

Contents

Glossary.....	39
Introduction	41
1. Type 2 diabetes mellitus	43
1.1. Pathophysiology of type 2 diabetes	44
1.1.1. Oxidative stress and insulin resistance	46
1.2. Diabetes definition and pre-diabetes markers	47
1.3. Diabetes complications and related outcomes	48
1.3.1. Kidney diseases	49
1.3.2. Cardiovascular diseases	50
2. Selenium, a key element for antioxidant defence	51
2.1. Sources of selenium, intake recommendations and global variation	51
2.2. Selenium metabolism	52
2.3. Selenium species	53
2.4. Selenium biomarkers	56
2.5. Selenium in human disease: type 2 diabetes	57
3. Genomics and Genome-wide association studies	58
3.1. Pre-GWAS steps.....	61
3.2. Genetic association test for GWAS	63
3.3. Meta-analysis and genetic consortia	64
3.4. Exploiting genomic variation to assess cause-effect associations in post-GWAS analyses	65
3.4.1. Polygenic score.....	66
3.4.2. Mendelian randomization.....	67
3.4.3. Genetic colocalization.....	74
Hypotheses and objectives	84
1. Hypotheses	86

2. Objectives.....	87
Differential association of selenium exposure with insulin resistance and β-cell function in middle age and older adults	90
Abstract	93
1. Introduction	95
2. Material and methods.....	96
2.1. Study populations.....	96
2.2. Blood selenium levels	98
2.3. HOMA-IR and HOMA- β outcomes	98
2.4. Other variables	99
2.5. External Study population (reference) for calibration of glucose and insulin and replicate cross-sectional results	100
2.6. Statistical analyses	102
2.7. Sensitivity analyses.....	106
3. Results.....	107
3.1. Participant characteristics.....	107
3.2. Association of selenium with HOMAs endpoints.....	111
3.3. Sensitivity analysis	115
4. Discussion	119
metaGWASmanager: a toolbox for an automated workflow from phenotypes to meta-analysis in GWAS consortia.....	126
Abstract	129
1. Introduction	130
2. Materials, methods and workflow	131
2.1. metaGWASmanager strategy and stages	131
2.2. Software availability.....	136
2.3. MetaGWASmanager application to selenium data.....	137
2.3.1. Study populations	137
2.3.2. Serum and plasma selenium determination.....	138
2.3.3. Genotyping and imputation.....	138

3. Results from metaGWASmanager application	139
3.1. Customize metaGWASmanager for selenium data	140
3.2. Phenotypes examination	140
3.3. Association analyses.....	144
3.4. Cross-study validation and meta-analysis	149
4. Discussion	154
Non-genetic and genetic determinants of selenium biomarkers in the Aragon Workers Health Study.....	158
Abstract	161
1. Introduction	163
2. Material and methods.....	165
2.1. Study population	165
2.2. Selenium species and biomarkers assessment	165
2.3. Genotyping and imputation	167
2.4. Other variables	167
2.5. Statistical analyses.....	168
3. Results.....	174
3.1. Descriptive analysis.....	174
3.2. Non-genetic determinants of selenium markers.....	180
3.3. Genetic determinants of selenium markers.....	185
3.4. Enrichment analysis	197
3.5. Relative contribution of genetic and lifestyle factors to selenium markers	205
4. Discussion	207
Exploitation of genomics to uncover underlying causal associations of selenium with diabetes-related endpoints: proof of concept analyses..	218
1. Introduction	221
2. Materials and methods.....	223
2.1. Study design and participants	223
2.2. Selenium species and biomarkers determinations.....	224

2.3.	Genetic data	224
2.4.	Outcome data sources	224
2.4.1.	Data sources for individual level observational and genetic doses-responses	224
2.4.2.	Data sources for colocalization analyses.....	226
2.4.3.	Data sources for MR analyses	226
2.5.	Definition of additional variables.....	227
2.6.	Statistical analyses.....	228
2.6.1.	GWAS of selenium biomarkers levels and polygenic scores.....	228
2.6.2.	Observational and genetic dose-response	228
2.6.3.	Bayesian colocalization analyses.....	229
2.6.4.	Mendelian randomization analyses.....	230
2.6.5.	Pathway enrichment analysis	233
3.	Results.....	233
3.1.	Observational and genetic dose-responses of selenium biomarkers and diabetes-related outcomes	233
3.2.	Diabetes-wide colocalization analyses of selenium biomarkers	241
3.3.	Enrichment analyses of genes associated with selenium biomarkers in colocalization analyses	248
3.4.	Mendelian randomization associations.....	253
3.5.	Causal effect of selenium biomarkers on diabetes-related outcomes.....	270
4.	Discussion	274
	General summary of findings.....	283
	Conclusions	291
	Future perspectives	297
	Scientific production and research dissemination activities during the PhD program	303
1.	Publications	305
2.	Conferences.....	307

