ISSN(O): 3107-7889

Jana Nexus: Journal of Health and Medicine

Publisher's Name: Jana Publication and Research LLP

https://www.jananexushm.com

REVIEWER'S REPORT

Manuscript No.: JNHM-136

Title: In Silico Analysis of Differentially Expressed Gene Sets Reveals LY6E as a Potential Candidate Gene in Oral Squamous Cell Carcinoma (OSCC) Progression.

Recommendation:

Accept after minor revision.....

Rating	Excel.	Good	Fair	Poor
Originality				
Techn. Quality				
Clarity				
Significance				

Reviewer's ID: JPR-Bilqees Hamza

Detailed Reviewer's Report

The submitted manuscript presents an *in silico* computational study designed to identify novel, stage-specific and stage-independent molecular markers for Oral Squamous Cell Carcinoma (OSCC). Recognizing that OSCC is often asymptomatic in its early phases and frequently diagnosed at advanced clinical stages—which drastically reduces the 5-year survival rate to below 50%—the authors utilize bioinformatics frameworks to uncover early diagnostic and prognostic tools. The primary focus of the work is the analysis of the public microarray dataset GSE85195, extracted from the NCBI Gene Expression Omnibus (GEO) database, consisting of 50 oral cavity samples mapped across 34 OSCC cases, 15 pre-cancerous oral leukoplakia (OLK) cases, and 1 normal control.

By sorting these data cohorts using the GEO2R platform alongside the GEOquery and limma statistical packages in R, the authors filtered differentially expressed genes (DEGs) using an absolute log fold change threshold ($|\log FC| > 1$) and adjusted p-values ($p < 0.05$). This computational workflow revealed distinct stage-wise alterations, highlighting 318 DEGs in Control vs. Stage 1, 10,012 DEGs in Control vs. Stage 2, 9,392 DEGs in Control vs. Stage 3, and 4,480 DEGs in Control vs. Stage 4. Utilizing a Venn diagram cross-comparison, the authors identified a baseline signature of 245 common DEGs altered across all malignant progression stages relative to the control.

To determine the functional and physical connections among these 245 core signatures, a protein-protein interaction (PPI) network was constructed via the STRING database and structurally mapped using Cytoscape software. Applying the Molecular Complex Detection (MCODE) sub-network clustering

ISSN(O): 3107-7889
Jana Nexus: Journal of Health and Medicine
Publisher's Name: Jana Publication and Research LLP
https://www.jananexushm.com

REVIEWER'S REPORT

plugin, the authors isolated a highly interconnected functional module containing 10 key genes displaying a maximum MCODE density score of 10. This core sub-network consists of *USP18*, *OAS2*, *LY6E*, *ISG15*, *IRF7*, *IFIT3*, *IFI6*, *IFI35*, *HERC6*, and *BST2*. Gene Ontology (GO) and KEGG functional enrichment analyses performed via the gProfiler tool revealed that these tightly clustered candidate targets are heavily implicated in host-interspecies interactions, negative viral regulation, and the positive regulation of interferon-beta ($IFN-\beta$) production pathways.

The manuscript then shifts to external validation and survival analysis using the GEPIA platform, utilizing TCGA and GTEx cancer genomics infrastructure. The authors emphasize *LY6E* as the most promising driver candidate and diagnostic biomarker due to its top MCODE ranking and its biological role in modulating immune evasion mechanisms within the tumor microenvironment. Furthermore, to explore the post-transcriptional and transcriptional landscapes governing *LY6E*, the authors mapped an interactive regulatory hub featuring 7 transcription factors (TFs) and 26 microRNAs (miRNAs). The study concludes that *LY6E* stands as a promising molecular driver that could enhance future target-dependent oncological therapies and diagnostic models for OSCC.

Improvements

While the study addresses an important clinical challenge and provides a foundational screening of the GSE85195 dataset, the manuscript exhibits severe structural contradictions, methodological flaws, and typographical discrepancies that compromise its scientific rigor and eligibility for indexation.

