Marquette University

e-Publications@Marquette

Mathematical and Statistical Science Faculty Research and Publications

Mathematical and Statistical Science, Department of

4-25-2019

A gene-based recessive diplotype exome scan discovers *FGF6*, a novel hepcidin-regulating iron-metabolism gene

Shicheng Guo

Shuai Jiang

Narendranath Epperla

Yanyun Ma

Mehdi Maadooliat

See next page for additional authors

Follow this and additional works at: https://epublications.marquette.edu/math_fac

Authors

Shicheng Guo, Shuai Jiang, Narendranath Epperla, Yanyun Ma, Mehdi Maadooliat, Zhan Ye, Brent Olson, Minghua Wang, Terrie Kitchner, Jeffrey Joyce, Peng An, Fudi Wang, Robert Strenn, Joseph J. Mazza, Jennifer K. Meece, Wenyu Wu, Li Jin, Judith A. Smith, Jiucan Wang, and Steven J. Schrodi

Marquette University

e-Publications@Marquette

Mathematics and Statistical Sciences Faculty Research and Publications/College of Arts and Science

***This paper is NOT THE PUBLISHED VERSION;* but the author's final, peer-reviewed manuscript.** The published version may be accessed by following the link in the citation below.

Blood, Vol. 133, No. 17 (April 25 2019) : 1888-1898. DOI. This article is © American Society of Hematology and permission has been granted for this version to appear in e-Publications@Marquette. American Society of Hematology does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from American Society of Hematology.

A gene-based recessive diplotype exome scan discovers *FGF6*, a novel hepcidin-regulating iron-metabolism gene

Shicheng Guo
Center for Human Genetics, Marshfield Clinic Research Institute, Marshfield, WI
Shuai Jiang
State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Science and Human Phenome Institute, Fudan University, Shanghai, China
Narendranath Epperla
Division of Hematology, The Ohio State University, Columbus, OH
Yanyun Ma
State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai, China
Mehdi Maadooliat
Center for Human Genetics, Marshfield Clinic Research Institute, Marshfield, WI
Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, WI
Zhan Ye

Biomedical Informatics Research Center, Marshfield Clinic Research Institute, Marshfield, WI
Brent Olson
Biomedical Informatics Research Center, Marshfield Clinic Research Institute, Marshfield, WI
Minghua Wang
Department of Biochemistry and Molecular Biology, Medical College, Soochow University, Suzhou, China
Terrie Kitchner
Center for Human Genetics, Marshfield Clinic Research Institute, Marshfield, WI
Jeffrey Joyce
Center for Human Genetics, Marshfield Clinic Research Institute, Marshfield, WI
Peng An
Beijing Advanced Innovation Center for Food Nutrition and Human Health, China Agricultural University, Beijing, China
Fudi Wang
Nutrition Discovery Innovation Center, School of Public Health, Zhejiang University School of Medicine, Hangzhou, China
Robert Strenn
Biomedical Informatics Research Center, Marshfield Clinic Research Institute, Marshfield, WI
Joseph J. Mazza
Clinical Research Center, Marshfield Clinic Research Foundation, Marshfield, WI
Jennifer K. Meece
Integrated Research and Development Laboratory, Marshfield Clinic Research Institute, Marshfield, WI
Wenyu Wu
Department of Dermatology, Huashan Hospital, Fudan University, Shanghai, China
Li Jin
State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai, China
Judith A. Smith
Department of Pediatrics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI
Jiucun Wang
State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai, China; Institute of Rheumatology, Immunology and Allergy, Fudan University, Shanghai, China
Steven J. Schrodi
Center for Human Genetics, Marshfield Clinic Research Institute, Marshfield, WI
Computation and Informatics in Biology and Medicine, University of Wisconsin-Madison, Madison, WI

Abstract

Standard analyses applied to genome-wide association data are well designed to detect additive effects of moderate strength. However, the power for standard genome-wide association study (GWAS) analyses to identify effects from recessive diplotypes is not typically high. We proposed and conducted a gene-based compound heterozygosity test to reveal additional genes underlying complex diseases. With this approach applied to iron overload, a strong association signal was identified between the fibroblast growth factor–

encoding gene, *FGF6*, and hemochromatosis in the central Wisconsin population. Functional validation showed that fibroblast growth factor 6 protein (FGF-6) regulates iron homeostasis and induces transcriptional regulation of hepcidin. Moreover, specific identified *FGF6*variants differentially impact iron metabolism. In addition, FGF6 downregulation correlated with iron-metabolism dysfunction in systemic sclerosis and cancer cells. Using the recessive diplotype approach revealed a novel susceptibility hemochromatosis gene and has extended our understanding of the mechanisms involved in iron metabolism.

Introduction

Genome-wide association studies (GWASs) are well designed to detect additive effects of modest effect sizes. We hypothesized that gene-based tests sensitive to recessive diplotypes, including recessive single-site effects and compound heterozygosity, may reveal additional genes underlying complex diseases. Carrying variants conferring a compromised function on both homologous chromosomes is likely to impact molecular physiological states. Deep-sequencing studies have conclusively shown a vast reservoir of rare variants segregating in human populations.[1] Rare variants in functional categories (eg, missense, regulatory motifs) may generate pathogenic effects through recessively acting diplotypes, and such effects are apt to remain concealed from standard GWAS analyses. Simple power calculations show that recessive diplotype inheritance produces signals that are difficult for standard GWAS methods to discover (supplemental Figure 1, available on the *Blood* Web site). Compound heterozygosity disease models also enjoy a high degree of biological plausibility, particularly if the alleles confer compromised protein function.[2-5] Recessive diplotype modes of inheritance are well established in Mendelian diseases, such as cystic fibrosis,[6] mevalonic aciduria,[7] β-thalassemia,[8]and Niemann-Pick disease.[9] Although not systematically examined in population-based studies, there is a sizable repository of genes underlying complex diseases with recessive, loss-of-function effects.[10-14] 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.[15,16]Importantly, the hepatic hormone hepcidin is a key regulator of iron homeostasis by controlling iron flux from enterocytes and macrophages to plasma through degradation of the cellular iron exporter ferroportin (*SLC40A1*). Within cells, ferritin is the iron-storage protein that can be used for indirect iron quantification. To investigate the inheritance of hemochromatosis, several segregation analyses were initially conducted, concluding that a recessive mode of inheritance is highly plausible.[17,18] Several human studies have investigated the genetics of iron overload, revealing several critically important genes. Notably, *HFE*, encoding the membrane-bound hereditary hemochromatosis protein, was mapped 2 decades ago through family-based linkage[19-22] and association approaches.[23,24] Additional studies have definitively placed the missense polymorphism C282Y (rs1800562) in *HFE* as the major susceptibility factor in adult-onset, type 1 hereditary hemochromatosis.[25,26] Additional genes have been identified through pathway-based genetic association studies and GWASs, including *BMP2*, *BMP4*, *HJV*, *TF*, *TMPRSS6*, *NAT2*, *FADS2*, and *TFR2*.[27-29]

