Assessing the genetic relationships between osteoarthritis and human plasma proteins: a large scale genetic correlation scan

Li Liu#, Sen Wang#, Yan Wen, Ping Li, Shiqiang Cheng, Mei Ma, Lu Zhang, Bolun Cheng, Xin Qi, Chujun Liang, Feng Zhang

Key Laboratory of Trace Elements and Endemic Diseases of National Health Commission, School of Public Health, Health Science Center, Xi’an Jiaotong University, Xi’an, China

Contributions: (I) Conception and design: Y Wen, F Zhang; (II) Administrative support: S Cheng, X Qi; (III) Provision of study materials or patients: F Zhang; (IV) Collection and assembly of data: S Wang, F Zhang; (V) Data analysis and interpretation: L Liu, C Liang, P Li, C Bolun, L Zhang, M Ma; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

Correspondence to: Feng Zhang. School of Public Health, Health Science Center, Xi’an Jiaotong University, Xi’an 710061, China. Email: fzhxjtu@mail.xjtu.edu.cn.

Background: Osteoarthritis (OA) is a multifactorial complex disease. The impact of plasma proteins on OA remains elusive now.

Methods: The UK Biobank genome-wide association study data of OA was used here. Genome-wide SNP genotyping was performed using the Affymetrix UK BiLEVE Axiom or UK Biobank Axiom array. Equally, the GWAS summary data of 3,622 plasma proteins was derived from a recently published study. Consequently, linkage disequilibrium score regression (LD score regression) analysis was performed to evaluate the genetic correlation between each plasma protein and different sites of OA.

Results: Several suggestive plasma proteins were identified for OA. For hand OA, evidence of genetic correlation was observed for inter-alpha-trypsin inhibitor heavy chain H1 (coefficient =−0.3854, P value =0.0198), multiple inositol polyphosphate phosphatase 1 (coefficient =0.0172, P value =0.0303). For hip OA, 7 suggestive genetic correlation signals were observed, such as Transmembrane glycoprotein NMB (coefficient =0.6944, P value =0.0098), Endothelial cell-specific molecule 1 (coefficient =0.6337, P value =0.03). For Knee OA, 12 suggestive genetic correlation signals were identified, including Elafin (coefficient =−0.5562, P value =0.0092), Interleukin-16 (coefficient =0.3949, P value =0.0435).

Conclusions: We investigated the genetic correlations between plasma proteins and different sites of OA in a systematic way. Our results provide novel evidence that OA is a heterogeneous disease.

Keywords: Osteoarthritis (OA); human plasma proteins; linkage disequilibrium score regression; pQTL

Submitted Dec 25, 2019. Accepted for publication Apr 27, 2020.
doi: 10.21037/atm-19-4643

View this article at: http://dx.doi.org/10.21037/atm-19-4643

Introduction

Osteoarthritis (OA) is a degenerative and chronic joint disease, mainly characterized by the breakdown of articular cartilage and underlying bone. OA was estimated to affect more than 20.7 million people in the United States, and to make the leading cause of disability across the whole world (1). The knee, hip, and hand are the areas most affected by OA, ultimately leading to joint failure with pain, stiffness and disability (2). The health economic burden of OA is increasing due to the current rapid ageing and increasing obesity evident in recent years, especially in the Asian region (3). Accumulated studies showed that genetic factors contributed greatly to the pathogenesis and disease progression of OA (4,5). Classic twin studies indicated primary OA had a strong...
hereditary component (5), with heritability estimating 30–46% according to the different sites (6). However, the genetic mechanism of OA is complicated and remain largely unknown now.

