Li et al: Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple *De Novo* assemblies

Supplemental references

1 Wang LY, Wang AG, Wang LX, Li K, Yang GS, He RG, Qian L, Xu NY, Huang RH, Peng ZZ, et al. 2011. Animal genetic resources in China: pigs. (ed. China National Commission of Animal Genetic Resources) , pp. 18-470. China Agricultural Press, Beijing.

2 Chen, K., Baxter, T., Muir, W. M., Groenen, M. A. & Schook, L. B. Genetic resources, genome mapping and evolutionary genomics of the pig (*Sus scrofa*). *Int. J. Biol. Sci.* **3**, 153-165 (2007).

3 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

4 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303 (2010).

5 Saeed, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374-378 (2003).

6 Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595 (1989).

7 Maillet, N., Lemaitre, C., Chikhi, R., Lavenier, D. & Peterlongo, P. Compareads: comparing huge metagenomic experiments. *BMC Bioinformatics* **13 Suppl 19**, S10 (2012).

8 Sennoune, S. R. & Martinez-Zaguilan, R. Vacuolar $H^{(+)}$-ATPase signaling pathway in cancer. *Curr. Protein Pept. Sc.* **13**, 152-163 (2012).

9 Yang, D. Q. *et al*. V-ATPase subunit *ATP6AP1* (Ac45) regulates osteoclast differentiation, extracellular acidification, lysosomal trafficking, and protease exocytosis in osteoclast-mediated bone resorption. *J. Bone Miner. Res.* **27**, 1695-1707 (2012).

10 Arif, S. *et al.* Blueprint of quartz crystal microbalance biosensor for early detection of breast cancer through salivary autoantibodies against *ATP6AP1*. *Biosens. Bioelectron.* **65C**, 62-70 (2014).

11 Bridle, K. R. *et al.* Disrupted hepcidin regulation in HFE-associated haemochromatosis

and the liver as a regulator of body iron homoeostasis. *Lancet* **361**, 669-673 (2003).

12 Vantyghem, M. C. *et al.* Phenotype and *HFE* genotype in a population with abnormal iron markers recruited from an Endocrinology Department. *Eur. J. Endocrinol.* **154**, 835-841 (2006).

13 Tussing-Humphreys, L. M. *et al.* Elevated systemic hepcidin and iron depletion in obese premenopausal females. *Obesity* **18**, 1449-1456 (2010).

14 Lauro, M., Grimes, C. & Bahnson, B. Molecular characterization and structural Determination of *Nod2*, an Innate Immune Receptor. *FASEB J.* **29**, 890-897 (2015).

15 Kobayashi, K. S. *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307**, 731-734 (2005).

16 Ogura, Y. *et al.* A frameshift mutation in *NOD2* associated with susceptibility to Crohn's disease. *Nature* **411**, 603-606 (2001).

17 Dugan, J. *et al.* Blau syndrome-associated *Nod2* mutation alters expression of full-length *NOD2* and limits responses to muramyl dipeptide in knock-in mice. *J. Immunol.* **194**, 349-357 (2015).

18 Zhao, H. *et al.* Association between *NOD2* single nucleotide polymorphisms and Grade III-IV acute graft-versus-host disease: A meta-analysis. *Hematology* **20**, 254-262 (2015).

19 Mizuguchi, Y. *et al.* Breast tumor kinase/protein tyrosine kinase 6 (Brk/PTK6) activity in normal and neoplastic biliary epithelia. *J. Hepatol.* **63**, 399-407 (2015).

20 Regan Anderson, T. M. *et al.* Breast tumor kinase (Brk/PTK6) is a mediator of hypoxia-associated breast cancer progression. *Cancer Res.* **73**, 5810 (2013).

21 Vogan, K. *RPGRIP1L*, *FTO* and obesity. *Nat. Genet.* **46**, 532-532 (2014).

22 Ashkinadze, E., Rosen, T., Brooks, S. S., Katsanis, N. & Davis, E. E. Combining fetal sonography with genetic and allele pathogenicity studies to secure a neonatal diagnosis of Bardet-Biedl syndrome. *Clin. Genet.* **83**, 553-559 (2013).

