REFERENCES

[1] Hurles, M. E., Dermitzakis, E. T. & Tyler-Smith, C. The functional impact of structural variation in humans. *Trends Genet* **24**, 238–45 (2008).

[2] Lupski, J. R. & Stankiewicz, P. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet* **1**, e49 (2005).

[3] Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–53 (2007).

[4] Millar, J. K. *et al.* Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet* **9**, 1415–23 (2000).

[5] Nakata, K. *et al.* DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. *Proc Natl Acad Sci U S A* **106**, 15873–8 (2009).

[6] Millar, J. K. *et al.* DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science* **310**, 1187–91 (2005).

[7] Shen, S. *et al.* Schizophrenia-related neural and behavioral phenotypes in transgenic mice expressing truncated Disc1. *J Neurosci* **28**, 10893–904 (2008).

[8] Kirov, G. *et al.* De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol Psychiatry* (2011).

[9] Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–72 (2010).

[10] Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nat Genet* **43**, 838–46 (2011).

[11] International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–41 (2008).

[12] Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–12 (2010).

[13] Wellcome Trust Case Control Consortium *et al.* Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713–20 (2010).

[14] Dang, V. T., Kassahn, K. S., Marcos, A. E. & Ragan, M. A. Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. *Eur J Hum Genet* **16**, 1350–7 (2008).

[15] Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nat Genet* **39**, 1256–60 (2007).

[16] Johnson, M. E. *et al.* Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**, 514–9 (2001).

[17] Zhang, J., Zhang, Y.-p. & Rosenberg, H. F. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet* **30**, 411–5 (2002).

[18] Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* **40**, 1253–60 (2008).

[19] Affymetrix. Affymetrix genome-wide human SNP array 6.0 data sheet. `http://media.affymetrix.com/support/technical/datasheets/genomewide_snp6_datasheet.pdf` (2009).

[20] Affymetrix. Manuals of Affymtrix Power Tools. `http://media.affymetrix.com/support/developer/powertools/changelog/index.html` (2009).

[21] Price, T. S. *et al.* SW-ARRAY: a dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data. *Nucleic Acids Res* **33**, 3455–64 (2005).

[22] Veltman, J. A. *et al.* High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. *Am J Hum Genet* **70**, 1269–76 (2002).

[23] Wang, P., Kim, Y., Pollack, J., Narasimhan, B. & Tibshirani, R. A method for calling gains and losses in array CGH data. *Biostatistics* **6**, 45–58 (2005).

[24] Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–72 (2004).

[25] Jong, K., Marchiori, E., Meijer, G., Vaart, A. V. D. & Ylstra, B. Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics* **20**, 3636–7 (2004).

[26] Hupé, P., Stransky, N., Thiery, J.-P., Radvanyi, F. & Barillot, E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* **20**, 3413–22 (2004).

[27] Pique-Regi, R. *et al.* Sparse representation and Bayesian detection of genome copy number alterations from microarray data. *Bioinformatics* **24**, 309–18 (2008).

[28] Fridlyand, J., Snijders, A., Pinkel, D., Albertson, D. & Jain, A. Hidden Markov models approach to the analysis of array CGH data. *Journal of Multivariate Analysis* **90**, 132–153 (2004).

[29] Marioni, J. C., Thorne, N. P. & Tavaré, S. BioHMM: a heterogeneous hidden Markov model for segmenting array CGH data. *Bioinformatics* **22**, 1144–6 (2006).

[30] Guha, S., Li, Y. & Neuberg, D. Bayesian hidden Markov Modeling of array CGH data. *Journal of the American Statistical Association* **103**, 485–497 (2008).

[31] Shah, S. P. *et al.* Integrating copy number polymorphisms into array CGH analysis using a robust HMM. *Bioinformatics* **22**, e431–9 (2006).

[32] Colella, S. *et al.* QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res* **35**, 2013–25 (2007).

[33] Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* **17**, 1665–74 (2007).

[34] McCarroll, S. A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* **40**, 1166–74 (2008).

[35] Marioni, J. C. *et al.* Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol* **8**, R228 (2007).

[36] The HDF5 group. Hierarchical data format version 5. `http://www.hdfgroup.org/HDF5` (2009).

