European Spine Journal

Integrative Analysis of Metabolomic, Genomic, and Imaging-based Phenotypes Identify Very-Low-Density Lipoprotein as a Potential Risk Factor for Lumbar Modic Changes

--Manuscript Draft--

Manuscript Number:	
Full Title:	Integrative Analysis of Metabolomic, Genomic, and Imaging-based Phenotypes Identify Very-Low-Density Lipoprotein as a Potential Risk Factor for Lumbar Modic Changes
Article Type:	ISSLS Prize
Keywords:	Modic; genetics; metabolomics; disc degeneration; spine
Corresponding Author:	Dino Samartzis, DSc, PhD, (C) MSc Rush University Medical Center Chicago, IL UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Rush University Medical Center
Corresponding Author's Secondary Institution:	
First Author:	Yiming Li
First Author Secondary Information:	
Order of Authors:	Yiming Li
	Jaro Karppinen
	Kathryn S.E. Cheah
	Danny Chan
	Pak Sham
	Dino Samartzis, DSc, PhD, (C) MSc
Order of Authors Secondary Information:	
Funding Information:	Research Grants Council, University Grants Committee (776613) — Dr. Dino Samartzis
	AOSpine (SZ) — Dr. Dino Samartzis
Abstract:	Purpose: Modic changes (MC) on MRI have been associated with the development and severity of LBP. The etiology of MC remains elusive, but it has been suggested that altered metabolism may be a risk factor. As such, this study aimed to identify metabolomic biomarkers for MC phenotypes of the lumbar spine via a combined metabolomic-genomic approach. Methods: A population cohort of 3,584 southern Chinese underwent lumbar spine MRI. Blood samples were genotyped with SNP arrays (n=2,482) and serum metabolomics profiling using magnetic resonance spectroscopy (n=757), covering 130 metabolites representing three molecular windows, were assessed. Genome-wide association studies (GWAS) were performed on each metabolite, to construct polygenic scores for predicting metabolite levels in subjects who had GWAS but not metabolomic data. Associations between predicted metabolite levels and MC phenotypes were assessed using linear/logistic regression and LASSO. Two-sample Mendelian randomization analysis tested for causal relationships between metabolic biomarkers and MC. Results: 20.4% had MC (10.6% type 1, 67.2% type 2, 22.2% mixed types). Significant MC metabolomic biomarkers were mean diameter of very-low-density lipoprotein (VLDL)/low-density lipoprotein (LDL) particles and cholesterol esters/phospholipids in

	large LDL. Mendelian randomization indicated that decreased VLDL mean diameter may lead to MC. Conclusions: This large-scale study is the first to address metabolomics in subject with/without lumbar MC. Causality studies implicate VLDL related to MC, noting a metabolic etiology. Our study substantiates the field of “spino-metabolomics” and illustrates the power of integrating metabolomics-genomics-imaging phenotypes to discover biomarkers for spinal disorders, paving the way for more personalized spine care for patients.

Integrative Analysis of Metabolomic, Genomic, and Imaging-based Phenotypes Identify Very-Low-Density Lipoprotein as a Potential Risk Factor for Lumbar Modic Changes

Authors:
[1] Yiming Li, PhD
[2] Jaro Karppinen, MD, PhD
[3] Kathryn S.E. Cheah, PhD
[3] Danny Chan, PhD
[1,4] Pak C. Sham, MBBS, PhD
[5,6,7] Dino Samartzis, DSc

Affiliations:
[1] Department of Psychiatry, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[2]Medical Research Center Oulu, University of Oulu and Oulu University Hospital, Oulu, Finland
[3] School of Biomedical Sciences, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[4] Centre for PanorOmic Sciences, State Key Laboratory of Brain and Cognitive Sciences, Pokfulam, Hong Kong SAR, China
[5] Department of Orthopaedics and Traumatology, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[6] International Spine Research and Innovation Initiative, Rush University Medical Center, Chicago, IL, USA
[7] Department of Orthopaedic Surgery, Rush University Medical Center, Chicago, IL, USA

Disclosures: The authors have no financial or competing interests related to this work.

Keywords: Modic; genetics; metabolomics; disc degeneration; spine; phenotyping; MRI

Correspondence: **Dino Samartzis**
Department of Orthopaedic Surgery
Rush University Medical Center
1611 W. Harrison Street
Chicago, IL 60612
USA
Email: Dino_Samartzis@rush.edu

Pak Sham
Centre for PanorOmic Sciences
State Key Laboratory of Brain and Cognitive Sciences
The University of Hong Kong
21 Sassoon Road, Pokfulam
Hong Kong SAR, China
Email: pcsham@hku.hk

Integrative Analysis of Metabolomic, Genomic, and Imaging-based Phenotypes Identify Very-Low-Density Lipoprotein as a Potential Risk Factor for Lumbar Modic Changes

ABSTRACT

Purpose: Modic changes (MC) on MRI have been associated with the development and severity of LBP. The etiology of MC remains elusive, but it has been suggested that altered metabolism may be a risk factor. As such, this study aimed to identify metabolomic biomarkers for MC phenotypes of the lumbar spine via a combined metabolomic-genomic approach.

Methods: A population cohort of 3,584 southern Chinese underwent lumbar spine MRI. Blood samples were genotyped with SNP arrays (n=2,482) and serum metabolomics profiling using magnetic resonance spectroscopy (n=757), covering 130 metabolites representing three molecular windows, were assessed. Genome-wide association studies (GWAS) were performed on each metabolite, to construct polygenic scores for predicting metabolite levels in subjects who had GWAS but not metabolomic data. Associations between predicted metabolite levels and MC phenotypes were assessed using linear/logistic regression and LASSO. Two-sample Mendelian randomization analysis tested for causal relationships between metabolic biomarkers and MC.

Results: 20.4% had MC (10.6% type 1, 67.2% type 2, 22.2% mixed types). Significant MC metabolomic biomarkers were mean diameter of very-low-density lipoprotein (VLDL)/low-density lipoprotein (LDL) particles and cholesterol esters/phospholipids in large LDL. Mendelian randomization indicated that decreased VLDL mean diameter may lead to MC.

Conclusions: This large-scale study is the first to address metabolomics in subject with/without lumbar MC. Causality studies implicate VLDL related to MC, noting a metabolic etiology. Our study substantiates the field of "spino-metabolomics" and illustrates the power of integrating metabolomics-genomics-imaging phenotypes to discover biomarkers for spinal disorders, paving the way for more personalized spine care for patients.

INTRODUCTION

Low back pain (LBP) is the world's most debilitating condition. [1] Lumbar disc degeneration is an important factor associated with LBP and has also been a precursor for other spinal phenotypes, such as disc herniation, endplate abnormalities and Modic changes (MC).[2-6] Throughout the years, many etiological factors for disc degeneration have been proposed, including age progression, abnormal physical loading, environmental determinants, hormonal influences, genetics, and circulatory or metabolic disease. [7-13]

The role of improper vascular supply or metabolism in the pathway of disc degeneration has been under heated debate. Although the disc in adults is one of the largest avascular structures in the human body, it receives nutrients and metabolites from diffusion from the vascular plexuses, through the endplate, surrounding the disc. Studies have noted that altered metabolism (e.g. oxygen, glucose etc.) may affect the normal integrity of the disc by affecting proteoglycan synthesis, thereby contributing to degenerative changes.[14,15] In addition, occlusion or insufficient arterial blood supply to the lumbar spine due to improper endplate permeability or vascular disorder may have direct implications upon the integrity of the disc, leading to degeneration. [9] Deficiencies in nutrient supply reduce the number of viable disc cells and their metabolic activity,[16] which may limit the ability of the disc to recover from an injury. In fact, studies noting elevated body mass index (BMI), increased lipid and apolipoprotein-E levels, Vitamin C deficiency, and increased lactate production may play a role in disc degeneration and LBP,[11-13,17-20] further providing credence of a potential metabolic component to the development of spine changes.

