Peer Review File

Manuscript Title: Identification and genetic dissection of convergent persister cell states

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

In this manuscript entitled “Discovery and genetic dissection of an archetypal persister cell State”, the authors aimed to dissect the molecular mechanisms of lag-dependent persisters in E. coli, using a combination of scRNAseq and CRISPRi screening. In an earlier work, the authors isolated a metG mutant that showed prolonged lag phase after transitioning from stationary phase, as well as increased fractions of persisters. In this study, a single-cell RNA atlas of the wild-type and the metG mutant at various growth phases was generated using scRNA-seq, leading to the identification of a “persister population” which is mostly composed of cells of metG at the lag phase, as well as comparing this to a hipA7 mutant to suggest that translational deficiency is a common feature of persister cells. By comparing persister cells to cells exposed to tetracycline, a translation-inhibiting antibiotic, the authors identified pathways and genes associated with persister formation and maintenance. Combining this with a genome wide CRISPRi screening, the authors finally focused on three target genes (all involved in translation), including lon, yqgE, and rpoH. The roles that lon and yqgE play in increasing survival and prolonged lag phase were further confirmed using knockout and overexpression strains.

Despite the tremendous interest in the phenomenon of persistence, understanding the mechanisms for generating and maintaining this state have been elusive; thus, the authors are to be commended for this major effort to apply the very latest genomic technologies to elucidate persisters. Nevertheless, the originality and significance are debatable as translation and even specifically, the lon protease, have been previously described to be implicated in persistence. Overall, the CRISPRi work is the more functional, and thus revealing approach, though without much follow up on individual genes to reveal new biology.

Major points:

1. There is something quite circular about the study in that the persistence mutants examined (metG and hipA7) involve tRNA-synthetases involved in translation, with results pointing to translational deficiency. The study of wild-type cells after 6 days in stationary phase may well involve starvation of cells at that point, which would also point to translational deficiency. This is not to say that translation is not part of persistence, but this work can certainly not claim to be unbiased or comprehensive, and may partly explain the inability to uncover something truly novel, with translation previously well implicated.

2. Another possible reason that new no unique, specific (much less new) pathways were found to be

upregulated in persisters looking at the original scRNA seq data could be related to the quality of the scRNAseq data. As the median number of unique genes in many of the samples is below 50 (Fig S3), and the numbers of cells queried in each sample ranges from ~800-13K, the likely sparse coverage of the genome would impede the ability to confidently identify transcriptional states and call out significant genes/pathways. It would be helpful to actually know what is the coverage of the genome in the samples. (Pooling of samples for analysis in order to get better coverage is biased as it already presumes that they can be grouped together.)

3. Although the functional approach more clearly identifies specific genes and pathways in persister formation, no clear advance in biology is appreciated. While confirmation is provided for the role of Lon, which previously has been described albeit with a disputed role in persistence, and YqgE, a new gene, is identified, their roles are most implicated in the metG mutant and not wild-type. What is its more general significance or is this only in the context of specific mutants? Knockdown or overexpression of lon or yqgE alter the persistence phenotype, which could, if followed by the authors, reveal new biology related to persister cells.

Minor points:

1. The persistence phenotype studied is in lab growing– all the references used by the authors are studies that used lab growing/in vitro techniques and as such, its relevance to human infection is overstated.

2. Phrases like “comprehensive single-cell atlas” and “precise physiological and genetic bases of persistence” seem like overstatements.

3. It is unclear whether the persisters are formed in stationary phase or in lag phase. Generally speaking, type-I persisters are thought to form during starvation in stationary phase. Surprisingly, the “persister population” identified in this study is from the early lag-phase (Fig. 1d). Can the authors explain why the “persister cluster” was rarely identified in the stationary phase (green cells in Fig. 1d)?

4. The survival assay was measured using cells directly diluted into fresh medium with antibiotics (Fig. 1b), which corresponds to the green cells in Fig. 1d. The survival assay should also be measured using cells diluted into fresh medium and cultured for 20 mins, which corresponds to the orange cells in Fig. 1d. If the authors are correct and the persister cluster is true, a much higher survival rate should be observed. Thus, this additional experiment will be crucial for the validation of the “persister population.”

5. No biological replicates were performed for the most important sampling point, which was 5.1 hours cells and diluted then culture for 20 min (Fig. 1d, upper right panel).

6. From the pooled analysis (all samples included), it appeared that “persister population” is a homogeneous population. Are there enough cells with enough genes detected to allow unsupervised analysis for this population to show heterogeneity in this population?

7. Fig. 2d, e, g-i: please provide the UMAP plots for biological replicates, and the percentages of the "persister population" in each replicate. Without this information, it is difficult to assess how reproducible the "persister population" is in hipA and wilt-type strains.

8. The FACS experiments (Fig. S4) were important experimental validation for the "persister population." However, the clusters were defined by multiple genes, rather than one. A double-marker RNA FISH experiment is needed to validate these marker genes at single-cell and RNA level.

