Reviewer 1 Comments for the Author...

Work by Guo et al. used a gene-based recessive diplotype screening approach to make an exome-wide comparison between controls and the patients with ICD-9 defined hemochromatosis in a cohort housed by the Marshfield Clinic Research Institute. Their analysis directed them to FGF6, which was previously unidentified in iron metabolism, and subsequently validated by the authors to have a functional role through various experiments. The paper was put together coherently and organized well. The writing is good quality. The overall scope of the paper was well stated. The figures and tables were labeled efficiently. The codes used in the analysis were uploaded and made available in Github, which was quite helpful. However, when the computational methodology is concerned, there’re certain details that can’t be overlooked and need to be more clearly stated, especially since the whole premise of the paper relies on how it was carried out. Here’re some questions and comments to the authors to clarify a few important points in the manuscript:

1/ The information and the utility of the cohort were very well stated. In concordance with such study data description, it’d be advisable to include statements about the data collection standards (Declaration of Helsinki principles…), informed consent, protocol review etc.

2/ Individual level filtering was done by matching the ancestry using Principal Component Analysis (PCA) as well as ICD-9 diagnostic codes:

2.1/ For readers who aren’t familiar with how an ancestry match was done, the authors should explicitly mention which reference populations were used, how the outliers were removed using the PCA and considering which ethnicity. Also, it’s usually advised to spell out any abbreviations used for the first time in the text (i.e. PCA).

HRC r1.1 2016

2.2/ What’s the extent of genotype-level filtering done on the individuals? These include: genotyping call rates, any relatedness measures (i.e. kinship coefficient) and sex-match between what was reported and found through genotyping.

Response: We thanks the comments of the reviewer. Genotyping-level filtering was conducted with default plink –-geno setting (missing rate<10%). We applied IBD to measure the relatedness with the pi\_hat <0.2. We didn’t apply sex-match between case and control since evidence shown hereditary hemochromatosis sees no obvious differences between men and women.

2.3/ Do the final numbers for cases and controls, 18 and 6,896, respectively, reflect these measures?

Response: We thanks the comments of the reviewer. Yes, all the filter and data processing is conducted both in case and control as the same way. Actually, before the association, we didn’t separate the data and all the data processing is case-control independently except the compound heterozigosity test.

3/ While the initial number of variants was stated as being ‘over 500,000’, the subsequent steps of quality control (i.e. genotype call rates, deviation from Hardy-Weinberg Equilibrium (HWE)) were just mentioned in the passing, and how these affected the final distribution of the variants based on frequency (i.e. how many is <= 1% (rare variants), 1-10% (moderately common variants) and >= 10% (common variants)) wasn’t mentioned. Can the authors include this information within text? This would be useful to get a sense on how many rare variants there were left before going into phasing as well as a the final number of variants used in the analysis.

Response: We thanks the comments of the reviewer. We have 540,101 initial SNPs and 10,861 (2% of total SNPs) variance were filter out (HWE <10-5). Since it is exom-array, the rate of rare allele usually higher than GWAS array and we found 47.8% probes is rare variants, 44.1% probes are common alleles while 8.1% are moderately common variants.

4/ How was the annotation done? How many variants were left after filtering for the functional variants (i.e. missense, nonsense, 5’UTR, 3’UTR, splice sites and “frameshift”)? Were the missense variants annotated for their pathogenicity and also filtered in/out as such?

Response: We thanks the comments of the reviewer. We applied the ANNOVAR to annotate the variants. We collected 13 different predictions including phaseConsElements46way, Polyphen2\_HDIV, Polyphen2\_HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, PROVEAN, VEST3, MetaLR, M-CAP, CADD and GWAS significant hits. We require at least two “D” preditions to define pathogenic variants. Yes. We define pathogenicity exactly same as the above definition.

5/ Has the unbalance between the number of cases and controls addressed? A lot of times, including rare variants in association studies when there’s such an imbalance would result in increased rates of false positives.

Response: We thanks the comments of the reviewer. We agree with the reviewer. Increased rates of false positive is always the side effect of increase the power. Therefore, we conducted the cellular functional screen in the result 2 and 3 section. We tried to knockout and increase the gene expression of *FGF6* and check the changes of iron metabolism related gene expression, especially hepcidin (*HAMP*, key regulator of the entry of [iron](https://en.wikipedia.org/wiki/Iron) into the circulation in mammals) and we also detected the iron deposition of different cells with the projection of FGF6 with different alleles.

5.1/ It’d be a good check to look at and compare the mutation burden between the cases and controls using maybe a gene-burden test. Would the findings be similar?

Response: We thanks the comments of the reviewer. Yes, we tried SKAT test with rvtest (Xuewei Zhan, [Bioinformatics](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4848408/). 2016) and we found the association is still significant with P-value=3.86x10-5 as well as permutation P-value=1.0x10-4.

6/ The most important part of the analysis revolves around the details surrounding the haplotype phasing step in the presence of both common and rare variants, which were not sufficiently explained. Here, the purpose of phasing is to ascertain compound heterozygosity from the genotyping data. Given a string of heterozygous calls, we actually don’t know which are in phase and anti-phase. The ultimate proof would be to experimentally show that the adjacent alleles are in phase by long-range PCR amplicon sequencing or single molecule long-read sequencing. Statistical phasing only works well for common variants for which one can tap into common haplotypes in a reference database. However, for rare and ultra-rare variants, statistical phasing is not as reliable. It is not clear in the manuscript which database or reference was used for phasing and how strong the statistical support was. A great deal depends on how large the reference sample is (that has been typed on the same platform) and whether the rare variants had many observations to do credible phasing. If phasing is not reliable, then the statistical uncertainty or the so-called switching error ought to be documented.