Methods

Central Wisconsin hemochromatosis sample set

The homogenous population in rural central Wisconsin is the source population for the Personalized Medicine Research Project (PMRP), a biobank linked to electronic health records (EHRs) housed by the Marshfield Clinic Research Institute.[30] Samples from over 20 000 individuals comprise the PMRP. The study was conducted in accordance with the Declaration of Helsinki. All samples were collected following written informed consent. All investigators using the PMRP samples had obtained Research Ethics and Compliance Training certification

through the Collaborative Institutional Training Initiative (CITI) program. The study protocol was reviewed and approved by the Marshfield Clinic Institutional Review Board (details in "Acknowledgments"). The Central Wisconsin population is largely stationary and primarily derived from Bavarian migrants in the late 1800s. The population carries high utility for disease gene mapping through reduction in confounding by population stratification and lower expected levels of allelic and locus heterogeneity. Additionally, environmental exposures are thought to be relatively uniform across this population. For these reasons, the PMRP has been effectively used in numerous human genetics studies.[31-34] PMRP DNA samples were collected and stored ~14 years ago and all individuals have longitudinal EHR information housed at the Marshfield Clinic, averaging >30 years. The EHR is composed of International Classification of Diseases, Ninth Revision (ICD-9) diagnostic codes, laboratory test results, clinical procedure data, prescription information, and physician notes. Hemochromatosis cases and controls were selected from the PMRP population. Hemochromatosis cases were selected on the basis of the percent transferrin saturation laboratory values (the ratio of serum iron to transferrin iron-binding capacity) exceeding 48% and having 2 or more instances of ICD-9 codes indicating the diagnosis of hemochromatosis: 275.0 (iron metabolism disorder, excluding anemia), 275.01 (hereditary hemochromatosis), 275.03 (unspecified hemochromatosis), and/or 275.09 (other iron-metabolism disorders). To reduce confounding by population stratification, a principal components analysis on the exome-genotyping data was implemented using all samples, blinded to disease status. Individuals considered genetic background outliers (>3 standard deviations from the centroid of the first 2 principal components) were excluded from the study. Following the removal of outliers, the resulting set of individuals was highly homogeneous based on the first 3 principal components. Exhaustive pairwise kinship coefficients were calculated and 1 individual from pairs of individuals exhibiting third-degree or closer relatedness were removed. Of the ~10 000 individuals previously subjected to the exome-genotyping array and quality control procedures, the phenotype algorithm identified 18 individuals who were selected as hemochromatosis cases. Controls (n = 6896) were individuals without abnormal saturation values and without any instances of hemochromatosis ICD-9 codes.

Genotyping

Of the full PMRP cohort, ~10 000 DNAs were interrogated by high-density genotyping on the Illumina HumanCoreExome beadchip. This exome array of >500 000 markers has ancestry informative markers, a panel of identity-by-descent single nucleotide polymorphisms (SNPs), coverage of markers found to be genome-wide significant in GWASs, and excellent coverage of exonic variants. The version of the beadchip was designed and used in the AMD Consortium.[34] Rare variants (<1% frequency) represented 47.8% of the markers, moderately common variants (1% to 10% frequency) were 8.1% of the variants, and 44.1% of the variants interrogated were common alleles (>10% frequency). The genotyping quality control measures were previously described (call rates for each variant or individual >0.985).[34] Variants exhibiting departure from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$) were excluded from subsequent analyses. Additional recent studies have used data generated from this genotyping to discover susceptibility genes for common diseases.[35] Following quality control procedures, 413 701 variants remained for analysis. The site frequency spectrum of the resulting variants is displayed in supplemental Figure 2 and supplemental Table 1.

Haplotype phasing

In general, gametic phasing is necessary to directly determine compound heterozygous individuals at a particular gene. Using all 10 000 exome-genotyped samples from the PMRP, the software package Beagle was applied to infer phased haplotypes from unphased genotype data using a localized haplotype-cluster model algorithm.[36] The calculations were performed on a high-performance computing cluster housed at the Marshfield Clinic. As the subsequent analyses were gene-based and the genotyping data were concentrated on exonic variants, each gene in the exome was phased separately using this approach. Although rare variants can present difficulties in phasing, the use of a large sample size from a highly homogeneous population aids in

mitigating the error rate. Notably, Beagle has been shown to have error rates in phasing between 0.77% to 0.94% for medium (n = 1000) to large (n = 5000) sample sizes using a 500K GWAS array.[36] Recently, the switch error rate was calculated for the Beagle applied to 2 sequencing data sets. Beagle attained a switch error rate of 1.525% and 0.488% for the 1000 Genomes Project and Haplotype Reference Consortium, respectively.[37]

Determination of putative functional variants

Following the phasing of the genotype data, putative functional variants were identified. The putative functional variants included in the analyses satisfied the quality control criteria as described previously.[34] Markers used in the analyses were either GWAS-significant as of June 2015 and/or annotated as missense, nonsense, 3' untranslated region, 5' untranslated region, or occurring within a splice site region. Additionally, on the resulting set of variants, annotation was performed using the ANNOVAR software.[38] Only variants that were also annotated as pathogenic by at least 2 ANNOVAR predictions were included in the scan. Following filtering for putative functional variants, 129 556 SNPs remained for use in our gene-based recessive diplotypes scan. The putative functional variants of *FGF6* are shown in supplemental Table 2.

Statistical tests of recessive diplotypes

At each gene, individuals were classified as having a recessive diplotype configuration if they carried at least 1 putative functional allele on each homolog (*PF*). Individuals carrying at least 1 homolog free from putative functional alleles were deemed as having a wild type (WT) diplotype (*W*). The total number of case/control individuals carrying a recessive diplotype was denoted by *PFcs* and *PFct*, respectively. Similarly, the total number of case/control individuals carrying a WT diplotype was denoted by *Wcs* and *Wct*, respectively. Following the determination of these counts, a Fisher exact test was applied. As the hypergeometric null density holds for all sample sizes, the Fisher exact test is robust to imbalance between case and control sample sizes. Simulations have recapitulated this finding showing that the Fisher exact test does not inflate type I error rates under unbalanced designs.[39]Individuals carrying 1 or more homozygous genotype(s) at a single site for a putative functional allele were included in the *PF* category. Genes without any high-quality, putative functional alleles across all samples were removed from the analyses. Across all genes with analyzable data, a conservative experiment-wise multiplicity correction was calculated using 15 900 gene-based tests. To compare the recessive diplotype analysis procedure to a standard rare variant gene-based test, the RVTESTS software,[40] which implements the sequence kernel association test, was also applied to the genotype data.[41] Additionally, to investigate the sex-specific effects, the Haldane odds ratio (OR) was calculated separately for female and male strata. Lastly, the Mantel-Haenszel test of homogeneity was calculated to determine the level of statistical evidence for sex-specific differences in effects.[42]

Power calculations

To explore the efficacy of the proposed approach, we performed analytic power calculations under the alternative model of compound heterozygosity/recessive inheritance of disease at 2 sites, each segregating 2 alleles. By doing so, we sought to compare the power of a standard GWAS analysis (Armitage trend test) to a log-likelihood ratio *G* test.[42] Supplemental Figure 1 shows the power of each of these tests across different sets of penetrances and haplotype frequencies. To consolidate the different sets of haplotype frequencies, the results are plotted as a function of linkage disequilibrium between the 2 sites. The power of the *G* test for recessive diplotypes exceeded the power for the Armitage trend test across virtually all of the parameter space. Additional work in this area was recently performed by Sanjak et al showing similar results.[43]

Comparative genomic analysis and protein-protein interaction inference

Amino acid sequencing of the core iron-metabolism genes were collected, including transferrin receptor 1 (*TFRC*), *FTH1*, *IREB1*, *SKP1*, *SKP1*, *ACO1*, *TFR2*, *TF*, *HMOX1*, *ACO2*, *HAMP*, *FGF6*, and *FGFR1*. The alignments were

derived from National Center for Biotechnology Information (NCBI) BLASTn database. Phylogeny for different genes were compared, showing the earliest evolutionary time point; then, occurrence for each gene was mapped to the phylogenetic tree.[44] Protein-protein interaction network inference was conducted to fibroblast growth factor 6 protein (FGF-6) and main iron-metabolism proteins. The finial network was tuned after removing nonnecessary nodes between FGF-6 and key iron molecules including hereditary hemochromatosis protein (HFE) and SLC40A1.