[37] Alted, F., Vilata, I. *et al.* PyTables: hierarchical datasets in Python. `http://www.pytables.org` (2002).

[38] Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–54 (2006).

[39] Weedon, M. N. *et al.* Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* **40**, 575–83 (2008).

[40] Zöllner, S. CopyMap: localization and calling of copy number variation by joint analysis of hybridization data from multiple individuals. *Bioinformatics* **26**, 2776–7 (2010).

[41] Williams, N. M. *et al.* Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet* **376**, 1401–8 (2010).

[42] Whitby, H. *et al.* Benign copy number changes in clinical cytogenetic diagnostics by array CGH. *Cytogenet Genome Res* **123**, 94–101 (2008).

[43] Bell, C. G., Walley, A. J. & Froguel, P. The genetics of human obesity. *Nat Rev Genet* **6**, 221–34 (2005).

[44] Rennie, K. L. & Jebb, S. A. Prevalence of obesity in Great Britain. *Obes Rev* **6**, 11–2 (2005).

[45] Allison, D. B. *et al.* The heritability of body mass index among an international sample of monozygotic twins reared apart. *Int J Obes Relat Metab Disord* **20**, 501–6 (1996).

[46] Maes, H. H., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* **27**, 325–51 (1997).

[47] Stunkard, A. J., Harris, J. R., Pedersen, N. L. & McClearn, G. E. The body-mass index of twins who have been reared apart. *N Engl J Med* **322**, 1483–7 (1990).

[48] Chen, H. *et al.* Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in db/db mice. *Cell* **84**, 491–5 (1996).

[49] Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131–41 (1997).

[50] Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–32 (1994).

[51] Cummings, D. E. & Schwartz, M. W. Genetics and pathophysiology of human obesity. *Annu Rev Med* **54**, 453–71 (2003).

[52] Walley, A. J., Asher, J. E. & Froguel, P. The genetic contribution to non-syndromic human obesity. *Nat Rev Genet* **10**, 431–42 (2009).

[53] Clément, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398–401 (1998).

[54] Krude, H. *et al.* Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat Genet* **19**, 155–7 (1998).

[55] Montague, C. T. *et al.* Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* **387**, 903–8 (1997).

[56] Farooqi, I. S. & O'Rahilly, S. Monogenic obesity in humans. *Annu Rev Med* **56**, 443–58 (2005).

[57] Farooqi, I. S. *et al.* Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. *J Clin Invest* **110**, 1093–103 (2002).

[58] Goldstone, A. P. Prader-Willi syndrome: advances in genetics, pathophysiology and treatment. *Trends Endocrinol Metab* **15**, 12–20 (2004).

[59] O'Rahilly, S. & Farooqi, I. S. Human obesity: a heritable neurobehavioral disorder that is highly sensitive to environmental conditions. *Diabetes* **57**, 2905–10 (2008).

[60] Farooqi, I. S. *et al.* Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med* **348**, 1085–95 (2003).

[61] Comuzzie, A. G. *et al.* A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nat Genet* **15**, 273–6 (1997).

[62] Mitchell, B. D. *et al.* A quantitative trait locus influencing BMI maps to the region of the beta-3 adrenergic receptor. *Diabetes* **48**, 1863–7 (1999).

[63] Kissebah, A. H. *et al.* Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc Natl Acad Sci U S A* **97**, 14478–83 (2000).

[64] Walder, K., Hanson, R. L., Kobes, S., Knowler, W. C. & Ravussin, E. An autosomal genomic scan for loci linked to plasma leptin concentration in Pima Indians. *Int J Obes Relat Metab Disord* **24**, 559–65 (2000).

[65] Dong, C. *et al.* Interacting genetic loci on chromosomes 20 and 10 influence extreme human obesity. *Am J Hum Genet* **72**, 115–24 (2003).

[66] Hunt, S. C. *et al.* Linkage of body mass index to chromosome 20 in Utah pedigrees. *Hum Genet* **109**, 279–85 (2001).

[67] Frayling, T. M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–94 (2007).

[68] Dina, C. *et al.* Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* **39**, 724–6 (2007).

