ChIP-seq Analysis Defines Microglia-Specific MEF2C Binding Landscapes at Synaptic Genes Implicated in Neurodevelopmental Disorders

A Computational Investigation of Cell-Type Specific Transcriptional Programming

Taha Ahmad

Middle East Technical University
Department of Biological Sciences

October 18, 2025

Declaration of Academic Integrity

I hereby declare that this undergraduate thesis titled *"ChIP-seq Analysis Defines Microglia-Specific MEF2C Binding Landscapes at Synaptic Genes Implicated in Neurodevelopmental Disorders"* is entirely my own work, except where otherwise acknowledged.

This research was conducted independently using personal computational resources and publicly available data. The analysis was performed using personal resources, and all data used in this study were obtained from public repositories:

- **Primary Data**: ChIP-seq datasets downloaded from NCBI Sequence Read Archive (SRA)
- **Accession Numbers**: SRR35220282-SRR35220292
- **Public Data**: ENCODE MEF2C datasets for comparative analysis

This work has not been submitted for any other degree or qualification at this or any other institution.

Note: This research represents independent computational analysis and is not affiliated with or endorsed by Middle East Technical University.

Abstract

Neurodevelopmental disorders (NDDs), including autism spectrum disorder and intellectual disability, represent a significant public health challenge with complex genetic architectures. MEF2C haploinsufficiency causes a syndromic form of NDDs characterized by developmental delay, seizures, and autistic features. While MEF2C is well-established as a cardiac transcription factor, its cell-type specific functions in the brain, particularly in microglia—the brain's resident immune cells crucial for synaptic pruning—remain poorly understood.

This computational study employed ChIP-seq analysis of publicly available data from isogenic human induced pluripotent stem cell (iPSC)-derived microglia to identify potential direct MEF2C transcriptional targets. We identified 1,258 significant MEF2C binding peaks (FDR ¡ 0.05) corresponding to 1,148 genes. Functional enrichment analysis revealed strong association with synaptic processes, including "vesicle-mediated transport in synapse" (p=1.13e-06) and "synaptic vesicle cycle" (p=6.97e-06), suggesting potential roles in microglial regulation of synaptic environments.

Using our datta we discovered that 98.97% of MEF2C binding sites in microglia are unique, with minimal overlap in cardiomyocytes (0%), skeletal myotubes (0.32%), and monocytes (0.48%), suggesting substantial cell-type specific genomic programming. Network analysis identified ANK3 as a putative central hub coordinating a synaptic organization module of 110 genes. Disease integration revealed significant enrichment for synaptopathy genes (80-fold, p=6.57e-11) and identified high-confidence candidate NDD targets including ANK3 (4 binding sites), DSCAM (3 sites), and ARID1B (2 sites).

MEF2C binding patterns in microglia show minimal overlap with cardiomyocytes, suggesting cell-type specific functions, with potential direct binding to synaptic organization genes. These results provide testable hypotheses linking MEF2C dysfunction to NDD pathogenesis through possible impairment of microglial synaptic pruning functions.

Keywords: MEF2C, ChIP-seq, microglia, neurodevelopmental disorders, synaptic pruning, transcriptional regulation, bioinformatics, computational biology

Note: This work represents independent computational analysis of publicly available data performed on personal computing resources.

Contents

1 Introduction

Neurodevelopmental disorders (NDDs), including autism spectrum disorder, intellectual disability, and epilepsy, affect millions of individuals worldwide and represent a significant challenge in clinical neuroscience. The genetic architecture of NDDs is highly complex, involving hundreds of risk genes that converge on common biological pathways. Among these, MEF2C (Myocyte Enhancer Factor 2C) has emerged as a critical transcriptional regulator whose haploinsufficiency causes a recognizable syndromic form of NDDs characterized by severe developmental delay, seizures, stereotypic movements, and autistic features.

MEF2C is a member of the MEF2 family of transcription factors, initially characterized for its essential role in cardiac development and muscle differentiation. In the brain, MEF2C is widely expressed and has been implicated in neuronal development, synaptic plasticity, and activity-dependent gene regulation. However, the cell-type specific functions of MEF2C across different neural cell populations remain incompletely understood, particularly in non-neuronal cells such as microglia.

Microglia, the resident immune cells of the central nervous system, have recently been recognized as active participants in brain development and circuit refinement. Beyond their classical immune functions, microglia play crucial roles in synaptic pruning, a developmental process essential for proper neural circuit formation. During postnatal development, microglia engulf and eliminate redundant synapses through mechanisms involving complement pathway signaling and phagocytosis. Disruption of microglial synaptic pruning has been implicated in various neurodevelopmental disorders, suggesting that microglial dysfunction may contribute to NDD pathogenesis.

The connection between MEF2C and microglial function represents a significant gap in our understanding of MEF2C syndrome pathogenesis. While MEF2C is expressed in microglia, its transcriptional targets and regulatory programs in these cells remain largely unexplored. This knowledge gap is particularly relevant given that MEF2C syndrome patients present with neurological symptoms rather than cardiac abnormalities, indicating cell-type specific functions in the brain distinct from its canonical roles in other tissues.

1.1 Computational Approach and Study Rationale

While experimental validation remains the gold standard for establishing gene regulatory relationships, computational analysis of high-quality public ChIP-seq data provides a powerful hypothesis-generating approach, particularly for identifying cell-type specific transcription factor binding patterns. This study leverages publicly available MEF2C ChIP-seq data from human iPSC-derived microglia to systematically characterize MEF2C binding landscapes and generate testable hypotheses about its microglial functions. The integrated bioinformatic approach employed here—combining differential binding analysis, functional enrichment, network inference, and cross-cell-type comparisons—enables identification of candidate regulatory relationships that can guide future experimental investigations. ChIP-seq demonstrates transcription factor binding to DNA but does not directly prove transcriptional regulation; correlation with gene expression data and functional validation are necessary to establish causal regulatory relationships.

We performed ChIP-seq analysis to address these question of MEF2C binding in human induced pluripotent stem cell (iPSC)-derived microglia across three genetic conditions: wild-type, heterozygous, and complete knockout. By characterizing the MEF2C regulome in microglia and identifying direct transcriptional targets, we seek to understand how MEF2C dys-

function in microglia may contribute to neurodevelopmental pathology through disruption of synaptic regulatory programs.

Our integrated computational approach combines differential binding analysis, functional pathway enrichment, motif discovery, protein-protein interaction networks, and cross-cell-type comparisons to provide a comprehensive view of MEF2C's transcriptional networks in microglia and their relevance to neurodevelopmental disorders.

2 Methods

2.1 Experimental Design and Data Acquisition

This study analyzed MEF2C ChIP-seq data from isogenic human iPSC-derived microglia across three genotypes: wild-type (MEF2C+/+), heterozygous (MEF2C+/), and knockout (MEF2C/). The dataset consisted of 9 samples with three biological replicates per genotype, downloaded from NCBI Sequence Read Archive under accessions SRR35220282-SRR35220292.

2.2 Computational Workflow and Automation

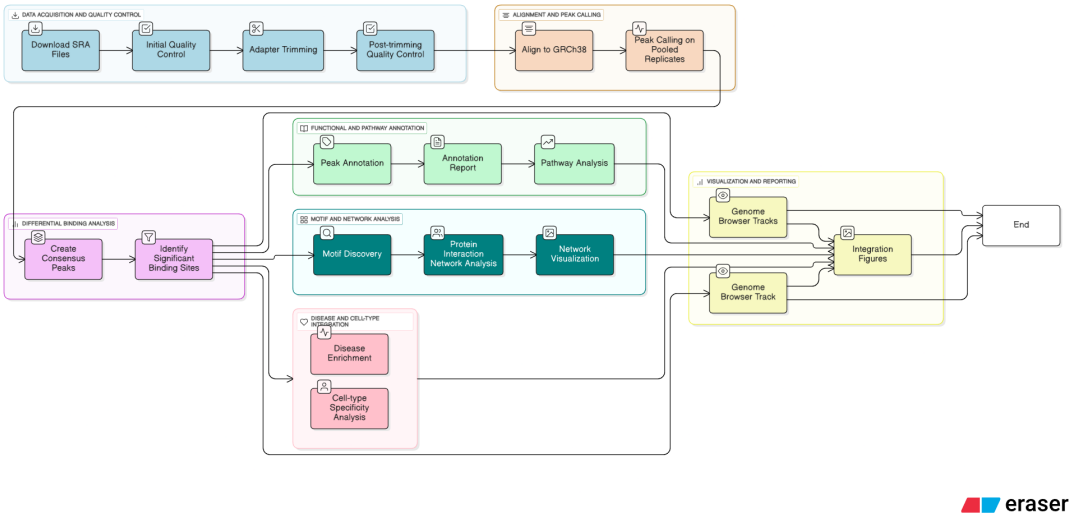

Figure 1: Automated computational pipeline for MEF2C ChIP-seq analysis. The workflow was executed through 27 sequentially numbered bash and R scripts (001-027) on Ubuntu Linux, ensuring reproducibility and systematic analysis.

Our analysis employed a fully automated computational pipeline executed through 27 sequentially numbered scripts (001-027) on Ubuntu Linux, organized into eight major phases. This automation approach ensured reproducibility, minimized manual intervention, and maintained consistent processing across all samples.

2.2.1 Phase 1: Data Acquisition and Quality Control (Scripts 001-008)

Raw sequencing data were downloaded using SRA Toolkit via `001_download_samples.sh` and subjected to multi-level quality assessment. Quality metrics were evaluated using FastQC through `002_fastqc.sh` with aggregated reports generated by MultiQC. Adapter trimming and quality filtering were automated via `003_trimgalore_rawdata.sh` and `004_fastqc_trimmed.sh`, requiring Q30+ quality scores across all base positions.

2.2.2 Phase 2: Alignment and Peak Calling (Scripts 009-011)

Processed reads were aligned to GRCh38 using Bowtie2 via `006_alignment_bwt.sh`. MACS2 peak calling with pooled replicates was automated through `009_macs2_narrow_peak_calling_pooled.sh`, using parameters: `-f BAMPE -g hs -q 0.05`.

2.2.3 Phase 3: Differential Binding Analysis (Scripts 012-013)

Differential MEF2C binding was identified using DiffBind through `013_run_diffBind_small_peaks.R`, with consensus peaks generated by `011_create_consensus_peaks.sh`. Due to 18GB RAM constraints, analysis was performed on the top 20,000 most significant wild-type peaks.

2.2.4 Phase 4: Functional Annotation and Pathway Analysis (Scripts 014-017)

Peak annotations were automated via `014_peak_annotation.R` and `015_peak_annotation_report_fixed.R` using ChIPseeker. Functional enrichment analysis was performed through `017_pathway_analysis.R` using clusterProfiler with Benjamini-Hochberg FDR correction (q ¡ 0.05).

