Name: Chinmay Raut **Dissertation Advisor:** Elizabeth Speliotes
Relevant Faculty: Mike Boehnke, Ryan Mills, Stephen CJ Parker, Maureen Sartor

Preliminary Exam (Specific Aims)

Metabolic syndrome (MetS) is a combination of several health risk factors, including abdominal obesity, insulin resistance, hypertension, and dyslipidemia, that together increase the risk for cardiovascular disease, type 2 diabetes, stroke, and other chronic conditions [1]. Up to 1 in 3 Americans currently satisfy the MetS conditions making it a significant public health challenge [1]. The underlying mechanisms of metabolic syndrome are complex and not fully understood, but are thought to involve a genetic component with MetS having an estimated heritability of about 20-30% [2]. There is a need for further research to determine the underlying causes of MetS and to develop effective strategies for prevention and treatment.

Previous studies have identified hundreds of associations between common genetic variants and MetS using Genome-Wide Association Studies (GWAS) on databases of up to 300,000 individuals [3]. However, little is known about the effects of rare variants on MetS, and no large-scale (>10,000 individuals) rare-variant associations have been conducted on MetS [4]. Rare genetic variants (minor allele frequency < 0.05) are worth studying for three reasons: they make up most of the sites of variation across the genome [5], they are predicted to result in significant phenotypic changes if they reside in protein-coding regions [5], and studying effects in protein-coding regions also provides a direct route for future functional studies and therapeutic targets [6]. Recently, with the integration of Next-Generation Sequencing and longer read technology, we have been able to perform Whole Genome (WGS) and Whole Exome Sequencing (WES) on Biobanks of individuals resulting in high-quality data on low-frequency genetic variants [5]. Currently, studies on rare variants in GWAS are limited by power due to low sample counts of individuals within a selected rare variant. A series of approaches have been developed to combat these limitations including the construction of the optimal Sequence Kernel Association Test (SKAT-O) [7] which combines a set of rare variants together to increase the power of the association test. When the SKAT-O test is paired with a variant annotator like the WGS Annotator (WGSA) [8] this results in the direct implication of genes to an outcome.

To understand the role genetic variants have on MetS, we want to identify and study its associations with rare, protein-coding variations using data from Biobanks. Exonic variants have the potential to alter proteins which then directly affect MetS phenotypes. Therefore, studying protein-altering variants offers direct insight into the biological mechanisms of MetS. We hypothesize that using WGS and WES data from large Biobanks such as the UK Biobank (UKBB) and the Michigan Genomics Initiative (MGI), we can detect exonic variants resulting in protein-coding changes which affect MetS. This will establish a series of protein-altering variants that provide insight into factors of MetS either for improvements in genetic testing or follow-ups for functional studies to investigate the biological mechanisms underlying the associations.

Aim 1: Determine Novel Gene-Based Associations with MetS.

Aim 1a: Identify genes associated with components of disease. MetS diagnosis is based on thresholds for 5 different quantitative traits, and individuals must surpass the thresholds on 3/5 of these traits for diagnosis resulting in 10 (5 choose 3) distinct trait combinations for disease qualification. Using support vectors we will quantify an individual's disease severity score against each of these 10 combinations. We then run gene-based analysis techniques to identify associations between genes and each of these MetS components.

Aim 1b: Integrate discovered signals across components to study the holistic effect on disease. After identifying significant genes across the 10 trait combinations, we will conduct enrichment analysis on the gene-based results from Aim 1a genes associated with multiple MetS profiles. We will also perform gene-based association tests against a joint severity score across the different trait combinations.

Based on the rare exonic variants we expect to identify and implicate specific genes that associate with components of MetS. We expect to construct a network of how these genes affect MetS as a whole, and if certain genes share similarities in effect on MetS providing further insights into its genetic basis.

Aim 2: Determine sequence-specific mechanisms of disease and progression.

Aim 2a: Within a gene, identify loci of individual variants affecting MetS. We will perform Leave One Variant Out (LOVO) studies on the effects of variants within an implicated gene to understand if the effects of the protein-altering variants are restricted to a specific region of the gene or not.

Aim 2b: Design CRISPR CAS9 KO guides based on exonic variants. Based on the specific type of rare variants identified (frameshift, splicing, missense, etc...) we will design a CAS9 guide library surrounding those regions to study the functional effects of the identified variants in vitro.

