nature portfolio

Peer Review File

Switch-like Gene Expression Modulates Disease Risk

Corresponding Author: Professor Omer Gokcumen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors present a comprehensive analysis of transcriptomes from 943 individuals across 27 tissues to identified 1,013 switch-like genes. They found that only 31 of these genes exhibit switch-like behavior across all tissues and that the universally switch-like genes appear to be genetically driven, with large exonic genomic structural variants explaining five of them. The remaining switch-like genes exhibit tissue-specific expression patterns and tend to be switched on or off in unison within individuals, likely under the influence of tissue-specific master regulators, including hormonal signals. They present a very robust sex specific analysis of the findings and explore concordance in the gene expression. They provide specific biological links between switched-off genes in the stomach and vagina that are linked to gastric cancer and vaginal atrophy with some experimental validation.

Overall the study is interesting impactful, the manuscript is clear and well written and the figures are effective. Please find some suggestions for points to be addressed below:

• Please define switch-like genes early on (both in the abstract and in the introduction)
• Exploring overall pathway or gene set enrichment analysis in the group of ~1013 identified genes might be informative in learning more about their overall function. Further characterization could be done of the individual clusters presented in Figure 2.
• I would suggest including the table of clusters in 2A and 2B in the main manuscript with gene descriptions and relevant statistics
• I suggest an organized section of the results to focus on the sex specific analysis and also highlight that in the abstract / introduction
• Figure 6 would further be enhanced by sex specific heatmaps showing the bimodal nature of in the expression of the genes providing a high level visualization and summary of the findings.
• Another analysis that might be interesting to explore is intersecting disease associated genes from OMIM or GWAS Catalog with the set of 1013 switch like genes generally tying them to disease before focusing on stomach and vaginal signals otherwise it's unclear how these were chosen as examples or experimental followup.
• Throughout the results section there is a lot of text that could be moved into discussion, including potential limitations of the analyses, keeping the results shorter and more to the point

Reviewer #2

(Remarks to the Author)
Overview: Aqil et al present work investigating switch-like behavior of gene expression using mRNA measurements from individuals in GTEx. To identify switch-like genes, they authors performed a test for bi-modality. Using this approach the authors identified a number of genes exhibited switch-like behavior. The central hypothesis is interesting, and the manuscript is well written, however I have concerns regarding the statistical methodology as well as confounders.

Major Comments:
1. It would be helpful for the authors to perform simulations or some empirical-based simulation/permutation to assess the calibration of p-values computed from the dip test.

2. The authors performed some analyses to demonstrate the effect of ischemic time on measurements, but I have additional

concerns. In practice, there may be a number of confounders that track to rna quality, batch, processing, etc.

For example, in eQTL studies, it is standard practice to adjust for PCs of gene expression measurements precisely due to technical artifacts. Similarly, age of the participants should be considered.

Reviewer #3

(Remarks to the Author)
The authors propose a nice way of looking at switch-like (aka bimodal) expression patterns in genes across tissues.

Larger points:

They show some nice results, my primary concern is that I'm not sure there is enough context for how surprising this should be relative to expectation. For example, about 5% of the genes are identified as being significantly bimodal even with BH correction. Is this distinct from a null? What's the proposed model here? Is it more than expected?

Another concern I had - there's not much discussion about whether this could be accounted for by epigenetic tissue specific patterns or other epigenetic patterns (like imprinting) that would cause bimodal expression patterns. It would be nice to see whether there is any relationship between existing understood epigenetic variability and switch like behavior.

Minor comments:

63: Would be helpful to actually explain this in the first paragraph of the paper.
97: Font is weirdly larger here than for the word 'results' which seems confusing. Sorry!!
115: Could clarify language there as "pairwise correlation of expression levels for the same gene in different tissues" and then reduce the subsequent clarification lines.
258: Where are these genes located? Are they on sex chromosomes? Are sex chromosomes more or less likley to contain switch like genes and is this related to the second larger concern point?
374: Clunky phrasing 'symptoms of vaginal atrophy appear'... maybe better as just 'causing vaginal atrophy'

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The reviewers have addressed my comments.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
I appreciate the authors' efforts towards addressing my previous comments. I have no new comments at this time.

