Integrating Public Transcriptomic Re-analysis and Molecular Docking to Explore Androgen Receptor Signaling in Skeletal Muscle and Candidate Ligand Binding

Author: [Anonymized for double-blind review]

Date: February 04, 2026

Keywords: androgen receptor, skeletal muscle, GEO, gene expression, pathway enrichment, Enrichr, molecular docking, AutoDock Vina, ostarine

Abstract

Androgen receptor (AR) signaling is a major regulator of skeletal muscle physiology. However, understanding how AR deficiency affects downstream gene programs can be difficult without a workflow that provides a link between transcriptomics and a mechanistic hypothesis. In this standalone research note, publicly available Gene Expression Omnibus (GEO) data (accession: GSE216372) are analyzed using a false discovery rate (FDR) cutoff (adjusted p-value < 0.05) and a gene effect size cutoff (log2 fold change ≥ 0.5). The up-regulated and down-regulated gene sets will then be evaluated using a pathway analysis (Enrichr). To generate a structural hypothesis using a reproducible workflow, docking is performed using OpenBabel and AutoDock Vina with ostarine as the ligand of interest and an androgen receptor structure. The docking box will be centered at 7.180, 27.291, 10.936 Å and will be 24 Å in all dimensions. This workflow generates reproducible results and figures for a manuscript, and provides a clear process that can be replicated by others. This work will be presented as a research note on an interdisciplinary problem, and will include a clear outline for future studies.

Introduction

Skeletal muscle responds to endocrine signaling by changing gene expression, metabolic profiles, and contractile function. The androgen receptor (AR) is a member of the nuclear hormone receptor superfamily and is involved in the anabolic actions of androgens in skeletal muscle. Altering AR signaling, such as by genetic knockout, can create a transcriptomic profile from which information about the altered phenotype might be gleaned, although such information is not necessarily immediate and is likely dependent upon gene set trends and mechanistic hypotheses.

This study has two aims: (1) to offer an understandable and replicable re-analysis of an existing publicly available transcriptomic dataset for AR deficiency in skeletal muscle tissue, and (2) to offer these findings with an understandable and replicable command-line-based process for molecular docking, which can offer structural-based hypotheses for the interaction of the AR and its ligands.

Methods

Data Source and Preprocessing. The gene expression data used in this study were sourced from the Gene Expression Omnibus database, a public functional genomics database that stores microarray, next-generation sequencing, and other functional genomics data (Edgar et al., 2002). The ‘all genes’ table was used was used to obtain the data on differential expression between the androgen receptor loss state and the control state. Genes were filtered based on the following

conditions: adjusted p-value < 0.05 and |log2 fold change| >= 0.5. The top 10 genes that were up-regulated and down-regulated were summarized.

Visualization and Enrichment. The data on gene expression was visualized using a volcano plot, which showed the relationship between the effect size and the statistical significance of the data. A heatmap of the top genes was used to show the data on gene expression. The data on gene expression was further enriched using the Enrichr gene set enrichment tool, a web-based platform that performs gene set enrichment analysis and utilizes a wide variety of functional gene sets for this purpose (Chen et al., 2013; Kuleshov et al., 2016). The data on the genes that were up-regulated and down-regulated were enriched separately and visualized using a bar plot of the most significant terms.

Software Environment and Reproducibility. The study used a Python/Jupyter notebook environment with Anaconda as the distribution. Dependencies were installed on macOS using Homebrew.

Preparation of Ligand and Receptor for Docking. The ligand of interest, ostarine, was downloaded from an online database as an SDF file. It was converted to a PDB file using Open Babel (O'Boyle et al., 2011). The PDB file was then converted to a PDBQT file, and partial charges were added using Open Babel with the Gasteiger algorithm. The protein structure of androgen receptor was obtained from the Protein Data Bank (PDB), and it was converted to a PDBQT file.

In addition to ostarine, testosterone and dihydrotestosterone (DHT) were docked as reference ligands. Docking was performed against the androgen receptor (AR) and, for specificity comparison, estrogen receptor alpha (ERα) and glucocorticoid receptor (GR) using the same docking protocol and search-space dimensions. Best-score (top-ranked pose) affinities were summarized for cross-target comparison.

