

nf-sarcope

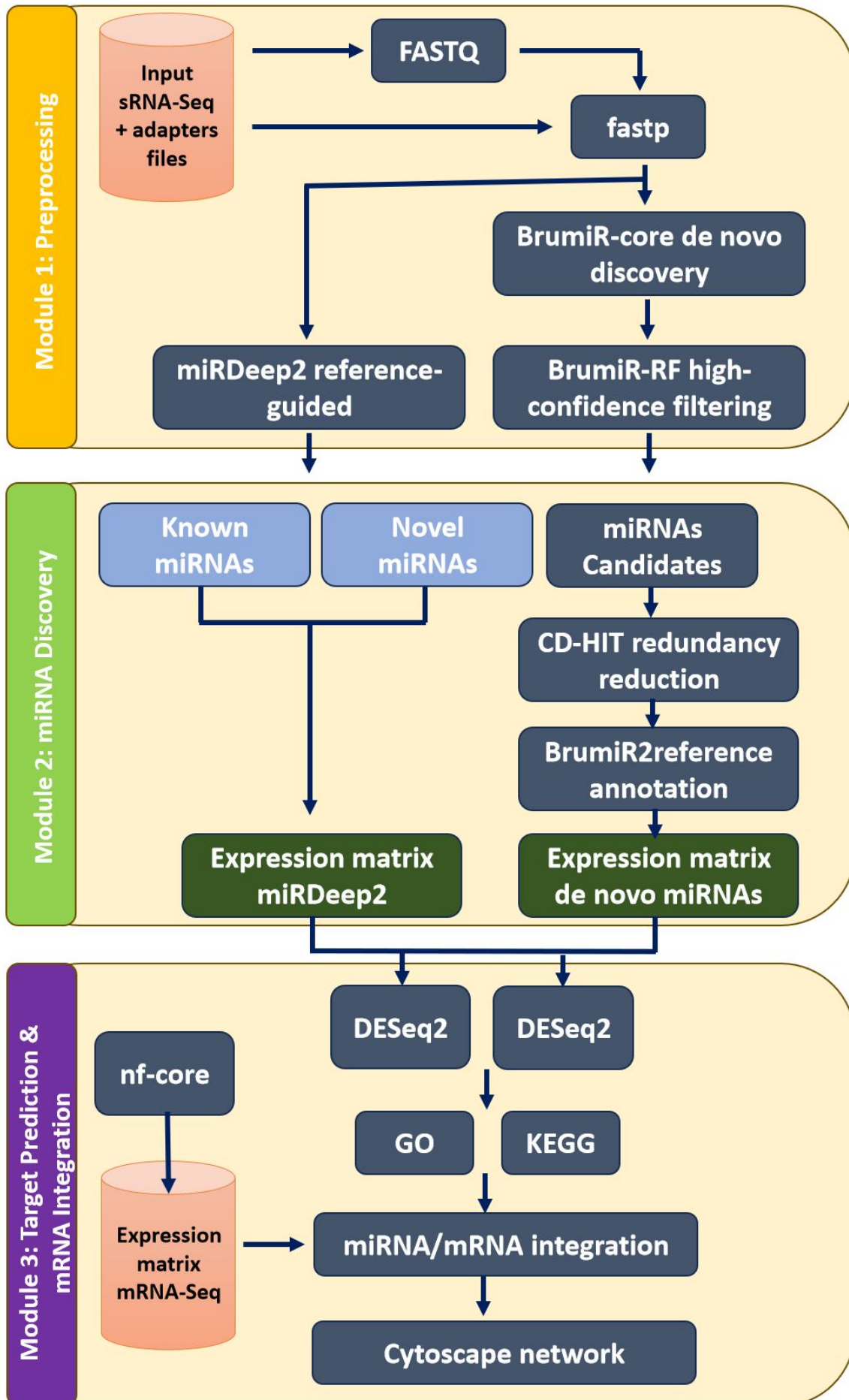


Figure 1. Overview of the *nf-Sarcope* workflow. Sarcope is a modular Nextflow DSL2 pipeline designed for the integrated analysis of sRNA-seq and mRNA-seq data to identify miRNA–mRNA regulatory interactions associated with exercise adaptation and skeletal muscle remodeling. The workflow is organized into three modules: (I) Preprocessing, which includes quality control and adapter trimming (*fastp*), *de novo* miRNA discovery using BrumiR, machine learning–based filtering (BrumiR-RF), and reference-guided detection using miRDeep2; (II) miRNA Discovery, which performs redundancy reduction (CD-HIT), genomic annotation (BrumiR2Reference), expression matrix construction, and identification of core miRNAs based on intra-group presence; and (III) Target Prediction & mRNA Integration, which includes differential expression analysis (DESeq2) and integrative miRNA–mRNA network reconstruction using mRNA-seq expression data.

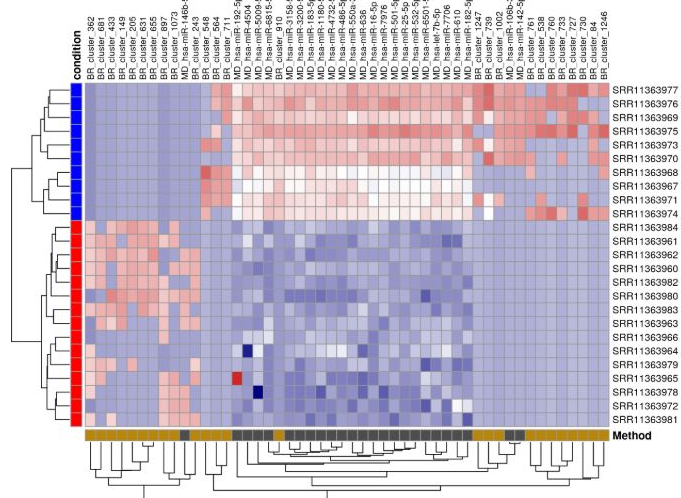
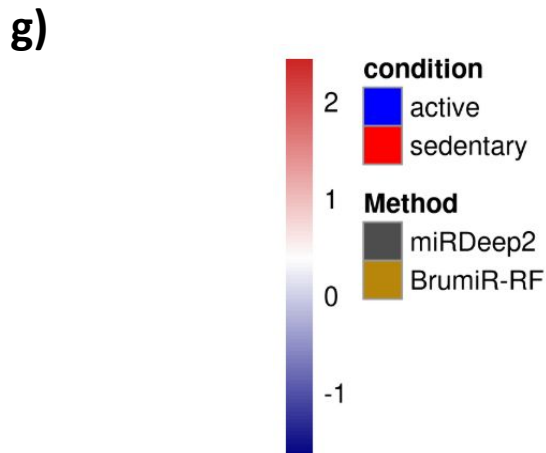
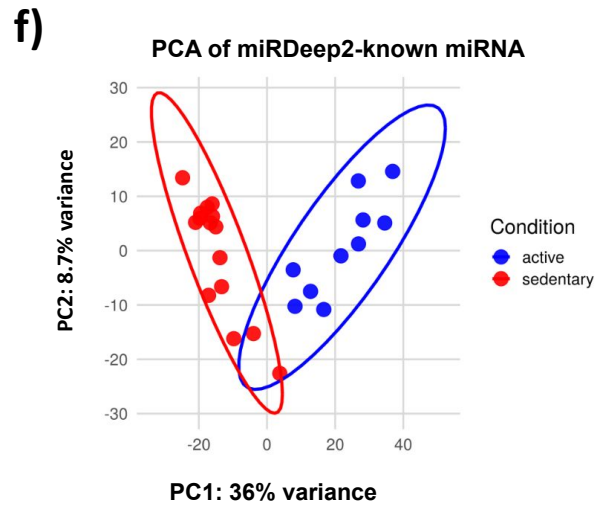
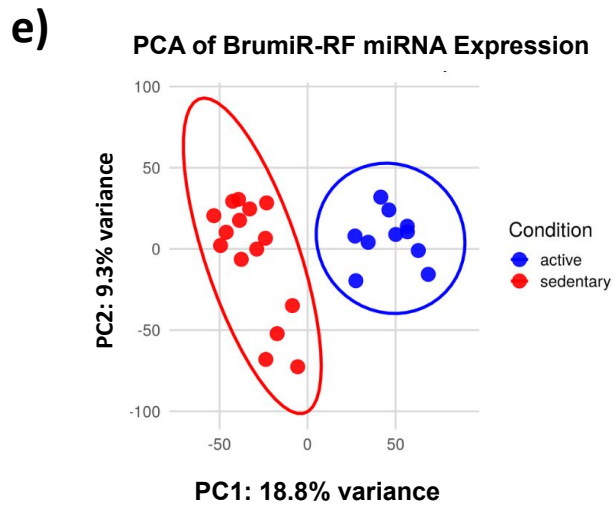
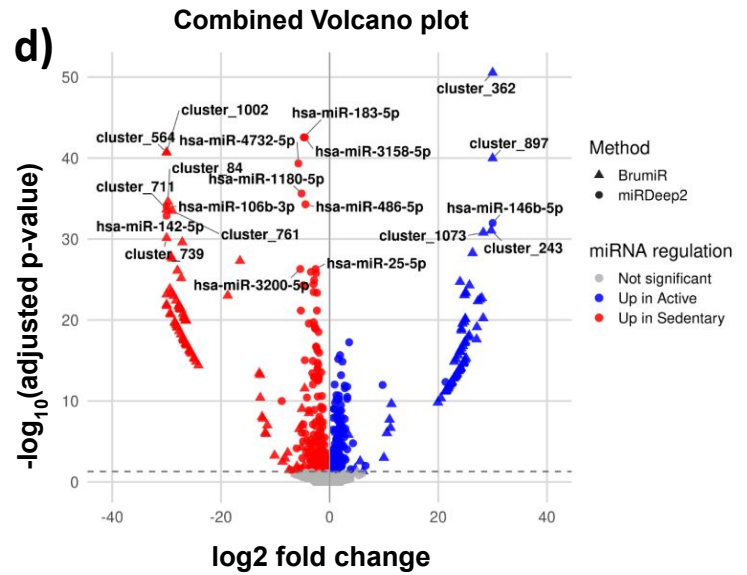
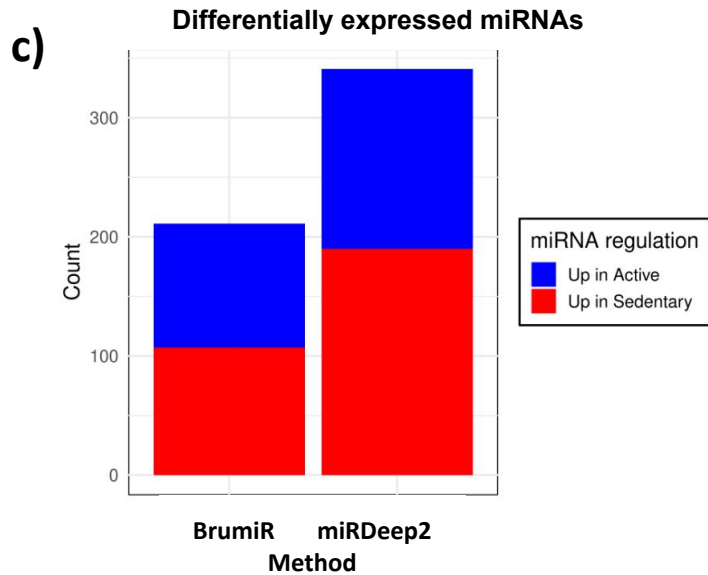
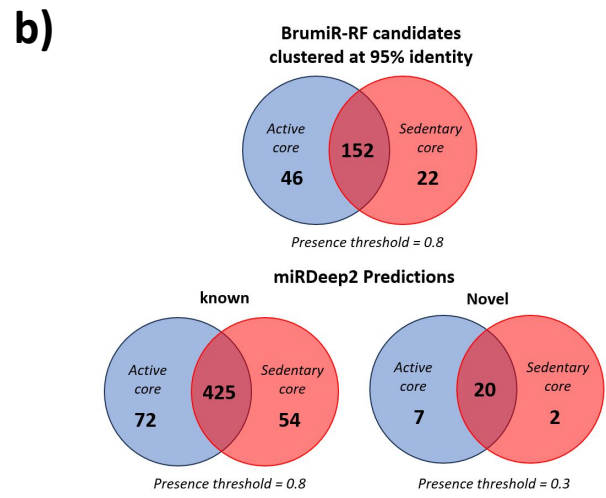
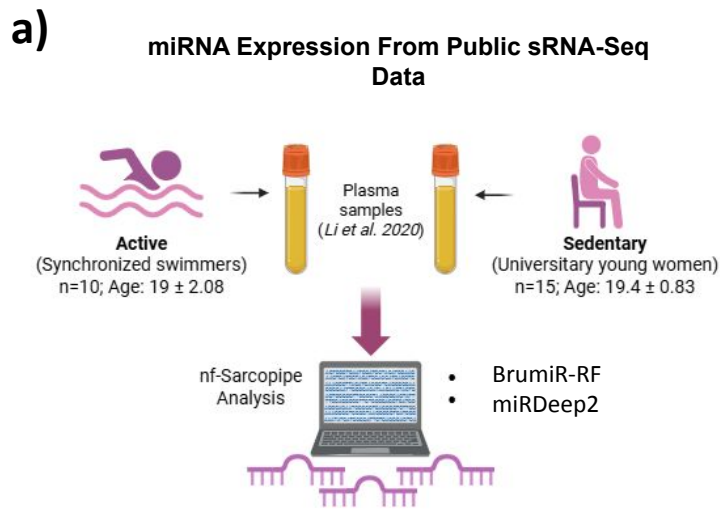
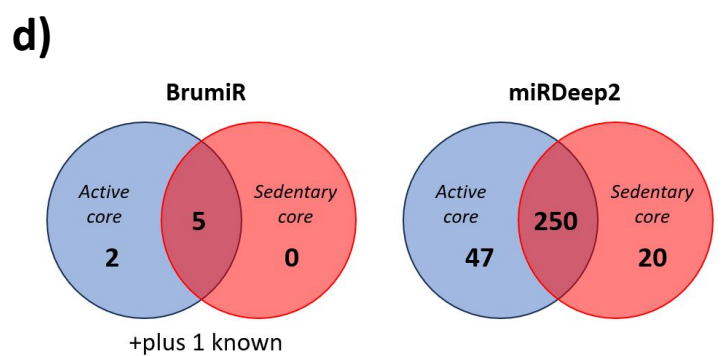
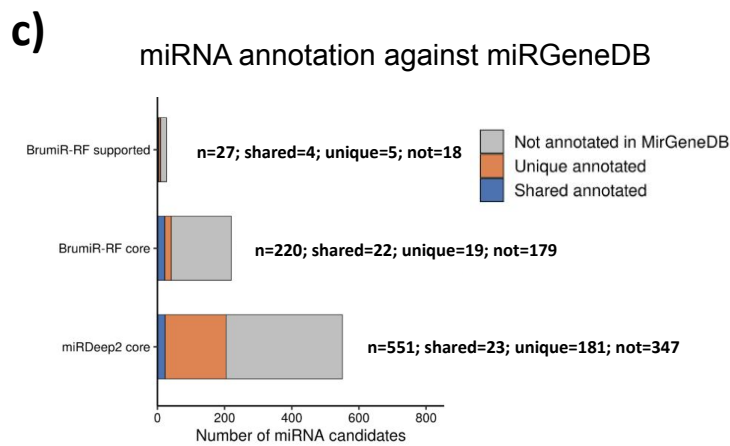
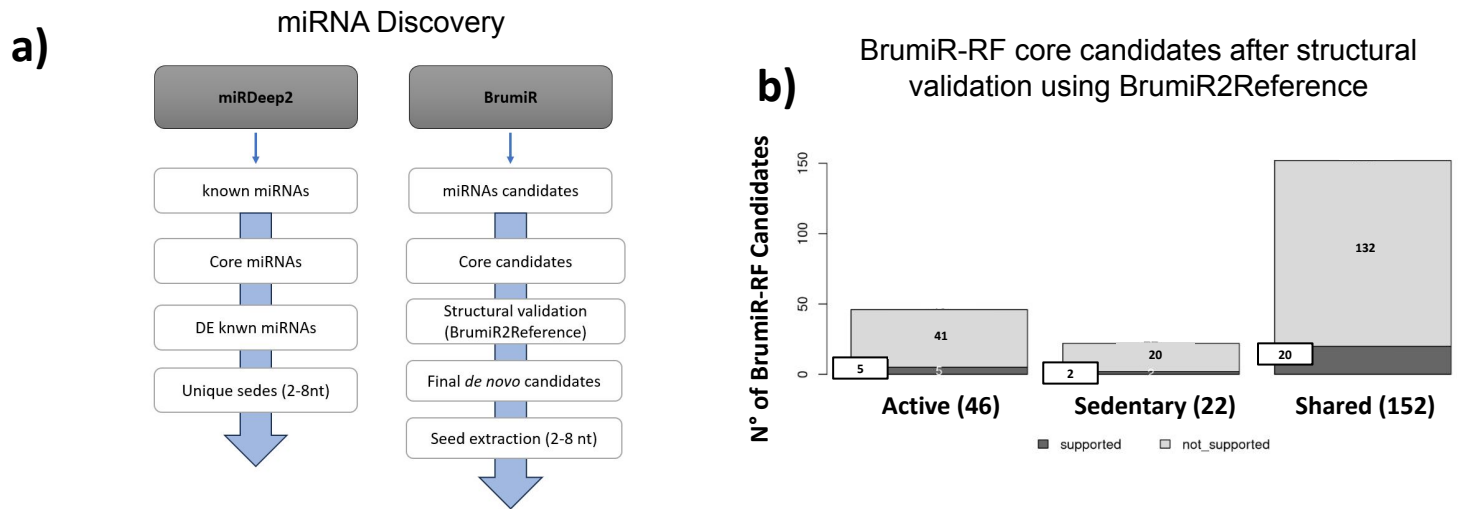