2.1. Oral presentation.....	307
2.2. Poster presentation	310
3. Software development	311
4. Grants.....	311
5. Projects participation	312
6. Undergraduate thesis supervision	312
7. Teaching experience	313
8. Other activities	313
References	316
Appendix A.....	372
Appendix B.....	442
Appendix C	462

List of tables

Table 1.1. Overview of differences between Mendelian randomization and colocalization methods.....	81
Table 3.1. β coefficients (standard errors) of glucose and insulin levels in the NHANES 2011-2018 external (reference) cohort for predictors selected for the recalibration approach.....	106
Table 3.2. Geometric means (geometric standard deviations) of blood selenium levels and recalibrated HOMA endpoints at baseline by participant characteristics in AWHs and SEN-2 studies.	107
Table 3.3. Geometric mean (geometric standard deviation) of originally measured (i.e., non-recalibrated) HOMA endpoints at baseline by participant characteristics in AWHs and SEN-2 studies.	109
Table 3.4. Geometric mean (geometric standard deviation) of blood selenium levels and HOMA endpoints by participant characteristics in NHANES 2011-2018 (participants aged <60 or $\geq$ 60 years).	110
Table 3.5. Cross-sectional geometric mean ratios (95% CIs) of baseline HOMA-IR and HOMA- β by blood selenium levels in AWHs and SEN-2 studies.	111
Table 3.6. Mean (standard deviation) of fasting serum glucose (mg/dL) and geometric mean (geometric standard deviation) of fasting serum insulin (μ U/mL) levels at baseline before and after recalibration in AWHs and in NHANES 2011-2018 non-Hispanic white participants aged 40-59 years (reference population).	116
Table 3.7. Mean (standard deviation) of fasting serum glucose (mg/dL) and geometric mean (geometric standard deviation) of fasting serum insulin (μ U/mL) levels at baseline before and after recalibration in SEN-2 and in	

NHANES 2011-2018 non-Hispanic white participants aged 60-80 years (reference population).....	117
Table 3.8. Cross-sectional geometric mean ratios (95% confidence intervals) of baseline HOMA-IR and HOMA- β calculated based on originally measured (i.e., non-recalibrated) fasting serum glucose and insulin by blood selenium levels in AWHs and SEN-2 studies.....	118
Table 3.9. Sensitivity cross-sectional analysis after additionally adjusting models by waist circumference. Geometric mean ratios (95% confidence intervals) of baseline HOMA-IR and HOMA- β calculated based on recalibrated measured fasting serum glucose and insulin by blood selenium levels in AWHs and SEN-2 studies.....	118
Table 4.1. Dummy-input file example for the metaGWASmanager application for selenium data.	141
Table 4.2. Summary-table of plasma selenium-related variables used in GWAS in the AWHs and The Hortegea study.....	144
Table 5.1. Limit of detection (LOD) and coefficient of variation (CV, %) of selenium biomarkers.	167
Table 5.2. NCBI-Gene search strategy for selenium-related genes conducted on April 29th 2023.....	171
Table 5.3. Median (interquartile range) of serum selenium and selenium species concentrations by participants characteristics.	175
Table 5.4. Median (interquartile range) by participants characteristics of serum, blood and urine selenium levels in the subsample of participants with available blood selenium (N=1,168); and urine selenium in the sample with available urine data (N=1,875).....	179

