

Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with colocRedRibbon

Author list

Anthony Piron; Florian Szymczak; Lise Folon; Daniel J M Crouch; Theodora Papadopoulou; Maria Lytrivi; Yue Tong; Maria Inês Alvelos; Maikel L. Colli; Xiaoyan Yi; Marcin L. Pekalski; Konstantinos Hatzikotoulas; Alicia Huerta-Chagoya; Henry J. Taylor; Type 2 Diabetes Global Genomics Initiative; Matthieu Defrance; John A. Todd; Décio L. Eizirik; Josep M. Mercader; and Miriam Cnop

Summary

Initial submission: Received : December 12th 2024

Scientific editor: Judith Nicholson

First round of review: Number of reviewers: 2
Revision invited : May 12th 2025
Revision received : June 29th 2025

Second round of review: Number of reviewers: 2
Accepted : 18th August 2025

Data freely available: Yes

Code freely available: Yes

This transparent peer review record is not systematically proofread, type-set, or edited. Special characters, formatting, and equations may fail to render properly. Standard procedural text within the editor's letters has been deleted for the sake of brevity, but all official correspondence specific to the manuscript has been preserved.

Referees' reports, first round of review

Reviewer 1

The manuscript by Piron and colleague presents a novel computational pipeline, called colocRedRibbon, for the identification of genes associated with nearby GWAS variants. The authors demonstrated that this pipeline is more sensitive in identifying more nuanced associations using pancreatic islet eQTL analysis from the TIGER study, and have identified > 400 colocalisations using this method, more than double the number achieved by previous methods, and many novel genes in type 1 diabetes. This pipeline is a valuable contribution to the field, with applications beyond diabetes research. The reviewer appreciates that this method is already online and available to be used. Below are some minor comments that would strengthen the manuscript:

1. The methodology is indeed the most significant aspect of the study, as effectively demonstrated using the pancreatic islet dataset. Consider reorganizing the results to emphasize the method's output (e.g., coloc vs colocRedRibbon, Figure 4A) prior to specific examples and biological interpretations (i.e., prior to the "The colocRedRibbon shortlisting steps solve the issue of multiple GWAS signals" section). This would highlight the pipeline's strengths early in the manuscript.
2. In the discussion, the authors state: "It is also applicable to quantitative trait loci for protein, open chromatin, and DNA methylation." However, this claim is not substantiated with any examples or discussion in the manuscript. Would be valuable to elaborate on this point to guide those who may wish to apply colocRedRibbon to these contexts.
3. Is the method also applicable to identifying trans-eQTLs?
4. Include a brief explanation of the H4 hypothesis in the results section to aid readers unfamiliar with colocalization metrics, including the cut-offs used.
5. The pathway enrichment analysis does not reveal many specific pathways. Have the authors tried alternative tools like Metascape or EnrichR? These might offer complementary insights or identify additional pathways.

6. In Figure 6, the siRNA experiment in EndoCBH1 is indeed interesting, but the KD efficiency of the gene is not shown. Have the authors also investigated the impact on insulin secretion?
7. It would be useful to include a table summarizing the most significant or biologically interesting co-localizations in the main manuscript. This would make it easier for readers to identify the highlights without relying solely on the supplementary.
8. Supplementary Tables:
 - o The authors cite Marselli et al., 2019 in Supplementary Table 1. Should this be Marselli et al., 2020? Please verify.
 - o It would help to include a column whether the fGSEA.t2d column/results reflect upregulated or downregulated genes with T2D
 - o Provide some explanation or clearer column names for supplementary tables. For example, not clear what "s" column in S7 represents.
 - o Ensure Supplementary Tables are in order in the manuscript.
9. Consistency with the spelling of "co-localization" or "colocalization".

Reviewer 2

The manuscript "Identification of novel type 1 and type 2 diabetes genes by colocalization of human islet eQTL and GWAS variants with colocRedRibbon" developed a new co-localization pipeline and applied it in T1D and T2D. In short, this is a methods-driven biological discovery paper that fills a gap in the field of co-localization.