Human plasma proteins have been demonstrated to be involved in many biological processes, including signaling, transport, repair, and defense against infection (7). They play a major role as signals or biological markers in disease diagnosis as well as therapeutic monitoring (8). Recently, the impact of plasma proteins on the development of OA has also been reported. Researchers found multiple plasma proteins, which could activate macrophages via the innate immune receptor TLR-4, contributing to the low-level grade inflammation in OA patients (9,10). Another study found that human C-reactive protein-transgenic mice developed more severe OA compared with wild-type controls under the same high-fat diet (HFD) regime, implicating the role of CRP as a trigger to aggravate HFD-induced OA (11). Screening of the plasma proteome is considered as a promising approach for the identification of metabolic and disease processes. However, limited efforts were paid to explore the functional relevance of plasma proteome with OA. Recently, Sun et al. conducted a large scale GWAS of human plasma proteome in 3,301 healthy participants, and identified 1,927 genotype-protein associations (pQTLs), providing new clues for understanding of the genetic regulation of plasma proteins (7).

Genetic correlation serves as a critical population parameter in describing the genetic correlation of complex traits and diseases. Linkage disequilibrium score regression (LD score regression) is one of the most popular approach for genetic correlation scan. A far more important use of LD Score regression is the estimation and correction of confounding, for which is robust to the confounding biases caused by cryptic relatedness and population stratification (12). If the genetic correlation is statistically and quantitatively significant, then total phenotypic correlations are not entirely attributable to confounders of complete environmental nature (13). The more genetic variation an index variant tags, the higher the probability that this index variant will tag a causal variant (14). The great power of LD score regression has been demonstrated in previous studies. For example, Treur et al. identified the genetic correlation between smoking behaviors and schizophrenia via comparing the genome-wide association study (GWAS) meta-analysis results from five smoking phenotypes and the GWAS results of schizophrenia using LD score regression (15).

In this study, utilizing the large scale GWAS datasets of OA and plasma proteome, for the first time, we conducted a genetic correlation scan of OA and plasma proteome to explore the functional relevance of plasma proteins with OA. Our study holds the potential to provide new insights into the genetic mechanism of OA. We present the following article in accordance with the STREGA reporting checklist (available at http://dx.doi.org/10.21037/atm-19-4643).

Methods

The GWAS summary data of hip OA and knee OA

The GWAS summary data sets of hip OA and knee OA were obtained from the published studies (16). Briefly, 2,396 hospital-diagnosed hip OA and 9,593 controls, 4,462 hospital-diagnosed knee OA and 17,885 controls were included in the GWAS, separately. The genotyping of samples was conducted either using the UK BiLEVE array or the UK Biobank Axiom array. Imputation was performed using UK10K/1000 Genomes Project haplotype reference panel by the IMPUTE3 program. After quality control, there are 16,122,076 SNPs for hip OA and 16,309,199 SNPs for knee OA were used in following analysis, respectively. Detailed description of sample characteristics, experimental design, and statistical analysis can be shown in the published study (16).

The GWAS summary data of hand OA

The GWAS summary of hand OA was derived from the UK Biobank (UK Biobank fields: 20002). Briefly, a total of 37,782 hand OA cases and 414,482 controls in UK Biobank participants of European ancestry were enrolled in our study. The UK Biobank participants were genotyped using the Affymetrix UK BiLEVE Axiom or UK Biobank Axiom array and augmented by imputation of ~90 million genetic variants from the Haplotype Reference Consortium, 1,000 Genomes and UK 10K projects (17). After filtering, this dataset contains 9,113,133 imputed variants. Detailed information about samples, imputation, genotyping can be found in the published study (18).

The GWAS data of human plasma proteome

The GWAS summary data of human plasma proteome...
was obtained from the published study (7). Briefly, a total of 3,622 plasma proteins were quantified in 3,301 healthy participants from 25 centers across England (19). Genotyping was performed using Affymetrix Axiom UK Biobank genotyping array (Santa Clara, California, USA). Totally, 654,966 high-quality variants were used for imputation. Using a combined 1,000 Genomes Phase 3-UK10K reference panel, imputation was performed via the Sanger Imputation Server (https://imputation.sanger.ac.uk) and resulted in 87,696,888 imputed variants.

