

Review

Candidate Gene Identification Approach: Progress and Challenges

Mengjin Zhu and Shuhong Zhao

Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan 430070, P. R. China

Correspondence to: Shuhong Zhao, Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan 430070, P. R. China. Tel: +86-27-87281306; Fax: +86-27-87280408; E-mail: shzhao@mail.hzau.edu.cn

Received: 2007.08.26; Accepted: 2007.10.24; Published: 2007.10.25

Although it has been widely applied in identification of genes responsible for biomedically, economically, or even evolutionarily important complex and quantitative traits, traditional candidate gene approach is largely limited by its reliance on the priori knowledge about the physiological, biochemical or functional aspects of possible candidates. Such limitation results in a fatal information bottleneck, which has apparently become an obstacle for further applications of traditional candidate gene approach on many occasions. While the identification of candidate genes involved in genetic traits of specific interest remains a challenge, significant progress in this subject has been achieved in the last few years. Several strategies have been developed, or being developed, to break the barrier of information bottleneck. Recently, being a new developing method of candidate gene approach, digital candidate gene approach (DigiCGA) has emerged and been primarily applied to identify potential candidate genes in some studies. This review summarizes the progress, application software, online tools, and challenges related to this approach.

Key words: candidate gene approach; information bottleneck; digital candidate gene approach

1. Introduction

Based on the polygenic hypothesis, classical quantitative genetics considers a black box to reveal the holistic status of all genes associated with variation of complex and quantitative traits by complicated statistical methods. Such strategy could not independently decompose individual genes, which usually follow the Mendel's law, from the polygenic system of the investigated traits. Advances in molecular methods and quantitative techniques have clearly changed this status, which are able to look inside the black box of polygenic control for complex and quantitative traits with a more accurate description of how genes act to determine the phenotypic variation. More recently, major progress has been made in this field with the advent of genomics and its potential contribution to development of quantitative genetics. One of the hot interests of current quantitative genetics is systematically exploring an exact genetic architecture of the number, distribution and interaction of loci affecting the variations of biomedically, economically, and evolutionarily important complex and quantitative traits.

There are two approaches for genetic dissections of complex and quantitative traits, i.e., genome-wide scanning and candidate gene approach, which each has specific advantages and disadvantages. Genome-wide scanning usually proceeds without any presuppositions regarding the importance of specific functional features of the investigated traits, but of

which the principal disadvantage is expensive and resource intensive. In general, genome-wide scanning only locates the glancing chromosomal regions of quantitative trait loci (QTLs) at cM-level with the aid of DNA markers under family-based or population-based experimental designs, which usually embed a large number of candidate genes. In comparison, the alternative candidate gene approach has been proven to be extremely powerful for studying the genetic architecture of complex traits, which is a far more effective and economical method for direct gene discovery. Nevertheless, the practicability of traditional candidate gene approach is largely limited by its reliance on existing knowledge about the known or presumed biology of the phenotype under investigation, and unfortunately the detailed molecular anatomy of most biological traits remains unknown. It is quite necessary to develop new strategies to break the restriction of information bottleneck, although considerable candidate genes have already been identified.

In this article, we review and summarize the main research advances in the subject, including the outline of candidate gene approach and the extended strategies for breaking the information bottleneck of traditional candidate gene approach. Finally, as a new development of candidate gene approach, digital candidate gene approach (DigiCGA) was discussed and some research outlooks were given to further promote this valuable research subject.

2. A glance of traditional candidate gene approach

The rationale of candidate gene approach states that a major component of quantitative genetic variation of phenotype under investigation is caused by functional mutation of putative gene. Candidate genes are generally the genes with known biological function directly or indirectly regulating the developmental processes of the investigated traits, which could be confirmed by evaluating the effects of the causative gene variants in an association analysis. Candidate gene approach has been ubiquitously applied for gene-disease research, genetic association studies, biomarker and drug target selection in many organisms from animals to humans [1]. To date, many candidate genes of economic traits or disease resistance/susceptibility were primarily or even repeatedly detected, although the total number of the publicly accepted genes is still absolutely small. Most importantly, candidate gene analysis is usually the indispensable procedure for subsequent positional cloning of QTLs controlling the major genetic variation of interested traits after initial genome scans. In general, significant components of QTLs in a chromosomal region affecting genetic variation of investigated traits are causative genes, so the ultimate pinpoint location of a QTL, with dozens or even hundreds of genes assembled in the about ~20cM confidence interval, to a specific polymorphic gene is inevitably involved in candidate gene analysis. However, candidate gene approach has been criticized owing to low replication of results and its limited ability to include all possible causative genes [1]. Moreover, this approach is by necessity highly subjective in the process of choosing specific candidates from numbers of potential possibilities. The main disadvantage is that it requires the information that comes from the existed well-known physiological, biochemical or functional knowledge such as hormonal regulation, biochemical metabolism pathway and etc., which is generally finite or sometimes not available at all. The actual absence of background knowledge for unscrambling the molecular stories of most complex and quantitative traits has obviously become an information bottleneck to clag its further application, and how to break the information bottleneck is thus one of the most important challenges represented to us.