(Remarks on code availability)
The repo has extensive documentation, which is appreciated.

Reviewer #3

(Remarks to the Author)
The authors have fully addressed my comments - thank you.

(Remarks on code availability)
The confounders code (GAM.r) is substantially more clearly commented than the (admittedly more brief) dip test or sex bias code - otherwise everything is very clear and well laid out.

Point-by-point response

We thank the editor and the reviewers for their extremely helpful suggestions. Based on the reviewers' comments, we have revised our manuscript and addressed all the concerns. The changes in the main text are marked with red. To describe the new results, we have added multiple figures (**Figures 1D-E, 6D, and 7**) and supplementary material (**Supplementary texts 1-3**, **Figures S1-3**).

Reviewer #1 (Remarks to the Author):

The authors present a comprehensive analysis of transcriptomes from 943 individuals across 27 tissues to identified 1,013 switch-like genes. They found that only 31 of these genes exhibit switch-like behavior across all tissues and that the universally switch-like genes appear to be genetically driven, with large exonic genomic structural variants explaining five of them. The remaining switch-like genes exhibit tissue-specific expression patterns and tend to be switched on or off in unison within individuals, likely under the influence of tissue-specific master regulators, including hormonal signals. They present a very robust sex specific analysis of the findings and explore concordance in the gene expression. They provide specific biological links between switched-off genes in the stomach and vagina that are linked to gastric cancer and vaginal atrophy with some experimental validation.

Overall the study is interesting impactful, the manuscript is clear and well written and the figures are effective. Please find some suggestions for points to be addressed below:

We really appreciate the positive feedback from the reviewer. We have now addressed all the reviewer's comments, which has further improved our manuscript, in our opinion.

• Please define switch-like genes early on (both in the abstract and in the introduction)

We have now added the descriptor "on-versus-off" with our first mention of switch-like genes in both the abstract and introduction to ensure clarity.

• Exploring overall pathway or gene set enrichment analysis in the group of ~1013 identified genes might be informative in learning more about their overall function. Further characterization could be done of the individual clusters presented in Figure 2.

Based on reviewer's suggestions, we performed a comprehensive enrichment analysis of all switch-like genes using GREAT, which evaluates functional enrichments across multiple biological processes and diseases. The results have now been incorporated into the main text and presented in the new **Figure 1D**.

Unfortunately, our cluster-specific enrichment analysis did not yield significant results for clusters 2A and 2B due to the small sample sizes. However, we note that the significant enrichments observed for all switch-like genes are largely driven by cluster 1, which comprises most genes.

• I would suggest including the table of clusters in 2A and 2B in the main manuscript with gene descriptions and relevant statistics.

We have shared these genes in **Tables S2-3** and the corresponding violin plots in **Figure S6**. Given that the main manuscript already contains eight multi-panel figures, we refrained from adding the full table to the main text to maintain readability.

• I suggest an organized section of the results to focus on the sex specific analysis and also highlight that in the abstract / introduction

We recognize the importance of emphasizing sex-biased switch-like gene expression. To address this, we have renamed the section on tissue-specific switch-like genes to "Sex is a frequent modulator of tissue-specific switch-like gene expression" to better highlight these findings. Additionally, we have added a sentence in the Introduction to draw attention to our results on sex-biased switch-like genes in the breast. However, we have opted not to include this in the abstract to maintain conciseness and ensure that the primary focus remains on the disease-related implications of our findings. We believe this approach allows us to keep the abstract clear and impactful while still addressing the reviewer's concern within the main text.

• Figure 6 would further be enhanced by sex specific heatmaps showing the bimodal nature of in the expression of the genes providing a high level visualization and summary of the findings.