Figure 2. Overview and statistical characterization of circulating miRNA expression profiles analyzed with nf-Sarcopipe. (a) Schematic representation of the study design and data processing workflow. Public plasma sRNA-seq data from active (synchronized swimmers, $n = 10$) and sedentary (university women, $n = 15$) individuals were processed using nf-Sarcopipe, integrating both reference-guided (miRDeep2) and *de novo* (BrumiR-RF) miRNA detection strategies. (b) Venn diagrams summarizing core miRNA sets for each method and condition. For BrumiR-RF, candidate miRNAs clustered at 95% sequence identity were filtered using a presence threshold of 0.8. For miRDeep2, known miRNAs (miRBase-annotated) were filtered at a presence threshold of 0.8, while novel predictions were filtered at 0.3. Presence thresholds were selected to ensure sufficient representation of detectable miRNA features across samples while maintaining consistency between methods; accordingly, BrumiR-RF candidates were analyzed using a presence threshold of 0.8. miRDeep2 novel predictions were not considered in downstream analyses. (c) Stacked bar plot showing the number of differentially expressed miRNAs identified by each method. Differential expression analysis was performed using DESeq2 with Wald test statistics and Benjamini–Hochberg correction for multiple testing (adjusted $p < 0.05$). (d) Combined volcano plot displaying log2 fold change versus $-\log_{10}$ adjusted p -value for miRNAs detected by BrumiR-RF and miRDeep2. Top 10 Significant miRNAs from different algorithms are colored according to regulation direction, and point shapes indicate the detection method. (e–f) Principal component analysis (PCA) of variance-stabilized miRNA expression values (DESeq2 VST transformation). PCA was performed independently for (e) BrumiR-RF candidates and miRDeep2-known miRNAs (f) using the top 500 most variable features. (g) Heatmap hierarchical clustering of the 25 most significant BrumiR-RF candidate clusters and the 25 most significant miRDeep2 annotated miRNAs, using variance-stabilized expression values scaled by feature.



e) BrumiR novel miRNAs

Active Candidates	Cluster_740 → hsa-miR-novel_A miRNA (seed): AGCCACC mRNA 3'UTR: ... UCGGUGG...		Cluster_986 → hsa-miR-novel_B miRNA (seed): GUGCUUU mRNA 3'UTR: ... CACGAAA...	
	Cluster_42 → hsa-miR-novel_C miRNA (seed): AGUGUGA mRNA 3'UTR: ... UCAUACU...		Cluster_374 → hsa-miR-novel_D miRNA (seed): CGCCAAU mRNA 3'UTR: ... GCGGUUA...	
Shared Candidates	Cluster_413 → hsa-miR-novel_E miRNA (seed): ACAGAUG mRNA 3'UTR: ... UGUCUAC...		Cluster_465 → hsa-miR-novel_F miRNA (seed): ACCAAUA mRNA 3'UTR: ... UGGUUAU...	
			Cluster_485 → hsa-miR-novel_G miRNA (seed): AGUGCCU mRNA 3'UTR: ... UCACGGA...	

Figure 3. Prioritization and functional characterization of *de novo* miRNA candidates identified by *nf-Sarcope*. (a) Overview of the miRNA discovery and prioritization workflow. miRDeep2-derived miRNAs were subjected to differential expression analysis followed by extraction of seed regions (nucleotides 2–8), yielding a set of unique seeds for downstream target prediction. In parallel, BrumiR-derived candidates were filtered through core set definition, structural validation (BrumiR2Reference), and seed-based target prediction, resulting in a refined set of high-confidence *de novo* miRNA candidates. (b) Stacked bar plot summarizing BrumiR-RF core candidates (presence ≥ 0.8) across conditions (active, sedentary, and shared) following structural validation using BrumiR2Reference. Dark and light segments indicate candidates supported or not supported by genomic mapping and secondary structure prediction (RNAfold), respectively. (c) Annotation of BrumiR-RF and miRDeep2 core miRNA candidates against MirGeneDB. Matches required $\geq 98\%$ identity, $\geq 75\%$ query coverage, and zero mismatches. BrumiR-RF candidates are shown before and after BrumiR2Reference structural validation. miRDeep2 candidates classified as *Not annotated in MirGeneDB* correspond to known miRNAs annotated in miRBase but lacking a qualifying MirGeneDB match. (d) Distribution of final differentially expressed miRNA candidates according to their original core-set membership. BrumiR-RF candidates correspond to structurally supported *de novo* miRNAs with adjusted $p < 0.05$, whereas miRDeep2 candidates correspond to differentially expressed annotated miRNAs within the known core union. Shared candidates were recurrently detected in both groups but differed significantly in abundance. (e) Schematic representation of seed-mediated target interactions for representative BrumiR-derived miRNA *de novo* candidates. Seed regions (nucleotides 2–8) are aligned to complementary sequences within predicted mRNA 3'UTRs, illustrating canonical base-pairing interactions. Candidate clusters were renamed using a standardized nomenclature (hsa-miR-novel_A–G).