23 Stratigopoulos, G. *et al.* Hypomorphism for *RPGRIP1L*, a ciliary gene vicinal to the *FTO* locus, causes increased adiposity in mice. *Cell Metab.* **19**, 767-779 (2014).

24 Wang, X. *et al.* Siva1 inhibits p53 function by acting as an ARF E3 ubiquitin ligase. *Nature Commun.* **4**, 1551 (2013).

25 Li, N. *et al.* Siva1 suppresses epithelial-mesenchymal transition and metastasis of tumor cells by inhibiting stathmin and stabilizing microtubules. *Proc. Natl. Acad. Sci. USA* **108**, 12851-12856 (2011).

26 Xu, Y., Malhotra, A., Ren, M. & Schlame, M. The enzymatic function of tafazzin. *J. Biol. Chem.* **281** (2006).

27 Schlame, M. *et al.* The physical state of lipid substrates provides transacylation specificity for tafazzin. *Nat. Chem. Biol.* **8**, 862-869 (2012).

28 Kirwin, S. M., Manolakos, A., Barnett, S. S. & Gonzalez, I. L. Tafazzin splice variants and mutations in Barth syndrome. *Mol. Genet. Metab.* **111**, 26-32 (2014).

29 Buroker, N. E. Regulatory SNPs and transcriptional factor binding sites in *ADRBK1*, *AKT3*, *ATF3*, *DIO2*, *TBXA2R* and *VEGFA*. *Transcription* **5**, e964559 (2014).

30 Jones, M. L. *et al.* Functional variations in genes encoding platelet G-protein coupled receptors in unselected and platelet function disorder populations. *Blood* **122**, 3511-3511 (2013).

31 Kavalar, M. S. *et al.* Association of *ORMDL3*, *STAT6* and *TBXA2R* gene polymorphisms with asthma. *Int. J. Immunogenet.* **39**, 20-25 (2012).

32 Reed, J. H. & Gordon, T. P. Autoimmunity: Ro60-associated RNA takes its toll on disease pathogenesis. *Nat. Rev. Rheumatol.* doi:10.1038/nrrheum.2015.,148 (2015).

33 Maier-Moore, J. S. *et al.* Passive transfer of antibodies to the linear epitope 60 kD Ro 273-289 induces features of Sjogren's syndrome in naive mice. *Clin. Exp. Immunol.* **180**, 19-27 (2015).

34 Hung, T. *et al.* The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science* **350**, 455-459 (2015).

35 Sievers, F. & Higgins, D. G. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods in Mol. Biol.* **1079**, 105-116 (2014).

36 Ai, H. *et al.* Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* **47**, 217-225 (2015).

37 Moon, S. *et al.* A genome-wide scan for signatures of directional selection in domesticated pigs. *BMC Genomics* **16**, 130 (2015).

38 Choi, J. W. *et al.* Whole-genome resequencing analyses of five pig breeds, including Korean wild and native, and three European origin breeds. *DNA Res.* **22**, 259-267 (2015).

39 Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564-567 (2010).

40 Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993 (2008).

41 Groenen, M. A. *et al.* Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393-398 (2012).

42 Li, M. *et al.* Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* **45**, 1431-1438 (2013).

43 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

44 Parra, G., Bradnam, K., Ning, Z., Keane, T. & Korf, I. Assessing the gene space in draft genomes. *Nucleic Acids Res.* **37**, 289-297 (2009).

45 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).

46 Annilo, T., Kepp, K. & Laan, M. Natural antisense transcript of natriuretic peptide precursor A (NPPA): structural organization and modulation of *NPPA* expression. *BMC Mol. Biol.* **10**, 81 (2009).

47 Newton-Cheh, C. *et al.* Association of common variants in *NPPA* and *NPPB* with circulating natriuretic peptides and blood pressure. *Nat. Genet.* **41**, 348-353 (2009).

48 Lynch, A. I. *et al.* Pharmacogenetic association of the *NPPA* T2238C genetic variant with cardiovascular disease outcomes in patients with hypertension. *JAMA* **299**, 296-307 (2008).