[69] Hinney, A. *et al.* Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. *PLoS One* **2**, e1361 (2007).

[70] Scuteri, A. *et al.* Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet* **3**, e115 (2007).

[71] Loos, R. J. F. *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* **40**, 768–75 (2008).

[72] Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* **41**, 25–34 (2009).

[73] Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* **42**, 937–48 (2010).

[74] McDermid, H. E. & Morrow, B. E. Genomic disorders on 22q11. *Am J Hum Genet* **70**, 1077–88 (2002).

[75] Boerkoel, C. F., Inoue, K., Reiter, L. T., Warner, L. E. & Lupski, J. R. Molecular mechanisms for CMT1A duplication and HNPP deletion. *Ann N Y Acad Sci* **883**, 22–35 (1999).

[76] Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E. & Pritchard, J. K. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* **38**, 75–81 (2006).

[77] Iafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nat Genet* **36**, 949–51 (2004).

[78] Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–8 (2004).

[79] Sharp, A. J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* **77**, 78–88 (2005).

[80] Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nat Genet* **37**, 727–32 (2005).

[81] McCarroll, S. A. *et al.* Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet* **40**, 1107–12 (2008).

[82] Gonzalez, E. *et al.* The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–40 (2005).

[83] McKinney, C. *et al.* Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann Rheum Dis* **67**, 409–13 (2008).

[84] McCarthy, S. E. *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* **41**, 1223–7 (2009).

[85] Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–6 (2008).

[86] Walsh, T. *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* **320**, 539–43 (2008).

[87] Glessner, J. T. *et al.* Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **459**, 569–73 (2009).

[88] Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* **82**, 477–88 (2008).

[89] Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316**, 445–9 (2007).

[90] Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* **358**, 667–75 (2008).

[91] Girirajan, S. *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* **42**, 203–9 (2010).

[92] Bochukova, E. G. *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* **463**, 666–70 (2010).

[93] Barnes, C. *et al.* A robust statistical method for case-control association testing with copy number variation. *Nat Genet* **40**, 1245–52 (2008).

[94] Raychaudhuri, S. *et al.* Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS Genet* **6** (2010).

[95] Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–50 (2005).

[96] Rankinen, T. *et al.* The human obesity gene map: the 2005 update. *Obesity (Silver Spring)* **14**, 529–644 (2006).

[97] Walters, R. G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463**, 671–5 (2010).

[98] Marchini, J., Cardon, L. R., Phillips, M. S. & Donnelly, P. The effects of human population structure on large genetic association studies. *Nat Genet* **36**, 512–7 (2004).

[99] Iype, T. *et al.* The transcriptional repressor Nkx6.1 also functions as a deoxyribonucleic acid context-dependent transcriptional activator during pancreatic beta-cell differentiation: evidence for feedback activation of the nkx6.1 gene by Nkx6.1. *Mol Endocrinol* **18**, 1363–75 (2004).

[100] Schisler, J. C. *et al.* Stimulation of human and rat islet beta-cell proliferation with retention of function by the homeodomain transcription factor Nkx6.1. *Mol Cell Biol* **28**, 3465–76 (2008).

[101] Thomas, M. A., Preece, D. M. & Bentel, J. M. Androgen regulation of the prostatic tumour suppressor NKX3.1 is mediated by its 3′ untranslated region. *Biochem J* **425**, 575–83 (2010).

[102] Jacquemont, S. *et al.* Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* **478**, 97–102 (2011).

[103] Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–9 (2006).

[104] Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* **42**, 949–60 (2010).

[105] Fisher, R. A. *The genetical theory of natural selection* (The Clarendon press, Oxford, 1930).

[106] Wright, S. *Evolution and the genetics of populations: a treatise* (University of Chicago Press, Chicago, 1968). URL `http://www.loc.gov/catdir/description/uchi051/67025533.html`.

[107] Kacser, H. & Burns, J. A. The molecular basis of dominance. *Genetics* **97**, 639–66 (1981).

[108] Amberger, J., Bocchini, C. A., Scott, A. F. & Hamosh, A. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res* **37**, D793–6 (2009).

[109] Wilkie, A. O. The molecular basis of genetic dominance. *J Med Genet* **31**, 89–98 (1994).