3. Extended strategies for breaking the information bottleneck

3.1. Position-dependent strategy

Until recently, large efforts have been focused on breaking the restriction of information bottleneck to which the traditional candidate gene approach faces. There are several developed or developing strategies. One is position-dependent strategy. Position-dependent strategy has integrated genome scans and candidate gene analyses, in which the identification of candidate gene is mainly based on the

physical linkage information in a QTL-identified chromosomal segment. Such strategy resulted in the emergence of positional candidate gene approach, the post-genomic version of the positional cloning method. This approach aims at the vicinity of known QTLs, and candidate genes are sought out from tens to hundreds of gene members harbored in the targeted chromosomal region. Some successful applications of position-dependent strategy have already been reported in different fields (including the classical examples of DGAT1 in cattle, GDF8 in sheep and IGF2 in swine) [2-10]. Using this strategy, a recent study has testified that a single-nucleotide polymorphism haplotype of IGF1 contributes to the control of body size in dogs [11]. In general, a combination of linkage studies and candidate gene analyses for promising chromosomal regions is a straightforward strategy, and of which the unifying can effectively improve the hitting accuracy [12].

However, the successful map-based positional cloning was mainly involved in the genes that are responsible for Mendelian traits with discrete phenotypic differences, while the studies that have attempted to identify the positionally causative genes responsible for typical quantitative traits have met with limited success. At the same time, many statistically positive genes detected by the gene-trait associations could not be verified to locate in or near to the known QTL region, which also hints that the position-dependent strategy can not always work well. Although there were some successful examples of positional cloning in animals, the pinpoint location of a causative gene or even underlying functional QTL nucleotide in a conserved block is highly challenged. Usually, there is no guarantee that an identified QTL represents a single gene [13] and there are also many false-positive QTLs that directly fail the application of position-dependent strategy. The difficulty of prioritization of positional candidate gene might be resulted in by the low penetrance of multiple contributing genes. Moreover, the commonly used linkage analysis often contains hundreds of genes in the LOD support interval for a QTL. High-density markers in the same region and alternative analytical methods such as linkage disequilibrium analysis can refine the span of confidence interval small enough to be physically mapped, but this reduced map units will still contain tens of genes [14]. Obviously, when applying the position-dependent strategy, it is difficult to prioritize functional candidates harbored in the targeted region, which is frequently scanned through the microsatellites markers. On the one hand, for a single gene consideration, if without combined information about gene position with clues about biological function, it is not ensured that the empirical speculation can hit the true gene in the face of too much interferential information from dozens to hundreds of genes; on the other hand, for multiple gene consideration, it is too time-consuming and expensive to identify all or most of candidates selected from the total genes in the targeted region. Moreover,

once a certain candidate has already been sought out to detect the polymorphisms, e.g., single nucleotide polymorphisms (SNPs), there is choice for individual site or multiple site detection. If individual rather than multiple mutation sites is detected, the really contributing mutation site might be missed when other mutation sites exist but separate from the effect on traits of interest. Unfortunately, individual mutation site detection strategy was commonly used in many actual applications. It is convinced that a pure position-dependent strategy is generally inefficient, and positional cloning of the underlying gene(s) of complex and quantitative traits still has a stumbling block, for which the whole genome association analysis might provide one of the ultimate solutions.

3.2. Comparative genomics strategy

Comparative genomics strategy makes the utility of cross-species approach to identify and characterize the effect of putative candidates. This strategy includes comparative functional genomics strategy and comparative structural genomics strategy, which results in comparative functional candidate gene approach and comparative positional candidate gene approach, respectively. In this strategy, candidate genes may be functionally conserved or structurally homologous genes identified from other related species. Comparative genomics strategy can rapidly work if functionally conserved or structurally homologous genes affecting phenotypic variation of interest have already been confirmed in other species. It is publicly known that animal models generally provide a comparative approach for identification of potential genes susceptible to human diseases [15-18]. It has been proved that comparative genomics strategy is a well-worked strategy on many occasions, e.g., the information from human, mouse, rat and other information-riched species was frequently used to make discovery of candidate genes of economically important traits in livestock [19-22]. In fact, such strategy has been broadly applied in the biological, agricultural and medical sciences [23].

Until nowadays, increasing accumulations of mammalian genomic data make this strategy more convenient. Nevertheless, this strategy has sporadically come up against difficulties in some applications, although it has many advantages [24]. For most complex and quantitative traits, the total number of identified genes in related species is still small, and furthermore, the phenotypic similar trait of different species maybe has a quite different genetic architecture, which could lead that the selected candidate genes have quite different genetic effects in the analytical species. Thus, comparative genomics strategy is occasionally inefficient because of the biological difference from one species to another due to the genetic heterogeneity or evolutionary differentiation.