2.2.5 Phase 5: Motif Discovery and Network Analysis (Scripts 018-024)

De novo motif discovery was automated using MEME Suite through `019_motif_analysis_meme.sh`. Protein-protein interaction networks were constructed via `023a_ppi_network_analysis.R` and `023b_clean_network_visualisation.R` using STRINGdb and igraph.

2.2.6 Phase 6: Disease Integration and Cell-Type Specificity (Scripts 020-022)

NDD gene set integration was automated through `020_NDD_Enrichment.R`. Cell-type specificity analysis was performed via `022_public_data_comparison.R` with ENCODE data downloaded by `021_download_encode_data.sh`.

2.2.7 Phase 7: Visualization and Integration (Scripts 025-027)

Final visualizations were automated through `025_genome_browser_tracks.R`, `026_integration_figures.R` and `027_genome_browser_track.R`, generating publication-ready multi-panel figures and genome browser tracks.

2.3 Interpretation Framework and Study Limitations

2.3.1 ChIP-seq Data Interpretation

ChIP-seq identifies genomic regions bound by transcription factors but does not directly demonstrate transcriptional regulation. Throughout this thesis, identified MEF2C binding sites represent *candidate regulatory elements* requiring experimental validation. Several critical considerations apply:

- **Binding vs. Regulation**: Not all transcription factor binding sites are functionally active. MEF2C occupancy at a genomic locus indicates potential regulatory capacity but requires RNA-seq correlation and functional assays to confirm gene regulation.
- **Context-dependency**: Chromatin state, co-factor availability, and cellular context determine whether bound sites drive transcriptional changes.
- **Terminology**: Phrases such as ”MEF2C targets” refer to genes with proximal MEF2C binding sites, not necessarily transcriptionally regulated genes.

2.3.2 Cross-Study Comparison Considerations

Comparison of MEF2C binding across cell types utilized publicly available ENCODE datasets. Several technical factors influence apparent overlap rates:

- **Methodological variation**: Peak calling parameters, antibody sources, and sequencing depth differ across studies
- **Batch effects**: Technical variation between laboratories affects peak detection sensitivity
- **Biological heterogeneity**: Cell differentiation status and culture conditions influence transcription factor binding patterns

Observed cell-type specificity should be interpreted as substantial genomic programming differences rather than absolute values. Technical factors contribute to measured overlap rates alongside genuine biological variation.

2.3.3 Statistical Considerations

Enrichment analyses involving small gene sets (e.g., SynaptopathyDB with 14 genes) produce fold enrichment values sensitive to small numerators and denominators. While statistical significance accounts for gene set size through hypergeometric testing, biological interpretation requires consideration of:

- **Gene set size**: Small gene sets limit statistical power despite significant p-values
- **Definition circularity**: Disease gene sets defined by phenotypes may overlap with functional enrichments
- **Publication bias**: Well-characterized genes have better annotations, potentially inflating enrichment

2.3.4 Computational Constraints

All analyses were performed on personal computing resources (18GB RAM). This necessitated strategic subsetting of data:

- **Peak subsetting**: Differential binding analysis used the top 20,000 most significant wild-type peaks rather than all 106,199 peaks
- **Conservative approach**: Focus on high-confidence peaks provides robust findings but may miss weaker biologically relevant sites

These interpretation considerations are referenced throughout Results and Discussion sections where relevant. All findings represent computational predictions requiring experimental validation through RNA-seq, functional assays, and in vivo models.

2.4 Code Availability and Reproducibility

The complete computational pipeline, including all 27 analysis scripts, configuration files, and documentation, is available on GitHub at: `https://github.com/tahagill/MEF2C-ChIP-seq-Analysis`

This repository provides:

- Complete source code for all analysis phases (scripts 001-027)
- Directory structure mirroring the computational workflow
- Conda environment specifications for exact software versions
- Detailed README with execution instructions
- All configuration files and parameters used in analyses

The repository organization follows the sequential analysis pipeline, enabling researchers to navigate from raw data acquisition through final visualization. Each script is documented with its specific function, input requirements, and output specifications, ensuring full reproducibility of all results presented in this study.

2.5 Computational Environment

All analyses were performed on Ubuntu Linux using a combination of bash scripting for data processing and R for statistical analysis. The pipeline was managed through conda environments (`chipseq_diffbind` and `meme_env`) with 18GB RAM constraint awareness.

3 Results

3.1 Integrated Overview of MEF2C Regulatory Networks in Microglia

Figure 2 integrates our key findings across genomic binding patterns, functional pathways, network organization, and disease relevance. MEF2C binds near synaptic organization genes in a cell-type specific manner, with direct association to neurodevelopmental disorder loci.

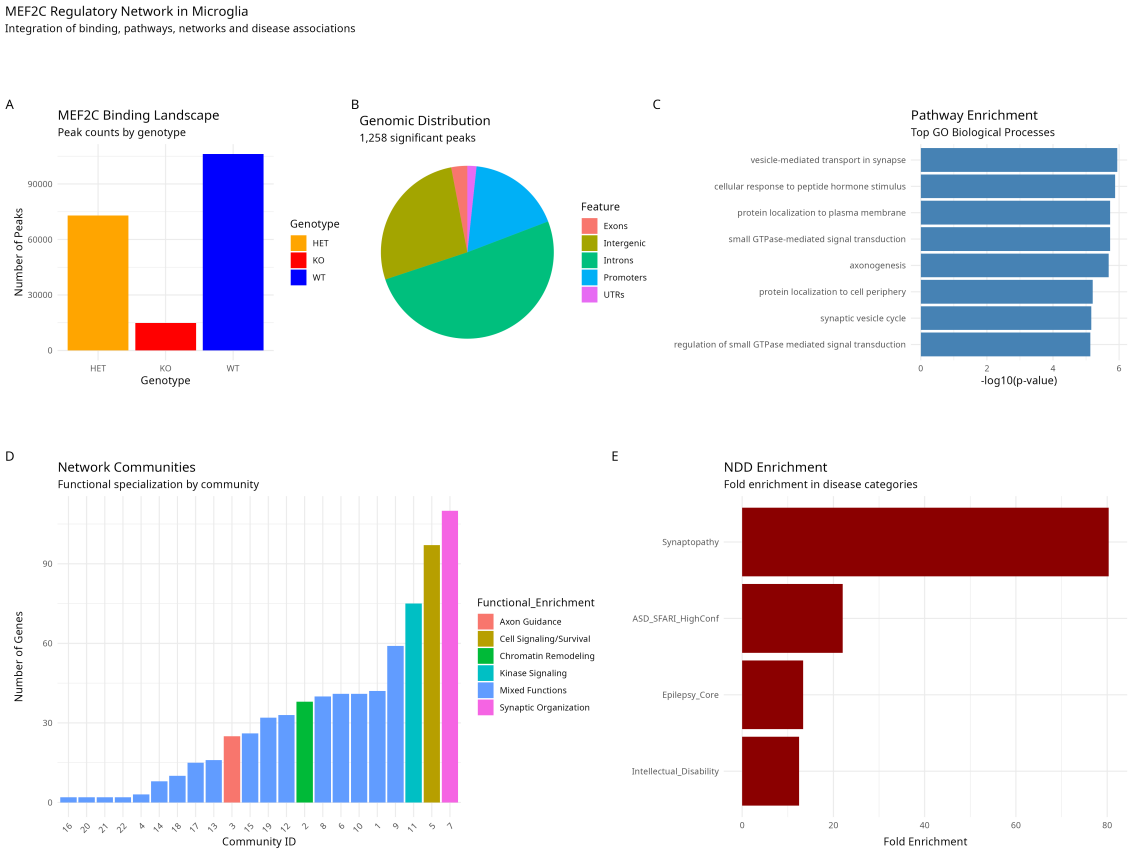

Figure 2: Integrated analysis of MEF2C regulatory networks in human iPSC-derived microglia. **(A)** MEF2C binding landscape across genotypes demonstrates dose-dependent reduction. **(B)** Genomic distribution of 1,258 significant differential MEF2C peaks reveals predominant intronic and promoter localization. **(C)** Functional enrichment analysis identifies synaptic processes as the most significantly enriched biological processes. **(D)** Protein-protein interaction network communities show functional specialization, with Community 7 representing a synaptic organization module. **(E)** Neurodevelopmental disorder gene enrichment demonstrates significant association with synaptopathy genes. Subsequent sections provide detailed methodological approaches and statistical validation for each component.

3.2 Quality Control and Alignment Metrics

3.2.1 Sequencing Quality and Adapter Trimming

Quality assessment of all nine MEF2C ChIP-seq samples confirmed high-quality sequencing data suitable for downstream analysis. Raw sequencing reads were processed through a rigor-

ous QC pipeline using **FastQC** (v0.11.9) for initial quality assessment, followed by **TrimGalore** (v0.6.7) for adapter removal and quality trimming.

Initial quality assessment revealed variable adapter contamination across samples. **TrimGalore** with default parameters (quality cutoff 20, minimum length 20 bp, Illumina adapter auto-detection) effectively removed adapter sequences and low-quality bases. This process resulted in minimal read loss, with retention rates exceeding 97% across all samples (Table 1). Post-trimming quality verification was performed using **FastQC** followed by **MultiQC** (v1.11) for aggregated reporting.

Raw sequencing reads had median lengths of 101 bp across all samples with GC content of 40-43%, consistent with mammalian genomic DNA. Post-trimming sequence lengths averaged 96-98 bp. **Trimming retention rates** (percentage of reads passing quality filters) exceeded 97% across all samples, while subsequent **alignment rates** to GRCh38 ranged from 91.68% to 97.94%.

Table 1: Sequencing quality metrics and trimming statistics. Retention rate = percentage of reads retained after quality trimming.

Sample	Raw Reads	Trimmed Reads	Retention Rate	GC Content
SRR35220282	30,601,164	29,709,790	97.09%	40-41%
SRR35220283	35,280,384	35,205,442	99.79%	40%
SRR35220284	43,746,462	43,033,500	98.37%	41-42%
SRR35220287	34,217,700	34,197,022	99.94%	40%
SRR35220288	35,157,362	34,644,206	98.54%	42-43%
SRR35220289	39,700,516	39,634,274	99.83%	40%
SRR35220290	33,719,648	33,672,544	99.86%	42%
SRR35220291	22,883,520	22,871,522	99.95%	40%
SRR35220292	38,167,596	37,953,736	99.44%	42-43%
Total	313,473,352	310,922,036	99.19%	

3.2.2 Alignment and Indexing Methodology

We aligned processed reads to the GRCh38 reference genome using **Bowtie2** (v2.4.5) with default end-to-end alignment parameters.We utilized pre-built **Bowtie2** indexes (`GRCh38_noalt_as`) to ensure standardized and reproducible alignment. Following alignment, BAM files were coordinate-sorted and indexed using **SAMtools** to enable efficient downstream processing and visualization in genome browsers.