We thank the reviewer for this suggestion. We have now added a new heatmap (**Figure 6B**) and agree that it greatly enhanced the visualization of our results.

• Another analysis that might be interesting to explore is intersecting disease associated genes from OMIM or GWAS Catalog with the set of 1013 switch like genes generally tying them to disease before focusing on stomach and vaginal signals otherwise it's unclear how these were chosen as examples or experimental followup.

We appreciate this valuable suggestion. Instead of OMIM, we used Human Disease Ontology (DO) for enrichment analysis, as DO already groups diseases into biological categories, making enrichment testing more straightforward. Our analysis confirmed strong associations with cancers, metabolic and immune-related diseases, and integumentary (skin-related) conditions. We now include a new figure (**Figure 1D**) and associated text to describe these results.

• Throughout the results section there is a lot of text that could be moved into discussion, including potential limitations of the analyses, keeping the results shorter and more to the point.

We appreciate the reviewer's comments and recognize the importance of maintaining a concise **Results** section. We have now moved the discussion on the potential limitation of the study in the context of protein versus RNA expression from **Results** to **Discussion**.

Reviewer #2 (Remarks to the Author):

Overview: Aqil et al present work investigating switch-like behavior of gene expression using mRNA measurements from individuals in GTEx. To identify switch-like genes, they authors performed a test for bi-modality. Using this approach the authors identified a number of genes exhibited switch-like behavior. The central hypothesis is interesting, and the manuscript is well written, however I have concerns regarding the statistical methodology as well as confounders.

We appreciate the positive comments. We implemented all the suggestions by the reviewer, and we believe they have improved our manuscript substantially.

Major Comments:

1. It would be helpful for the authors to perform simulations or some empirical-based simulation/permutation to assess the calibration of p-values computed from the dip test.

Thank you for this valuable suggestion. To examine the calibration of our dip test p-values, we now performed the test on:

1. Simulated unimodal expression data approximating the mean and standard deviation observed in GTEx, under a normal distribution; and
2. Empirical expression data from housekeeping genes, which are reliably unimodal.

Instead of observing uniformly distributed p-values under the null hypothesis, the p-values were heavily skewed toward one, implying that our dip test implementation was much more conservative than anticipated by our 5% FDR threshold. To address this, we performed an empirical null calibration using the housekeeping-gene p-value distribution. We have now updated all of our results throughout the manuscript based on these new analyses. After incorporating both of the reviewer's suggestions (p-value recalibration and more comprehensive confounder-correction) into our methodology, the number of the identified switch-like genes has changed from 1,013 to 473. The methodological details about these steps are available in **Methods** in the section "Dip test", in **Supplementary text 2**, and **Figure S2.**

2. The authors performed some analyses to demonstrate the effect of ischemic time on measurements, but I have additional concerns. In practice, there may be a number of confounders that track to rna quality, batch, processing, etc.

We agree. We have now controlled for six confounders, i.e., RIN (RNA integrity number), the Hardy death score, Cohort type (post-mortem vs. surgical vs. organ donor), type of nucleic acid isolation batch (RNA_Extraction_from_Paxgene-derived_Lysate_Plate_Based vs. RNA_isolation_PAXgene_Tissue_miRNA vs. RNA_isolation_PAXgene_Blood_RNA), ischemic time (post-mortem interval), and time spent by tissue in PAX fixative. Specifically, we used residuals from generalized additive models for each of these potential confounders, as we have now explained in the "Correcting expression levels for confounders" section in **Methods**. We found that several hundred genes were strongly influenced by these confounders (**Figure S3**). We have removed those genes from further analyses and updated our manuscript accordingly.