Cell culture, reagents, and protein treatment

Colon cancer cell lines (HCT8 and HCT116), a kidney cancer cell line (786-O), a liver cancer cell line (HepG2), and a fibroblast cell line (HFF-1) were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) at 37°C in a 5% CO_2 humidified incubator. To investigate the change in iron uptake under different protein treatments or plasmid transfection, the ferric ammonium citrate (FAC) cell culture method was used. Cells were cultured in normal DMEM and FBS medium with the presence of 10 μM FAC and 500 μM ascorbate for 48 hours during detection of cellular iron concentration. Cells cultured in normal medium exhibited very low iron concentrations. Total intracellular iron concentration in cells cultured with FAC for 48 hours was dramatically increased over cells cultured in normal medium. Recombinant human FGF-6 protein (active) and anti-ferritin were purchased from Abcam, Flag tag antibody was purchased from Proteintech Group, and anti–glyceraldehyde-3-phosphate dehydrogenase (GAPDH) antibody was purchased from Shanghai Yeasen Biotechnology. FAC (1 mM) and ascorbate (50 mM) were dissolved in distilled water. NaOH, HCl, $KMnO_4$, ferrozine, neocuproine, ammonium acetate, ascorbic acid, and $FeCl_3$ were purchased from Beijing Oka Biological Technology. Plasmid with raw *FGF6* sequence was purchased from PPL-Shanghai Co, Ltd, which was constructed in an N-terminal p3XFLAG-CMV vector, whereas 3 different *FGF6* mutations (E127X, D174V, R188Q) were synthesized with overlapping polymerase chain reaction (PCR).

Quantification of iron content by ferrozine assay

Total intracellular iron content was measured by the ferrozine assay.[45] Cells were cultured in 12- or 24-well plate for 48 hours and washed 3 times with cold phosphate-buffered saline (PBS). After being lysed for 2 hours with 50 mM NaOH, 100 μL of cell lysates was mixed with 10 mM HCl, and 100 μL of the iron-releasing reagent (a freshly mixed solution of equal volumes of 1.4 M HCl and 4.5% [wt/vol] $KMnO_4$ in H_2O). The mixtures were incubated for 2 hours and 30 μL of iron-detection reagent (6.5 mM ferrozine, 6.5 mM neocuproine, 2.5 M ammonium acetate, and 1 M ascorbic acid) was added; after 30-minute incubation, 280 μL of solution was added to a 96-well plate and read 550 nm on a microplate reader. In addition, $FeCl_3$ (0-100 μM) was used as iron standards and protein quantification was determined by a Lowry protein assay.

Western blot

Cell lysates were harvested when incubated with iron for 48 hours, then equal amounts of protein from every sample were subjected to 12% sodium dodecyl sulfate–polyacrylamide gel electrophoresis and then transferred to polyvinylidene difluoride membranes. After blocking with 5% bovine serum albumin (BSA), the membranes were incubated with GAPDH (1:10000), ferritin (1:1000), and Flag (1:2000) at 4°C overnight. Then, membranes were washed 3 times with Tris-buffered saline plus polysorbate 20, incubated with anti-rabbit or anti-mouse secondary antibody. The bands were visualized using Image QuantTL software.

Perls staining

Cells were washed with PBS 3 times, fixed with 4% glutaraldehyde for 10 minutes, and incubated at 37°C for 60 minutes with 2 mL of Prussian blue solution comprising equal volumes of 2% hydrochloric acid aqueous solution and 2% potassium ferrocyanide (II) trihydrate. After the cells were stained with 0.5% neutral red for 3 minutes,

iron staining was visualized by Nikon microscope. Iron-positive high positive staining cells divided by total cell number was used to evaluate the iron-deposition levels.

RT-PCR and quantitative RT-PCR analysis

Total RNA was extracted from the cells using TRIzol (Invitrogen). One microgram of total RNA was subjected to complementary DNA synthesis using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems) according to the manufacturer's instructions. The specific primers for each gene were designed using Primer 5 and synthesized by Generay Biotech Co, Ltd. The reverse transcription polymerase chain reaction (RT-PCR) amplification was conducted using a SYBR Green I PCR kit (TaKaRa) according to the manufacturer's instructions. The reaction was carried out on an ABI Prism 7900 Detector System (Applied Biosystems). RT-PCR conditions were 95°C for 3 minutes, followed by 40 cycles of 95°C for 15 seconds, 60°C for 40 seconds; the conditions for obtaining the dissociation curve were 95°C for 15 seconds, 60°C for 15 seconds, 95°C for 15 seconds. The data obtained from the assays were analyzed with SDS 2.3 software (Applied Biosystems). For each sample, the relative gene expression was calculated using a relative ratio to GAPDH. RT-PCR primers can be found in the supplemental Table 3.

Immunohistochemical staining of FGF-6

The primary antibody used was anti-FGF-6 (1:200; BBI) and anti-ferritin (1:100; abcam). Liver and skin tissues from 4 liver cancer patients and 6 systemic sclerosis (SSc) patients, respectively, and normal controls were formalin-fixed and paraffin-embedded. Sections were deparaffinized and incubated with 5% BSA for 60 minutes. Cells positive for FGF-6 were detected by incubation with the primary antibody for 2 hours at room temperature followed by incubation with 3% hydrogen peroxide for 10 minutes. Rabbit anti-rabbit immunoglobulin G labeled with horseradish peroxidase was used as secondary antibody. The expression of FGF-6 or ferritin was visualized with 3,3-diaminobenzidinetetrahydrochloride (DAB-4HCl). The expression of FGF-6 in SSc and tumor tissues was quantitated by the average optical density (AOD) of the positive signal in each sample using NIH ImageJ software (Windows and Java-1.8.0).

Results

Gene-based compound heterozygosity scan identifies a novel hemochromatosis-susceptibility gene

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 linked to electronic medical records obtained from a rural, genetically homogeneous population in central Wisconsin. Of the 10 000 samples evaluated, our transferrin saturation and diagnostic code-based phenotype algorithm identified 18 case individuals and 6896 controls. We estimated gametic phase on all individuals and restricted our analyses of diplotypes to putative functional variants. Our recessive diplotype scan identified 2 exome-wide significant genes (**Figure 1**; **Table 1**; supplemental Figure 3), *HFE* ($P = 1.29 \times 10^{-8}$; OR = 28.7) and *FGF6* ($P = 1.99 \times 10^{-6}$; OR = 22.8). For comparison, the SKAT/rvtest procedure on the *FGF6* genotype data yielded an asymptotic $P = 3.86 \times 10^{-5}$ and permuted $P = 1.0 \times 10^{-4}$. Notably, the recessive diplotype scan result exceeded exome-wide significance, whereas the rare variant test did not. Additionally, there was no statistical evidence of effect differences between females and males for the *FGF6*data (Mantel-Haenszel test of homogeneity $P = .728$). These results motivated our investigation of FGF-6 function and the impact of specific *FGF6* variants on iron metabolism.