3.3. Function-dependent strategy

Tracing of gene expression process of the investigated trait in different stages or genetic

background, including signaling pathway, regulatory network and complex genome-wide transcriptional profiles can contribute to a better understanding of the molecular architecture and find out the detailed clues that candidate gene tells. Although functional information from gene knock-out and transgenic animal and cellular models can also provide us with distinct clues about candidate genes responsible for phenotypes of interest, there is little practical information available because of the difficulty of producing gene knock-out and transgenic animals in livestock. In general, important biological features of traits are directly reflected by transcript pattern, and quantitative traits were usually the consequence of the structure of genetic regulatory networks and the parameters that control the dynamics of those networks [25]. The genetic analysis of variation in gene expression would provide valuable models for studying complex and quantitative traits [26]. Considering that environmental factors affecting gene expression process are also mediated with products of specific genes such as heat shock protein [27, 28], both genetic and environmental factors affect phenotypic variation of trait through gene expression process. Apparently, the variations of traits are directly responsible for the variations of transcriptome and proteome rather than the variome of genomic DNAs. The rationale of function-dependent strategy states that those genes responsible for the variation of gene expression process are also responsible for the variation of trait, and the candidate gene governing the major genetic component of trait variation can be mined from the pattern of gene expression profiles. In fact, gene expression profiles are increasingly analyzed in the search for candidate genes. Generally speaking, there are two types of gene expression variation, the inheritable one and the non-inheritable one [29]. The genes directly transferring or decoding the environmental factors inside and outside usually arouse the non-inheritable components of gene expression variation. By contrast, the genes determining the inheritable components of gene expression variation naturally control the inheritable components of phenotypic variation. There have been hundreds of literatures to sustain the aforementioned viewpoints concerning inheritance of gene expression [30-35].

The function-dependent strategy resulted in the functional candidate gene approach, in which a putative candidate gene is the one that could be statistically detected from the genes controlling large components of inheritable gene expression variation. To date, some researchers began to consider or use this approach for seeking candidate genes in different fields. For instances, by using this strategy, functional candidate genes for "eye muscle area" in pigs [36], genetic resistance for *mastitis* in cows [37], cancer, obesity and diabetes in human beings [38], nutrient transformation in cattle [39], responses for anabolic agents in heifers [40], muscle development in bovine fetuses [41] and other candidate genes with causative

allelic variant that may be of biomedical, economic and evolutionary interest were mined in succession.

High-throughput technologies have produced massive expression data that are invaluable for identifying candidate genes associated with traits of specific interest. However, when using the function-dependent strategy, challenge remains. Especially for earlier simple applications of function-dependent strategy, there was a trend for misemploying. In many cases, the differentially expressed genes were directly taken as candidate genes in a nutshell [42, 43], and such hypothesis is usually improper and befall failure [44]. Nowadays it is clear that candidate gene is far beyond differentially expressed gene. In general, there are too many differentially expressed genes presented in the expression process, and, without additional supporting evidence, the aforementioned hypothesis ineluctably meets the following dilemma: the comprehensive identification of all differentially expressed genes is too arduous and expensive to be feasible, while the random identification of single differentially expressed gene could capture the true candidate gene only in a very small probability. At a large extent, candidate genes underlying the large inheritable components of gene expression variation are usually the key genes impacting on the vital cogs between the neighboring developmental phases or key node genes in the topologic structure of gene expression network. The coming systems biology might provide an ultimate understanding of this problem.

3.4. Combined strategy

Every strategy mentioned previously is conditionally effective and not universal. In such circumstances, combined strategy, which combines at least two strategies together to mine candidate genes, has begun to show its onset in some applications. Recently, it is increasingly common to combine genome-wide expression profiles and linkage analysis to search for candidate genes and such newly developed genetical genomic approach originating from function-dependent strategy provides a particularly powerful means to identify candidates underlying complex phenotypic variation of economic importance [45-47]. In chicken, Marek's disease resistance genes were identified through the gene expression differences between disease resistant and susceptible chickens in which microarrays analysis and QTL mapping were jointly used [48]. By investigating the expression pattern of genes harbored in a genomic interval including a known QTL, candidate genes for alcohol preference were identified in a rat model [49]. Weibel et al. (2006) [50] have combined QTL mapping with proteomics approach to discover six candidate genes for longevity. The study that 34 candidate genes in the control of ovariole number were identified from 548 positional candidate genes through linkage associated with microarray analyses in *Drosophila melanogaster* provided another successful application for combined strategy [6]. These

studies mainly provided the successful applications of combination of function-dependent strategy with position-dependent strategy. Any other types of combinations, e.g., combined function-dependent and comparative mapping strategy [51], combined linkage and linkage disequilibrium strategy [52] and combined RNAi-microarray strategy [53], could more effectively work despite few actual applications of other type combined strategies have been reported. It is anticipated that the promising combined strategy would provide a more powerful comprehensive means to solve the problem of information bottleneck because it could congregate the advantages of each single strategy.