49 Song, W., Wang, H. & Wu, Q. Atrial natriuretic peptide in cardiovascular biology and disease (NPPA). *Gene* **569**, 1-6 (2015).

50 Shi, Y.-C. *et al.* Arcuate NPY controls sympathetic output and BAT function via a relay of tyrosine hydroxylase neurons in the PVN. *Cell Metab.* **17**, 236-248 (2013).

51 Loktev, A. V. & Jackson, P. K. Neuropeptide Y family receptors traffic via the Bardet-Biedl syndrome pathway to signal in neuronal primary cilia. *Cell Rep.* **5**, 1316-1329 (2013).

52 Krashes, M. J., Shah, B. P., Koda, S. & Lowell, B. B. Rapid versus delayed stimulation of feeding by the endogenously released AgRP neuron mediators *GABA*, *NPY*, and *AgRP*. *Cell Metab.* **18**, 588-595 (2013).

53 Mashiko, S. *et al.* Synergistic interaction between neuropeptide Y 1 and Y 5 receptor pathways in regulation of energy homeostasis. *Eur. J. Pharmacol.* **615**, 113-117 (2009).

54 Kanakadurga, S. *et al.* Neuropeptide Y is produced by adipose tissue macrophages and regulates obesity-induced inflammation. *PLoS ONE* **8 (3)**: e57929 (2013).

55 Lin, X. *et al.* Neuropeptide Y genotype, central obesity, and abdominal fat distribution: the POUNDS LOST trial. *Am. J. Clin. Nutr.* **102**, 514-519 (2015).

56 Fred, R. G., Boddeti, S. K., Lundberg, M. & Welsh, N. Imatinib mesylate stimulates low-density lipoprotein receptor-related protein 1-mediated ERK phosphorylation in insulin-producing cells. *Clin. Sci.* **128**, 17-28 (2015).

57 Kang, B. Y., Bae, H. G., Jhin, H. K., Lee, K. S. & Lee, K. O. Association between I/D polymorphism of human *LRPAP1* gene and body mass index in Korean general population. *Toxicol. Res.* **19**, 205-210 (2003).

58 Lockyer, P., Mao, H. & Pi, X. Low-density lipoprotein receptor-related protein in the endothelium regulates metabolic responses. *Arterioscl. Throm. Vas.* **35**, A695-A695 (2015).

59 Liu, R. *et al.* Identification of loci and genes for growth related traits from a genome-wide association study in a slow-× fast-growing broiler chicken cross. *Genes Genom.* **37**, 829-836 (2015).

60 Cheepala, S. B. *et al.* The *ABCC4* membrane transporter modulates platelet aggregation. *Blood* **126**, 2307-2319 (2015).

61 Huynh, T., Norris, M. D., Haber, M. & Henderson, M. J. ABCC4/MRP4: a MYCN-regulated transporter and potential therapeutic target in neuroblastoma. *Front. Oncol.* **2:**178 (2012).

62 Chen, Y. *et al.* Global gene expression profiling reveals genes expressed differentially in cattle with high and low residual feed intake. *Anim. Genet.* **42**, 475-490 (2011).

63 More, V. R. & Slitt, A. L. Alteration of hepatic but not renal transporter expression in diet-induced obese mice. *Drug Metab. Dispos.* **39**, 992-999 (2011).

64 Eto, M. *et al.* Interaction between gene polymorphism of *α* adducin (*ADD1*) and alcohol habit on the risk of essential hypertension in a Japanese population. *Circulation* **132**, A12535-A12535 (2015).

65 Zhang, L. N. *et al.* Lower *ADD1* gene promoter DNA methylation increases the risk of essential hypertension. *PloS ONE* **8**, e63455 (2013).

66 Han, L. *et al.* The interactions between alcohol consumption and DNA methylation of the *ADD1* gene promoter modulate essential hypertension susceptibility in a population-based, case–control study. *Hypertens. Res.* **38**, 284-290 (2015).

67 Vanhaesebroeck, B., Stephens, L. & Hawkins, P. PI3K signalling: the path to discovery and understanding. *Nat. Rev. Mol. Cell Bio.* **13**, 195-203 (2012).