**Genetic correlation analysis**

Following the standard approach recommended by the developers, LD score regression software (https://github.com/bulik/ldsc) was used here to evaluate the genetic correlation between OA and each of the human plasma proteins. LD score regression is capable of estimating genetic correlation by using GWAS summary statistics rather than individual level genotype data (14,20,21). The idea of LD score regression is that, for a polygenic trait, LD to an annotation that is enriched for heritability will increase the $\chi^2$ statistic of a SNP more than LD to an annotation that is not enriched for heritability. And SNPs with lower LD have significantly larger per-SNP heritability. Consequently, the LD score regression is capable of distinguishing genuine polygenicity from the confounding bias (such as cryptic relatedness and population stratification) and more powerful than $\lambda_{GC}$, especially at large sample sizes (14). After correcting for multiple testing, the significant threshold of this study should be $P<1.38\times10^{-4}$.

**Results**

For GWAS analysis of OA, the significant SNPs associated with hand OA, hip OA and knee OA were summarized in (Figure 1A, Table 1), such as Elafin (coefficient =-0.5562, P value =0.0092), Interleukin-16 (coefficient =0.3949, P value =0.0435), leukotriene A-4 hydrolase (coefficient =0.8366, P value =0.0114), muscle and skeletal receptor tyrosine-protein kinase (coefficient =0.434, P value =0.0219), fructose-1,6-bisphosphatase 1 (coefficient =0.8139, P value =0.0344).

**Genetic correlation analysis between hip OA and plasma proteome**

For hip OA, we also detected 7 suggestive plasma proteins (Figure 1B, Table 1), such as transmembrane glycoprotein NMB (coefficient =0.6944, P value =0.0098), endothelial cell-specific molecule 1 (coefficient =0.6337, P value =0.03) and Cytochrome P450 3A4 (coefficient =-0.8474, P value =0.0357).

**Genetic correlation analysis between hand OA and plasma proteome**

For hand OA, 7 suggestive plasma proteins were detected (Figure 1C, Table 1), such as inter-alpha-trypsin inhibitor heavy chain H1 (coefficient =-0.3854, P value =0.0198), multiple inositol polyphosphate phosphatase 1 (coefficient =-1.1721, P value =0.0303), protein kinase C-binding protein NELL1 (coefficient =-0.7416, P value =0.0427), neurexophilin-3 (coefficient =0.717, P value =0.0241), neurotrophin-4 (coefficient =0.722, P value =0.0292).

**Discussion**

Recent studies observed the impact of plasma proteins on the development of OA. However, there was no systematic study online investigating the genetic correlation among plasma proteins and different sites of OA. In this study, utilizing the large scale GWAS datasets of OA and plasma proteome, LD score regression scan was conducted to investigate the functional relevance between OA and plasma proteome. We identified several candidate plasma proteins for hip OA, knee OA and hand OA, respectively. Our study results hold great potential in promoting the understanding of the genetic pathophysiology of OA and paving the way towards the development of effective therapies against OA given the different genetic mechanisms of different sites.

Inter-alpha-trypsin inhibitor heavy chain H1 showed the strongest correlation evidence with hand OA in this study. Inter-alpha-trypsin inhibitor heavy chain H1
Figure 1  Plot of genetic correlation analysis results of OA and Human Plasma proteins suggestive correlations are labelled with different colors. Each dot denotes a plasma protein. The X-axis denotes the plasma proteins. The Y-axis denotes the negative of P values. (A) Plot of genetic correlation analysis results of knee OA and 1,259 Human Plasma proteins. The 12 suggestive correlations are found. Among them, Elafin (coefficient =−0.5562, P value =0.0092) showed the most suggestive signal. (B) Plot of genetic correlation analysis results of Hip OA and 1,231 Human Plasma proteins. The 7 suggestive correlations are found. Transmembrane glycoprotein NMB (coefficient =0.6944, P value =0.0098) is the most suggestive one. (C) Plot of genetic correlation analysis results of Hand OA and 1,272 Human Plasma proteins. The 7 suggestive correlations are labelled with different colors. Inter-alpha-trypsin inhibitor heavy chain H1 (coefficient =−0.3854, P value =0.0198) is the most suggestive one.