Up to date, many candidate genes or linked markers have been identified but few of them have been successfully verified and made an endpoint usage ultimately. It is common phenomenon that candidate genes did not provide accurate and consistent evidence in each gene-trait association analysis. So, the facticity of a primary association necessarily need to be further verified in some feasible way, which, for animals, usually includes validations in more future generations of the same population or other different populations, and even quantitative complementation test [54] or other functional mutation analyses to the site-specific mutation of candidate gene that brings the phenotypic mutant effect. Quantitative complementation test is a validating method for the candidate gene at a QTL, which was designed originally for QTL work in model animal [55] but usually difficult in livestock.

4. Digital candidate gene approach

The most remarkable progress in this field is the emergence of digital candidate gene approach (DigiCGA). DigiCGA, which also named in *silico* candidate gene approach or computer facilitated candidate gene approach, is a novel web resource-based candidate gene identification approach. In this section, we address a recapitulation of DigiCGA concerning its birth background, concept and some other related issues.

4.1. Background and concept

It is well known that the prosperous projects of mammalian genome mapping accelerate researches on the anatomy of molecular architecture of complex and quantitative traits. The completion and development of the animal genome projects have revealed a multitude of potential avenues for identifying candidate genes in which digital approach is an attention-getting one and as such could enable the systematic identification of genes underlying biological traits [56]. Especially, when the advent and development of Biological Ontology (BO) has well established, the digital resources make it possible to identify candidates by some certain principles, e.g., functional similarity [57]. In such circumstance, with increasing accumulations of web resources, DigiCGA emerges and comes into some use in practice.

As a new development of candidate gene

approach, DigiCGA can be defined as an approach that objectively extract, filter, (re)assemble, or (re)analyze all possible resources available derived from the public web databases mainly in accordance with the principles of biological ontology (e.g., anatomy ontology, cell & tissue ontology, developmental ontology, gene ontology, and phenotype & trait ontology) and complex statistical methods to make computational identification of the potential candidate genes of specific interest, which is generally followed a subsequent validation of actual association analysis.

4.2. Classification of existing methods

Up to date, in our opinion, the present reported approaches related with DigiCGAs could be primarily classified as ontology-based identification approach, computation-based identification approach and integrated identification approach (including literature-based meta-analysis).

The ontology-based identification approach is mainly involved in the bioinformatic analyses for *in silico* identification of candidate genes for specific interest in case of the semistructured, structured and controlled vocabularies for systematic annotation of gene functional information from biological ontology sources available through Internet. A typical example of this approach is the prioritization of positional candidate gene by using gene ontology [58]. The computation-based identification approach includes those computational candidate gene identification methods that describe computational framework to prioritize the most likely candidate genes through a variety of web resource-based data sets. There were many statistical algorithms or computational methods, and of which some included data-mining analysis [59], hidden Markov analysis [60], cluster analysis

(similarity-based method) [61], kernel-based data fusion analysis [62], machine learning [63], KNN classification algorithm [64] and others. Tiffin et al. (2006) had compared seven independent computational methods for disease gene identification [65]. The integrated identification approach comprises most of the combined methods for prioritizing candidate genes through more than one avenues available or integration of relevant information from many sources, including converging actual experimental data, web database-based resources (including literature-based resources [66] and biological ontology resources) or the theoretical assembling of molecular features or molecular interaction principles, e.g., gene structure variation, homologs, orthologs, SNPs data, protein-DNA interactions, protein-protein interactions (interactome), molecular module, pathway and gene regulatory network [67-71]. There have been reported many candidate genes prioritized by the integrated identification approach such as pathway and gene ontology combined analysis [72], text- and data-mining integrated method [73], genetic maps and QTL combined analysis [74] and mutome network modeling integrative analysis [75].

Currently, some application software or online tools for prioritizing candidate genes such as GFSST, ENDEAVOUR, POCUS, G2D, SUSPECTS and others have been developed and released to public [57, 76-81] (see Table 1). In addition, a series of software or online tools such as TAMAL, SNPsfinder, SNPselector, QuickSNP, SNP Hunter, SNP-VISTA, CLUSTAG, WCLUSTAG, CASCAD, LS-SNP, QualitySNP, SNP-PHAGE and MAVIANT could be taken as auxiliary tools to redound to the downstream validation steps of DigiCGA [82, 83].

