1500 words excluding abstract, methods

Intro Paragraph 150 words

Max 6 figures/tables

No headings

Online Methods

**A Gene-Based Recessive Diplotype Exome Scan Discovers *FGF6* as a Novel Iron Metabolism Gene**

Shicheng Guo1, Shuai Jiang2,3, Mehdi Maadooliat1,4, Yanyun Ma2,3, Narendranath Epperla5, Zhan Ye6, Brent Olson6, (Fudan Collaborator)2,3, Terrie Kitchner1, Jeffrey Joyce1, Robert Strenn6, Joseph J. Mazza7, Jennifer K. Meece8, Jiucun Wang2,3,\*, Steven J. Schrodi1,9,\*

1Center for Precision Medicine Research, Marshfield Clinic Research Institute, Marshfield, WI, USA.

2State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China.

3Institute of Rheumatology, Immunology and Allergy, Fudan University, Shanghai, China.

4Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, WI, USA.

5Division of Hematology and Oncology, Medical College of Wisconsin, Milwaukee, WI, USA.

6Biomedical Informatics Research Center, Marshfield Clinic Research Institute, Marshfield, WI, USA.

7Clinical Research Center, Marshfield Clinic Research Institute, Marshfield, WI, USA.

8Integrated Research and Development Laboratory, Marshfield Clinic Research Institute, Marshfield, WI, USA.

9Computation and Informatics in Biology and Medicine, University of Wisconsin-Madison, Madison, WI, USA.

\*Correspondence:

Jiucun Wang, Ph.D.

School of Life Sciences

Fudan University

Shanghai, China

Email: [jcwang@fudan.edu.cn](mailto:jcwang@fudan.edu.cn)

Steven J. Schrodi, Ph.D.

Center for Precision Medicine Research

Marshfield Clinic Research Institute

1000 N Oak Ave--MLR

Marshfield, WI 54449

Tel: 715-221-6443

Email: [Schrodi.steven@mcrf.mfldclin.edu](mailto:Schrodi.steven@mcrf.mfldclin.edu)

[Schrodi@wisc.edu](mailto:Schrodi@wisc.edu)

Genome-wide Association Studies (GWAS) are well-designed to detect additive effects of moderate strength. We hypothesized that gene-based tests sensitive to compound heterozygosity may reveal additional genes underlying complex disease heritability, for compromised function on both homologous chromosomes is likely to profoundly impact molecular physiological states. Here we present a systematic case/control gene-based scan for recessive diplotypes for hemochromatosis using exome array data on 10,000 individuals from a biobank composed of genetically-homogeneous samples from Central Wisconsin. The study yielded the well-established hereditary hemochromatosis gene, *HFE*, and also generated a significant signal at the fibroblast growth factor, *FGF6*. Functional studies and homology with known iron metabolism genes demonstrate the role of Fgf6 in iron metabolism, highlighting a novel mechanism for iron pathophysiology. We anticipate that analogous investigations in other complex diseases may reveal additional susceptibility genes previously concealed from typical GWAS analyses.

As predicted by population genetics theory, deep sequencing studies have conclusively shown a vast reservoir of rare variants segregating in human populations.1 Simple power calculations show that modes of inheritance such as recessive diplotypes produce disease genetic signals that are difficult for standard GWAS methods to discover. Recently, a thorough investigation of compound heterozygosity disease models has provided ample evidence that human geneticists should consider these effects.2 These models also enjoy a high degree of biological plausibility, particularly if the alleles confer compromised protein function and/or expression. Indeed, substantial enrichment of pathogenic exonic variants over other variant classes has been observed across a wide variety of disease phenotypes.3-6 Further, recessive diplotype modes of inheritance are well-established in numerous Mendelian diseases, such as cystic fibrosis7, mevalonic aciduria8, beta-thalassemia9 and Niemann-Pick disease.10 Although not systematically examined in population-based disease studies, there is a sizable and growing repository of genes underlying complex diseases with recessive, loss-of-function effects.11-16 Hence, we posited that an exome-wide, gene-based screen of recessive diplotypes using putative functional variants in both oligogenic and complex diseases may expand our knowledge of disease genes.