Modic changes are subchondral bone marrow lesions adjacent to the endplate and are commonly observed alongside disc degeneration **(Figure 1).** Modic changes also have a high specificity for discogenic LBP,[21] and LBP patients with MC have longer, more frequent and more painful LBP episodes and seek care more often than patients without MC. [22-25] Studies have also noted that pain and disability may exist, irrespective of MC type.[4] Elevated BMI has been associated with MC, perhaps serving as a surrogate of more complex metabolic disruption.[24] Genetic risk factors have also been linked with MC but remain questionable and validation in different ethnicities remains.[26] Infection, lifestyle and environmental factors have also been suggested to be associated with MC but fail to be replicated because the studies have limitations in the study design, analyses and/or narrow understanding of the phenotype.[27,28] As such, the etiology and pathobiology of MC remains under heated debate. Therefore, studying the underlying biological mechanism of MC will help determine their role in LBP.[29]

Genomics, transcriptomics and proteomics represent the genome-oriented main disciplines in life sciences. The information in genomes operates across three diverse time spans: evolution (millions of years), development (hours to years), and physiology (milliseconds to weeks). Physiology constitutes the triggering of specific functional pathways by environmental signals, and thereby the phenotype of a biological system is largely reflected by its metabolite composition and their interactions. An essential and complementary "omics" approach in understanding biomolecular function is **metabolomics** – the quantitative measurement of the time-related multiparametric metabolic responses of multicellular systems to (patho)physiological stimuli or genetic modification.[30-32] Metabolomics has shed light and directed new targeted therapeutics for various conditions, such as Alzheimer's disease, cardiovascular disease, diabetes and knee osteoarthritis (OA).[30,31,33] To date, no study has addressed the metabolomic risk factors of spine degeneration, especially in the context of MC.

The following study addressed the possible link between underlying metabolomic changes with MC using a purely data-driven approach. Through a set of integrative analyses of genomic, metabolomic and phenotypic data from a large population-based cohort, we assessed the association between metabolomic features, predicted via a novel polygenic scoring scheme, and MC phenotypes. In addition, we performed Mendelian randomization, which allows to establish "causation" of an exposure (metabolite) to the outcome (MC). Our main goal was to identify potential metabolomic biomarkers related to the etiology of MC, which could aid personalized diagnosis and targeted treatment of LBP in future.

METHODS

Study sample

The Hong Kong Disc Degeneration Cohort Study [4-6,11,12,24] is a population-based initiative consisting of 3,584 Southern Chinese volunteers at baseline (age range: 10-88 years) whereby subjects were recruited by open invitation and not based on their pain profiles. Individuals with a known history of spinal tumor, infection or deformities were excluded. Age, sex, weight (kg), height (m) and cigarette smoking pack-years were noted of all subjects. Body mass index (kg/m^2) was calculated of each individual. The different stages of this study are illustrated in **Figure 2**.

Imaging assessment

All subjects underwent 3T MRI of L1-S1, which was performed using axial T1-weighted and sagittal T2-weighted fast spin-echo sequences (repetition time = 3,000 milliseconds; echo time = 92 milliseconds; slice thickness = 5 millimeters). [4-6,11,12,24] Utilizing the second follow-up images of the longitudinal assessment of the cohort subjects since baseline, two experienced individuals assessed the lumbar region of the MRI scans, blinded to the clinical data. The first rater evaluated the presence/absence of MC at each lumbar level (L1 to L5) of 1,416 cohort participants. The second examined the lumbar MC types (types 1, 2, 3 and mixed types) of 1,713 volunteers assessed at a later time point than the first rater whereby the image availability of the cohort had increased. Reliability scores of our rating method have been previously reported and were good to excellent. [4,5]

For each individual, we defined 5 composite MC phenotypes based on the MRI readings of these raters. Three composite scores were defined utilizing the first rater's readings. The original MC readings were first transformed into continuous scores using truncated normal conversion, and overall MC was defined as the sum of MC scores at all levels, upper MC sum scores (L1-L3) and lower MC sum scores (L4 –S1). Two binary composite phenotypes of type 1 and 2 MC were also utilized.

Metabolomic measurements

On January 1, 2010, a cardiovascular assessment of all cohort subjects ≥40 years old (n=1,800 out of N=3,584 overall subjects) from the cohort was initiated. Around the same period of imaging, blood serum samples of approximately 50 ul were obtained, and stored at -80 degrees Celsius, from 814 subjects during the morning fasting state for metabolomic analysis. Metabolomic profiling of 130 metabolic measures was conducted on three platforms: lipoprotein lipids (LIPO), low molecular weight metabolites (LMWM) and lipid extracts (LIPID) **(Supplement A).** Proton (1H) serum NMR spectroscopy, via a Bruker AVANCE 500 DRX spectrometer operating at 500.13 MHz machine, analyzed the samples. Among the 814 volunteers with serum samples collected, 757 had complete metabolomic measurements (i.e. no missing data). After data filtering, data normalization (sample-wise: normalization by sum; measurement-wise: auto-scaling) was performed using MetaboAnalyst.[34]

Genotyping

The blood samples of 2,482 subjects were obtained for genotyping. Deoxyribonucleic acid (DNA) was extracted from the blood samples and underwent concentration quality control. The DNA samples were genotyped using Illumina's OmniZhongHua-8 BeadChip, which is a population-specific whole genome array covering 77% of common variation (minor allele frequency, or MAF >

5%), 73% of intermediate variation (MAF > 2.5%) as well as 65% of low frequency variation (MAF > 1%) in the Chinese population. In total, 900,015 single nucleotide polymorphisms (SNPs) on chromosomes 1 to 22 and chromosome X were genotyped. The raw data was converted to PLINK[35] format using Illumina's GenomeStudio.

GWAS and polygenic scoring of metabolomic measures

To guarantee the quality of samples used in genome-wide association studies (GWAS), we filtered out 369 low-quality samples with high SNP missing rates, sample mislabeling, gender inconsistencies, sample contamination or high relatedness to other samples. After conducting quality control (QC), variant QC was performed. 88,625 SNPs of poor quality were removed, including those of low call rates, low minor allele frequencies (MAF) and deviating from the Hardy-Weinberg equilibrium (HWE). The GWAS QC was performed using PLINK.[35] After QC, the genotype data contained 2,113 individuals and 805,525 SNPs.