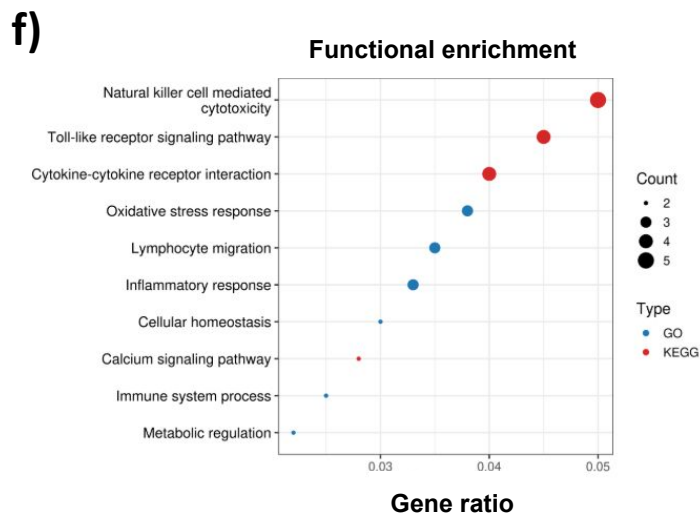
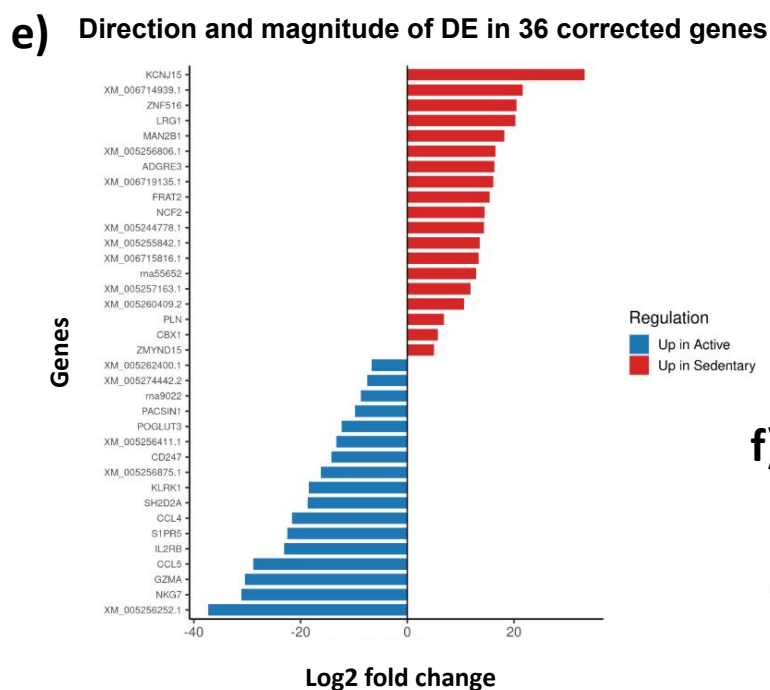
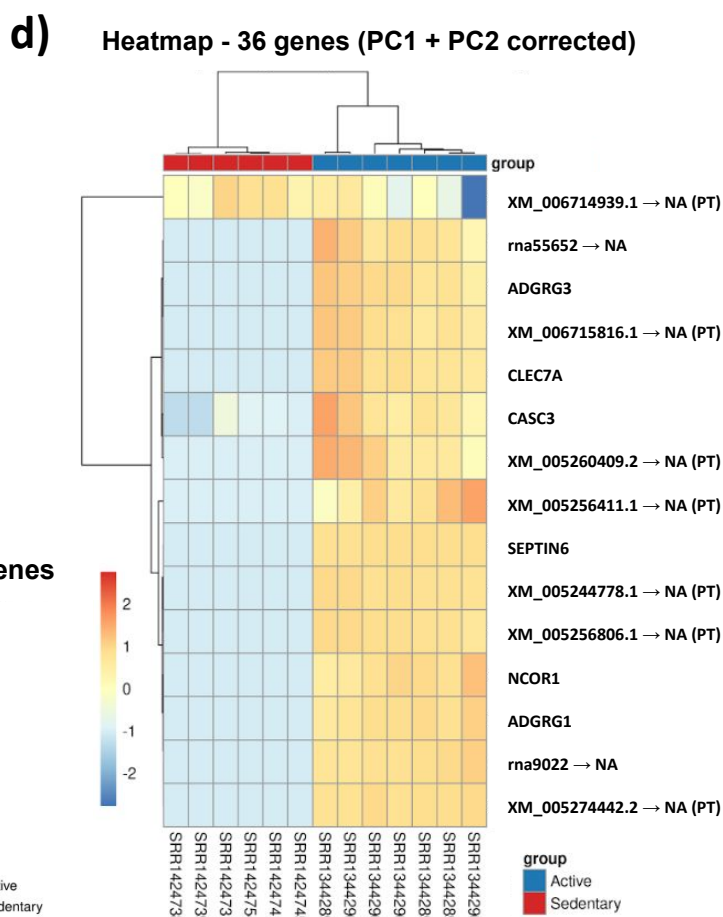
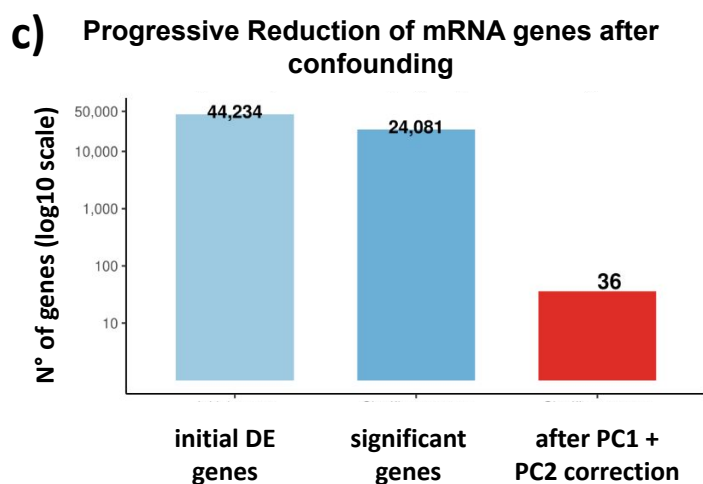
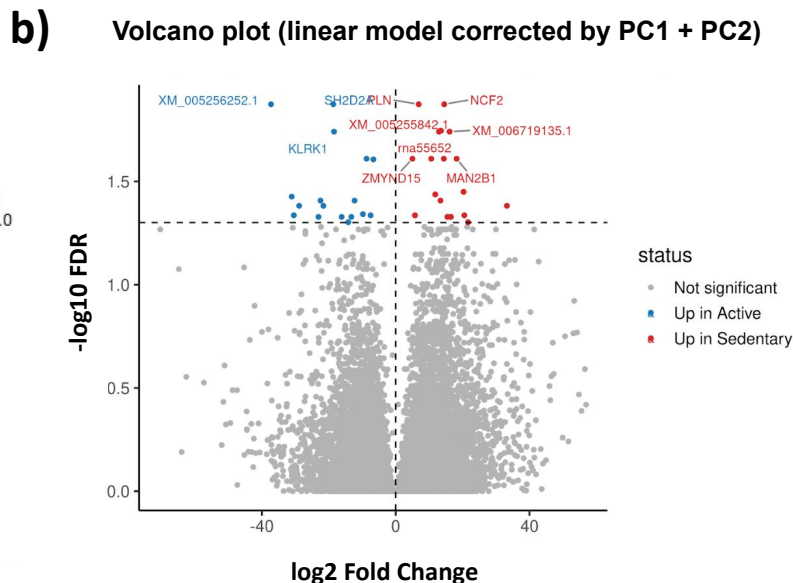
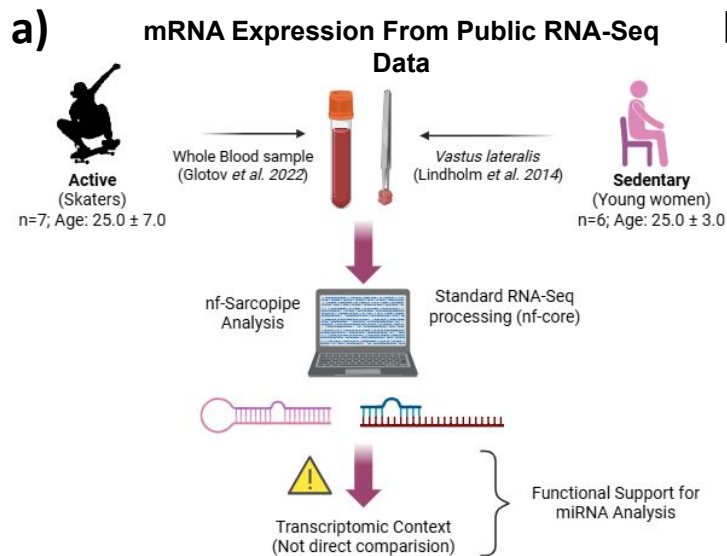
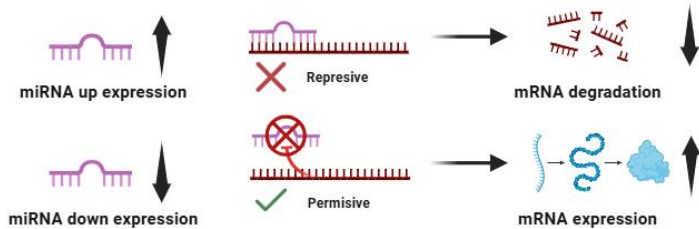
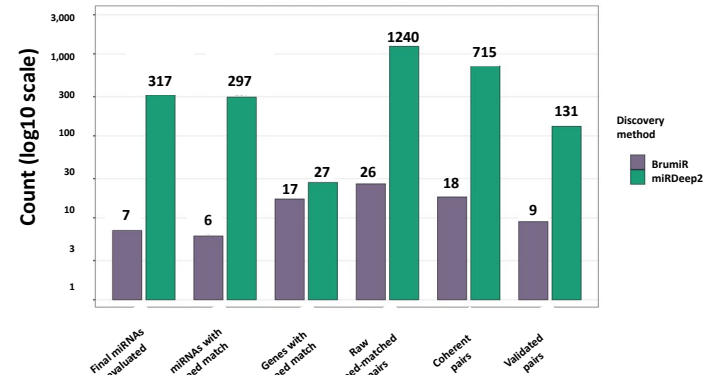


Figure 4. Correction of tissue-driven confounding reveals a robust transcriptional signature associated with physical activity. (a) Schematic overview of the study design, integrating RNA-seq data from whole blood (Active group) and skeletal muscle (Sedentary group), processed through a standardized pipeline. (b) Volcano plot of differential expression results after correction using a linear model including PC1 and PC2 as covariates. A limited number of genes remain significant ($FDR < 0.05$), indicating removal of confounding-driven signals. (c) Barplot illustrates the progressive reduction in candidate genes from 44,234 filtered transcripts, to 24,081 significant genes prior to correction, and finally to 36 genes after correction for confounding effects. (d) Heatmap of the corrected gene set ($n = 36$), showing clear separation between Active and Sedentary individuals. Due to filtering and annotation constraints, a subset of genes ($n = 15$) is displayed. NA = Not Annotated; PT = Predicted Transcript. (e) Diverging barplot showing the direction and magnitude of differential expression for the 36 corrected genes. Positive values indicate genes upregulated in Sedentary individuals, while negative values indicate genes upregulated in Active individuals. (f) Curated functional enrichment analysis of the 36 corrected genes, highlighting pathways related to immune regulation, oxidative stress, inflammation, and cellular homeostasis. Only biologically relevant terms were retained after manual curation to exclude unrelated disease- or infection-driven pathways.

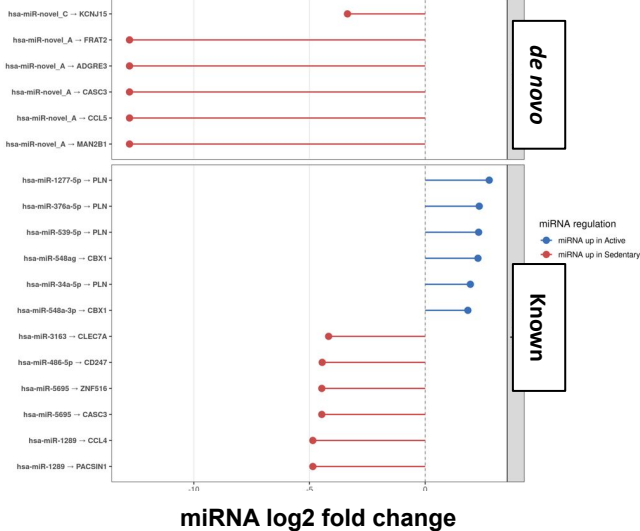
a) Coherent-pair integration miRNA/mRNA



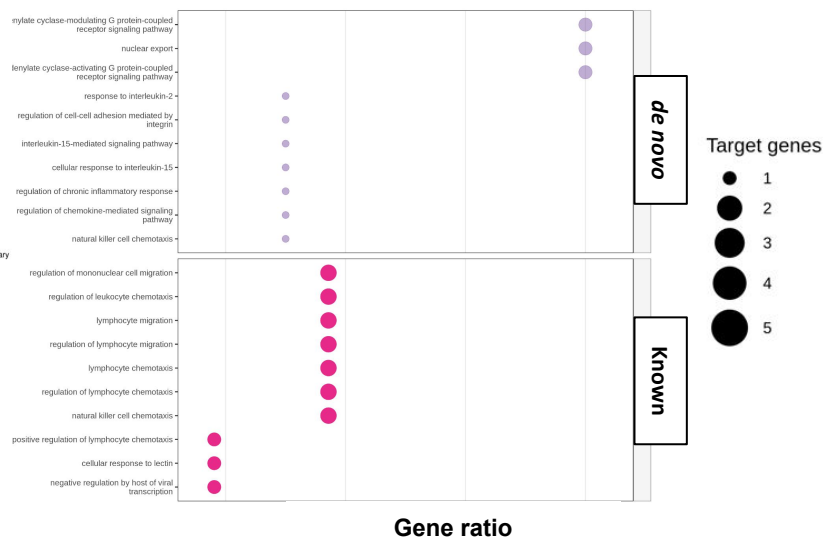
b) Progressive prioritization of miRNA/mRNA interaction



