J Clin Periodontol 2011; 38: 103–107 doi: 10.1111/j.1600-051X.2010.01653.x

Journal of Clinical Periodontology

Guest Editorial

Periodontal genetics: a decade of genetic association studies mandates better study designs

Arne S. Schäfer[1], Søren Jepsen[2] and Bruno G. Loos[3]

[1]Institute for Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Schleswig-Holstein, Germany [2]Department of Conservative and Preventive Dentistry, Rheinische–Friedrichs–Wilhelm University, Bonn, Germany [3]Department of Conservative and Preventive Dentistry, Academic Centre for Dentistry, Amsterdam, The Netherlands

Schäfer AS, Jepsen S, Loos BG: Periodontal genetics: a decade of genetic association studies mandates better study designs. J Clin Periodontol 2011; 38: 103–107. doi: 10.1111/j.1600-051X.2010.01653.x.

Periodontitis develops in a limited subset of humans. About 10% of the population will develop severe forms of destructive periodontal disease. The disease is initiated by microorganisms, and perhaps viruses, in the subgingival biofilm (environmental factors), and further affected by lifestyle factors such as smoking, stress and diet. It can also be influenced by acquired systemic diseases, which reduce or hamper an ''optimal'' host response. On top of the above risk factors are modifying disease genes, responsible for susceptibility to periodontitis. For these disease-modifying genes, Mendelian principles do not apply, because for a given genetic variation in a gene or locus, both heterozygous as well as homozygous subjects, may not necessarily develop the disease; other genetic risk factors (gene–gene interactions) and environmental and lifestyle risk factors (gene–environmental, gene–lifestyle and environmental–lifestyle interactions) also need to be present simultaneously for the phenotype to develop (definition of complex disease) (Craig 2008) (Fig. 1). It is clear that any study population suffers from a great deal of such inherent heterogeneity, but accumulated empirical data from large genetic association studies indicate that commonly expressed concerns about such ''noise'' may be overstated, and hitherto, no excess of false-positive or -negative association with common genetic variants was revealed. This indicates that selection biases based on these issues have a minimal impact on genotype distributions, if the patients are well diagnosed and the statistical power is sufficient (McCarthy et al. 2008).

Despite tremendous efforts and published papers in the field of genetic association studies for periodontitis over the last decade, the causative gene polymorphisms of periodontitis and their pathophysiological effects are still very controversial. What went wrong? A couple of important questions need to be raised. Most sensitive is the question: do association studies actually provide the right tools to elucidate the underlying genetic factors of periodontitis? Next, what have researchers experienced in the fields of other complex human diseases? Here, we should ask if they experienced the same shortcomings, and if so, what did they do about them, and most importantly, what can we learn from them?

To answer these questions straightforwardly: yes, association studies are effective tools to elucidate the genetic risk factors of complex diseases, but particularly for the more frequent variants. Association studies have a limited power to detect the rare genetic risk factors. Secondly, researchers of other complex human diseases have also experienced comparable shortcomings but their study designs have been adapted over the last years. Likewise, after decades of controversial literature (Morgan et al. 2007), it is considered that most of the common genetic risk factors have been unveiled for coronary heart disease (CHD) and type 2 diabetes (T2D), and there are at least 11 for CHD (Schunkert et al. 2010) and 18–24 for T2D (Prokopenko et al. 2008, Stolerman & Florez 2009). Interestingly, most of them had not been thought of as likely candidate genes before.

