Dean&Francis
ISSN 2959-409X

Multi-omics Integration in AMD: the CFH Gene Case Study

Zheqin Yang

Faculty of Science, McGill University, Montréal, Canada
Corresponding author: zheqin.yang@mail.mcgill.ca

Abstract:

Age-related macular degeneration (AMD) is a progressive retinal disease and a leading cause of irreversible blindness in older adults, imposing a substantial global health burden. Genome-wide association studies (GWAS) have identified over 50 loci associated with AMD, yet most lie in non-coding regions, obscuring the underlying molecular mechanisms. This paper highlights the application of multi-omics integration—a framework that combines genetic associations with regulatory and functional outcomes across molecular layers—to elucidate the genetic risk of AMD, with a particular focus on the CFH gene, a central player in AMD pathogenesis. This study first summarizes key omics data types (genomics, transcriptomics, epigenomics) and core analytical tools, such as colocalization analysis and expression quantitative trait loci (eQTL) mapping. This then examines CFH-centered multi-omics findings, including evidence of shared causal variants between AMD GWAS and CFH eQTLs, and tissue-specific regulatory effects observed in the liver and retinal pigment epithelium. The discussion further extends to other AMD loci (e.g., ARMS2/HTRA1, C3) and outlines technical challenges such as data heterogeneity, tissue accessibility and translational potential. Overall, despite existing complexities, multi-omics integration provides an essential bridge between genetic associations and disease mechanisms, offering a roadmap for improved risk prediction and the development of mechanism-driven, biomarker-based therapeutics for AMD.

Keywords: Age-related macular degeneration; CFH gene; multi-omics integration; GWAS; eQTL; colocalization.

1. Introduction

Age-related macular degeneration (AMD) is a complex retinal degenerative disorder characterized predominantly by the deterioration of the macula, which can lead to the loss of central vision in individuals

ISSN 2959-409X

older than 60 years of age. AMD comprises nearly 9% of total blindness worldwide according to the World Health Organization [1]. According to clinical criteria, AMD can be divided into early and late stages with two subtypes: geographic atrophy, and neovascular AMD (also known as wet AMD). The etiologies associated with AMD appear to be multifactorial, involving aging, environmental exposures, and genetic predisposition. The underlying mechanistic pathology associated with AMD is characterized by the loss of retinal Photoreceptors, drusen core deposits, and chronic retinal pigment equilibrium (RPE) Infiltration [2]. While several anti-VEGF therapies have been developed to address wet AMD, there is still no effective therapy available to address geographic atrophy. It is crucial to understand the underlying genetics and molecular mechanisms of AMD to develop novel preventive and therapeutic interventions to combat AMD.

Genetic basis of AMD has been extensively studied for over two decades. Through Genome-wide association studies (GWAS), over 50 risk loci have been identified, represented by genes in the complementary pathway, lipid metabolism, extracellular matrix remodeling, and angiogenesis [3,4]. Among all, variants located at complement factor H (CFH) and ARMS2/HTRA1 loci show strongest and most reproducible associations. Recently, however most of the risk variants located on non-coding regions, suggesting that their effects may relatively be exerted through regulatory mechanisms versus protein coding effectors. This outcome entails a prime interpretational gap: can GWAS identify statistical associations but cannot clarify the causative genes or the underlying molecular mechanism? In addition, genetic effects are tissue-and cell-type-dependent, so therefore the classic GWAS approach alone can no longer comprehensively answer how this different risk alleles ultimately drive disease pathogenesis.

Multi-omics integration is the coordinated combination of data from various biological layers, including genomics, epigenomic, transcriptomics, proteomics and metabolomics. In the study of AMD, integrating these layers provides a more in-depth understanding of the regulatory cascade from genetic variation to cellular phenotype. Examples include eQTL mapping that associates genetic variants with differential gene expression, and combining chromatin accessibility (ATAC-seq, DNase-seq) and histone modification data to unravel the underlying epigenetic landscape that mediates gene expression [5]. Multi-omics analysis aids in the identification of causal variants through colocalization and fine-mapping techniques, ultimately identifying potential regulatory elements regulating disease susceptibility. Multi-omics approaches compared to single-layer analyses have advantages in interpretability, causal inference, and higher reproducibility within cohorts. Particularly, they hold power in unraveling complex diseases like AMD, wherein the interplay of gene environmental and tissue specific effects is dynamic.