In PLINK, we performed 130 genome-wide association analyses (one for each metabolomic measurement) on the quality-controlled genotype data of the 571 individuals with both genomic and metabolomic data. Sex, age and the first ten principal components (PC's) calculated using EIGENSTRAT [36] were considered as covariates. To increase power and reduce false positives, meta-analysis of our GWAS was next performed in PLINK using a random effect approach for the 116 metabolomic measurements present in both studies.[37]

The summary statistics from meta-analysis were then used for polygenic scoring through the standard pipeline of Lassosum.[38] Only chromosomes 1 to 22 were used in polygenic risk score (PRS) construction, and pseudo-validation was performed to select the best set of parameters for Lassosum when calculating the PRS. Since GWAS loci for metabolomic traits typically have a large effect size, we could treat the metabolomic PRS as "predicted" metabolomic measurements.[39,40]

Regression analysis of Modic changes (MC) phenotypes - standard

A set of linear and logistic regression analyses were conducted in order to look into the relationship between metabolomic polygenic risk scores (PRS) and the five MC composite phenotypes. The composite scores consisted of the following: continuous - overall MC (sum of all disc MC scores); upper MC (sum MC scores L1-L4); lower MC (sum MC scores L4-S1); binary - presence of type 1 MC; presence type 2 MC. To avoid over-fitting, we excluded the samples used in previous GWAS. In total, 580 regression models (5 MC phenotypes * 116 metabolomic PRS) were fitted, and the issue of multiple testing arose. Certain MC phenotypes may have a number of associated metabolites with relatively low, but insignificant p-values. To take advantage of this group

information, we grouped the resulting p-values by MC phenotype and controlled the false discovery rate (FDR) at 0.1 through the adaptive group Benjamini-Hochberg (B-H) procedure.

Regression analysis of Modic changes (MC) phenotypes - penalized

Five LASSO models (each corresponding to one of the 5 MRI composite phenotypes) were fitted using the R package "glmnet". The LASSO tuning parameter was selected via ten-fold cross validation — the loss used for cross validation was squared error for Gaussian models (i.e. continuous phenotypes) and deviance for logistic models (i.e. binary phenotypes). To avoid over-fitting, the models were fitted using a subset of individuals with both GWAS and MC phenotypic data, but no metabolomic data.

Mendelian randomization analysis

To test for a potential causal relationship between the *metabolomic measurement (exposure) and MC phenotype (outcome),* pairs with significant associative findings were identified. We further performed two-sample Mendelian randomization analysis with metabolomic measurements as the exposure and composite MC phenotypes as the outcome **Figure 3.**

First, we performed GWAS on all the individuals with MRI (outcome), BMI (confounder) or smoking (confounder) and genetic data. Age, sex and the first ten PCs were selected as the covariates for all the studies. To avoid bias due to participant overlap in two- sample Mendelian randomization, we excluded the subjects who were among the 571 individuals in the metabolomic GWAS. Following, for each exposure-case pair with significant associative findings, we retrieved all the GWAS hits ($p<5x10^{-8}$) of the metabolomic measurement. These SNPs were clumped in PLINK using the default parameters of the R package TwoSampleMR. The physical distance threshold for clumping was set to be 10,000 kb, the LD threshold was chosen to be 0.001, and both the significance threshold for index SNPs and the secondary significance threshold for clumped SNPs was set to be 1. After clumping, we removed the SNPs that were also GWAS hits ($p< 5x10^{-8}$) of the confounders in our model (BMI and smoking), if any. To avoid pleiotropy, we also removed any SNPs significantly associated with the outcome (the MC phenotype) at the significance level of $p=10^{-5}$. If there were <= 1 SNPs (i.e. instruments) left after the previous steps, we dropped the pair *(metabolomic measurement, MC phenotype)* from subsequent analysis.

For the remaining *(metabolomic measurement, MC phenotype)* pairs, we performed two-sample Mendelian randomization with the R package TwoSampleMR, utilizing five different methods — inverse variance weighted (IVW), Egger regression, weighted median, simple mode and

weighted mode. To control for multiple testing, the FDR, stratified by the Mendelian randomization method used, was controlled at 0.2.

RESULTS

The mean age of the subjects was 50 years and mean BMI was 22.9 kg/m^2. The majority were female (59.7%) and the prevalence of MC was 20.4% (10.6% type 1, 67.2% type 2, 0% type 3, 22.2% mixed types), mainly occurring at L4-S1.

Regression modeling and polygenic risk score (PRS)

The MC scoring scheme was based on types 1 and 2 and not on mixed types since a component was to determine biomarker association with specific individual types and mixed types prevented that distinction. Out of the 580 standard regression models fitted, only one model had a significant b_3 at the FDR cut-off of 0.1 – as the PRS of the mean "diameter" of very-low-density lipoprotein particles (VLDL.D) decreased, MC in the lower lumbar region ($b_3 \approx -0.1148$; q-value ≈ 0.0396) became more severe on average.

The LASSO coefficients (β's) of the metabolomic PRS with a non-zero β in at least one of the fitted LASSO models were plotted in **Figure 4.** The negative influence of VLDL.D PRS on MC, which we identified through standard regression studies, was supported by the LASSO results (overall MC: $\beta \approx -0.0387$; lower MC: $\beta \approx -0.1118$; type 2 MC: $\beta \approx -0.0469$). Similarly, the PRS of the mean "diameter" of low-density lipoprotein (LDL.D)/intermediate-density lipoprotein (IDL.D) particles also had a negative effect on type 2 MC ($\beta \approx -0.1981$).

Additionally, the fitted LASSO models showed that the PRS of cholesterol esters (CE) and phospholipids (PL) in large LDL could potentially lead to overall, lower lumbar region and type 2 MC **(Figure 4).** These findings were consistent with the VLDL.D results, since a low VLDL.D is typically associated with higher VLDL particle levels in patients, and VLDL/LDL levels are closely related to each other.[41] It is worth noting, though, that the PRS of phospholipids (PL) in chylomicrons and extremely large VLDL was negatively associated with lower MC at L4-S1 ($\beta \approx -0.0089$) and type 2 MC ($\beta \approx -0.3265$). This may indicate that PL in large LDL and extremely large VLDL play different roles in the development of MC.

Mendelian randomization

In our MR studies, no significant results were found when using the IVW, weighted median, simple mode and weighted mode approaches. Utilizing the Egger regression method, there were three *(metabolomic measurement, MC phenotype)* pairs with a significant causal effect, as shown in **Figure 5.** We identified a negative causal effect of VLDL.D on overall MC ($\beta \approx -0.4592$, p-value ≈ 0.0499, q-value$\approx$0.1747) and lower MC ($\beta \approx -0.4812$, p-value$\approx$0.0397, q-value$\approx$0.1747) **(Figure 5).** No other *(metabolomic measurement, MC phenotype)* pairs had significant results at 0.2 FDR cut-off.

DISCUSSION

Our large-scale study is, to our knowledge, the first to address the link between metabolomic signatures with degenerative spine pathology, in particular focusing on MC. Employing Mendelian randomization techniques that can establish causation, our study noted a potential metabolic etiology of MC and noted decreased diameter of VLDL as the significant risk factor in that pathway. Secondly, our study is unique in its utilization of an 'integrative" framework" approach analyzing "panOmics data (metabolomics, genomics) and imaging-based phenotype data to discover unique biomarkers related to the spine. This approach has demonstrated significant viability and can serve as a model for any future initiative targeting complex traits, such as degenerative pathology of the spine.