Development in Recent Years

The year 2007 was the turning point in the genetics of complex human diseases, ushered into the era of genome-wide association studies (GWAS) by the milestone publication of the Wellcome Trust Case Control Consortium (WTCCC 2007). To reach this era, two paradigm shifts were necessary. Firstly, it was realized that the literature-based hypotheses on candidate genes do not usually reflect the situation in nature, and it was right in time, that technical advances allowed the hypothesis-free approach of simultaneously testing 500,000–1,000,000 and more polymorphisms spread across the genome in the GWAS. Secondly, it was recognized that large case–control populations were the indispensible prerequisite for any association study to overcome the inherent heterogeneity in populations; this resulted in the formation of extensive international consortia for the recruitment of the appropriate numbers. To identify the common genetic risk variants within the population, case–control panels of over 1000 well-defined cases and many more controls are considered the standard in the genetics of complex diseases like CHD (Erdmann et al. 2009, Samani et al. 2009) and T2D (Saxena et al. 2007, Scott et al. 2007, Steinthorsdottir et al. 2007, Zeggini et al. 2007). Although these studies proved successful in identifying the common genetic susceptibility variants, the power of an individual study

that encompasses over 1000 cases is still limited to detecting small or modest effects of common SNPs. To detect such variants, or even low frequent variants with a large effect at the individual level, the investigators of these studies joined forces and combined data for complex disease GWAS meta-analysis, which eventually included over tens of thousands of cases and controls (Zeggini et al. 2008, Erdmann et al. 2009, Samani et al. 2009, Schunkert et al. 2010). These efforts boosted the power to detect common susceptibility loci with modest effects, although the power to detect smaller effects still remained low (Fig. 2). Besides, this issue is an inherent limitation to the study design of GWAS. They are only sharp tools to identify frequent variants, most often with a limited genetic effect, but they usually miss rare variants with greater effects. This is mainly due to the fact that most GWAS are necessarily performed in an underpowered situation due to financial limitations. Here, candidate gene studies that can use much higher sample numbers in the explorative study provide a more adequate study design to identify rare variants with stronger genetic effects. Consequently, GWAS will not eliminate the necessity of well-designed candidate gene studies in the near future.

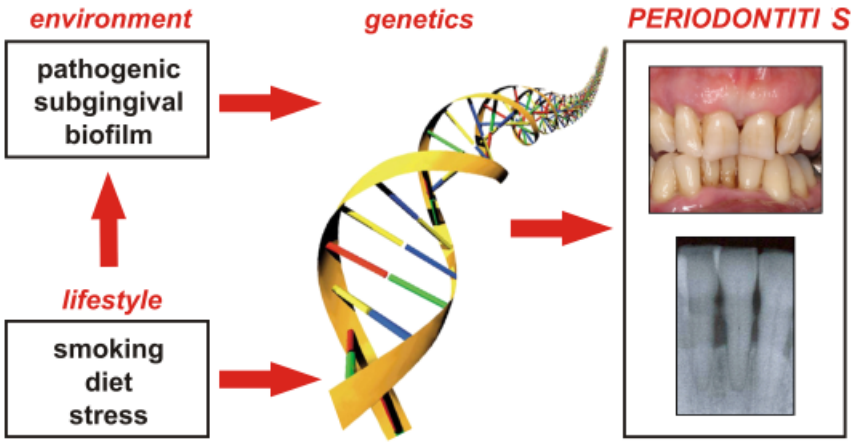

Fig. 1. Periodontitis is a complex disease (for definition see Craig 2008). The modifying disease genes play a central role and determine susceptibility to periodontitis. With this intrinsic characteristic, lifestyle factors pave the way to the disease, which is eventually initiated by an infection (environmental factors). Both have an individual as well as an additive effect on the (patho)physiological response of the patient, which is, in the final step, determined by the individual genetic predisposition. Therefore, in the complexity of periodontitis, gene–gene interactions as well as gene–environmental, gene–lifestyle and environmental–lifestyle interactions play a role in the development of the phenotype.