68 Rasmussen, B. A. *et al.* Jejunal leptin-PI3K signaling lowers glucose production. *Cell Metab.* **19**, 155-161 (2014).

69 Shi, X. *et al.* Central GLP-2 enhances hepatic insulin sensitivity via activating PI3K signaling in POMC neurons. *Cell Metab.* **18**, 86-98 (2013).

70 Selli, M., Wraith, D. & Newby, A. HSP60 as a novel autoantigen in obesity. *Atherosclerosis* **244**, e10-e11 (2016).

71 Molina, M. N., Ferder, L. & Manucha, W. Emerging role of nitric oxide and heat shock proteins in insulin resistance. *Current hypertens. Rep.* **18**, 1-13 (2016).

72 Marker, T. *et al.* Heat shock protein 60 as a mediator of adipose tissue inflammation and insulin resistance. *Diabetes* **61**, 615-625 (2012).

73 Quan, B., Seo, H. S., Blobel, G. & Ren, Y. Vesiculoviral matrix (M) protein occupies nucleic acid binding site at nucleoporin pair (Rae1 * Nup98). *Proc. Natl. Acad. Sci. USA* **111**, 9127-9132 (2014).

74 Ren, Y., Seo, H.-S., Blobel, G. & Hoelz, A. Structural and functional analysis of the interaction between the nucleoporin *Nup98* and the mRNA export factor *Rae1*. *Proc. Natl. Acad. Sci. USA* **107**, 10406-10411 (2010).

75 Joo, W. S. *et al.* Polymeric delivery of therapeutic RAE-1 plasmid to the pancreatic islets for the prevention of type 1 diabetes. *J. Control. Release* **162**, 606-611 (2012).

76 Comartin, D. *et al.* CEP120 and SPICE1 cooperate with CPAP in centriole elongation. *Curr. Biol.* **23**, 1360-1366 (2013).

77 Shaheen, R. *et al.* A founder *CEP120* mutation in Jeune asphyxiating thoracic dystrophy expands the role of centriolar proteins in skeletal ciliopathies. *Hum. Mol. Genet.* **24**, 1410-1419 (2015).

78 Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18-24 (2009).

79 Kosobud, A. E. *et al.* Adaptation of subjective responses to alcohol is affected by an interaction of *GABRA2* genotype and recent drinking. *Alcohol. Clin. Exp. Res.* **39**, 1148-1157 (2015).

80 Duka, T. *et al.* Motivational effects of methylphenidate are associated with *GABRA2* variants conferring addiction risk. *Front. Behav. Neurosci.* **9**:304 (2015).

81 Bauer, L. O., Yang, B. Z., Houston, R. J., Kranzler, H. R. & Gelernter, J. *GABRA2* genotype, impulsivity, and body mass. *Am. J. Addict.* **21**, 404-410 (2012).

82 Hu, J. *et al.* Endothelial cell-derived angiopoietin-2 controls liver regeneration as a spatiotemporal rheostat. *Science* **343**, 416-419 (2014).

83 Cullberg, K. B. *et al.* Effect of weight loss and exercise on angiogenic factors in the circulation and in adipose tissue in obese subjects. *Obesity* **21**, 454-460 (2013).

84 Xue, Y. *et al.* *FOXC2* controls *Ang-2* expression and modulates angiogenesis, vascular patterning, remodeling, and functions in adipose tissue. *Proc. Natl. Acad. Sci.*

USA **105**, 10167-10172 (2008).

85 Hardie, D. G., Ross, F. A. & Hawley, S. A. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nat. Rev. Mol. Cell Biol.* **13**, 251-262 (2012).

86 Xu, X. J. *et al.* Insulin sensitive and resistant obesity in humans: AMPK activity, oxidative stress, and depot-specific changes in gene expression in adipose tissue. *J. Lipid Res.* **53**, 792-801 (2012).

87 Milan, D. *et al.* A mutation in *PRKAG3* associated with excess glycogen content in pig skeletal muscle. *Science* **288**, 1248-1251 (2000).

88 Lindholm-Perry, A. *et al.* DNA polymorphisms and transcript abundance of *PRKAG2* and phosphorylated AMP-activated protein kinase in the rumen are associated with gain and feed intake in beef steers. *Anim. Genet.* **45**, 461-472 (2014).