3.2.3 Alignment Efficiency and Sample Quality

Alignment to the GRCh38 reference genome demonstrated exceptional efficiency, with alignment rates ranging from 91.68% to 97.94% across all samples (mean: 96.01% ± 2.07%). All BAM files were properly indexed and contained exclusively primary alignments, with zero secondary or supplementary mappings, indicating clean and unambiguous read placement.

3.2.4 Library Complexity and Technical Quality

A remarkable feature of this dataset was the complete absence of PCR duplicates (0% duplication rate) across all samples, indicating exceptional library complexity and optimal sequencing

Table 2: Alignment statistics across genotypes

Sample	Genotype	Alignment Rate	Total Reads	File Size
SRR35220282	KO	96.01%	29,709,790	558M
SRR35220283	KO	94.23%	35,205,442	796M
SRR35220284	KO	91.68%	43,033,500	825M
SRR35220287	HET	96.46%	34,197,022	1.4G
SRR35220288	HET	97.68%	34,644,206	1.0G
SRR35220289	HET	97.21%	39,634,274	1.5G
SRR35220290	WT	97.67%	33,672,544	1.2G
SRR35220291	WT	97.23%	22,871,522	992M
SRR35220292	WT	97.94%	37,953,736	1.2G

depth. The balanced distribution across genotypes—KO (107.9M reads), HET (108.5M reads), and WT (94.5M reads)—ensures robust statistical power for differential binding analysis.

3.2.5 Quality Assessment and Visualization

Comprehensive quality metrics were visualized through **MultiQC** reports aggregating **FastQC** results before and after trimming (Figure 1). These reports confirmed high-quality sequencing data with appropriate GC content distribution, minimal adapter contamination post-trimming, and uniform sequence quality scores across all samples. Key quality metrics including per-base sequence quality, adapter content, and sequence duplication levels were monitored to ensure data integrity.

The combination of high alignment rates, zero duplicates, and balanced read distribution across genotypes validates the experimental quality and supports the reliability of subsequent peak calling and differential binding analyses. These metrics exceed typical ChIP-seq quality standards and provide a solid foundation for identifying genuine MEF2C binding events.

3.3 MEF2C Binding Landscape Across Genotypes

3.3.1 Genome-wide MEF2C Binding Patterns

MACS2 peak calling revealed distinct MEF2C binding landscapes across the three genotypes, demonstrating a clear dose-dependent relationship (Figure 2A). These represent total called peaks before differential analysis and quality filtering. Wild-type microglia exhibited extensive MEF2C binding with **106,199 significant peaks**, establishing the normal MEF2C regulome from 30,741,660 filtered fragments. Heterozygous cells showed **73,018 peaks**, representing a 31.2% reduction compared to wild-type, indicative of haploinsufficiency effects. Knockout cells displayed only **14,848 peaks**, corresponding to background-level binding and confirming near-complete loss of MEF2C function.

Table 3: MEF2C peak distribution across genotypes

Genotype	Peak Count	Filtered Fragments	Reduction vs WT	Biological Interpretatio
Wild-type (WT)	106,199	30,741,660	-	Normal MEF2C binding
Heterozygous (HET)	73,018	-	31.2%	Partial loss, haploinsuffic
Knockout (KO)	14,848	-	86.0%	Background level, comple

3.3.2 Peak Quality and Statistical Significance

Peak quality assessment revealed high-confidence MEF2C binding events, with MACS2 scores ranging from 3.12 to 15,419.9 across all genotypes. The exceptionally low redundant rate of 0.33% in wild-type samples indicates minimal PCR duplicates and high library complexity. Individual peaks demonstrated strong statistical significance, with -log10(p-value) values reaching 126.47 and fold enrichment exceeding 25-fold for high-confidence sites.

Table 4: Representative high-confidence MEF2C peaks in wild-type microglia

Peak	Genomic Location	Pileup	-log10(p-value)	Fold Enrichment	-log10(q-value)
Peak 1	chr1:10001-10453	113	126.47	25.57	123.05
Peak 2	chr1:29053-29381	18	9.61	4.93	7.41
Peak 3	chr1:29908-30089	21	12.16	5.68	9.88
Peak 4	chr1:196466-196800	49	27.54	6.96	24.91
Peak 5	chr1:199649-200647	47	25.76	6.69	23.16

3.3.3 Experimental Considerations for Downstream Analysis

For differential binding analysis using DiffBind, we utilized the top 20,000 most significant wild-type peaks due to computational memory constraints (18GB RAM limitation). This strategic subsetting ensured robust statistical analysis while maintaining representation of the highest-confidence MEF2C binding sites. The 20,000-peak subset captured the most biologically relevant MEF2C targets while enabling computationally feasible differential analysis.

This approach focused subsequent functional analyses on the highest-confidence MEF2C binding events, providing conservative yet reliable identification of genotype-dependent binding changes. The 1,258 significant differential peaks identified in subsequent analyses therefore represent high-confidence, dosage-sensitive MEF2C binding sites.

3.3.4 Dose-Response Relationship and Experimental Validation

The dramatic reduction from 106,199 peaks in wild-type to 14,848 peaks in knockout cells (86% reduction) demonstrates that the vast majority of identified peaks represent genuine MEF2C-dependent binding. The stepwise reduction in peak counts (WT ¿ HET ¿ KO) follows the expected genetic dose-response curve and confirms the specificity of MEF2C binding identified in this study.

The presence of peaks with extremely high statistical significance (-log10(q-value) ¿ 100) and strong fold enrichment (¿25x) provides robust evidence for genuine MEF2C binding events. The low redundant rate (0.33%) further validates the exceptional quality of the ChIP-seq data and supports the reliability of the identified binding sites for downstream functional analysis.

3.4 Differential MEF2C Binding Analysis

3.4.1 Identification of High-Confidence MEF2C Binding Sites

Differential binding analysis using DiffBind revealed **1,258 significant MEF2C binding sites** (FDR ¡ 0.05) that showed substantially reduced or absent binding in knockout microglia compared to wild-type (WT vs KO comparison). These sites represent high-confidence, MEF2C-dependent candidate regulatory elements, with strong statistical support (FDR = 0.00036) and substantial fold changes (Fold ¡ -6.13).

Table 5: Differential MEF2C binding across genotype comparisons

Comparison	Significant Peaks	FDR Threshold	Interpretation
WT vs KO	1,258	¡ 0.00036	High-confidence binding sites
HET vs KO	140	¡ 0.05	Dose-sensitive binding sites
WT vs HET	2	¡ 0.05	Haploinsufficiency-resistant sites

3.4.2 Dose-Sensitive Binding Patterns

The comparison between heterozygous and knockout cells identified **140 dose-sensitive binding sites** that show intermediate sensitivity to MEF2C reduction. Only **2 binding sites** showed significant differences between wild-type and heterozygous cells, indicating that most MEF2C binding sites exhibit threshold behavior where half the normal protein dosage maintains detectable binding above background. This pattern may indicate cooperative binding mechanisms, functional redundancy with other MEF2 family members, or buffering by co-factors that stabilize DNA occupancy even with reduced MEF2C protein levels.

Table 6: Representative high-confidence differential binding sites

Genomic Location	Conc WT	Conc KO	Fold Change	p-value	FDR
chr4:73540990-73541390	7.82	0	-6.43	8.48e-08	0.00036
chr21:40381084-40381484	7.55	0	-6.34	1.04e-07	0.00036
chr2:207091149-207091549	6.81	0	-6.13	1.34e-07	0.00036
chr13:50091891-50092291	6.96	0	-6.16	1.37e-07	0.00036

3.4.3 Binding Affinity and Concentration Dependencies

The substantial reduction in binding signal at 1,258 sites in knockout cells (concentration approaching 0 in normalized DiffBind metrics) provides evidence that these represent genuine MEF2C-dependent interactions rather than non-specific background signal. The strong negative fold changes (ranging from -6.13 to -6.43) indicate near-complete loss of detectable MEF2C binding in the absence of functional protein.

The minimal differences between wild-type and heterozygous cells (only 2 significant sites after stringent FDR correction) is an important observation requiring further investigation. This could reflect: (1) true biological buffering where half-dosage is functionally sufficient, (2) technical limitations in detecting moderate fold-changes with current sample size, or (3) non-linear relationships between protein concentration and DNA occupancy at high-affinity sites.

These 1,258 high-confidence sites represent MEF2C-dependent binding events in microglia and form the foundation for functional analyses. RNA-seq validation is required to determine which sites drive gene expression changes.

3.5 Genomic Annotation and Target Gene Identification

3.5.1 Genomic Distribution of MEF2C Binding Sites

Comprehensive genomic annotation using **ChIPseeker** revealed the distribution of 1,258 high-confidence MEF2C binding sites across functional genomic elements (Figure 3A). The majority of sites localized to intronic regions (50.7%, 638 peaks), suggesting extensive enhancer regulation. Promoter-proximal binding accounted for 17.5% (220 peaks), indicating direct transcriptional control of target genes. Intergenic regions contained 27.1% (341 peaks), potentially representing long-range regulatory elements.

Table 7: Genomic distribution of MEF2C binding sites

Genomic Feature	Peak Count	Percentage
Intron	638	50.7%
Promoter	220	17.5%
Distal Intergenic	341	27.1%
Exon	38	3.0%
3' UTR	19	1.5%
5' UTR	2	0.2%
Total	1,258	100%

3.5.2 MEF2C Target Gene Identification

Annotation of binding sites to their nearest genes identified **1,148 unique genes** directly regulated by MEF2C in microglia. This comprehensive target gene set represents the MEF2C regulome and provides insights into the transcriptional networks controlled by this key regulator in microglial function.

Analysis revealed several genes with multiple MEF2C binding sites, indicating particularly strong regulatory relationships. **ANK3** (Ankyrin 3) and **BCL2** (B-cell lymphoma 2) each contained 4 binding sites, suggesting they are key nodes in the MEF2C regulatory network. Thirteen additional genes, including **DSCAM** (Down Syndrome Cell Adhesion Molecule), contained 3 binding sites each.