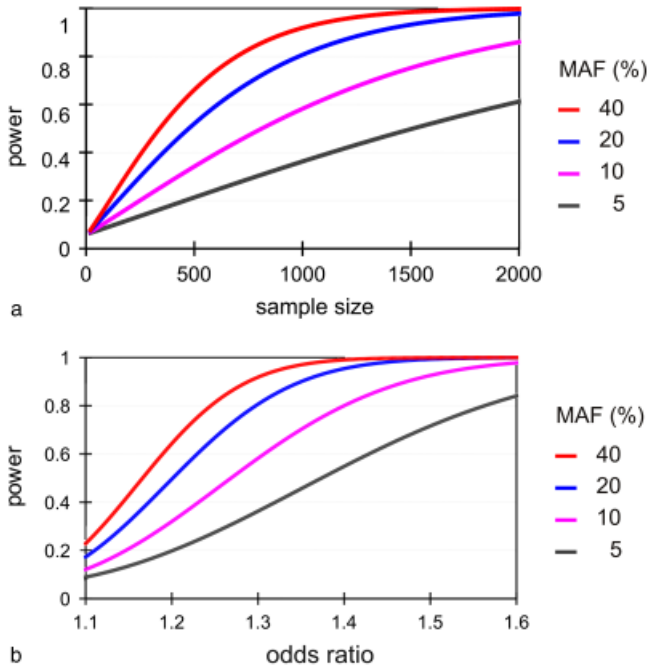

Fig. 2. Statistical power in relation to the sample size, allele frequency and odds ratio. (a) To identify a genetic risk variant with a minor allele frequency (MAF) of 20% in the general population, 1000 cases and 2000 controls were required to achieve the necessary statistical power of 0.8 [The statistical power was calculated as described in Dupont & Plummer 1998, for an average odds ratio (OR) of 1.3, and 2 times as many controls as cases were considered. A power of 0.8 is regarded as statistically significant.]. (b) Statistical power calculation was taken from 1000 cases and 2000 controls as a basis. A population of this size would allow the identification of rare risk variants with relatively large genetic effects (e.g. a genetic effect of >1.5 would be required to detect a risk allele with 5% MAF).

Limitations of Genetic Association Studies

In this context, an important issue of the current discussion in human genetics has to be touched upon. Although the recent GWAS in complex diseases provided valuable insight into the genetic basis and identified many genetic susceptibility genes, they could only explain a small proportion of the heritability of these traits (Frazer et al. 2009). Where this heritability is likely to lie and how research strategies for uncovering the missing genetic risk factors in these diseases might best be developed, are currently heavily debated (Eichler et al. 2010). There is strong evidence that rare variants have an important role (Bodmer & Bonilla 2008), that genetic variance may to some extent also be explained by structural variations (Conrad et al. 2010, Craddock et al. 2010), gene interactions (Phillips 2008, Moore & Williams 2009) and we certainly need to learn more about how the environment modifies allelic effects (Christensen et al. 2009, Check Hayden 2010). Only if the major contributory elements are fully understood, which jointly predisposes people to the specific individual disease phenotype, a risk prediction or an individualized therapy will be possible. Here, to meet with the current discussion on the best future research strategies to elucidate the missing herit-

ability would be inadequate, and we would like to refer to the cited literature. Detecting associations with common genetic variants is still considered as the first and also the most straightforward and cost-efficient step, towards a better understanding of the genetic architecture of a complex disease.

Which criteria are currently considered as compulsory for any effective candidate gene association or GWA study at a recognized scientific level? Four issues are of chief importance: these are case selection; the statistical power and replication; and lastly the capture of the complete genetic information.

Case Selection

The first indispensable prerequisite is a case-selection strategy, which is designed for enrichment of specific disease-predisposing alleles. These should include efforts to minimize phenotypic heterogeneity by stringent diagnosis criteria, and should focus on extreme cases, defined for example by a particularly early age of onset or level of severity. In most circumstances, and particularly when the total sample size has financial or operational constraints, efforts to enrich case selection are very likely to improve the statistical power (McCarthy et al. 2008).