Metabolomics, Human Disease and Clinical Relevance

With respect to the spine, magnetic resonance (MR) spectroscopy has been utilized to assess the disc's metabolic profile.[17] Although findings have been promising, in particular assessing painful discs, this approach can be extremely costly and does not provide an extensive metabolic profile of the disc or systems. In the field of metabolomics, "serum-based" NMR spectroscopy has become a key technology, with validated utility in many diseases, that is specific yet non-selective and cost-effective. Particularly, ^{1}H NMR has the advantage of efficiently obtaining information on large numbers of metabolites in biofluids.[42] As such, a metabolome-wide association study (MWAS) is plausible.

The application of metabolomics has noted promise in elaborating upon Alzheimer's disease,[43] diabetes,[44] and atherosclerosis [45] among other conditions. For example, with regards to Alzheimer's disease, a multi-metabolite assessment has identified that the interaction of inflammation and the metabolic syndrome increases the risk of cognitive decline.[43] In atherosclerosis, metabolomic assessment has targeted metabolic phenotypes that will complement and might even replace

conventional lipid measurements for comprehensive cardiovascular risk assessment.[45] In the context of diabetes, metabolomic longitudinal studies have noted the complex interactions between diabetes, insulin resistance, and the metabolic syndrome.[44] Serum-based NMR spectroscopy was able to identify the polydiagnostic metabolite profile of type 1 diabetes and how its variations may translate to clinical phenotypes, potential vascular complications, and mortality. In knee OA, serum-based metabolomics has found significant analytes related to the presence and severity of OA changes, independent of age and BMI.[33] As such, metabolomics has elaborated upon the understanding of various disease processes and has provided a new perspective upon the assessment of disease with direct clinical implications that would lead to improving health outcomes. Furthermore, the combination of metabolomic data with genome-wide and gene expression information has added clinical value in various diseases.[46] These findings show the importance and opportunity for systematic molecular investigation in human populations.[47]

Very-Low-Density Lipoprotein and its Implications

In this study, we found that as the mean diameter of VLDL particles decreased, the degree of MC was on average more severe. Through Mendelian randomization analysis, we obtained evidence that this association may be causal. VLDL are transporters of triglycerides (synthesized in the liver/intestines) to adipose tissue capillary beds and muscles. At that point, they can be converted to provide fatty acids to ultimately produce adenosine triphosphate (ATP) for cellular energy or they can be converted to glycerol and stored as fat; whereas remaining VLDL can then be converted to LDL. VLDL/LDL have been implicated in arterial plaque build-up, atherosclerosis, ischemic heart disease, macular degeneration, and others.[48,49] A low mean diameter VLDL is typically found in retinopathy, with higher VLDL/LDL particle levels in patients with retinopathy.[41] High amounts of oxidized LDLs activate TLR2/4 (toll-like receptor 2/4), and chronic stimulation of TLRs facilitates fatty marrow conversion as observed in type 2 MC.[29] Mechanistic and functional studies can further elaborate in understanding the underlying pathway of VLDL "particle dimensions" to MC.

In our regression analysis, we also found that high levels of CEs and PLs in large LDL were positively associated with MC. These results were in line with those regarding diameter of VLDL, indicating that both VLDL and LDL lipid levels may be potential biomarkers for MC. An association between atheromatous lesions in the aorta and lumbar disc degeneration has been detected, and LBP has been found to be associated with aortic calcification and stenosis of lumbar arteries.[9,50] Therefore, our results may indicate shared metabolomic components of MC and arterial damage or a disruption in the pathway of triglyceride transport and fat storage that can alter the integrity of the vertebral bone marrow as seen in MC.

Strengths and Limitations

One limitation of our study is sample size – only a modest proportion of the population had metabolomics measurements, limiting the accuracy of the metabolomic PRS derived from GWAS. Nevertheless, by integrating genomic, metabolomic and phenotypic data and improving, our analyses provided insights into the possible underlying metabolomic mechanism of MC. Cohorts with metabolomic data are typically small compared to those with genomic data. Furthermore, it will be important to test the generalizability of our findings in other ethnic populations.

CONCLUSIONS

Our large-scale study has identified a potential metabolic etiology of lumbar MC, noting specific biomarkers. In fact, 130 genome-wide association studies (GWAS) for the metabolomic traits were performed with polygenic scoring noting a link between MC and metabolic traits. Furthermore, Mendelian randomization studies further noted that a decreased mean diameter of VLDL particles may lead to the development of MC. As such, our study, employing metabolic, genomic and imaging phenotyping is the "first" in the spine field to utilize such an integrative approach towards big data. Our large-scale study provides the foundation for new and exciting research into assessing the relationship between our body's metabolic state and pathological changes in the spine. Although further studies are needed, the new field of "spino-metabolomics" as introduced by this study has the potential to contribute to personalized spine care, and in that how it can be used for diagnostic, prognostic and novel therapeutic considerations to treat MC and resulting LBP as well as other spinal disorders.

ACKNOWLEDGEMENTS

This work was supported by grants from the Hong Kong Theme-Based Research Scheme (T12-708/12N), Hong Kong Research Grants Council (776613) and the Hansjorg Wyss Award by AO Spine International. The authors would like to thank Ms. Cora Bow, Yu Pei and Kenneth MC Cheung from the Department of Orthopaedics and Traumatology, Hong Kong for their assistance with the cohort. The authors would also like thank Drs. Mika Ala-Korpela, Pasi Soininen and Antti J Kangas from the University of Eastern Finland, NMR Metabolomics Laboratory, Finland for their assistance in analyzing the blood samples to obtain the metabolomic analytes.

REFERENCES

1. Coste J, Paolaggi JB, Spira A. Reliability of interpretation of plain lumbar spine radiographs in benign, mechanical low-back pain. *Spine* 1991; 16(4): 426-8.

2. de Schepper EI, Damen J, van Meurs JB, et al. The association between lumbar disc degeneration and low back pain: the influence of age, gender, and individual radiographic features. *Spine (Phila Pa 1976)* 2010; 35(5): 531-6.

3. Luoma K, Riihimaki H, Luukkonen R, Raininko R, Viikari-Juntura E, Lamminen A. Low back pain in relation to lumbar disc degeneration. *Spine* 2000; 25(4): 487-92.

4. Maatta JH, Karppinen J, Paananen M, et al. Refined Phenotyping of Modic Changes: Imaging Biomarkers of Prolonged Severe Low Back Pain and Disability. *Medicine (Baltimore)* 2016; 95(22): e3495.

5. Maatta JH, Karppinen JI, Luk KD, Cheung KM, Samartzis D. Phenotype profiling of Modic changes of the lumbar spine and its association with other MRI phenotypes: a large-scale population-based study. *Spine J* 2015; 15(9): 1933-42.

6. Samartzis D, Mok FPS, Karppinen J, Fong DYT, Luk KDK, Cheung KMC. Classification of Schmorl's nodes of the lumbar spine and association with disc degeneration: a large-scale population-based MRI study. *Osteoarthritis Cartilage* 2016; 24(10): 1753-60.

7. Battie MC, Videman T, Gill K, et al. 1991 Volvo Award in clinical sciences. Smoking and lumbar intervertebral disc degeneration: an MRI study of identical twins. *Spine* 1991; 16(9): 1015-21.

8. Eskola PJ, Lemmela S, Kjaer P, et al. Genetic association studies in lumbar disc degeneration: a systematic review. *PLoS One* 2012; 7(11): e49995.

9. Kauppila LI. Atherosclerosis and disc degeneration/low-back pain--a systematic review. *Eur J Vasc Endovasc Surg* 2009; 37(6): 661-70.