2. Technical Basis and Core Tools of Multi-omics Integration

2.1 Data Types and Characteristics in Multi-omics Analysis

Multi-omics analysis integrates diverse biological data layers—genomics, transcriptomics, epigenomics, proteomics, and metabolomics—to achieve a systematic understanding of complex biological processes. Genomics provides static, inherited information and typically utilizes SNP genotyping, whole-genome sequencing, and GWAS to map the genetic architecture of disease risk. Transcriptomics captures dynamic gene expression patterns through bulk or single-cell RNA sequencing, enabling quantification of expression levels and alternative splicing. Epigenomics encompasses DNA methylation, histone modifications, and chromatin accessibility (e.g., ATAC-seq, DNase-seq), revealing the regulatory landscape that shapes transcription. Proteomics and metabolomics measure downstream molecular phenotypes, with mass spectrometry-based approaches quantifying protein abundance and metabolic fluxes. Together, these layers describe the cascade from genetic variation to cellular and physiological phenotypes.

Each omics layer differs substantially in resolution, dynamic range, temporal behavior, and biological interpretability. Genomic variation is stable across tissues, whereas transcriptomic and epigenomic signals are highly context-specific and responsive to environmental cues. Proteomic and metabolomic traits, in turn, represent cumulative outputs influenced by upstream regulation. Integrating such heterogeneous datasets offers substantial value: it enables investigators to infer causal mechanisms, delineate regulatory hierarchies, and enhance interpretability beyond single-layer analyses. However, the integration process poses technical challenges, including batch effects, inconsistent data structures, variable sample sizes, and difficulties in harmonizing information across spatial and temporal scales. Despite these complexities, multi-omics strategies greatly increase analytical robustness and reproducibility when supported by well-structured cohorts and standardized pipelines [6].

In age-related macular degeneration (AMD), multi-omics integration has become indispensable for disentangling the interplay between genetic predisposition and tissue-specif-

ic regulation, particularly in the retina and retinal pigment epithelium (RPE). Genomic signals identified through GWAS can be mapped to transcriptomic or epigenomic changes via QTL mapping, chromatin accessibility profiling, and fine-mapping approaches, enabling identification of candidate causal variants and regulatory elements. By jointly analyzing genomic, retinal transcriptomic, and epigenomic data, researchers can uncover functional pathways—such as complement activation, lipid metabolism, and inflammatory responses—that contribute to AMD pathogenesis. This integrated framework therefore provides a mechanistic bridge between association signals and biological interpretation, serving as a foundation for downstream functional validation in loci such as CFH, ARMS2/HTRA1, and other AMD-associated regions.

2.2 Key Technologies and Computational Tools

Integrating multi-omics data demands statistical and biogenetic similarity. Common analytical frameworks for integration include colocalization analysis, fine-mapping, eQTL mapping, and sQTL mapping, functional annotation, and integration frameworks. Pathway-wise assays for colocalization analyses assess whether a GWAS and QTL signal co-share a causal variant. Probabilistically-inferred causal variants are spotted in fine-mapping analysis, which works by annotating the locus with candidate SNPs that map the causal variant. eQTL and sQTL mapping quantify the strength, direction, and effect size of the variants on a gene's expression or splicing levels. Functional annotation probes the variation control abilities of each variant. Functional integration into omics datasets provides a networked representation of latent factors across omics layers.

Data homogenization and normalization remains pivotal to mitigate batch effects. Integrations can be horizontal, from across studies of the same omics type, or, vertically, across omics types in the same samples. For the case of AMD, vertical, across omics integrations, say GWAS, transcriptomics as well as epigenomics, is exceedingly advantageous as this allows statistical associations to be shaped biologically by underlaying mechanisms.

3. Multi-omics Integration of the CFH Gene Region in AMD

3.1 Biological Function and Genetic Association of CFH

CFH Complement Factor H (the pleiotropic) gene whose main function is as the key negative inhibitor of the alternate pathway of the innate response complement pathway (ACP)-one of the main branches of innate immunity-CFH plays an important physiological role in tissue homeostasis. Allows pathogen elimination, resisting host cellular hyperactivation-proteolytic activity [2,5,6]. Structurally, CFH consists of 15 complement control protein(CCP)-domains. CCPs 1-4 mediate binding with C3b, inhibiting cleavages into pro-inflammatory components, and CCP 19-20 allows cell surface 10,12 and C Reactive protein (CRP), 12 extracellular glycoprotein/heparan sulfate interactions 10,12This feature allows CFH to selectively suppress ACP activation at host tissue-air interfaces as CFH activation at cell interfaces, a physiological mechanism for retinal health, is critical due to the retina's 40 high metabolic and oxidative load conditions.