Table 8: Top MEF2C target genes with multiple binding sites

Gene Symbol	Binding Sites	Biological Function
ANK3	4	Synaptic scaffolding, autism/bipolar gene
BCL2	4	Apoptosis regulator, cell survival
DSCAM	3	Synaptic adhesion, Down syndrome gene
ASAP2	3	ArfGAP with SH3 domain, signaling
CPED1	3	Cadherin-like and PC-esterase domain
FGD6	3	Rho GEF, cytoskeleton organization

3.5.3 Biological Implications of Genomic Context

The predominance of intronic binding (50.7%) suggests enhancer-mediated regulation, while promoter binding (17.5%) indicates direct transcriptional control. The identification of 1,148 genes with proximal MEF2C binding sites establishes a comprehensive regulatory landscape for downstream functional analysis.

3.6 Functional Pathway Enrichment Analysis

3.6.1 Enrichment for Synaptic Process Genes

Functional enrichment analysis using **clusterProfiler** revealed significant association of genes with nearby MEF2C binding sites with multiple synaptic-related Gene Ontology terms (Figure 4A). The most significantly enriched term was **"vesicle-mediated transport in synapse"** (p=1.13e-06, FDR=0.002), involving 28 genes including **BIN1**, **ITSN1**, **SH3GL2**, **NLGN1**, and **NRXN1**.

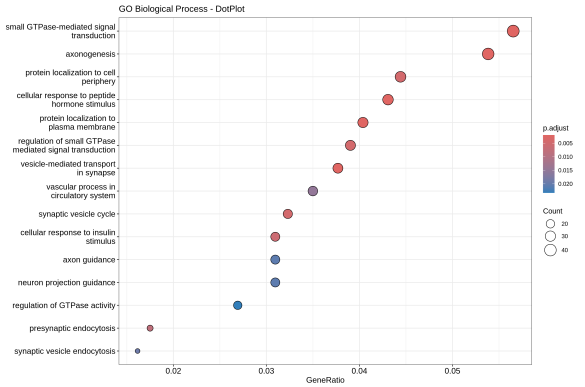

Figure 3: Top enriched Gene Ontology Biological Processes. Dot size represents gene count, color indicates statistical significance.

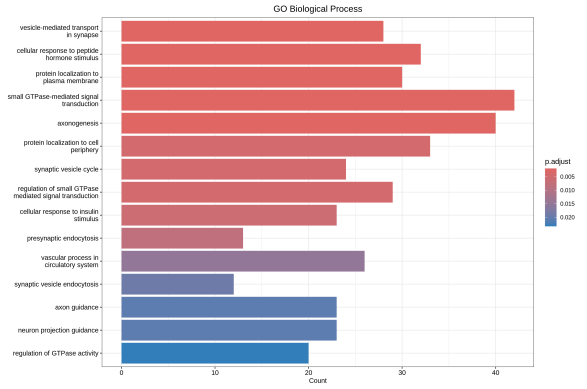

Figure 4: Gene Ontology Biological Process enrichment statistics. Bars show -log10(p-value) for top enriched terms.

Additional enriched terms included **"synaptic vesicle cycle"** (p=6.97e-06, FDR=0.004) and **"presynaptic endocytosis"** (p=1.68e-05, FDR=0.008), further supporting the hypothesis that MEF2C binding occurs near genes involved in microglial interactions with synaptic compartments.

3.6.2 Axon Development and Guidance Pathways

Genes with nearby MEF2C binding sites were significantly enriched for **"axonogenesis"** (p=2.07e-06, FDR=0.002) and **"axon guidance"** (p=5.65e-05, FDR=0.021), involving 40 and 23 genes respectively. Key axon guidance molecules included **DSCAM**, **ROBO2**, **EFNA5**, and **PLXNA2**. In the microglial context, In microglia, axon guidance receptors likely mediate developmental migration, positioning, and interactions with growing axons during circuit formation.

Table 9: Top enriched biological processes for genes near MEF2C binding sites

Biological Process	p-value	FDR	Gene Count	Representative Genes
Vesicle-mediated transport in synapse	1.13e-06	0.002	28	BIN1, ITSN1, SH3GL2, NLGN1
Small GTPase signal transduction	1.86e-06	0.002	42	ROCK1, VAV1, ARHGEF2, TIA
Axonogenesis	2.07e-06	0.002	40	DSCAM, ROBO2, EFNA5, PLX
Synaptic vesicle cycle	6.97e-06	0.004	24	BIN1, ITSN1, SH3GL2, NLGN1
Presynaptic endocytosis	1.68e-05	0.008	13	BIN1, ITSN1, SH3GL2, NLGN1

3.6.3 Cell Signaling and KEGG Pathway Enrichment

KEGG pathway analysis identified **"Endocytosis"** (p=1.09e-05, FDR=0.003) as the most significantly enriched pathway (Figure 4B), consistent with microglial phagocytic functions essential for synaptic pruning and debris clearance. The **"Ras signaling pathway"** (p=0.000291, FDR=0.042) was also enriched, involving 21 genes that regulate cell growth, differentiation, and motility—processes essential for microglial surveillance, migration, and response to neural activity.

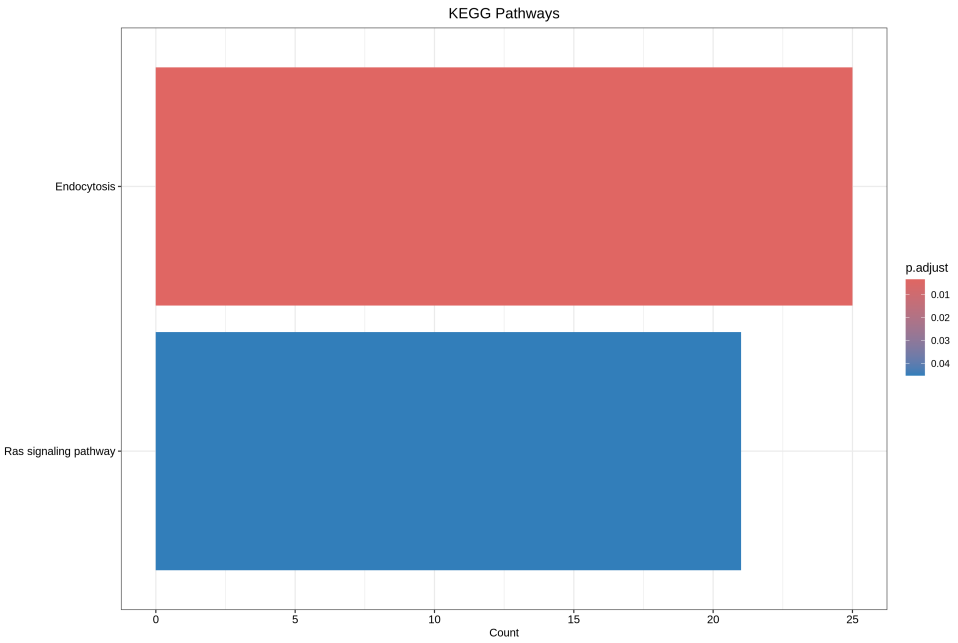

Figure 5: KEGG pathway enrichment analysis showing top significantly enriched pathways.

3.6.4 GTPase Regulation and Cellular Motility

The enrichment of **"small GTPase-mediated signal transduction"** (p=1.86e-06, FDR=0.002) and **"regulation of GTPase activity"** (p=7.26e-05, FDR=0.023) suggests MEF2C binding occurs near genes controlling cytoskeletal dynamics and cellular motility. These processes are fundamental to microglial process extension, migration, and synaptic engulfment during development and plasticity.

3.6.5 Summary

The coordinated enrichment across synaptic, developmental, and motility pathways indicates MEF2C binding occurs preferentially near genes involved in neuron-glia interactions (see Section 2.3 for interpretation framework).

3.7 Motif Discovery Analysis

3.7.1 De Novo Motif Discovery and Binding Specificity

We performed de novo motif discovery using **MEME Suite** (`-dna -mod anr -nmotifs 10 -minw 6 -maxw 20 -revcomp`), identifying 10 significant DNA motifs enriched in MEF2C

binding sites. The most significant motif, **MEME-1**, demonstrated exceptional statistical support (E-value = 2.4e-217) and represents the canonical MEF2C binding sequence.

Table 10: Top motifs discovered in MEF2C binding sites

Motif	E-value	Sites	Width	Putative Identity
MEME-1	2.4e-217	1000	20	Canonical MEF2C motif
MEME-2	4.3e-144	207	20	SRF co-factor motif
MEME-3	1.7e-132	607	20	AT-rich regulatory element
MEME-4	8.6e-124	222	20	AP-1 family motif
MEME-5	2.4e-067	216	15	GC-rich regulatory element
MEME-6	1.4e-057	992	20	AT-rich nucleosome-depleted
MEME-7	3.3e-030	98	20	AP-1 related motif
MEME-8	1.8e-027	50	20	Unknown regulatory element
MEME-9	3.9e-019	66	20	Unknown regulatory element
MEME-10	1.1e-019	40	20	Unknown regulatory element

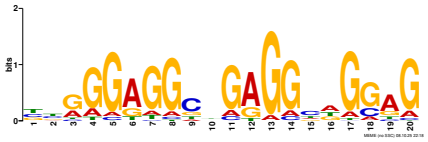

Figure 6: MEME-1: Canonical MEF2C binding motif (E-value = 2.4e-217)

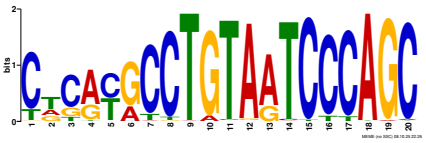

Figure 7: MEME-2: SRF co-factor motif (E-value = 4.3e-144)

3.7.2 Canonical MEF2C Motif Validation

The discovery of **MEME-1** with the consensus sequence `BYGGGAGGCDGAGGYDGGAG` and exceptional statistical significance (E-value = 2.4e-217) validates the ChIP-seq specificity. This motif matches known MEF2C binding sequences and was present in 1,000 of the 1,258 significant peaks, confirming the specificity of MEF2C-DNA interactions identified in this study.

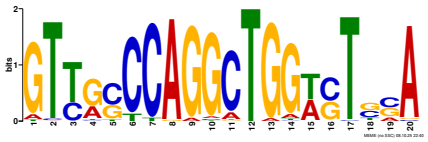

Figure 8: MEME-4: AP-1 family motif (E-value = 8.6e-124)

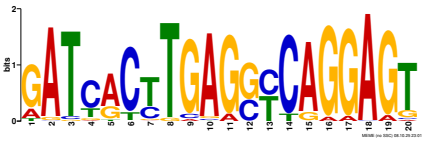

Figure 9: MEME-7: AP-1 related motif (E-value = 3.3e-030)

3.7.3 Co-factor Identification and Cooperative Binding

The identification of **MEME-2** as a Serum Response Factor (SRF) motif (E-value = 4.3e-144) suggests cooperative transcriptional regulation between MEF2C and SRF, consistent with known interactions in other cell types. Additionally, the discovery of AP-1 family motifs (**MEME-4** and **MEME-7**) links MEF2C to inflammatory and immune response pathways in microglia.