In the risk allele effect step and RedRibbon overlap step, the reviewer wants the author to address the following questions:

- 1)"the ranking of variants by P-values for both GWAS or eQTLs and statistically examining potential overlaps between the ranked lists." The P-value is a statistical value that acts as a baseline for ranking whether there is risk. As the author says, "A downside is that it comes at the cost of non-determinism in the minimal P-value finding algorithm."(PMID: 38081640).
- 2)In combination with steps 1 and 2, the authors need to explain why Z(beta) values are not considered for up-down classification and ranking.
- 3)In addition to quantitative glycemic traits, whether to consider BMI.
- 4)In the section "The colocRedRibbon shortlisting steps solve the issue of

multiple GWAS signals," the authors give an example of the CCDC67 gene that, although interesting, needs to be explained. Whether adjacent, highly significant (though weaker) signals violate the principle of COLOC. And whether the signal is indeed biologically significant in subsequent analysis identified in colocRedRibbon.

Reviewers are more concerned about the portability of the method.

5) Whether the disease background of GWAS can affect the effect. Here, the authors mainly analyze T1D and T2D, and whether the method can be transferred to other complex diseases or traits.

6) Another keypoint is the dimension of QTL. First of all, whether different QTL types (such as mQTL, pQTL, etc.) are applicable to the hypothesis of this framework. The author should provide a thorough explanation and explanation, at least in the discussion section. Second, whether background information about QTL, including tissue-specific, disease-specific, and cell-type specific, would interfere with the method.

7) Finally, whether the co-localization method chosen will affect biological discovery.

Authors' response to the first round of review

Editor's comment:

Both reviewers would like to see if this works for multiple classes of QTLs, and reviewer 2 is also concerned this wouldn't apply across multiple diseases.

Thank you for underlining this relevant request. To address the applicability across different diseases, we have expanded the analyses beyond type 1 and type 2 diabetes and 4 glycemic traits. We have performed new experiments on obesity as an additional disease and body mass index (BMI) as an additional trait, also in response to the request of Reviewer 2 to perform analyses on BMI. The manuscript now describes analyses for a total of 3 diseases and 5 traits.

To address the question about the applicability of *colocRedRibbon* to multiple QTL classes, we have complemented our original islet eQTL studies with new experiments using plasma protein QTL (pQTL) and adipose tissue DNA methylation QTL (meQTL) data. We have performed new co-localization analyses between:

- Obesity GWAS and islet eQTL,
- Diabetes GWAS and plasma pQTL,
- BMI GWAS and adipose tissue meQTL.
- Obesity and type 2 diabetes GWAS for the *FTO* region

These new experiments, described in greater detail below, demonstrate the adaptability of *colocRedRibbon* to diverse QTL types and its applicability across diseases and traits. In total, we now report 17 co-localization analyses spanning 3 distinct tissues and 3 classes of QTLs. These new analyses highlight the versatility and robustness of *colocRedRibbon* which will be of interest to the genomics research communities at large.

Reviewers' comments:

Reviewer #1: *The manuscript by Piron and colleague presents a novel computational pipeline, called colocRedRibbon, for the identification of genes associated with nearby GWAS variants. The authors demonstrated that this pipeline is more sensitive in identifying more nuanced associations using pancreatic islet eQTL analysis from the TIGER study, and have identified > 400 colocalisations using this method, more than double the number achieved by previous methods, and many novel genes in type 1 diabetes. This pipeline is a valuable contribution to the field, with applications beyond diabetes research. The reviewer appreciates that this method is already online and available to be used. Below are some minor comments that would strengthen the manuscript:*

We are very grateful to the Reviewer for the very positive evaluation of the manuscript, and for the constructive comments and helpful suggestions that we have addressed below.

1. *The methodology is indeed the most significant aspect of the study, as effectively demonstrated using the pancreatic islet dataset. Consider reorganizing the results to emphasize the method's output (e.g., coloc vs colocRedRibbon, Figure 4A) prior to specific examples and biological interpretations (i.e., prior to the "The colocRedRibbon shortlisting steps solve the issue of multiple GWAS signals" section). This would highlight the pipeline's strengths early in the manuscript.*

Thank you for your valuable feedback. We appreciate your suggestion to emphasize the methodological comparison (e.g., coloc vs. colocRedRibbon) earlier in the manuscript to better highlight the strengths of our pipeline. The Results section is structured as follows: 1. Description of the “co-localization missing link” that provides the rationale for *colocRedRibbon* development and overview of the datasets to be analyzed; 2. Brief summary of the results of the diabetes and glycemic traits GWAS and islet eQTL co-localization; 3. Examples from the diabetes and glycemic traits GWAS and islet eQTL co-localization to illustrate the advances of *colocRedRibbon* versus other co-localization approaches. This is then followed by an in-depth description of the novel findings from the diabetes and glycemic traits GWAS and islet eQTL co-localization studies, and finally by the application of *colocRedRibbon* to other diseases, traits and QTLs.