We carefully evaluated whether to correct for the two batch IDs (SMNABTCH for nucleic acid isolation and SMGEBTCH for expression). However, in most tissues, each "batch" contained only one or a few samples from a given tissue, which would drastically reduce our statistical power and potentially eliminate genuine bimodality signals if we treated these sparse

groupings as separate factors. Therefore, we instead employed a broader “batch type” confounder (nucleic acid isolation batch type; SMNABTCHT). Meanwhile, the expression batch type (SMGEBTCHT) was identical across all samples in this study, so we did not use it. Additionally, we did not correct for the autolysis score (SMATSSCR) because the effects of certain pathologies on the tissue may be similar to autolysis. In such cases, correcting for autolysis may obscure the biological signals of interest. We believe that not correcting for autolysis score will not significantly bias our results since we have already controlled for the related confounders of RNA integrity and post-mortem interval.

As a result of the new adjustments in p-value calibration (described above) and confounder correction, our dataset is now more biologically robust, while retaining our principal findings. For instance, after confounder correction, the number of Cluster 1 genes in the vagina decreased from 214 to 34, yet genes linked to vaginal atrophy rose from six to seven. Similarly, whereas we previously attributed 43% of switch-like expression in Cluster 2A to genetic variants, this percentage has now increased to 69%. However, we did lose the stomach-specific bimodal genes that were previously linked to gastric cancer, as RIN differences across individuals drove their patterns. Therefore, we have removed the corresponding section in the revised manuscript. Although RIN may itself be connected to biological phenomena in the stomach, exploring this possibility is beyond the scope of the current study.

For example, in eQTL studies, it is standard practice to adjust for PCs of gene expression measurements precisely due to technical artifacts. Similarly, age of the participants should be considered.

We appreciate the reviewer’s suggestion to consider principal component (PC) adjustment, which is standard in many eQTL studies. However, our goal differs in that we aim to capture all biologically relevant variation, including factors such as sex, age, endocrine signaling, and epigenetic regulation. In eQTL detection, these factors are often removed to isolate purely genetic effects. In contrast, our emphasis is on identifying any bimodal expression, regardless of whether it arises from genetic or other biological influences.

To address the age concern, we examined correlations between switch-like gene expression and participants’ ages (**Table S8**) but observed only two significant associations (**Figure 6D**). We also performed similar analysis for BMI (**Table S8**). The results are now described in the **Results** under the section “Sex is a frequent modulator of tissue-specific switch-like gene expression.” We believe that our targeted approach to confounder correction, together with the specific biological context (e.g., vaginal atrophy), reduces false positives arising from technical artifacts while preserving genuine biological signals.

Reviewer #3 (Remarks to the Author):

The authors propose a nice way of looking at switch-like (aka bimodal) expression patterns in genes across tissues.

We thank the reviewer for positive remarks and thorough reading of our manuscript.

Larger points:

They show some nice results, my primary concern is that I'm not sure there is enough context for how surprising this should be relative to expectation. For example, about 5% of the genes are identified as being significantly bimodal even with BH correction. Is this distinct from a null? What's the proposed model here? Is it more than expected?

That is an excellent point that resonates with comments from Reviewer 2. To ensure the robustness of our findings, we refined our switch-like gene identification pipeline by (1) calibrating p-values for the dip test and (2) incorporating additional confounder corrections. These refinements are detailed in **Supplementary Texts 2–3**, **Figures S2–4**, and **Methods** in the revised manuscript. To address the reviewer's question directly, identifying 473 switch-like genes is indeed a surprising result. Under the null hypothesis—modeled using both housekeeping gene-tissue pairs and simulated unimodal expression distributions (described in **Supplementary Text 2, Figure S2**)—we observed zero switch-like genes at an FDR < 5% threshold. This finding confirms that the observed switch-like expression is highly unlikely to arise from random transcriptional variation coming from unimodal distributions.

Another concern I had - there's not much discussion about whether this could be accounted for by epigenetic tissue specific patterns or other epigenetic patterns (like imprinting) that would cause bimodal expression patterns. It would be nice to see whether there is any relationship between existing understood epigenetic variability and switch-like behavior.