89 Jing, L. *et al.* Transcriptome analysis of mRNA and miRNA in skeletal muscle indicates an important network for differential Residual Feed Intake in pigs. *Sci. Rep.* **5** (2015).

90 Tian, G. & Cowan, N. J. Tubulin-specific chaperones: components of a molecular machine that assembles the *α/β* heterodimer. *Methods Cell Biol.* **115**, 155-171 (2012).

91 Glessner, J. T. *et al.* A genome-wide study reveals copy number variants exclusive to childhood obesity cases. *Am. J. Hum. Genet.* **87**, 661-666 (2010).

92 Chesi, A. & Grant, S. F. The genetics of pediatric obesity. *Trends Endocrin. Met.* **26**, 711-721 (2015).

93 Khokha, R., Murthy, A. & Weiss, A. Metalloproteinases and their natural inhibitors in inflammation and immunity. *Nat. Rev. Immunol.* **13**, 649-665 (2013).

94 Dufour, A. & Overall, C. M. Missing the target: matrix metalloproteinase antitargets in inflammation and cancer. *Trends Pharmacol. Sci.* **34**, 233-242 (2013).

95 Miksztowicz, V. *et al.* Adiponectin predicts MMP-2 activity independently of obesity. *Eur. J. Clin. Invest.* **44**, 951-957 (2014).

96 Belo, V. A. *et al.* Effect of metabolic syndrome risk factors and *MMP-2* genetic variations on circulating MMP-2 levels in childhood obesity. *Mol. Biol. Rep.* **40**, 2697-2704, doi:10.1007/s11033-012-2356-7 (2013).

97 Morgan, A. R., Han, D. Y., Thompson, J. M., Mitchell, E. A. & Ferguson, L. R. Analysis of *MMP2* promoter polymorphisms in childhood obesity. *BMC Res. Notes* **4**, 253 (2011).

98 Caunt, C. J., Sale, M. J., Smith, P. D. & Cook, S. J. MEK1 and MEK2 inhibitors and cancer therapy: the long and winding road. *Nat. Rev. Cancer* **15**, 577-592 (2015).

99 Xiao, B. *et al.* Structure of mammalian AMPK and its regulation by ADP. *Nature* **472**, 230-233 (2011).

100 Zhou, S., Lechpammer, S., Greenberger, J. S. & Glowacki, J. Hypoxia inhibition of adipocytogenesis in human bone marrow stromal cells requires transforming growth factor-*β*/Smad3 signaling. *J. Biol. Chem.* **280**, 22688-22696 (2005).

101 Arakawa, H. Netrin-1 and its receptors in tumorigenesis. *Nat. Rev. Cancer* **4**, 978-987 (2004).

102 Williams, M. E. *et al.* *UNC5A* promotes neuronal apoptosis during spinal cord development independent of netrin-1. *Nat. Neurosci.* **9**, 996-998 (2006).

103 Rosenberger, P. *et al.* Hypoxia-inducible factor–dependent induction of netrin-1 dampens inflammation caused by hypoxia. *Nat. Immunol.* **10**, 195-202 (2009).

104 Morgan, M., Iaconcig, A. & Muro, A. F. *CPEB2*, *CPEB3* and *CPEB4* are coordinately regulated by miRNAs recognizing conserved binding sites in paralog positions of their 3'-UTRs. *Nucleic Acids Res.* **38**, 7698-7710 (2010).

105 Hagele, S., Kuhn, U., Boning, M. & Katschinski, D. Cytoplasmic polyadenylation-element-binding protein (*CPEB*) 1 and 2 bind to the *HIF-1α* mRNA 3'-UTR and modulate *HIF-1α* protein expression. *Biochem. J* **417**, 235-246 (2009).

106 Chen, P. J. & Huang, Y. S. CPEB2-eEF2 interaction impedes *HIF-1α* RNA translation. *EMBO J.* **31**, 959-971 (2012).

107 Morgan, A. J. *et al.* *CXCR6* identifies a putative population of retained human lung T cells characterised by co-expression of activation markers. *Immunobiology* **213**, 599-608 (2008).