We would prefer to keep this order as it provides strong scientific and methodological arguments for the importance of this novel analytical approach. We are committed to ensuring that the presentation best serves the readers and are happy to make adjustments if the Reviewer and Editor felt it would improve the clarity and impact of the manuscript.

2. In the discussion, the authors state: “It is also applicable to quantitative trait loci for protein, open chromatin, and DNA methylation.” However, this claim is not substantiated with any examples or discussion in the manuscript. Would be valuable to elaborate on this point to guide those who may wish to apply colocRedRibbon to these contexts.

The point is very well taken. To substantiate our claim, we have performed new analyses beyond our original co-localizations of islet eQTLs with diabetes and glycemic traits GWAS. We performed new co-localization analyses for plasma protein QTL (pQTL) and subcutaneous adipose tissue methylation QTL (meQTL), using diabetes-related GWAS as well as new GWAS for obesity and BMI, also in response to the request by Reviewer 2. In these new analyses we have thus examined co-localization for:

1. Obesity GWAS and islet eQTL,
2. All diabetes-related GWAS and plasma protein QTL,
3. BMI GWAS and adipose tissue meQTL,
4. Obesity and type 2 diabetes GWAS for the *FTO* region.

In these analyses, we have found 12 co-localizations for analysis 1, 40 co-localizations for analysis 2, 5 co-localizations for analysis 3, and co-localization between GWAS in the *FTO* region for analysis 4. These new analyses are described in the Results section (pages 3, 5-6), new Figures 3G, S6, S7 and S8, Tables S13-16 and in the Discussion (page 11). They demonstrate the adaptability of *colocRedRibbon* to diverse QTL types and its applicability across diseases and traits. In total, we now report 17 co-localization analyses spanning three distinct tissues and three classes of QTLs. The full R code to generate these analyses is provided in the zenodo repository via accession number 10.5281/zenodo.13987433 to guide those who may wish to apply it to similar contexts.

3. Is the method also applicable to identifying trans-eQTLs?

Yes, the method is applicable to identifying trans-eQTLs. *ColocRedRibbon* can be used with any well-defined genomic region or a union of multiple regions. In our current analyses, we focused on two-

megabase regions surrounding the QTL gene of interest. The method is equally suitable for investigating chromosomal regions located farther away from the gene—such as those implicated by significant trans-eQTL signals. Users can specify any region of interest, including the union of regions around trans-eQTL variants associated with the gene under study. This flexibility allows colocRedRibbon to be readily adapted for both cis- and trans-eQTL co-localization analyses. We now mention this possibility in the Discussion (page 11).

4. Include a brief explanation of the H4 hypothesis in the results section to aid readers unfamiliar with colocalization metrics, including the cut-offs used.

Thank you for the excellent suggestion. The H4 hypothesis and cut-offs are now explained in the Results section (page 3).

5. The pathway enrichment analysis does not reveal many specific pathways. Have the authors tried alternative tools like Metascape or EnrichR? These might offer complementary insights or identify additional pathways.

Following this helpful comment, we have used the pathway enrichment analysis tools Metascape and EnrichR. We have included these new analyses in Figure S2C-F and refer to them in the Results section (page 7).

Across all platforms tested, the results were highly consistent, likely because these tools draw from similar core databases and ontologies for pathway annotation and enrichment analysis. Consequently, the additional tools did not reveal substantially different or more specific pathways beyond those identified in our original analysis. This suggests that the pathway specificity is due to the underlying gene set rather than the choice of enrichment tool.

6. In Figure 6, the siRNA experiment in EndoCBH1 is indeed interesting, but the KD efficiency of the gene is not shown. Have the authors also investigated the impact on insulin secretion?

Thank you for the interesting comment. We have now included the knockdown efficiency of FUT2 siRNA in EndoC-βH1 cells in Figure S3. Following the Reviewer's suggestion, we have examined the impact of FUT2 silencing on insulin content and insulin secretion. These experiments showed that FUT2 depletion does not alter insulin content nor glucose- or glucose+forskolin-stimulated insulin secretion. These new data have been included in Figure S4 and the results are described on page 8.

7. It would be useful to include a table summarizing the most significant or biologically interesting co-localizations in the main manuscript. This would make it easier for readers to identify the highlights without relying solely on the supplementary.

Thank you for your helpful suggestion. We have now added a summary table to the main text (page 15-16), highlighting the most significant and biologically interesting gene/protein co-localizations discussed in the text. We agree that this will make it easier for readers to quickly identify some of the key findings of the study.