(ITIH1), is encoded by ITIH1 gene, acting as a carrier of hyaluronan in serum or as a binding protein between hyaluronan and other matrix protein. ITIH1 has been defined as an important protein biomarker for OA in the previous studies (22). An existing study revealed the protein level of ITIH1 in serum was significantly elevated in OA patients compared with the control subjects (23). A large scale GWAS of OA identified ITIH1 as a susceptibility gene for OA (24), which was consistent with our finding.

Protein kinase C-binding protein, also known as NELL-like protein 1 (NELL1) or Nel-related protein 1, was identified for hand OA in our study. The important role of NELL-1 in driving differentiation and growth of bone and cartilage tissue has been well documented in recent years (25). A recent study showed that adding NELL1 to the human perivascular stem cells (hPSCs) could increase the gene and protein expression of chondrogenic markers, suggesting the role of NELL1 in enhancing and accelerating cartilage repair (26). Another study suggested that NELL1 was a novel osteogenic factor that transduced the osteogenic signals through the activation of certain Tyr-kinases, leading to the osteogenic differentiation (27). Desai, Shannon et al. observed that NELL1-deficient mice showed reduced expression of extracellular matrix proteins (ECM), cell adhesion and communication proteins, causing cranial and vertebral defects (28). Another animal study suggested the significant improvement of NELL1 in treating diverse pathologies of cartilage defects and degeneration (29).

Transmembrane glycoprotein NMB (GPNMB), a protein involved in the bone formation, showed the strongest correlation evidence with Hip OA. GPNMB
### Table 1  The suggestive genetic correlations between different sites of OA and Human Plasma Protein (significant P value <1.38×10^{-5})

<table>
<thead>
<tr>
<th>OA sites</th>
<th>Plasma protein</th>
<th>Genes</th>
<th>Genetic correlation</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand OA</td>
<td>Inter-alpha-trypsin inhibitor heavy chain H1</td>
<td>ITIH1</td>
<td>−0.3854</td>
<td>0.0198</td>
</tr>
<tr>
<td></td>
<td>Neurexophilin-3</td>
<td>NXPH3</td>
<td>0.717</td>
<td>0.0241</td>
</tr>
<tr>
<td></td>
<td>Neurotrophin-4</td>
<td>NTF4</td>
<td>0.722</td>
<td>0.0292</td>
</tr>
<tr>
<td></td>
<td>Multiple inositol polyphosphate phosphatase 1</td>
<td>MINP1</td>
<td>−1.1721</td>
<td>0.0303</td>
</tr>
<tr>
<td></td>
<td>NA</td>
<td>IgG4, Kappa</td>
<td>0.9761</td>
<td>0.0312</td>
</tr>
<tr>
<td></td>
<td>Protein O-linked-mannose beta-1,4-N-acetylglucosaminyltransferase 2</td>
<td>POMGNT2</td>
<td>−0.768</td>
<td>0.0385</td>
</tr>
<tr>
<td></td>
<td>Protein kinase C-binding protein NELL1</td>
<td>NELL1</td>
<td>−0.7416</td>
<td>0.0427</td>
</tr>
<tr>
<td>Hip OA</td>
<td>Transmembrane glycoprotein NMB</td>
<td>GPNMB</td>
<td>0.6944</td>
<td>0.0098</td>
</tr>
<tr>
<td></td>
<td>Protein FAM163A</td>
<td>FAM163</td>
<td>0.6769</td>
<td>0.0251</td>
</tr>
<tr>
<td></td>
<td>Endothelial cell-specific molecule 1</td>
<td>ESM1</td>
<td>0.6337</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Trypsin-2</td>
<td>PRSS2</td>
<td>−0.4843</td>
<td>0.0317</td>
</tr>
<tr>
<td></td>
<td>Pulmonary surfactant-associated protein C</td>
<td>SFTPC</td>
<td>0.5255</td>
<td>0.0344</td>
</tr>
<tr>
<td></td>
<td>Cytochrome P450 3A4</td>
<td>CYP3A4</td>
<td>−0.8474</td>
<td>0.0357</td>
</tr>
<tr>
<td></td>
<td>A disintegrin and metalloproteinase with thrombospondin motifs 1</td>
<td>ADAMTS1</td>
<td>0.9372</td>
<td>0.05</td>
</tr>
<tr>
<td>Knee OA</td>
<td>Elafin</td>
<td>PI3</td>
<td>−0.5562</td>
<td>0.0092</td>
</tr>
<tr>
<td></td>
<td>Leukotriene A-4 hydrolase</td>
<td>LTA4H</td>
<td>0.8366</td>
<td>0.0114</td>
</tr>
<tr>
<td></td>
<td>Muscle, skeletal receptor tyrosine-protein kinase</td>
<td>MUSK</td>
<td>0.434</td>
<td>0.0219</td>
</tr>
<tr>
<td></td>
<td>Fructose-1,6-bisphosphatase 1</td>
<td>FBP1</td>
<td>0.8139</td>
<td>0.0344</td>
</tr>
<tr>
<td></td>
<td>NADPH--cytochrome P450 reductase</td>
<td>POR</td>
<td>0.5025</td>
<td>0.0355</td>
</tr>
<tr>
<td></td>
<td>Vesicle transport protein USE1</td>
<td>USE1</td>
<td>−0.5008</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>Creatine kinase M-type</td>
<td>CKM</td>
<td>0.8022</td>
<td>0.0375</td>
</tr>
<tr>
<td></td>
<td>Rho-associated protein kinase 2</td>
<td>ROCK2</td>
<td>−0.8409</td>
<td>0.0377</td>
</tr>
<tr>
<td></td>
<td>GDNF family receptor alpha-1</td>
<td>GFRα-1</td>
<td>0.5034</td>
<td>0.0399</td>
</tr>
<tr>
<td></td>
<td>Transmembrane gamma-carboxyglutamic acid protein 1</td>
<td>PRRG1</td>
<td>−0.6206</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>Interleukin-16</td>
<td>IL-16</td>
<td>0.3949</td>
<td>0.0435</td>
</tr>
<tr>
<td></td>
<td>Tuftelin</td>
<td>TUFT1</td>
<td>−0.7689</td>
<td>0.05</td>
</tr>
</tbody>
</table>