The most critical area requiring immediate improvement concerns a fundamental, systemic error regarding the tissue datasets used for external validation and stage plotting. In the materials and methods section, the authors state that they validated their OSCC-derived biomarkers using TCGA and GTEx datasets of "Liver hepatocellular carcinoma" (LIHC). Subsequently, in the results section under the validation and stage plot subheadings, the text states that they investigated differential expression in "Esophageal carcinoma" (ESCA) utilizing "286 normal liver tissue samples". Finally, the network analysis section introduces a completely new third malignancy, claiming the network reveals interconnectedness in "gallbladder cancer".

This is a profound logical and biological mismatch. A candidate biomarker identified from oral squamous epithelial cells cannot be validated using liver, esophageal, or gallbladder datasets, nor can normal liver tissues serve as a control group for upper aerodigestive tract malignancies. The authors must consistently use the Head and Neck Squamous Cell Carcinoma (HNSCC or HNSC) dataset within TCGA to validate targets found in an oral cancer dataset. Validating oral cavity findings against unrelated digestive organs undermines the biological relevance of the entire validation framework.

ISSN(O): 3107-7889
Jana Nexus: Journal of Health and Medicine
Publisher's Name: Jana Publication and Research LLP
https://www.jananexushm.com

REVIEWER'S REPORT

Methodologically, the baseline dataset composition presents a severe statistical limitation that must be addressed with transparency. The GSE85195 dataset utilizes only a single ($n = 1$) normal control sample. Conducting a differential expression analysis using the limma linear modeling framework against a single control sample lacks statistical variance and drastically increases the false-positive rate. The authors fail to justify this or acknowledge it as a limitation in their discussion. Furthermore, the reported stage-wise numbers contain major formatting errors; for instance, the text records "10012 DEGs" for Stage 2 and "9392 DEGs" for Stage 3. Given that microarrays typically capture around 20,000 to 44,000 probes total, having over 10,000 distinct genes differentially expressed between a stage and a single control suggests a loose statistical threshold or an unfiltered background noise problem.

The functional annotation and discussion regarding *LY6E* also require refinement. The authors repeatedly state that *LY6E* showed the "highest MCODE value" indicating its critical involvement. However, by definition, an MCODE score is calculated for an entire extracted *network module* rather than an isolated node. As illustrated in the provided network diagram, all 10 genes belong to a single, completely interconnected clique where every node is joined to almost every other node, sharing an identical module score of 10. The text needs to clarify what exact topological metric (e.g., degree centrality, betweenness centrality, or bottleneck score) uniquely positioned *LY6E* above the other 9 interferon-stimulated partner genes like *ISG15* or *USP18*.

Furthermore, the manuscript suffers from internal text contradictions within its results section. When describing the stage-wise violin plots, the authors note that *USP18* and *OAS2* displayed progressive upregulation, whereas *LY6E*, *IFIT3*, *IFI35*, *BST2*, and *HERC6* "showed less pronounced expression changes across stages, suggesting their potentially limited impact on disease progression". This directly contradicts the abstract and discussion, where *LY6E* is elevated as the most critical progression driver. If *LY6E* shows flat expression dynamics across stages, its role as a *progression* marker is diminished, and it must instead be framed purely as an early diagnostic marker.

Finally, the presentation of the figures requires significant quality improvements. The labeling on the axis components for Figures 1, 3, 5, and 6 contains low-resolution elements and overlapped text. Crucially, Figure 7 displays a transcription factor and miRNA network where the absolute central hub is clearly labeled as *USF1* (Upstream Transcription Factor 1), while *LY6E* is relegated to a single peripheral red node attached to it. Yet, the accompanying text describes *LY6E* as emerging as the "central hub" interacting directly with 7 TFs and 26 miRNAs. This explicit contradiction between the network visualization and the written narrative must be resolved.