Iron metabolism disorders, including adult hereditary hemochromatosis, collectively are common conditions with considerable public health implications.17,18 Under the rubric of iron metabolism disorders, there are multiple clinically-defined types of iron overload pathologies with variable ages of onset, symptomatology, and presentation. Although our understanding of iron metabolism and the pathophysiology of iron overload as accelerated over the past few decades, much remains enigmatic about their etiologies and clinical courses. Iron homeostasis is achieved by regulating iron absorption and storage to balance excretion flux. Absorption of heme iron and non-heme iron occurs in the small intestine via enterocyte-mediated uptake and regulated by the ferroxidase hephastin. The resulting iron-containing molecules are primarily used in myoglobin synthesis, the bone marrow, and reticulocytes/erythrocytes.19 Recycling of iron and storage within ferritin is a critical mechanism for modulating the overall iron stores and distribution. Importantly, the hepatic hormone hepcidin is a key regulator for maintaining iron homeostasis through controlling iron flux from enterocytes and macrophages to plasma and degradation of the cellular iron exporter ferroportin. Sustained perturbation of iron homeostasis resulting in iron overload can cultivate hepatic cirrhosis, hepatocellular carcinoma, cardiomyopathy, neurodegeneration and septicemia related in infections from pathogens such as *Yersinia enterocolitica*, *Listeria monocytogenes*, and *Vibrio vulnificus*. With early treatment, progression of many types of iron overload disorders can be halted and the pathology reversed.

To investigate the inheritance of hemochromatosis, several segregation analyses were conducted several decades ago, primarily concluding that a recessive mode of inheritance is highly plausible.20,21 A number of human studies have investigated the genetics of iron overload, revealing several critically important genes. Notably, the *HFE* gene on chromosome 6p, encoding for the membrane-bound hereditary hemochromatosis protein, was mapped approximately two decades ago through substantial efforts using family-based linkage22-25 and population-based association approaches.26,27 Additional studies have definitively placed the missense polymorphism C282Y in *HFE* as the major susceptibility factor in adult-onset, type 1 hereditary hemochromatosis.28,29

In an effort to discover novel iron overload-predisposing genes, we conducted a gene-based scan for recessive diplotypes composed of putative functional alleles across the exome using biobanked samples housed by the Marshfield Clinic Research Institute linked to electronic medical records (averaging in excess of 30 years) obtained from the adult Central Wisconsin population—a rural, genetically-homogeneous, stationary population, largely of Bavarian descent. Individuals were included as hemochromatosis cases based on percent transferrin saturation laboratory values (ratio of serum iron to transferrin iron-binding capacity) exceeding 48% and having two or more instances of ICD-9 codes indicating the diagnosis of hemochromatosis. Of the 10,000 samples evaluated, the phenotype algorithm identified 18 case individuals. Controls (n=6,896) were individuals without abnormal saturation values and without any instances of hemochromatosis ICD-9 codes. We estimated gametic phase on all individuals and restricted our analyses of diplotypes to putative functional variants. Our recessive diplotype association scan identified *FGF6* as an exome-wide significant finding. *FGF6* resides on chromosome 12p13.32 and encodes for the fibroblast growth factor 6.

To further explore the involvement of Fgf6 in iron metabolism, we investigated the homology…

Functional studies

Proposed pathway

We investigated the FGF6 pathway

Discussion

**Online Methods**

**Figure**

**Supplementary Tables**

**Supplementary Figures**

**References**

1. Lek M, Karczewski KJ, Minikel EV, Samocha KE, et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536(7616):285-291.

2. Sanjak JS, Long AD, Thornton KR (2017) A model of compound heterozygous, loss-of-function alleles is broadly consistent with observations from complex-disease GWAS datasets. PLoS Genet 13(1):e1006573.