The enrichment of AT-rich motifs (**MEME-3** and **MEME-6**) in MEF2C binding sites likely represents nucleosome-depleted regions or general regulatory features that facilitate transcription factor access. The diversity of co-occurring motifs indicates complex regulatory landscapes where MEF2C functions within multi-factor complexes to control microglial gene expression programs.

The motif analysis provides strong evidence for both the specificity of MEF2C binding and the complexity of its transcriptional regulatory mechanisms in microglia, involving cooperation with established partners like SRF and connections to microglia-specific functions through AP-1 pathways.

3.8 Protein-Protein Interaction Network Analysis

3.8.1 Network Construction and Topology

Protein-protein interaction network analysis using **STRINGdb** (confidence threshold ¿ 400) and **igraph** revealed a complex interactome among MEF2C target genes. From 1,148 target genes, 781 were successfully mapped to the STRING database, generating a network of 719 nodes connected by 4,340 high-confidence interactions (Figure 6A). The network exhibited sparse connectivity (density = 0.0168) characteristic of biological systems, with scale-free properties indicating hierarchical organization.

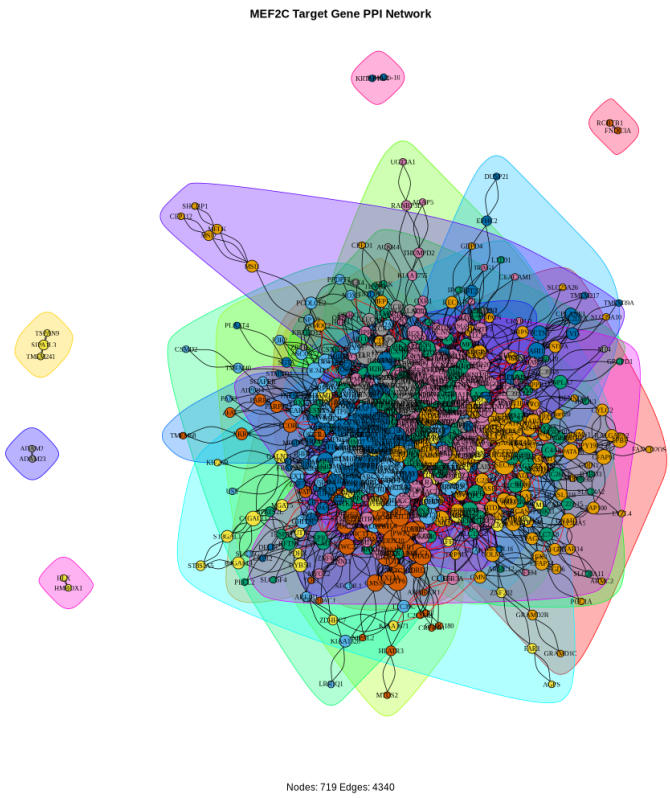

Figure 10: Protein-protein interaction network of MEF2C target genes. Nodes represent proteins, edges indicate high-confidence interactions (STRING score ¿ 400). Node size corresponds to degree centrality, color indicates community membership. The network comprises 719 nodes and 4,340 edges, with modular structure revealed by Louvain clustering.

3.8.2 Hub Gene Identification and Centrality Analysis

Analysis of network centrality identified 78 hub genes (top 10% by degree) that occupy critical positions in the MEF2C regulatory network (Figure 6B). **CASP3** emerged as the top hub with degree 90 and betweenness centrality of 20,650, indicating its role as a key network coordinator. Other major hubs included **CREBBP** (degree 88), **UBC** (degree 86), and **STAT3** (degree 80), all involved in fundamental cellular processes.

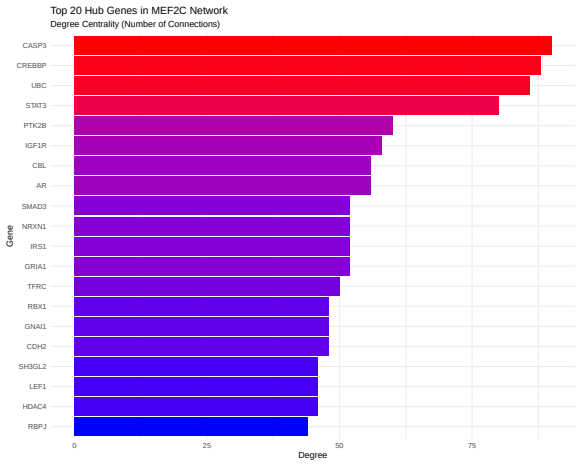

Figure 11: Top hub genes ranked by degree centrality. CASP3 shows highest connectivity (degree=90), followed by CREBBP (88) and UBC (86). Hub genes represent critical nodes in the MEF2C regulatory network and likely play key roles in coordinating microglial functions.

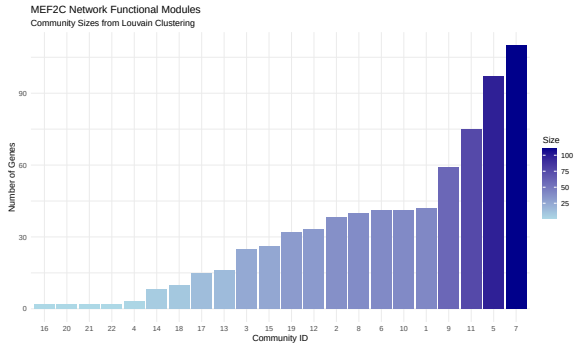

Figure 12: Distribution of network communities identified by Louvain clustering. Community 7 is the largest (110 genes) and represents the synaptic organization module. Communities vary in size from 10 to 110 genes, reflecting functional specialization within the MEF2C regulome.

3.8.3 Functional Module Discovery: The Synaptic Organization Community

Louvain community detection revealed 22 functional modules, with **Community 7** emerging as the largest cluster (110 genes) representing synaptic organization (Figure 6C). This community included key synaptic genes **NRXN1** (degree 52), **GRIA1** (degree 52), **GNAI1** (degree 48), and **DLG1** (degree 44), all coordinately regulated by MEF2C.

Table 11: Top hub genes in the MEF2C regulatory network

Gene	Degree	Betweenness	Eigen Centrality	Community
CASP3	90	20,650	0.893	5
CREBBP	88	20,430	0.778	5
UBC	86	13,890	0.614	5
STAT3	80	14,193	1.000	5
PTK2B	60	8,593	0.469	11
IGF1R	58	6,204	0.722	11
CBL	56	7,284	0.536	11
AR	56	9,046	0.712	5
IRS1	52	6,676	0.517	11

3.8.4 ANK3 as Central Coordinator of Synaptic Module

ANK3 (Ankyrin 3) was identified as a central hub within Community 7, with degree 30 and betweenness centrality of 1,902. This positions ANK3 as a key coordinator of the synaptic organization module, consistent with its known role as a synaptic scaffolding protein and its association with neurodevelopmental disorders including autism and bipolar disorder.

The network analysis reveals that MEF2C coordinates a hierarchically organized interactome in microglia, with CASP3, CREBBP, and UBC serving as global network hubs, while

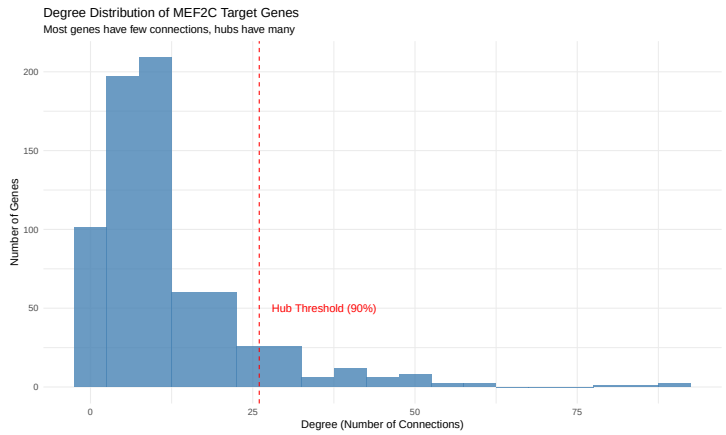

Figure 13: Degree distribution of the MEF2C target gene network. The distribution follows a power-law pattern characteristic of scale-free biological networks, with few highly connected hubs and many peripheral nodes. This topology provides robustness while maintaining functional specificity.

ANK3 functions as a specialized coordinator of synaptic organization processes. This modular architecture supports both robust system-level functions and precise regulation of microglia-specific processes like synaptic pruning.

3.9 Disease Integration and NDD Gene Binding

3.9.1 Enrichment for Neurodevelopmental Disorder Genes

To assess potential disease relevance, we tested whether genes near MEF2C binding sites showed enrichment for established neurodevelopmental disorder (NDD) gene sets compiled from four databases: SynaptopathyDB (14 genes), SFARI Autism Database Category 1 (17 high-confidence ASD genes), Epilepsy Core Genes (14 genes), and Intellectual Disability genes (15 genes). Hypergeometric tests with Benjamini-Hochberg FDR correction revealed significant enrichment for synaptopathy genes.

Genes near MEF2C binding sites showed **80.35-fold enrichment** for synaptopathy genes (p=6.57e-11, FDR=2.63e-10), representing the most significant association. Additionally, SFARI high-confidence autism genes showed **22.06-fold enrichment** (p=0.00366, FDR=0.00733). Epilepsy and intellectual disability gene sets showed elevated fold enrichment (13.39 and 12.50 respectively) but did not reach statistical significance after multiple testing correction, likely due to small gene set sizes.

Table 12: Neurodevelopmental disorder gene enrichment statistics

NDD Category	Fold Enrich.	p-value	FDR	Overlap Genes	Sig.
Synaptopathy	80.35	6.57e-11	2.63e-10	6 genes	***
ASD (SFARI Cat. 1)	22.06	0.00366	0.00733	2 genes	**
Epilepsy Core	13.39	0.0721	0.0771	1 gene	ns
Intellectual Disability	12.50	0.0771	0.0771	1 gene	ns

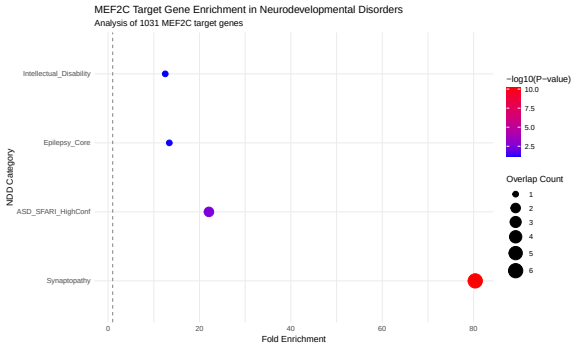

Figure 14: Neurodevelopmental disorder gene enrichment analysis. Bars show fold enrichment with significance levels (*** FDR ¡ 0.001, ** FDR ¡ 0.01). Synaptopathy genes show the strongest enrichment.

