The past five years have seen vibrant and rapid developments of methods for studying epistasis in human complex traits/diseases1–6. The wide range of methods spans from conventional regression-based methods to nature-inspired algorithms (FIG 1). Most methods use SNP-based tests for pairwise or high order interactions in GWAS data via either an exhaustive search of all SNP combinations or testing of a reduced, preselected set. In addition, methods have been developed to assess interactions between groups of SNPs (e.g. genes) or functional modules (e.g. pathway or network). One of the key recent achievements during this course is that an exhaustive search for pairwise interactions becomes a routine exercise because the computational barrier has been greatly reduced7–11. Here we provide an overview of the developments since Cordell’s review12 with a focus on genome-wide methods.

Regression based methods. Regression-based methods are commonly used to assess SNP interactions in either diseases or quantitative traits12,13 (BOX 1). In GWAS where billions of pairwise SNP combinations need to be assessed, the primary goal is to identify interacting SNPs from the huge search space. One can use SNP genotype models to test interactions directly by comparing the saturated model including interactions (LS) against the reduced model without (LR) using four degree-of-freedom (df) and thus save computing time in estimating genetic parameters (BOX 1). When concerning only additive effects, the genotype model is reduced to an allelic (i.e. the minor alleles of each SNP) model where the interaction is tested in the same way but using only one df14. Previously using regression-based methods in exhaustive search for pairwise interactions at the genome-wide level was computationally prohibitive12.

Various approaches have been taken to reduce the computational barrier. First, taking advantage of modern computing infrastructure and technologies including clusters of computers equipped with multiple CPU cores and/or graphic processing units (GPU)8,10,11,15, parallelization7,9,16 and bitwise computing where SNP genotype data are stored in bitwise data structures to achieve great memory efficiency and computing speed7,17,18. Second, applying approximate interaction tests that can be quickly computed without missing any important epistatic SNP pairs18, e.g. F ratio and Kirkwood Superposition Approximation approximate the (LS vs. LR) tests under the Hardy-Weinberg Equilibrium (HWE) assumption for quantitative20 and disease traits18 respectively and can be quickly computed from contingency tables based on SNP genotypes. For convenience, we list some recent applications based on regression and other approaches that can perform fast genome-wide screening of epistasis in GWAS (TABLE 1). Considering varied strengths and weaknesses in these applications, we recommend an extra step of re-examining the screening results in the full regression models (e.g. conditional tests) to avoid false positives or redundant signals9,19,20.

Nevertheless, regression-based approaches are known to suffer from low power in detection of epistasis when applying a genome-wide threshold adjusted for billions of pair-wise tests12,21,22. Big sample sizes are generally required for success, e.g. meaningful contingency table based tests require each cell to have (e.g. > 10) samples23 leading to removal of low frequency SNPs that might tag some causal variants9. Focusing on interactions involving SNPs with important marginal effects may be a practical compromise24,25 in light of limited samples available in most individual GWAS cohorts, because a much less stringent threshold can be applied owing to much reduced multiple tests9,26–29. Using high dense SNPs can make such focused interaction search more fruitful as power is function of interaction effects and LD between the SNP and causal variant at both loci20,23.

LD and haplotype based methods. In disease traits, methods based on the difference of inter-locus associations between cases and controls may be more powerful than the logistic regression mainly because such difference can be tested using 1 df χ2 statistics in contrast to 4 df in regression12,14,30. Intuitively, if a haplotype of two SNPs tagging causal variants of a disease with a higher frequency in cases than in controls, i.e. the inter-locus associations differ cases from controls, it can generate apparent epistasis illustratable in a genotype-phenotype map27 or a contingency table where each joint genotype is a combination of two of the four possible haplotypes31,32. Unfortunately haplotypes are not directly observed in GWAS and require linkage phase of SNP genotypes to be estimated in advance.

Assuming HWE and linkage phase known an LD-based statistics was first used to measure inter-locus associations and indeed had a power gain32, which was recently implemented in SIXPAC to perform fast pairwise genome scans17 but with several issues to be addressed to become more applicable for GWAS: HWE does not always hold14m; phasing SNP genotype data is computational expensive; intra-locus interactions are possible20,33. Correlation-based measures of LD do not require HWE or phasing34 and have been increasingly used in studying epistasis in GWAS14,15,35. For example, for each pair of SNPs one can compute their Pearson’s correlations in cases and controls separately and derive a Z-score statistics based on the difference to test interactions15,34. Unfortunately, the Z-score statistics is still subject to inflated false positives when two SNPs in LD and/or with marginal effects14, which may be cured in the logistic regression model35. Therefore a two-step solution is implemented in a GPU enhanced tool EPIBlASTER15 using Z-score statistics for screening and logistic regression models for testing and thus controls false positives without losing the power advantage in the LD-based methods.

Haplotype-based methods face an additional issue: uncertainty in estimating haplotype frequencies especially when the HWE assumption does not hold14,36u. An improved haplotype-based statistics that adopts correct variance calculation and incorporates a weight average of the joint effects of two SNPs is as powerful as the Z-score statistics34 and can control false positives when only one SNP has marginal effects, but still generates inflated false positive particularly if both SNPs have marginal effects and are in LD14. Arguably such inflation could be viewed as strengthen of the haplotype-based methods in terms of identifying ‘co-associated’ marginal loci without interactions35,37 if they are missed in conventional GWAS. The two-step solution15 may be an effective fix of the inflation issue here.