108 Paust, S. *et al.* Critical role for the chemokine receptor *CXCR6* in NK cell-mediated antigen-specific memory of haptens and viruses. *Nat. Immunol.* **11**, 1127-1135 (2010).

109 La Porta, C. A. CXCR6: the role of environment in tumor progression. Challenges for therapy. *Stem Cell Rev. Rep.* **8**, 1282-1285 (2012).

110 Lin, S. *et al.* Chemokine C-X-C motif receptor 6 contributes to cell migration during hypoxia. *Cancer Lett.* **279**, 108-117 (2009).

111 Zhang, W. *et al.* Lactotransferrin could be a novel independent molecular prognosticator of nasopharyngeal carcinoma. *Tumour Biol.* **36**, 675-683 (2015).

112 Siqueiros-Cendón, T. *et al.* Immunomodulatory effects of lactoferrin. *Acta Pharmacol. Sin.* **35**, 557-566 (2014).

113 Gharib, S. A., Luchtel, D. L., Madtes, D. K. & Glenny, R. W. Global gene annotation analysis and transcriptional profiling identify key biological modules in hypoxic pulmonary

hypertension. *Physiol. Genomics* **22**, 14-23 (2005).

114 Zakharova, E. T., Kostevich, V. A., Sokolov, A. V. & Vasilyev, V. B. Human apo-lactoferrin as a physiological mimetic of hypoxia stabilizes hypoxia-inducible factor-1 *α*. *Biometals* **25**, 1247-1259 (2012).

115 Ni, Z. *et al. RPRD1A* and *RPRD1B* are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nat. Struct. Mol. Biol.* **21**, 686-695 (2014).

116 Wu, Y. *et al.* p15RS attenuates Wnt/*β*-catenin signaling by disrupting *β*-catenin.TCF4 Interaction. *J. Biol. Chem.* **285**, 34621-34631 (2010).

117 Liu, C. *et al.* p15RS/RPRD1A (p15INK4b-related sequence/regulation of nuclear pre-mRNA domain-containing protein 1A) interacts with *HDAC2* in inhibition of the Wnt/beta-catenin signaling pathway. *J. Biol. Chem.* **290**, 9701-9713 (2015).

118 Yoshimatsu, H. *et al.* Functional involvement of RFVT3/SLC52A3 in intestinal riboflavin absorption. *Am. J. Physiol.* **306**, G102-110 (2014).

119 Fujimura, M. *et al.* Functional characteristics of the human ortholog of riboflavin transporter 2 and riboflavin-responsive expression of its rat ortholog in the small intestine indicate its involvement in riboflavin absorption. *J. Nutr.* **140**, 1722-1727 (2010).

120 Storz, J. F. & Moriyama, H. Mechanisms of hemoglobin adaptation to high altitude hypoxia. *High Alt. Med. Biol.* **9**, 148-157 (2008).

121 Bartkowiak, B. & Greenleaf, A. L. Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex. *J. Biol. Chem.* **290**, 1786-1795 (2015).

122 Joshi, P. M., Sutor, S. L., Huntoon, C. J. & Karnitz, L. M. Ovarian cancer-associated mutations disable catalytic activity of *CDK12*, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly (ADP-ribose) polymerase inhibitors. *J. Biol. Chem.* **289**, 9247-9253 (2014).

123 Bajrami, I. *et al.* Genome-wide profiling of genetic synthetic lethality identifies *CDK12* as a novel determinant of PARP1/2 inhibitor sensitivity. *Cancer Res.* **74**, 287-297 (2014).

124 Ekumi, K. M. *et al.* Ovarian carcinoma *CDK12* mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. *Nucleic Acids Res.* **43**, 2575-2589 (2015).

125 Yonekura, H. *et al.* Antisense display—a method for functional gene screening: evaluation in a cell-free system and isolation of angiogenesis-related genes. *Nucleic Acids Res.* **27**, 2591-2600 (1999).

126 Wiegand, N. *et al.* Comparative calorimetric analysis of 13 different types of human

healthy and pathologic collagen tissues. *Thermochim. Acta* **568**, 171-174 (2013).