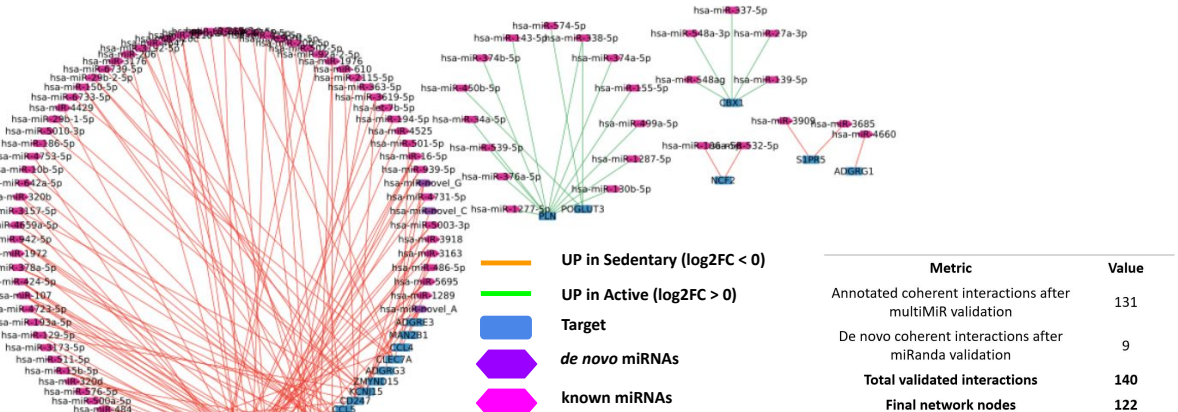
c) Coherent miRNA-mRNA target pairs supported by miRanda or MultiMiR



d) Filtered GO enrichment of coherent predicted target genes



e) Integrated miRNA-mRNA regulatory network from tissue-driven confounding validated genes



f) Integrated miRNA-mRNA regulatory network from sarcopenia-associated validated genes

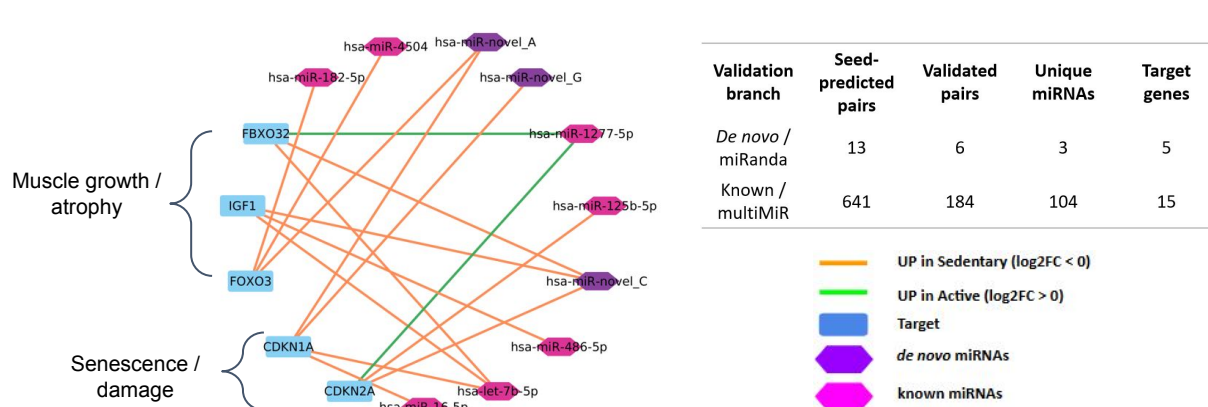
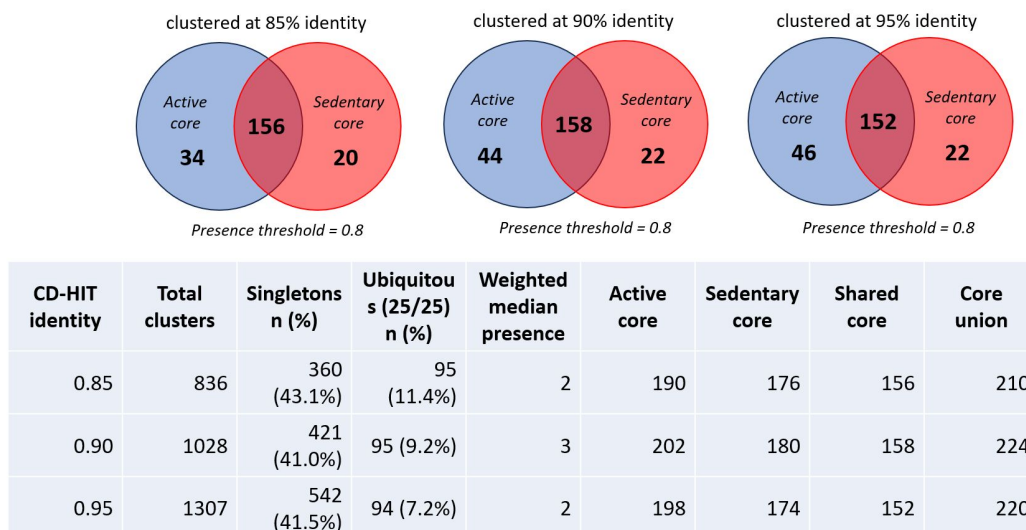


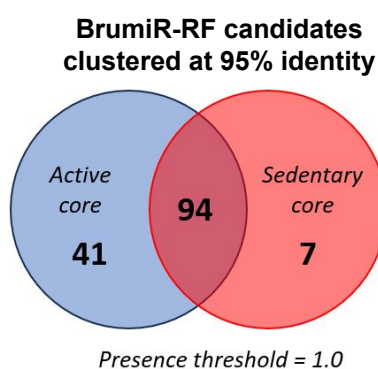
Figure 5. Integrated miRNA–mRNA regulatory landscape combining annotated and *de novo* miRNAs identified by nf-Sarcope. (a) Conceptual framework for coherent miRNA–mRNA integration. Inverse expression relationships were used to define biologically coherent interactions, where miRNA upregulation is associated with target mRNA repression and vice versa. (b) Progressive prioritization of miRNA–mRNA interactions identified by nf-Sarcope. Sequential reduction of candidate interactions through the target prediction workflow for BrumiR-derived *de novo* miRNAs and miRDeep2 annotated miRNAs. Bars indicate the number of differentially expressed miRNAs evaluated, miRNAs with at least one predicted seed match, target genes containing canonical seed matches, raw seed-matched miRNA–mRNA pairs, coherent interactions exhibiting inverse miRNA–mRNA expression patterns, and independently validated interactions retained after miRanda validation (BrumiR) or multiMiR validation (miRDeep2). The y-axis is displayed on a logarithmic scale. (c) Representative validated coherent miRNA–mRNA interactions. The strongest validated pairs from each discovery branch were ranked according to absolute miRNA log₂ fold change. BrumiR-derived *de novo* interactions were exclusively associated with miRNAs upregulated in sedentary individuals, whereas validated miRDeep2 interactions included both active- and sedentary-associated miRNAs. (d) Gene Ontology Biological Process enrichment analysis performed separately on validated target genes regulated by *de novo* and annotated miRNAs. Enriched biological processes were predominantly associated with immune regulation, leukocyte migration, cytokine signaling, and inflammatory responses, reflecting systemic adaptations associated with physical activity. (e) Integrated regulatory network reconstructed from all validated coherent miRNA–mRNA interactions. The network combines interactions supported by miRanda (*de novo* miRNAs) and multiMiR (annotated miRNAs) following seed-based target prediction and inverse-expression filtering. Edge colors indicate whether the corresponding miRNA was upregulated in the active or sedentary group, whereas node colors distinguish annotated miRNAs, *de novo* miRNAs, and target genes. The final high-confidence network comprises 140 validated interactions connecting 122 nodes. (f) Validated interactions between discovered miRNAs and sarcopenia-associated genes. A subset of validated interactions involving genes implicated in muscle growth, atrophy, extracellular matrix remodeling, and cellular senescence is shown. *De novo* miRNAs recovered by BrumiR target canonical regulators including FOXO3, IGF1, FBXO32, CDKN1A, and CDKN2A, illustrating the biological relevance of newly identified candidates within musculoskeletal regulatory pathways.

a)

BrumiR-RF Candidates

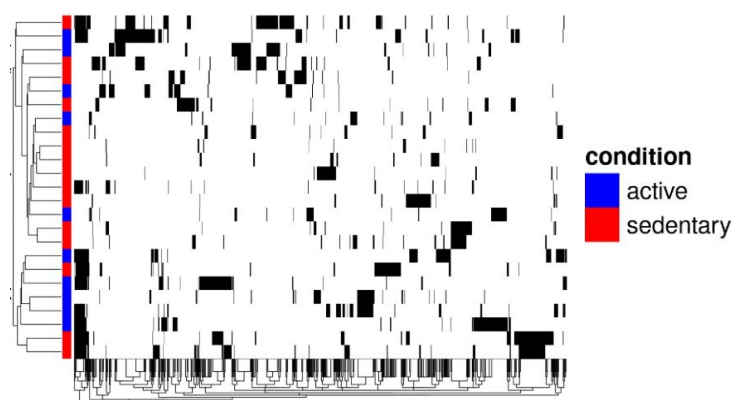


b)



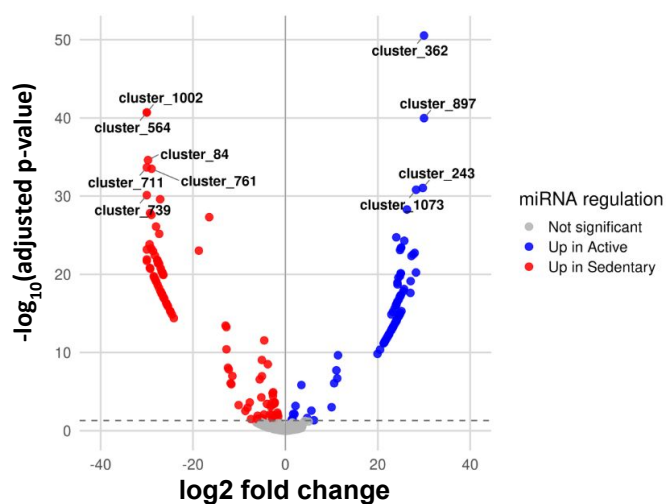
c)

Binary Presence of miRDeep2 novel miRNAs (p0.3)



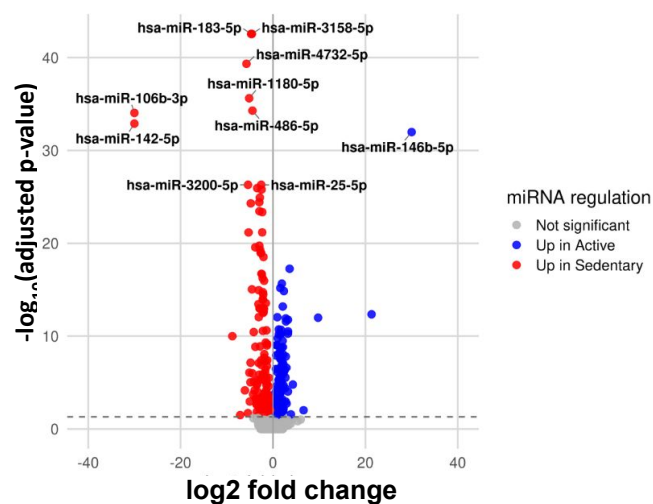
d)

BrumiR-RF miRNA Differential Expression



e)