Docking Procedure. AutoDock Vina (Trott and Olson, 2010; Eberhardt et al., 2021) was used as the docking tool. The exhaustiveness of the search algorithm was set to 16, and a 24 Å cube with its center at the centroid of the protein (center_x = 7.180, center_y = 27.291, center_z = 10.936 Å), defined by the ATOM and HETATM records, was used.

Workflow Timeline (Audit Trail)

1. Developed a directory structure for the project, including transcriptomics results (tables and figures), as well as docking results and inputs/outputs.
2. Transcriptomics: downloaded the all genes differential expression table for the GEO accession GSE216372.
3. Transcriptomics: filtered the results for the differentially expressed genes using an adjusted p-value < 0.05 and |log2 fold change| >= 0.5, and exported the results for the unregulated and down-regulated genes.
4. Transcriptomics: generated figures (volcano plot and heatmap for the top genes), as well as enrichment results (Enrichr bar plots).

5. Docking preparation: installed Homebrew and Xcode Command Line Tools, and Open Babel using Homebrew.

6. Docking preparation: converted the ligand SDF file to PDB using Open Babel, and generated the ligand PDBQT file.

7. Docking preparation: generated the receptor PDBQT file for the androgen receptor structure.

8. Docking execution: computed the centroid for the receptor, ran AutoDock Vina using a 24 Å box and an exhaustiveness of 16, and exported the results for the ligand.

9. Docking summary: computed the best mode affinity, and generated a docking table and figure for inclusion in the manuscript.

10. Final packaging: figures and tables are packaged, and the manuscript is drafted, including an anonymized version for SPARC submission.

Results

The differentially expressed genes were further filtered with the cut-off criterion of adjusted $p < 0.05$ and $|\log2FC| \geq 0.5$, which identified 330 significant expression changes in the ARKO compared to the WT muscle, consisting of 131 upregulated and 199 downregulated genes. The most upregulated genes were Dnajc5g (log2FC = 5.93, padj = 2.75E-4), Abcc2 (log2FC = 5.52, padj = 3.49E-3), and Fgl1 (log2FC = 5.16, padj = 3.81E-3), and the most downregulated were Dnase1 (log2FC = -6.58, padj = 1.40E-5), Gm12240 (log2FC = -6.04, padj = 2.68E-3), and Fscn2 (log2FC = -4.30, padj = 2.29E-12). The KEGG pathway enrichment analysis with the Enrichr bioinformatics tool, with the KEGG 2021 database for the human genome, revealed the most significant pathways to be folate biosynthesis ($p = 0.0074$; combined score = 82.83),

followed by ABC transporters ($p = 0.0213$; combined score = 36.27), with other significant associations including neomycin/kanamycin/gentamicin biosynthesis ($p = 0.0248$; combined score = 185.84), prolactin signaling ($p = 0.0480$; combined score = 18.08), and cytokine–cytokine receptor interaction ($p = 0.0607$; combined score = 7.87). Docking of ostarine into the androgen-receptor ligand-binding domain produced a top-ranked pose with a predicted affinity of −10.1 kcal/mol (AutoDock Vina score). Comparative docking suggested weaker predicted affinities for ERα (−9.0 kcal/mol) and GR (−9.4 kcal/mol) than for AR under the same fixed search space.

Discussion

This paper proposes an end-to-end workflow where the connection between the androgen receptor (AR) deficiency transcriptomics signature and the pathway analysis and docking methodologies is established with the possibility of deriving hypotheses. The enrichment part of the analysis provides a brief overview of the biological processes most affected, while the docking part offers a transparent framework for determining the likelihood of a proposed ligand binding with the receptor binding pocket.

The paper’s limitations should be considered. The enrichment analysis results are based on correlation and database annotation accuracy. The differential expression analysis results depend on the chosen criteria and initial processing pipeline. The docking results are not binding free energies and do not directly imply biological activity. However, as a research note, the paper is significant for its clarity and transparency and for the integrative use of several methodologies.

