
Genome-wide Association Study of RSPO3 Protein Expression: A Survey of Sex-Specific Genetic Variations

www.surveyx.cn

Abstract

This survey paper presents a comprehensive examination of genome-wide association studies (GWAS) focusing on RSPO3 protein expression, emphasizing sex-specific genetic variations. RSPO3, a key modulator in the Wnt signaling pathway, plays critical roles in developmental processes, cellular homeostasis, and cancer biology. By integrating advanced methodologies such as empirical Bayes regression models and machine learning techniques, this study enhances the detection of genetic associations, particularly those contributing to sex-specific differences in RSPO3 expression. The survey highlights the dual roles of RSPO3 in bone homeostasis and tumorigenesis, suggesting its potential as a therapeutic target. Methodological advancements, including the use of multi-tissue expression quantitative trait loci (eQTL) data, enable a nuanced understanding of the genetic architecture influencing RSPO3. Despite these advancements, challenges remain in addressing population stratification and computational intensity. Future research should focus on refining analytical methods, exploring gene-environment interactions, and expanding study populations to improve the generalizability of findings. This survey underscores the significance of RSPO3 in genetic research, offering valuable insights into the biological mechanisms underlying complex traits and diseases.

1 Introduction

1.1 Overview of Genome-Wide Association Studies (GWAS)

Genome-wide association studies (GWAS) are fundamental in genetic research, elucidating the genetic underpinnings of complex traits and diseases through the identification of single nucleotide polymorphisms (SNPs) [1, 2]. By linking genetic variants to phenotypic traits, GWAS enhance our comprehension of biological processes and contribute significantly to gene discovery and clinical implications for complex diseases [3].

Methodological advancements have addressed challenges such as noise in genetically inferred relationship matrices, particularly among distant relatives [4]. Understanding population structure's impact on causal variant estimation is also crucial, as it can confound GWAS results [5]. The integration of empirical Bayes approaches has improved the identification of genetic variants influencing gene expression across tissues, as exemplified by the Genotype-Tissue Expression (GTEx) project.

GWAS are particularly valuable for studying complex traits shaped by nonlinear interactions among high-dimensional covariates, such as obesity [6]. As the field evolves, GWAS remain essential for deciphering the genetic architecture of complex traits, thereby deepening our understanding of the genetic and biological mechanisms that underlie human health and disease.

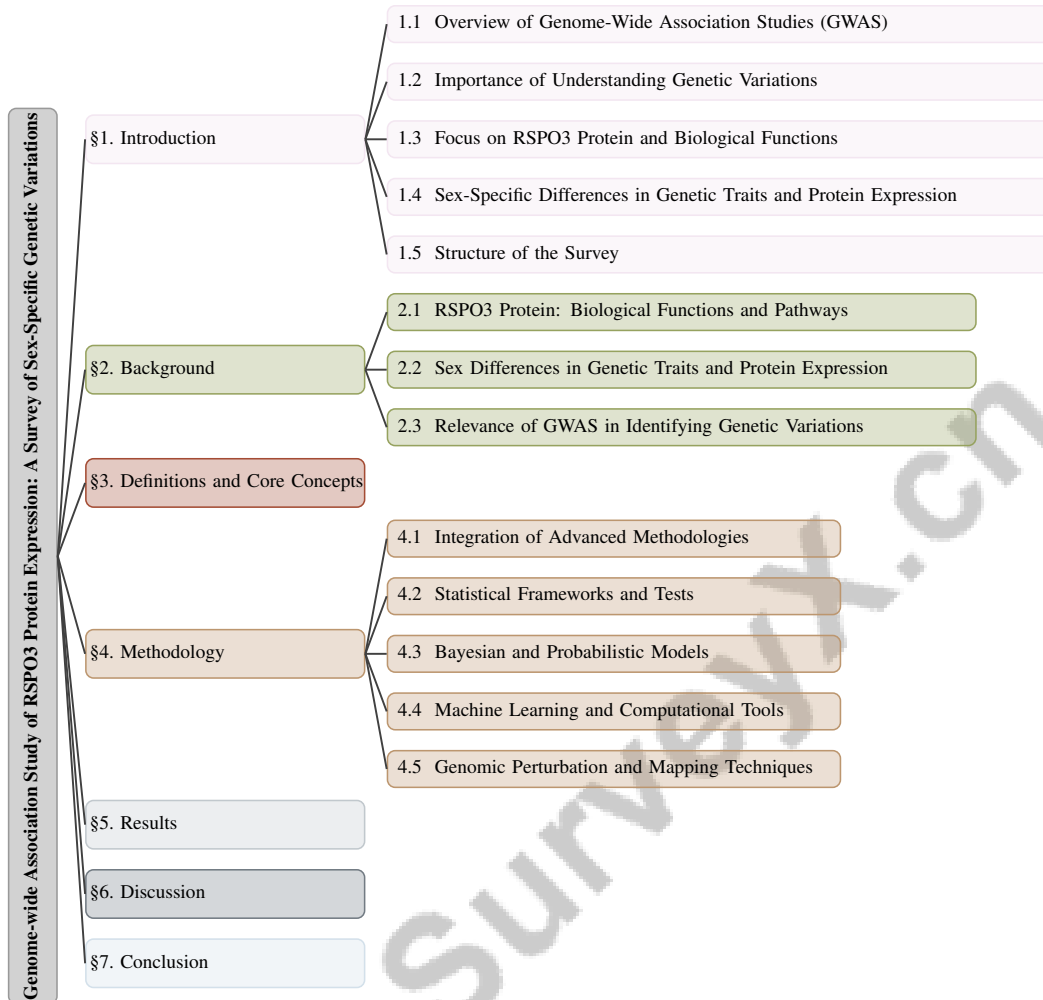


Figure 1: chapter structure

1.2 Importance of Understanding Genetic Variations

Understanding genetic variations is crucial for advancing biological and medical research, as it sheds light on the mechanisms driving complex traits and diseases. Genetic interactions, or epistasis, significantly contribute to phenotypic variation, necessitating thorough analyses to unravel these complexities [7]. Identifying SNPs associated with phenotypic differences is vital for elucidating genetic architecture [2], yet these variants often account for only a fraction of genetic variation, leading to the "missing heritability" problem [1]. This highlights the need to differentiate between deleterious and benign genetic variations, particularly in complex disease contexts [8].

Refining genetically inferred relationships among distantly related individuals is essential to mitigate inaccuracies stemming from noise in empirical genetic data [4]. Traditional expression quantitative trait loci (eQTL) analyses, which typically focus on single tissue data, may overlook insights from multi-tissue approaches [9]. The evolution of GWAS in identifying genetic variants linked to diseases and traits remains a central theme in genetic research [3].

In terms of data security, understanding genetic variations is vital for safeguarding genomic data privacy, notably through techniques like homomorphic encryption [10]. Discriminative patterns derived from genetic data provide essential insights for classification and subgroup discovery, underscoring the importance of accurately characterizing genetic variations in elucidating the genetic basis of complex traits and diseases [11].

1.3 Focus on RSPO3 Protein and Biological Functions

RSPO3 protein is crucial in various biological functions, significantly influencing developmental processes and cellular homeostasis. It modulates the Wnt signaling pathway, essential for regulating stem cell pluripotency, differentiation, and skeletal homeostasis [12]. Additionally, RSPO3 serves as a negative regulator of ERK/FGF signaling, impacting cell survival and development, which is pertinent in both physiological and pathological contexts, including tumorigenesis [13, 14].

Immunohistochemistry studies reveal RSPO3's localization in the hypothalamus, emphasizing its diverse roles in neurobiological processes [15]. Furthermore, RSPO3's involvement in gene expression regulation at the single-cell level indicates its integral role in mediating complex gene interactions that affect protein expression [16]. Despite its established significance across biological contexts, the role of wild-type RSPO3 in colorectal cancer growth remains unclear, necessitating further investigation to clarify its functions in oncogenesis [17].

RSPO3's multifaceted roles are crucial for advancing the understanding of genetic variations and their implications for sex-specific differences in protein expression. Consequently, RSPO3 is pivotal in GWAS aimed at uncovering genetic variants associated with complex traits and diseases, enhancing our grasp of genetic architecture and biological mechanisms that inform targeted therapies [18, 19].

1.4 Sex-Specific Differences in Genetic Traits and Protein Expression

Sex-specific differences in genetic traits and protein expression are essential for understanding the biological mechanisms that differentiate male and female phenotypes. These differences have significant implications for the diagnosis, treatment, and management of diseases. Sex-specific recombination rates and allele frequencies influence the dynamics of sexually antagonistic alleles, contributing to genetic diversity within populations [20]. Variations in recombination patterns, such as increased rates in females with age compared to males, underscore the complexity of genetic inheritance and its impact on disease susceptibility [21].

The influence of sex on genetic traits extends to various physiological and pathological contexts. For example, notable sex differences in genetic influences have been observed in endocrine functions, with females exhibiting higher heritability estimates for thyroid-stimulating hormone (TSH) levels than males [22]. This necessitates incorporating sex-specific genetic factors into endocrine research and clinical practice. In metabolic regulation, biological sex affects the prevalence and manifestation of metabolic disorders, such as type 2 diabetes [23].