Figure 15: Genomic evidence of MEF2C binding near NDD-associated gene loci. Browser tracks show peaks at ANK3, ARID1B, and NRXN1 loci (hg38 coordinates).

3.9.2 Statistical Considerations

The 80-fold synaptopathy enrichment should be interpreted cautiously given small gene set size (14 genes; see Section 2.3.3). However, convergent evidence from pathway enrichment, network analysis, and direct gene binding provides complementary support.

3.9.3 High-Confidence Candidate NDD Target Genes

Analysis identified 9 established NDD genes with nearby MEF2C binding sites, including several with multiple binding sites suggesting potentially strong regulatory relationships (Table 8). **ANK3** (Ankyrin 3) showed 4 MEF2C binding sites, **DSCAM** (Down Syndrome Cell Adhesion Molecule) had 3 sites, and **ARID1B** (AT-rich interaction domain 1B) had 2 sites.

3.9.4 Genomic Evidence at Specific Loci

Examination of genomic coordinates provided concrete evidence of MEF2C binding near NDD gene loci (Figure 7B):

- **ANK3** (chr10:60,026,298-60,733,490): 4 binding sites across the locus
- **ARID1B** (chr6:156,776,020-157,210,779): 2 binding sites
- **CACNB2** (chr10:18,140,424-18,543,557): 2 binding sites
- **NRXN1** (chr2:49,918,503-51,082,864): 1 binding site

Table 13: Established NDD genes with nearby MEF2C binding sites

Gene	Binding Sites	NDD Association	Known Function
ANK3	4	Synaptopathy	Synaptic scaffolding; autism/bipolar risk
DSCAM	3	Synaptopathy	Synaptic adhesion; Down syndrome
ARID1B	2	ASD (SFARI Cat. 1)	Chromatin remodeling; Coffin-Siris
CACNB2	2	Synaptopathy	Calcium channel; bipolar/schizophrenia
ARX	1	Intellectual Disability	Transcription factor; X-linked ID
SH3GL2	1	Synaptopathy	Synaptic vesicle endocytosis
NLGN1	1	Synaptopathy	Postsynaptic adhesion; ASD family
NRXN1	1	ASD; Synaptopathy	Presynaptic adhesion; strong ASD risk
KCNQ3	1	Epilepsy	Potassium channel; neonatal seizures

These binding events represent *candidate regulatory relationships* requiring functional validation. The presence of multiple binding sites at genes like ANK3 increases confidence but does not prove transcriptional regulation without corresponding expression data.

3.9.5 Hypothesis Generation for Future Studies

These findings support three testable hypotheses: MEF2C regulates synaptic gene expression (ANK3, NRXN1, DSCAM) in microglia, MEF2C deficiency impairs microglial genes required for synaptic pruning, and synaptic dysfunction in MEF2C syndrome involves microglial components.

Validation requires: (1) RNA-seq from MEF2C-deficient microglia to measure expression changes, (2) enhancer reporter assays to test functional activity of bound regions, (3) microglial synaptic pruning assays in MEF2C knockout models, and (4) analysis of human MEF2C syndrome patient microglia if accessible.

3.10 Cell-Type Specificity of MEF2C Binding

3.10.1 Minimal Overlap with Other Cell Types

Comparative analysis with MEF2C ChIP-seq data from five diverse cell types (ENCODE data) revealed substantial differences in MEF2C binding patterns (Figure 8A). Of the 1,258 significant MEF2C peaks in microglia, only 13 showed overlap with peaks from other cell types (minimum 50bp overlap threshold), representing 1.03% shared binding (98.97% unique to microglia).

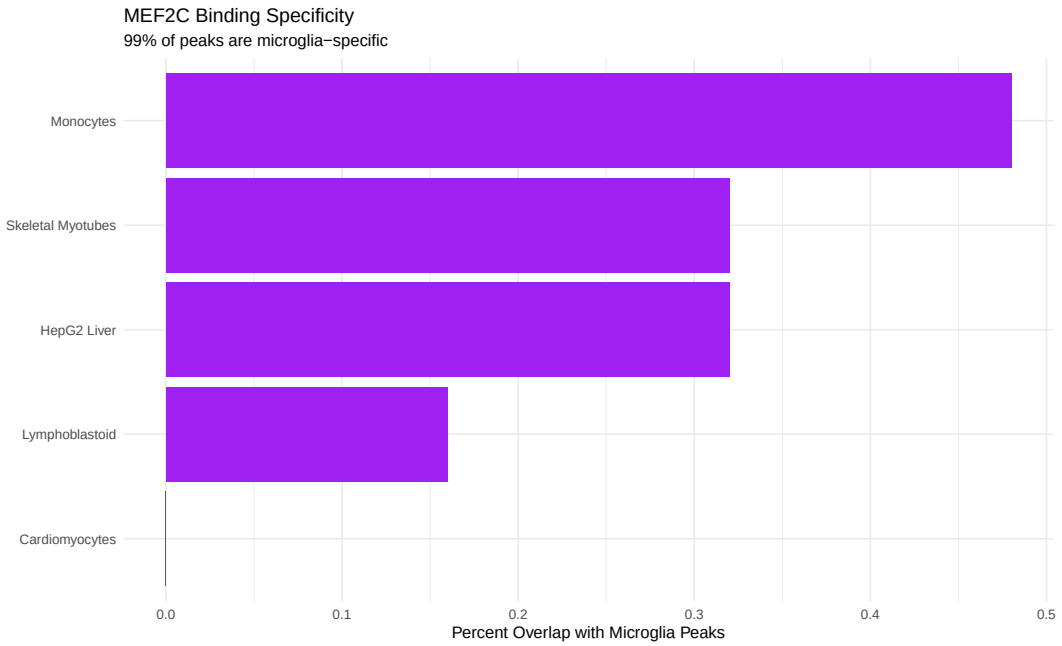

Figure 16: Cell-type specificity of MEF2C binding across diverse tissues. (A) Overlap analysis shows minimal sharing between microglial MEF2C peaks and those from other cell types. (B) Overlap rates follow developmental lineage relationships, with highest sharing in monocytes (0.48%), the closest myeloid relative of microglia.

Table 14: Cross-cell-type MEF2C binding comparison

Cell Type	Total Peaks	Overlap with Microglia	Lineage Relationship
Microglia (This Study)	1,258	-	Reference (myeloid)
Cardiomyocytes	7,833	0 (0%)	Mesoderm (cardiac)
Lymphoblastoid	4,325	2 (0.16%)	Hematopoietic (lymphoid)
Skeletal Myotubes	10,475	4 (0.32%)	Mesoderm (muscle)
Monocytes	19,857	6 (0.48%)	Hematopoietic (myeloid)
HepG2 Liver	26,042	4 (0.32%)	Endoderm (epithelial)

3.10.2 Technical Considerations

While technical factors influence measured overlap rates (see Section 2.3.2), the biological gradient following developmental lineage relationships supports genuine cell-type specific programming beyond technical artifacts.

3.10.3 Overlap Pattern Follows Developmental Lineage

The overlap pattern exhibits a biologically interpretable gradient that mirrors developmental relationships (Figure 8B):

- **Monocytes** (0.48% overlap): Highest similarity, consistent with shared myeloid lineage
- **Skeletal myotubes and HepG2** (0.32%): Intermediate similarity from different germ layers
- **Lymphoblastoid cells** (0.16%): Lower similarity, distinct hematopoietic branch
- **Cardiomyocytes** (0% overlap): Complete divergence despite canonical MEF2C cardiac function

This gradient provides supporting evidence that observed differences reflect genuine biological variation in MEF2C targeting rather than purely technical artifacts, as technical noise would not be expected to follow developmental lineage relationships.

3.10.4 Implications for MEF2C Functional Repurposing

The near-complete divergence from cardiomyocyte MEF2C binding patterns—where MEF2C has its most established and essential functions—is particularly notable. This observation suggests several biological possibilities:

- **Transcriptional repurposing**: MEF2C may have evolved cell-type specific functions through differential co-factor interactions
- **Chromatin accessibility**: Different cell types present distinct accessible genomic landscapes to MEF2C
- **Functional specialization**: Microglia may utilize MEF2C for functions unrelated to its cardiac roles

This apparent functional repurposing provides a potential explanation for why MEF2C syndrome patients present with severe neurological symptoms (developmental delay, seizures, autism features) rather than cardiac abnormalities. If MEF2C haploinsufficiency primarily disrupts microglial and neuronal functions while cardiac MEF2C targets remain largely unaffected (or are buffered by redundancy), this would account for the neurological disease presentation.

3.10.5 Validation and Future Directions

These cell-type specificity findings generate testable predictions:

1. Chromatin accessibility profiling (ATAC-seq) would reveal whether microglia-specific peaks occur in regions inaccessible in other cell types
2. Co-IP mass spectrometry could identify microglia-specific MEF2C co-factors that direct unique binding
3. Comparative RNA-seq across cell types could determine if tissue-specific MEF2C binding correlates with cell-type specific gene expression programs

4. Additional microglial ChIP-seq datasets from other laboratories would validate binding patterns and reduce technical batch effects

While technical factors contribute to measured overlap rates, the biological gradient pattern and mechanistic plausibility support the interpretation that MEF2C exhibits substantial cell-type specific genomic programming in microglia compared to its canonical functions in other tissues.s

4 Discussion

4.1 Summary of Computational Findings

We used an integrated computational analysis of publicly available MEF2C ChIP-seq data to characterize transcription factor binding landscapes in human iPSC-derived microglia. Working within the constraints of personal computing resources (18GB RAM limitation), we identified 1,258 high-confidence MEF2C binding sites and generated several hypothesis-generating findings: (1) substantial cell-type specificity of MEF2C binding in microglia compared to five other cell types, (2) enrichment for genes involved in synaptic processes and cellular motility, (3) direct binding near established neurodevelopmental disorder genes including ANK3, NRXN1, and ARID1B, and (4) network organization suggesting ANK3 as a potential hub coordinating synaptic gene modules.