Future research directions include performing sensitivity analysis (e.g., varying criteria and/or gene set library), improving receptor preparation, and validating trajectories (e.g., triangulation with literature, experimental data, and comparisons with known AR ligands).

Data and Code Availability

All analyses were performed using publicly available data (NCBI GEO: GSE216372). To preserve the integrity of double-blind review, code and intermediate files are available upon request after editorial decisions.

Acknowledgments

This project was completed independently as a research and methods exercise.

References

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Research 28, 235–242.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research 30, 207–210.

Eberhardt, J., Santos-Martins, D., Tillack, A.F., and Forli, S. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. Journal of Chemical Information and Modeling 61, 3891–3898.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research 44, W90–W97.

O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., and Hutchison, G.R. (2011). Open Babel: An open chemical toolbox. Journal of Cheminformatics 3, 33.

Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry 31, 455–461.

Figures

Figure 1. Workflow schematic for transcriptomic differential-expression analysis, pathway enrichment, and docking.

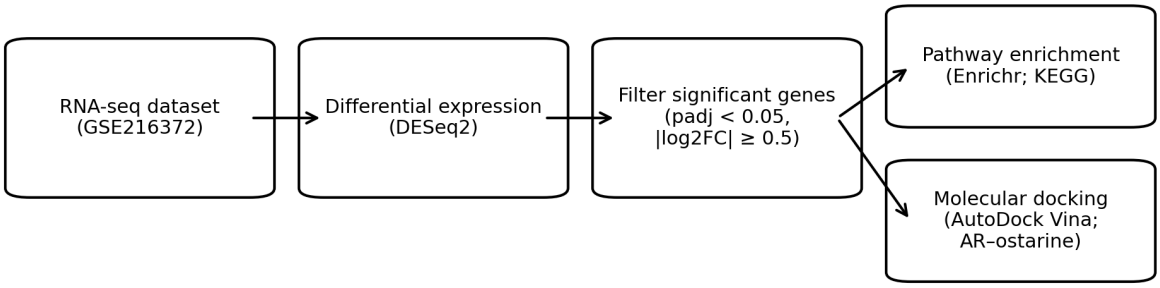

Figure 2. Volcano plot of differential expression (ARKO vs WT muscle; GSE216372) showing significance versus log2 fold change.

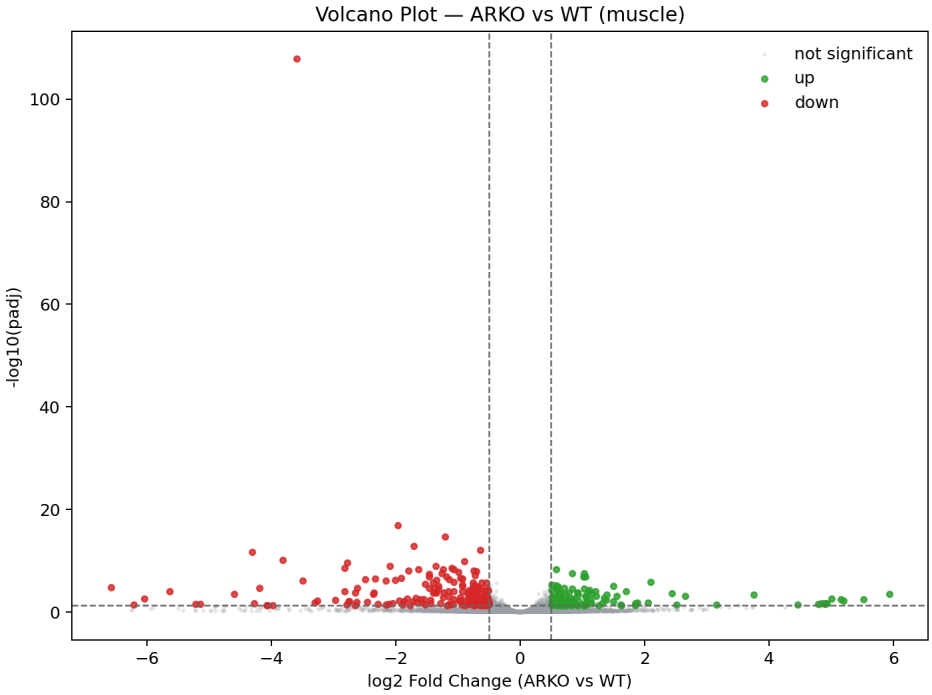

Figure 3. Heatmap of top differentially expressed genes (z-scored per gene) across WT and ARKO samples.