10. Li Y, Samartzis D, Campbell DD, et al. Two subtypes of intervertebral disc degeneration distinguished by large-scale population-based study. *Spine J* 2016; 16(9): 1079-89.

11. Samartzis D, Karppinen J, Chan D, Luk KD, Cheung KM. The association of lumbar intervertebral disc degeneration on magnetic resonance imaging with body mass index in overweight and obese adults: a population-based study. *Arthritis Rheum* 2012; 64(5): 1488-96.

12. Samartzis D, Karppinen J, Mok F, Fong DY, Luk KD, Cheung KM. A population-based study of juvenile disc degeneration and its association with overweight and obesity, low back pain, and diminished functional status. *J Bone Joint Surg Am* 2011; 93(7): 662-70.

13. Teraguchi M, Yoshimura N, Hashizume H, et al. Metabolic Syndrome Components Are Associated with Intervertebral Disc Degeneration: The Wakayama Spine Study. *PLoS One* 2016; 11(2): e0147565.

14. Bibby RL, Webster-Brown JG. Characterisation of urban catchment suspended particulate matter (Auckland region, New Zealand); a comparison with non-urban SPM. *Sci Total Environ* 2005; 343(1-3): 177-97.

15. Ohshima H, Urban JP. The effect of lactate and pH on proteoglycan and protein synthesis rates in the intervertebral disc. *Spine (Phila Pa 1976)* 1992; 17(9): 1079-82.

16. Urban JP, Smith S, Fairbank JC. Nutrition of the intervertebral disc. *Spine (Phila Pa 1976)* 2004; 29(23): 2700-9.

17. Keshari KR, Lotz JC, Link TM, Hu S, Majumdar S, Kurhanewicz J. Lactic acid and proteoglycans as metabolic markers for discogenic back pain. *Spine (Phila Pa 1976)* 2008; 33(3): 312-7.

18. Cameron BM, VanderPutten DM, Merril CR. Preliminary study of an increase of a plasma apolipoprotein E variant associated with peripheral nerve damage. A finding in patients with chronic spinal pain. *Spine (Phila Pa 1976)* 1995; 20(5): 581-9; discussion 9-90.

19. VanderPutten DM, Cameron BM, Merril CR. Increased apolipoprotein-E concentrations in individuals suffering chronic low back syndrome identified by two-dimensional gel electrophoresis. *Appl Theor Electrophor* 1993; 3(5): 247-52.

20. Smith VH. Vitamin C deficiency is an under-diagnosed contributor to degenerative disc disease in the elderly. *Med Hypotheses* 2010; 74(4): 695-7.

21. Thompson KJ, Dagher AP, Eckel TS, Clark M, Reinig JW. Modic changes on MR images as studied with provocative diskography: clinical relevance--a retrospective study of 2457 disks. *Radiology* 2009; 250(3): 849-55.

22. Jensen TS, Karppinen J, Sorensen JS, Niinimaki J, Leboeuf-Yde C. Vertebral endplate signal changes (Modic change): a systematic literature review of prevalence and association with non-specific low back pain. *Eur Spine J* 2008; 17(11): 1407-22.

23. Maatta JH, Wadge S, MacGregor A, Karppinen J, Williams FM. ISSLS Prize Winner: Vertebral Endplate (Modic) Change is an Independent Risk Factor for Episodes of Severe and Disabling Low Back Pain. *Spine (Phila Pa 1976)* 2015; 40(15): 1187-93.

24. Mok FP, Samartzis D, Karppinen J, Fong DY, Luk KD, Cheung KM. Modic changes of the lumbar spine: prevalence, risk factors, and association with disc degeneration and low back pain in a large-scale population-based cohort. *Spine J* 2016; 16(1): 32-41.

25. Takatalo J, Karppinen J, Niinimaki J, et al. Association of modic changes, Schmorl's nodes, spondylolytic defects, high-intensity zone lesions, disc herniations, and radial tears with low back symptom severity among young Finnish adults. *Spine (Phila Pa 1976)* 2012; 37(14): 1231-9.

26. Freidin M, Kraatari M, Skarp S, et al. Genome-wide meta-analysis identifies genetic locus on chromosome 9 associated with Modic changes. *J Med Genet* 2019; 56(7): 420-6.

27. Wu J, Huang J, Battie MC, Wang Y. Lifestyle and lifetime occupational exposures may not play a role in the pathogenesis of Modic changes on the lumbar spine MR images. *Spine J* 2020; 20(1): 94-100.

28. Braten LCH, Grovle L, Espeland A, et al. Clinical effect modifiers of antibiotic treatment in patients with chronic low back pain and Modic changes - secondary analyses of a randomised, placebo-controlled trial (the AIM study). *BMC Musculoskelet Disord* 2020; 21(1): 458.

29. Dudli S, Fields AJ, Samartzis D, Karppinen J, Lotz JC. Pathobiology of Modic changes. *Eur Spine J* 2016; 25(11): 3723-34.

30. Bogl LH, Kaye SM, Ramo JT, et al. Abdominal obesity and circulating metabolites: A twin study approach. *Metabolism* 2016; 65(3): 111-21.

31. Karjalainen MK, Holmes MV, Wang Q, et al. Apolipoprotein A-I concentrations and risk of coronary artery disease: A Mendelian randomization study. *Atherosclerosis* 2020; 299: 56-63.

32. Wurtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol* 2017; 186(9): 1084-96.

33. Zhai G, Wang-Sattler R, Hart DJ, et al. Serum branched-chain amino acid to histidine ratio: a novel metabolomic biomarker of knee osteoarthritis. *Annals of the rheumatic diseases* 2010; 69(6): 1227-31.

34. Xia J, Mandal R, Sinelnikov IV, Broadhurst D, Wishart DS. MetaboAnalyst 2.0--a comprehensive server for metabolomic data analysis. *Nucleic Acids Res* 2012; 40(Web Server issue): W127-33.

35. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81(3): 559-75.

36. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38(8): 904-9.

37. Kettunen J, Demirkan A, Wurtz P, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016; 7: 11122.

38. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol* 2017; 41(6): 469-80.

39. Rhee EP, Ho JE, Chen MH, et al. A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab* 2013; 18(1): 130-43.

40. Gieger C, Geistlinger L, Altmaier E, et al. Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS genetics* 2008; 4(11): e1000282.

41. Colhoun HM, Otvos JD, Rubens MB, Taskinen MR, Underwood SR, Fuller JH. Lipoprotein subclasses and particle sizes and their relationship with coronary artery calcification in men and women with and without type 1 diabetes. *Diabetes* 2002; 51(6): 1949-56.

42. Ala-Korpela M. Critical evaluation of 1H NMR metabonomics of serum as a methodology for disease risk assessment and diagnostics. *Clin Chem Lab Med* 2008; 46(1): 27-42.

43. Tukiainen T, Tynkkynen T, Makinen VP, et al. A multi-metabolite analysis of serum by 1H NMR spectroscopy: early systemic signs of Alzheimer's disease. *Biochem Biophys Res Commun* 2008; 375(3): 356-61.

44. Makinen VP, Soininen P, Forsblom C, et al. 1H NMR metabonomics approach to the disease continuum of diabetic complications and premature death. *Mol Syst Biol* 2008; 4: 167.