Table 1. Summary of application software and online tools related to digital candidate gene approach

Name	Literature source	Web Site
GeneSeeker	van Driel MA, et al. Nucleic Acids Res. 2005;33:W758-61	http://www.cmbi.ru.nl/GeneSeeker/
GFSST	Zhang P, et al. BMC Bioinformatics. 2006; 7: 135	http://gfsst.nci.nih.gov
Endeavour	Aerts S, et al. Nat Biotechnol. 2006;24:537-44	http://www.esat.kuleuven.be/endeavour
POCUS	Turner FS, et al. Genome Biol. 2003; 4: R75.	http://www.hgu.mrc.ac.uk/Users/Colin.Semple/
G2D	Perez-Iratxeta C, et al. Nucleic Acids Res. 2007;35:W212-6.	http://www.ogic.ca/projects/g2d_2/
SUSPECTS	Adie EA, et al. Bioinformatics. 2006; 22: 773-4.	http://www.genetics.med.ed.ac.uk/suspects/
TOM	Rossi S, et al. Nucleic Acids Res. 2006; 34: W285-92.	http://www-micrel.deis.unibo.it/~tom/
BioMercator	Arcade A, et al. Bioinformatics. 2004; 20: 2324-26	http://moulon.inra.fr/~bioinfo/BioMercator
FunMap	Ma CX, et al. Bioinformatics. 2004; 20: 1808-11.	
GFINDER	Masseroli M, et al. Nucleic Acids Res. 2005;33:W717-23	http://www.bioinformatics.polimi.it/GFINDER/
PROSPECTR	Adie EA, et al. BMC Bioinformatics. 2005;6:55.	http://www.genetics.med.ed.ac.uk/prospectr/
eVOC	Tiffin N, et al. Nucleic Acids Res. 2005; 33:1544-52	
QTL Mixer	Serrano-Fernández P, et al. Bioinformatics. 2005;21:1737-8	http://qtl.pzr.uni-rostock.de/qtlmix.php
DGP	Lopez-Bigas N, Ouzounis CA. Nucleic Acids Res. 2004;32:3108-14	
CoGenT++	Goldovsky L, et al. Bioinformatics. 2005; 21:3806-10	http://cgg.ebi.ac.uk/cogentpp.html
KNN classifier	Xu J, Li Y. Bioinformatics. 2006;22:2800-05	available on request: jianz xu@hotmail.com
SNPs3D	Yue P, et al. BMC Bioinformatics. 2006;7:166	http://www.SNPs3D.org
PhD-SNP	Capriotti E, et al. Bioinformatics. 2006;22:2729-34	http://gpcr.biocomp.unibo.it/cgi/predictors/PhD-SNP/PhD-SNP.cgi

4.3. Outstretched issues

In comparison with traditional candidate gene approach, DigiCGA is a rational inferring rather than empirical speculation. In usual, the technical framework of DigiCGA includes the upstream web resource-based operational procedures and the downstream validation procedures similar with the actual procedures of association analysis in traditional candidate gene approach. Additionally, in order to heighten the veracity of candidate gene identification, DigiCGA would be essentially open to utilize multifarious available information despite the main analytical source is web resource-based. To date, DigiCGA has given positive results in some cases but failed to identify candidate genes in others, and the consummation and in-depth applications of DigiCGA remain large challenges. Currently, it is urgent to establish the theoretical building blocks and the mature framework for common applications of DigiCGA, which would capture the eyeballs of some computational geneticists and bioinformaticians.

From a practical viewpoint, the pursuit of successful application of DigiCGA has still been problematic because the detailed information of molecular architecture with respect to most biological traits in public web databases is still fragmentary, which suggests DigiCGA is still in its infancy. Although the human and other mammalian genome projects have produced a vast magnitude of digital resources including maps, clones, sequences, expression data and phenotypic data, the public databases provide more sequence data rather than functional data. As for most animals, the gene expression data in public web databases is still needed to supplement on large scale. It should be strongly suggested that the authoritative public databases should subdivide specific sub database to accept and offer the more detailed functional resources for mass identifications of candidate genes underlying traits of biomedical, economic and evolutionary importance. Moreover, the mature methodology and easily used tools compatible with this approach are still being under development. There is still a long way to reach the broader applications. For the development of DigiCGA, it is just the beginning but not the end of the story. It is our view that, with the further development of functional genomics and consummations of mature methodologies and tools, DigiCGA would undoubtedly become more important for various fields to address a wide range of biological questions in near future.

5. Conclusion

Although the candidate gene approach is useful for quickly determining the association of a specific genetic variant with phenotype, the proportion of causative genes governing traits of biomedical, economic and evolutionary importance that have been confirmed is still small and consequently, the number in the list of candidate genes is limited. Current methods for solving the problem of information

bottleneck have complemented and consummated the efforts of traditional candidate gene approach in identifying causative genes, in which much progress has been achieved, but there are still lots to be done. Here we generalized the representative methods in order to be able to promote the efficiency for evaluating the gene-phenotype relations. For the future landscape of candidate gene approach, to meet near-complete or complete solutions to current problems, the ultimate development is to integrate traditional mapping data, fine mapping data, cross-species resources, literature resources, bioinformatics resources on the internet, and even high-through genome-wide resources including sequence-based and gene expression data as comprehensively as possible.