Recommendations

ISSN(O): 3107-7889
Jana Nexus: Journal of Health and Medicine
Publisher's Name: Jana Publication and Research LLP
https://www.jananexushm.com

REVIEWER'S REPORT

To elevate this manuscript to a standard suitable for international peer-reviewed publication, the following strategic changes must be implemented:

First, the authors must completely re-execute the external validation, stage plots, and survival analyses on the GEPIA/TCGA platform using exclusively the Head and Neck Squamous Cell Carcinoma (HNSC) cohort. All references to liver hepatocellular carcinoma, esophageal carcinoma, and gallbladder cancer must be purged from the text, tables, and figure legends to correct this fundamental data mismatch.

Second, the authors need to carefully re-verify and re-filter their GEO2R processing settings for the GSE85195 dataset. They must apply a stringent, uniform false discovery rate (FDR) correction (such as Benjamini-Hochberg adjusted $p < 0.01$) to reduce the unusually high gene counts reported for Stages 2 and 3. The exact number of up- and down-regulated genes per stage should be presented in a clean, comprehensive summary table.

Third, the text must include a dedicated "Study Limitations" section within the Discussion. This section must explicitly address the statistical constraints of relying on a single control sample in the discovery dataset and outline how validating against larger, multi-sample external cohorts (like TCGA-HNSC) helps mitigate this initial design deficit.

Fourth, the authors must fix the MCODE terminological inaccuracy. Instead of stating that *LY6E* possessed the highest individual MCODE score, they should compute individual node centrality metrics—such as Degree, Closeness, or Betweenness centrality—using Cytoscape's NetworkAnalyzer tool to mathematically demonstrate why *LY6E* represents the most critical hub within that interferon-regulated module.

Fifth, the contradiction regarding the stage plots must be rectified. The authors must align their description of Figure 6 with their primary hypothesis. If the violin plots show that *LY6E* expression remains stably elevated from Stage 1 through Stage 4, the narrative must clarify that *LY6E* acts as an early-onset pan-stage diagnostic marker rather than a driver of advanced clinical transformation.

Sixth, Figure 7 must be redesigned or its description rewritten. If *USF1* is the actual topological hub of the target network as shown in the diagram, the text must correctly identify *USF1* as the primary upstream transcription factor regulating *LY6E* expression, rather than erroneously claiming *LY6E* is the structural center of that specific visual plot.

Finally, a meticulous proofreading of the entire text is mandatory. The authors must correct basic typographical errors, such as "mostprevalent" in line 32, "indicates" instead of "indicating" in line 44, "IFN- " missing its beta symbol in line 101, and the duplicated text snippet "geo/)." in line 111. Ensuring consistent nomenclature and structural accuracy will make the manuscript cohesive, rigorous, and scientifically sound.

ISSN(O): 3107-7889
Jana Nexus: Journal of Health and Medicine
Publisher's Name: Jana Publication and Research LLP
https://www.jananexushm.com

REVIEWER'S REPORT

Final Recommendation

Decision: Recommendation for Publication with Minor Revision

Justification

The manuscript addresses a highly significant clinical challenge by establishing an *in silico* discovery pipeline for targetable molecular networks in Oral Squamous Cell Carcinoma (OSCC). By identifying a core baseline signature of 245 common DEGs and isolating an influential interferon-stimulated functional module, the authors have successfully highlighted *LY6E* as a central molecule of therapeutic and diagnostic interest. The integrated look into its regulatory network featuring specific transcription factors and microRNAs provides a valuable foundation for future translational medicine.

The severe errors identified during evaluation—specifically the incorrect use of liver, esophageal, and gallbladder cancer datasets for the validation of an oral malignancy—are structural, administrative, and editorial in nature rather than fatal conceptual flaws. Because these contradictions can be completely resolved through targeted data revision without undermining the core primary screening findings, the manuscript is recommended for publication in the *Journal of Natural Health and Medicine (JNHM)*, conditional upon the meticulous execution of these minor corrections.