[110] Veitia, R. A. Exploring the etiology of haploinsufficiency. *Bioessays* **24**, 175–84 (2002).

[111] Willing, M. C. *et al.* Osteogenesis imperfecta type I: molecular heterogeneity for COL1A1 null alleles of type I collagen. *Am J Hum Genet* **55**, 638–47 (1994).

[112] Ebert, B. L. *et al.* Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335–9 (2008).

[113] Devriendt, K. *et al.* Haploinsufficiency of the HOXA gene cluster, in a patient with hand-foot-genital syndrome, velopharyngeal insufficiency, and persistent patent Ductus botalli. *Am J Hum Genet* **65**, 249–51 (1999).

[114] Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol* **5**, 89–99 (2004).

[115] Vissers, L. E. L. M. *et al.* Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* **36**, 955–7 (2004).

[116] Ng, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272–6 (2009).

[117] Ng, P. C. *et al.* Genetic variation in an individual human exome. *PLoS Genet* **4**, e1000160 (2008).

[118] Xue, Y. *et al.* Adaptive evolution of UGT2B17 copy-number variation. *Am J Hum Genet* **83**, 337–46 (2008).

[119] Blekhman, R. *et al.* Natural selection on genes that underlie human disease susceptibility. *Curr Biol* **18**, 883–9 (2008).

[120] Kondrashov, F. A. & Koonin, E. V. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet* **20**, 287–90 (2004).

[121] Nguyen, D.-Q., Webber, C. & Ponting, C. P. Bias of selection on human copy-number variants. *PLoS Genet* **2**, e20 (2006).

[122] Thomas, P. D. & Kejariwal, A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci U S A* **101**, 15398–403 (2004).

[123] Tavtigian, S. V., Byrnes, G. B., Goldgar, D. E. & Thomas, A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat* **29**, 1342–54 (2008).

[124] Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073–81 (2009).

[125] Yue, P., Melamud, E. & Moult, J. SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics* **7**, 166 (2006).

[126] Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* **30**, 3894–900 (2002).

[127] Ng, P. C. & Henikoff, S. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* **7**, 61–80 (2006).

[128] Hehir-Kwa, J. Y. *et al.* Accurate distinction of pathogenic from benign CNVs in mental retardation. *PLoS Comput Biol* **6**, e1000752 (2010).

[129] Firth, H. V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* **84**, 524–33 (2009).

[130] Stenson, P. D. *et al.* The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum Genomics* **4**, 69–72 (2009).

[131] Itsara, A. *et al.* Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* **84**, 148–61 (2009).

[132] Zhang, J., Feuk, L., Duggan, G. E., Khaja, R. & Scherer, S. W. Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome. *Cytogenet Genome Res* **115**, 205–14 (2006).

[133] Breckpot, J. *et al.* Challenges of interpreting copy number variation in syndromic and non-syndromic congenital heart defects. *Cytogenet Genome Res* **135**, 251–9 (2011).

[134] Buysse, K. *et al.* Challenges for CNV interpretation in clinical molecular karyotyping: lessons learned from a 1001 sample experience. *Eur J Med Genet* **52**, 398–403 (2009).

[135] Goldgar, D. E. *et al.* Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet* **75**, 535–44 (2004).

[136] Hubbard, T. J. P. *et al.* Ensembl 2009. *Nucleic Acids Res* **37**, D690–7 (2009).

[137] Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15**, 901–13 (2005).

[138] Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **101**, 6062–7 (2004).

[139] Assou, S. *et al.* A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells* **25**, 961–73 (2007).

[140] Smith, C. M. *et al.* The mouse Gene Expression Database (GXD): 2007 update. *Nucleic Acids Res* **35**, D618–23 (2007).

[141] Brown, K. R. & Jurisica, I. Online predicted human interaction database. *Bioinformatics* **21**, 2076–82 (2005).

[142] Chatr-aryamontri, A. *et al.* MINT: the Molecular INTeraction database. *Nucleic Acids Res* **35**, D572–4 (2007).

[143] Keshava Prasad, T. S. *et al.* Human Protein Reference Database–2009 update. *Nucleic Acids Res* **37**, D767–72 (2009).