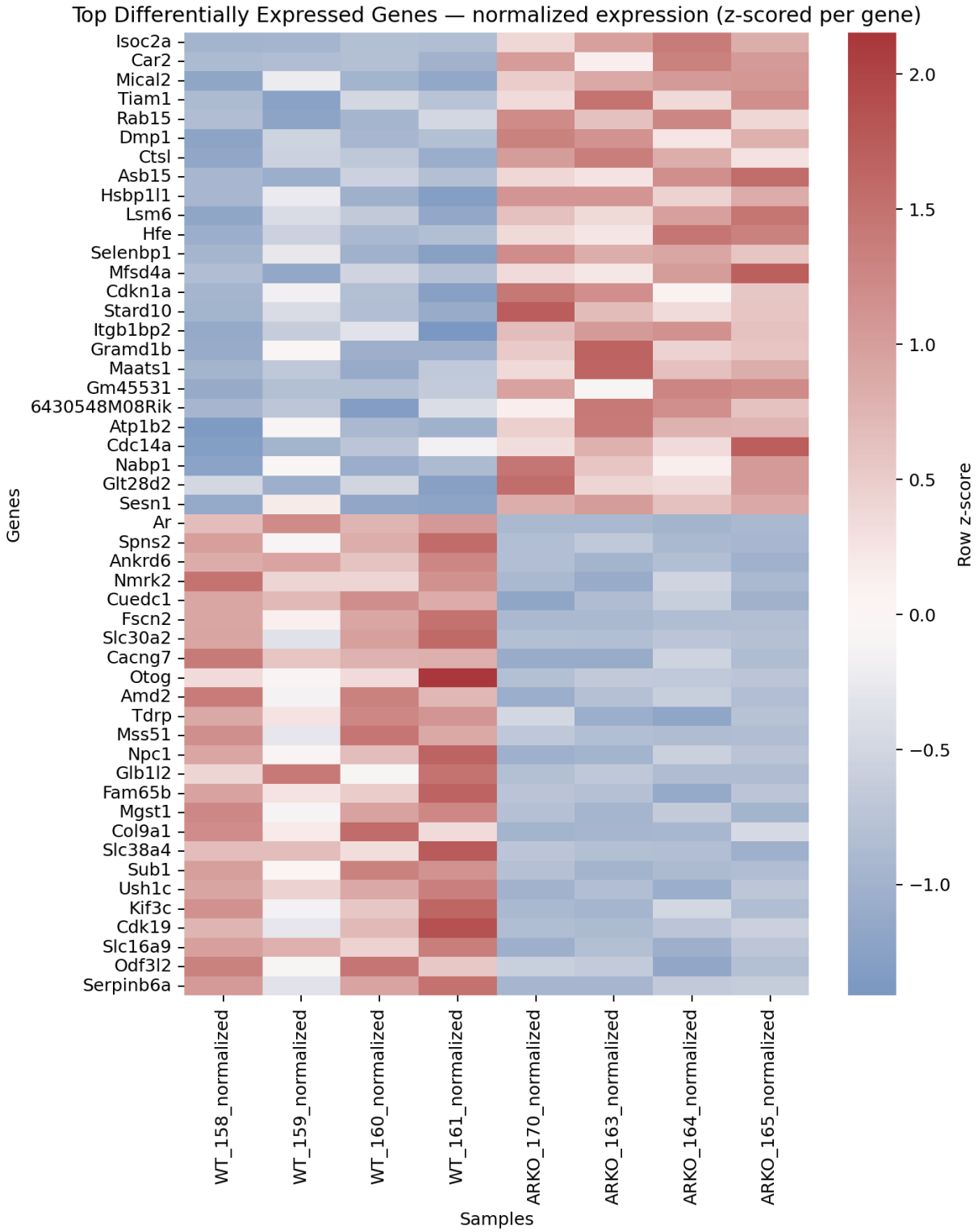

Figure 4. Top enriched KEGG pathways from Enrichr (KEGG 2021 Human) for the ARKO vs WT differential gene set.

