechnol*, 19:S67–71, 2001. 1.1.1.4

Jimenez-Sanchez G, Childs B and Valle D. Human disease genes. *Nature*, 409:853–855, 2001. 3.3.4.3, 1

Jones S and Thornton JM. Principles of protein-protein interactions. *Proc Natl Acad Sci USA*, 93:13–20, 1996. 1.1

Jothi R, Cherukuri PF, Tasneem A and Przytycka TM. Co-evolutionary analysis of domains in interacting proteins reveals insights into domain–domain interactions mediating protein–protein interactions. *J Mol Biol*, 362:861–875, 2006. 2.1, 2.4.1, 2.4.3

Junttila MR, Saarinen S, Schmidt T, Kast J and Westermarck J. Single-step Strep-tag purification for the isolation and identification of protein complexes from mammalian cells. *Proteomics*, 5:1199–1203, 2005. 1.1.1.1

Kafatos FC, Efstratiadis A, Forget BG and Weissman SM. Molecular evolution of human and rabbit beta-globin mRNAs. *Proc Natl Acad Sci USA*, 74:5618–5622, 1977. 1.2.1

Karev GP, Wolf YI, Rzhetsky AY, Berezovskaya FS and Koonin EV. Birth and death of protein domains: a simple model of evolution explains power law behavior. *BMC Evol Biol*, 2:18, 2002. 1.1.4

Kellis M, Birren BW and Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, 428:617–624, 2004. 4.1

Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H and Phillips DC.

A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181:662–666, 1958. 1.1.1.4

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM and Haussler D. The human genome browser at UCSC. *Genome Res*, 12:996–1006, 2002. 4.4

Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A *et al.* IntAct–open source resource for molecular interaction data. *Nucleic Acids Res*, 35:D561–565, 2007. 1.1.3, 2.2.1, 3.3.3, 4.2.3

Kersey P, Bower L, Morris L, Horne A, Petryszak R *et al.* Integr8 and genome reviews: integrated views of complete genomes and proteomes. *Nucleic Acids Res*, 33:D297–302, 2005. 2.2.3

Kestler HA. ROC with confidence — a Perl program for receiver operator characteristic curves. *Comput Methods Programs Biomed*, 64:133–136, 2001. 3.7

Khanin R and Wit E. How scale–free are biological networks? *J Comput Biol*, 13:810–818, 2006. 1.1.4

Kind J, Vaquerizas JM, Gebhardt P, Gentzel M, Luscombe NM, Bertone P and Akhtar A. Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell*, 133:813–828, 2008. 4.4.2

Kissinger CR, Rejto PA, Pelletier LA, Thomson JA, Showalter RE *et al.* Crystal structure of human ABAD/HSD10 with a bound inhibitor: implications for design of Alzheimer's disease therapeutics. *J Mol Biol*, 342:943–952, 2004. 3.3.6

Klein C, Philippe N, Le Deist F, Fraitag S, Prost C, Durandy A, Fischer A and Griscelli C. Partial albinism with immunodeficiency (Griscelli syndrome). *J Pediatr*, 125:886–895, 1994. 3.3.6.1

Kleywegt GJ. Validation of protein crystal structures. *Acta Crystallogr D Biol Crystallogr*, 56:249–265, 2000. 1.1.1.4

Klotz IM, Langerman NR and Darnall DW. Quaternary structure of proteins. *Annual Review of Biochemistry*, 39:25–62, 1970. 1.1

Koegl M and Uetz P. Improving yeast two-hybrid screening systems. *Brief Funct Genomic Proteomic*, 6:302–312, 2007. 1.1.1.2

Kondrashov FA and Koonin EV. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet*, 20:287–290, 2004. 1.2.3.1, 4.1

Kondrashov FA, Rogozin I, Wolf Y and Koonin E. Selection in the evolution of gene duplications. *Genome Biol*, 3:2, 2002. 4.1

Kouranov A, Xie L, de la Cruz J, Chen L, Westbrook J, Bourne PE and Berman HM. The RCSB PDB information portal for structural genomics. *Nucleic Acids Res*, 34:D302–305, 2006. 1.1.1.4, 2.1, 3.3.1

Krissinel E and Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, 372:774–797, 2007. 1.1.1.4

Krogan NJ, Cagney G, Yu H, Zhong G, Guo X *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*, 440:637–643, 2006. 4.2.3

Krogh A, Brown M, Mian IS, Sjolander K and Haussler D. Hidden Markov models in computational biology. applications to protein modeling. *J Mol Biol*, 235(5):1501–1531, 1994. 1.3

