Interaction Effects between Established Non-genetic Risk Factors and Genetically Regulated Protein Levels Influencing the Risk of Coronary Artery Disease: A Large-Scale Analysis of the All of Us Cohort

Nubaira Rizvi, M.S. (LSU-Health, SPH); Shuai Liu, PhD (LSU-LCMC Health Cancer Center); Lang Wu, PhD (LSU-Health, SPH; LSU-LCMC Health Cancer Center)*; Jingjing Zhu, PhD (LSU-LCMC Health Cancer Center, Bioinformatics and Genomics Core)* (*joint corresponding authors)

Background

- Recent proteogenomic studies have identified specific circulating proteins as causal risk factors for **Coronary Artery Disease (CAD)**.
- Traditional models assume these protein-disease associations are static across all individuals, ignoring the complexity of human biology.
- It remains unknown if **non-genetic risk factors** (e.g., BMI, smoking, diabetes etc.) modify the impact of these proteins on CAD risk.
- Identifying **potential interactions** is critical to facilitate the development of improved risk assessment strategies.

Objective

This study aimed to comprehensively characterize such potential interaction effects between genetically regulated protein levels and established risk factors including age, sex, and metabolic state on CAD risk in European populations within the large-scale All of Us cohort.

Methodology

Study Population: **All of Us Research Program** whole-genome sequencing (WGS) data, focusing on a cohort of **162,509 participants** of European ancestry (25,511 CAD cases; 136,998 controls).

Genetic Prediction Models:

- **Weights Source:** Proteomic prediction weights derived from the **Multi-Ethnic Study of Atherosclerosis (MESA)**, a validated longitudinal cohort.
- **Protein Selection:** Focused on **355 proteins** previously identified as causal or highly associated with Coronary Artery Disease (CAD) risk in European populations.

Statistical Framework:

Interaction Testing: Performed multivariable logistic regression testing interactions between protein levels and **7 established risk factors**:

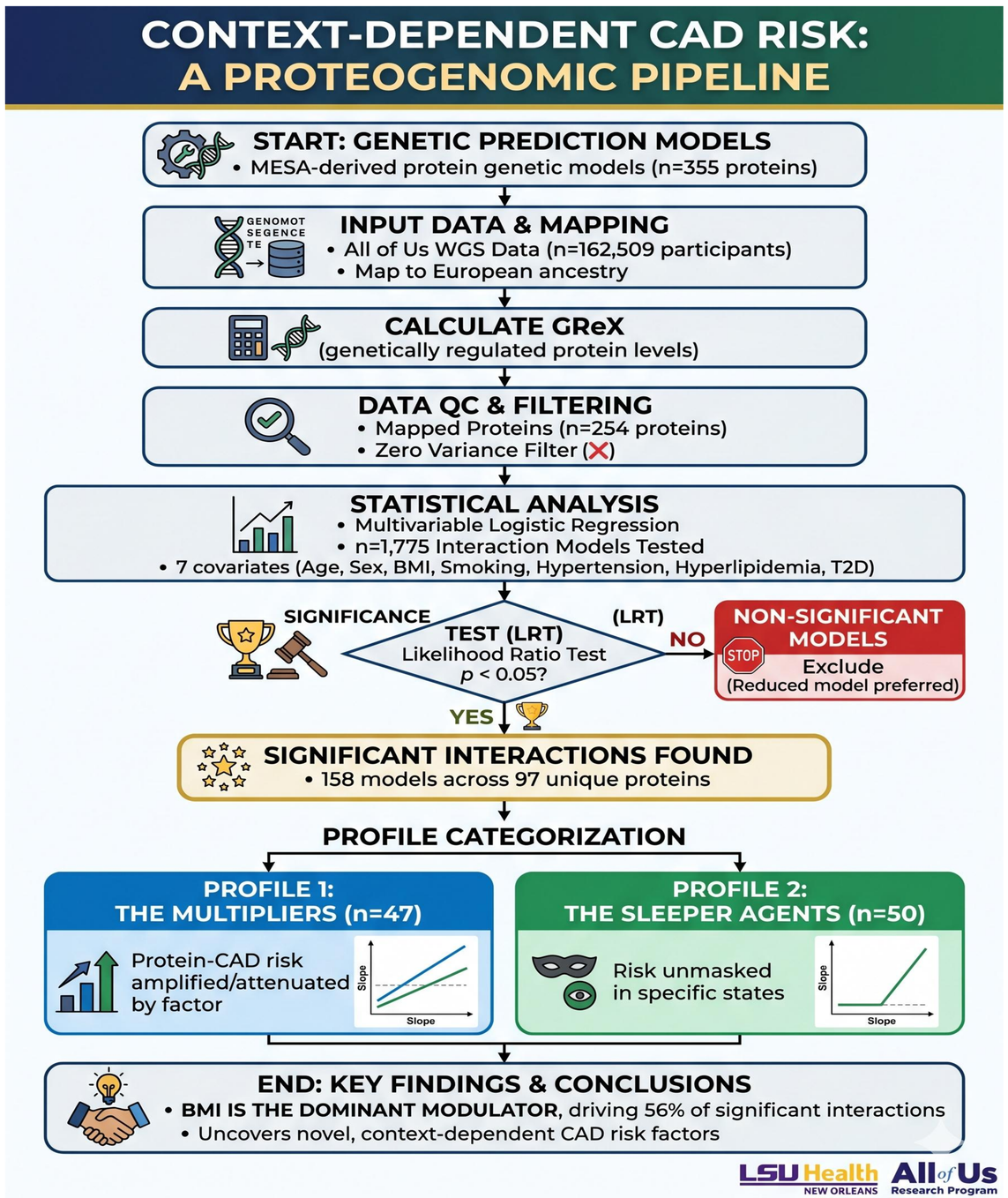
Demographic: Age, Sex.