Statistical Power and Replication

The second prerequisite for any association study is, as pointed out above, a sufficiently large case–control analysis population which provides the necessary statistical power to identify a disease associated polymorphism of a given allele frequency. Variants that are associated with complex diseases increase the disease risk rather modestly. The odds ratio (OR) of such a variant usually does not exceed 1.3. Accordingly, to identify a variant with an OR of 1.3 and of an allele frequency of 20% in the general population, 1000 well-defined cases are necessary to reach the necessary statistical power of 0.8. (Fig. 2). To match these cases, at least the same number of controls is necessary, ideally twice as many. With a sufficiently powered clinical analysis population, the hypothesis of the study can be generated. In other words, the explorative association study merely observes an association (or does not observe), thus creating the hypothesis, that a specific variant is/is not associated with the disease. The hypothesis obviously requires a test, which is the replication. This replication is of utmost importance, and considered as the gold standard, which defines that each association is only as good as its replication. It goes without saying that the replication needs to be performed in an independent case–control population of the same phenotype (diagnosis criteria), and to be sampled from the same ethnic background. A repetition of the experiment with samples from different ethnic groups, with different diagnosis criteria, or independent cases but the same controls is not a replication and does not test the hypothesis properly. Only after being confirmed, can the hypothesis be validated in different sub-phenotypes or in different ethnic groups.

It is a further essential prerequisite to have the equal size of the initial case–control population or ideally, twice the size. Why so? Estimates of the genetic effect based on new association findings tend to be upwardly biased due to a phenomenon known as the ''winner's curse'' (Lohmueller et al. 2003). If the sample size calculation for a subsequent study is based on an overestimated effect size, this overestimation in the initial study may cause follow-up studies to be underpowered and so they fail (Ioannidis et al. 2001). Although overestimation of the genetic effect decreases as the statistical power of the association study increases (Xiao & Boehnke 2009), the ''winner's curse'' bias is an issue of serious concern, and as a consequence replication case–control populations should be larger than the explorative cohort to increase the statistical power for the replication experiment. Accordingly, this upward bias of a new association finding usually appears to give less significant association signals in the replication than in the explorative study. This effect can be evaluated and corrected for by various approaches (Xiao & Boehnke 2009), but it is easiest avoided by an increased power in the replication. These descriptions illustrate why several thousands of cases and controls are usually required for an association study, and why large scientific consortia had to be formed to jointly recruit the necessary samples.

For the improvement of the statistical power, an alternative to the performance of de novo large-scale experiments could lie in a meta-analysis of already existing studies. If some studies propose a genetic effect on the phenotype, but others do not, in a meta-analysis the majority of these studies can show the same trend and a statistically significant p-value in the test. This would surely provide new information. But a meta-analysis cannot be considered as the replication of a hypothesis, because it does not create nor analyse new data. It only refines the hypothesis. Before the potentially significant disease association can be considered as replicated, it also needs a test in an independent analysis population of sufficient statistical power (Fig. 2). Necessary prerequisites for a meta-analysis are the same diagnosis criteria between the different studies, and if various genetic markers were independently tested, a correction for multiple testing.

Capture of the Complete Genetic Information

Apart from insufficient statistical power of most genetic association studies in periodontitis within the last decade, most studies were unable to draw an unambiguous conclusion from their findings, even when the outcome was negative. Apart of the reasons mentioned above, this was also because most genetic association studies on periodontitis did not capture the complete genetic information of the region of interest. Only one or a few variants were genotyped in these studies, and usually, their selection had reasons that were more historical than biological. But a precise statement about a possible association of a genetic locus with a disease can only be made, if the full haplotype information has been assessed (Slatkin 2008). Testing one polymorphism might at best be sufficient to indicate an association with every other polymorphism in the same haplotype block, and genotyping of a single genetic variant alone usually does not allow the conclusion, whether the gene of interest is disease associated or not. Genes are usually patchworks of different haplotype blocks, being mostly in poor-to-moderate linkage disequilibrium. Thus, information on the potential association of one haplotype block provides little to no information on the association with another (Slatkin 2008). As findings of negative associations of single variants or haplotype blocks cannot rule out a potential disease association of the gene of interest, such studies

have a very limited value. Therefore, as the outcome of an association study is always unclear, it is necessary that each genetic study needs to assess the haplotype information as completely as possible. Genotyping one variant simply provides no information on a possible disease association of the genetic locus when the outcome is negative.