45. Goonewardena SN, Prevette LE, Desai AA. Metabolomics and atherosclerosis. *Curr Atheroscler Rep* 2010; 12(4): 267-72.

46. Inouye M, Kettunen J, Soininen P, et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol Syst Biol* 2010; 6: 441.

47. Inouye M, Ripatti S, Kettunen J, et al. Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS genetics* 2012; 8(8): e1002907.

48. Wurtz P, Soininen P, Kangas AJ, et al. Characterization of systemic metabolic phenotypes associated with subclinical atherosclerosis. *Mol Biosyst* 2011; 7(2): 385-93.

49. Acar IE, Lores-Motta L, Colijn JM, et al. Integrating Metabolomics, Genomics, and Disease Pathways in Age-Related Macular Degeneration: The EYE-RISK Consortium. *Ophthalmology* 2020.

50. Kauppila LI, McAlindon T, Evans S, Wilson PW, Kiel D, Felson DT. Disc degeneration/back pain and calcification of the abdominal aorta. A 25-year follow-up study in Framingham. *Spine* 1997; 22(14): 1642-7; discussion 8-9.

FIGURE LEGENDS

Figure 1: Sagittal T1- and T2-weighted MRIs of the lumbar spine illustrating Modic changes types 1, 2, and 3.

Figure 2: Flow-chart of the different stages of analyses based on the population-based cohort.
QC: quality control, GWAS: genome-wide association study, SNP: single nucleotide polymorphism, MC: Modic changes

Figure 3: Flow chart of our Mendelian randomization model.
GWAS: genome-wide association study, SNP: single nucleotide polymorphism, MC: Modic changes, BMI: body mass index

Figure 4: Plot of the LASSO coefficients. Of the 130 metabolomic measurements, the plot shows only those with a non-zero coefficient in at least one of the fitted models. Metabolites were clustered using hierarchical clustering based on Kendall's correlation, and the dendrogram was cut into three subtrees by height. Positive (Negative) β's were colored blue (red), and the β's shrunk to zero were colored white. The color was darker when the regression coefficient was of a larger magnitude.

Figure 5: Mendelian randomization scatterplots of the (metabolomic measurement, MC phenotype) pairs with significant Egger regression results. **(A)** The causal effect of VLDL particle diameter on overall MC (regression coefficient, standard error and p-value). **(B)** The causal effect of VLDL particle diameter on lower MC (regression coefficient, standard error and p-value).

SUPPLEMENTAL A

Table: List of 130 metabolic biomarkers assessed in this study.

Figure 1

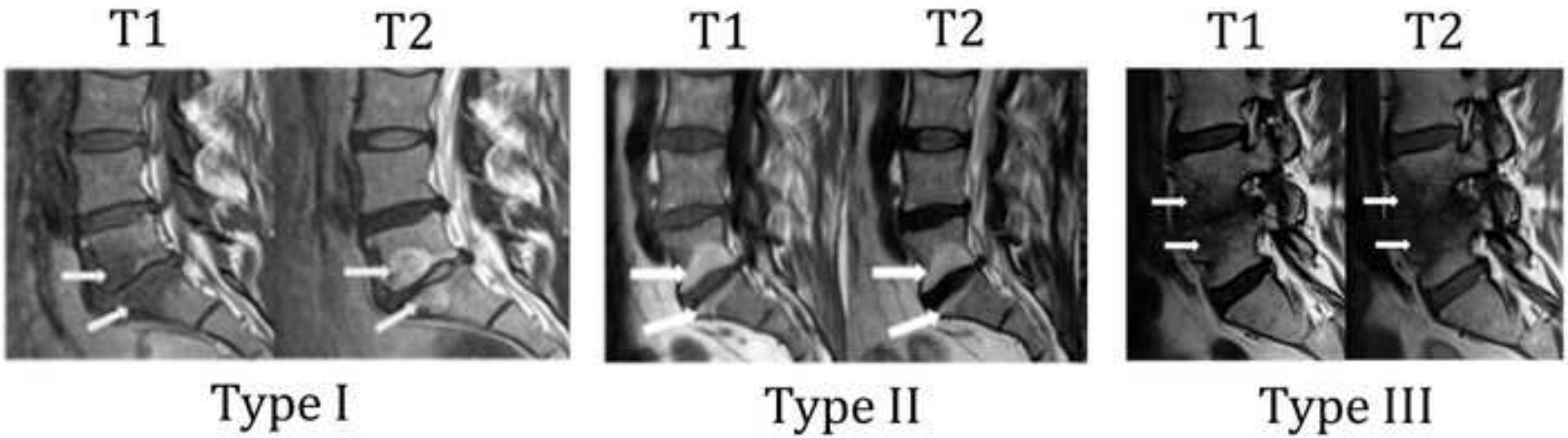


Figure 2

Click here to access/download;Figure;Figure 2 - Flow Chart of Metabolomic Study - UPDATED.jpg