Exploration of sex-specific differences is crucial for understanding disease mechanisms and outcomes, as seen in chronic obstructive pulmonary disease (COPD) and COVID-19 studies. Research indicates that genetic and environmental contributions to COPD differ by sex, influencing disease mechanisms and outcomes [24]. Similarly, sex differences in severity and mortality rates of COVID-19 highlight the impact of gender on clinical outcomes [25]. These findings emphasize the need to incorporate sex as a biological variable in medical research to bridge knowledge gaps and enhance personalized healthcare strategies.

In evolutionary biology, sex-specific differences in life history, physiology, and behavior across taxa are crucial for understanding the pace-of-life syndrome (POLS) hypothesis, suggesting that these differences are shaped by evolutionary pressures [26]. Improved inference in high-dimensional contexts further emphasizes the study of sex-specific genetic traits, enhancing the ability to discern complex genetic interactions and their phenotypic consequences [27]. The significance of studying interactions among items in discriminative patterns, which can vary across classes akin to sex-specific differences in genetic traits, is well-documented [11]. As the field advances, integrating diverse datasets and innovative methodologies will be vital for overcoming challenges such as population structure confounding and analytical limitations, thereby enhancing the understanding of sex-specific genetic and protein expression differences [28].

1.5 Structure of the Survey

This survey is systematically structured to provide a comprehensive examination of genome-wide association studies (GWAS) focusing on RSPO3 protein expression, particularly regarding sex-specific genetic variations. It begins with an **Introduction**, highlighting the significance of GWAS in genetic research and the critical role of RSPO3 in biological functions, alongside the importance of investi-

gating sex-specific differences in genetic traits and protein expression. Following the introduction, a detailed **Background** section explores the biological functions and pathways associated with RSPO3, delves into sex differences in genetic traits, and discusses GWAS's relevance in identifying genetic variations.

The survey then presents **Definitions and Core Concepts**, clarifying key terms and scientific methodologies pertinent to the study. The **Methodology** section elaborates on the approaches employed in conducting GWAS, including advanced methodologies, statistical frameworks, Bayesian models, machine learning tools, and genomic mapping techniques. The **Results** section highlights the findings related to genetic variations contributing to sex-specific differences in RSPO3 protein expression and their functional impacts.

The **Discussion** interprets the results within the context of existing literature, addressing implications for biological pathways, comparing methodologies, and discussing the integration of diverse datasets. It also considers methodological advances and limitations encountered. Finally, the **Conclusion** summarizes key findings and their significance, suggesting future research directions. This survey aims to enhance our understanding of sex-specific genetic variations in RSPO3 protein expression, which may significantly impact genetic research, particularly in GWAS analyzing complex diseases influenced by multiple genetic variants and their interactions with environmental factors [29, 18]. The following sections are organized as shown in Figure 1.

2 Background

2.1 RSPO3 Protein: Biological Functions and Pathways

RSPO3 protein is integral to numerous biological functions and pathways, significantly impacting physiological and pathological processes. As a key modulator of the Wnt signaling pathway, RSPO3 regulates stem cell pluripotency, differentiation, and skeletal homeostasis [12]. It negatively regulates osteogenic differentiation in human adipose-derived stem cells, influencing bone formation [13]. RSPO3 also inhibits Erk phosphorylation and stabilizes β -catenin, activating Wnt signaling, which affects cellular proliferation and differentiation [12].

In cancer biology, RSPO3, alongside RSPO2, is implicated in intestinal tumor initiation and development, indicating its role in oncogenic pathways [14]. Conversely, RSPO3 can inhibit colorectal cancer growth and enhance NK cell infiltration, suggesting a dual role as both an oncogene and tumor suppressor, contingent on the cellular context [17]. Its localization in hypothalamic nuclei underscores its role in neurobiological processes, supported by immunohistochemical studies [15]. This distribution suggests involvement in regulating neural functions and behaviors through interactions with other signaling pathways. Neural network analyses of protein expression have further elucidated RSPO3's complex regulatory networks [16].

RSPO3's multifaceted roles underscore its significance in GWAS, which seek to clarify genetic variations leading to sex-specific differences in RSPO3 expression. These studies employ advanced statistical methods to identify SNPs associated with these differences, enhancing our understanding of genetic factors influencing disease susceptibility and informing targeted therapeutic strategies [18, 1, 30, 19].

2.2 Sex Differences in Genetic Traits and Protein Expression

Sex differences in genetic traits and protein expression are crucial for understanding the biological mechanisms differentiating male and female phenotypes. These differences are vital for elucidating fundamental biological processes and their implications for disease diagnosis, treatment, and management. The dynamics of sexually antagonistic alleles, shaped by sex-specific recombination rates and allele frequencies, maintain genetic diversity and provide insights into sexual selection and evolutionary biology [20]. Maternal age affects recombination rates and crossover interference, influencing genetic inheritance [21].

Sex-specific differences manifest in various physiological and pathological contexts. For instance, significant sex differences in genetic influences on TSH and fT4 levels reveal that females exhibit higher heritability estimates for TSH levels, emphasizing the need to incorporate sex-specific genetic factors into endocrine research and clinical practice [22]. Additionally, sex differences play a crucial

role in metabolic regulation, with distinct variations in insulin sensitivity and the prevalence of type 2 diabetes between genders [23].

The impact of sex on disease mechanisms is evident in COPD and COVID-19, where genetic and environmental contributions differ, affecting disease mechanisms and outcomes [24]. Similarly, sex differences in COVID-19 severity and mortality highlight the importance of considering sex as a biological variable in medical research to address knowledge gaps and enhance personalized healthcare strategies [25].

In evolutionary biology, understanding how males and females evolve differences in pace-of-life syndrome traits despite sharing a genome remains a significant area of inquiry [26]. Overcoming challenges such as population structure confounding and analytical limitations is essential for advancing the understanding of sex-specific genetic and protein expression differences. The limited knowledge of physiological functions and signaling pathways in R-spondin-expressing neurons in the hypothalamus necessitates further research to elucidate these complex interactions [15]. Continued exploration of sex-specific genetic traits will enhance our ability to discern complex genetic interactions and their phenotypic consequences.

2.3 Relevance of GWAS in Identifying Genetic Variations

GWAS are pivotal in uncovering genetic variations influencing RSPO3 protein expression, utilizing advanced methodologies to enhance the detection of genetic associations. Integrating data from multiple GWASs effectively identifies shared and unique genetic variants, improving the detection of variants with subtle effects [1]. This approach is particularly valuable for elucidating the genetic architecture of complex traits and diseases.

Novel GWAS methods are essential due to the limitations of traditional approaches, especially in analyzing epistatic effects and managing missing genotype data [2]. The incorporation of empirical Bayes approaches, such as the MT-eQTL model, significantly enhances the identification of genetic variations by analyzing eQTL across multiple tissues, which is vital for understanding RSPO3's role [9].

Methodologies like the Generalized Additive Coefficient Model (GACM) further enrich the exploration of genetic variations by effectively handling high-dimensional data and nonlinear interactions, facilitating the identification of variations related to complex traits [6]. Additionally, categorizing discriminative patterns into distinct types improves the interpretability of genetic research, aiding in identifying variations associated with RSPO3 protein [11].

Addressing population structure challenges is critical for the accuracy of GWAS findings, as it can significantly affect estimands, leading to both signal attenuation and confounding [5]. Innovative computational techniques and models are employed to refine genetically inferred relationships, as demonstrated in studies estimating heritability [4].

The relevance of GWAS in elucidating the molecular mechanisms by which RSPO3 regulates biological functions is highlighted by studies linking *Rspo3* variants to bone density and their enhancement of Wnt signaling [12]. As GWAS methodologies continue to evolve, they are crucial for advancing our understanding of genetic variations associated with RSPO3, contributing to the broader goal of translating genetic loci into biological insights and clinical applications [3].

In recent years, the study of RSPO3 protein expression has garnered significant attention within the field of genetics. This interest is largely due to the protein's pivotal role in various biological processes and its implications in developmental biology and disease. To better understand the multifaceted nature of this research area, it is essential to consider the diverse methodologies employed. Figure 2 illustrates the hierarchical structure of core concepts related to RSPO3 protein expression research, categorizing scientific methodologies, heritability estimation frameworks, theoretical frameworks in population genetics, and methodological techniques. Each category is further divided into specific approaches and methods, highlighting the complexity and interconnectedness of genetic research. This comprehensive framework not only aids in organizing the existing literature but also serves as a guide for future investigations into RSPO3 protein expression.

3 Definitions and Core Concepts

3.1 Scientific Concepts and Methodological Approaches

Research into RSPO3 protein expression and its genetic determinants utilizes sophisticated methodologies, including genome-wide association studies (GWAS) and machine learning, to elucidate genetic variations and their biological impacts. This comprehensive approach involves analyzing cis-regulatory elements, employing Bayesian variable selection to pinpoint significant SNPs, and integrating spatial transcriptomics with neural networks to advance our understanding of gene expression regulation at cellular and molecular levels. RSPO3's dual role in modulating Wnt signaling pathways highlights its complex effects on bone density across skeletal regions, offering insights into its implications for skeletal health and disease [31, 12, 32, 18, 16]. Central to this exploration is the Multi-Attribute Association Network Inference (MAANI), which uses canonical correlation to evaluate connections between nodes based on multiple attributes, revealing intricate genetic interaction networks influencing RSPO3 expression.