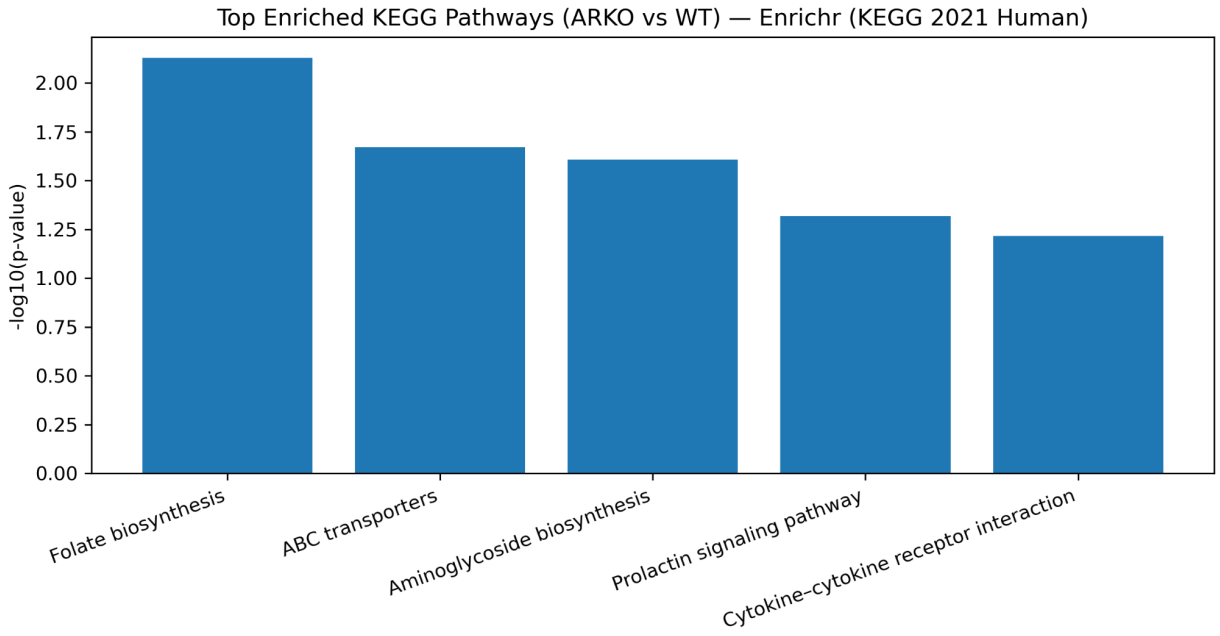

Figure 5. Docking score summary across ligands and receptors (binding energy; kcal/mol).