[144] Rual, J.-F. *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437**, 1173–8 (2005).

[145] Vastrik, I. *et al.* Reactome: a knowledge base of biologic pathways and processes. *Genome Biol* **8**, R39 (2007).

[146] Lee, I., Li, Z. & Marcotte, E. M. An improved, bias-reduced probabilistic functional gene network of baker's yeast, Saccharomyces cerevisiae. *PLoS One* **2**, e988 (2007).

[147] Lee, I. *et al.* A single gene network accurately predicts phenotypic effects of gene perturbation in Caenorhabditis elegans. *Nat Genet* **40**, 181–8 (2008).

[148] Van Dongen, S. Graph clustering via a discrete uncoupling process. *Siam Journal On Matrix Analysis and Applications* **30**, 121–141 (2008).

[149] Forbes, S. *et al.* COSMIC 2005. *Br J Cancer* **94**, 318–22 (2006).

[150] Deutschbauer, A. M. *et al.* Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**, 1915–25 (2005).

[151] Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters* **27**, 861–874 (2006).

[152] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. & Nielsen, H. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* **16**, 412–424 (2000).

[153] Van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Software* **45**, 1–67 (2011).

[154] Seal, R. L., Gordon, S. M., Lush, M. J., Wright, M. W. & Bruford, E. A. genenames.org: the HGNC resources in 2011. *Nucleic Acids Res* **39**, D514–9 (2011).

[155] Tsuruoka, Y. *et al.* Developing a robust part-of-speech tagger for biomedical text. *Advances In Informatics, Proceedings* **3746**, 382–392 (2005).

[156] Seidman, J. G. & Seidman, C. Transcription factor haploinsufficiency: when half a loaf is not enough. *J Clin Invest* **109**, 451–5 (2002).

[157] Jensen, L. J. *et al.* STRING 8–a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**, D412–6 (2009).

[158] Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153–7 (2005).

[159] Lohmueller, K. E. *et al.* Proportionally more deleterious genetic variation in European than in African populations. *Nature* **451**, 994–7 (2008).

[160] Boyko, A. R. *et al.* Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* **4**, e1000083 (2008).

[161] Pickering, D. L. *et al.* Array-based comparative genomic hybridization analysis of 1176 consecutive clinical genetics investigations. *Genet Med* **10**, 262–6 (2008).

[162] Hrabé de Angelis, M. H. *et al.* Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet* **25**, 444–7 (2000).

[163] D'Adamo, P. *et al.* Does epidermal thickening explain GJB2 high carrier frequency and heterozygote advantage? *Eur J Hum Genet* **17**, 284–6 (2009).

[164] Guastalla, P. *et al.* Detection of epidermal thickening in GJB2 carriers with epidermal US. *Radiology* **251**, 280–6 (2009).

[165] Hurst, L. D. & Randerson, J. P. Dosage, deletions and dominance: simple models of the evolution of gene expression. *J Theor Biol* **205**, 641–7 (2000).

[166] Orr, H. A. A test of Fisher's theory of dominance. *Proc Natl Acad Sci U S A* **88**, 11413–5 (1991).

[167] Magrane, M. & Uniprot Consortium. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* **2011**, bar009 (2011).

[168] López-Bigas, N. & Ouzounis, C. A. Genome-wide identification of genes likely to be involved in human genetic disease. *Nucleic Acids Res* **32**, 3108–14 (2004).

[169] Smith, N. G. C. & Eyre-Walker, A. Human disease genes: patterns and predictions. *Gene* **318**, 169–75 (2003).

[170] Bruneau, B. G. *et al.* A murine model of holt-oram syndrome defines roles of the t-box transcription factor tbx5 in cardiogenesis and disease. *Cell* **106**, 709–21 (2001).

[171] Xu, B. *et al.* Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* **40**, 880–5 (2008).

[172] Vermeesch, J. R., Balikova, I., Schrander-Stumpel, C., Fryns, J.-P. & Devriendt, K. The causality of de novo copy number variants is overestimated. *Eur J Hum Genet* **19**, 1112–3 (2011).