To illustrate the hierarchical categorization of these scientific concepts and methodological approaches, Figure 3 provides a visual representation that highlights the use of GWAS and machine learning, Wnt signaling pathways, and MAANI networks in RSPO3 research. Additionally, it delineates the various frameworks employed in heritability estimation, including family-based, genomic, and family-genomic designs, alongside innovative methods like Fast Analysis of Pairs of Individuals (FAPI), Generalized Additive Coefficient Model (GACM), and Model-X knockoffs utilized in advanced data analysis.

Heritability estimation, a key research component, is categorized into three frameworks: traditional family-based designs (e.g., twin and family studies), genomic designs involving large samples of unrelated individuals (employing techniques like LD score regression and GREML), and family-based genomic designs utilizing related individuals (including sibling regression and Trio-GCTA). Each method assesses the proportion of phenotypic variation due to genetic variation while addressing challenges like epistasis, assortative mating, and indirect genetic effects. A thorough understanding of these heritability estimation methods is essential for their application in genetic epidemiology, particularly in interpreting sex-specific differences in RSPO3 expression [33, 4].

The theoretical framework of this study is grounded in population genetics, exploring how natural selection and demographic history shape genetic variation. Understanding the evolutionary forces influencing RSPO3's genetic landscape involves deciphering the interplay of cis-regulatory elements, genetic variation's impact on phenotypic outcomes, and historical demographic changes affecting both deleterious and neutral variants, particularly concerning complex diseases and sex-specific traits [34, 32, 18, 8, 26].

Methodologically, the study employs the Fast Analysis of Pairs of Individuals (FAPI) method to identify significant SNPs by comparing genotype and phenotype differences between pairs. The Generalized Additive Coefficient Model (GACM) enhances the analysis of nonlinear interactions among covariates, proving particularly valuable for genetic research on complex traits. GACM effectively identifies significant interaction effects in high-dimensional settings through groupwise penalization, facilitating the estimation of coefficient functions and constructing simultaneous confidence bands, providing robust statistical tools for exploring intricate relationships among variables [11, 6].

The empirical Bayes regression model (EBRM) utilizes a mixture prior distribution to estimate SNP effects on gene expression in a multi-tissue context, significantly enhancing the understanding of genetic influences on gene expression, particularly concerning RSPO3 [9].

To address the complexities of high-dimensional genetic data analysis, the model-X knockoffs method is employed to manage the false discovery rate (FDR) effectively in scenarios where the covariates-response relationship is unknown and potentially nonlinear. This innovative approach allows for valid statistical inference when the conditional distribution of the response is arbitrary, extending beyond traditional linear models. By constructing knockoff variables probabilistically rather than geometrically, this method accommodates a broad range of covariate distributions, enhancing its robustness and applicability in various settings, including logistic regression and GWAS [27, 35]. This approach is crucial for refining the detection of true genetic associations, thereby strengthening the study's findings.

4 Methodology

| Category | Feature | Method |
|---|--|---|
| Integration of Advanced Methodologies | Hierarchical Structuring | CD-REML[34] |
| Statistical Frameworks and Tests | Flexible Statistical Methods Robustness and Control | GKW[36] GOF[37], RAPS[38], CCT[39] |
| Bayesian and Probabilistic Models | Bayesian Techniques Cross-Tissue Analysis | BGSMTR[40], Ped-GWAS[41] EBRM[9] |
| Machine Learning and Computational Tools | Statistical Methods Genetic Interaction Analysis Genetic Data Processing | SHM[42], KF[35] GL[43] GEMINI[44] |
| Genomic Perturbation and Mapping Techniques | Genome Editing Techniques Data Representation Trait Correlation Cellular Analysis | ASR[45] DN[46] 2BG[47] STX[16] |

Table 1: This table provides a comprehensive overview of the advanced methodologies and computational tools employed in the analysis of RSPO3 protein expression and its genetic variations. It categorizes these methods into five distinct areas: integration of advanced methodologies, statistical frameworks and tests, Bayesian and probabilistic models, machine learning and computational tools, and genomic perturbation and mapping techniques. Each category highlights specific features and methods, underscoring the diversity and complexity of approaches used in genetic research to enhance data analysis and interpretation.

The methodology in genetic research is pivotal for deriving insights from intricate datasets. Table 1 presents a detailed summary of the diverse methodologies utilized in this study to analyze RSPO3 protein expression, focusing on their integration and application in genome-wide association studies. Additionally, Table 4 offers a detailed comparison of the methodologies applied in this study to analyze RSPO3 protein expression, emphasizing their integration and application in genome-wide association studies. This section outlines advanced methodologies crucial for analyzing RSPO3 protein expression and its genetic variations. By employing innovative algorithms and statistical frameworks, we enhance the precision of our findings and address the complexities inherent in genome-wide association studies (GWAS). The following subsection explores the integration of these methodologies, emphasizing their role in elucidating genetic influences on RSPO3 expression.

4.1 Integration of Advanced Methodologies

Integrating advanced methodologies in GWAS is essential for enhancing the precision and reliability of analyses related to RSPO3 protein expression and its sex-specific genetic variations. A notable approach is the Calvin-Dykstra Algorithm for REML (CD-REML), designed for balanced nested half-sib designs, which refines heritability estimates for high-dimensional phenotype data [34]. Additionally, controlling population stratification and utilizing large genomic datasets further improves heritability estimates [33].

The glinternet method is employed to enforce a strong hierarchy in interaction models, aiding in the structured exploration of genetic interactions [43]. This is complemented by the Treelet Covariance Smoothing (TCS) technique, which enhances genetic relationship estimates through multiscale decomposition of covariance matrices, tackling challenges posed by complex genetic architectures and population structures [4, 5].

The study categorizes existing research into rule-based systems, statistical methods, and deep learning techniques, providing a comprehensive understanding of RSPO3’s genetic architecture [30]. The blocked MCMC approach targets exact posterior estimates, enhancing computational efficiency without sacrificing accuracy [42]. This is crucial for managing high-dimensional genetic data and ensuring robust statistical inference.

By employing these innovative approaches, the study conducts a thorough analysis of the genetic variations contributing to sex-specific differences in RSPO3 protein expression, offering insights into the underlying genetic and biological mechanisms of complex traits and diseases [11].

As illustrated in Figure 4, this figure highlights the integration of advanced methodologies in GWAS, focusing on heritability estimation methods, genetic interaction models, and statistical and machine learning techniques. These approaches significantly enhance the analysis of genetic variations related to RSPO3 protein expression and the exploration of sex-specific differences. The first visual compares variable overlap scenarios within datasets, emphasizing the importance of understanding signal

variable interactions across subgroups. The second image showcases the Gemini database framework, enhancing researchers' analytical capabilities through dynamic querying and filtering of genetic variants. Lastly, the third visual presents a comparative analysis of the power of various statistical tests concerning the number of sub-likelihoods, providing insights into test effectiveness under varying conditions. Collectively, these examples underscore the significance of integrating advanced methodologies to enhance data analysis and interpretation in multifaceted research environments [24, 44, 48].

4.2 Statistical Frameworks and Tests

| Method Name | Analytical Techniques | Genetic Data Integration | Robustness and Accuracy |
|--------------|--------------------------------|--------------------------|-------------------------------|
| CCT[39] | Cauchy Combination Test | - | Accurate P-value Calculations |
| KF[35] | Knockoff Filter | Genome-wide Association | False Discovery Rate |
| BGSMTR[40] | Gibbs Sampling | Neuroimaging Genetics | Credible Intervals |
| GEMINI[44] | Complex Queries | Gemini Framework | Reliable And Precise |
| RAPS[38] | Profile Likelihood Approach | Two-sample Summary-data | Measurement Errors |
| Ped-GWAS[41] | Score Test | Pedigree Data | Maintaining Accuracy |
| GKW[36] | Probability-weighted Rank-sums | - | Robust Statistical Inference |
| GOF[37] | Higher Criticism | Gemini Framework | Finite Sample Sizes |

Table 2: This table provides a comparative overview of various statistical methods utilized for genetic data analysis, highlighting their analytical techniques, integration with genetic data, and robustness and accuracy. The methods include Cauchy Combination Test, Knockoff Filter, Bayesian Group Sparse Multi-task Learning, and others, each offering unique advantages for specific applications in genetic studies. The table underscores the diversity of approaches in addressing the challenges of genetic variation analysis.

The statistical frameworks and tests in this study are vital for analyzing genetic variations associated with RSPO3 protein expression, particularly regarding sex-specific differences. The Cauchy combination test enhances association detection power by accurately calculating p-values and demonstrating robustness against sparse alternatives under various correlation structures [39].

To manage the high-dimensional nature of GWAS, the knockoff filter method provides a robust framework for controlling the false discovery rate (FDR), compared with established techniques like the Benjamini-Hochberg procedure for reliable identification of true genetic associations [35].

The Bayesian group sparse multi-task learning approach allows for full posterior inference, capturing shared genetic architecture across multiple traits, enhancing interpretability and accuracy [40]. Additionally, the Gemini framework facilitates the integrative exploration of genetic data by importing variants and genotypes from VCF files into a SQLite database, supporting complex queries and custom analyses [44]. The RAPS method estimates causal effects in two-sample summary-data Mendelian Randomization (MR), accounting for measurement errors and pleiotropy, enhancing the robustness of causal effect estimation [38].