Kuhn K, Baker SC, Chudin E, Lieu MH, Oeser S *et al.* A novel, high-performance random array platform for quantitative gene expression profiling. *Genome Res*, 14:2347–2356, 2004. 4.2.5

Lamolet B, Pulichino AM, Lamonerie T, Gauthier Y, Brue T, Enjalbert A and Drouin J. A pituitary cell-restricted T box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. *Cell*, 104:849–859, 2001. 3.3.6.2

Leandro J, Nascimento C, de Almeida IT and Leandro P. Co-expression of different subunits of human phenylalanine hydroxylase: Evidence of negative interallelic complementation. *Biochim Biophys Acta*, 1762:544–550, 2006. 3.4.3

Lee H, Deng M, Sun F and Chen T. An integrated approach to the prediction of domain-domain interactions. *BMC Bioinformatics*, 7, 2006. 2.1, 2.4.1

Lee JA and Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*, 52:103–121, 2006. 4.1

Lehner B, Crombie C, Tischler J, Fortunato A and Fraser AG. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat Genet*, 38:896–903, 2006. 1.1.1.5

Lemos B, Meiklejohn CD and Hartl DL. Regulatory evolution across the protein interaction network. *Nat Genet*, 36:1059–1060, 2004. 4.4.1

Lewontin RC. The apportionment of human diversity. *Evolutionary Biology*, 6:391–398, 1972. 1.2.3

Littler SJ and Hubbard SJ. Conservation of orientation and sequence in protein domain–domain interactions. *J Mol Biol*, 345:1265–1279, 2005. 2.1

Livingstone F. Anthropological implications of sickle cell gene distribution in west africa. *Am Anthropol*, 60:533–562, 1958. 1.2.3

Logsdon NJ, Jones BC, Allman JC, Izotova L, Schwartz B, Pestka S and Walter MR. The IL-10R2 binding hot spot on IL-22 is located on the N-terminal

helix and is dependent on N-linked glycosylation. *J Mol Biol*, 342:503–514, 2004. 3.2.6

Lu X, Shaw CA, Patel A, Li J, Cooper ML *et al.* Clinical implementation of chromosomal microarray analysis: summary of 2513 postnatal cases. *PLoS ONE*, 2:e327, 2007. 3.1

Luscombe NM, Qian J, Zhang Z, Johnson T and Gerstein M. The dominance of the population by a selected few: power-law behaviour applies to a wide variety of genomic properties. *Genome Biol*, 3:R8, 2002. 1.1.4

Lynch M and Conery JS. The evolutionary fate and consequences of duplicate genes. *Science*, 290:1151–1155, 2000. 4.1

MacBeath G and Schreiber SL. Printing proteins as microarrays for high-throughput function determination. *Science*, 289:1760–1763, 2000. 1.1.1.5

Mani R, St Onge RP, Hartman JLt, Giaever G and Roth FP. Defining genetic interaction. *Proc Natl Acad Sci USA*, 105:3461–3466, 2008. 1.1.1.5

Markham K, Bai Y and Schmitt-Ulms G. Co-immunoprecipitations revisited: an update on experimental concepts and their implementation for sensitive interactome investigations of endogenous proteins. *Anal Bioanal Chem*, 389:461–473, 2007. 1.1.1.5

Maslov S and Sneppen K. Specificity and stability in topology of protein networks. *Science*, 296:910–913, 2002. 2.2.6

Menasche G, Pastural E, Feldmann J, Certain S, Ersoy F *et al.* Mutations in RAB27A cause Griscelli syndrome associated with haemophagocytic syndrome. *Nat Genet*, 25:173–176, 2000. 3.3.6.1

Mendel J. Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn*, 4:3–47, 1865. 1.2

von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S and Bork P. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, 417:399–403, 2002. 1.1.2, 2.1

Mewes HW, Dietmann S, Frishman D, Gregory R, Mannhaupt G *et al.* MIPS: analysis and annotation of genome information in 2007. *Nucleic Acids Res*, 36:D196–201, 2008. 1.1.1.3, 1.1.3

Mintseris J and Weng Z. Structure, function, and evolution of transient and obligate protein–protein interactions. *Proc Natl Acad Sci USA*, 102:10930–10935, 2005. 4.3.4, 4.4.3

Mishra GR, Suresh M, Kumaran K, Kannabiran N, Suresh S *et al.* Human protein reference database–2006 update. *Nucleic Acids Res*, 34:D411–414, 2006. 1.1.3, 2.2.1, 3.3.3

