Referee reports from *Nature Reviews Genetics*

Dear Chris, Gib and Wen-Hua

I'm writing to let you know that we’ve received (all three) referee reports for your Review article for *Nature Reviews Genetics*. Many thanks for your patience during the delay with these reports. Acceptance for publication will be based on you addressing the points raised by the referees, which are shown below (my comments are shown in bold). As you will see, the reports are fairly positive overall, in favour of publication of a revised manuscript, but had useful suggestions for improvements.

I have also attached to the covering e-mail the latest version of your manuscript [Haley V2\_DB]. As mentioned in previous email correspondence, before sending to peer review, I carried out some minor editing for readability, which can be seen as tracked changes. A clean version of this manuscript (without changes tracked or comments enabled) was sent to referees. If you view all revisions as balloons, the page and line numbers should match up with those mentioned by referees. Please use this latest manuscript when revising your manuscript, and please consider my editorial suggestions (in comment bubbles in the manuscript) when you address the referee suggestions.

Your figure, as rendered by our art editor is attached. Please send me any corrections or suggestions (either by email or uploaded alongside your revised manuscript).

We would like you to return the revised manuscript by 8th July (two weeks from now). Please let me know as soon as possible if this date is not feasible. To submit your revised manuscript, please click on the link that is shown in the covering e-mail. With your revised version, please submit an annotated version of the referees’ reports, in which you detail how you have addressed each comment; if there are any changes that you decide not to make, please include your reasons.

Other comments/requests:

1) Review length and number of references. When making the necessary revisions, please could I ask you to keep in mind both the length of your article and the number of references in your reference list? Please keep within **5,700** words and **160** references; the version of your article that was sent to referees was **5,454** words with **145** references. The word count is for the main text, that is, excluding the abstract, legends, display items and references. If you are having trouble meeting the word limit, please let me know.

**Response:** We have revised the text in places and added some new references. We think we are still within the limits!

2) Reference format: Please ensure that references are in the correct format, which should be straightforward using your reference management software. This looks fine in your latest version, so please make sure this is maintained in your resubmission. Note, however, that referees spotted duplicated references, so please make sure that is tidied up in the resubmission. For your information, the correct format is as follows:

1. Author, A. B. & Author, B. C. Title of the article. *Nature Cell Biol.* **6**, 123–131 (2001).

If there are six or more authors to a reference, only the first author should be listed followed by ‘*et al.*’.

Please ensure that references are cited sequentially in the following order: main text, boxes, figure legends and then tables.

**Response:** Done

3) Highlighted references: We ask that you select the top 5–10% of references in your list that those entering the field must read and write a single sentence in bold beneath each one to explain the significance of the work.

**Response:** Done

4) Copyright: If any of the display items in your manuscript (for example - figures, tables, images, videos or text boxes) include images that are the same as (or are

adaptions of) images that have previously been published elsewhere, please fill

in the Third Party Rights Table [<http://www.nature.com/licenceforms/npg/thirdpartyrights-table.doc>] and return to us when you resubmit your revised manuscript. This information will enable us to obtain the necessary rights to re-use such previously published material. If we are unable to obtain the necessary rights to use or adapt any of the material that you wish to use in the display items in your manuscript, we will contact you to discuss the sourcing of alternative images. For these display items that have been previously published, please acknowledge the source in the figure legends, and that the reference is included in the reference list.

**Response:** Not applicable

5) Biographies: We need from you brief (literally just 100 words or so) biographies for you and your coauthors, detailing your research career and the interests of your lab. This will be an online-only feature of the journal. Please see any of our previous articles for examples of typical formats.

**Response:** Done

6) Online summary: We need from you a summary, which is an online-only feature that should comprise a bullet-pointed list of the contents of the article (~6 points, each 1–2 sentences long). The summary should provide the reader with a quick overview of the content, and should also act as a reminder once the article has been read. Again, please see any of our previous Review articles for examples of typical formats.

**Response:** Done

7)Nomenclature: Please be aware that we use standard gene and protein nomenclature. These changes will be made at the copy-editing stage, but it would help us if any relevant changes are made now. This will help to avoid disruption at the proof stage. Please consult the appropriate nomenclature databases for standard gene names and symbols. A useful resource is Entrez Gene.