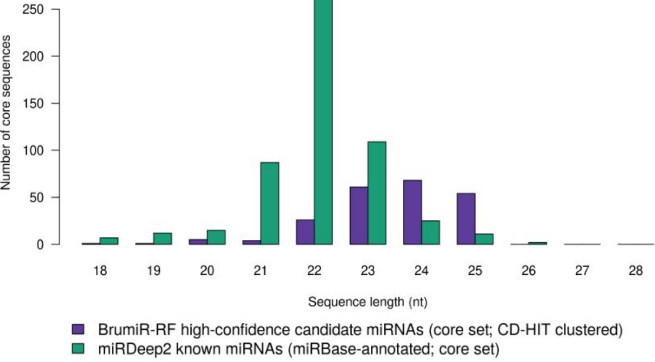
miRDeep2 miRNA Differential Expression



Supplementary Figure 1. Robustness of de novo candidate detection and evaluation of miRDeep2 novel predictions. (a) BrumiR-RF candidate miRNAs clustered at 85%, 90% and 95% sequence identity showing the overlap between active and sedentary core sets (presence ≥ 0.8). Increasing clustering stringency increases the total number of clusters while maintaining stable core set composition across conditions. Summary statistics include total clusters, proportion of singletons, ubiquitous clusters (detected in all samples), and core set sizes. (b) Core miRNA candidates identified by BrumiR-RF at 95% identity using a stringent presence threshold (≥ 1.0), showing reduced set size compared to the ≥ 0.8 threshold while preserving all of shared candidates. (c) Binary presence heatmap of miRDeep2 novel miRNAs detected at a relaxed presence threshold (≥ 0.3). The sparse and heterogeneous distribution of these candidates across samples indicates limited reproducibility compared to core miRNA sets used in downstream analyses. (d–e) Volcano plots showing differential expression results for BrumiR-RF candidates (d) and miRDeep2 known miRNAs highlighting the distribution of effect sizes and statistical significance across methods.

a)

Length distribution (BrumiR vs miRDeep2)



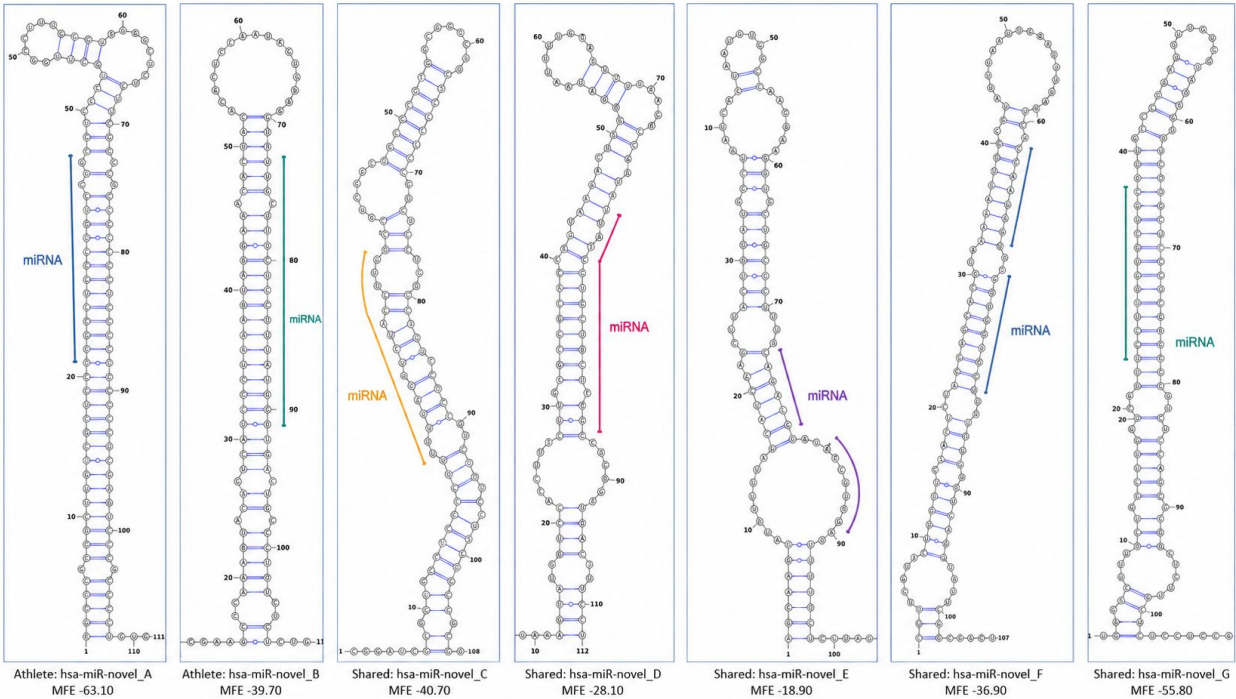
b)

Cluster size of BrumiR passfilter *de novo* miRNA candidates

Final candidate	Original cluster	Group	Cluster size	Classification
hsa-miR-novel_A	cluster_740	active	9	Cluster
hsa-miR-novel_B	cluster_986	active	11	Cluster
hsa-miR-novel_C	cluster_42	shared	27	Cluster
hsa-miR-novel_D	cluster_374	shared	29	Cluster
hsa-miR-novel_E	cluster_413	shared	26	Cluster
hsa-miR-novel_F	cluster_465	shared	32	Cluster
hsa-miR-novel_G	cluster_485	shared	23	Cluster

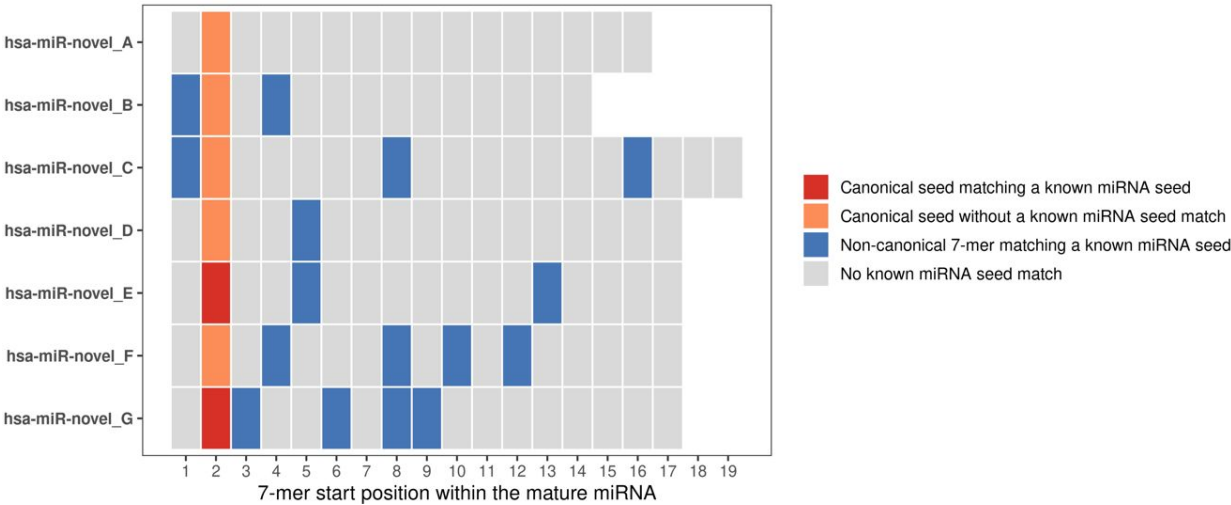
c)

BrumiR2Reference passfilter *de novo* catalog precursors

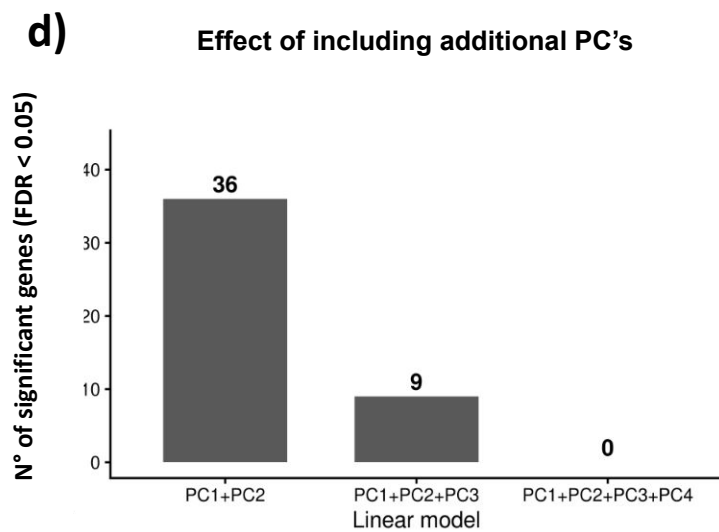
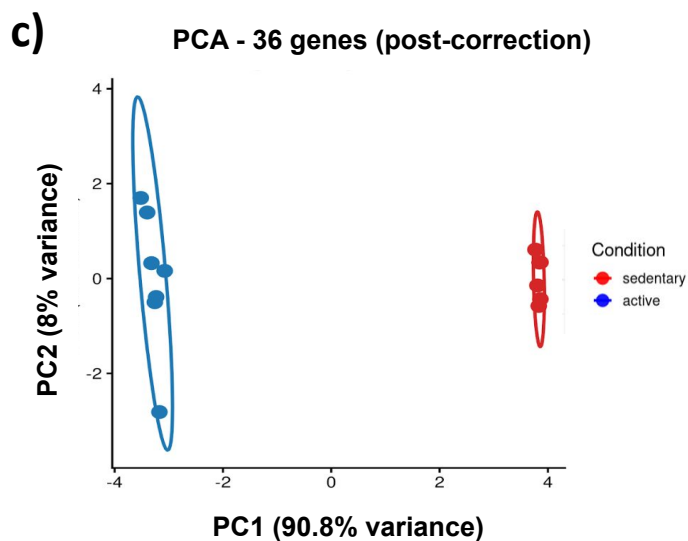
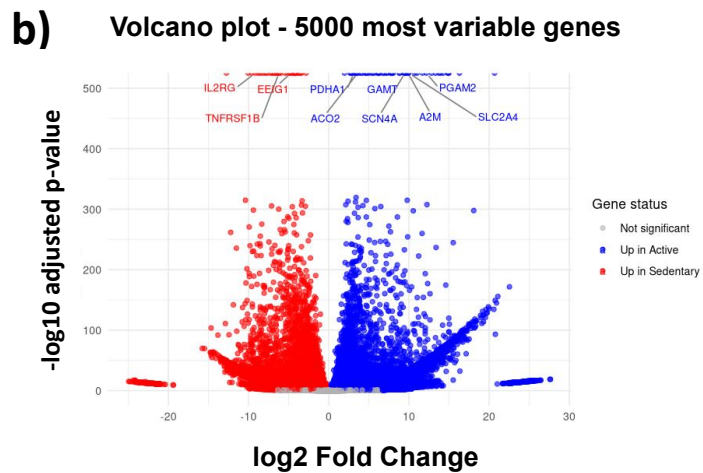
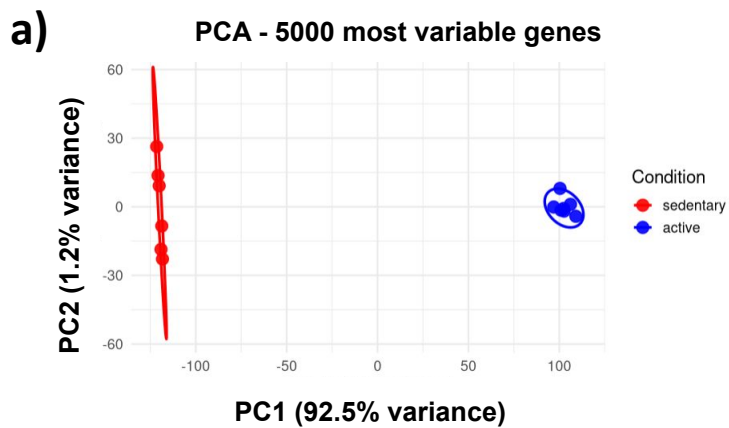


d)