We expect to produce a series of specific genomic loci potentially responsible for changes in MetS along with a CAS9 library for future functional experiments. This provides a direct target for future studies regarding the effects of genetic modification on MetS.

By studying new next-generation sequencing data to understand the role of rare genetic variants on MetS, we expect to identify genes that affect MetS and also provide biological insight into how they relate to the components of the condition. This study would lay the groundwork for further functional tests to investigate the biology behind the genetics of MetS and provide a genetics-based informatory tool to understand how one's genome contributes to risks in MetS.

References

[1] M. G. Saklayen, "The Global Epidemic of the Metabolic Syndrome," *Curr Hypertens Rep*, vol. 20, p. 12, Feb 2018.

[2] F. Graziano, G. Biino, M. T. Bonati, B. M. Neale, R. Do, M. P. Concas, S. Vaccargiu, M. Pirastu, O. Terradura-Vagnarelli, M. Cirillo, M. Laurenzi, M. Mancini, A. Zanchetti, and M. Grassi, "Estimation of metabolic syndrome heritability in three large populations including full pedigree and genomic information," *Human Genetics*, vol. 138, pp. 739–748, Jul 2019.

[3] L. Lind, "Genome-Wide Association Study of the Metabolic Syndrome in UK Biobank," *Metab Syndr Relat Disord*, vol. 17, pp. 505–511, Dec 2019.

[4] H.-S. Lee, Y. Kim, and T. Park, "New common and rare variants influencing metabolic syndrome and its individual components in a korean population," *Scientific Reports*, vol. 8, p. 5701, Apr 2018.

[5] J. D. Backman, A. H. Li, A. Marcketta, D. Sun, J. Mbatchou, M. D. Kessler, C. Benner, D. Liu, A. E. Locke, S. Balasubramanian, A. Yadav, N. Banerjee, C. E. Gillies, A. Damask, S. Liu, X. Bai, A. Hawes, E. Maxwell, L. Gurski, K. Watanabe, J. A. Kosmicki, V. Rajagopal, J. Mighty, M. Jones, L. Mitnaul, E. Stahl, G. Coppola, E. Jorgenson, L. Habegger, W. J. Salerno, A. R. Shuldiner, L. A. Lotta, J. D. Overton, M. N. Cantor, J. G. Reid, G. Yancopoulos, H. M. Kang, J. Marchini, A. Baras, G. R. Abecasis, and M. A. R. Ferreira, "Exome sequencing and analysis of 454,787 UK Biobank participants," *Nature*, vol. 599, pp. 628–634, Nov 2021.

[6] N. Verweij, M. E. Haas, J. B. Nielsen, O. A. Sosina, M. Kim, P. Akbari, T. De, G. Hindy, J. Bovijn, T. Persaud, L. Miloscio, M. Germino, L. Panagis, K. Watanabe, J. Mbatchou, M. Jones, M. LeBlanc, S. Balasubramanian, C. Lammert, S. Enhörning, O. Melander, D. J. Carey, C. D. Still, T. Mirshahi, D. J. Rader, P. Parasoglou, J. R. Walls, J. D. Overton, J. G. Reid, A. Economides, M. N. Cantor, B. Zambrowicz, A. J. Murphy, G. R. Abecasis, M. A. Ferreira, E. Smagris, V. Gusarova, M. Sleeman, G. D. Yancopoulos, J. Marchini, H. M. Kang, K. Karalis, A. R. Shuldiner, G. Della Gatta, A. E. Locke, A. Baras, and L. A. Lotta, "Germline mutations in cideb and protection against liver disease," *New England Journal of Medicine*, vol. 387, no. 4, pp. 332–344, 2022. PMID: 35939579.

[7] S. Lee, M. J. Emond, M. J. Bamshad, K. C. Barnes, M. J. Rieder, D. A. Nickerson, D. C. Christiani, M. M. Wurfel, and X. Lin, "Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies," *Am J Hum Genet*, vol. 91, pp. 224–237, Aug 2012.

[8] X. Liu, S. White, B. Peng, A. D. Johnson, J. A. Brody, A. H. Li, Z. Huang, A. Carroll, P. Wei, R. Gibbs, R. J. Klein, and E. Boerwinkle, "WGSA: an annotation pipeline for human genome sequencing studies," *J Med Genet*, vol. 53, pp. 111–112, Feb 2016.