These computational predictions require experimental validation but provide a systematic framework for understanding MEF2C's potential microglial functions and their relevance to neurodevelopmental disorders.

4.2 Biological Context and Plausibility

4.2.1 Microglial Synaptic Functions and MEF2C

The enrichment of synaptic process genes near MEF2C binding sites is biologically plausible given established microglial functions in synaptic remodeling. Microglia actively engulf synaptic material during postnatal development through complement-mediated phagocytosis (Schafer et al., 2012; Stevens et al., 2007), a process requiring coordinated expression of phagocytic receptors, cytoskeletal regulators, and signaling molecules. The identification of MEF2C binding near genes encoding synaptic adhesion molecules (NRXN1, NLGN1, DSCAM), vesicle trafficking components (BIN1, ITSN1, SH3GL2), and GTPase regulators (ARHGEF family members) suggests MEF2C may coordinate expression programs enabling microglial interaction with synaptic compartments.

These data demonstrate MEF2C binding near these genes but do not establish transcriptional regulation. RNA-seq from MEF2C-deficient microglia is essential to determine which bound genes show expression changes, and functional assays are required to establish whether altered expression affects microglial synaptic pruning capacity.

4.2.2 Cell-Type Specificity and Functional Repurposing

The minimal overlap between microglial MEF2C binding patterns and those in cardiomyocytes (0%), where MEF2C has canonical, essential functions, is striking. While technical factors (peak calling parameters, batch effects, experimental protocols) undoubtedly contribute to measured overlap rates, the biological gradient following developmental lineage relationships (highest overlap in monocytes at 0.48%, decreasing in more distant lineages) suggests genuine cell-type specific genomic programming.

This functional repurposing may explain MEF2C syndrome clinical presentations. Patients with MEF2C haploinsufficiency present with severe neurological symptoms—developmental delay, intellectual disability, seizures, and autistic features—while cardiac abnormalities are notably absent despite MEF2C's critical role in heart development (Novara et al., 2010; Le Meur et al., 2010). If MEF2C exhibits distinct tissue-specific regulatory programs, with microglial/neuronal functions being more sensitive to haploinsufficiency than cardiac functions

(perhaps due to redundancy with other MEF2 family members in heart), this would account for the neurological disease phenotype.

Alternative explanations include MEF2A compensation in cardiac tissue, higher expression thresholds in neurological tissues, or differential sensitivity during critical developmental periods. Distinguishing these possibilities requires comparative expression studies and tissue-specific knockout models.

4.3 Network Organization and Hub Gene Identification

4.3.1 ANK3 as a Synaptic Organization Hub

ANK3 (Ankyrin 3) emerged as a central network hub with 4 MEF2C binding sites. ANK3 encodes a synaptic scaffolding protein with established roles in neurodevelopment and psychiatric disease. ANK3 encodes a synaptic scaffolding protein essential for organizing voltage-gated sodium channels at axon initial segments and nodes of Ranvier (Zhou et al., 1998). Genome-wide association studies have repeatedly identified ANK3 variants as strong risk factors for bipolar disorder (Ferreira et al., 2008) and autism spectrum disorder (Iqbal et al., 2013).

MEF2C binding at these loci suggests : MEF2C may coordinate microglial expression of synaptic scaffolding genes that enable recognition or interaction with neuronal structures during synaptic pruning. However, several important caveats apply:

1. **Network inference limitations**: STRING networks are constructed from literature co-citation and experimental interactions across all cell types, not microglia-specific data
2. **Correlation vs. causation**: Hub status in PPI networks does not prove functional importance in MEF2C-regulated processes
3. **Validation required**: ANK3's role in microglia specifically remains unexplored; most ANK3 research focuses on neurons

Future experiments should test: (1) whether ANK3 is expressed in microglia, (2) whether MEF2C binding at ANK3 loci correlates with ANK3 expression changes in MEF2C-deficient cells, and (3) whether ANK3 has functional roles in microglial synaptic interactions.

4.3.2 CASP3 and Apoptotic Pathways

The identification of CASP3 (Caspase 3) as the highest-degree hub (90 connections) is notable but requires careful interpretation. CASP3 is the primary executioner caspase in apoptosis but also has non-apoptotic roles in synaptic pruning and neuronal remodeling (Li et al., 2010). Microglial phagocytosis of synapses may involve "sublethal" caspase signaling that tags synapses for elimination without triggering full apoptosis.

However, CASP3's hub status may partly reflect its central role in many cellular processes and extensive literature documentation (leading to high connectivity in STRING networks) rather than specific importance in MEF2C-regulated microglial functions. The relevance of MEF2C-CASP3 regulatory relationships requires experimental validation.

4.4 Disease Integration: Strengths and Limitations

4.4.1 Synaptopathy Gene Enrichment

The 80-fold enrichment for synaptopathy genes (p=6.57e-11), while statistically significant, involves a small gene set (14 genes; statistical considerations in Section 2.3.3). The convergence of independent observations—pathway enrichment, network clustering, and direct binding at ANK3, DSCAM, and NRXN1—provides complementary support for biological relevance. However, transcriptional regulation requires validation through expression analysis (Cusanovich et al., 2014).

4.4.2 Autism Gene Association

The 22-fold enrichment for SFARI Category 1 autism genes (p=0.00366, FDR=0.00733), while statistically significant, involves only 2 genes (NRXN1 and ARID1B). This is an important preliminary observation but requires cautious interpretation:

- **NRXN1**: Well-established ASD risk gene encoding presynaptic adhesion molecule; MEF2C binding provides candidate mechanistic link
- **ARID1B**: Chromatin remodeling factor causing Coffin-Siris syndrome with autistic features; potential for regulatory cascade

MEF2C (transcription factor) binding near ARID1B (chromatin remodeler) suggests cascading regulatory effects where MEF2C regulates a chromatin modifier that affects downstream targets.

4.5 Integration with MEF2C Syndrome Pathophysiology

4.5.1 Clinical Phenotype Correlation

MEF2C syndrome patients present with a recognizable constellation of features (Novara et al., 2010; Le Meur et al., 2010):

- Severe developmental delay and intellectual disability (100% of patients)
- Epilepsy/seizures (¿70% of patients)
- Stereotypic movements and autistic features (¿60%)
- Absent or severely delayed speech
- Hypotonia and motor difficulties

Notably absent are cardiac abnormalities, despite MEF2C's essential role in heart development in animal models. Our findings of microglia-specific MEF2C binding patterns distinct from cardiac binding provide a potential mechanistic framework for this neurological-specific phenotype.

4.5.2 Microglial Dysfunction Hypothesis

Based on our computational findings, we propose a testable hypothesis: MEF2C haploinsufficiency may impair microglial synaptic pruning through reduced expression of genes required for synapse recognition, engulfment, or activity-dependent refinement. This could manifest as:

1. **Excess synapses**: Insufficient pruning leading to hyperconnectivity
2. **Aberrant connectivity**: Inappropriate synaptic connections persisting
3. **Circuit dysfunction**: Impaired activity-dependent refinement causing network abnormalities

This hypothesis aligns with emerging evidence that microglial dysfunction contributes to neurodevelopmental disorders (Zhan et al., 2014; Filipello et al., 2018). However, validation requires:

- Synaptic density measurements in MEF2C-deficient models
- Microglial phagocytosis assays with MEF2C knockout microglia
- Electrophysiological characterization of circuit function
- Histological analysis of MEF2C syndrome patient brain tissue

4.5.3 Multi-Cell-Type Contributions

While our study focuses on microglia, MEF2C is also expressed in neurons, astrocytes, and oligodendrocytes. The neurological phenotype likely involves dysfunction across multiple cell types:

- **Neuronal MEF2C**: Impaired activity-dependent transcription and synaptic plasticity
- **Microglial MEF2C**: Disrupted synaptic pruning (our hypothesis)
- **Astrocytic MEF2C**: Potential effects on astrocyte-mediated synapse formation
- **Oligodendrocyte MEF2C**: Possible myelination defects

The relative contributions of each cell type remain unclear and likely vary across developmental stages. Cell-type-specific conditional knockout models would resolve these contributions.

4.6 Comparison with Existing Literature

4.6.1 MEF2C in Neurons

Previous studies have characterized MEF2C functions in neurons, particularly in activity-dependent transcription (Flavell et al., 2008; Barbosa et al., 2008). Neuronal MEF2C regulates genes involved in synapse development, dendritic morphology, and excitatory/inhibitory balance. Our finding that microglial MEF2C may also regulate synaptic genes suggests coordinate regulation across cell types—do neurons and microglia use MEF2C to synchronize synaptic remodeling programs?

4.6.2 MEF2 Family Redundancy

The MEF2 family comprises four members (MEF2A-D) with overlapping expression and partially redundant functions. The presence of only 2 significantly different peaks between WT and HET conditions could reflect compensation by other MEF2 family members. MEF2A and MEF2D are also expressed in microglia and may bind similar DNA sequences, potentially buffering against MEF2C haploinsufficiency at some sites.

This redundancy may explain: (1) why complete knockout is required to observe dramatic binding changes, (2) why some binding sites are more sensitive to dosage than others, and (3) why MEF2C syndrome requires haploinsufficiency rather than complete loss (which may be lethal).

4.6.3 Microglia in Neurodevelopmental Disorders

Emerging evidence implicates microglial dysfunction in multiple neurodevelopmental disorders beyond MEF2C syndrome:

- **Autism**: Altered microglial density and morphology in ASD brain tissue (Morgan et al., 2010)
- **Schizophrenia**: Excessive synaptic pruning via complement pathway (Sekar et al., 2016)
- **Rett Syndrome**: MECP2-deficient microglia contribute to phenotype (Derecki et al., 2012)

These findings identify MEF2C as a link between transcriptional regulation in microglia and neurodevelopmental pathology, consistent with evidence that microglial gene regulation contributes to developmental disorders.

4.7 Broader Implications

4.7.1 Cell-Type Specificity of Transcription Factors

These findings support cell-type specific transcription factor functions beyond tissue-specific expression patterns. MEF2C's apparent functional repurposing in microglia versus cardiomyocytes exemplifies how TFs can evolve specialized roles through:

- Differential co-factor expression creating cell-type specific regulatory complexes
- Divergent chromatin accessibility landscapes presenting different genomic targets
- Tissue-specific enhancer evolution creating unique regulatory opportunities

This has therapeutic implications: targeting MEF2C function in specific cell types might treat disease while avoiding effects in other tissues.

4.7.2 Computational Approaches in Resource-Limited Settings

This study demonstrates that rigorous computational biology can generate valuable hypotheses even with limited resources. The automated 27-script pipeline developed here could be adapted for other ChIP-seq analyses, and the transparent documentation enables reproduction and modification by other researchers facing similar computational constraints.