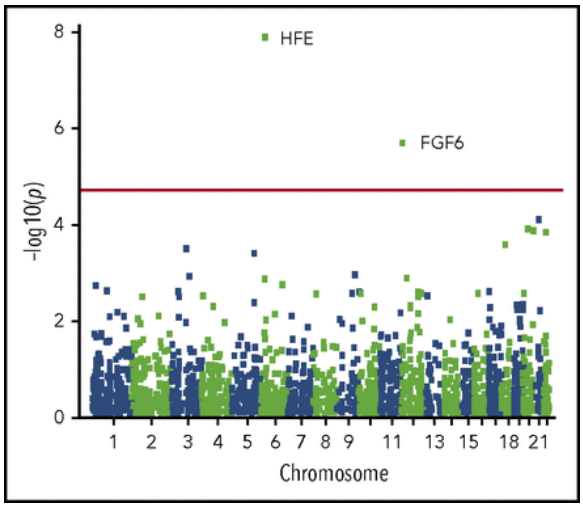

Figure 1. Manhattan plot of the gene-based recessive diplotype association results. The association *P*-value testing hemochromatosis association for each gene ($-log_{10}\ p$ plotted on the ordinate) on different chromosomes is shown in alternating navy blue and yellow along the abscissa, with the experiment-wise significance level for the gene-based analyses across the exome (experiment-wise $\alpha = 3.14 \times 10^{-6}$) depicted in red.

Table 1. Significant genes identified by recessive diplotype scanning

Chromosome	Gene	*P*	OR	SNPs	Case+	Case−	Control+	Contro
chr6	*HFE*	1.29×10^{-8}	28.6	14	8	10	189	6707
chr12	*FGF6*	1.99×10^{-6}	22.8	10	6	12	153	6743
chr21	*KRTAP15-1*	7.55×10^{-5}	6.78	5	11	7	1271	5625
chr20	*XKR7*	1.18×10^{-4}	43.6	7	3	15	35	6861
chr20	*CABLES2*	1.28×10^{-4}	42.4	7	3	15	36	6860
chr22	*THOC5*	1.38×10^{-4}	6.24	9	13	5	1945	4951

The 6 most significant genes identified in the recessive diplotype scan are displayed. *P* values are from a 2-tailed Fisher exact test.

Case+, number of iron-overload case individuals carrying recessive diplotypes with putative functional alleles; Case−, number of cases carrying at least 1 homolog at the gene without a putative functional allele; Control+, number of control individuals carrying recessive diplotypes with putative functional alleles; Control−, number of controls carrying at least 1 homolog at the gene without a putative functional allele; OR, Haldane odds ratio; SNP, number of genotyped variants per gene that were polymorphic in the samples studied.

Protein-protein interaction indicates FGF-6 is involved in iron-metabolism network

To explore the involvement of *FGF6* in iron metabolism, we found evidence for FGF-6 interactions with FGFR1, MAPK1/3, INS, FN1, and involvement in the iron-metabolism subnetwork involving transferrin (TF), HFE, hepcidin antimicrobial peptide (HAMP), and SLC40A1 (supplemental Figure 4) by investigating FGF-6 protein-protein interactions. FGF-6, also known as heparin secretory-transforming protein 2 or heparin-binding growth factor 6, has multiple heparin-binding sites (HBSs). Three known nonsynonymous variants located in the HBSs (R188Q) or flanking sites (D174V and E172X) were speculated to be important for FGF-6 function. Furthermore, D174V and E172X are located in the regions between FGFR-binding region (FGFR-BR-3) and HBS-1 (**Figure 2**). Hence, we studied these 3 variants in functional studies to further investigate the involvement of FGF-6 in iron metabolism.

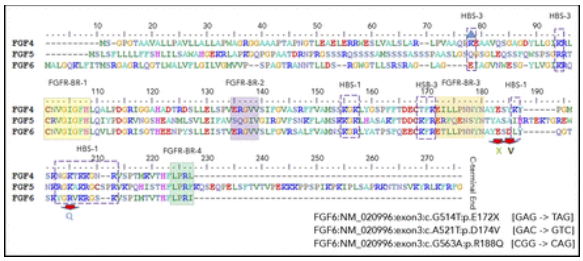

Figure 2. Protein sequence alignment for FGF-4, FGF-5, and FGF-6 with heparin and FGFR-binding domains. Protein domains summarized from a previous FGF-6 functional study.[58] Alignment and heparin and FGFR-binding sites/regions (HBS and FGFR-BR, respectively) are shown for FGF-4, FGF-5, and FGF-6 proteins.

FGF-6 modulation of hepcidin expression and iron uptake

To investigate the potential mechanism linking FGF-6 to iron metabolism, the effects of FGF6 on iron uptake and the expression of iron-metabolism genes in HepG2, HCT8, HCT116, 786-O, and HFF1 cells were evaluated. Using cultured cells and a ferrozine assay to detect iron, total intracellular iron concentration was significantly decreased in HepG2, 786-O, HCT8, HCT116, and HFF-1 cells when treating with active FGF-6 protein in a dose-dependent manner (**Figure 3**; supplemental Figure 5). Testing the effect of FGF6 on the expression of a set of genes involved in iron metabolism (*HAMP*, *HDAC2*, *HMOX1*, *TFRC*, and *HEPH*), HepG2 cells were subjected to treatment from control or FGF-6 protein and *FGF6*messenger RNA (mRNA) or control and mRNA expression relative to *GAPDH* was measured in the 5 iron-metabolism genes. RT-PCR analysis revealed that *HAMP* and *HDAC2* mRNA levels were significantly increased after the FGF-6 active protein introduction in HepG2 cells compared with treatment with PBS as control (**Figure 4A**). *FGF6* plasmid transfection significantly increased *HAMP*, *HDAC2*, and *HMOX1*levels, whereas *TFRC* levels significantly decreased in HepG2 compared with a vector without *FGF6* (**Figure 4B**). *HEPH* expression did not change with either *FGF6* plasmid or FGF-6 protein, suggesting that the effect of FGF-6 may be independent of *HEPH*(**Figure 4A-B**).