7-mer seed-space analysis of *de novo* miRNA candidates

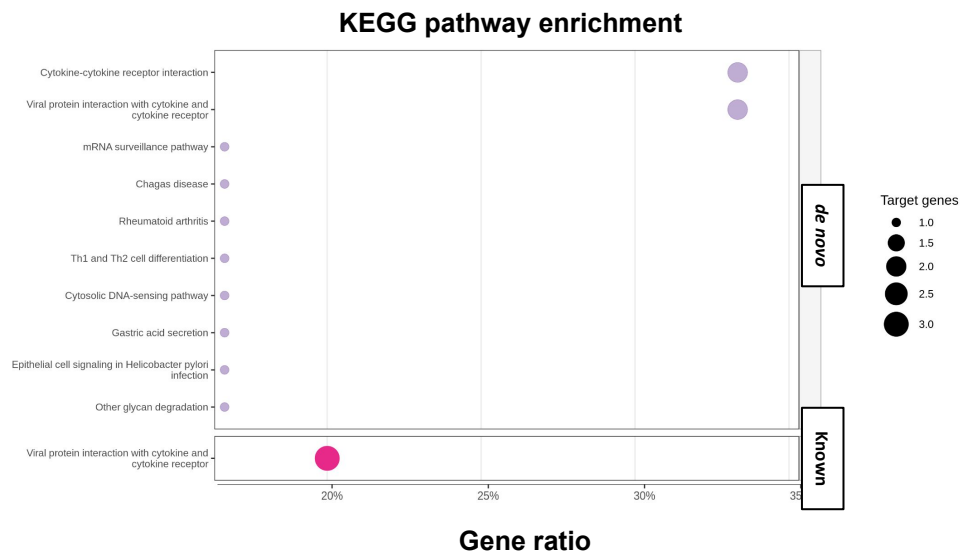


Supplementary Figure 2. Structural and sequence-based characterization of *de novo* miRNA candidates identified by nf-Sarcopipe. (a) Length distribution of miRNA candidates detected by BrumiR-RF and miRDeep2. BrumiR high-confidence *de novo* candidates show a broader size distribution enriched at 22–25 nucleotides, whereas miRDeep2-known miRNAs display a peak at 22 nucleotides, consistent with canonical Dicer processing. (b) Summary of the final high-confidence *de novo* miRNA candidates retained after BrumiR2Reference structural validation. For each candidate, the original CD-HIT cluster, group specificity, cluster size, and candidate classification are shown. Cluster size corresponds to the number of sequences grouped within each representative CD-HIT cluster. (c) Predicted secondary structures of the 7 precursor sequences for the BrumiR2Reference passfilter *de novo* miRNA candidates visualized using VARNA (red dots indicate miRNA site). Mature miRNA regions are highlighted within the stem region of the hairpin structures, and MFE values indicate stable folding compatible with canonical miRNA precursors. (d) 7-mer seed-space analysis of BrumiR passfilter candidates miRNAs. All possible 7-mer subsequences derived from mature miRNA sequences were compared against canonical human miRNA seeds from miRBase. The heatmap shows the presence or absence of matches across all possible 7-mer positions within each sequence. The canonical seed region (nucleotides 2–8) is highlighted and classified as either known or novel.

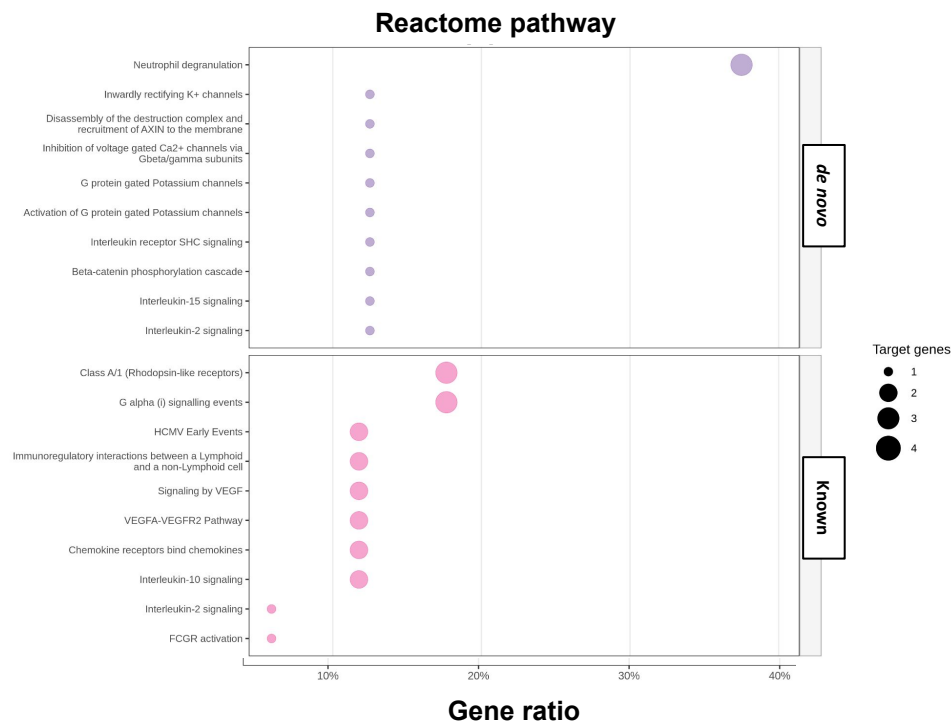


Supplementary Figure 3. Evidence of confounding effects and their correction in RNA-seq analysis. (a) Principal component analysis (PCA) of the 5,000 most variable genes showing strong separation between samples driven by tissue origin, with PC1 explaining 92.5% of the variance. (b) Volcano plot of differential expression analysis without correction, revealing a large number of significant genes consistent with a confounded signal dominated by tissue-specific effects. (c) PCA of the 36 genes retained after correction (expression $\sim$ group + PC1 + PC2), showing separation of samples according to biological condition rather than tissue origin, indicating effective removal of the dominant confounding effect. (d) Sensitivity analysis showing the effect of including additional principal components in the linear model. The number of significantly differentially expressed genes (FDR < 0.05) decreases from 36 (PC1 + PC2) to 9 (PC1 + PC2 + PC3) and 0 (PC1 + PC2 + PC3 + PC4), demonstrating over-adjustment and loss of biological signal when higher-order components are included.

a)

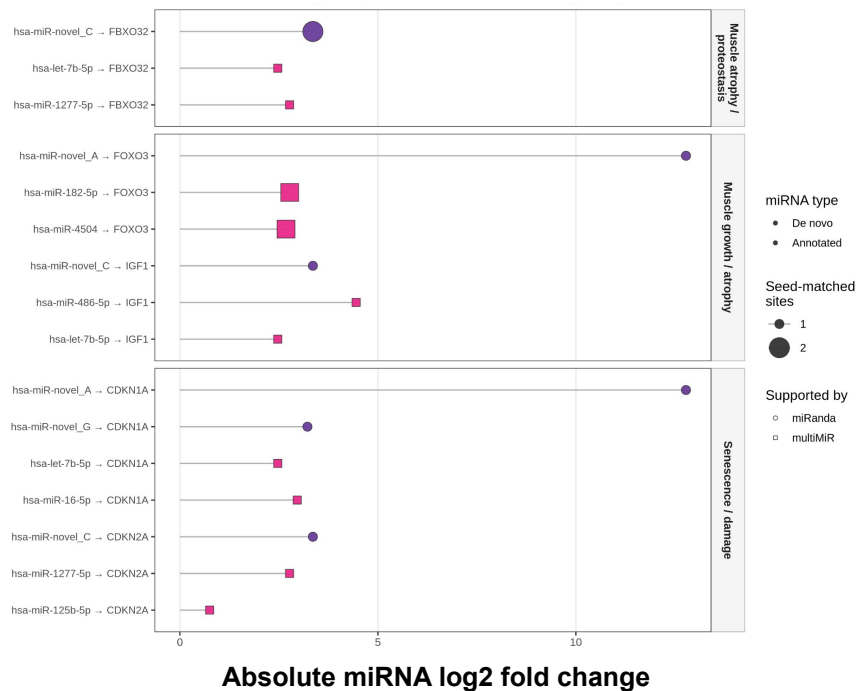


b)



Validated miRNA interaction with sarcopenia-associated genes

c)



Supplementary Figure 4. Extended analysis of predicted miRNA–mRNA interactions and pathway enrichment. (a) KEGG pathway enrichment analysis performed separately on validated target genes regulated by *de novo* and annotated miRNAs. The *de novo* miRNA targets were associated with several immune- and signaling-related pathways that did not remain significant after multiple-testing correction, whereas the annotated miRNA target set showed significant enrichment for the cytokine–cytokine receptor interaction pathway. (b) Reactome pathway enrichment analysis performed separately for validated targets of *de novo* and annotated miRNAs. Although no pathways remained significant after multiple-testing correction, the highest-ranking Reactome terms for *de novo* miRNA targets included interleukin-2/interleukin-15 signaling, G protein-coupled receptor signaling, β -catenin-related processes, and neutrophil degranulation, supporting the biological plausibility of the prioritized candidates. (c) Validation of predicted interactions involving a curated set of musculoskeletal genes associated with muscle growth, atrophy, extracellular matrix remodeling, mitochondrial metabolism, inflammation, and cellular senescence. Only interactions supported by miRanda (*de novo* miRNAs) or multiMiR (annotated miRNAs) are shown. Point size represents the number of canonical seed-matched sites, whereas color and shape distinguish the validation strategy and miRNA discovery branch.

ID	Sex	Exercise	State	Health Status	miRNA Biopsy	Reference	raw reads	Processed reads (fastp + miRNA adapters [small RNA-seq])	Processed reads (fastp + miRNA adapters + splitter)		Brumir-core Short			Brumir-RF		miRDeep2 Short	
									Reads between 18–28 nt (short reads)	>28	UNITIGS	Candidate s	Other sequence	High-confidence miRNAs	miRNAs novel	miRNAs known	
SRR11363960	Female	no	Sedentary 10	Healthy	Plasma	Li et al. 2020	19,022,725	9,845,947	9,084,749	761,198	61,668	615	834	402	1,378	850	
SRR11363961	Female	no	Sedentary 9	Healthy	Plasma	Li et al. 2020	20,590,468	8,501,859	8,136,652	365,207	28,873	484	236	365	1,214	804	
SRR11363962	Female	no	Sedentary 8	Healthy	Plasma	Li et al. 2020	23,587,262	13,814,661	13,071,752	742,909	63,781	691	504	480	1,422	755	
SRR11363963	Female	no	Sedentary 7	Healthy	Plasma	Li et al. 2020	20,654,567	6,128,209	5,577,336	550,873	74,782	547	690	360	1,219	788	
SRR11363964	Female	no	Sedentary 6	Healthy	Plasma	Li et al. 2020	23,241,588	2,871,766	2,258,482	613,284	88,711	348	950	224	907	647	
SRR11363965	Female	no	Sedentary 5	Healthy	Plasma	Li et al. 2020	20,028,716	12,889,765	12,609,379	280,386	33,253	480	310	362	1,193	791	
SRR11363966	Female	no	Sedentary 4	Healthy	Plasma	Li et al. 2020	19,089,860	5,144,366	4,741,002	403,364	49,296	426	516	303	1,099	736	
SRR11363967	Female	Swimmer	Active 10	Healthy	Plasma	Li et al. 2020	29,155,468	8,514,649	7,504,116	1,010,533	97,826	570	1,282	340	1,225	782	
SRR11363968	Female	Swimmer	Active 9	Healthy	Plasma	Li et al. 2020	19,739,731	5,560,122	5,006,241	553,881	46,135	414	842	294	1,092	754	
SRR11363969	Female	Swimmer	Active 8	Healthy	Plasma	Li et al. 2020	24,382,847	10,664,351	10,004,135	660,216	71,921	536	790	363	1,296	822	
SRR11363970	Female	Swimmer	Active 7	Healthy	Plasma	Li et al. 2020	28,448,347	13,723,589	12,829,709	893,880	62,432	587	760	377	1,397	847	
SRR11363971	Female	Swimmer	Active 6	Healthy	Plasma	Li et al. 2020	29,317,758	12,122,149	11,106,501	1,015,648	121,754	708	1,424	409	1,317	793	
SRR11363972	Female	no	Sedentary 3	Healthy	Plasma	Li et al. 2020	20,973,415	3,032,547	2,767,538	265,009	21,396	283	250	216	869	630	
SRR11363973	Female	Swimmer	Active 5	Healthy	Plasma	Li et al. 2020	22,113,129	8,891,970	8,211,132	680,838	57,064	503	666	357	1,201	781	
SRR11363974	Female	Swimmer	Active 4	Healthy	Plasma	Li et al. 2020	22,734,855	9,682,360	8,983,974	698,386	76,987	551	878	360	1,314	805	
SRR11363975	Female	Swimmer	Active 3	Healthy	Plasma	Li et al. 2020	21,908,930	15,070,142	13,153,194	1,916,948	107,962	537	1,492	330	1,209	672	
SRR11363976	Female	Swimmer	Active 2	Healthy	Plasma	Li et al. 2020	25,232,880	10,955,312	10,094,551	860,761	82,524	570	958	357	1,291	715	
SRR11363977	Female	Swimmer	Active 1	Healthy	Plasma	Li et al. 2020	16,552,141	7,801,434	7,232,046	569,388	44,155	462	418	326	1,204	770	
SRR11363978	Female	no	Sedentary 15	Healthy	Plasma	Li et al. 2020	14,839,099	1,752,536	1,553,728	198,808	13,192	221	228	175	767	595	
SRR11363979	Female	no	Sedentary 14	Healthy	Plasma	Li et al. 2020	17,608,864	12,656,961	10,040,941	2,616,020	193,623	713	1,838	401	1,233	677	
SRR11363980	Female	no	Sedentary 13	Healthy	Plasma	Li et al. 2020	20,914,126	6,552,894	6,260,245	292,649	24,286	425	246	332	1,130	761	
SRR11363981	Female	no	Sedentary 12	Healthy	Plasma	Li et al. 2020	21,476,762	9,228,210	8,574,123	654,087	62,937	607	760	395	1,373	830	
SRR11363982	Female	no	Sedentary 11	Healthy	Plasma	Li et al. 2020	19,411,413	8,231,392	7,787,322	444,070	40,941	521	478	369	1,288	837	
SRR11363983	Female	no	Sedentary 2	Healthy	Plasma	Li et al. 2020	11,695,923	11,676,363	11,145,105	531,258	49,776	636	422	430	1,440	850	
SRR11363984	Female	no	Sedentary 1	Healthy	Plasma	Li et al. 2020	17,510,813	11,106,736	10,709,914	396,822	49,164	552	508	386	1,324	826	
Total							530,231,687	226,420,290	208,443,867	17,976,423	1,624,439	12,987	18,280 Unique sequences	8,713	30,402	19,118	
													3,346	3,461	5,868		