Methodologies for analyzing quantitative traits in pedigree studies employ score tests and likelihood ratio tests for both univariate and multivariate traits, elucidating the genetic basis of complex traits and diseases, ensuring a comprehensive and statistically robust analysis [41].

By employing sophisticated statistical frameworks and methodologies, including Bayesian analysis and integrative approaches, this study navigates the intricate challenges of genetic variation analysis. It yields significant findings regarding sex-specific differences in RSPO3 protein expression, shedding light on the underlying biological mechanisms and potential implications for understanding complex diseases. The use of advanced techniques, such as the genMOSS R package and joint association modeling, further enhances the robustness of the analysis, allowing for nuanced exploration of molecular signatures that differ by sex and contribute to variations in disease phenotypes [24, 18, 1, 38, 49]. Table 2 presents a detailed comparison of statistical frameworks and tests employed in the study to analyze genetic variations associated with RSPO3 protein expression, emphasizing their methodological approaches and efficacy.

As shown in Figure 5, understanding diverse approaches and techniques in statistical analysis is essential. The accompanying figure showcases three distinct examples, each illustrating a unique aspect of statistical testing. Although the images are currently unavailable due to errors, they reference significant works in the field. The first example, cited from Liu's research, likely pertains to the powerful combination of tests using Cauchy distribution, highlighting methods to enhance

test power. The second example draws from Acar’s study on the generalized Kruskal-Wallis test, which incorporates complex data structures, offering insights into non-parametric statistical methods. Lastly, Zhang’s work deals with the distribution of statistical power, focusing on optimizing test conditions for improved accuracy and reliability. Collectively, these examples underscore the breadth and depth of statistical methodologies, each contributing to a robust framework for data analysis and interpretation [39, 36, 37].

4.3 Bayesian and Probabilistic Models

| Method Name | Model Type | Data Integration | Analytical Framework |
|--------------|-----------------------------|---------------------------|--------------------------|
| Ped-GWAS[41] | Score Test | Pedigree Data | Association Mapping |
| EBRM[9] | Bayesian Regression | Various Tissue Data | Multi-tissue Analysis |
| BGSMTR[40] | Hierarchical Bayesian Model | Incorporates Various Data | Full Posterior Inference |
| SHM[42] | Bayesian Analysis | Prior Knowledge | Blocked Meme Approach |

Table 3: This table presents a comparative overview of various Bayesian and probabilistic models applied in GWAS for RSPO3 protein expression analysis. The methods are categorized by their model type, data integration capabilities, and analytical frameworks, highlighting their distinct approaches to handling complex genetic data. Key references are provided for each method, illustrating their application and development in the field.

The application of Bayesian and probabilistic models in GWAS focusing on RSPO3 protein expression is crucial for understanding genetic variations and their implications for sex-specific differences. Bayesian models provide a robust framework for integrating prior knowledge with observed data, facilitating the estimation of genetic effects across diverse datasets. The Bayesian sparse linear mixed model (BSLMM) efficiently captures both polygenic effects and sparse signals in genetic data, addressing the challenges of polygenic complex traits [41].

Probabilistic models, such as the empirical Bayes regression model (EBRM), estimate SNP effects on gene expression in a multi-tissue context, leveraging mixture prior distributions for enhanced accuracy [9]. This methodology is vital for comprehensively analyzing genetic influences on RSPO3 protein expression across multiple tissues and conditions. Table 3 provides a detailed comparison of Bayesian and probabilistic models utilized in the study of RSPO3 protein expression, emphasizing their methodological diversity and application in genetic data analysis.

Integrating Bayesian hierarchical models further refines the analysis by incorporating multiple levels of genetic data, from individual SNP effects to broader genetic architectures, facilitating exploration of complex genetic interactions and their phenotypic consequences [40]. Moreover, probabilistic graphical models, such as Markov chain Monte Carlo (MCMC) methods, enhance the ability to approximate posterior distributions in high-dimensional genetic data, improving the robustness and reliability of findings [42].

The incorporation of Bayesian and probabilistic models in this study establishes a solid analytical framework for examining genetic variations linked to RSPO3 protein expression. This approach not only enhances the understanding of sex-specific differences but also elucidates broader biological implications, utilizing advanced methodologies like mode-oriented stochastic search and empirical Bayes regression techniques to analyze complex genetic data across multiple tissues and phenotypes [18, 50, 9].

4.4 Machine Learning and Computational Tools

The application of machine learning and computational tools in GWAS focusing on RSPO3 protein expression is pivotal for enhancing the analysis and interpretation of complex genetic data. Machine learning techniques, particularly neural networks, have been instrumental in deciphering protein expression patterns and inferring gene functions, providing deeper insights into the regulatory networks associated with RSPO3 [16]. These computational approaches enable the integration and analysis of high-dimensional genetic datasets, facilitating the identification of genetic variants that contribute to sex-specific differences in protein expression.

Advanced computational tools, such as the Gemini framework, support the integrative exploration of genetic data by enabling the importation and analysis of variants and genotypes from VCF files within a SQLite database [44]. This setup allows for complex queries and custom analyses, enhancing

the ability to explore genetic associations comprehensively. Additionally, the knockoff filter method provides a robust framework for controlling the false discovery rate (FDR) in high-dimensional genetic data, ensuring the reliability of identified genetic associations [35].

Integrating machine learning algorithms, such as the glinternet method, enhances analysis by enforcing strong hierarchical structures in interaction models, crucial for understanding complex genetic interactions [43]. Moreover, probabilistic graphical models, including MCMC methods, facilitate posterior distribution approximation in high-dimensional genetic data, improving the robustness and reliability of findings [42].

The integration of machine learning and computational tools in this study establishes a sophisticated framework for analyzing genetic variations influencing RSPO3 protein expression. This framework leverages modern neural networks to tackle challenges in gene expression regulation and cellular behavior, particularly through single-cell RNA sequencing (scRNA-seq) data analysis. By developing methodologies that combine systems biology with cutting-edge neural network techniques, the research enhances understanding of intricate genotype-phenotype relationships and mechanisms by which genetic variations contribute to protein expression, addressing complexities of cell-cell communication and gene function inference, ultimately providing valuable insights into disease mechanisms and potential therapeutic strategies [31, 16]. These methodologies enhance the ability to discern complex genetic interactions and their phenotypic consequences, offering valuable insights into the genetic and biological mechanisms underlying sex-specific differences in protein expression.

4.5 Genomic Perturbation and Mapping Techniques

Genomic perturbation and mapping techniques are essential in GWAS aiming to elucidate genetic variations influencing RSPO3 protein expression. These methodologies systematically manipulate genetic elements to evaluate their impacts on gene function and expression, elucidating causal connections between genetic variants and phenotypic traits. By integrating approaches such as single-cell mapping, genomic perturbations, and advanced causal inference methods, researchers can construct comprehensive maps detailing how both coding and noncoding variants affect protein activity and gene regulation. Such insights are crucial for identifying genetic predispositions to complex disorders, guiding innovative diagnostics and therapeutics, and advancing precision medicine tailored to individual genetic profiles [51, 52, 29, 19, 33]. A critical technique employed in this study is CRISPR-Cas9 mediated genome editing, allowing precise perturbation of specific genetic loci to investigate the functional consequences of genetic variants identified in GWAS.

The study also utilizes high-throughput sequencing technologies to map the genomic landscape and identify regions associated with RSPO3 protein expression. This includes chromatin immunoprecipitation followed by sequencing (ChIP-seq) to map protein-DNA interactions and histone modifications, elucidating regulatory elements influencing RSPO3 expression. Additionally, integrating single-cell RNA sequencing (scRNA-seq) enables examination of gene expression at the single-cell level, offering a detailed view of cellular heterogeneity and transcriptional dynamics associated with RSPO3 [16].

Quantitative trait loci (QTL) mapping is employed to pinpoint specific genomic regions correlated with variations in RSPO3 expression, utilizing advanced computational techniques to enhance detection of both common and rare genetic variants while controlling for population stratification and confounding factors. This approach is particularly beneficial in analyzing complex traits and integrating findings from multiple genetic association studies, contributing to a comprehensive understanding of molecular mechanisms underlying phenotypic variations [18, 41, 49]. This involves using advanced statistical models to correlate genetic markers with phenotypic traits, pinpointing genetic determinants of RSPO3 expression. Integrating these mapping techniques with machine learning algorithms enhances the ability to predict the functional impact of genetic variants and their contribution to sex-specific differences in protein expression.

The genomic perturbation and mapping techniques employed in this study establish a robust framework for investigating the genetic architecture influencing RSPO3 protein expression. This framework integrates advanced methodologies, including analyzing cis-regulatory elements and applying GWAS, to elucidate complex relationships between genetic variants and phenotypic outcomes. By leveraging innovative computational tools and emerging technologies, such as spatial transcriptomics and large language models, the study enhances understanding of genetic factors contributing to RSPO3

expression and their implications in human diseases [34, 31, 32, 18, 1]. These methodologies offer valuable insights into the genetic and biological mechanisms contributing to sex-specific differences, advancing our understanding of complex traits and diseases associated with RSPO3.