Müller CW and Herrmann BG. Crystallographic structure of the T domain-DNA complex of the brachyury transcription factor. *Nature*, 389:884–888, 1997. 3.3.6.2

Nair SK and Burley SK. X-ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors. *Cell*, 112:193–205, 2003. 3.3.6.3

Neduva V and Russell RB. Linear motifs: evolutionary interaction switches. *FEBS Lett*, 579:3342–3345, 2005. 2.3.5

Ng PC and Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*, 31:3812–3814, 2003. 3.3.2.2

Ng SK, Zhang Z, Tan SH and Lin K. InterDom: a database of putative interacting protein domains for validating predicted protein interactions and complexes. *Nucleic Acids Res*, 31:251–4, 2003. 2.1

Nguyen DQ, Webber C and Ponting CP. Bias of selection on human copy-number variants. *PLoS Genet*, 2:e20, 2006. 4.1, 4.3.3, 4.3.4, 4.4.3

Nye TMW, Berzuini C, Gilks WR, Babu MM and Teichmann SA. Statistical analysis of domains in interacting protein pairs. *Bioinformatics*, 21:993–1001, 2005. 2.1

Ofman R, Ruiter JPN, Feenstra M, Duran M, Poll-The BT *et al.* 2-Methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency is caused by mutations in the HADH2 gene. *Am J Hum Genet*, 72:1300–1307, 2003. 3.3.6

Ofran Y and Rost B. Analysing six types of protein–protein interfaces. *J Mol Biol*, 325:377–387, 2003. 3.3.4.4

Ofran Y and Rost B. Protein–protein interaction hotspots carved into sequences. *PLoS Comput Biol*, 3:e119, 2007. 3.3.2.4

Orr H. A test of Fisher's theory of dominance. *Proc Natl Acad Sci USA*, 88:11413–11415, 1991. 4.1

Ostermeier C and Brunger AT. Structural basis of Rab effector specificity: crystal structure of the small G protein Rab3A complexed with the effector domain of rabphilin-3a. *Cell*, 96:363–374, 1999. 3.3.6.1

Pagel P, Wong P and Frishman D. A domain interaction map based on phylogenetic profiling. *J Mol Biol*, 344:1331–46, 2004. 2.1

Papp B, Pal C and Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. *Nature*, 424:194–197, 2003. 3.4.3, 4.1

Park J, Lappe M and Teichmann SA. Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast. *J Mol Biol*, 307:929–38, 2001. 2.1

Peng K, Obradovic Z and Vucetic S. Exploring bias in the protein data bank using contrast classifiers. *Pac Symp Biocomput*, 435–446, 2004. 2.1

Perutz M, Rossmann M, Cullis A, Muirhead H, Will G and North A. Structure of haemoglobin: A three-dimensional fourier synthesis at 5.5Å resolution, obtained by X-ray analysis. *Nature*, 185:416–422, 1960. 1.1.1.4

Platzer P, Upender MB, Wilson K, Willis J, Lutterbaugh J *et al.* Silence of chromosomal amplifications in colon cancer. *Cancer Res*, 62:1134–1138, 2002. 4.4.2

Powell AJ, Read JA, Banfield MJ, Gunn-Moore F, Yan SD *et al.* Recognition of structurally diverse substrates by type II 3-hydroxyacyl-CoA dehydrogenase (HADH II)/amyloid-beta binding alcohol dehydrogenase (ABAD). *J Mol Biol*, 303:311–327, 2000. 3.3.6

Prince VE and Pickett FB. Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet*, 3:827–837, 2002. 4.1

Pulichino AM, Vallette-Kasic S, Couture C, Gauthier Y, Brue T *et al.* Human and mouse TPIT gene mutations cause early onset pituitary ACTH deficiency. *Genes Dev*, 17:711–716, 2003. 3.3.6.2

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0. 3.2.8, 4.2.12

Randles LG, Lappalainen I, Fowler SB, Moore B, Hamill SJ and Clarke J. Using model proteins to quantify the effects of pathogenic mutations in Ig-like proteins. *J Biol Chem*, 281:24216–24226, 2006. 3.3.2.4

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH *et al.* Global variation in

copy number in the human genome. *Nature*, 444:444–454, 2006. 1.2.3, 3.1, 4.4, 4.3.1, 4.7, 4.3.3

Reguly T, Breitkreutz A, Boucher L, Breitkreutz B, Hon G *et al.* Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *J Biol*, 5:11, 2006. 1.1.1.3

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M and Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol*, 17:1030–1032, 1999. 1.1.1.1