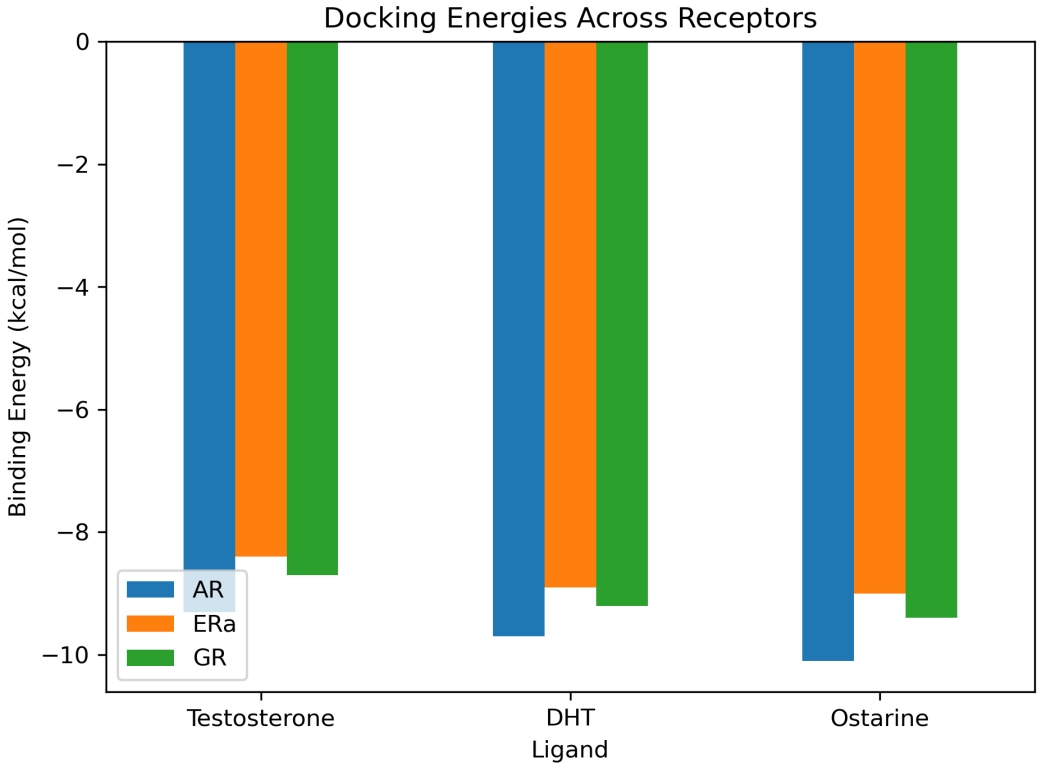

Table 1. Top differentially expressed genes after filtering.

Table 1A. Top 10 up-regulated genes

Rank	Gene	log2FoldChange	Adjusted p-value (padj)
1	Dnajc5g	5.932169	0.000275
2	Abcc2	5.524149	0.003490
3	Gm15857	5.194389	0.006949
4	Fgl1	5.156612	0.003806
5	Gm5083	5.012379	0.002688
6	A4gnt	4.929938	0.023471
7	Ankrd36	4.904175	0.027476
8	Gm12354	4.902406	0.018165
9	Ush2a	4.844273	0.019439
10	Uox	4.790881	0.027422

Table 1B. Top 10 down-regulated genes

Rank	Gene	log2FoldChange	Adjusted p-value (padj)
1	Dnase1	-6.577513	1.402773E-05
2	Blk	-6.207067	3.507575E-02
3	Gm12240	-6.040573	2.677282E-03
4	Ly6g6e	-5.631952	8.620348E-05
5	Cd8b1	-5.219840	2.546452E-02
6	Ly6d	-5.142809	2.546452E-02
7	Irx3os	-4.597744	2.700356E-04
8	Fscn2	-4.304822	2.290942E-12
9	Opn1mw	-4.273940	1.746640E-02
10	Slc15a5	-4.193670	2.136836E-05

Table 2. Top enriched KEGG pathways ranked by nominal p-value (unadjusted).

Rank	Pathway Name	P-value	Adjusted P-value	Odds Ratio	Combined Score
1	Folate biosynthesis	0.0074	0.6769	16.90	82.83
2	ABC transporters	0.0213	0.7123	9.42	36.27
3	Neomycin, kanamycin and gentamicin biosynthesis	0.0248	0.7123	50.24	185.84
4	Prolactin signaling pathway	0.0480	0.7123	5.95	18.08
5	Cytokine–cytokine receptor interaction	0.0607	0.7123	2.81	7.87

(Source: Enrichr, KEGG 2021 Human Results)

Table 3. Best docking affinity (kcal/mol) per ligand–receptor pair (AutoDock Vina)

Ligand	AR Affinity (kcal/mol)	ERα Affinity (kcal/mol)	GR Affinity (kcal/mol)
Testosterone	-9.3	-8.4	-8.7
DHT	-9.7	-8.9	-9.2
Ostarine	-10.1	-9.0	-9.4

Docking under an identical fixed search space predicted that ostarine had the strongest affinity for AR (mode 1: −10.1 kcal/mol), compared with DHT (−9.7 kcal/mol) and testosterone (−9.3 kcal/mol); cross-receptor docking yielded weaker predicted affinities for ERα (−9.0 kcal/mol) and GR (−9.4 kcal/mol) (Table 3; Figure 5).