Supplementary Table 1. Sample metadata and preprocessing summary for the plasma sRNA-seq dataset used for initial small RNA candidate discovery.

The table summarizes sample metadata and preprocessing statistics for the only publicly available plasma small RNA-seq dataset suitable for candidate miRNA discovery in the context of exercise physiology. The dataset consists of young healthy female participants, including active synchronized swimmers and sedentary individuals, as originally reported by Li et al. (2020). Raw sequencing reads were processed using the nf-Sarcopipe preprocessing module, including FastQC quality assessment, adapter trimming and filtering with *fastp*, and miRNA and small RNA candidate annotation using BrumiR-core and miRDeep2. For each sample, the table reports sample identifiers, participant characteristics, tissue source (plasma), sequencing depth before and after trimming, and the number of small RNA candidates detected by each discovery method.

Sample	Raw reads	Reads retained	Retained (%)	Lost (%)	Low-quality reads	Too short reads	Adapter-trimmed reads	Adapter	Overrepresented seq. (%)	Duplication (%)
SRR111363960	19,022,725	9,845,947	51.76	48.24	101	9,176,677	9,177,871	PASS	33.19	97.37
SRR111363961	20,590,468	8,501,859	41.29	58.71	114	12,088,495	12,093,331	PASS	38.39	99.01
SRR111363962	23,587,262	13,814,661	58.57	41.43	150	9,772,451	9,773,629	PASS	25.85	98.11
SRR111363963	20,654,567	6,128,209	29.67	70.33	189	14,526,169	14,525,810	PASS	46.73	96.75
SRR111363964	23,241,588	2,871,766	12.36	87.64	109	20,369,713	20,360,471	PASS	56.58	97.21
SRR111363965	20,028,716	12,889,765	64.36	35.64	134	7,138,817	7,146,667	PASS	22.32	98.83
SRR111363966	19,089,860	5,144,366	26.95	73.05	168	13,945,326	13,948,065	PASS	46.34	98.04
SRR111363967	29,155,468	8,514,649	29.20	70.80	192	20,640,627	20,606,544	PASS	45.34	97.32
SRR111363968	19,739,731	5,560,122	28.17	71.83	116	14,179,493	14,181,180	PASS	46.98	97.92
SRR111363969	24,382,847	10,664,351	43.74	56.26	244	13,718,252	13,708,970	PASS	34.98	97.90
SRR111363970	28,448,347	13,723,589	48.24	51.76	281	14,724,477	14,729,371	PASS	32.15	98.41
SRR111363971	29,317,758	12,122,149	41.35	58.65	141	17,195,468	17,186,558	PASS	37.75	97.26
SRR111363972	20,973,415	3,032,547	14.46	85.54	127	17,940,741	17,953,953	PASS	51.49	99.17
SRR111363973	22,113,129	8,891,970	40.21	59.79	78	13,221,081	13,225,909	PASS	37.47	98.18
SRR111363974	22,734,855	9,682,360	42.59	57.41	104	13,052,391	13,055,588	PASS	35.54	97.47
SRR111363975	21,908,930	15,070,142	68.79	31.21	304	6,838,484	6,819,095	PASS	16.17	95.80
SRR111363976	25,232,880	10,955,312	43.42	56.58	258	14,277,310	14,261,068	PASS	35.66	97.68
SRR111363977	16,552,141	7,801,434	47.13	52.87	170	8,750,537	8,748,388	PASS	32.09	97.96
SRR111363978	14,839,099	1,752,536	11.81	88.19	86	13,086,477	13,088,705	PASS	57.70	99.15
SRR111363979	17,608,864	12,656,961	71.88	28.12	131	4,951,772	4,950,477	PASS	18.00	91.60
SRR111363980	20,914,126	6,552,894	31.33	68.67	242	14,360,990	14,368,331	PASS	40.45	99.11
SRR111363981	21,476,762	9,228,210	42.97	57.03	330	12,248,222	12,247,111	PASS	35.67	97.72
SRR111363982	19,411,413	8,231,392	42.40	57.60	97	11,179,924	11,182,153	PASS	38.44	98.40
SRR111363983	11,695,923	11,676,363	99.83	0.17	139	19,421	0	PASS	9.74	97.30
SRR111363984	17,510,813	11,106,736	63.43	36.57	283	6,403,794	6,409,746	PASS	23.03	97.69

Supplementary Table 2. Summary of raw library quality and fastp filtering statistics across the 25 small RNA-seq libraries. Summary of sequencing quality and preprocessing metrics obtained from FastQC and *fastp* reports for all plasma small RNA-seq libraries. For each sample, the table shows the number of raw reads, reads retained after adapter trimming, retention and loss percentages, reads discarded because of low quality or insufficient length, reads subjected to adapter trimming, FastQC adapter-content status, percentage of the most abundant overrepresented sequence, and estimated sequence duplication level. *fastp* statistics were extracted from the corresponding JSON reports generated using the same preprocessing parameters applied in the nf-Sarcopipe workflow.

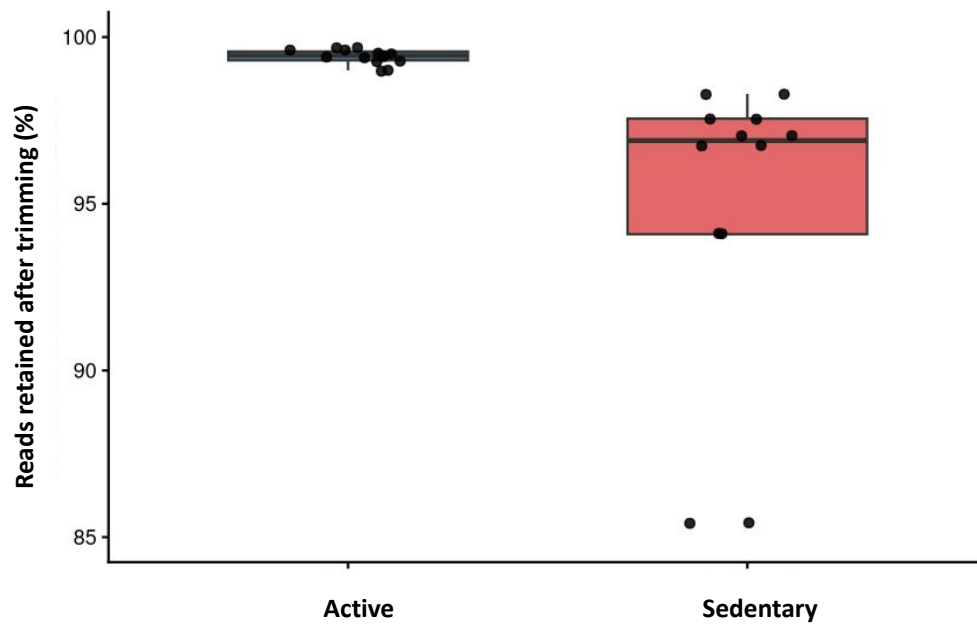
Candidate	Group	Chromosome	Start	End	Mature sequence (5'→3')	Mature length (nt)	Precursor length (nt)	BruniR2Reference MFE (kcal/mol)	Known exact match	De novo exact
cluster_270	athlete	17	59137755	59137868	CAGCAGAGACAAUUAUUAUAGGGU	24	113	-18.90	0	1
cluster_396	athlete	X	50013235	50013347	UACCAUUGCAUUAUCGGAGUUGU	23	112	-34.60	1	0
cluster_740	athlete	22	40956683	40956794	GAGCCACCAUGUCCGACCUCAA	22	111	-63.10	0	1
cluster_962	athlete	11	122152203	122152311	CACAAGUUCGGAUUCUACGGG	20	108	-31.70	0	1
cluster_986	athlete	17	58331214	58331324	AGUCUUUUCUACUUUAUUGGG	20	110	-39.70	0	1
cluster_421	sedentary	14	101027080	101027192	AACACACUGGUUAACCUUUUU	23	112	-36.30	0	1
cluster_481	sedentary	14	101047878	101047990	UUAAUAUCGGACACCAUUGUUU	23	112	-31.00	0	1
cluster_2	shared	3	49020131	49020240	ACUCAACGGGAGUGAUCGUGUUAU	25	109	-43.50	0	1
cluster_42	shared	7	129770427	129770535	CAGUGUGAGUUCUACCAUUGCCAAA	25	108	-40.70	0	1
cluster_70	shared	X	151958557	151958671	CUAAACGGAAACACUAGUGACUUGA	25	114	-44.50	0	1
cluster_83	shared	14	101055392	101055504	AGAGAGGCGUGGCCGUGAUGAAUUG	25	112	-38.50	0	1
cluster_106	shared	1	40754346	40754452	CUGUAAACAUCUUUGACUGGGAAGCU	25	106	-54.40	0	1
cluster_164	shared	9	83969750	83969861	AACAACAAAAUCACUAGUCUCCA	24	111	-36.60	1	0
cluster_167	shared	14	100109644	100109757	UCUCACACAGAAAAUCGCACCCGUC	24	113	-50.30	0	1
cluster_244	shared	1	198859037	198859150	ACUACCCGACAGCGUUGAUGUUC	24	113	-28.00	0	1
cluster_374	shared	13	50048955	50049067	ACGCCAAUAUUUACGUGUGCUA	23	112	-28.10	0	1
cluster_386	shared	11	57641181	57641293	CAGUGCAAUGUUAAAAAGGCAUU	23	112	-47.30	0	1
cluster_413	shared	X	134169310	134169417	UACAGAUGGAUACCGUGCAUUU	23	107	-18.90	0	1
cluster_430	shared	9	124692458	124692571	AACCACUGACCGUUGACUGUACC	23	113	-46.30	0	1
cluster_431	shared	9	109046210	109046321	UGCAACUUAAGUAAUUGUCCAUAU	23	111	-23.60	0	1
cluster_465	shared	3	160404733	160404845	CACCAUAUUACUGUGUGCUUUU	23	112	-36.90	0	1
cluster_485	shared	7	32732980	32733092	CAGUGCCUGAGGGAGUAAGAGCC	23	112	-55.80	0	1
cluster_554	shared	8	134800510	134800622	UGAAGUAAAAUCAUCCACCUCCAG	23	112	-29.30	0	1
cluster_612	shared	2	135665375	135665486	UCACAGUAAACGGGUCUCUUUU	22	111	-39.50	0	1
cluster_650	shared	7	100093981	100094094	UAUCUGCACUGUCAGCACUUUA	22	113	-48.60	0	1
cluster_652	shared	16	1734966	1735077	GUGCACGGCACUGGGGACACGU	22	111	-50.40	0	1
cluster_819	shared	17	12081888	12081998	UGCGGGCGUAGGGCUAACAGC	21	110	-41.10	0	1

Supplementary Table 3. Unique structurally supported BrumiR-RF core candidate miRNAs identified after BrumiR2Reference analysis. The table summarizes 27 unique BrumiR-RF core candidates retained after structural validation. For candidates with multiple predicted genomic loci, the locus associated with the most stable precursor, defined as the lowest minimum free energy (MFE), was retained. Reported coordinates correspond to the predicted genomic location of the precursor hairpin. The mature sequence (5'→3') represents the predicted mature miRNA, whereas precursor length indicates the length of the predicted hairpin sequence. **Known exact match = 1** identifies candidates with an exact mature-sequence match to miRBase and/or MirGeneDB, whereas **de novo exact = 1** identifies structurally supported candidates lacking an exact reference match and retained as high-confidence *de novo* miRNA candidates.