127 Pepin, M., Schwarze, U., Superti-Furga, A. & Byers, P. H. Clinical and genetic features of Ehlers-Danlos syndrome type IV, the vascular type. *New Engl. J. Med.* **342**, 673-680 (2000).

128 Jørgensen, A. *et al.* Vascular Ehlers–Danlos Syndrome in siblings with biallelic *COL3A1* sequence variants and marked clinical variability in the extended family. *Eur. J. Hum. Genet.* **23**, 796-802 (2015).

129 Mayer, K., Kennerknecht, I. & Steinmann, B. Clinical utility gene card for: Ehlers-Danlos syndrome types I-VII and variants-update 2012. *Eur. J. Hum. Genet.* **21** (2013).

130 Parapia, L. A. & Jackson, C. Ehlers-Danlos syndrome–a historical review. *Brit. J. Haematol.* **141**, 32-35 (2008).

131 Zong, F. Y. *et al.* The RNA-binding protein *QKI* suppresses cancer-associated aberrant splicing. *PLoS Genet.* **10**, e1004289 (2014).

132 Conn, S. J. *et al.* The RNA binding protein quaking regulates formation of circRNAs. *Cell* **160**, 1125-1134 (2015).

133 Guo, W. *et al.* ASSA14-03-03 QKI deficiency promotes *FoxO1* mediated nitrosative stress and endoplasmic reticulum stress contributing to increased vulnerability to ischaemic injury in diabetic heart. *Heart* **101**, A9-A10 (2015).

134 van der Veer, E. P. *et al.* Quaking, an RNA-binding protein, is a critical regulator of vascular smooth muscle cell phenotype. *Circ. Res.* **113**, 1065-1075 (2013).

135 Winters, T., McNicoll, F. & Jessberger, R. Meiotic cohesin *STAG3* is required for chromosome axis formation and sister chromatid cohesion. *EMBO J.*, e201387330 (2014).

136 Caburet, S. *et al.* Mutant cohesin in premature ovarian failure. *New Engl. J. Med.* **370**, 943-949 (2014).

137 PEZZI, N. *et al.* *STAG3*, a novel gene encoding a protein involved in meiotic chromosome pairing and location of STAG3-related genes flanking the Williams-Beuren syndrome deletion. *FASEB J.* **14**, 581-592 (2000).

138 Rudolph, J., Crawford, J. J., Hoeflich, K. P. & Wang, W. Inhibitors of p21-activated kinases (PAKs). *J. Med. Chem.* **58**, 111-129 (2015).

139 Ke, Y., Wang, X., Jin, X. Y., Solaro, R. J. & Lei, M. PAK1 is a novel cardiac protective signaling molecule. *Front. Med.* **8**, 399-403 (2014).

140 Davis, R. T., Simon, J. N., Ke, Y., Wolska, B. M. & Solaro, R. J. p21-activated kinase-1 (*Pak1*) Mediates Exercise-induced Cardiac Remodeling Through Calcineurin Signaling.

Circ. Res. **117**, A226-A226 (2015).

141 Kelly, M. L., Astsaturov, A., Rhodes, J. & Chernoff, J. A Pak1/Erk signaling module acts through Gata6 to regulate cardiovascular development in zebrafish. *Dev. Cell* **29**, 350-359 (2014).

142 Yin, J.-w. & Wang, G. The Mediator complex: a master coordinator of transcription and cell lineage development. *Development* **141**, 977-987 (2014).

143 Schiano, C., Casamassimi, A., Vietri, M. T., Rienzo, M. & Napoli, C. The roles of mediator complex in cardiovascular diseases. *BBA-Gene Regul. Mech.* **1839**, 444-451 (2014).

144 Spitler, K. M., Ponce, J. M., Hall, D. D. & Grueter, C. E. Cardiac *Med1* is necessary for postnatal survival in mice. *Circ. Res.* **117**, A278-A278 (2015).

145 Goraya, T. A., Masada, N., Ciruela, A. & Cooper, D. M. Sustained entry of Ca^{2+} is required to activate Ca^{2+}-calmodulin-dependent phosphodiesterase 1A. *J. Biol. Chem.* **279**, 40494-40504 (2004).