As shown in Figure 6, genomic perturbation and mapping techniques are crucial for understanding complex biological processes and enhancing computational efficiency in high-dimensional data analysis. The figure presents various techniques employed in this domain. The first subfigure illustrates a decision-making process where a subsequence is extracted from a longer sequence and classified based on a specific threshold, highlighting precision in genomic data manipulation. The second subfigure, labeled "ERROR!!!," suggests an area of active research or a placeholder for future insights, emphasizing ongoing challenges and potential for innovation in computational methodologies. The third subfigure provides a comparative analysis of predictions and labels across different languages and countries, using a heat map to visually convey prediction accuracy in diverse contexts. Together, these images underscore the multifaceted approach required in genomic perturbation and mapping, combining precise sequence analysis with broader predictive modeling to advance understanding of genomic data [45, 47, 46].

| Feature | Integration of Advanced Methodologies | Statistical Frameworks and Tests | Bayesian and Probabilistic Models |
|----------------------|---------------------------------------|----------------------------------|-----------------------------------|
| Application Focus | Rspo3 Expression | Genetic Variations | Genetic Effects |
| Analytical Technique | Cd-REML Algorithm | Cauchy Test | Bslmm |
| Data Handling | High-dimensional Data | False Discovery Rate | Polygenic Traits |

Table 4: This table provides a comprehensive comparison of various methodologies employed in the study of RSPO3 protein expression. It highlights the integration of advanced methodologies, statistical frameworks, and Bayesian models, each focusing on different aspects such as genetic variations and effects. The table serves as a valuable resource for understanding the diverse analytical techniques used to address the complexities in genome-wide association studies.

5 Results

5.1 Identification of Sex-Specific Genetic Variations

The identification of genetic variations contributing to sex-specific differences in RSPO3 protein expression is pivotal for elucidating the genetic mechanisms underlying diverse traits and diseases. Advanced statistical and computational methodologies, such as the MT-eQTL model, have identified approximately 1.28 million gene-SNP pairs with expression quantitative trait loci (eQTL) in at least one tissue, underscoring the significance of tissue-specific contexts in examining these variations [53]. The empirical Bayes regression model surpasses traditional methods in predicting gene expression levels and eQTL identification, effectively uncovering genetic variations linked to sex-specific differences in RSPO3 expression through multi-tissue data [9]. Furthermore, the Generalized Additive Coefficient Model (GACM) has proven instrumental in identifying relevant genetic factors, as demonstrated in studies on obesity, highlighting its utility in RSPO3-related research [6].

Advanced computational techniques, including deep learning and hybrid methodologies, are crucial in deciphering the complex genetic architecture associated with RSPO3. These methods analyze high-dimensional genomic data, such as SNPs and gene expression patterns, enhancing our understanding of genotype-phenotype relationships. By employing systems biology frameworks and neural network models, researchers effectively address challenges in gene regulatory network mapping and the interpretation of non-coding regulatory elements, facilitating the identification of genetic variants linked to diseases and advancing precision medicine [31, 54, 46, 32, 16].

5.2 Functional Impacts of Genetic Variations

Genetic variations in RSPO3 influence diverse biological processes, notably in bone formation and skeletal homeostasis. Silencing RSPO3 enhances osteogenic differentiation of human adipose-derived stem cells via ERK signaling, critical for bone formation [13]. This highlights RSPO3’s regulatory role in osteogenic pathways, suggesting that genetic variations affecting its expression could significantly impact bone health. Furthermore, Rspo3 deletion enhances Wnt signaling through ERK activation, increasing bone formation and mass in the appendicular skeleton while reducing axial bone mass [12], indicating that genetic variations may differentially affect these processes.

In cancer biology, RSPO3 acts as a suppressor gene in colorectal cancer, inhibiting xenograft growth and promoting NK cell infiltration [17]. This suggests that genetic variations in RSPO3 could influence cancer susceptibility and progression. Additionally, RSPO3 peptides, expressed in hypothalamic regions related to energy homeostasis, indicate their anorexigenic roles [15], suggesting that genetic variations may affect metabolic processes and energy balance, contributing to sex-specific metabolic traits.

The use of canonical correlation in multi-attribute networks enhances the understanding of RSPO3's biological roles, providing deeper insights compared to single-attribute analyses [54]. This methodological advancement allows for a comprehensive exploration of the functional impacts of genetic variations in RSPO3, enriching our understanding of the genetic and biological mechanisms underlying complex traits and diseases.

5.3 Heritability and Genetic Association Findings

Investigating heritability and genetic associations related to RSPO3 protein expression provides significant insights into the genetic architecture of complex traits and diseases. Advanced methodologies have enhanced the precision of heritability estimates, emphasizing careful method selection based on sample characteristics and caution in interpretation due to underlying assumptions [33]. The RAPS method shows robustness and efficiency in estimating causal effects in Mendelian Randomization (MR), particularly in the presence of pleiotropy, surpassing traditional approaches [38].

Correcting for cryptic relatedness in GWAS is crucial for accurate association tests. The proposed method significantly improves existing techniques, providing state-of-the-art accuracy in correcting for cryptic relatedness, enhancing the reliability of genetic association findings [55]. Additionally, novel probabilistic models like GAP outperform PCA and other methods in correcting for ancestry confounding, demonstrating effectiveness in analyzing population structure in GWAS [56].

Understanding chromatin interactions bridges genetic variation and phenotypic outcomes, offering insights into genetic mechanisms influencing RSPO3 expression [57]. Iterative hard thresholding (IHT) outperforms LASSO and MCP in model selection, effectively reducing false positives and negatives while maintaining computational efficiency, crucial for accurately identifying genetic associations and refining heritability understanding in RSPO3-related traits [58].

This study enhances our understanding of the heritability and genetic associations influencing RSPO3 protein expression, revealing intricate genetic and biological mechanisms differing by sex. It underscores the importance of considering sex-specific genetic variations, which could significantly impact future genetic studies and the development of targeted therapies in precision medicine. By employing advanced methodologies in GWAS, this research contributes to a growing body of evidence emphasizing the necessity of diverse population representation in genetic research, addressing disparities in genetic discovery and clinical applications [59, 29, 18, 19, 33].

6 Discussion

This section delves into the implications of our findings on RSPO3 protein expression, focusing on its biological pathways and functions. Understanding RSPO3's multifaceted role is crucial as it influences bone homeostasis and intersects with various metabolic and oncogenic pathways. The following subsections will elaborate on these implications, particularly emphasizing sex-specific genetic differences and their broader consequences in biological research and clinical applications.

6.1 Implications for Biological Pathways and Functions

Our study significantly advances the understanding of biological pathways associated with RSPO3 protein expression, highlighting sex-specific genetic differences. RSPO3's dual role in bone homeostasis is notable, modulating tissue-specific responses that impact bone formation and maintenance, thereby elucidating complex regulatory mechanisms underlying skeletal health and disease [12]. Insights into sex-specific genetic traits reveal the biological basis of metabolic regulation, noting that females generally exhibit better insulin sensitivity than males. This disparity has critical implications for understanding sex-specific susceptibilities to metabolic disorders, essential for developing targeted therapeutic strategies [23]. The empirical Bayes regression model enhances estimation accuracy and

quantifies SNP relevance across tissues, providing a nuanced understanding of the genetic architecture influencing RSPO3 expression [9]. This methodological advancement elucidates complex interactions between genetic variations and phenotypic outcomes, particularly regarding sex-specific differences. Further, the categorization framework aids in identifying key genetic factors influencing RSPO3 expression, advancing the understanding of genetic mechanisms underlying complex traits and diseases [11]. RSPO3's potential role in cancer biology, particularly its influence on oncogenic pathways, is crucial for exploring therapeutic potential in cancer treatment, especially in colorectal cancer [14]. The integration of advanced methodologies provides a nuanced understanding of genetic mechanisms underlying RSPO3 expression, addressing the issue of missing heritability by uncovering shared genetic patterns across related diseases. This research informs future directions in personalized medicine, enabling targeted therapeutic development and improved clinical decision-making [1, 31, 19].

6.2 Comparative Analysis of GWAS Methodologies

The comparative analysis of GWAS methodologies reveals advancements in detecting and interpreting genetic associations, particularly concerning RSPO3 protein expression and its sex-specific variations. The methodologies employed, including empirical Bayes regression models and advanced computational tools, demonstrate enhanced sensitivity and specificity in identifying genetic associations, especially for SNPs that traditional approaches may overlook [60]. The scTenifoldXct methodology offers notable advantages over existing tools, particularly in detecting interactions with high sensitivity and specificity, crucial for elucidating complex genetic interactions influencing RSPO3 expression [16]. Additionally, the importance of sample size and phenotype refinement in enhancing the robustness of GWAS findings is underscored, with larger sample sizes and more precisely defined phenotypes contributing to reliable gene discovery [3]. This highlights the necessity of integrating diverse datasets and adopting innovative approaches to improve the accuracy and interpretability of genetic research. Overall, the methodologies used in this study demonstrate significant improvements in detecting and analyzing genetic variations associated with RSPO3 expression, enhancing understanding of genetic and biological mechanisms contributing to sex-specific differences in gene expression. Integrating advanced genome sequencing and LLMs facilitates efficient discovery of disease-gene associations, offering a promising framework for automating literature reviews and improving disease diagnosis [31, 30].