We fully agree that epigenetic factors warrant further exploration. To investigate this, we analyzed GTEx methylation data for the six tissues in which both methylation and expression profiles were available for at least 30 individuals. This analysis identified dozens of genes whose switch-like expression is at least partially regulated by epigenetic modifications, either through genetic or environmental influences. We have now incorporated these results in **Figure 7** and the **Results** section under "Epigenetic regulation of bimodal gene expression by DNA methylation", which we believe significantly strengthens the biological interpretation of our findings.

In response to the reviewer's specific question about imprinting, we examined all identified switch-like genes in the Geneimprint database (https://www.geneimprint.com/site/genes-by-species). We found only one switch-like gene (*ERAP2*) annotated as imprinted. However, its bimodal expression can be explained by an eQTL (rs2248374; **Supplementary text 6**), indicating that imprinting is not required to account for its bimodality. Therefore, we do not discuss imprinting the revised manuscript.

Minor comments:

63: Would be helpful to actually explain this in the first paragraph of the paper.

We have now added the descriptive modifier “on-versus-off” to our first mention of the term switch-like genes. However, since the initial context relates to bacterial gene regulation, we have retained the full definition later in the manuscript, where it is most relevant.

97: Font is weirdly larger here than for the word 'results' which seems confusing. Sorry!!

We had mistakenly left it in uppercase. We have corrected this now.

115: Could clarify language there as "pairwise correlation of expression levels for the same gene in different tissues" and then reduce the subsequent clarification lines.

We have clarified this in the text as “Therefore, for each gene, we calculated the pairwise tissue-to-tissue correlation of expression levels for all tissue pairs.”

258: Where are these genes located? Are they on sex chromosomes? Are sex chromosomes more or less likely to contain switch-like genes, and is this related to the second larger concern point?

Among Cluster 2A switch-like genes, only one gene (*GPX1P1*) is located on the X chromosome, and its bimodal expression does not correlate with either methylation at nearby sites or genetic polymorphisms. Y chromosome genes make up most of cluster 2B (7 out of 8), as mentioned in the section “Genetic variation underlies universally switch-like genes” in **Results**. In Cluster 1, two genes are located on sex chromosomes:

- *FAM9C* (X chromosome): A female-biased switch-like gene that escapes X chromosome inactivation in females. We now discuss this gene explicitly in the Results section: “Sex is a frequent modulator of tissue-specific switch-like gene expression.”
- ENSG00000273906 (Y chromosome): This gene should theoretically belong to Cluster 2B but was grouped into Cluster 1 due to slight clustering imperfections. This is now discussed in **Supplementary text 5**.

374: Clunky phrasing 'symptoms of vaginal atrophy appear'... maybe better as just 'causing vaginal atrophy'

We have changed the wording here to say “causing vaginal atrophy.”

Response to reviewers

Reviewer #1 (Remarks to the Author):

The reviewers have addressed my comments.

Thank you. We are glad!

Reviewer #2 (Remarks to the Author):

I appreciate the authors' efforts towards addressing my previous comments. I have no new comments at this time.

Thank you for the comments. We are glad we addressed them.

Reviewer #2 (Remarks on code availability):

The repo has extensive documentation, which is appreciated.
We thank the reviewer for reviewing the GitHub page.

Reviewer #3 (Remarks to the Author):

The authors have fully addressed my comments - thank you.
We thank the reviewer for the helpful comments. We are glad the reviewer's concerns were addressed.

Reviewer #3 (Remarks on code availability):

The confounders code (GAM.r) is substantially more clearly commented than the (admittedly more brief) dip test or sex bias code - otherwise everything is very clear and well laid out.
Thank you for reviewing the GitHub page. We have added more comments to the diptest.r script to address this concern. This can be viewed at https://github.com/AlberAqil/Switch_like_gene_expression_modulates_disease_risk_2025/blob/main/Dip_test/Diptest.r. In particular, we have added the following comments:
#This script applies the dip test of the distribution of log(TPM + 1) as opposed to the distribution of raw TPMs.
#This script calculates the dip statistic D, the associated p-value, mean of the distribution, median of the distribution, the first quartile, the third quartile, and the standard deviation.