8. Supplementary Tables:

o The authors cite Marselli et al., 2019 in Supplementary Table 1. Should this be Marselli et al., 2020? Please verify.

This has been corrected in the table.

o Provide some explanation or clearer column names for supplementary tables. For example, not clear what "s" column in S7 represents.

Thank you for this excellent advice. We have added a legend of the fields in the first sheet of the Excel spreadsheet.

o Ensure Supplementary Tables are in order in the manuscript.

We have reorganized the order the supplementary tables following the Reviewer's suggestion.

9. Consistency with the spelling of "co-localization" or "colocalization".

We now consistently use "co-localization" throughout the manuscript.

Reviewer #2: *The manuscript "Identification of novel type 1 and type 2 diabetes genes by colocalization of human islet eQTL and GWAS variants with colocRedRibbon" developed a new co-localization pipeline and applied it in T1D and T2D. In short, this is a methods-driven biological discovery paper that fills a gap in the field of co-localization.*

We are very grateful to the Reviewer for the very positive evaluation of the manuscript, and for the constructive comments and helpful suggestions.

In the risk allele effect step and RedRibbon overlap step, the reviewer wants the author to address the following questions:

1) "the ranking of variants by P-values for both GWAS or eQTLs and statistically examining potential overlaps between the ranked lists." The P-value is a statistical value that acts as a baseline for ranking whether there is risk. As the author says, "A downside is that it comes at the cost of non-determinism in the minimal P-value finding algorithm." (PMID: 38081640).

To address the non-determinism, we re-ran the co-localization analyses multiple times, a standard practice in evolutionary algorithm optimization. Through these repeated runs, we identified an additional 5% of co-localizations. Our Zenodo repository (accession number 10.5281/zenodo.13987433, the code is in the src/ directory) includes code to facilitate re-running the co-localization analyses in borderline cases. Importantly, our method detected all co-localizations identified by the original *coloc* approach, as well as additional ones, making our results a strict superset.

2) In combination with steps 1 and 2, the authors need to explain why Z(beta) values are not considered for up-down classification and ranking.

Thank you for your thoughtful comment. In designing the *colocRedRibbon* tool, we intentionally separated "significativity" and "direction of effect" into two distinct components, rather than relying on an aggregate statistic such as the z-score or beta value. The only requirement for the "significativity"

metric is that it ranks SNPs appropriately (with lower values indicating higher significance), while the "direction of effect" indicates whether the association is up-regulating (positive) or down-regulating (negative).

Our rationale for this approach is two-fold:

- Flexibility across datasets: GWAS and QTL studies often report different summary statistics, and z-scores or beta values are not always available. By decoupling significance from direction of effect, *colocRedRibbon* can accommodate a wider range of input formats and study designs.
- Generalizability: This separation allows users to flexibly define their own significance and direction metrics. For example, if z-scores are available, users can set "significativity" to $|zscore|^{-1}$ and "direction of effect" to the sign of the z-score, achieving the same as with a combined metric.

In practice, we have used z-scores and/or beta values in several of our analyses, and this flexible approach is documented in the tool's user guide (<https://antpiron.github.io/colocRedRibbon.html>) and illustrated in the manuscript's code repository on Zenodo (accession number 10.5281/zenodo.13987433, the code is in the `src/` directory).

3) In addition to quantitative glycemic traits, whether to consider BMI.

Thank you for this interesting suggestion. To address the question regarding applicability across traits and diseases (see also points 5 and 6 below), we have expanded our analyses beyond type 1 and type 2 diabetes and 4 glycemic traits, including obesity as an additional disease and BMI as an additional trait. The results of these new analyses are described on pages 9-10, and presented in the new Figures S7 and S8 and Tables S14-16.

4) In the section "The colocRedRibbon shortlisting steps solve the issue of multiple GWAS signals," the authors give an example of the CCDC67 gene that, although interesting, needs to be explained. Whether adjacent, highly significant (though weaker) signals violate the principle of COLOC. And whether the signal is indeed biologically significant in subsequent analysis identified in colocRedRibbon.

Thank you for the insightful comment. It is correct that *CCDC67* does not pertain to biological pathways identified in our subsequent analyses of type 2 diabetes and glycemic trait co-localizing genes; its function in islets is unknown. For this reason, we have replaced the *CCDC67* gene by another example of missed co-localization, namely *RPL39L* (previously presented in Figure S2). Similar to *CCDC67*, *coloc* is confounded in the *RPL39L* region by a neighboring more significant GWAS peak (see new Figure 3A-C, described on pages 4-5). The lead SNP and *RPL39L* gene are further discussed on page 7 in biological terms.