**Response:** Done

8) Glossary: If you notice any additional specialist terms when revising your manuscript, please add glossary definitions to the existing list at the end of your manuscript.

**Response:** Done

9) Weblinks: if you would like to include weblinks to useful databases or resources that are relevant to the article, please paste the URLs at the end of your manuscript. Additionally, if you would like to include weblinks to your laboratories, please paste the URL into your biographies.

**Response:** Done.

Thanks again for writing for *Nature Reviews Genetics*. I look forward to receiving your revised manuscript and to seeing the article in print. Please feel free to contact me if you wish to discuss any of the points raised in this letter or by the referees.

Best wishes,

Darren

**Darren Burgess, Ph.D.  
Associate Editor  
*Nature Reviews Genetics***

**4 Crinan Street, London N1 9XW**[d.burgess@nature.com](mailto:d.burgess@nature.com)**[Au: please address all referee requests, unless indicated otherwise. I have added my additional comments in bold]**

**Referee report: Referee #1**

Immediate interest: Yes

Logically presented: Yes

Balance: Yes

Fair discussion: Yes

Figures clear: Yes

References cited: Yes

(Remarks to the Author)

Minor comments:

The "boxes" seem to be sometimes referenced in small letters (e.g. "Box 2" on page 3 line 5) and sometimes in capitals (e.g. "BOX 3" on page 3 line 13).

**[Au: this will automatically be standardized and hyperlinked during the production stages, so this is not important to fix]**

Reference 101 (which is incomplete) seems to be the same as reference 119.

**Response:** We have resolved this.

Page 2 line 40: "whose existence cannot be verified empirically due to technical limitations". This is a bit confusing - what technical limitations are you talking about? This statement is not really enlightened by reading Box 1 - can you explain it better either here or in one of the boxes?

**Response:** We have expanded on this in Box 1.

Page 4 lines 7-8: You say that an exhaustive search for pairwise interactions at the genome-wide level was considered computationally prohibitive, referencing Cordell 2009 (Ref 13),

but in fact this statement is not supported by reference 13 - reference 13 actually states "Exhaustive search of all two-locus interactions from a genome-scan is time-consuming but computationally feasible."

**Response:** We agree we mis-stated this and hence have changed “prohibitive” to “time consuming” in our text.

Page 5 line 5-6: You state that a collection of statistics (Suppl Inf Box S1) has been derived to measure the inter-locus association differences between cases and controls, but actually this is only covered in the last two of the 4 sections of Suppl Inf Box S1. Could be clarified. **[Au: I’m not sure that there is a way to easily clarify this without making the mention of the suppl box too cumbersome in the main text]**

**Response:** We moved ‘(Supplementary information box S1)’ to the end of the sentence and simplified the sentence to highlight that the collection of statistics were assessed in the paper cited as ref 15 and further details could be found in the supplementary. This summary sentence is also clarified by examples later in the paragraph.

Page 5 3-4 lines from bottom: I was not aware that BEAM was based on any kind of rationale to do with the difference of inter-locus genotype frequency distributions between cases and controls - I thought BEAM fitted a standard logistic regression type model? Of course you could say that a logistic regression model is actually to do with the difference of inter-locus genotype frequency distributions between cases and controls - but then that means this statement would apply to regression models too, whereas in the previous 2 paragraphs you seem to be making a distinction between regression methods and methods that directly consider genotype frequency distributions between cases and controls? Could be clarified.

**Response:** The rationale was quoted in the review of Bayesian models by the BEAM authors (ref 7, page 184) that reads: “The basic rationale behind the BEAM model is that, if some SNPs are associated with the disease, the distribution of their genotypes (or alleles) should be different between cases and controls, otherwise there is no evidence of disease association at those SNPs. To distinguish between interactive and marginal associations of multiple SNPs, the BEAM model defines a set of SNPs to be interactive if the joint distribution of these SNPs fits the data better than the independence model (i.e., the product of their respective marginal distributions).” Hence we have not altered our text.