1. MacArthur DG, Balasubramanian S, Frankish A, Huang N, et al. (2012) A systematic survey of loss-of-function variants in human protein-coding genes. Science 335(6070):823-828.
2. Zou J, Valiant G, Valiant P, Karczewski K, et al. (2016) Quantifying unobserved protein-coding variants in human population provides a roadmap for large-scale sequencing projects. Nat Commun 7:Article number 13293.
3. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305(5685):869-872.
4. Andreoletti G, Shakhnovich V, Christenson K, Coelho T, et al. (2017) Exome analysis of rare and common variants within the NOD signaling pathway. Sci Report 7:46454.
5. De Braekeleer M, Allard C, Leblanc JP, Simard F, Aubin G (1997) Genotype-phenotype correlation in cystic fibrosis patients compound heterozygous for the A455E mutation. Hum Genet 101(2):208-211.
6. Prietsch V, Mayatepek E, Krastel H, Haas D, et al. (2003) Mevalonate kinase deficiency: enlarging the clinical and biochemical spectrum. Pediatrics 111(2):258-261.
7. Thein SL (2005) Genetic modifiers of beta-thalassemia. Haematologica 90:649-660.
8. Bauer P, Knoblich R, Bauer C, Finckh U, et al. (2002) NPC1: complete genomic sequence, mutation analysis, and characterization of hapolotypes. Hum Mutat 19:30-38.
9. Khetarpal SA, Schjoldager KT, Christoffersen C, Raghavan A, et al. (2016) Loss of function of GALNT2 lowers high-density lipoproteins in humans, nonhuman primates, and rodents. Cell Metab 24(2):234-245.
10. Singh T, Kurki MI, Curtis D, Purcell SM, et al. (2016) Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci 19(4):571-577.
11. Adam R, Spier I, Zhao B, Kloth M, et al. (2016) Exome sequencing identifies biallelic MSH3 germline mutations as a recessive subtype of colorectal adenomatous polyposis. Am J Hum Genet 99(2):337-351.
12. Hague S, Rogaeva E, Hernandez D, Gulick C, et al. (2003) Early-onset Parkinson’s disease cause by a compound heterozygous DJ-1 mutation. Ann Neurol 54(2):271-274.
13. Onoufriadis A, Simpson MA, Pink MA, Di Meglio P, et al. (2011) Mutations in IL36RN/IL1F5 are associated with the severe episodic inflammatory skin disease known as generalized pustular psoriasis. Am J Hum Genet 89(3):432-437.
14. Dewey FE, Murray MF, Overton JD, Habegger L, et al. (2016) Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 354(6319) pii: aaf6814.
15. Adams PC, Barton JC (2007) Haemochromatosis. Lancet 370:1855-1860.
16. Andrews NC, Schmidt PJ (2007) Iron homeostasis. Annu Rev Physiol 69:69-85.
17. Andrews NC (1999) Disorders of iron metabolism. N Engl J Med 341(26):1986-1995.
18. Saddi R, Feingold J (1974) Idiopathic haemochromatosis: an autosomal recessive disease. Clin Genet 5:234-241.
19. Borecki IB, Rao DC, Yaouanq J, Lalouel JM (1989) Segregation of genetic hemochromatosis indexed by latent capacity of transferrin. Am J Hum Genet 45:465-470.
20. Simon M, Alexandre JL, Bourel M, Le Marec B, Scordia C (1977) Heredity of idiopathic haemochromatosis: a study of 106 families. Clin Genet 11(5):327-341.
21. Cartwright GE, Skolnick M, Amos DB, Edwards CQ, Kravitz K, Johnson A (1978) Inheritance of hemochromatosis: linkage to HLA. Trans Assoc Am Physicians 91:273-281.
22. Edwards CQ, Griffen LM, Dadone MM, Skolnick MH, Kushner JP (1986) Mapping the locus for hereditary hemochromatosis: localization between HLA-B and HLA-A. Am J Hum Genet 38(6):805-811.
23. Jazwinska EC, Lee SC, Webb SI, Halliday JW, Powell LW (1993) Localization of the hemochromatosis gene close to D6S105. Am J Hum Genet 53(2):347-352.
24. Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, et al. (1996) A novel MHC class-I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 13(4):399-408.
25. Jazwinska EC, Cullen LM, Busfield F, Pyper WR, et al. (1996) Haemochromatosis and HLA-H. Nat Genet 14(3):249-251.
26. Griffiths W, Cox T (2000) Haemochromatosis: novel gene discovery and the molecular pathophysiology of iron metabolism. Hum Mol Genet 9(16):2377-2382.
27. Allen KJ, Gurrin LC, Constantine CC, Osborne NJ, et al. (2008) Iron-overload-related disease in HFE hereditary hemochromatosis. N Engl J Med 358:221-230.

**Acknowledgements**

In 2013 and approved until Dec 20, 2018, the study was reviewed and approved by the Marshfield Clinic Research Institute Institutional Review Board. FWA00000873, IRB00000673, Title: *Two Allele Loss of Function Genotype Array Study*. The project described was supported by the Clinical and Translational Science Award (CTSA) program, previously through the National Center for Research Resources (NCRR) grant 1UL1RR025011 and the National Center for Advancing Translational Sciences (NCATS) grant 9U54TR000021, and now by the NCATS grant UL1TR000427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Additional funding was provided by generous donors to the Marshfield Clinic Research Institute. (funding from JW lab)

**Author contributions**

SG performed analyses, interpreted results, designed the functional studies, and aided in drafting the manuscript. MM aided in the analyses and reviewed the manuscript. NE provided clinical advice and aided in drafting the manuscript. ZY performed analyses. BO implemented the phenotyping algorithms. TK and JJ aided in the regulatory paperwork and reviewed the manuscript. RS performed data management tasks. JJM provided clinical advice and reviewed the manuscript. JKM supervised the management of biological samples for genotyping and reviewed the manuscript. JW supervised the functional studies, reviewed the manuscript and provided biological advice. SJS designed the experiment, supervised the genetic analyses, developed phenotyping algorithms, developed analysis methods, interpreted results and aided in drafting the manuscript.

**Competing interests**