Following your excellent suggestion, we have revised the Results section (page 4) to clarify how *colocRedRibbon* addresses the "single causal variant per trait" assumption inherent to *coloc*. Taking the example of the *RPL39L* gene, we explain how narrowing the region under scrutiny allows for keeping the assumption under control by reducing confounding from nearby signals.

Reviewers are more concerned about the portability of the method.

5) Whether the disease background of GWAS can affect the effect. Here, the authors mainly analyze T1D and T2D, and whether the method can be transferred to other complex diseases or traits.

The point is very well taken. To address the concern regarding the applicability of *colocRedRibbon* across different traits and diseases, we have performed new analyses on obesity (as an additional disease) and BMI (as an additional trait). A description of these new data is provided below (see response to point 6). Altogether, we now have applied *colocRedRibbon* to 3 diseases and 5 traits, showing the portability of the pipeline.

6) Another key point is the dimension of QTL. First of all, whether different QTL types (such as mQTL, pQTL, etc.) are applicable to the hypothesis of this framework. The author should provide a thorough explanation and explanation, at least in the discussion section. Second, whether background information about QTL, including tissue-specific, disease-specific, and cell-type specific, would interfere with the method.

Thank you for underlining this excellent comment. To address it, we performed new co-localization analyses for plasma protein QTL (pQTL) and subcutaneous adipose tissue methylation QTL (meQTL), using diabetes-related GWAS as well as new GWAS for obesity and BMI. In these new analyses we have thus examined co-localization for:

1. Obesity GWAS and islet eQTL,
2. All diabetes-related GWAS and plasma protein QTL,
3. BMI GWAS and adipose tissue meQTL,
4. Obesity and type 2 diabetes GWAS for the *FTO* region.

In these analyses, we have found 12 co-localizations for analysis 1, 40 co-localizations for analysis 2, 5 co-localizations for analysis 3, and co-localization between GWAS in the *FTO* region for analysis 4. These new analyses are described in the Results section (pages 3, 5-6), new Figures 3G, S6, S7 and S8, Tables S13-16 and in the Discussion (page 11). They demonstrate the adaptability of *colocRedRibbon* to diverse QTL types and its applicability across diseases and traits. In total, we now report 17 co-localization analyses spanning three distinct tissues and three classes of QTLs. The full R code to generate these analyses is provided in the zenodo repository via accession number 10.5281/zenodo.13987433 to guide those who may wish to apply it to similar contexts.

For the second point related to background information, we do not expect interference with the *colocRedRibbon* method because of tissue, disease or cell type. Indeed, the new analyses described above show the broad applicability of the approach across diseases, QTL types and tissues. The co-localization results are of course dependent on the tissue and disease analyzed. This is illustrated by the number of co-localizations for obesity GWAS and islet eQTLs (12) as compared with the yield of co-localizations for type 2 diabetes and related trait GWAS and islet eQTLs (268). Islets are highly disease-relevant in the latter condition, while brain and adipose tissue are central in obesity. This is mentioned in the results section (page 9) and we now touch upon the limitation related to the current scarcity of datasets in relevant tissues in the Discussion section (page 11).

7) Finally, whether the co-localization method chosen will affect biological discovery.

We posit that the choice of the co-localization method (*colocRedRibbon* versus other currently available tools) will have a significant impact on biological discovery for several reasons:

- Increased detection power: *ColocRedRibbon* substantially increases the number of detected co-localizations. For example, in the co-localization analyses of diabetes and glycemic traits GWAS and islet eQTLs, *ColocRedRibbon* effectively doubled the number of candidate loci for further biological investigation. This broader detection provides more opportunities to uncover novel gene-trait relationships.
- Enhanced resolution: By separating variants based on direction of effect, *colocRedRibbon* reduces the number of variants in the 99% credible set for each co-localization. This refined resolution helps focus follow-up studies on more relevant variant sets, streamlining experimental validation and functional characterization.

These improvements make *colocRedRibbon* a powerful tool for accelerating and refining biological discovery in complex trait genetics. We have clarified this in the Discussion section (page 10, and 11-12).

Referees' report, second round of review

Reviewer 1

The authors have fully addressed my previous comments, and I have no further concerns. This is a valuable manuscript for the scientific community, and I fully support its publication.

Reviewer 2

The author addressed the reviewers' concerns.

Authors' response to the second round of review