Table 5.5. Geometric mean ratio (GMR) and 95% confidence intervals (CI) of serum selenium and selenium species concentrations by sociodemographic and lifestyle factors.	182
Table 5.6. Progressively adjusted multiple-determinants models for serum selenium.	183
Table 5.7. Progressively adjusted multiple-determinants models for selenoalbumin.....	183
Table 5.8. Geometric mean ratios (GMR) and 95% confidence intervals (CI) of serum, blood and urine selenium levels by sociodemographic and lifestyle determinants in the subset of participants with available blood selenium ^a (N=1,168); and, urine selenium in the sample with available urine data ^b (N=1,875).	184
Table 5.9. SNPs associated with the concentration of selenium biomarkers in the candidate gene approach.....	186
Table 5.10. Geometric mean ratios and 95% confidence interval for quintiles categories of the polygenic score for selenium biomarkers concentrations.	206
Table 5.11. Percentage of variance explained by non-genetic and genetic factors for each selenium biomarker concentration.	207
 Table 6.1. Genetic instrumental variables selection based on the number of SNPs at baseline along the different filters for each specific outcomes and selenium biomarkers.	231
Table 6.2. Evidence for a shared causal SNP or two distinct causal SNPs between selenium biomarkers and diabetes-related outcomes in FinnGen database.	242
Table 6.3. Evidence for a shared causal SNP or two distinct causal SNPs between selenium biomarkers and diabetes-related outcomes in UKBB database.	244

Table 6.4. Evidence for a shared causal SNP or two distinct causal SNPs between selenium biomarkers and gene expression in the kidney-eQTL database.	245
Table 6.5. Biological enrichment analysis of GTEx- and kidney eQTL-genes associated with selenium biomarkers in colocalization analyses.	249
Table 6.6. Statistically significant causal associations between selenium biomarkers and diabetes-outcomes using mendelian randomization approach.	253
Table 6.7. Mendelian randomization sensitivity analyses: heterogeneity (Cochran's Q test).	261
Table 6.8. Mendelian randomization sensitivity analyses: horizontal pleiotropy in MR-Egger.	262
Table 6.9. Comparison of the results derived from three different methods to assess the causal effect of selenium biomarkers on diabetes-related outcomes.	271

List of figures

Figure 1.1. Prevalence of diabetes worldwide (20-79 years old).	43
Figure 1.2. Insulin resistance mechanism.	45
Figure 1.3. Selenium metabolism in humans: absorption, transformation, distribution, and excretion.	54
Figure 1.4. Outline of GWAS steps.	60
Figure 1.5. Analogy between Mendelian randomization and randomized controlled trials.	68
Figure 1.6. Overview of the instrumental variable assumptions.	69
Figure 1.7. Overview of the two-sample Mendelian randomization approach.	72
Figure 1.8. Configuration under colocalization hypothesis for trait A and B.	76
Figure 1.9. Example of the impact of prior probabilities on colocalization results.	79
 Figure 3.1. Cross-sectional dose-response trends of blood selenium levels with baseline HOMAs-endpoints in AWHs and NHANES participants aged < 60 years.	112
Figure 3.2. Cross-sectional dose-responses trends of blood selenium levels with baseline HOMAs-endpoints in SEN-2 and NHANES participants aged ≥ 60 years.	113
Figure 3.3. Longitudinal dose-response trends of blood selenium levels with annual relative changes in HOMA-IR and HOMA-β in AWHs and SEN-2 studies.	114
 Figure 4.1. metaGWASmanager workflow.	131

Figure 4.2. Summary of diagnostic plots generated during metaGWASmanager-Phase 2 for AWHs.	142
Figure 4.3. Summary of diagnostic plots generated during metaGWASmanager-Phase 2 for the Hortega Study.....	143
Figure 4.4. Allele frequency comparison between Spanish studies and 1000 Genome reference dataset for GWAS of plasma selenium.	145
Figure 4.5. Allele frequency comparison between Spanish studies and TOPMed dataset for GWAS of plasma selenium.	146
Figure 4.6. P-value correlation plot for GWAS of plasma selenium in Spanish studies.	147
Figure 4.7. Distribution of the plasma selenium GWAS-parameters.	148
Figure 4.8. QQ plots for GWAS of plasma selenium in Spanish studies. ...	150
Figure 4.9. Manhattan plots for plasma selenium.	151
Figure 4.10. Across-study boxplots of genomic parameters for GWAS of plasma selenium	152
Figure 4.11. Meta-analysis results of plasma selenium.	153
Figure 5.1. Flux diagram of selected SNPs for genetic association studies of selenium biomarkers.	172
Figure 5.2. Distribution of serum selenium species percent in the AWHs.	177
Figure 5.3. Spearman correlation coefficients between pairs of selenium biomarkers.	178
Figure 5. 4. Genome-wide association $-\log_{10} P$ -values for adjusted GMR of serum selenium.	188
Figure 5.5. Genome-wide association $-\log_{10} P$ -values for adjusted GMR of glutathione peroxidase.	189
Figure 5.6. Genome-wide association $-\log_{10} P$ -values for adjusted GMR of selenoprotein P.	190