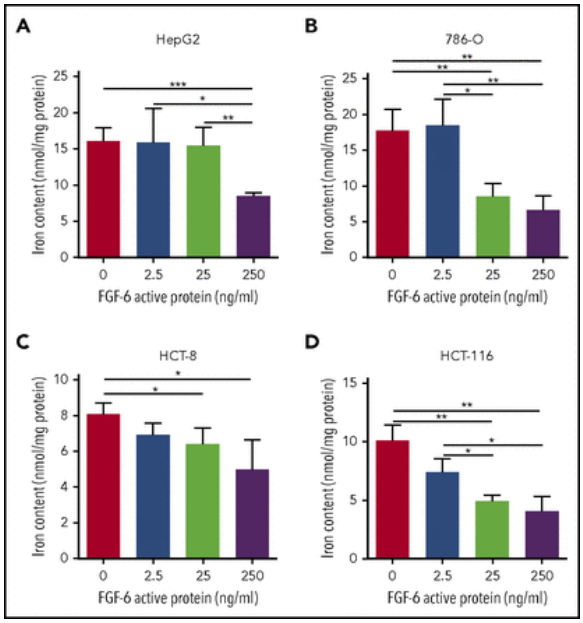

Figure 3. FGF-6 active protein dosage effect on intracellular iron concentration. A ferrozine assay was applied for the evaluation of total cell iron content in HepG2 (human liver hepatocellular carcinoma cell line), 786-O (human kidney adenocarcinoma cell line), HCT-8 (human ileocecal colorectal adenocarcinoma cell line), HCT116 (human colon carcinoma cell line), and HFF-1 (human skin fibroblast cell line) with 10 µM FAC and 500 µM ascorbate in cell culture media, respectively, with different concentrations of FGF-6 active protein (0 ng/mL, 2.5 ng/mL, 25 ng/mL, and 250 ng/mL). Control group was treated with ascorbate alone. After 48-hour incubation,

cells were lysed and iron contents were determined with the ferrozine assay. (A) Total iron content in HepG2 cells with increasing FGF-6 protein concentration. (B) Total iron content in 786-O cells with increasing FGF-6 protein concentration. (C) Total iron content in HCT-8 cells with increasing FGF-6 protein concentration. (D) Total iron content in HCT-116 cells with increasing FGF-6 protein concentration. *P< .05, **P < .01, ***P < .001.

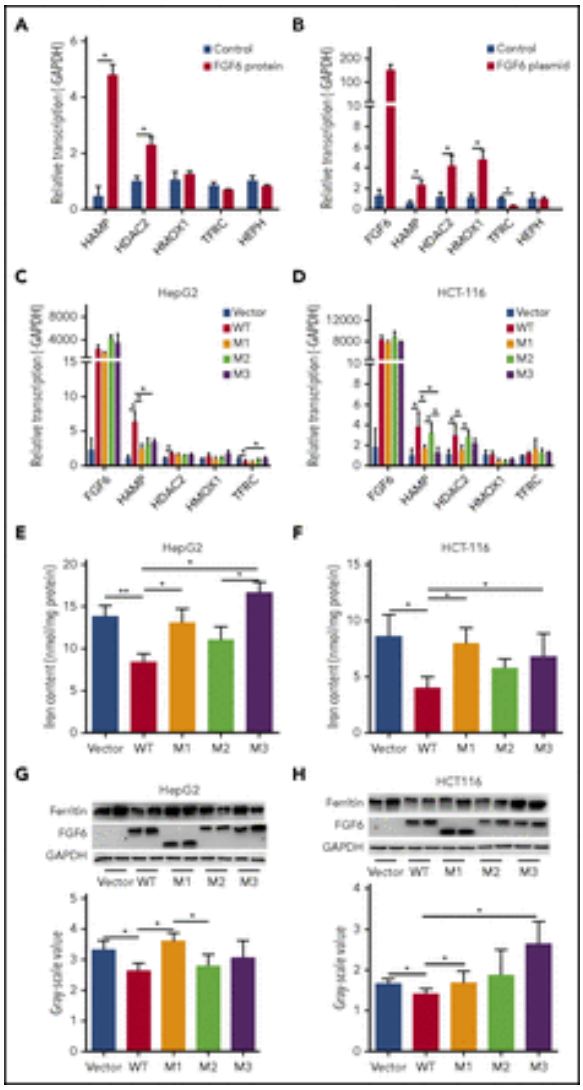

Figure 4. The effect of *FGF6* nonsynonymous variants on hepcidin expression and intracellular iron concentration. (A) The effect of FGF-6 active protein treatment on mRNA expression of several iron metabolism genes in HepG2 liver hepatocellular carcinoma cell culture media compared with control. Protein concentration was 250 ng/mL and the incubation time was 24 hours with 10 µM FAC and 500 µM ascorbate in the cell culture media. *HAMP* encodes for hepcidin. *HDAC2* encodes for histone deacetylase 2. *HMOX1* encodes for heme oxygenase 1. *TFRC* encodes for transferrin receptor 1 and HEPH encodes hephaestin. mRNA expression was quantified relative to *GAPDH* expression. Treatment with PBS served as control. A Student *t* test was used test for pairwise differences between sets of observations. *P < .05. Results are the mean ± standard deviation (SD) of 3 observations in a single experiment. (B) Iron-metabolism gene expression changes with *FGF6* mRNA transfection in the HepG2 cell culture media after 24 hours. Vector without *FGF6* served as control. A Student *t* test was used to test for pairwise differences between sets of observations. *P < .05. Results are the mean ± SD of 3 observations in a single experiment. (C-D) Iron-metabolism gene-expression changes after the transfection by *FGF6* mRNA into various cell types with WT and the identified variants E172X (M1), D174V (M2), and R188Q (M3). Cell lines: HepG2 are liver hepatocellular carcinoma cells, HCT116 are ileocecal colorectal adenocarcinoma cells, and HFF-1 are human normal skin fibroblasts. A Student *t*test was used to test for

pairwise differences between sets of observations. *P < .05. Results are the mean ± SD of 3 observations in a single experiment. (E-F) Total intracellular iron-concentration changes after the transfection with *FGF6* mRNA into 3 cell types with WT and the identified M1, M2, and M3 variants in the presence of FAC for 48 hours. A Student t test was used to test for pairwise differences between sets of observations. *P < .05; **P < .01. Results are the mean ± SD of 3 observations in a single experiment. (G-H) Ferritin protein level changes after the transfection by *FGF6*mRNA into the 3 cell types with WT and the identified M1, M2, and M3 variants in the presence of FAC for 48 hours. A Student t test was used to test for pairwise differences between sets of observations. *P < .05. Results are the mean ± SD of 3 observations in a single experiment.

Evaluation of *FGF6* variants on HAMP expression and iron concentration compared with WT *FGF6*

To investigate the effects of the *FGF6* alleles on FGF-6 function, we transfected plasmids carrying either the WT *FGF6* or variant *FGF6* with each of the 3 point mutations described in **Figure 2** (supplemental Figure 6). The M1 (E172X) and M3 (R188Q) variants exhibited a significant downregulation of *HAMP* mRNA compared with WT in HepG2 cells (**Figure 4C**), HCT-116 cells (**Figure 4D**), 786-O (supplemental Figure 7), A498 (supplemental Figure 7), but not HCT-8 cells (supplemental Figure 7). Evaluating the effect of M2 (D174V) on *HAMP* expression compared with WT only yielded a significant reduction in HepG2 (**Figure 4C**), but not in any of the other cell lines (**Figure 4D**; supplemental Figure 7). Furthermore, we noted *HAMP* mRNA levels in M1 and M3 transfections were comparable to control levels, which illustrated a strong attenuation of *FGF-6* function for M1 and M3 variants (**Figure 4C-D**; supplemental Figure 7). Examining the impact of specific variants on intracellular iron concentration in HepG2 and HCT-116 cells, M1 and M3 produced significantly elevated iron deposition (**Figure 4E-F**; supplemental Figure 8) and ferritin expression (**Figure 4G-H**; supplemental Figure 8), indicating a deficiency in M1/M3 *FGF6*-mediated iron homeostasis. In addition, the intracellular iron-accumulation pattern was confirmed by immunohistochemistry (IHC) using Perls stain (supplemental Figure 9). In contrast, M2 did not produce a significant departure from WT in iron concentration and ferritin expression in HepG2 and HCT-116 (**Figure 4**). Furthermore, *TFRC* expression was significantly upregulated in the presence of M3 compared with WT (**Figure 4C**). The functions mentioned were also validated in HFF-1 (supplemental Figure 8).