The argument about how in a string of k variable sites, how it is unlikely all rare variants are on one side, needs to be examined in a case-by-case basis. Because if the cases have more suspected heterozygosity because they just have more rare variants; then the hypothesis of rare variant burden difference between cases and controls need to be revisited. Therefore, it would be useful to see a lot more details and summary information on how the phasing was done and how its performance was.

Response: We thanks the comments of the reviewer. We are agree with the authors that switching error is one of most important issue in compound heterozygosity test and can be validated by long-range PCR. However, it is not practical in clinical research since the distance between two pathogenic variants could be 40-50K or longer. In our study, we didn’t discuss about the optimization of the string k variable sites since exom-array is not high dense array and we are agree with the review, it will be interesting to discussion the ‘k’ variable in genome-wide sequencing data. We prefer to apply the ‘k’ to the whole haplotype so that it will be consistent with the biological model of “compound heterozygosity”. The switching error and mistake phasing of ‘all variants in one side’ caused false-positive is one of important issues for compound heterozygosity test, especially as the reviewer mentioned, for rare variants. However, in our study, we found the significant association after the multiple test correction is quite limited and we can apply some other validation approach to validate the result such as gene co-expression network, pathway analysis, cellular and molecular experiences. What’s more, as the CRIPS-CAS9 technique, gene and variants functions can be validated in parallel and high-throughput way.

7/ Comparative genomic analysis of the iron metabolism genes can be moved to the discussion in a supportive fashion rather than part of the results as it may not be completely a validation point or proof for the function of FGF6.

Response: We thanks the comments of the reviewer. We moved the Comparative genomic analysis to the discussion section to be supportive evidence.

Reviewer 2 Comments for the Author...

By using a gene-based compound heterozigosity test on 10,000 DNAs the authors

identified HFE and FGF6 as susceptibility genes of iron overload. The finding

is interesting and so is the comparative genomic analysis showing that FGF6

evolved together with major genes of iron metabolism. However, the cellular

studies to support the role of FGF6 in cellular iron homeostasis seem preliminary and not fully convincing.

1- The studies of cellular iron supplementation should be done more carefully.

The iron concentration and time of incubation should be stated. Moreover the

quantification of cellular iron uptake by counting the Perl’s positive cells is

not straightforward. Perl’s stains mainly hemosiderin iron which is only a

fraction of total iron and that is deposited only when iron is in excess.

Moreover, I do not see the blue cells in the strong red background in fig 3 and

4. More direct methods would be more convincing, such as measurement of total

cellular iron, or of ferritin protein, ferritin-iron or also transferrin

receptor mRNA.

2- The differential uptake of Fe(II) and Fe(III) cannot be used as an evidence

of hepcidin-dependent iron uptake. Fe(III) at neutral pH readily forms

polynuclear insoluble complexes the absorption of which is not studied. Fe(II)

is supposed to be taken up mainly by ZIP14, which is unrelated to hepcidin

activity. Most studies of iron supplementation use ferric ammonium citrate in

the presence of ascorbate to maintain it in a mononuclear and soluble form.

3- The finding that FGF6 overexpression induces hepcidin mRNA is rather

convincing, less clear is the effect of the three mutants. The claim that M2

(D174V) differs from M1 (E172X) and M3 (R188Q) is supported only by fig 4B and

4D, and not by fig 4E-G.

4- A list of abbreviations would help the reader.

我仔细看过意见以后，发现reviewer 2主要是进行的功能实验，我对reviewer 2的解决办法如下：

1- The studies of cellular iron supplementation should be done more carefully.The iron concentration and time of incubation should be stated. Moreover the quantification of cellular iron uptake by counting the Perl’s positive cells is not straightforward. Perl’s stains mainly hemosiderin iron which is only a fraction of total iron and that is deposited only when iron is in excess. Moreover, I do not see the blue cells in the strong red background in fig 3 and 4. **More direct methods would be more convincing, such as measurement of total** **cellular iron, or of ferritin protein, ferritin-iron or also transferrin** **receptor mRNA.**

我会优化一下实验条件，之前因为没有ferric ammonium citrate(意见2)，我就直接用的Fe2+和Fe3+的盐代替，所以需要意见2中的方法，首先重复细胞铁染色的结果，另查阅文献后可以用Ferrozine assay(Blood. 2017 Oct 26;130(17):1923-1933. )检测**total** **cellular iron**和**ferritin-iron**，用Anti-Ferritin抗体[EPR3004Y] (ab75973) ([J Neurochem.](https://www.ncbi.nlm.nih.gov/pubmed/?term=Journal+of+Neurochemistry%2C+2004%2C+88%2C+1194%E2%80%931202) 2004 Mar;88(5):1194-202.)检测**ferritin protein**的蛋白水平。**transferrin** **receptor mRNA的mRNA**我们在Figure 4C-G中有体现，我们是否也要检测一下蛋白水平？

2- The differential uptake of Fe(II) and Fe(III) cannot be used as an evidence of hepcidin-dependent iron uptake. Fe(III) at neutral pH readily forms polynuclear insoluble complexes the absorption of which is not studied. Fe(II) is supposed to be taken up mainly by ZIP14, which is unrelated to hepcidin activity. Most studies of iron supplementation use ferric ammonium citrate in the presence of ascorbate to maintain it in a mononuclear and soluble form.

**购买ferric ammonium citrate和ascorbate，重复之前的细胞铁染色**

3- The finding that FGF6 overexpression induces hepcidin mRNA is rather convincing, less clear is the effect of the three mutants. The claim that M2 (D174V) differs from M1 (E172X) and M3 (R188Q) is supported only by fig 4B and 4D, and not by fig 4E-G.

**这个问题我不太会回答，只有786-O的细胞系支持这个观点，剩下两个细胞系的数据该怎么说呢。。**