Acknowledgements

This work was supported by Key Project of National Basic Research and Developmental Plan (2006CB102105) of China, National High Technology Research and Development Program (2007AA10Z148) of China and Hubei Province Natural Science Creative Team Project (2006ABC008). We also thank Dr. Deng and three anonymous reviewers for their constructive and insightful comments and suggestions to the improvement of this review.

Conflict of interest

The authors have declared that no conflict of interest exists.

References

1. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet.* 2002; 3: 391-397.
2. Fujii J, Otsu K, Zorzato F, et al. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science.* 1991; 253: 448-451.
3. Johnson PL, McEwan JC, Dodds KG, et al. A directed search in the region of GDF8 for quantitative trait loci affecting carcass traits in Texel sheep. *J Anim Sci.* 2005;83:1988-2000.
4. Bellamy R. Identifying genetic susceptibility factors for tuberculosis in Africans: a combined approach using a candidate gene study and a genome-wide screen. *Clin Sci (Lond).* 2000; 98: 245-250.
5. Grisart B, Coppieters W, Farnir F, et al. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res.* 2002; 12: 222-231.
6. Wayne ML, McIntyre LM. Combining mapping and arraying: An approach to candidate gene identification. *Proc Natl Acad Sci USA.* 2002; 99: 14903-14906.
7. Thaller G, Kuhn C, Winter A, et al. DGAT1, a new positional and functional candidate gene for intramuscular fat deposition in cattle. *Anim Genet.* 2003; 34: 354-357.
8. Clop A, Marcq F, Takeda H, et al. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics.* 2006; 38: 813-818.
9. Stratil A, Geldermann H. Analysis of porcine candidate genes from selected QTL regions affecting production traits. *Anim Sci Pap Rep.* 2004; 22: 123-125.
10. Van Laere AS, Nguyen M, Braunschweig M, et al. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature.* 2003; 425:832-836.