Page 6 lines 20-36: It might be worth referencing the following paper in this section. In particular this paper addresses the issue of significance thresholds at the second stage only incurring the multiple testing correction for the number of tests performed at the second stage (as the second stage test is independent of the first stage):

Efficient Two-Step Testing of Gene-Gene Interactions in Genome-Wide Association Studies

1. Juan Pablo Lewinger1,\*,

2. John L. Morrison1,

3. Duncan C. Thomas1,

4. Cassandra E. Murcray1,2,†,

5. David V. Conti1,

6. Dalin Li1,3,4,† and

7. W. James Gauderman1

Genetic Epidemiology

Volume 37, Issue 5, pages 440-451, July 2013

**Response:** We have included this reference.

Page 13 line 35: "searching for epistatic effects might be one way to uncover loci with additive variance with more power..." - is it really relevant that the loci have (any) additive variance? Is it not the larger epistatic variance that is increasing the power? Would you be better just to remove the words "with additive variance" here?

**Response:** The “missing heritability” in essence relates to additive genetic variance. Here we are making the point that searching for epistasis may help capture more of the true additive genetic variance. Hence we have left unaltered.

Page 17 (Box 4) line 23: Can you explain your rationale for recommending 1x10-13 as the appropriate threshold? The following paper rather suggested 1.0×10−12:

Significance Levels in Genome-Wide Interaction Analysis (GWIA)

1. Tim Becker1,2,\*,

2. Christine Herold1,

3. Christian Meesters1,

4. Manuel Mattheisen1,3,4 and

5. Max P. Baur1,2

Annals of Human Genetics

Volume 75, Issue 1, pages 29-35, January 2011

**Response:** We have added this reference which derives the 1 x 10-12 threshold for a specific 500k SNP array rather than a dense map where the number of independently segregating regions is germane, as we now clarify in the text of the box.

Page 20 glossary); Bit confusing to have 3 terms included in one definition? E.g. "observed scale" and "liability scale" appear within the definition for "binary phenotype", rather than being defined separately. This also occurs in a few other definitions within the glossary.

**[Au: note that it is a journal restriction that each glossary definition must be standalone (i.e. every glossary term must have its own definition). Apologies that I didn’t spot that earlier. Where relevant, the definitions can mention other terms, but these other terms must have their own, dedicated definitions separately]**

**Response:** We have now made these stand-alone as requested.

Page 22 line 25: "gen" should read "genetic"

**Response:** Done.

Suppl Box S1 line 3: "to differ the" should read "to differ from the"

**Response:** Done.

Suppl Box S1 2nd table: It is not clear what the unit of analysis is in this allelic table. Are you assuming that haplotypes have been measured and can be determined with certainty? I recommend you remove this table and instead just use the usual logistic regression equation (with genotypes coded additively) to illustrate the concept of a 1df test of interaction in an allelic model.

**Response:** We have made the suggested modifications..

**Referee report: Referee #2**

Immediate interest: Yes

Logically presented: Yes

Balance: Yes

Fair discussion: Yes

Figures clear: Yes

References cited: Yes

(Remarks to the Author)

The manuscripts reviews the topic of SNP-SNP interaction. It is timely as a lot have happened in the field since the last comprehensive review by Cordell, and specifically, genomewide scans for such interactions have shifted from the realm of computationally prohibitive tasks to the grey zone, where one accomplish such scans given sufficient algorithmic sophistication and computational infrastructure.

Major comments:

1. In my mind, a good review, in contrast to a lit. survey, is one with a message that emerges from the body of work included. I was missing that here.

**Response:** We have clarified the major messages and included them in the online summary.

2. I found Box 2 going off-topic describing polygenic SNP scores that have no parallel that I know of in epistatic analysis. It is further complicated by the vague definition of "highly" vs. "lowly" heritable, and perhaps a table with two or three representative phenotypes would clarify. Otherwise, it is confusing: Sure, if h2=0.9, epistatic contribution to prediction is tightly bounded, but how do you call a trait with h2=0.6? Finally, the conclusion from Box 2 paragraph between pages 12 and 13) ignores the quadratic, cubic etc. number of potential interaction signals, that together may provide substantial predictive value even if none individually does.

**Response:** We have modified the text of Box 2 slightly to clarify our point but we feel that the general point about the difficulty and potential value of the inclusion of two-way interactions stands and applies even more to higher order interactions.