The CFH locus on chromosome 1q31 was one of the earliest results to indicate AMD risk association loci and, among Y402H, has emerged as one of the strongest and most reproducible findings in diverse genomes. The missense variant (CCP 7) substitutes a tyrosine with a histidine residue which markedly reduces binding to both HSPGs and CRP (by~50%) [7-9]. Within retinal pigment epithelium (RPE) cells–distinctive epithelial cellular components that sustain photoreceptors and constitute the blood retina barrier—this change in binding impacts CFH localization to the RPE basement membrane thereby leading to an escape of ACP activation. C3a accumulation (an inflammatory anaphylatoxin) and C5b-9 (membrane attack complex) in turn, impels oxidative stress, apoptosis, and secretion of pro-angiogenic factors in RPE cells. and ultimate causation of both dry and wet AMD (through RPE atrophy and choroidal neovascularization, respectively).

Remarkably, subsequent fine-mapping studies based on multi-ethnic GWAS datasets (e.g. the International AMD Genomics Consortium, IAMDGC) saw that additional to Y402H, at a minimum 3-4 independent signals exist underlining genetic loci. Virtually all of these variants are non-coding and appear to largely exert effects via transcriptional or post-transcriptional regulation (as opposed to altering protein structure). This intricate genetic topography, along with the central role played by CFH in retinal immunity, have made it an ideal test bed for addressing how multi-omics data integration leads to resolving "missing mechanism" between non-coding variants and Ahmed pathobiology. Beyond its general role in complement regulation, CFH exerts a highly tissue-specific protective function in AMD. At the retinal pigment epithelium–Bruch's membrane interface, CFH prevents excessive accumulation of C3b and controls continuous low-grade complement turnover. When CFH activity is reduced—due to genetic variants such as Y402H—complement activation products accumulate, leading to oxidative stress,

ISSN 2959-409X

chronic para-inflammation, and drusen formation. This dysregulated complement environment ultimately drives RPE atrophy in dry AMD and promotes choroidal neovascularization in wet AMD.

3.2 Standard Workflow for CFH-centered Multi-omics Analysis

CFH-focused multi-omics workflows share a sequence of steps. Initially, GWAS signal extraction identifies SNPs that may map to CFH-associated AMD through meta-analyses on large samples. Subsequently, linkage disequilibrium extension further expands the coverage with ±500 kb and extracts other variants showing correlation. Colocalization with QTL data is replicated to determine the spatial relationship between AMD signals and eQTLs/sQTLs affecting CFH expression in tissues of the tissue in question. Epigenomic contextualization overlay variants on chromatin accessibility maps and enhancer marks from retinal datasets. Functional annotation evaluates the candidate variants considering governance databases and in silico prediction tools. Integration and prioritization finally integrates statistical/functional evidence to prioritize causal variants and their targets. Altogether, the above pipeline facilitates the extension of GWAS association finding to generating functional hypothesis. For instance, colocalization analyses demonstrated potential modulatory roles of CFH-associated AMD risk variants upstream of CFH expression level either in the liver or retinal pigment equilibrium (RPE) influencing systemic complement function [10].

3.3 Key Findings and Ongoing Debates

Multi-omics analyses consistently replicate causal roles of CFH dysregulation to assessing AMD risk which are still currently under debate. Highlights include overlapping population genetic signals, tissue specificity, epigenetics landscape features and alternative splicing. In addition, high posterior probabilities of colocalization between AMD-GWAS and CFEQTLs supports a shared causal variant [11]. Finally, the evidence supporting the regulatory role of CFH variants may vary from liver and RPE. Moreover, associations between well associated AMD variants have shown regular overlap with functional enhancers characterized by H3K27ac and open chromatin strengthens further support for their regulatory role. The (slow) QTL experimental evidence indicates that the variant risk allele in turn could modulate CFH isoform ratios, affecting complement regulatory efficiency. However, there are still controversies around which variant has a dominant effect, which precise target cell type and about variations effect of environmental interactions on CFH expression levels. Single cell-resolution based and spatial transcriptomic approaches are likely to solve these ambiguities in the near future.