was found to expressed in various cell types, including melanocytes, osteoclasts, osteoblasts, dendritic cells. A previous study reported that the expressional level of GPNMB is significantly higher in OA cartilage compared with normal donor cartilage (4). Moreover, Jiang et al. found the expressional level of GPNMB was associated with bone loss and the pathology of OA (30). GPNMB is expressed in normal articular chondrocytes, and expression level is higher in damaged cartilage, suggesting the anti-inflammatory role of GPNMB in OA. Furthermore, data also suggested that GPNMB inhibits pERK and pNFκB-65 proteins through the CD44 receptor (31). GPNMB was also known as osteoactivin due to the role in osteoblast differentiation and bone remodeling. Previous literatures have shown that osteoactivin (OA/GPNMB) inhibited RANKL-induced osteolysis in vivo and this process was
regulated through CD44 and ERK activation (32). The same conclusion about the role of osteoactivin as a positive regulator of osteoblastogenesis was also drawn by another finding (33).

Another notable protein identified in hip OA was a disintegrin and metalloproteinase with thrombospondin motifs 1 (ADAMTS1). ADAMTS1, a member of the ADAMTS family of proteases, is a multifunctional metalloproteinase necessary for the development and function of a variety of organs and pathological processes. ADAMTS families, including ADAMTS4, ADAMTS5, ADAM9, ADAM10, ADAM12 and so on, are found to be expressed in cartilage and several of them showed significantly altered expression in OA (34). Wachsmuth et al. confirmed the expression and presence of ADAMTS1 in normal and OA human cartilage, and found it was the first matrix-degrading enzyme downregulated by the catabolic factor IL-1 beta (IL-1β) in vitro (35). Interestingly, different conclusions were also drawn by some other studies. Data from previous study suggested there was no increase in expression of ADAMTS1 after destabilization of the medial meniscus (DMM) in mice (36). A comparative analysis found that ADAMTS1 was downregulated in both cartilage and synovial tissues of OA compared with the fracture to the neck of femur samples (37). Further biological researches on the specific role of ADAMTS1 on OA are needed.