Minor comments:

3 The overall organization of the manuscript is compelling,and I liked the attempt to visually classify methodologies (fig 1). The presentation can be improved if each bode/level of the hierarchy is annotated by the question it would answer: "What is the search space considered?" or "How is a particular interaction evaluated?". I am not sure the hierarchical representation is appropriate, given that one can apply any of several filtering criteria to any of several search spaces, scan them with any of several algorithms and then score interactions with any of several statistical models. Perhaps a pipeline representation would be better suited

**Response:** We appreciate the referee’s comment and suggestions. We have tried very hard to find a better presentation but feel that the one submitted is the simplest layout of the overall picture of recent method developments capturing the great complexity, high diversity, and their interlinks. Indeed, a number of methods used mixed algorithms which however are all covered in the picture. A pipeline representation is possible for certain method types but hard for everything else.

4. Literature that should also be covered: A method by Elston, and recent review by Ritchie (including some references within)

**Response:** We have included these references..

5. Table 1 should include reference to relevant publications

**Response:** Done.

6. Some of the current supplementary material should be in main text boxes.**[Au: we are already at almost the size limit, so it won’t be possible to import any substantial text into the (non-supplementary) boxes or main text]**

**Response:** Given the editor’s comment, we have not implemented the referee’s suggestion.

7. Technically, it is important to communicate that interaction can appear or disappear due to a non-linear transformation of the phenotype value.

**Response:** We had already pointed out that epistasis should not be scale-dependant, and we have now emphasised this point in Box 4.

8. The world of somatic mutations in cancer reports a lot of what germ line geneticsists would call epistatic interactions

**Response:** We feel that the role of functional epistasis in evolution in populations or in tumours is not directly relevant to the issue of statistical epistasis maintained in segregating populations and so have not modified our text.

9. The field of group-based tests has seen significant interest due to rare alleles, that should be mentioned (with reference to an appropriate separate review).

**Response:** We have discussed group-based interaction tests in the ‘Group- and module-based methods’ section and highlighted potential importance of intra-locus interactions that may be related to rare/hidden causal variants. We are not aware of a review specific on testing interactions involving rare variants.

10. Pleiotropic association analysis is a field with an orthogonal set of complexities. When discussing pleiotropic epistasis, the authors should refer to (a review detailing) those,

**Response:** We totally agree that pleiotropy is another difficult field. Pleitropic epistasis appears to be even harder and was not covered at all in a recent review of pleiotropy (Solovieff et al. Nat Rev Genet 14, 483 – 95, 2013) so we decided to skip such a review to focus on epistasis.

11. Citations 101, 119 are the same

**Response:** Done.

12. Please number your sections and subsections in a nested fashion. It is annoying to rely on heading font size to navigate the document. **[Au: we can’t accommodate numbered headings in journal style. Your current headings and subheadings are fine]**

**Response:** No changes made.

**Referee report: Referee #3**

Immediate interest: Yes

Logically presented: Yes

Balance: Yes

Fair discussion: Yes

Figures clear: Yes

References cited: Yes

(Remarks to the Author)

Wei et al. provide a truly excellent review of a) recent methods and b) emerging results in studies of epistasis in human complex traits.

The review covers a very large body of material yet manages to organise the presentation in a logical and eminently digestible manner. The material is also sufficiently distinct from previous

reviews.

The authors strike an appropriately cautious tone when considering the 'big picture' of what has been learned (i.e. the 'cold hard look'). The first section provides an up-to-date reference list for investigators interested in using these methods.

Pretty much all the relevant conceptual and statistical issues seem to be covered. Of the few things I felt were missing whilst reading through, most were subsequently fleshed out in the Boxes (e.g. the greater impact of LD on non-additive components of variance, the role of family studies to delimit components of genetic architecture, etc).

Below are only some relatively minor comments that the authors may wish to consider. None are particularly critical to address.

1) The initial contrast between functional and statistical interaction, whilst a necessary note, is perhaps likely to confuse many readers unless followed up more fully. Perhaps room for another Box item? I recognize this issue has been clearly dealt with in other similar articles, but particularly for biologists coming to this manuscript, it still may be worth a fuller treatment (i.e. connecting it with the issues of scaling effects mentioned later, etc) as it still seems to cause an inordinate amount of confusion.**[Au:I’ll leave this to your judgement. I think that this issue is addressed enough, and you point to where readers can get more information. We can accommodate 1 extra display item (up to a maximum of 7 print display items) if you want to develop this]**

**Response:** We agree with the editor that this is sufficiently dealt with in the text and the references therein and so have made no changes.