Altered FGF6 gene expression in SSc and cancer

We hypothesized that FGF-6 might be involved in human autoimmune diseases and cancers because abnormal iron metabolism has been reported in numerous studies.[46-50] More specifically, decreased hepcidin has been implicated in the anemia of chronic disease, which frequently accompanies these systemic inflammatory states. To explore the relationship between *FGF6* expression and iron deposition in autoimmune tissues, *FGF6* expression and iron deposition in the skin lesions from SSc patients and healthy controls were examined. We found significantly decreased FGF-6 protein by immunohistochemistry assay (**Figure 5A**) and elevated iron deposition in SSc skin tissue by ferrozine assay (**Figure 5B**), especially in the epidermis. Increased iron deposition was confirmed by Perls stain in SSc skin tissues (supplemental Figure 10A). We also investigated the relationship between FGF6 protein expression with iron deposition in liver cancer tissues. We found that FGF-6 was significantly decreased in nonmetastatic cancer lesion tissues (**Figure 5C**) and the increased iron deposition (**Figure 5D**; supplemental Figure 10B). However, increased FGF-6 expression was observed in metastatic liver carcinoma tissue (supplemental Figure 11), suggesting that FGF-6 plays different roles in oncogenesis and metastasis, analogous to transforming growth factor β.[51,52]

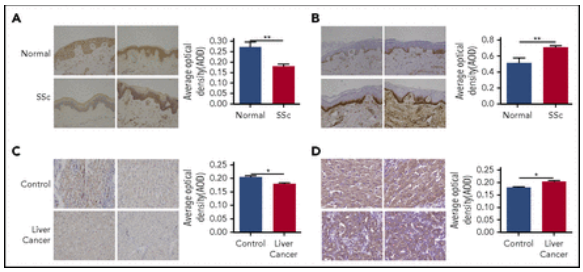

Figure 5. Perls stain and ferritin expression. (A) FGF-6 protein level was evaluated by IHC assay (IHC) in skin tissues from SSc patients and healthy controls (Normal). Staining was visualized by Nikon microscopy; original magnification ×200. A Student *t* test was used to test for pairwise differences between AOD values between SSc and normal observations. The ratio of positive stain areas to the total area was used to evaluate protein levels. AODs were quantified by ImageJ software. ***P* < .01. (B) IHC with Perls Prussian Blue stain for ferritin protein was applied to evaluate the iron deposition in SSc skin tissues and healthy skin tissue. AOD values were quantified by ImageJ software. Staining was visualized by Nikon microscopy; original magnification ×200. ***P* < .01. (C) IHC of FGF-6 protein in liver cancer tissue and control tissue. AODs were quantified by ImageJ software. Staining was visualized by Nikon microscopy; original magnification ×200. **P* < .05. (D) IHC of ferritin protein using Perls Prussian Blue stain in liver cancer tissue and control tissue. AODs were quantified by ImageJ software. Staining was visualized by Nikon microscopy; original magnification ×200. **P* < .05.

Discussion

Iron homeostasis results from a combination of pathways and 4 main cell types: enterocyte, hepatocyte, macrophage, and erythroblast. The epidermal growth factor/epidermal growth factor receptor signaling pathway, heme production, STAT signaling, cyclic adenosine monophosphate signaling, ferritin storage, and bone morphogenetic protein–SMAD signaling are all involved in iron regulation. We conducted an exome-wide, gene-based recessive diplotype scan using putative functional variants to reveal additional genes underlying hemochromatosis susceptibility, an approach that can be widely applied to investigate complex disease susceptibility generated by compound heterozygosity and recessive single-site effects using existing exome-wide association genotype and sequencing data. Although the case sample size was very small, this novel scan identified *FGF6* as being significantly associated with hemochromatosis following correction for multiple testing. *FGF6* belongs to the paracrine FGF gene family and is largely expressed in skeletal muscle, which plays an important role in iron metabolism as it contains 10% to 15% of iron stores. We conducted the evolutionary analysis of *FGF6*and known iron-metabolism genes including *FGFR1*, *TFRC*, *FTH1*, *IREB1*, *TF*, *HMOX1*, *ACO2*, and *HAMP* (encoding hepcidin). The appearance of iron-metabolism genes can be separated into 2 stages. *TF* and *HMOX1*, which are found in animals from *Caenorhabditis elegans* to *Homo sapiens*, indicate an origin in early Bilateria evolution (~635 million years ago [Mya]). *FGF6*, *FGFR1*, *ACO2*, and *HAMP* can be found from *Danio rerio* to *H sapiens*, but are not present in *C elegans* and *Drosophila*, indicating emergence in early Vertebrata (~485 Mya). The coappearance of these genes suggests possible coregulatory functions (supplemental Figure 4A). Functional experiments demonstrated that FGF-6 strongly impacted hepcidin expression, thereby playing a role in regulation of iron homeostasis. These results suggest FGF-6 mediates its effect on iron metabolism via hepcidin. The induction of hepcidin expression by FGF-6 leads to degradation of ferroportin through binding and internalization of ferroportin by hepcidin as shown in **Figure 6**. We additionally found that 3 *FGF6* nonsynonymous variants increased intracellular iron concentrations and reduced hepcidin levels compared with WT *FGF6*, indicating loss of function. Interestingly, a genome-wide RNA interference–profiling study reported that knockdown of *FGF6* increased transferrin-mediated endocytosis.[53]Rs12368351, ~8 kb downstream of *FGF6*, has been associated with phosphorus levels[54]; and 2 SNPs, rs140668749 and rs10849061, within 20 kb downstream of *FGF6*, are associated with migraine.[55,56] Previous studies have indicated that iron plays a role in autoimmunity and a study examining pulmonary arterial hypertension in SSc

noted iron deposition in lung elastin fibers and giant cells,[57] however, epidermal iron deposition in SSc has not been previously investigated. We observed that FGF-6 is involved with iron deposition in SSc and liver cancer. Together, these results demonstrate that FGF receptor (FGFR) signaling through FGF-6 is a critically important mechanism in iron metabolism.