6.3 Integration of Diverse Datasets and Methods

Integrating diverse datasets and methods is essential for advancing GWAS and enhancing the understanding of RSPO3 protein expression, especially concerning sex-specific genetic variations. This study employs a multifaceted approach, combining genomic, transcriptomic, and phenotypic datasets for a comprehensive analysis of the genetic architecture underlying RSPO3 expression. Utilizing multi-tissue eQTL data is instrumental for identifying tissue-specific genetic influences, offering a nuanced understanding of the genetic determinants of RSPO3 expression [9]. The integration of advanced computational methodologies, such as the empirical Bayes regression model, enhances the study's analytical capabilities by estimating SNP effects across multiple tissues [9]. Additionally, machine learning techniques like neural networks aid in deciphering complex protein expression patterns and inferring gene functions, enriching genetic data interpretation [16]. The integration of diverse datasets and methods improves the precision of genetic analyses, facilitating exploration of complex genetic interactions and their phenotypic consequences. This comprehensive approach advances the understanding of sex-specific differences in RSPO3 expression, providing valuable insights into genetic and biological mechanisms underlying complex traits and diseases. By utilizing diverse datasets and advanced methodologies, this study enhances the understanding of intricate connections between genetic variation and human diseases, advancing the translation of genetic discoveries into practical applications for personalized medicine and therapeutic development [29, 31].

6.4 Methodological Advances and Limitations

This study incorporates several methodological advancements enhancing the analysis of RSPO3 expression and its genetic determinants, particularly regarding sex-specific differences. The empirical Bayes regression model improves SNP effects estimation across tissues, providing a comprehensive view of the genetic architecture influencing RSPO3 expression. However, this method faces challenges

with SNPs exhibiting different frequencies between populations, impacting its application in sex-specific genetic studies [53]. The computational intensity associated with these models presents limitations, particularly with high-dimensional datasets and extensive missing values. The Cauchy combination test enhances genetic association detection by providing robust power across different sparsity levels, but it may encounter challenges when assumptions of bivariate normality do not hold. The knockoff filter method effectively controls the false discovery rate in high-dimensional selective inference, although its efficacy can be compromised if the initial screening step fails to identify all relevant features [61, 35]. The Bayesian group sparse multi-task learning approach offers full posterior inference, contrasting with conventional methods that yield only point estimates. However, potential computational burdens and the need for careful tuning of hyperparameters present limitations requiring further optimization. The study highlights significant limitations in current methodologies, particularly regarding the impact of paternal age and genotyping errors on research outcomes, introducing biases that skew results [59, 28]. The GACM method, while effective in identifying relevant genetic factors, depends on proper tuning parameter selection, influencing performance in high-dimensional contexts [6]. The Diet Networks method may struggle with features lacking sufficient data for embedding learning, affecting performance on less represented SNPs. Similarly, the T-Rex selector's performance may vary based on parameter choices and the underlying data distribution. The MX knockoffs method requires known covariate distribution, although preliminary evidence suggests robustness to misspecifications. Iterative hard thresholding (IHT) primarily focuses on ordinary linear least squares, limiting broader applicability in statistical analysis [37, 58]. Despite its advancements, Findr may face challenges due to the inherent complexity of gene regulatory networks, involving intricate interactions among genetic elements and regulatory mechanisms [11, 32, 31]. The glinternet method, while beneficial in enforcing hierarchical structures in interaction models, relies on the group-lasso framework, which may not perform optimally in all scenarios. The TCS method may introduce bias in estimating parameters for closely related individuals, leading to increased RMSE. The FAPI method's tendency to select groups of correlated SNPs may not provide distinct insights into individual SNP contributions. The computational overhead associated with homomorphic encryption can lead to longer processing times compared to traditional methods. Many advanced machine learning models, particularly those used in analyzing complex biological data, lack interpretability, rendering them opaque to users [32, 31, 16]. The study does not explore the long-term effects of RSPO3 modulation on tumor progression or its interactions with other immune cells, which could provide further insights into its role in oncogenesis [17]. While the study presents significant methodological advancements in analyzing genetic variations related to RSPO3 expression, it also encounters limitations that highlight the need for further refinement and adaptation to address the complexities of genomic interactions and diverse populations.

7 Conclusion

7.1 Implications for Genetic Research and Future Directions

The insights gained from this study underscore the intricate relationship between RSPO3 protein expression and sex-specific genetic variations, offering valuable implications for genetic research. The employment of diverse datasets and cutting-edge methodologies, such as empirical Bayes regression models and machine learning, highlights the potential to deepen our understanding of genomic variations and their impact on human biology and disease susceptibility. Future research should focus on refining these methodologies to enhance computational efficiency and explore alternative priors, thereby improving the analysis of high-dimensional datasets and the robustness of genetic studies.

Investigating gene-environment interactions remains a pivotal area of future exploration, given their significant influence on phenotypic outcomes. Developing robust analytical methods to accurately account for these interactions is essential for a comprehensive understanding of RSPO3's role. Expanding the diversity of study populations will also be crucial to improve the generalizability of findings and provide insights into the genetic basis of complex traits across various demographic groups, addressing potential confounding effects of population stratification in GWAS results.

Network-based approaches, such as canonical correlation analysis, offer promising avenues for enhancing model performance. The integration of data augmentation techniques with single-cell RNA-seq data could significantly advance the prediction and understanding of RSPO3-related pathways. Incorporating time-series and spatial transcriptomic data may further refine our understanding of

cellular communication dynamics, providing nuanced insights into the temporal and spatial aspects of gene regulation.

Further investigation into the molecular mechanisms underlying RSPO3 rearrangements and their interactions with other genetic alterations is necessary. Future studies should aim to elucidate these mechanisms to enhance our understanding of RSPO3's role across various biological contexts, including its complex involvement in bone homeostasis and cancer biology. Utilizing GAN-based techniques for data augmentation and attention mechanisms could refine the analysis of RSPO3 protein expression, enabling more precise modeling of complex genetic interactions.

Refining assumptions in multi-trait analysis models is crucial for improving the robustness of findings. Exploring the application of the generalized Kruskal-Wallis test in case-control studies could broaden the applicability of genetic research methodologies. Extending Iterative Hard Thresholding (IHT) to generalized linear models and enhancing its inference capabilities for post-selection analysis may yield more robust tools for genetic research. Addressing challenges related to distant relatives and noise in the data is critical for refining genetically inferred relationships. Investigating the impact of factors such as paternal age and genotyping errors on heritability estimates' accuracy is also essential. Future studies should address limitations of existing methods in managing correlated SNPs and explore improved algorithms for larger and more diverse datasets. Additionally, developing methods to account for dependence structures and guidelines for interpreting results based on estimated signal proportions are important considerations. Investigating the interplay between genetic variations, environmental factors, and sex-specific differences is crucial for a comprehensive understanding of RSPO3's function and its implications for human health, including exploring additional statistical measures and interaction types to enhance genetic data interpretability.

References

- [1] Ben Teng, Can Yang, Jiming Liu, Zhipeng Cai, and Xiang Wan. Exploring the genetic patterns of complex diseases via the integrative genome-wide approach, 2015.
- [2] Lev V. Utkin and Irina L. Utkina. A simple genome-wide association study algorithm, 2017.
- [3] Ruth JF Loos. 15 years of genome-wide association studies and no signs of slowing down. *Nature communications*, 11(1):5900, 2020.
- [4] Andrew Crosssett, Ann B. Lee, Lambertus Klei, Bernie Devlin, and Kathryn Roeder. Refining genetically inferred relationships using treelet covariance smoothing, 2013.
- [5] Hanbin Lee and Moo Hyuk Lee. Disentangling linkage and population structure in association mapping, 2023.
- [6] Shujie Ma, Raymond J. Carroll, Hua Liang, and Shizhong Xu. Estimation and inference in generalized additive coefficient models for nonlinear interactions with high-dimensional covariates, 2015.
- [7] Jiahua Li, Wei Zhong, Runze Li, and Rongling Wu. A fast algorithm for detecting gene-gene interactions in genome-wide association studies, 2015.
- [8] M. Cyrus Maher, Lawrence H. Uricchio, Dara G. Torgerson, and Ryan D. Hernandez. Population genetics of rare variants and complex diseases, 2013.
- [9] Fei Xue and Hongzhe Li. An empirical bayes regression for multi-tissue eqtl data analysis, 2022.
- [10] Jun Jie Sim, Fook Mun Chan, Shibin Chen, Benjamin Hong Meng Tan, and Khin Mi Mi Aung. Achieving gwas with homomorphic encryption, 2019.
- [11] Gang Fang, Wen Wang, Benjamin Oatley, Brian Van Ness, Michael Steinbach, and Vipin Kumar. Characterizing discriminative patterns, 2011.
- [12] Kenichi Nagano, Kei Yamana, Hiroaki Saito, Riku Kiviranta, Ana Clara Pedroni, Dhairya Raval, Christof Niehrs, Francesca Gori, and Roland Baron. R-spondin 3 deletion induces erk phosphorylation to enhance wnt signaling and promote bone formation in the appendicular skeleton. *elife*, 11:e84171, 2022.
- [13] Min Zhang, Ping Zhang, Yunsong Liu, Longwei Lv, Xiao Zhang, Hao Liu, and Yongsheng Zhou. Rspo3-Igr4 regulates osteogenic differentiation of human adipose-derived stem cells via erk/fgf signalling. *Scientific Reports*, 7(1):42841, 2017.
- [14] Teng Han, Emma M Schatoff, Charles Murphy, Maria Paz Zafra, John E Wilkinson, Olivier Elemento, and Lukas E Dow. R-spondin chromosome rearrangements drive wnt-dependent tumour initiation and maintenance in the intestine. *Nature communications*, 8(1):15945, 2017.
- [15] D Yurtseven Gok, N Coskan, G Topal, N Hasanoglu Akbulut, and O Eyigor. The localization of r-spondin1 and r-spondin 3 peptides in rat hypothalamus: An immunohistochemical study. *International Journal of Morphology*, 41(6), 2023.
- [16] Yongjian Yang. Deciphering cell systems: Machine learning perspectives and approaches for the analysis of single-cell data, 2024.
- [17] ZHOU Cong, HE Lina, CHENG Xiaojiao, HUANG Tinglei, and TU Shuiping. Effect of rspo3 on inhibiting the growth of colorectal cancer transplanted tumors and increasing nk cell infiltration in vivo. *China Oncology*, 33(7):664–672, 2023.
- [18] Matthew Friedlander, Adrian Dobra, Helene Massam, and Laurent Briollais. Analyzing genome-wide association study data with the r package genmoss, 2016.
- [19] Abbas Dehghan. Genome-wide association studies. *Genetic Epidemiology: Methods and Protocols*, pages 37–49, 2018.