Riley R, Lee C, Sabatti C and Eisenberg D. Inferring protein domain interactions from databases of interacting proteins. *Genome Biol*, 6:R89, 2005. 2.1, 2.4.1

Risch N and Merikangas K. The future of genetic studies of complex human diseases. *Science*, 273:1516–1517, 1996. 1.2.3.1

Ross ED, Minton A and Wickner RB. Prion domains: sequences, structures and interactions. *Nat Cell Biol*, 7:1039–1044, 2005. 3.1

Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature*, 437:1173–1178, 2005. 1.1.2, 4.2.3

Ruepp A, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C *et al.* CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res*, 36:D646–650, 2008. 4.2.2

Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU and Eisenberg D. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res*, 32:D449–451, 2004. 1.1.3, 2.2.1, 3.3.3

Sammut SJ, Finn RD and Bateman A. Pfam 10 years on: 10,000 families and still growing. *Brief Bioinform*, 9:210–219, 2008. 1.3

Sanchez C, Lachaize C, Janody F, Bellon B, Roder L, Euzenat J, Rechenmann F and Jacq B. Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an internet database. *Nucleic Acids Res*, 27:89–94, 1999. 1.1.4

Sankar P and Cho MK. Genetics. toward a new vocabulary of human genetic variation. *Science*, 298:1337–1338, 2002. 1.2.2

Savas S, Tuzmen S and Ozcelik H. Human SNPs resulting in premature stop codons and protein truncation. *Hum Genomics*, 2:274–286, 2006. 3.1

Schuster-Böckler B and Bateman A. Visualizing profile–profile alignment: pairwise HMM logos. *Bioinformatics*, 21:2912–2913, 2005. 1.4

Schuster-Böckler B and Bateman A. An introduction to hidden Markov models. *Curr Protoc Bioinformatics*, A3, 2007a. 1.3

Schuster-Böckler B and Bateman A. Reuse of structural domain-domain interactions in protein networks. *BMC Bioinformatics*, 8:259, 2007b. 1.4

Schuster-Böckler B and Bateman A. Protein interactions in human genetic diseases. *Genome Biol*, 9:9, 2008. 1.4

Schuster-Böckler B, Schultz J and Rahmann S. HMM Logos for visualization of protein families. *BMC Bioinformatics*, 5, 2004. 3.2.5

Sebat J, Lakshmi B, Troge J, Alexander J, Young J *et al.* Large-scale copy number polymorphism in the human genome. *Science*, 305:525–528, 2004. 1.2.3

Seto ML, Lee SJ, Sze RW and Cunningham ML. Another TWIST on Baller-Gerold syndrome. *Am J Med Genet*, 104:323–330, 2001. 3.3.6.3

Shany E, Saada A, Landau D, Shaag A, Hershkovitz E and Elpeleg ON. Lipoamide dehydrogenase deficiency due to a novel mutation in the interface domain. *Biochem Biophys Res Commun*, 262:163–166, 1999. 3.3.4.2

Sharan R and Ideker T. Modeling cellular machinery through biological network comparison. *Nat Biotech*, 24:427–433, 2006. 1

She X, Jiang Z, Clark RA, Liu G, Cheng Z *et al.* Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature*, 431:927–930, 2004. 4.2.8

Shinawi M and Cheung SW. The array CGH and its clinical applications. *Drug Discov Today*, 13:760–770, 2008. 1.2.3

Shy ME, Jani A, Krajewski K, Grandis M, Lewis RA *et al.* Phenotypic clustering in MPZ mutations. *Brain*, 127:371–384, 2004. 3.1

Sidhu SS, Fairbrother WJ and Deshayes K. Exploring protein-protein interactions with phage display. *Chembiochem*, 4:14–25, 2003. 1.1.1.5

Sing T, Sander O, Beerenwinkel N and Lengauer T. ROCR: visualizing classifier performance in R. *Bioinformatics*, 21:3940–3941, 2005. 3.6

Siva N. 1000 Genomes project. *Nat Biotechnol*, 26:256, 2008. 1.2.2

Smith EA and Corn RM. Surface plasmon resonance imaging as a tool to monitor biomolecular interactions in an array based format. *Appl Spectrosc*, 57:320A–332A, 2003. 1.1.1.5

Sopko R, Huang D, Preston N, Chua G, Papp B *et al.* Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell*, 21:319–330, 2006. 4.1

Sprinzak E, Sattath S and Margalit H. How reliable are experimental protein–protein interaction data? *J Mol Biol*, 327:919–923, 2003. 1.1.2