4. Expanded Applications and Challenges of Multi-omics Integration in AMD Research

4.1 Application to Other AMD Risk Loci

In addition to CFH, various other AMD loci have seen applications for multi-omics integration. These include ARMS2/HTRA1, C3, CFB, and APVOE loci. For example, integrative analyses at the ARMS2/HTRA1 locus demonstrated that non-coding variants in the HTRA1 proximal region influence promoter activity and gene expression in RPE cells [12]. At the C3 locus, colocalization of AMD risk SNPs with C3 expression elucidates contribution to complement amplification. Finally, integrative protein quantitative trait locus (pQTL) studies link AMD-related variants with circulating complement protein levels [13]. Collectively, these studies have shown the potential to extend the CFH workflow to other risk locus and foster an integrative mechanistic model of AMD genetics.

4.2 Technical Challenges in AMD Multi-omics Research

Despite this progress, there are still several hurdles that prevent full integration of multi-omics data. One major risk arises from data heterogeneity. Different omics layers may originate from different cohorts, leading to potential batch effects and population stratification. Analysis time and reproducibility Weights are an additional constraint, as omics experimental resolutions do not scale linearly with resolution. We advocate for the use of a combination of approaches, considering that integrated omics analyses may yield a high-resolution dataset that is beneficial in providing an integrative view of AMD. Multi-omics integration in clinical practice is a promising approach but needs to be translated into practice. Rapidly emerging research includes Cloud-based computational environments, standardized pipelines and harmonization across multiple cohorts, and efforts to expand cohorts for integrative analyses.

4.3 Translational Potential of Multi-omics Integration in Clinical Practice

Integrative omics also shows promise for linking discovery genetics with clinical development. Such integrative approaches can develop multi-omics-informed biomarker

panels that improve AMD risk prediction, warranting earlier screening and more personalized surveys of at-risk couples. For instance, functional annotations re-calibrating polygenic risk scores exhibit better performance to the conventional GWAS based estimates [14]. Additionally, causal gene identification facilitated by integration can lead to relatively new therapeutic intervention targets. In the context of CFH, therapeutic interventions, for example, complement inhibitors (avancincaptad pegol, pegcetacoplan) have been developed and are directly benefited by genetic and functional evidence for mechanisms linking complement dysregulation to AMD pathology [15]. However, achieving integrative frameworks transposing genomics, transcriptomics, and clinical findings may allow precision ophthalmology in which treatment is explicitly guided by molecular risk profiles. Nonetheless, clinical practice implementation of integrated multi-omics findings requires standardization and validation across varying populations and the provision of long-term longitudinal data to establish predictive power.

4.4 Mechanistic Insights and Future Directions from Multi-omics Integration

Regarding mechanistic implications of multi-omics integration, striking changes in our understanding of AMD pathogenesis was achieved revealing interplay mechanisms across multitudes of genes and transcripts instead of isolated gene-gene associations. One example includes convergence of CFH, C3, and CFB multi-omics-based evidence highlighting complement dysregulation at the heart of AMD etiology, whereas analyses of ARMS2/HTRA1 provide causal insight of crosstalk across two targeted pathways: extracellular matrix remodeling versus oxidative stress response pathways, two largely independent investigations to-date. Such an integrative functional perspective explains why interventions of single gene or single pathway specificity achieved limited benefit as AMD is most likely arising from dysregulation of multiple, interconnected biological processes.

An additional important lesson from analytical integration involves tissue and cell-type specificity. Although genome-wide association studies of liver-derived eQTL data relate CFH variants to systemic levels of complement, whereas the examination of epigenomic signals of retinal pigment epithelial cells (RPE) specific changes in CCR1138 and HL-182, p-ADC, and CTR1 loci provides the most mechanistic insight into the retinoscope vascular interplay underlying local retinal pathology. Such findings highlight the benefit to translational research and intervention targeting RPE-specific regulators as a path to prevent cellular senescence or death and to restore local tissue functional integrity in the context of AMD. This therapeutic strategy would be unthinkable via single omics analyses alone.

Potential future directions towards unlocking the untapped potential of multi-omics in studying Alzheimer's disease (AMD) include integrating single-cell multi-omics data, such as scRNA-seq combined with scATAC-seq or spatial proteomics to overcome the current limitations by this sub-field which only analyzes bulk tissues and no single cell. Second, an environmental omics profile data (i.e. metabolomics based on food consumption, and epigenomics based on smoking exposure, respectively) should be included to elucidate gene-environment interactions due to its major contributions to understanding of the highly penetrant features of AMD. Third, machine Learning pipelines which synthesize multi-omics and clinical variables to anticipate disease progression could be developed in order to provide real-world clinical translation and assist clinicians to identify those patients at high risk of disease progression who may benefit from early intervention.