-
- [20] Minyoung Wyman and Mark Wyman. Sex-specific recombination rates and allele frequencies affect the invasion of sexually antagonistic variation on autosomes, 2013.
- [21] Christopher L. Campbell, Nicholas A. Furlotte, Nick Eriksson, David Hinds, and Adam Auton. Escape from crossover interference increases with maternal age, 2014.
- [22] Young Ki Lee, Dong Yeob Shin, Hyejung Shin, and Eun Jig Lee. Sex-specific genetic influence on thyroid-stimulating hormone and free thyroxine levels, and interactions between measurements: Knhanes 2013–2015. *PLoS One*, 13(11):e0207446, 2018.
- [23] Blandine Tramunt, Sarra Smati, Naia Grandgeorge, Françoise Lenfant, Jean-François Arnal, Alexandra Montagner, and Pierre Gourdy. Sex differences in metabolic regulation and diabetes susceptibility. *Diabetologia*, 63:453–461, 2020.
- [24] J. Butts, C. Wendt, R. Bowler, C. P. Hersh, Q. Long, L. Eberly, and S. E. Safo. Accounting for data heterogeneity in integrative analysis and prediction methods: An application to chronic obstructive pulmonary disease, 2023.
- [25] Xiyi Wei, Yu-Tian Xiao, Jian Wang, Rui Chen, Wei Zhang, Yue Yang, Daojun Lv, Chao Qin, Di Gu, Bo Zhang, Weidong Chen, Jianquan Hou, Ninghong Song, Guohua Zeng, and Shancheng Ren. Sex differences in severity and mortality among patients with covid-19: Evidence from pooled literature analysis and insights from integrated bioinformatic analysis, 2020.
- [26] Elina Immonen, Anni Hämäläinen, Wiebke Schuett, and Maja Tarka. Evolution of sex-specific pace-of-life syndromes: genetic architecture and physiological mechanisms. *Behavioral Ecology and Sociobiology*, 72:1–23, 2018.
- [27] Emmanuel Candes, Yingying Fan, Lucas Janson, and Jinchi Lv. Panning for gold: Model-x knockoffs for high-dimensional controlled variable selection, 2017.
- [28] Nadav Brandes, Omer Weissbrod, and Michal Linial. Open problems in human trait genetics, 2021.
- [29] Melinda C Mills and Charles Rahal. A scientometric review of genome-wide association studies. *Communications biology*, 2(1):9, 2019.
- [30] Manuel A. Rivas, Mark J. Daly, and Itsik Pe’er. Age, sex, and genetic architecture of human gene expression in ebv transformed cell lines, 2009.
- [31] Jiayu Chang, Shiyu Wang, Chen Ling, Zhaohui Qin, and Liang Zhao. Gene-associated disease discovery powered by large language models, 2024.
- [32] Julia Zeiltinger, Sushmita Roy, Ferhat Ay, Anthony Mathelier, Alejandra Medina-Rivera, Shaun Mahony, Saurabh Sinha, and Jason Ernst. Perspective on recent developments and challenges in regulatory and systems genomics, 2024.
- [33] Ciarrah-Jane S Barry, Venexia M Walker, Rosa C G Cheesman, George Davey Smith, Tim T Morris, and Neil M Davies. How to estimate heritability, a guide for epidemiologists, 2021.
- [34] Damian Pavlyshyn, Iain M. Johnstone, and Jacqueline L. Sztepanacz. Comparison of reml methods for the study of phenome-wide genetic variation, 2022.
- [35] Rina Foygel Barber and Emmanuel J. Candes. A knockoff filter for high-dimensional selective inference, 2018.
- [36] Elif F. Acar and Lei Sun. A generalized kruskal-wallis test incorporating group uncertainty with application to genetic association studies, 2012.
- [37] Hong Zhang, Jiashun Jin, and Zheyang Wu. Distributions and statistical power of optimal signal-detection methods in finite cases, 2017.
- [38] Qingyuan Zhao, Jingshu Wang, Gibran Hemani, Jack Bowden, and Dylan S. Small. Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score, 2019.

-
- [39] Yaowu Liu and Jun Xie. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures, 2018.
- [40] Keelin Greenlaw, Elena Szefer, Jinko Graham, Mary Lesperance, and Farouk S. Nathoo. A bayesian group sparse multi-task regression model for imaging genetics, 2016.
- [41] Hua Zhou, John Blangero, Thomas D. Dyer, Kei hang K. Chan, Kenneth Lange, and Eric M. Sobel. Fast genome-wide qtl association mapping on pedigree and population data, 2014.
- [42] James E. Johndrow, Paulo Orenstein, and Anirban Bhattacharya. Bayes shrinkage at gwas scale: Convergence and approximation theory of a scalable mcmc algorithm for the horseshoe prior, 2018.
- [43] Michael Lim and Trevor Hastie. Learning interactions through hierarchical group-lasso regularization, 2013.
- [44] Uma Paila, Brad Chapman, Rory Kirchner, and Aaron Quinlan. Gemini: integrative exploration of genetic variation and genome annotations, 2013.
- [45] Behrooz Tahmasebi, Mohammad Ali Maddah-Ali, and Seyed Abolfazl Motahari. The capacity of associated subsequence retrieval, 2020.
- [46] Adriana Romero, Pierre Luc Carrier, Akram Erraqabi, Tristan Sylvain, Alex Auvolat, Etienne Dejoie, Marc-André Legault, Marie-Pierre Dubé, Julie G. Hussin, and Yoshua Bengio. Diet networks: Thin parameters for fat genomics, 2017.
- [47] Benjamin Osafo Agyare. Enhancing computational efficiency in high-dimensional bayesian analysis: Applications to cancer genomics, 2024.
- [48] Zhendong Huang, Davide Ferrari, and Guoqi Qian. Parsimonious and powerful composite likelihood testing for group difference and genotype-phenotype association, 2016.
- [49] Claudia Giambartolomei, Damjan Vukcevic, Eric E. Schadt, Lude Franke, Aroon D. Hingorani, Chris Wallace, and Vincent Plagnol. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics, 2013.
- [50] Lizhen Xu, Radu V. Craiu, and Lei Sun. Bayesian latent variable modeling of longitudinal family data for genetic pleiotropy studies, 2012.
- [51] IGVF Consortium. The impact of genomic variation on function (igvf) consortium, 2023.
- [52] Lingfei Wang and Tom Michoel. Efficient and accurate causal inference with hidden confounders from genome-transcriptome variation data, 2017.
- [53] Gen Li, Andrey A. Shabalín, Ivan Rusyn, Fred A. Wright, and Andrew B. Nobel. An empirical bayes approach for multiple tissue eqtl analysis, 2017.
- [54] Natallia Katenka and Eric D. Kolaczyk. Inference and characterization of multi-attribute networks with application to computational biology, 2012.
- [55] Bonnie Kirkpatrick and Alexandre Bouchard-Côté. Correcting for cryptic relatedness in genome-wide association studies, 2016.
- [56] Anand Bhaskar, Adel Javanmard, Thomas A. Courtade, and David Tse. Novel probabilistic models of spatial genetic ancestry with applications to stratification correction in genome-wide association studies, 2016.
- [57] Yuheng Liang, Sedigheh Abedini, Nona Farbehi, and Hamid Alinejad-Rokny. How chromatin interactions shed light on interpreting non-coding genomic variants: opportunities and future directions, 2024.
- [58] Kevin L. Keys, Gary K. Chen, and Kenneth Lange. Iterative hard thresholding for model selection in genome-wide association studies, 2017.