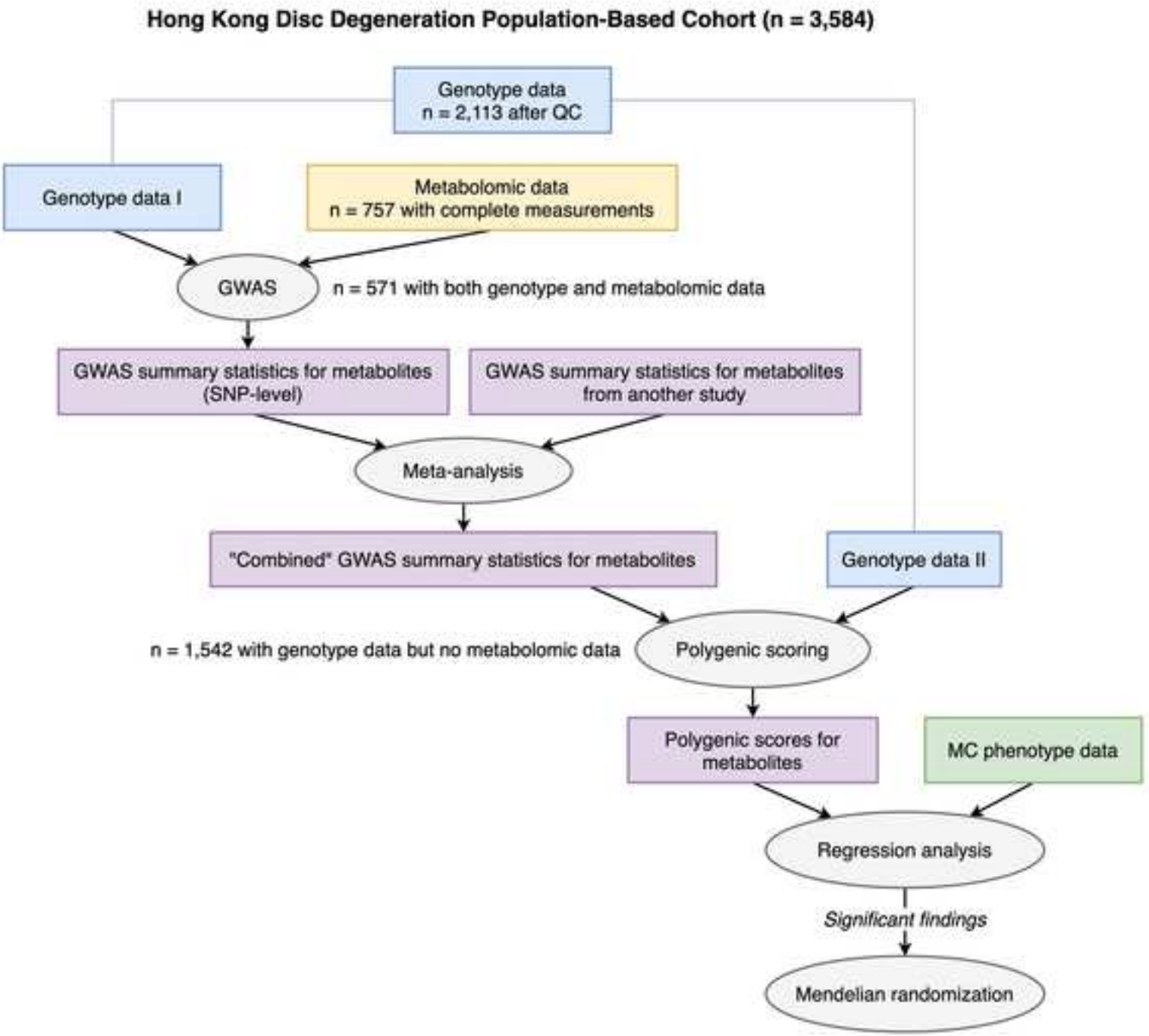


Figure 3

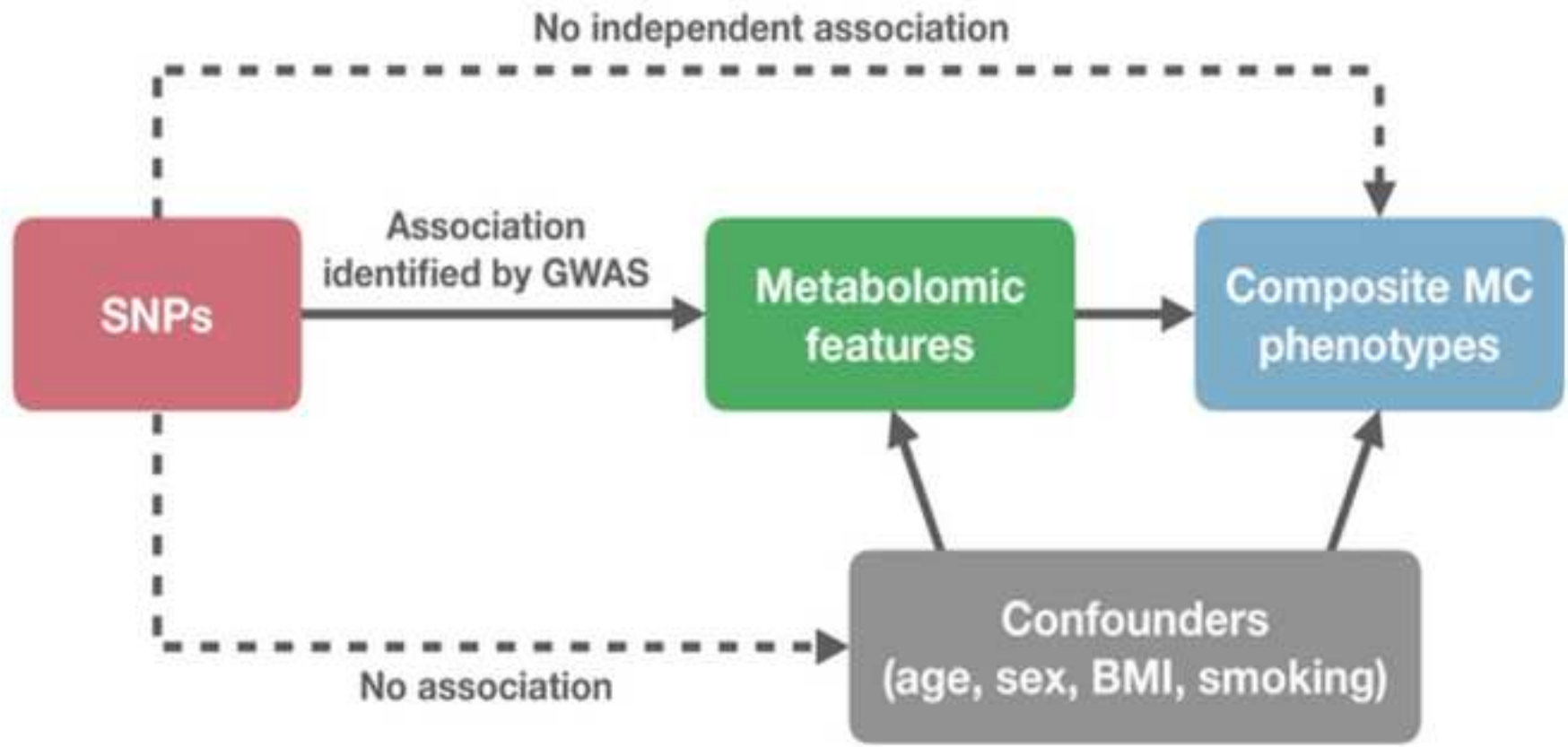


Figure 4

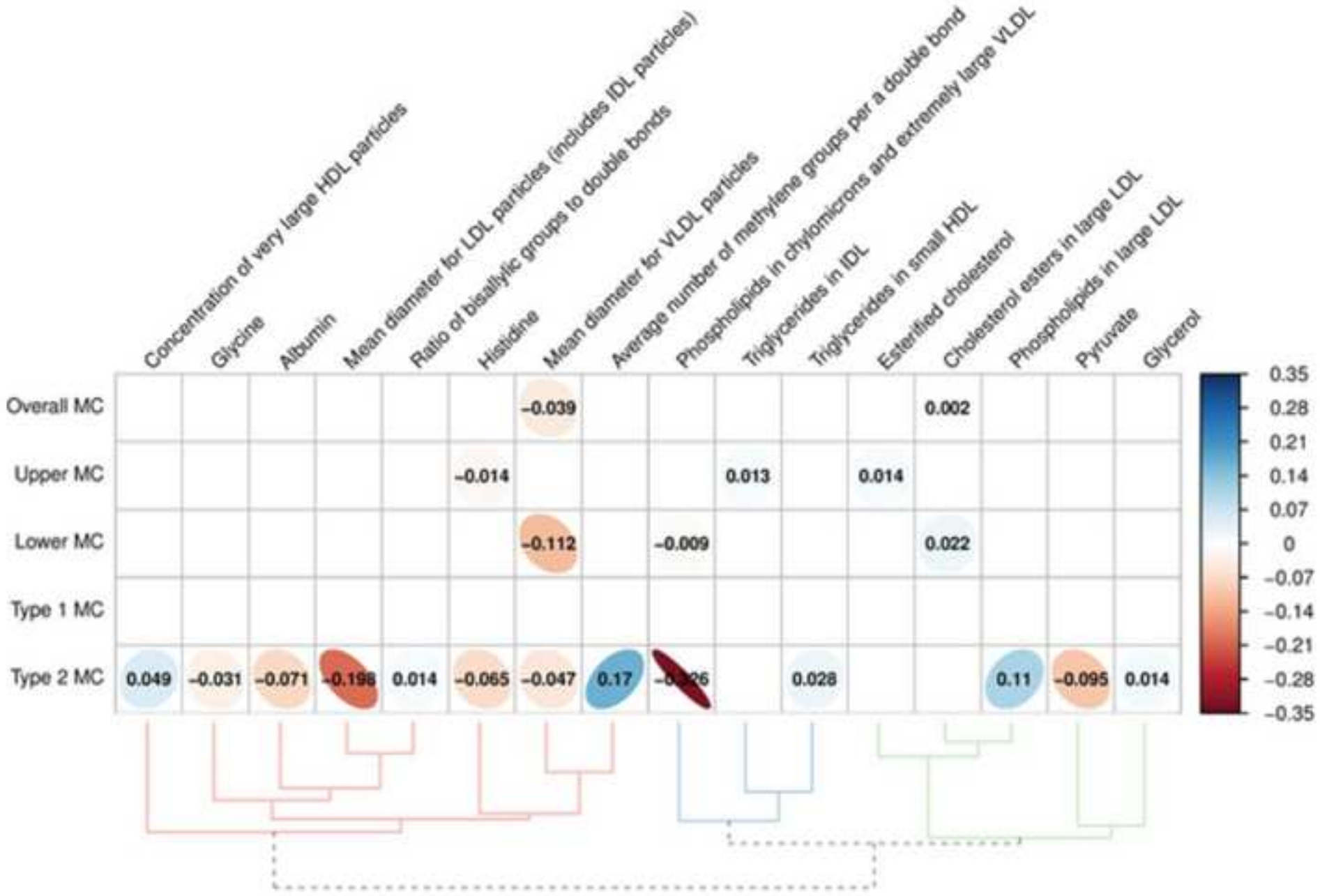
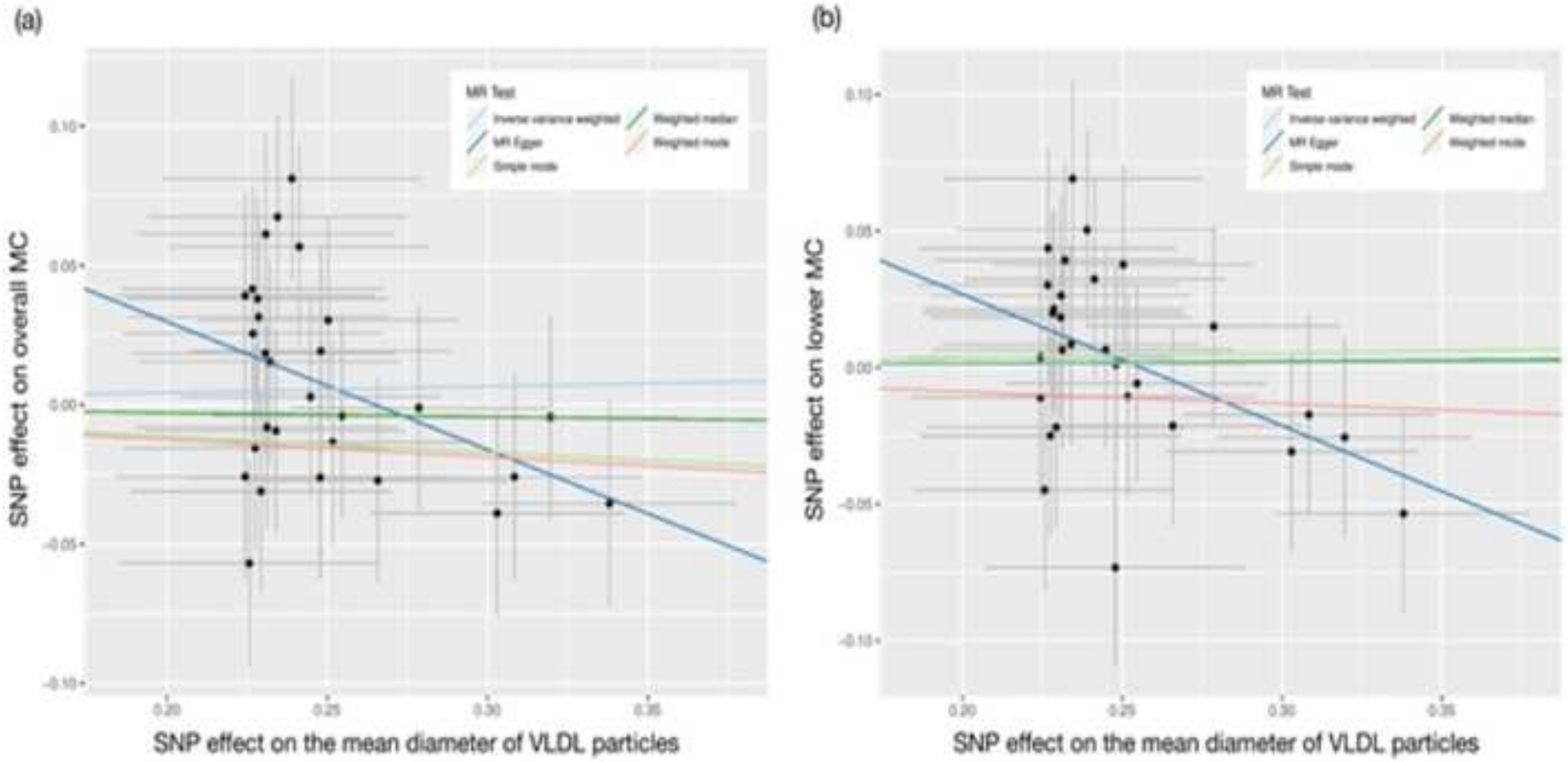


Figure 5

Click here to access/download;Figure;Figure 5 - MR Scatterplots.jpg



European Spine Journal

Springer
the language of science

Conflict of Interest Disclosure Form

It is the policy of the *European Spine Journal* to ensure balance, independence, objectivity, and scientific rigor in the journal. All authors are expected to disclose to the readers any real or apparent conflict(s) of interest that may have a direct bearing on the subject matter of the article. This pertains to relationships with pharmaceutical companies, biomedical device manufacturers or other corporations whose products or services may be related to the subject matter of the article or who have sponsored the study.

The intent of the policy is not to prevent authors with a potential conflict of interest from publication. It is merely intended that any potential conflict should be identified openly so that the readers may form their own judgements about the article with the full disclosure of the facts. It is for the readers to determine whether the authors' outside interest may reflect a possible bias in either the exposition of the conclusions presented.

The corresponding author will complete and submit this form to the Editorial Office on behalf of all authors listed below.

Article Title Integrative Analysis of Metabolomic, Genomic, and Imaging-based Phenotypes identify Very-Low-Density ... etc.

Authors Yiming Li, JARO KARPPINEN, Kathryn Cheah, Pak Sham, Dino Samartzis

Please note that a conflict of interest statement is published with each paper.

I certify that there is no actual or potential conflict of interest in relation to this article. If any conflict exists, please define hereafter:

Conflict (if none, "None" or describe financial interest/arrangement with one or more organizations that could be perceived as a real or apparent conflict of interest in the context of the subject of this article):

.....

.....

.....

Name DINO SAMARTZIS

Signature ..... Date October 19, 2020

Please upload the completed and signed form together with your submission via Editorial Manager.

Editorial Office:
Mrs. Irene Zintel
Pfrundstrasse 36
3176 Neuenegg, Switzerland