Nonetheless, these advancements must also tackle the lingering challenges in data standardization and reproducibility. Inter-study variability in multi-omics data generation and analytical workflows represent challenging restrictions for making cross-cohort comparisons credible, along with the limited availability of rich, diverse populations in released datasets. Collaborative efforts on global multi-omics consortia studying AMD should become the focus of future initiatives with efforts toward data collection standardization and an open-access repository to facilitate transparency and collaboration. Through untapping these potentials, the area of multi-omics integration to unlock devastating potential promise of advancements in studying AMD and bring us closer toward putting personalized preventive strategies to the forefront of this research field.

5. Conclusion

This review demonstrates how multi-omics integration has advanced our understanding of AMD pathogenesis. By bridging association signals with disease-relevant molecular traits, multi-omics approaches improve the interpretation of genetic liability from genotype to phenotype, and enhance our ability to link environmental exposures, regulatory pathways, and functional consequences across the genome. As AMD is an adult-onset disease influenced by both genetic and environmental factors, recent technological progress in functional genomics, tissue-specific profiling, and computational pipelines has transformed the study of complex traits, despite persistent challenges related to data harmonization, reproducibility, and cell-type

ISSN 2959-409X

heterogeneity.

Importantly, the CFH locus—one of the strongest and most consistent GWAS signals for AMD—illustrates the utility of multi-omics integration. Convergent evidence from expression QTL analyses, epigenomic annotations, and colocalization studies supports the presence of shared causal variants that jointly influence CFH transcriptional regulation and AMD risk. Multi-layer data further highlight tissue-specific regulatory mechanisms, including differential effects in the liver and retinal pigment epithelium, providing deeper insight into complement-mediated pathways central to AMD pathogenesis. This case study exemplifies how multi-omics integration refines causal inference and moves beyond association to mechanistic explanation.

Looking forward, advances in single-cell multi-omics, longitudinal profiling, and spatial transcriptomics will enable increasingly fine-grained dissection of cellular, temporal, and tissue-specific mechanisms. Integrative multi-omics frameworks will not only continue to uncover pathogenic mechanisms in AMD but will also pave the way toward personalized prevention strategies, therapeutic target prioritization, and precision intervention in AMD and related retinal diseases.

References

[1] Wong, W. L., et al. (2021). Global prevalence of age-related macular degeneration. The Lancet Global Health, 9(2), e106–e116.

[2] Ambati, J., & Fowler, B. J. (2012). Mechanisms of age-related macular degeneration. Neuron, 75(1), 26–39.

[3] Fritsche, L. G., et al. (2016). A large meta-analysis of AMD identifies 52 genetic loci. Nature Genetics, 48(2), 134–143.

[4] Fritsche, L. G., et al. (2014). Complement factor H-related protein genes in AMD susceptibility. Human Molecular Genetics, 23(22), 5832–5843.

[5] Gamazon, E. R., & Stranger, B. E. (2015). The impact of human regulatory variation on complex traits. Science, 347(6226), 1518–1521.

[6] Li, X., et al. (2022). Multi-omics approaches for complex disease studies. Trends in Genetics, 38(6), 561–576.

[7] Zipfel, P. F., et al. (2020). Complement factor H and related proteins in AMD. Frontiers in Immunology, 11, 608.

[8] Edwards, A. O., et al. (2005). Complement factor H polymorphism and age-related macular degeneration. Science, 308(5720), 421–424.

[9] Haines, J. L., et al. (2005). Complement factor H variant increases the risk of age-related macular degeneration. Science, 308(5720), 419–421.

[10] Ratnapriya, R., et al. (2019). Retinal transcriptome and eQTL analyses identify genes associated with AMD. Nature Genetics, 51(4), 606–616.

[11] Chu, C., et al. (2021). Integrative analysis of genetic and transcriptomic data identifies causal genes for AMD. Nature Communications, 12, 7031.

[12] Deangelis, M. M., Silveira, A. C., Carr, E. A., & Kim, I. K. (2018). Genetics of age-related macular degeneration: current concepts, future directions. Seminars in Ophthalmology, 33(5), 705–712.

[13] Sun, B. B., et al. (2018). Genomic atlas of the human plasma proteome. Nature, 558(7708), 73–79.

[14] Khawaja, A. P., et al. (2020). Genetic risk prediction for age-related macular degeneration using functional genomic annotations. PLoS Genetics, 16(4), e1008955.

[15] Liao, D. S., et al. (2023). Complement inhibition in AMD: clinical development and prospects. Ophthalmology Retina, 7(3), 202–215.