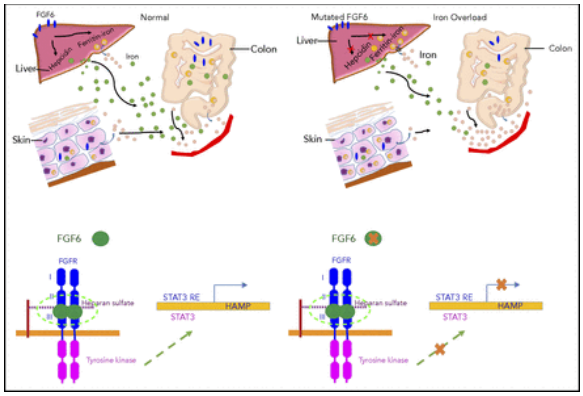

Figure 6. The proposed mechanism of FGF-6 in the regulation of hepcidin expression and iron concentrations. Paracrine FGF-6 interacts with FGFR with heparin or heparin sulfate proteoglycan (HPSG) as the cofactor to initial FGF pathway.[59] Activated FGFRs have the ability to phosphorylate specific tyrosine residues and activate STAT3 pathway.[60] Iron overload and inflammation could positively regulate hepcidin by BMP/Smad pathway[61] and inflammatory IL-6/STAT3 pathways.[62,63]However, loss-of-function *FGF6* variants will silence the FGF6-FGFR pathway, increase free heparin, and reduce expression of hepcidin, thereby decreasing the inhibition of ferroportin-mediated iron transfer from the intracellular compartment to the blood (ie, increasing plasma levels of iron). In SSc patients, IL-6 is increased so that hepcidin will be positively regulated which suppresses iron release to the plasma generating higher iron levels in skin cells.

Acknowledgments

On 15 April 2014, 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). Marshfield Clinic received a Certificate of Confidentiality from the National Institutes of Health.

This work was supported by the National Institutes of Health Clinical and Translational Science Award program, previously through National Center for Research Resources grant 1UL1RR025011 and National Center for Advancing Translational Sciences grant 9U54TR000021, and now by National Center for Advancing Translational Sciences grant UL1TR000427. This work was also funded by the National Natural Science Foundation of China (31521003, 81770066) and the 111 Project (B13016). Additional funding was provided by Marshfield Clinic Research Institute grant SCH10218 and generous donors to the Marshfield Clinic.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authorship

Contribution: S.G. performed analyses, interpreted results, designed the functional experiments, and aided in drafting the manuscript; S.J. conducted molecular and cell biology experiments; N.E. interpreted results, provided hematology and pathway expertise, and reviewed and edited the manuscript; M.M. aided in the analyses, and with the experimental design, and reviewed the manuscript; M.W., Y.M., and W.W. provided clinical and biochemistry advice and aided in drafting the manuscript; Z.Y. performed initial genetic and statistical analyses, managed data, and reviewed the manuscript; B.O. implemented and refined the phenotyping algorithms; T.K. and J.J. aided in the regulatory paperwork and reviewed the manuscript; P.A. and

F.W. generated data and aided with experiments; R.S. performed data management tasks; J.J.M. provided clinical advice and reviewed the manuscript; J.K.M. supervised the management of biological samples for genotyping and reviewed the manuscript; L.J. reviewed the manuscript and provided general scientific advice; J.A.S. provided molecular and cellular biology advice as well as clinical advice and reviewed and edited the manuscript; J.W. supervised the functional experiments, reviewed the manuscript, and provided biological advice; and S.J.S. designed the study, supervised the genetic analyses, developed phenotyping algorithms, developed analysis methods and power calculations, interpreted results, and aided in drafting and editing the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Jiucun Wang, School of Life Sciences, Fudan University, Shanghai, China; e-mail: jcwang@fudan.edu.cn; or Steven J. Schrodi, Center for Human Genetics, Marshfield Clinic Research Institute, 1000 N Oak Ave–MLR Marshfield, WI 54449; e-mail: schrodi.steven@mcrf.mfldclin.edu or schrodi@wisc.edu.

Footnotes

Submitted 9 October 2018; accepted 20 February 2019. Prepublished online as Blood First Edition paper, 27 February 2019; DOI 10.1182/blood-2018-10-879585.

*S.G. and S.J. contributed equally to this work.

The authors uploaded the analysis code to github: https://github.com/

Shicheng-Guo/marshfield/blob/master/2ALOF/readme.md.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

REFERENCES

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

2. MacArthur DG, Balasubramanian S, Frankish A, et al. ; 1000 Genomes Project Consortium. A systematic survey of loss-of-function variants in human protein-coding genes. Science. 2012;335(6070):823-828.
3. Zou J, Valiant G, Valiant P, et al. Quantifying unobserved protein-coding variants in human populations provides a roadmap for large-scale sequencing projects. Nat Commun. 2016;7(1):13293.
4. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science. 2004;305(5685):869-872.
5. Andreoletti G, Shakhnovich V, Christenson K, et al. Exome analysis of rare and common variants within the NOD signaling pathway. Sci Rep. 2017;7(1):46454.
6. De Braekeleer M, Allard C, Leblanc JP, Simard F, Aubin G. Genotype-phenotype correlation in cystic fibrosis patients compound heterozygous for the A455E mutation. Hum Genet. 1997;101(2):208-211.
7. Prietsch V, Mayatepek E, Krastel H, et al. Mevalonate kinase deficiency: enlarging the clinical and biochemical spectrum. Pediatrics. 2003;111(2):258-261.
8. Thein SL. Genetic modifiers of beta-thalassemia. Haematologica. 2005;90(5):649-660.
9. Bauer P, Knoblich R, Bauer C, et al. NPC1: complete genomic sequence, mutation analysis, and characterization of haplotypes. Hum Mutat. 2002;19(1):30-38.