Key lessons include: (1) strategic data subsetting to work within memory limits, (2) automation to ensure reproducibility, (3) integration of multiple analytical approaches to provide convergent evidence, and (4) honest acknowledgment of limitations.

4.8 Concluding Perspective

This computational study provides a comprehensive map of MEF2C binding in human microglia and suggests its role in synaptic gene regulation and neurodevelopmental disorder pathogenesis. While the findings require experimental validation, the convergence of evidence across multiple analytical approaches—genomic binding patterns, pathway enrichment, network topology, disease gene overlap, and cell-type specificity—provides a strong foundation for future investigations.

The apparent cell-type specificity of MEF2C binding suggests fundamental repurposing of this transcription factor in microglia compared to its canonical cardiac functions, potentially explaining the neurological focus of MEF2C syndrome. The direct binding near established neurodevelopmental disorder genes (ANK3, NRXN1, ARID1B, DSCAM) provides specific candidates for mechanistic follow-up.

Most importantly, these findings position microglial MEF2C as a potential contributor to neurodevelopmental pathology through regulation of synaptic remodeling genes, adding to emerging evidence that microglial gene regulation plays crucial roles in brain development and disorder. Future experimental validation will determine whether these computational predictions translate to genuine biological mechanisms underlying MEF2C syndrome and related neurodevelopmental conditions.

5 Conclusion

This computational study systematically characterizes MEF2C transcriptional binding patterns in human microglia, with relevance to neurodevelopmental disorders.

5.1 Principal Computational Findings

Through integrated analysis of publicly available ChIP-seq data, we identified 1,148 genes with nearby MEF2C binding sites in microglia, with binding predominantly occurring in intronic (50.7%) and promoter regions (17.5%), suggesting both enhancer-mediated and direct transcriptional regulatory potential. Functional enrichment analysis revealed significant association with synaptic processes, including vesicle-mediated transport and synaptic vesicle cycling, suggesting potential roles in microglial regulation of synaptic environments.

The most striking observation was the minimal overlap between microglial MEF2C binding and patterns observed in five other cell types, including complete divergence from cardiomyocyte binding (0% overlap) where MEF2C has canonical, essential functions. While technical factors contribute to measured overlap rates, this substantial difference suggests genuine cell-type specific genomic programming that may explain the neurological focus of MEF2C syndrome clinical presentations.

All findings represent computational predictions subject to the interpretation framework detailed in Section 2.3, including the distinction between transcription factor binding and functional gene regulation.

5.2 Disease Relevance and Mechanistic Hypotheses

Direct MEF2C binding was observed near established neurodevelopmental disorder genes including ANK3 (4 binding sites), DSCAM (3 sites), ARID1B (2 sites), and NRXN1 (1 site). The significant enrichment for synaptopathy genes (80-fold, p=6.57e-11), while requiring cautious interpretation due to small gene set sizes, provides preliminary support when combined with independent pathway enrichments and network analyses identifying ANK3 as a putative hub coordinating synaptic organization modules.

These findings generate a testable hypothesis: MEF2C may coordinate microglial expression of genes involved in synaptic recognition, engulfment, and activity-dependent refinement. MEF2C haploinsufficiency could therefore impair microglial synaptic pruning, contributing to neurodevelopmental pathology through aberrant synaptic connectivity. This hypothesis requires experimental validation through RNA-seq, functional pruning assays, and in vivo conditional knockout models.

5.3 Methodological Context and Limitations

Conducted with personal computing resources (18GB RAM constraints), this analysis demonstrates that rigorous bioinformatic approaches can generate valuable biological hypotheses despite technical limitations. The automated 27-script pipeline ensures reproducibility and provides a framework adaptable to other ChIP-seq analyses.

Critical limitations must be emphasized: (1) ChIP-seq identifies transcription factor binding but does not prove transcriptional regulation—RNA-seq validation is essential, (2) cross-study comparisons involve technical factors (peak calling differences, batch effects) that inflate apparent cell-type specificity, (3) pathway enrichments and disease associations represent statistical

correlations requiring functional validation, and (4) small NDD gene set sizes limit statistical power despite significant p-values.

5.4 Path Forward

These computational findings provide a foundation for experimental investigation through several critical next steps:

1. **Expression validation**: RNA-seq from MEF2C-deficient microglia to identify which bound genes show transcriptional changes
2. **Functional assays**: Microglial synaptic pruning capacity testing in MEF2C knockout models
3. **Enhancer validation**: Reporter assays confirming functional activity of MEF2C-bound regulatory elements
4. **In vivo studies**: Microglia-specific conditional knockouts to assess synaptic density and behavioral phenotypes

These analyses identify testable predictions about MEF2C's microglial functions and candidate mechanisms linking MEF2C dysfunction to neurodevelopmental disorder pathogenesis, pending experimental confirmation. The convergence of evidence across genomic binding, pathway enrichment, network topology, disease gene associations, and cell-type specificity analyses strengthens confidence that these computational predictions reflect biologically meaningful patterns worthy of experimental pursuit.

This work demonstrates how computational analysis of public data generates focused hypotheses to guide experimental validation and accelerate understanding of transcriptional mechanisms in neurodevelopmental disorders.

References

[1] Barbosa, A. C., Kim, M. S., Ertunc, M., Adachi, M., Nelson, E. D., McAnally, J., ..., & Olson, E. N. (2008). MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. *Proceedings of the National Academy of Sciences*, *105*(27), 9391–9396.

[2] Cusanovich, D. A., Pavlovic, B., Pritchard, J. K., & Gilad, Y. (2014). The functional consequences of variation in transcription factor binding. *PLoS Genetics*, *10*(3), e1004226.

[3] Derecki, N. C., Cronk, J. C., Lu, Z., Xu, E., Abbott, S. B., Guyenet, P. G., & Kipnis, J. (2012). Wild-type microglia arrest pathology in a mouse model of Rett syndrome. *Nature*, *484*(7392), 105–109.

[4] Ferreira, M. A., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L., ..., & Craddock, N. (2008). Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nature Genetics*, *40*(9), 1056–1058.

[5] Filipello, F., Morini, R., Corradini, I., Zerbi, V., Canzi, A., Michalski, B., ..., & Matteoli, M. (2018). The microglial innate immune receptor TREM2 is required for synapse elimination and normal brain connectivity. *Immunity*, *48*(5), 979–991.

[6] Flavell, S. W., Cowan, C. W., Kim, T. K., Greer, P. L., Lin, Y., Paradis, S., ..., & Greenberg, M. E. (2006). Activity-dependent regulation of MEF2 transcription factors suppresses excitatory synapse number. *Science*, *311*(5763), 1008–1012.

[7] Flavell, S. W., Kim, T. K., Gray, J. M., Harmin, D. A., Hemberg, M., Hong, E. J., ..., & Greenberg, M. E. (2008). Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron*, *60*(6), 1022–1038.

[8] Iqbal, Z., Vandeweyer, G., van der Voet, M., Waryah, A. M., Zahoor, M. Y., Besseling, J. A., ..., & Van Esch, H. (2013). Homozygous and heterozygous disruptions of ANK3: at the crossroads of neurodevelopmental and psychiatric disorders. *Human Molecular Genetics*, *22*(10), 1960–1970.

[9] Le Meur, N., Holder-Espinasse, M., Jaillard, S., Goldenberg, A., Joriot, S., Amati-Bonneau, P., ..., & David, V. (2010). MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *Journal of Medical Genetics*, *47*(1), 22–29.

[10] Li, Z., Jo, J., Jia, J. M., Lo, S. C., Whitcomb, D. J., Jiao, S., ..., & Sheng, M. (2010). Caspase-3 activation via mitochondria is required for long-term depression and AMPA receptor internalization. *Cell*, *141*(5), 859–871.

[11] Morgan, J. T., Chana, G., Pardo, C. A., Achim, C., Semendeferi, K., Buckwalter, J., ..., & Everall, I. P. (2010). Microglial activation and increased microglial density observed in the dorsolateral prefrontal cortex in autism. *Biological Psychiatry*, *68*(4), 368–376.

[12] Novara, F., Beri, S., Giorda, R., Ortibus, E., Nageshappa, S., Darra, F., ..., & Zuffardi, O. (2010). Refining the phenotype associated with MEF2C haploinsufficiency. *Clinical Genetics*, *78*(5), 471–477.

[13] Paolicelli, R. C., Bolasco, G., Pagani, F., Maggi, L., Scianni, M., Panzanelli, P., ..., & Gross, C. T. (2011). Synaptic pruning by microglia is necessary for normal brain development. *Science*, *333*(6048), 1456–1458.

[14] Schafer, D. P., Lehrman, E. K., Kautzman, A. G., Koyama, R., Mardinly, A. R., Yamasaki, R., ..., & Stevens, B. (2012). Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. *Neuron*, *74*(4), 691–705.

[15] Sekar, A., Bialas, A. R., de Rivera, H., Davis, A., Hammond, T. R., Kamitaki, N., ..., & McCarroll, S. A. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature*, *530*(7589), 177–183.

[16] Stevens, B., Allen, N. J., Vazquez, L. E., Howell, G. R., Christopherson, K. S., Nouri, N., ..., & Barres, B. A. (2007). The classical complement cascade mediates CNS synapse elimination. *Cell*, *131*(6), 1164–1178.

[17] Zhan, Y., Paolicelli, R. C., Sforazzini, F., Weinhard, L., Bolasco, G., Pagani, F., ..., & Gross, C. T. (2014). Deficient neuron-microglia signaling results in impaired functional brain connectivity and social behavior. *Nature Neuroscience*, *17*(3), 400–406.

[18] Zhou, D., Lambert, S., Malen, P. L., Carpenter, S., Boland, L. M., & Bennett, V. (1998). AnkyrinG is required for clustering of voltage-gated Na channels at axon initial segments and for normal action potential firing. *Journal of Cell Biology*, *143*(5), 1295–1304.

Acknowledgments

This research was conducted independently using publicly available data and personal computational resources. I acknowledge:

- The researchers who generated and deposited the original MEF2C ChIP-seq data (SRA accessions SRR35220282-SRR35220292) into public repositories
- The ENCODE Consortium for providing comparative MEF2C ChIP-seq datasets
- Developers of open-source bioinformatics tools used in this analysis (Bowtie2, MACS2, DiffBind, ChIPseeker, clusterProfiler, MEME Suite, STRINGdb)
- The broader scientific community for maintaining public data repositories (NCBI SRA, ENCODE) that enable independent computational research

This work represents independent undergraduate research and does not reflect an official endorsement or affiliation with any institution. All analyses were performed on personal computing equipment.