Metabolic/Lifestyle: BMI, Smoking, Hypertension, Hyperlipidemia, and Type 2 Diabetes (T2D).

Primary Outcome: The presence or absence of **Coronary Artery Disease (CAD)**.

Definition: CAD cases were identified within the *All of Us* Electronic Health Records (EHR) using a combination of:

ICD-9/ICD-10 Codes: Specifically focusing on myocardial infarction, stable/unstable angina, and coronary atherosclerosis.



Results

Figure 1: Distribution of 158 Protein-CAD Interactions across 7 Covariates

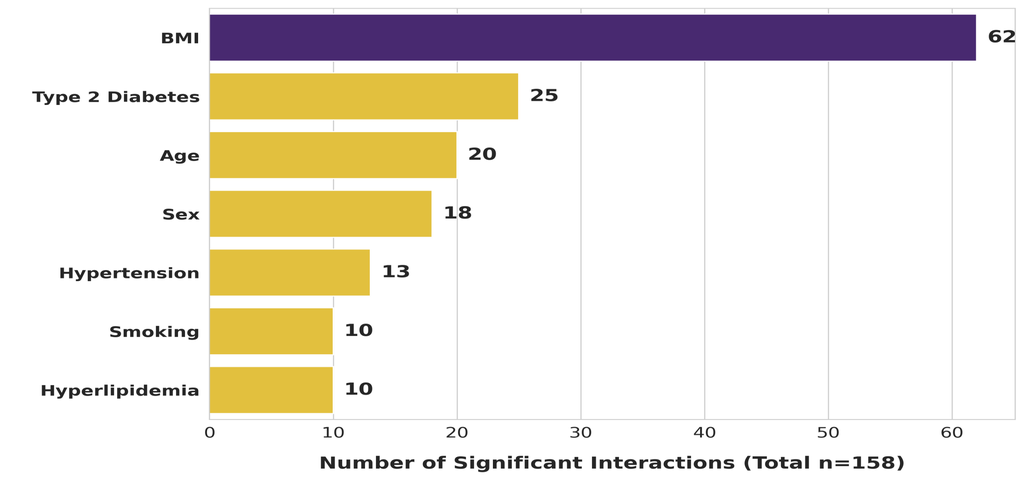


Table 1: Top Significant Protein-Covariate Interactions for CAD Risk

Protein ID	Modulating Factor	Profile Type	Interaction β	LRT P-value
seq.3054.3	Age	Multiplier	-0.25	0.0011
seq.3290.50	BMI	Sleeper Agent	+0.29	0.0026
seq.4154.57	Age	Multiplier	+0.19	0.0021
seq.9253.52	BMI	Multiplier	+0.38	0.00014
seq.9314.9	Smoking	Sleeper Agent	+0.51	0.0032
seq.10613.33	BMI	Multiplier	+0.42	0.000028
seq.15606.19	T2D	Sleeper Agent	+0.31	0.0048
seq.22796.17	T2D	Sleeper Agent	+0.33	0.0045

Key Findings

- Out of **1,775** interaction models tested, **158** demonstrated significant interaction effects (**LRT p < 0.05**).
- These interactions involve **97 unique proteins**, suggesting that a substantial portion of the CAD-related proteome is modulated by non-genetic factors.
- **BMI** emerged as the primary driver of interaction, accounting for **56%** of the most significant context-dependent findings.

Profile 1: The Multipliers (n = 47 Proteins)

- This profile identifies proteins where an established CAD association is significantly amplified or attenuated by external risk factors.
- The association between **seq.10613.33** and CAD risk was most strongly modified by **BMI** ($p = 2.8*10^{-5}$), showing an increase in risk.
- Proteins such as **seq.3054.3** and **seq.4154.57** showed **age-dependent** risk shifts ($p < 0.005$), indicating that genetic protein risk varies across the lifespan.

Profile 2: The Sleeper Agents (n = 50 Proteins)

- These proteins demonstrated **no significant main effect** on CAD but emerged as potent risk factors within specific strata.
- **seq.9314.9** was identified as a risk factor specifically in **smokers** ($p = 0.0032$), showing no effect in non-smokers.
- A distinct cluster of proteins (e.g., **seq.22796.17, seq.15606.19**) only demonstrated significant associations with CAD in patients with **Type 2 Diabetes**, highlighting pathways that require a specific metabolic "state" to become pathological.

Discussion

In the current work, associations of many CAD related proteins tend to show differences according to specific strata of key metabolic and demographic related risk factors. BMI acts as a primary modulator, unmasking potent risk factors within specific states like Type 2 Diabetes.

If validated by future independent studies, such findings could improve the etiologic understanding of this common disease and facilitate its risk assessment by integrating these context-dependent interactions.

Acknowledgements

We express our deepest gratitude to the participants of the All of Us Research Program.

This work was conducted under the guidance of Dr. Lang Wu (LSU-LCMC Health Cancer Center) and Dr. Jingjing Zhu (LSU-LCMC Health Cancer Center, Bioinformatics and Genomics core).

This project is funded by the following grants U54HG013243, U01CA293883 and R01CA263494.