10. Singh T, Kurki MI, Curtis D, et al. ; UK10 K Consortium. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci. 2016;19(4):571-577.
11. Adam R, Spier I, Zhao B, et al. Exome sequencing identifies biallelic MSH3 germline mutations as a recessive subtype of colorectal adenomatous polyposis. Am J Hum Genet. 2016;99(2):337-351.
12. Hague S, Rogaeva E, Hernandez D, et al. Early-onset Parkinson's disease caused by a compound heterozygous DJ-1 mutation. Ann Neurol. 2003;54(2):271-274.
13. Onoufriadis A, Simpson MA, Pink AE, et al. Mutations in IL36RN/IL1F5 are associated with the severe episodic inflammatory skin disease known as generalized pustular psoriasis. Am J Hum Genet. 2011;89(3):432-437.
14. Dewey FE, Murray MF, Overton JD, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science. 2016;354(6319).
15. Adams PC, Barton JC. Haemochromatosis. Lancet. 2007;370(9602):1855-1860.
16. Andrews NC, Schmidt PJ. Iron homeostasis. Annu Rev Physiol. 2007;69(1):69-85.
17. Saddi R, Feingold J. Idiopathic haemochromatosis: an autosomal recessive disease. Clin Genet. 1974;5(3):234-241.
18. Borecki IB, Rao DC, Yaouanq J, Lalouel JM. Segregation of genetic hemochromatosis indexed by latent capacity of transferrin. Am J Hum Genet. 1989;45(3):465-470.
19. Simon M, Alexandre JL, Bourel M, Le Marec B, Scordia C. Heredity of idiopathic haemochromatosis: a study of 106 families. Clin Genet. 1977;11(5):327-341.
20. Cartwright GE, Skolnick M, Amos DB, Edwards CQ, Kravitz K, Johnson A. Inheritance of hemochromatosis: linkage to HLA. Trans Assoc Am Physicians. 1978;91:273-281.
21. Edwards CQ, Griffen LM, Dadone MM, Skolnick MH, Kushner JP. Mapping the locus for hereditary hemochromatosis: localization between HLA-B and HLA-A. Am J Hum Genet. 1986;38(6):805-811.
22. Jazwinska EC, Lee SC, Webb SI, Halliday JW, Powell LW. Localization of the hemochromatosis gene close to D6S105. Am J Hum Genet. 1993;53(2):347-352.
23. Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet. 1996;13(4):399-408.
24. Jazwinska EC, Cullen LM, Busfield F, et al. Haemochromatosis and HLA-H. Nat Genet. 1996;14(3):249-251.
25. Griffiths W, Cox T. Haemochromatosis: novel gene discovery and the molecular pathophysiology of iron metabolism. Hum Mol Genet. 2000;9(16):2377-2382.
26. Allen KJ, Gurrin LC, Constantine CC, et al. Iron-overload-related disease in HFE hereditary hemochromatosis. N Engl J Med. 2008;358(3):221-230.
27. Milet J, Dehais V, Bourgain C, et al. Common variants in the BMP2, BMP4, and HJV genes of the hepcidin regulation pathway modulate HFE hemochromatosis penetrance. Am J Hum Genet. 2007;81(4):799-807.
28. Benyamin B, Esko T, Ried JS, et al. ; InterAct Consortium. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis [published correction appears in Nat Commun. 2015;6:6542]. Nat Commun. 2014;5:4926.
29. de Tayrac M, Roth MP, Jouanolle AM, et al. Genome-wide association study identifies TF as a significant modifier gene of iron metabolism in HFE hemochromatosis. J Hepatol. 2015;62(3):664-672.
30. McCarty CA, Wilke RA, Giampietro PF, Wesbrook SD, Caldwell MD. Marshfield Clinic Personalized Medicine Research Project (PMRP): design, methods and recruitment for a large population-based biobank. Per Med. 2005;2(1):49-79.
31. Carter TC, Rein D, Padberg I, et al. Validation of a metabolite panel for early diagnosis of type 2 diabetes. Metabolism. 2016;65(9):1399-1408.
32. Ye Z, Vasco DA, Carter TC, Brilliant MH, Schrodi SJ, Shukla SK. Genome wide association study of SNP-, gene-, and pathway-based approaches to identify genes influencing susceptibility to Staphylococcus aureus infections. Front Genet. 2014;5:125.

33. Ritchie MD, Denny JC, Zuvich RL, et al. ; Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) QRS Group. Genome- and phenome-wide analyses of cardiac conduction identifies markers of arrhythmia risk. Circulation. 2013;127(13):1377-1385.
34. Fritsche LG, Igl W, Bailey JN, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nat Genet. 2016;48(2):134-143.
35. Grassmann F, Kiel C, Zimmermann ME, et al. ; International AMD Genomics Consortium (IAMDGC). Genetic pleiotropy between age-related macular degeneration and 16 complex diseases and traits. Genome Med. 2017;9(1):29.
36. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007;81(5):1084-1097.
37. Choi Y, Chan AP, Kirkness E, Telenti A, Schork NJ. Comparison of phasing strategies for whole human genomes. PLoS Genet. 2018;14(4):e1007308.
38. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
39. Ma C, Blackwell T, Boehnke M, Scott LJ; GoT2D investigators. Recommended joint and meta-analysis strategies for case-control association testing of single low-count variants. Genet Epidemiol. 2013;37(6):539-550.
40. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics. 2016;32(9):1423-1426.
41. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89(1):82-93.
42. Sokal RR, Rohlf FJ. Biometry: The Principles and Practice of Statistics in Biological Research. 3rd ed. San Francisco, CA: W.H. Freeman; 1995.
43. Sanjak JS, Long AD, Thornton KR. A model of compound heterozygous, loss-of-function alleles is broadly consistent with observations from complex-disease GWAS datasets. PLoS Genet. 2017;13(1):e1006573.
44. dos Reis M, Thawornwattana Y, Angelis K, Telford MJ, Donoghue PC, Yang Z. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr Biol. 2015;25(22):2939-2950.
45. Riemer J, Hoepken HH, Czerwinska H, Robinson SR, Dringen R. Colorimetric ferrozine-based assay for the quantitation of iron in cultured cells. Anal Biochem. 2004;331(2):370-375.
46. Healy S, McMahon JM, FitzGerald U. Modelling iron mismanagement in neurodegenerative disease in vitro: paradigms, pitfalls, possibilities & practical considerations. Prog Neurobiol. 2017;158:1-14.
47. Recalcati S, Locati M, Gammella E, Invernizzi P, Cairo G. Iron levels in polarized macrophages: regulation of immunity and autoimmunity. Autoimmun Rev. 2012;11(12):883-889.
48. Bowlus CL. The role of iron in T cell development and autoimmunity. Autoimmun Rev. 2003;2(2):73-78.
49. Dixon SJ, Lemberg KM, Lamprecht MR, et al. Ferroptosis: an iron-dependent form of nonapoptotic cell death. Cell. 2012;149(5):1060-1072.
50. Torti SV, Torti FM. Iron and cancer: more ore to be mined. Nat Rev Cancer. 2013;13(5):342-355.
51. Derynck R, Akhurst RJ, Balmain A. TGF-beta signaling in tumor suppression and cancer progression [published correction appears in Nat Genet. 2001;29(3):351]. Nat Genet. 2001;29(2):117-129.
52. Wakefield LM, Roberts AB. TGF-beta signaling: positive and negative effects on tumorigenesis. Curr Opin Genet Dev. 2002;12(1):22-29.
53. Collinet C, Stöter M, Bradshaw CR, et al. Systems survey of endocytosis by multiparametric image analysis [published correction appears in Nature. 2014;513(7518):444]. Nature. 2010;464(7286):243-249.
54. Kanai M, Akiyama M, Takahashi A, et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat Genet. 2018;50(3):390-400.

55. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits [published correction appears in Nat Genet. 2016;48(10):1296]. Nat Genet. 2016;48(7):709-717.
56. Anttila V, Winsvold BS, Gormley P, et al. ; UK Brain Expression Consortium. Genome-wide meta-analysis identifies new susceptibility loci for migraine. Nat Genet. 2013;45(8):912-917.
57. Overbeek MJ, Vonk MC, Boonstra A, et al. Pulmonary arterial hypertension in limited cutaneous systemic sclerosis: a distinctive vasculopathy. Eur Respir J. 2009;34(2):371-379.
58. Li Y, Sun C, Yates EA, Jiang C, Wilkinson MC, Fernig DG. Heparin binding preference and structures in the fibroblast growth factor family parallel their evolutionary diversification. Open Biol. 2016;6(3):150275.
59. Itoh N, Ornitz DM. Fibroblast growth factors: from molecular evolution to roles in development, metabolism and disease. J Biochem. 2011;149(2):121-130.
60. Turner N, Grose R. Fibroblast growth factor signalling: from development to cancer. Nat Rev Cancer. 2010;10(2):116-129.
61. De Domenico I, Ward DM, Kaplan J. Hepcidin regulation: ironing out the details. J Clin Invest. 2007;117(7):1755-1758.
62. Bartnikas TB, Fleming MD. A tincture of hepcidin cures all: the potential for hepcidin therapeutics. J Clin Invest. 2010;120(12):4187-4190.
63. Wrighting DM, Andrews NC. Interleukin-6 induces hepcidin expression through STAT3. Blood. 2006;108(9):3204-3209.