146 Chan, S. & Yan, C. *PDE1* isozymes, key regulators of pathological vascular remodeling. *Curr. Opin. Pharmacol.* **11**, 720-724 (2011).

147 Nino, P. K. B. *et al.* Phosphodiesterase 1 regulation is a key mechanism in vascular aging. *Clin. Sci.*, DOI: 10.1042/CS20140753 (2015).

148 Kunej, T. *et al.* Obesity gene atlas in mammals. *J. Genomics* **1**, 45-55 (2013).

149 Butler, J. *et al.* ALLPATHS: *de novo* assembly of whole-genome shotgun microreads. *Genome Res.* **18**, 810-820 (2008).

150 Li, R. *et al. De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265-272 (2010).

151 Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).

152 Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).

153 Kapitonov, V. V. & Jurka, J. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* **9**, 411-412 (2008).

154 Bao, Z. & Eddy, S. R. Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269-1276 (2002).

155 Price, A. L., Jones, N. C. & Pevzner, P. A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **21 Suppl 1**, i351-358 (2005).

156 Li, Y. *et al.* Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome *de novo* assembly. *Nat. Biotechnol.* **29**, 723-730 (2011).

157 Abyzov, A., Urban, A.E., Snyder, M., & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974-984 (2011).

158 Warr, A. *et al.* Identification of low-confidence regions in the pig reference genome (Sscrofa10.2). *Front. Genet.* **6**, 338 (2015).

159 Paudel, Y. *et al.* Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. *BMC Genomics* **14**, 449 (2013).

160 Molnar, J. *et al.* Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary. *BMC Genomics* **15**, 761 (2014).

161 Paudel, Y. *et al.* Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC Genomics* **16**, 330 (2015).

162 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44-57 (2009).

163 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81:** 559-575 (2007).

164 Patterson, N., Price, A.L., & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2:** e190 (2006).

165 Barrett, J.C., Fry, B., Maller, J., & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21:** 263-265 (2005).

166 Atanur, S. S. *et al.* Genome sequencing reveals loci under artificial selection that underlie disease phenotypes in the laboratory rat. *Cell* **154**, 691-703 (2013).

167 Strimmer, K. FDRtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461-1462 (2008).

168 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).

169 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562-578 (2012).

170 Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988-995 (2004).

171 Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19 Suppl 2**, ii215-225 (2003).

172 Parra, G., Blanco, E. & Guigo, R. GeneID in *Drosophila*. *Genome Res.* **10**, 511-515 (2000).

173 Salamov, A. A. & Solovyev, V. V. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516-522 (2000).

174 Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878-2879 (2004).

175 Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).

176 Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res.* **12**, 656-664 (2002).

177 Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).

178 Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365-370 (2003).

179 Mitchell, A. *et al.* The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**, D213-221 (2015).

180 Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279-285 (2016).

181 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240 (2014).

182 Gene Ontology, C. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049-1056 (2015).

183 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457-462 (2016).

184 Schug, J. *et al.* Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol.* **6**, R33 (2005).

185 Excoffier, L., & Lischer, H.E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10:** 564-567 (2010)

Supplemental URLs

Domestic Animal Diversity Information System, http://www.fao.org/dad-is/;

dbSNP database, http://www.ncbi.nlm.nih.gov/projects/SNP/; Ensembl, http://ensembl.fugu-sg.org/index.html; Pig reference genome (Sscrofa10.2), http://asia.ensembl.org/Sus_scrofa/Info/Index; TimeTree, http://www.timetree.org/; Mouse Genome Informatics, http://www.informatics.jax.org/; Meryl, http://kmer.sourceforge.net/wiki/index.php/Getting_Started_with_Meryl; RepeatMasker, RepeatProteinMask and RepeatModeler, http://www.RepeatMasker.org; LASTZ (at the Miller Lab website), http://www.bx.psu.edu/miller_lab/; Picard, http://sourceforge.net/projects/picard/; SOAP, http://soap.genomics.org.cn/; DAVID, http://david.abcc.ncifcrf.gov/; NCBI RefSeqGene, http://www.ncbi.nlm.nih.gov/refseq/rsg/.