Currently, almost all of the genetic variants associated with periodontitis that have been published are questionable. This is essentially due to – as we see it now – the poor designs of most of these studies. Knowing this, what can we do about it? The first and most important step will be a stringent publication policy by dental scientific journals, which will have to ban mere observational studies from publication. But as a consequence, science and public oral health will benefit from this policy, because it will swiftly catalyse the creation of the necessary research consortia, that allows the sampling of sufficiently powered and well-designed case–control populations. This will be a considerable challenge for the scientific community in dentistry, as the recruitment of an appropriate, well-defined case–control population and the genetic experiments to come, are both exceedingly money- and time consuming. But only if this challenge is taken, will it be possible to elucidate the genetic basis of periodontitis, with the translational effects on diagnosis, classification, prevention and the development of advanced medicine for oral health. The progress in the other disciplines of human medicine in the last 3 years has successfully proved that this path is worth the efforts.

References

Bodmer, W. & Bonilla, C. (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics* **40**, 695–701.

Check Hayden, E. (2010) Human genome at ten: life is complicated. *Nature* **464**, 664–667.

Christensen, B. C., Houseman, E. A., Marsit, C. J., Zheng, S., Wrensch, M. R., Wiemels, J. L., Nelson, H. H., Karagas, M. R., Padbury, J. F., Bueno, R., Sugarbaker, D. J., Yeh, R. F., Wiencke, J. K. & Kelsey, K. T. (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genetics* **5**, e1000602.

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C. H., Kristiansson, K., Macarthur, D. G., Macdonald, J. R., Onyiah, I., Pang, A. W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N. P., Lee, C., Scherer, S. W. & Hurles, M. E. (2010) Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712.

Craig, J. (2008) Complex diseases: Research and applications. *Nature Education* **1**.