Steinmetz LM, Scharfe C, Deutschbauer AM, Mokranjac D, Herman ZS *et al.* Systematic screen for human disease genes in yeast. *Nat Genet*, 31:400–404, 2002. 4.1

Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH *et al.* A human protein–protein interaction network: a resource for annotating the proteome. *Cell*, 122:957–968, 2005. 4.2.3

Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*, 315:848–853, 2007. 4.2.5, 4.3.1

Strom M, Hume AN, Tarafder AK, Barkagianni E and Seabra MC. A family of Rab27-binding proteins. melanophilin links Rab27a and myosin Va function in melanosome transport. *J Biol Chem*, 277:25423–25430, 2002. 3.3.6.1

Stumpf MPH, Thorne T, de Silva E, Stewart R, An HJ, Lappe M and Wiuf C. Estimating the size of the human interactome. *Proc Natl Acad Sci USA*, 105:6959–6964, 2008. 1.1.2

Stumpf MPH, Wiuf C and May RM. Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proc Natl Acad Sci USA*, 102:4221–4224, 2005. 1.1.4

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA*, 101:6062–6067, 2004. 4.2.5

Svedberg T. Mass and size of protein molecules. *Nature*, 123:871, 1929. 1.1

The International HapMap Consortium. The international HapMap project. *Nature*, 426:789–796, 2003. 1.2.3

Thorn KS and Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics*, 17:284–285, 2001. 1.1, 3.2.6, 3.3.2

Tong AHY, Lesage G, Bader GD, Ding H, Xu H *et al.* Global mapping of the yeast genetic interaction network. *Science*, 303:808–813, 2004. 1.1.1.5

Uetz P, Giot L, Cagney G, Mansfield T, Judson R *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature*, 403:623–627, 2000. 1.1.2, 1.1.4

Veitia RA. Exploring the etiology of haploinsufficiency. *BioEssays*, 24:175–184, 2002. 3.4.3

Veitia RA. Gene dosage balance: deletions, duplications and dominance. *Trends Genet*, 21:33–35, 2005. 4.1

Velankar S, McNeil P, Mittard-Runte V, Suarez A, Barrell D, Apweiler R and Henrick K. E-MSD: an integrated data resource for bioinformatics. *Nucleic Acids Res*, 33:D262–265, 2005. 1.3.1

Vitkup D, Sander C and Church GM. The amino-acid mutational spectrum of human genetic disease. *Genome Biol*, 4:R72, 2003. 3.3.4.4

Walsh STR and Kossiakoff AA. Crystal structure and site 1 binding energetics of human placental lactogen. *J Mol Biol*, 358:773–784, 2006. 3.2.6

Wang Z and Moult J. SNPs, protein structure, and disease. *Hum Mutat*, 17:263–270, 2001. 3.1

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447:661–678, 2007. 1.2.3.1

Williams AD, Shivaprasad S and Wetzel R. Alanine scanning mutagenesis of Aβ(1–40) amyloid fibril stability. *J Mol Biol*, 357:1283–1294, 2006. 3.2.6

Wolf YI, Rogozin IB and Koonin EV. Coelomata and not ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res*, 14:29–36, 2004. 1.3

Wong JMS, Ionescu D and Ingles CJ. Interaction between BRCA2 and replication protein A is compromised by a cancer-predisposing mutation in BRCA2. *Oncogene*, 22:28–33, 2003. 3.3.4.2

Wright S. Physiological and evolutionary theories of dominance. *Am Nat*, 68:24–53, 1934. 1.2.3.1, 4.1

Wu C, Apweiler R, Bairoch A, Natale D, Barker W *et al.* The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res*, 34:D187–191, 2006. 1.3, 3.1, 3.2.1

Yang J, Lusk R and Li WH. Organismal complexity, protein complexity, and gene duplicability. *Proc Natl Acad Sci USA*, 100:15661–15665, 2003. 4.1, 4.4.3

Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*, 13:555–556, 1997. 4.2.11

Yue P, Li Z and Moult J. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol*, 353:459–473, 2005. 3.1

Zdobnov EM, Lopez R, Apweiler R and Etzold T. The EBI SRS server — Recent developments. *Bioinformatics*, 18:368–373, 2002. 3.2.1

Zhu H, Domingues FS, Sommer I and Lengauer T. NOXclass: prediction of protein-protein interaction types. *BMC Bioinformatics*, 7:27, 2006. 2.2.5, 3.2.3

Zuckerkandl E and Pauling L. Molecular disease, evolution, and genetic heterogeneity. *Horiz Biochem*, 189–225, 1962. 1.2.1