Figure 5.7. Genome-wide association $-\log_{10} P$ -values for adjusted GMR of selenoalbumin.....	191
Figure 5.8. Genome-wide association $-\log_{10} P$ -values for adjusted GMR of selenometabolites.....	192
Figure 5.9. Summary of results from genetic association.	193
Figure 5.10. Genome-wide association $-\log_{10} P$ -values for adjusted GMR of blood selenium.....	195
Figure 5.11. Genome-wide association $-\log_{10} P$ -values for adjusted GMR of urine selenium.	196
Figure 5.12. Network of biological pathways for genes annotated to the most relevant SNPs in genetic association analysis of serum selenium.	197
Figure 5.13. Network of biological pathways for genes annotated to the most relevant SNPs in genetic association analysis of glutathione peroxidase. ..	198
Figure 5.14. Network of biological pathways for genes annotated to the most relevant SNPs in genetic association analysis of selenoprotein P.....	199
Figure 5.15. Network of biological pathways for genes annotated to the most relevant SNPs in genetic association analysis of selenoalbumin.....	200
Figure 5.16. Network of biological pathways for genes annotated to the most relevant SNPs in genetic association analysis of selenometabolites.	201
Figure 5.17. Upset plot of enriched biological pathways from GDA for selenium biomarkers-related genes.....	202
Figure 5.18. Network of biological pathways for genes annotated to the most relevant SNPs in genetic association analysis of blood selenium.	203
Figure 5.19. Network of biological pathways for genes annotated to the most relevant SNPs in genetic association analysis of urine selenium.	204
Figure 5. 20. Histograms of polygenic score for selenium biomarkers.	205

Figure 6.1. Cross-sectional dose-response of observational (left) and genetic predicted (right) selenium biomarkers levels with HOMA-IR and HOMA- β	235
Figure 6.2. Cross-sectional dose-response of observational (left) and genetic predicted (right) selenium biomarkers levels with general plaque and Agatston scores.	237
Figure 6.3. Cross-sectional dose-response of observational (left) and genetic predicted (right) selenium biomarkers levels with eGFR and ACR.	239
Figure 6.4. Summary of colocalization results between selenium biomarkers and PheWAS public databases (UKBB and FinnGen).	247
Figure 6.5. Upset plot of the enriched biological pathways from genes associated with selenium biomarkers in colocalization analyses.	251
Figure 6.6. Pathway enrichment network of GTEx- and kidney eQTL-genes associated with selenium biomarkers in colocalization analyses.	252
Figure 6.7. Scatter plot inferring the causal effect of selenoprotein P on rapid eGFRcrea decline (CKDi25).	254
Figure 6.8. Scatter plot inferring the causal effect of serum selenium on rapid eGFRcrea decline (Rapid3).	255
Figure 6.9. Scatter plot inferring the causal effect of urine selenium on rapid eGFRcrea decline (CKDi25).	256
Figure 6.10. Scatter plot inferring the causal effect of serum selenium on cardiovascular diseases.	257
Figure 6.11. Scatter plot inferring the causal effect of serum selenium on major coronary heart disease event.	258
Figure 6.12. Scatter plot inferring the causal effect of GPx on ischemic heart disease.	259
Figure 6.13. Heatmap summary of mendelian randomization results between diabetes-related endpoints and selenium biomarkers.	260

Figure 6.14. Forest plots as sensitivity analysis to evaluate the causal association of serum selenium on rapid eGFR _{crea} decline (Rapid3).	263
Figure 6.15. Forest plots as sensitivity analysis to evaluate the causal association of selenoprotein P on rapid eGFR _{crea} decline (CKDi25).	264
Figure 6.16. Forest plots as sensitivity analysis to evaluate the causal association of urine selenium on rapid eGFR _{crea} decline (CKDi25).	265
Figure 6.17. Forest plots as sensitivity analysis to evaluate the causal association of serum selenium on cardiovascular diseases.	266
Figure 6.18. Forest plots as sensitivity analysis to evaluate the causal association of serum selenium on major coronary heart disease event.	267
Figure 6.19. Forest plots as sensitivity analysis to evaluate the causal association of GPx on ischemic heart disease.	268
Figure 6.20. Mendelian randomization sensitivity funnel plots for individual causal effect estimates.	269