11. Sutter NB, Bustamante CD, Chase K, et al. A single IGF1 allele is a major determinant of small size in dogs. *Science*. 2007; 316: 112-115.
12. Lou XY, Ma JZ, Yang MCK, et al. Improvement of mapping accuracy by unifying linkage and association analysis. *Genetics*. 2006; 172: 647-661.
13. Pasyukova EG, Vieira C, Mackay TFC. Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Genetics*. 2000; 156:1129-1146.
14. Ron M, Weller JL. From QTL to QTN identification in livestock - winning by points rather than knock-out: a review. *Anim Genet*. 2007;38:429-439.
15. Moore KJ. Utilization of mouse models in the discovery of human disease genes. *Drug Discov Today*. 1999; 4: 123-128.
16. Young LJ. Oxytocin and vasopressin as candidate genes for psychiatric disorders: lessons from animal models. *Am J Med Genet*. 2001; 105: 53-54.
17. Phillips TJ, Belknap JK, Hitzemann RJ, et al. Harnessing the mouse to unravel the genetics of human disease. *Genes Brain Behav*. 2002; 1: 14-26.
18. Ewart-Toland A, Balmain A. The genetics of cancer susceptibility: from mouse to man. *Toxicol Pathol*. 2004; 32 (Suppl): 26-30.
19. Mosher DS, Quignon P, Bustamante CD, et al. A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genet*. 2007; 3:e79.
20. Smith TP, Showalter AD, Sloop KW, et al. Identification of porcine Lhx3 and SF1 as candidate genes for QTL affecting growth and reproduction traits in swine. *Anim Genet*. 2001; 32: 344-350.
21. Grobet L, Martin LJ, Poncelet D, et al. A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nat Genet*. 1997; 17: 71-74.
22. Rothschild M, Jacobson C, Vaske D, et al. The estrogen receptor locus is associated with a major gene influencing litter size in pigs. *Proc Natl Acad Sci USA*. 1996; 93: 201-205.
23. Harris S, Foord SM. Transgenic gene knock-outs: functional genomics and therapeutic target selection. *Pharmacogenomics*. 2000; 1: 433-443.
24. Rigby RJ, Fernando MM, Vyse TJ. Mice, humans and haplotypes--the hunt for disease genes in SLE. *Rheumatology (Oxford)*. 2006; 45:1062-1067.
25. Frank SA. Genetic variation of polygenic characters and the evolution of genetic degeneracy. *J Evol Biol*. 2003; 16: 138-142.
26. Cheung VG, Spielman RS. The genetics of variation in gene expression. *Nat Genet*. 2002; 32: 522-525.
27. Edwards JL, King WA, Kawarsky SJ, et al. Responsiveness of early embryos to environmental insults: potential protective roles of HSP70 and glutathione. *Theriogenology*. 2001; 55: 209-223.
28. Piano A, Valbonesi P, Fabbri E. Expression of cytoprotective proteins, heat shock protein 70 and metallothioneins, in tissues of *Ostrea edulis* exposed to heat and heavy metals. *Cell Stress Chaperones*. 2004; 9: 134-142.
29. Gibson G, Weir B. The quantitative genetics of transcription. *Trends Genet*. 2005; 21:616-623.
30. Decanini LI, Collins AM, Evans JD. Variation and heritability in immune gene expression by diseased honeybees. *J Hered*. 2007; 98:195-201.
31. Kerr CA, Bunter KL, Seymour R, et al. The heritability of the expression of two stress-regulated gene fragments in pigs. *J Anim Sci*. 2005; 83:1753-1765.
32. Maatz H, Kren V, Pravenec M, et al. Heritability and tissue specificity of expression quantitative trait loci. *PLoS Genet*. 2006; 2:e172.
33. Manly KF, Wang J, Williams RW. Weighting by heritability for detection of quantitative trait loci with microarray estimates of gene expression. *Genome Biol*. 2005;6:R27.
34. Morley M, Molony CM, Weber TM, et al. Genetic analysis of genome-wide variation in human gene expression. *Nature*. 2004; 430: 743-747.
35. Schadt EE, Monks SA, Drake TA, et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature*. 2003; 422: 297-302.
36. Ponsuksili BS, Wimmers K, Schmol F, et al. Porcine ESTs detected by differential display representing possible candidates for the trait "eye muscle area". *J Anim Breed Genet*. 2000; 117: 25-35.
37. Schwerin M, Czernek-Schafer D, Goldammer T, et al. Application of disease-associated differentially expressed genes--mining for functional candidate genes for mastitis resistance in cattle. *Genet Sel Evol*. 2003; 35 (Suppl 1): S19-34.
38. Kaput J, Klein KG, Reyes EJ, et al. Identification of genes contributing to the obese yellow Avy phenotype: caloric restriction, genotype, diet x genotype interactions. *Physiol Genomics*. 2004; 18: 316-324.
39. Schwerin M, Kuehn C, Wimmers S, et al. Trait-associated expressed hepatic and intestine genes in cattle of different metabolic type--putative functional candidates for nutrient utilization. *J Anim Breed Genet*. 2006; 123:307-314.
40. Reiter M, Walf VM, Christians A, et al. Modification of mRNA expression after treatment with anabolic agents and the usefulness for gene expression-biomarkers. *Anal Chim Acta*. 2007; 586:73-81.
41. Crosier AE, Farin CE, Rodriguez KF, et al. Development of skeletal muscle and expression of candidate genes in bovine fetuses from embryos produced in vivo or in vitro. *Biol Reprod*. 2002; 67:401-408.
42. Lee SJ, Cicila GT. Functional genomics in rat models of hypertension: using differential expression and congenic strains to identify and evaluate candidate genes. *Crit Rev Eukaryot Gene Expr*. 2002; 12: 297-316.
43. Lee SJ, Liu J, Qi N, et al. Use of a panel of congenic strains to evaluate differentially expressed genes as candidate genes for blood pressure quantitative trait loci. *Hypertens Res*. 2003; 26: 75-87.
44. Okuda T, Sumiya T, Mizutani K, et al. Analyses of differential gene expression in genetic hypertensive rats by microarray. *Hypertens Res*. 2002; 25: 249-255.
45. de Koning DJ, Cabrera CP, Haley CS. Genetical genomics: combining gene expression with marker genotypes in poultry. *Poult Sci*. 2007;86:1501-1509.
46. Kadarmideen HN, von Rohr P, Janss LL. From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. *Mamm Genome*. 2006; 17:548-564.
47. Hubner N, Wallace CA, Zimdahl H, et al. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet*. 2005; 37: 243-253.
48. Liu HC, Cheng HH, Tirunagaru V, et al. A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Anim Genet*. 2001; 32: 351-359.
49. Walker JR, Su AI, Self DW, et al. Applications of a rat multiple tissue gene expression data set. *Genome Res*. 2004; 14: 742-749.
50. Weibel J, Sorensen MD, Kristensen P. Identification of genes involved in healthy aging and longevity. *Ann N Y Acad Sci*. 2006; 1067: 317-322.
51. Ron M, Israeli G, Seroussi E, et al. Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle. *BMC Genomics*. 2007; 8:183.
52. Olsen HG, Nilsen H, Hayes B, et al. Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk

- composition of dairy cattle. *BMC Genet.* 2007; 8:32.
53. Weissschuh N, Alavi MV, Bonin M, et al. Identification of genes that are linked with optineurin expression using a combined RNAi-microarray approach. *Exp Eye Res.* 2007; 85: 450-461.
 54. Fanara JJ, Robinson KO, Rollmann SM, et al. Vanaso is a candidate quantitative trait gene for *Drosophila* olfactory behavior. *Genetics.* 2002;162:1321-1328.
 55. Long AD, Mullaney SL, Mackay TFC, et al. Genetic interactions between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number in *Drosophila melanogaster*. *Genetics.* 1996; 144:1497-1510.
 56. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. *Science.* 2002; 298: 2345-2349.
 57. Zhang P, Zhang J, Sheng H, et al. Gene functional similarity search tool (GFSST). *BMC Bioinformatics.* 2006; 7: 135.
 58. Harhay GP, Keele JW. Positional candidate gene selection from livestock EST databases using Gene Ontology. *Bioinformatics.* 2003; 19: 249-255.
 59. Perez-Iratxeta C, Bork P, Andrade MA. Association of genes to genetically inherited diseases using data mining. *Nature Genet.* 2002; 31: 316-319.
 60. Pellegrini-Calace M, Tramontano A. Identification of a novel putative mitogen-activated kinase cascade on human chromosome 21 by computational approaches. *Bioinformatics.* 2006; 22:775-778.
 61. Freudenberg J, Propping P. A similarity-based method for genome-wide prediction of disease-relevant human genes. *Bioinformatics.* 2002;18 (Suppl 2):S110-115.
 62. De Bie T, Tranchevent LC, van Oeffelen LM, et al. Kernel-based data fusion for gene prioritization. *Bioinformatics.* 2007; 23: i125-132.
 63. Adie EA, Adams RR, Evans KL, et al. Speeding disease gene discovery by sequence based candidate prioritization. *BMC Bioinformatics.* 2005; 6:55.
 64. Xu J, Li Y. Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics.* 2006; 22:2800-2805.
 65. Tiffin N, Adie E, Turner F, et al. Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. *Nucleic Acids Res.* 2006; 34: 3067-3081.
 66. Hristovski D, Peterlin B, Mitchell JA, et al. Using literature-based discovery to identify disease candidate genes. *Int J Med Inform.* 2005; 74: 289-298.
 67. Sugaya N, Ikeda K, Tashiro T, et al. An integrative in silico approach for discovering candidates for drug-targetable protein-protein interactions in interactome data. *BMC Pharmacol.* 2007;7:10
 68. Franke L, Bakel H, Fokkens L, et al. Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *Am J Hum Genet.* 2006; 78: 1011-1025.
 69. Rossi S, Masotti D, Nardini C, et al. TOM: a web-based integrated approach for identification of candidate disease genes. *Nucleic Acids Res.* 2006; 34(Web Server issue): W285-W292.
 70. George RA, Liu JY, Feng LL, et al. Analysis of protein sequence and interaction data for candidate disease gene prediction. *Nucleic Acids Res.* 2006; 34:e130.
 71. Yonan AL, Palmer AA, Smith KC, et al. Bioinformatic analysis of autism positional candidate genes using biological databases and computational gene network prediction. *Genes Brain Behav.* 2003;2:303-320.
 72. Feng Z, Davis DP, Sásik R, et al. Pathway and gene ontology based analysis of gene expression in a rat model of cerebral ischemic tolerance. *Brain Res.* 2007; 1177: 103-123.
 73. Tiffin N, Kelso JF, Powell AR, et al. Integration of text- and data-mining using ontologies successfully selects disease gene candidates. *Nucleic Acids Res.* 2005; 33: 1544-1552.
 74. Arcade A, Labourdette A, Falque M, et al. BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics.* 2004; 20: 2324-2326.
 75. Hernández P, Solé X, Valls J, Moreno V, Capellá G, Urruticoechea A, Pujana MA. Integrative analysis of a cancer somatic mutome. *Mol Cancer.* 2007; 6: 13.
 76. Aerts S, Lambrechts D, Maity S, et al. Gene prioritization through genomic data fusion. *Nat Biotechnol.* 2006; 24: 537-544.
 77. Turner FS, Clutterbuck DR, Semple CA. POCUS: mining genomic sequence annotation to predict disease genes. *Genome Biol.* 2003; 4: R75.
 78. Perez-Iratxeta C, Wjst M, Bork P, et al. G2D: a tool for mining genes associated with disease. *BMC Genet.* 2005; 6: 45.
 79. Adie EA, Adams RR, Evans KL, et al. SUSPECTS: enabling fast and effective prioritization of positional candidates. *Bioinformatics.* 2006; 22: 773-774.
 80. van Driel MA, Cuelenaere K, Kemmeren PP, et al. A new web-based data mining tool for the identification of candidate genes for human genetic disorders. *Eur J Hum Genet.* 2003; 11: 57-63.
 81. Ma CX, Wu R, Casella G. FunMap: functional mapping of complex traits. *Bioinformatics.* 2004; 20: 1808-1811.
 82. Hemminger BM, Saelim B, Sullivan PF. TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits. *Bioinformatics.* 2006; 22:626-627.
 83. Xu H, Gregory SG, Hauser ER, et al. SNPselector: a web tool for selecting SNPs for genetic association studies. *Bioinformatics.* 2005; 21:4181-4186.