2) I may have missed it, but I think the case-only approach (very standard in epidemiological studies of interaction) is not covered. In many ways, it is ideally suited to screening for epistasis, at least in homogeneous populations between unlinked loci.... Think it should at least be mentioned as another analytic option (with its own type 1/type 2 trade-off).

**Response:** The case-only approach may be ideal for epidemiological studies but is not necessarily ideal for analysing epistasis in GWAS. While all LD- and haplotype-based methods are subject to the HWE assumption, as showed in recent simulation studies, case-only statistics are more likely to generate false positives when LD is present between a pair of SNPs, which is not uncommon in GWAS situations. We have added these points in the supplementary Box S1.

3) In discussing some of the machine learning/ AI approaches: it is mentioned that imputed/dosage data cannot necessarily be handled easily. Do many of these methods also suffer from problems of missing data (i.e. that may become severe when considering combinations of genotypes beyond pairs).

**Response:** We assume the referee is asking about empty cells in the contingency tables due to insufficient samples per joint genotype (as missing genotypes can be imputed). Indeed, contingency table based methods would suffer from the issue, which can be eased (but not cured) if raw data are available to combine in a meta-analysis. We have added this point in the main text (lines 33 to 36, page 9) and supplementary Box S1.

4) "Arising from these strategies are some examples of epistasis with strong statistical support (Box 4)..." After reading this, I was hoping that Box 4 would actually contain a list of these "strong" findings, tabulating the disease/QT, the type of interaction effect, the study design, allele frequencies, genes, etc. Box 4 as it stands is great (and certainly should remain), but I thought the former also could be informative (if indeed the authors wish to 'put their money where their mouth is', so to speak, and claim that these results are worthy of such highlighting). **[Au: there is room in this box for a small embedded table if you choose]**

**Response:** Our conclusion, now included at the start of this box, is that we know of no studies that fully meet these criteria.

5) Just to note that the discussion of scale effects (p12) is

excellent, especially noting that issues can arise when one of the loci has a very large effect.

6) Perhaps a brief treatment of, or connection to, intra-locus

non-additivity for complex traits (i.e. dominant/recessive gene action at the single locus level): the extent to which GWAS have or have not found evidence for such effects as a simpler model for what we might expect for epistasis (and with conceptually similar issues, e.g. relating to reduced power, LD, etc, but only many levels a far, far simpler problem of course).

**Response:** We have covered intra-locus interactions p6 lines 22-23 and p9 lines 1-2 and think that this is sufficient.

7) When describing some of the patterns of epistasis (p11), is it

worth relating to the classical terms (i.e. duplicate and complementary gene action, etc) where appropriate?

**Response:** We considered this but decided it would add more confusion than illumination. The named types of epistasis are really only relevant to major locus segregation in the case of two locus interactions and even then do not cover all types of interaction.

8) One thing I would have liked more treatment of, following from the discussion on p11, is the type of interaction model we might expect. Are certain methods better at detecting different types of interaction models? Should variable priors be put on different classes of interaction, for example, if we think that cross-over interactions are a priori less likely to be biologically plausible, and/or that multiplicative interactions may be more likely to represent scale-effects? Do any methods focus on the type of 'qualitative' interaction of no-effect versus some-effect, enhancing power for that class by discounting other types of interaction? Would it conceptually be desirable to prioritize statistical interaction tests on the nature of interaction over-and-above the p-value? (Most of the

top interaction results arising from false positives will be extreme cross-over type interactions.) In general, is a more model-constrained type of epistasis search warranted and/or implemented in available methods?

**Response:** The issue of epistasis patterns has been well addressed before. We have briefly mentioned Bayesian model averaging approach in the main text given these patterns unknown in advance (page 7 lines 15-17). Other questions raised by the referee are interesting but are unlikely to have a simple answer in our opinion so it would be better leave them to future research.

9) Typo on p6. "...will suffer from increasingly computational

complexity..."

**Response:** Done.