Craddock, N., Hurles, M. E., Cardin, N., Pearson, R. D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D. F., Giannoulatou, E., Holmes, C., Marchini, J. L., Stirrups, K., Tobin, M. D., Wain, L. V., Yau, C., Aerts, J., Ahmad, T., Andrews, T. D., Arbury, H., Attwood, A., Auton, A., Ball, S. G., Balmforth, A. J., Barrett, J. C., Barroso, I., Barton, A., Bennett, A. J., Bhaskar, S., Blaszczyk, K., Bowes, J., Brand, O. J., Braund, P. S., Bredin, F., Breen, G., Brown, M. J., Bruce, I. N., Bull, J., Burren, O. S., Burton, J., Byrnes, J., Caesar, S., Clee, C. M., Coffey, A. J., Connell, J. M., Cooper, J. D., Dominiczak, A. F., Downes, K., Drummond, H. E., Dudakia, D., Dunham, A., Ebbs, B., Eccles, D., Edkins, S., Edwards, C., Elliot, A., Emery, P., Evans, D. M., Evans, G., Eyre, S., Farmer, A., Ferrier, I. N., Feuk, L., Fitzgerald, T., Flynn, E., Forbes, A., Forty, L., Franklyn, J. A., Freathy, R. M., Gibbs, P., Gilbert, P., Gokumen, O., Gordon-Smith, K., Gray, E., Green, E., Groves, C. J., Grozeva, D., Gwilliam, R., Hall, A., Hammond, N., Hardy, M., Harrison, P., Hassanali, N., Hebaishi, H., Hines, S., Hinks, A., Hitman, G. A., Hocking, L., Howard, E., Howard, P., Howson, J. M., Hughes, D., Hunt, S., Isaacs, J. D., Jain, M., Jewell, D. P., Johnson, T., Jolley, J. D., Jones, I. R., Jones, L. A., Kirov, G., Langford, C. F., Lango-Allen, H., Lathrop, G. M., Lee, J., Lee, K. L., Lees, C., Lewis, K., Lindgren, C. M., Maisuria-Armer, M., Maller, J., Mansfield, J., Martin, P., Massey, D. C., McArdle, W. L., McGuffin, P., McLay, K. E., Mentzer, A., Mimmack, M. L., Morgan, A. E., Morris, A. P., Mowat, C., Myers, S., Newman, W., Nimmo, E. R., O'Donovan, M. C., Onipinla, A., Onyiah, I., Ovington, N. R., Owen, M. J., Palin, K., Parnell, K., Pernet, D., Perry, J. R., Phillips, A., Pinto, D., Prescott, N. J., Prokopenko, I., Quail, M. A., Rafelt, S., Rayner, N. W., Redon, R., Reid, D. M., Renwick, A., Ring, S. M., Robertson, N., Russell, E., St Clair, D., Sambrook, J. G., Sanderson, J. D., Schuilenburg, H., Scott, C. E., Scott, R., Seal, S., Shaw-Hawkins, S., Shields, B. M., Simmonds, M. J., Smyth, D. J., Somaskantharajah, E., Spanova, K., Steer, S., Stephens, J., Stevens, H. E., Stone, M. A., Su, Z., Symmons, D. P., Thompson, J. R., Thomson, W., Travers, M. E., Turnbull, C., Valsesia, A., Walker, M., Walker, N. M., Wallace, C., Warren-Perry, M., Watkins, N. A., Webster, J., Weedon, M. N., Wilson, A. G., Woodburn, M., Wordsworth, B. P., Young, A. H., Zeggini, E., Carter, N. P., Frayling, T. M., Lee, C., McVean, G., Munroe, P. B., Palotie, A., Sawcer, S. J., Scherer, S. W., Strachan, D. P., Tyler-Smith, C., Brown, M. A., Burton, P. R., Caulfield, M. J., Compston, A., Farrall, M., Gough, S. C., Hall, A. S., Hattersley, A. T., Hill, A. V., Mathew, C. G., Pembrey, M., Satsangi, J., Stratton, M. R., Worthington, J., Deloukas, P., Duncanson, A., Kwiatkowski, D. P., McCarthy, M. I., Ouwehand, W., Parkes, M., Rahman, N., Todd, J. A., Samani, N. J. & Donnelly, P. (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713–720.

Dupont, W. D. & Plummer, W. D. Jr. (1998) Power and sample size calculations for studies involving linear regression. *Control Clinical Trials* **19**, 589–601.

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H. & Nadeau, J. H. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics* **11**, 446–450.

Erdmann, J., Grosshennig, A., Braund, P. S., Konig, I. R., Hengstenberg, C., Hall, A. S., Linsel-Nitschke, P., Kathiresan, S., Wright, B., Tregouet, D. A., Cambien, F., Bruse, P., Aherrahrou, Z., Wagner, A. K., Stark, K., Schwartz, S. M., Salomaa, V., Elosua, R., Melander, O., Voight, B. F., O'Donnell, C. J., Peltonen, L., Siscovick, D. S., Altshuler, D., Merlini, P. A., Peyvandi, F., Bernardinelli, L., Ardissino, D., Schillert, A., Blankenberg, S., Zeller, T., Wild, P., Schwarz, D. F., Tiret, L., Perret, C., Schreiber, S., El Mokhtari, N. E., Schafer, A., Marz, W., Renner, W., Bugert, P., Kluter, H., Schrezenmeir, J., Rubin, D., Ball, S. G., Balmforth, A. J., Wichmann, H. E., Meitinger, T., Fischer, M., Meisinger, C., Baumert, J., Peters, A., Ouwehand, W. H., Deloukas, P., Thompson, J. R., Ziegler, A., Samani, N. J. & Schunkert, H. (2009) New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nature Genetics* **41**, 280–282.

Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. (2009) Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics* **10**, 241–251.

Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A. & Contopoulos-Ioannidis, D. G. (2001) Replication validity of genetic association studies. *Nature Genetics* **29**, 306–309

Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. & Hirschhorn, J. N. (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease, **33**, 177–182.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. & Hirschhorn, J. N. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* **9**, 356–369.

Moore, J. H. & Williams, S. M. (2009) Epistasis and its implications for personal genetics. *American Journal Human Genetics* **85**, 309–320.

Morgan, T. M., Krumholz, H. M., Lifton, R. P. & Spertus, J. A. (2007) Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. *Journal of the American Medical Association* **297**, 1551–1561.

Phillips, P. C. (2008) Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics* **9**, 855–867.

Prokopenko, I., McCarthy, M. I. & Lindgren, C. M. (2008) Type 2 diabetes: new genes, new understanding. *Trends in Genetics* **24**, 613–621.

Samani, N. J., Deloukas, P., Erdmann, J., Hengstenberg, C., Kuulasmaa, K., McGinnis, R., Schunkert, H., Soranzo, N., Thompson, J., Tiret, L. & Ziegler, A. (2009) Large scale association analysis of novel genetic loci for coronary artery disease. *Arteriosclerosis, thrombosis, and vascular biology* **29**, 774–780.

Saxena, R., Voight, B. F., Lyssenko, V., Burtt, N. P., de Bakker, P. I., Chen, H., Roix, J. J., Kathiresan, S., Hirschhorn, J. N., Daly, M. J., Hughes, T. E., Groop, L., Altshuler, D., Almgren, P., Florez, J. C., Meyer, J., Ardlie, K., Bengtsson Bostrom, K., Isomaa, B., Lettre, G., Lindblad, U., Lyon, H. N., Melander, O., Newton-Cheh, C., Nilsson, P., Orho-Melander, M., Rastam, L., Speliotes, E. K., Taskinen, M. R., Tuomi, T., Guiducci, C., Berglund, A., Carlson, J., Gianniny, L., Hackett, R., Hall, L., Holmkvist, J., Laurila, E., Sjogren, M., Sterner, M., Surti, A., Svensson, M., Tewhey, R., Blumenstiel, B., Parkin, M., Defelice, M., Barry, R., Brodeur, W., Camarata, J., Chia, N., Fava, M., Gibbons, J., Handsaker, B., Healy, C., Nguyen, K., Gates, C., Sougnez, C., Gage, D., Nizzari, M., Gabriel, S. B., Chirn, G. W., Ma, Q., Parikh, H., Richardson, D., Ricke, D. & Purcell, S. (2007)

Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336.

Schunkert, H., Erdmann, J. & Samani, N. J. (2010) Genetics of myocardial infarction: a progress report. *European Heart Journal* **31**, 918–925.

Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L., Erdos, M. R., Stringham, H. M., Chines, P. S., Jackson, A. U., Prokunina-Olsson, L., Ding, C. J., Swift, A. J., Narisu, N., Hu, T., Pruim, R., Xiao, R., Li, X. Y., Conneely, K. N., Riebow, N. L., Sprau, A. G., Tong, M., White, P. P., Hetrick, K. N., Barnhart, M. W., Bark, C. W., Goldstein, J. L., Watkins, L., Xiang, F., Saramies, J., Buchanan, T. A., Watanabe, R. M., Valle, T. T., Kinnunen, L., Abecasis, G. R., Pugh, E. W., Doheny, K. F., Bergman, R. N., Tuomilehto, J., Collins, F. S. & Boehnke, M. (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345.

Slatkin, M. (2008) Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* **9**, 477–485.

Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G. B., Styrkarsdottir, U., Gretarsdottir, S., Emilsson, V., Ghosh, S., Baker, A., Snorradottir, S., Bjarnason, H., Ng, M. C., Hansen, T., Bagger, Y., Wilensky, R. L., Reilly, M. P., Adeyemo, A., Chen, Y., Zhou, J., Gudnason, V., Chen, G., Huang, H., Lashley, K., Doumatey, A., So, W. Y., Ma, R. C., Andersen, G., Borch-Johnsen, K., Jorgensen, T., van Vliet-Ostaptchouk, J. V., Hofker, M. H., Wijmenga, C., Christiansen, C., Rader, D. J., Rotimi, C., Gurney, M., Chan, J. C., Pedersen, O., Sigurdsson, G., Gulcher, J. R., Thorsteinsdottir, U., Kong, A. & Stefansson, K. (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nature Genetics* **39**, 770–775.

Stolerman, E. S. & Florez, J. C. (2009) Genomics of type 2 diabetes mellitus: implications for the clinician. *Nature Reviews Endocrinology* **5**, 429–436.

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

Xiao, R. & Boehnke, M. (2009) Quantifying and correcting for the winner's curse in genetic association studies. *Genetic Epidemiology* **33**, 453–462.

Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H., Timpson, N. J., Perry, J. R., Rayner, N. W., Freathy, R. M., Barrett, J. C., Shields, B., Morris, A. P., Ellard, S., Groves, C. J., Harries, L. W., Marchini, J. L., Owen, K. R., Knight, B., Cardon, L. R., Walker, M., Hitman, G. A., Morris, A. D., Doney, A. S., McCarthy, M. I. & Hattersley, A. T. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341.

Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T., de Bakker, P. I., Abecasis, G. R., Almgren, P., Andersen, G., Ardlie, K., Bostrom, K. B., Bergman, R. N., Bonnycastle, L. L., Borch-Johnsen, K., Burtt, N. P., Chen, H., Chines, P. S., Daly, M. J., Deodhar, P., Ding, C. J., Doney, A. S., Duren, W. L., Elliott, K. S., Erdos, M. R., Frayling, T. M., Freathy, R. M., Gianniny, L., Grallert, H., Grarup, N., Groves, C. J., Guiducci, C., Hansen, T., Herder, C., Hitman, G. A., Hughes, T. E., Isomaa, B., Jackson, A. U., Jorgensen, T., Kong, A., Kubalanza, K., Kuruvilla, F. G., Kuusisto, J., Langenberg, C., Lango, H., Lauritzen, T., Li, Y., Lindgren, C. M., Lyssenko, V., Marvelle, A. F., Meisinger, C., Midthjell, K., Mohlke, K. L., Morken, M. A., Morris, A. D., Narisu, N., Nilsson, P., Owen, K. R., Palmer, C. N., Payne, F., Perry, J. R., Pettersen, E., Platou, C., Prokopenko, I., Qi, L., Qin, L., Rayner, N. W., Rees, M., Roix, J. J., Sandbaek, A., Shields, B., Sjogren, M., Steinthorsdottir, V., Stringham, H. M., Swift, A. J., Thorleifsson, G., Thorsteinsdottir, U., Timpson, N. J., Tuomi, T., Tuomilehto, J., Walker, M., Watanabe, R. M., Weedon, M. N., Willer, C. J., Illig, T., Hveem, K., Hu, F. B., Laakso, M., Stefansson, K., Pedersen, O., Wareham, N. J., Barroso, I., Hattersley, A. T., Collins, F. S., Groop, L., McCarthy, M. I., Boehnke, M. & Altshuler, D. (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics* **40**, 638–645.

Address:
Arne S. Schäfer
Institute for Clinical Molecular Biology
Christian-Albrechts-University
Schittenhelmstrasse 12 24105
Kiel, *Schleswig-Holstein 24105*
Germany
E-mail: a.schaefer@ikmb.uni-kiel.de