As is known to all, OA is a highly heterogeneous disease, which can also be seen from the clinical presentation. Several stratification methods of OA have been reported, including mechanism of onset, stage of disease progression, involved joints, inflammatory levels, and effusion (38). Specifically, OA is stratified by several subgroups considering the involved joints, such as generalized OA (39), bilateral OA, specific location OA (e.g., hand, hip, knee, and ankle) (16,40). Genetic evidence also confirms the heterogeneity of OA. Previous studies have identified many risk loci for OA (41). Nevertheless, most of these significant variants were not shared among different OA sites (42). Another study conducted in 992 monozygotic and dizygotic female twin participants showed that genetic influence on OA differs by anatomical site, indicating OA is a joint-specificity disease (43). Except from genetic loci, the causal effects of some phenotypes on OA are also site-specific, including lumbar spine bone mineral density and height (16).

Many reasons accounted for these observations. For example, studies showed there were higher serum concentrations of cartilage degradation markers [e.g., cartilage oligomeric matrix protein (COMP), hyaluronan (HA)] while lower serum concentrations of cartilage formation markers (e.g., a collagen II synthesis marker) in patients with hand OA compared with other joints OA (44,45). In addition, bone shape might be a risk factor for the occurrence or development of OA (46). Havercamp et al. investigated the role of joint shape in knee OA by using a statistical shape model, it turned out that knees with OA were wider, more extended during radiography, and had an elevated lateral tibial plateau (47). This kind change of bone shape could serve as a biomarker to predict the development or progression of knee OA (47). Furthermore, there might be histological differences in articular cartilage between joints. Biochemical properties of chondrocytes differed from knee and ankle joints, the ankle chondrocytes synthesized more proteoglycans (PGs) and collagens than knee chondrocytes (48).

In our study, we performed a LD score regression analysis to explore the association between different sites of OA and plasma proteins via repurposing the previous GWAS data of OA and pQTL data of human plasma proteome. We found 12, 7, 7 suggestive proteins associated with knee OA, hip OA and hand OA, respectively. In line with previous studies, there was no common plasma protein shared among these three sites of OA. Our results provide novel evidence that OA is a heterogeneous disease. In addition, the association of plasma proteins and OA was discussed from the perspective of genetics, which was not easily affected by the environmental confounding factors. Moreover, the large sample sizes of the GWAS data ensured the accuracy of our study results, which were consistent with previous study results. Finally, given the great power of LD score regression in estimating genetic correlation, our finding holds great potential in providing novel clues for the genetic mechanism of OA.

One issue should be noted in our study. After correcting for multiple testing, the significant threshold should be P<1.38×10^-6. Unfortunately, there was no significant genetic correlation under this threshold based on our results. Since all the plasma proteins identified in the present study were suggestive, our study results should be interpreted with caution. Additionally, given the complexity of OA, we cannot sure the specific role of these identified plasma proteins in disease progression, which needs more mechanism-based experiments to further confirm the biological rationality and to clarify the biological mechanism of the identified plasma proteins involving in the development of OA.
analysis with larger, and more homogeneous populations may help untangle the genetic architecture of OA. In our future study, we will conduct further studies to explore the functional roles of these proteins in the development of OA via biological functional analysis such as cells as well as animal experiments. Then we can confirm whether these proteins play a direct or indirect role in the development of OA.

Taken together, we investigated the genetic correlations between plasma proteins and different sites of OA in a systematic way by using the large summary GWAS datasets, respectively. And we found several suggestive plasma proteins on the different sites of OA in our study. Our results provide a novel clue for validating the heterogeneity of OA.

Acknowledgments

Funding: This study is supported by the National Natural Scientific Foundation of China (81472925, 81673112), the Key projects of international cooperation among governments in scientific and technological innovation (2016YFE0119100), the Natural Science Basic Research Plan in Shaanxi Province of China (2017JZ024), and the Fundamental Research Funds for the Central Universities.

Footnote

Reporting Checklist: The authors have completed the STREGA reporting checklist. Available at http://dx.doi.org/10.21037/atm-19-4643.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-19-4643). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

14. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from