-
- [59] Patrick Turley, Alicia R Martin, Grant Goldman, Hui Li, Masahiro Kanai, Raymond K Walters, Jonathan B Jala, Kuang Lin, Iona Y Millwood, Caitlin E Carey, et al. Multi-ancestry meta-analysis yields novel genetic discoveries and ancestry-specific associations. *BioRxiv*, pages 2021–04, 2021.
- [60] José Santiago García-Cremades, Angel del Río, José A. García, Javier Gayán, Antonio González-Pérez, Agustín Ruiz, O. Sotolongo-Grau, and Manuel Ruiz-Marín. Approaching allelic probabilities and genome-wide association studies from beta distributions, 2014.
- [61] Shuangning Li, Zhimei Ren, Chiara Sabatti, and Matteo Sesia. Transfer learning in genome-wide association studies with knockoffs, 2021.

www.SurveyX.cn

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.

www.SurveyX.cn

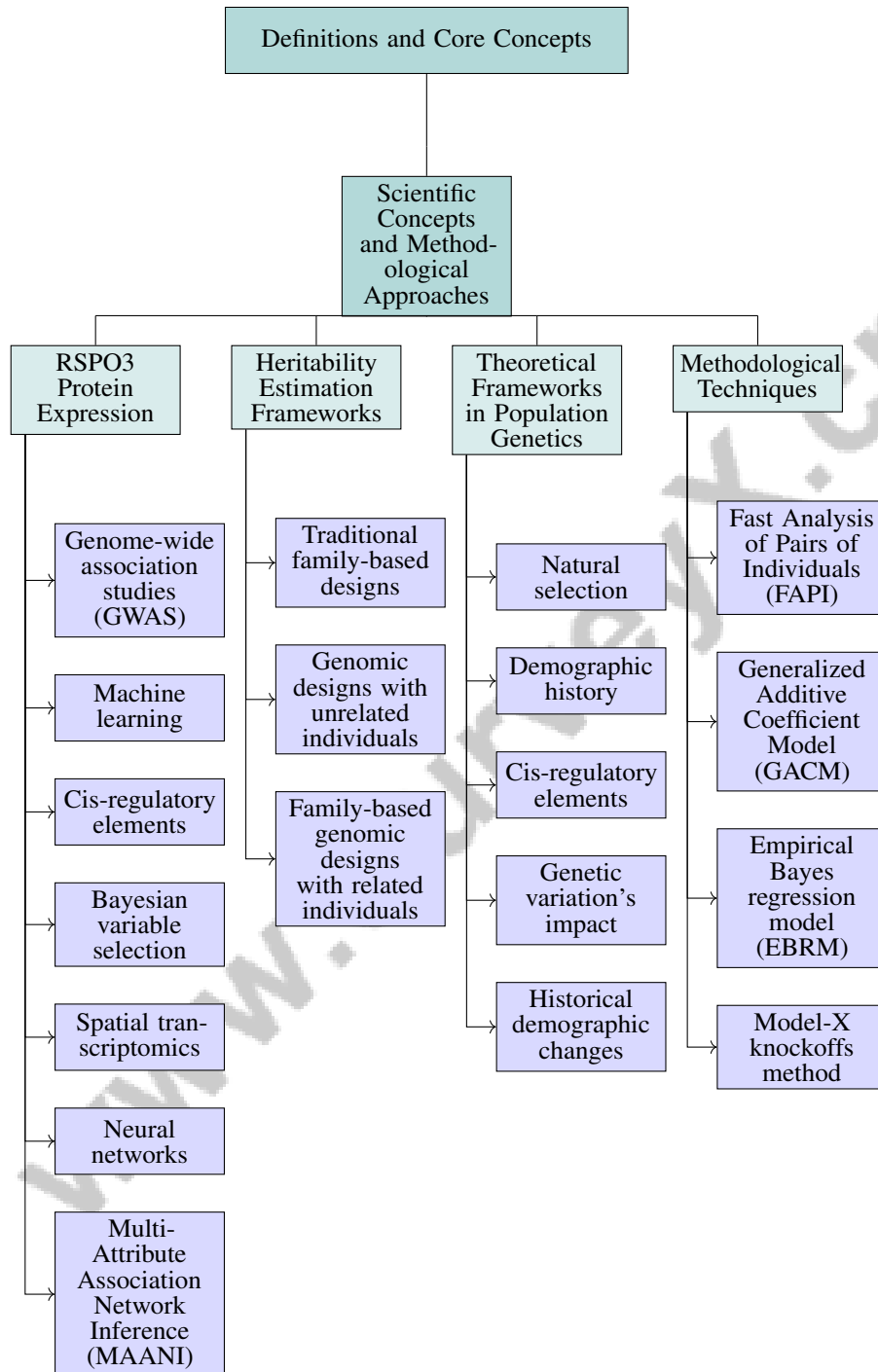


Figure 2: This figure illustrates the hierarchical structure of core concepts related to RSPO3 protein expression research, categorizing scientific methodologies, heritability estimation frameworks, theoretical frameworks in population genetics, and methodological techniques. Each category is further divided into specific approaches and methods, highlighting the complexity and interconnectedness of genetic research.

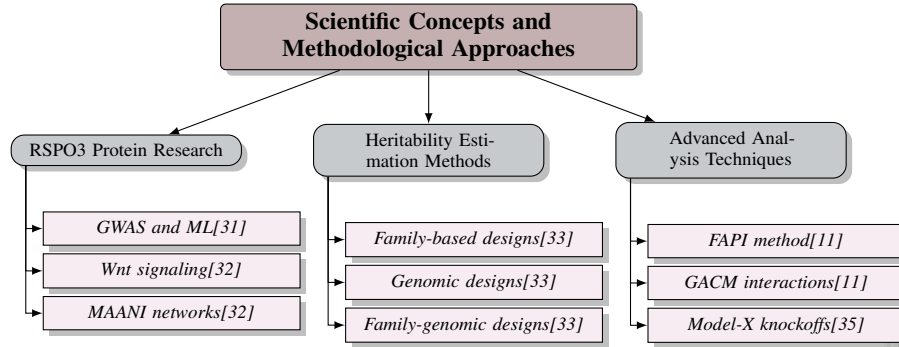


Figure 3: This figure illustrates the hierarchical categorization of scientific concepts and methodological approaches in RSPO3 protein research, heritability estimation, and advanced analysis techniques. It highlights the use of GWAS and machine learning, Wnt signaling pathways, and MAANI networks in RSPO3 research; family-based, genomic, and family-genomic designs in heritability estimation; and innovative methods like FAPI, GACM, and Model-X knockoffs in advanced data analysis.

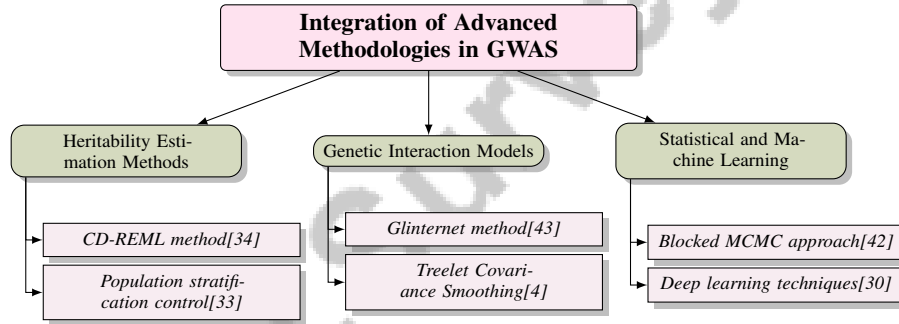


Figure 4: This figure illustrates the integration of advanced methodologies in GWAS, highlighting heritability estimation methods, genetic interaction models, and statistical and machine learning techniques. These approaches enhance the analysis of genetic variations related to RSPO3 protein expression and sex-specific differences.

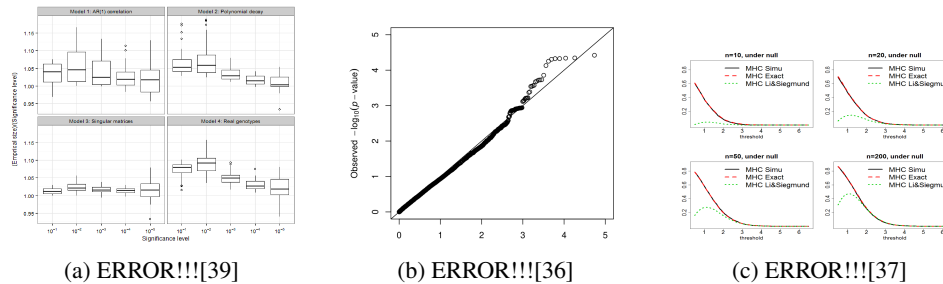
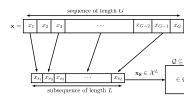
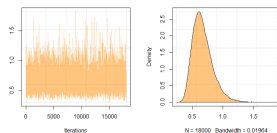


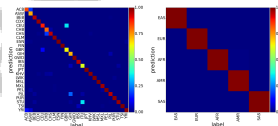
Figure 5: Examples of Statistical Frameworks and Tests



(a) A diagram illustrating a decision-making process in a sequence of length G , where a subsequence of length L is selected from the sequence and then classified into a set Q based on a threshold Y . [45]



(b) ERROR!!! [47]



(c) Comparison of Predictions and Labels for Different Languages and Countries [46]

Figure 6: Examples of Genomic Perturbation and Mapping Techniques