Candidate	Cluster	Length (nt)	Mature sequence (5'→3')	Canonical seed (2-8)	log2FC	Adjusted p value	Regulation	Exact miRBase seed match	Matching miRBase family	Total 7-mers	Matching known 7-mers	Matching 7-mer(s)	Families matched across all positions
hsa-miR-novel_A	cluster_740	22	GAGCCACCAUG UCCGACCUCAA	AGCCACC	-12.780	5.60×10^{-14}	Up in sedentary	No	–	16	0	–	–
hsa-miR-novel_B	cluster_986	20	AGUGCUUUCU ACUUUAUGGG	GUGCUUU	-6.138	3.15×10^{-2}	Up in sedentary	No	–	14	2	AGUGCUU; GCUUUCU	miR-3160-5p; miR-520f-3p
hsa-miR-novel_C	cluster_42	25	CAGUGUGAGU UCUACCAUUGC CAAA	AGUGUGA	-3.359	8.23×10^{-3}	Up in sedentary	No	–	19	3	AGUUCUA; CAGUGUG; CAUUGCC	miR-3168; miR-4302; miR-4317
hsa-miR-novel_D	cluster_374	23	ACGCCAAUUAU UAGGUGGUGC UA	CGCCAAU	-2.339	3.40×10^{-4}	Up in sedentary	No	–	17	1	CAAUAAU	miR-16-2-3p; miR-195-3p
hsa-miR-novel_E	cluster_413	23	UACAGAUUGA UACCGUGCAAU UU	ACAGAUG	-2.697	3.55×10^{-5}	Up in sedentary	Yes	miR-5683	17	3	ACAGAUG; CCGUGCA; GAUGGAU	miR-1973; miR-3134; miR-5683
hsa-miR-novel_F	cluster_465	23	CACCAUAUUA CUGUGGUGCU UU	ACCAUAU	-3.810	3.14×10^{-9}	Up in sedentary	No	–	17	4	AUUACUG; CAAUAUU; CUGUGCU; UACUGUG	miR-12132; miR-16-2-3p; miR-195-3p; miR-4693-5p; miR-5008-3p; miR-6737-3p; miR-7157-3p
hsa-miR-novel_G	cluster_485	23	CAGUGCCUGAG GGAGUAAGAG CC	AGUGCCU	-3.221	7.64×10^{-4}	Up in sedentary	Yes	miR-33b-3p; miR-515-3p; miR-519e-3p	17	5	AGUGCCU; CCUGAGG; GAGGGAG; GUGCCUG; UGAGGGA	miR-1271-3p; miR-33b-3p; miR-4510; miR-515-3p; miR-519e-3p; miR-544b; miR-550a-3-5p; miR-550a-5p; miR-6127; miR-6129; miR-6130; miR-6133; miR-6834-5p

Supplementary Table 4. *Characteristics of the final high-confidence de novo miRNA candidates identified by nf-Sarcope*. Final BrumiR-derived de novo miRNA candidates retained after structural prioritization and differential expression analysis. For each candidate, the table reports the provisional miRNA identifier, original cluster, mature sequence length, mature sequence, canonical seed sequence (positions 2–8), differential expression statistics (log2 fold change, adjusted *p* value, and regulation), canonical seed match to known human miRNAs in miRBase, corresponding matched miRNA family (when applicable), and the results of the comprehensive 7-mer seed-space analysis, including the total number of possible 7-mers, the number of known matching 7-mers, the matched 7-mer sequences, and the corresponding miRNA families identified across all positions of the mature sequence.

ID	Sex	Exercise	State	Health status	Biopsy mRNA	Raw reads (paired-end)	Reads after trimming (fastp)	Reads removed during trimming	Reference
SRR13442895_1.fastq	Female	Skater 1 - before exercise	Active	Healthy	Whole blood	29488902	29306765	182137	Glotov et al. 2022
SRR13442895_2.fastq	Female	Skater 1 - before exercise	Active	Healthy	Whole blood	29488902	29306765	182137	Glotov et al. 2022
SRR13442897_1.fastq	Female	Skater 2 - before exercise	Active	Healthy	Whole blood	29895997	29679635	216362	Glotov et al. 2022
SRR13442897_2.fastq	Female	Skater 2 - before exercise	Active	Healthy	Whole blood	29895997	29679635	216362	Glotov et al. 2022
SRR13442899_1.fastq	Female	Skater 3 - before exercise	Active	Healthy	Whole blood	38315307	38102268	213039	Glotov et al. 2022
SRR13442899_2.fastq	Female	Skater 3 - before exercise	Active	Healthy	Whole blood	38315307	38102268	213039	Glotov et al. 2022
SRR13442901_1.fastq	Female	Skater 4 - before exercise	Active	Healthy	Whole blood	40888862	40719776	169086	Glotov et al. 2022
SRR13442901_2.fastq	Female	Skater 4 - before exercise	Active	Healthy	Whole blood	40888862	40719776	169086	Glotov et al. 2022
SRR13442903_1.fastq	Female	Skater 5 - before exercise	Active	Healthy	Whole blood	38266305	38077449	188856	Glotov et al. 2022
SRR13442903_2.fastq	Female	Skater 5 - before exercise	Active	Healthy	Whole blood	38266305	38077449	188856	Glotov et al. 2022
SRR13442905_1.fastq	Female	Skater 6 - before exercise	Active	Healthy	Whole blood	42600787	42459876	140911	Glotov et al. 2022
SRR13442905_2.fastq	Female	Skater 6 - before exercise	Active	Healthy	Whole blood	42600787	42459876	140911	Glotov et al. 2022
SRR13442907_1.fastq	Female	Skater 7 - before exercise	Active	Healthy	Whole blood	12447565	12323017	124548	Glotov et al. 2022
SRR13442907_2.fastq	Female	Skater 7 - before exercise	Active	Healthy	Whole blood	12447565	12323017	124548	Glotov et al. 2022
SRR1424731_1.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	14977661	14611225	366436	Lindholm et al. 2014
SRR1424731_2.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	14977661	14611225	366436	Lindholm et al. 2014
SRR1424738_1.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	14174465	13336400	838065	Lindholm et al. 2014
SRR1424738_2.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	14174465	13336400	838065	Lindholm et al. 2014
SRR1424739_1.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	26183362	25737390	445972	Lindholm et al. 2014
SRR1424739_2.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	26183362	25737390	445972	Lindholm et al. 2014
SRR1424741_1.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	14288905	13868580	420325	Lindholm et al. 2014
SRR1424741_2.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	14288905	13868580	420325	Lindholm et al. 2014
SRR1424745_1.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	18265631	15605785	2659846	Lindholm et al. 2014
SRR1424745_2.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	18265631	15605785	2659846	Lindholm et al. 2014
SRR1424754_1.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	31172495	30154181	1018314	Lindholm et al. 2014
SRR1424754_2.fastq	Female	no	Sedentary	Healthy	Vastus lateralis	31172495	30154181	1018314	Lindholm et al. 2014



Supplementary Table 5. Metadata and sequencing quality metrics for the RNA-seq datasets used in the integrative analysis. The table summarizes sample metadata and sequencing quality metrics for RNA-seq datasets retrieved from the Gene Expression Omnibus (GEO) using GEOquery (GEPREP) and processed with the nf-core/rnaseq pipeline using the GRCh38 human reference genome. For each sample, the table reports sample identifiers (SRA accession), sex, exercise condition, activity status, health status, tissue source (whole blood or vastus lateralis muscle biopsy), and sequencing depth before and after adapter trimming using fastp for paired-end reads. The difference column indicates the number of reads removed during trimming. The datasets include RNA-seq samples from healthy young active female skaters prior to exercise and healthy sedentary female individuals, derived from previously published studies indicated in the reference column. The lower panel shows the percentage of reads retained after trimming, highlighting consistently high sequencing quality across samples.

Gene	Functional category	Biological role	Key reference
AKT1	Anabolism	Central node in protein synthesis pathway	Glass 2005
CDKN1A	Senescence	Cell cycle inhibition (p21)	Kim et al. 2025
CDKN2A	Senescence	Cell cycle arrest (p16)	Sousa-Victor et al. 2014
COL1A1	Fibrosis / ECM	Collagen deposition, fibrosis	Kim et al. 2025
COL3A1	Fibrosis / ECM	Extracellular matrix remodeling	Stearns-Reider et al. 2017
COX7A1	Mitochondrial metabolism	Electron transport chain (Complex IV)	García-Poyatos et al. 2024
FBXO32	Atrophy / proteolysis	Ubiquitin ligase (Atrogin-1), muscle protein degradation	Gomes et al. 2001
FOXO3A	Atrophy regulation	Transcription factor activating atrophy genes	Bogućka et al. 2025
IGF1	Anabolism	Supports muscle growth signaling	Ruan et al. 2005
IL1B	Inflammation	Cytokine driving chronic inflammation	Schaap et al. 2006
IL6	Inflammation	Pro-inflammatory cytokine (inflammaging)	Visser et al. 2002
MYOD1	Regeneration	Myogenic differentiation regulator	Raue et al. 2006
MYOG	Regeneration	Late myogenesis marker	Raue et al. 2006
PPARGC1A	Mitochondrial metabolism	Master regulator of mitochondrial biogenesis	Wenz et al. 2009
TNF	Inflammation	Pro-inflammatory signaling, catabolic effects	Visser et al. 2002
TRIM63	Atrophy / proteolysis	Ubiquitin ligase (MuRF1), sarcomere breakdown	Bodine et al. 2001

Supplementary Table 6. *Functional classification of sarcopenia-associated genes.* Genes were grouped into biologically relevant categories associated with sarcopenia, including muscle atrophy, mitochondrial dysfunction, inflammation, fibrosis, regeneration, neuromuscular junction integrity, and cellular senescence. This classification is based on well-established molecular mechanisms reported in the literature, where aging skeletal muscle is characterized by increased proteolysis (e.g., FBXO32, TRIM63), chronic inflammation (e.g., IL6, TNF), and fibrosis (e.g., COL1A1, TGFB1), together with reduced mitochondrial function (e.g., PPARGC1A) and impaired regenerative capacity (e.g., PAX7, MYOD1). References supporting each functional category are provided in the corresponding column.

[illegible]

Supplementary Table 7 (page 1 & 2). *Curated list of microRNAs associated with exercise responses in humans.* The table summarizes microRNAs (miRNAs) previously reported to be associated with exercise responses in young and older adults, including both cardiorespiratory endurance and strength exercise paradigms. For each miRNA, the table reports the study population or condition, functional annotations from miRBase, functional associations described in the literature, exercise modality, intervention type (acute or chronic), direction of expression change (↑ upregulated, ↓ downregulated), sample source (e.g., plasma, serum, exosomes, skeletal muscle, or whole blood), and the original reference reporting the observation. This curated catalog provides a literature-based reference framework to contextualize exercise-associated miRNAs and to facilitate comparison with candidate miRNAs identified using the nf-Sarcope workflow.