Bayesian methods. Bayes’ theorem offers a great flexibility to model and stochastically search epistasis without enumerating all SNP combinations12. In BEAM the first Bayesian tool for GWAS38, detection of interacting SNPs is equivalent to partition independent SNPs (i.e. no LD) into predefined groups according to their posterior probabilities without explicitly testing interactions6,12. Improved BEAM methods use new variables to account for LD among SNPs and thus allow a full analysis of GWAS data6,39, but may benefit more from additional tests for interactions among SNPs partitioned in the target group5. Hybrid Bayesian methods appear to be able to improve detection of epistasis in GWAS40–42, e.g. combining the strengths of Bayesian framework and generalized linear model allows fast and stable tests of SNP or haplotype interactions while considering covariates, marginal effects and gene-environment interaction simultaneously41,43. Besides, the Bayesian model averaging approach may increase power of detection by averaging evidence from multiple plausible models given unknown actual interaction types44.

Data filtering methods. A given GWAS data can be filtered to select a subset for interaction tests based on either existing biological knowledge (e.g. databases of pathways and protein-protein interactions4,23,45), or statistical features (e.g. marginal effects23 and SNP genotype frequencies46,47), or fast algorithms17,48–51. Methods based on variance heterogeneity can effectively select potentially interacting SNPs for quantitative traits but could miss interacting SNPs lack of variance heterogeneity52,53. Besides the apparent speed advantage, filtering based methods can be better than exhaustive search in power because of much reduced multiple tests as well as functional interpretation when considering only functional SNPs. However, cautions are recommended when applying filtering because of potential biases (either upwards or downwards) caused by limitations in the algorithms and existing knowledge that may be subject to publication bias4 and specific contexts54,55. Furthermore, it is debatable what threshold is appropriate after filtering as it might alter the NULL distributions56.

Machine learning, data mining and other algorithms. Continuous efforts have been made to adopt/improve algorithms from other disciplines to address the large P small N problem in detecting epistasis2,4,5,12, particularly high-order interactions where regression-based methods3 may suffer from increasingly computational complexity and data sparsity50,57. These algorithms often employ certain classifiers for data reduction and/or feature selection to reduce both the computational and statistical burden in exhaustive search5,58,59, but previously were not scaled up for GWAS or explicitly testing interactions and could miss interactions without marginal effects12,58,59. While most existing algorithms (e.g. Multifactor Dimensionality Reduction, tree-based, entropy-based) being scaled up by modern computing technologies60–64, their classifiers are also improved to be applicable for complex quantitative traits60,65–68, allow using risk scores aggregating multiple interactions in classification72, and account for interactions without marginal effects69,70.

Improved computing efficiency also allows an easy fix of the stability issue of the RelifF-based filtering methods71 by aggregating multiple runs72 and facilitates the development of combining complementary algorithms. For example, using tree-based methods for screening and Multifactor Dimensionality Reduction for interaction testing can improve the overall performance73–75; Reconstructability Analysis method uses entropy-based methods to construct and interpret interaction structures and graph theory heuristics to traverse64. New algorithms such as Ant Colony Optimization mimicking how ant colonies find the shortest route to foods76 are continuously adopted into epistasis studies but need to be clear how interactions are tested. Nevertheless, detection of high-order interactions appears very challenging to machine learning and data mining methods too considering that interactions are not explicitly tested in many cases. When interactions are explicitly tested, e.g. using Reconstructability Analysis, detection of three-way interactions requires multilayer hypothesis tests and thus is as difficult as regression methods3,77, in addition to issues of excessive computing demand, exponentially increased multiple tests and insufficient sample sizes12. Novel ideas are needed to genome-wide detect high-order interactions.

Group and module based methods. Testing interactions based on groups of SNPs or functional modules can dramatically reduce the multiple test burden (e.g. only ~108 pairwise tests required for 20,000 genes) and thus increase power of detection22,78, including high-order interactions79. In addition, properly grouping SNPs may collectively capture casual variants that are not well tagged by individual SNPs80 but could be complicated by a number of factors (e.g. group definitions, correlations among SNPs and SNP pairs)36. A common practice is to use genes to group SNPs and derive gene-based variables factoring in SNP correlations for interaction tests via regression81 or analog LD-based14 (i.e. testing the difference of correlation of a pair of genes between cases and controls) approaches82,83. Alternatively, one can first compute all pairwise interactions between SNPs in each group and derive gene-based interaction P values by integrating the pairwise interaction P values78.

These gene-based studies suggest it is important to incorporate external LD (e.g. via imputation) to capture unobserved causal variants and use a set of signals instead of the single best to detect multiple interactions between a pair of genes. Nonetheless, the gene-based methods implicitly assume no intragenic interactions which may not necessarily hold20,33 and could potentially jeopardize the power. One can also use pre-computed pairwise SNP interactions to build statistical interaction network and aggregate multiple gene-gene interactions and marginal effects using network analysis algorithms such as SNPrank84 to identify functional important pathways2. However, the issues associated with the gene-based approaches also need to be addressed here.

Multi-trait and multi-level integration. Collectively analyzing multiple complementary traits may help detection pleiotropic epistasis for better biological interpretation but remain fundamentally unexplored due to the amounted complexity85,86. Pleiotropic epistasis could be identified in human complex traits by simply looking for SNP-SNP interactions shared across related traits20 or using a gene-based method with a latent variable representing multiple traits87. On the other hand, emerging evidence of epistasis from eQTL and other -omics studies can help understand the underlying molecular mechanisms88,89. However, integration statistical and functional interactions is not straightforward90 and may require system biology approaches91. Further efforts are needed to better integrate interaction signals at both directions.

So far we have briefly discussed only the major method developments in the past five years. Obviously, it is very difficult to recommend the ‘best’ applications without careful comparisons because of huge variation among them, including interaction definitions and NULL distributions56. Several attempts have been made to evaluate different methods previously58,59,92–96, suggesting community-wide efforts97 may be essential to fulfill the task. Furthermore, consensus significance thresholds and standards for replication and reporting have become fundamentally important to improve the overall quality of future epistasis studies including result sharing.