9. Since the single-cell data is available, the expression of identified marker genes, e.g., lon, yqgE, and rpoH in Fig. 4c, could be shown with expression and percentage of cells expressed in each cluster (FeaturePlot and dotplot function using Seurat). This information is important to assess the distribution of these marker genes and how specific they are expressed in the "persister population."

10. The authors concluded that persister cells are translational deficient, but it seems that these cells also have lower numbers of transcript per cell (Fig. S3c). How do you exclude the possibility that the translational deficiency was not caused by transcriptional deficiency?

11. Testing existence of a similar persister sub-population in other in vitro models such as the hipA7 mutant strain and 6-day WT culture is valuable and makes the persister cluster definition much more robust. What was less clear was how the "occupancy" in the persister cluster was defined. Did all 3.6% of hipA7 persister cells express all the markers that are expressed in the metG persisters or was it a certain fraction of markers that were common to both types of persisters?

12. Using tetracycline-induced persistence model is a clever way to distinguish between transient persisters and self-maintained persisters. It is interesting that there are only 6 genes common to hipA7 and metG persisters and only 2 common to all 3 models of self-maintained persisters (Fig S7a). Are all or some of these genes essential? Did the authors observe any phenotypes for knock down of these genes in the CRISPRi dataset described in the next section? Including CRISPRi or gene deletion data on some or all of these genes before moving to the genome-wide approach can help improve the flow and continuity of this manuscript. The readers now need to dig into supplemental data to figure out if any of the dysregulated genes in persisters were picked up the CALM screen.

13. The handful of genes described appear to be cherry-picked genes; it is hard to know the overall performance of the screen. It would be helpful to have some performance metric of the screen in order to get a sense of whether identified genes contain a lot of false positives or are likely real.

14. It is interesting that there were only 3 significant hits from the CRISPRi experiment and about 60 when stringency was reduced. Is it possible that most of the genes in these other pathways are essential and guides towards them were depleted before persisters testing steps (lag phase or lag phase-abx) in Fig 4a? Did the guide library include guides that caused a milder downregulation of essential genes and hence reduced dropout in the dcas9 induction phase in the CRISPRi protocol?

15. How many of the 61 CALM screen hits with the metG strain also confirm in the 6-day WT model or in the hipA7 model of persisters? This data can give a better sense of which genes are most

important to persister creation and survival.

Referee #2 (Remarks to the Author):

This study uses current technologies to investigate the physiology and underlying mechanism of E. coli persister cells. It is exciting to see that the authors could identify with single-cell scRNA a distinct subset of cells that pre-exists prior to antibiotic exposure. Additional CRISPRi screening revealed some mechanisms involved in forming persisters in a high-persister mutant.

I have a few comments:
1) A previous study shows that E. coli persisters are cells that have undergone a higher number of divisions when entering stationary phase (DOI: 10.1038/s41564-021-00900-4). These cells have a smaller size and thus retain fewer resources for re-initiating replication. This could explain the low transcription and translational activity of the persisters as well as their unremarkable gene expression profile. It seems important to clarify the mechanistic link between this previous study and the current findings.

2) Limited insights into the persister state. The authors observe an unremarkable transcriptome signature with no clear activation of any specific pathways. This might be in part because the main lag-promoting mechanisms might be a translational impairment according to the authors' model, and this is not directly visible in RNA data. Thus, more interesting differences might exist at the protein levels and turnover. The authors might consider FACS sorting of GFP reporters as shown in Fig. S4 followed by proteome analysis of subsets to characterize such differences. This would provide deeper insights into the persister state.

3) If the main difference of persisters is indeed impaired induction of defense against redox stress and iron uptake, low-oxygen iron-replete conditions would reduce the lag of these cells. This seems testable.

4) Universal state. The authors propose a universal ("archetypal") persister state but it remains unclear how similar the various persister subsets really are. The striking differences in upstream mechanisms could actually suggest that the metG* persister population are quite different from persisters in wild-type E. coli. The UMAP positions are interesting but might be driven by just a few genes. The pathway expressions are seemingly quite different (Fig. 3c) and this would need to be compared to respective non-persisters. The main markers for the metG* and hipA7 persisters are not shared with WT 6d persisters (Fig. S7a). The partial overlap shown in this figure could also be driven by comparing the different persister cells against tetracyclin-treated cells which are clearly heavily disturbed. Comparing to this special condition might let the various other persisters appear

more similar among each other. A more relevant comparison might be the overlap between the different strains for significantly different genes of persister cells vs. concurrent non-persisters.

5) E. coli K-12. The strain MG1655 used throughout the study has been passaged under laboratory conditions since its isolation in 1922. Lag-phases of this laboratory strain are likely not representative for more recent wild-type isolates as shown recently (doi: 10.1038/s41586-021-04114-w). This raises questions about the general relevance that the authors might want to discuss.

Referee #3 (Remarks to the Author):

In their manuscript, Blattman et al. study the phenomenon of antibiotic persistence in the lab strain of E. coli MG1655. They carry out two unbiased approaches, namely scRNAseq and a CRISRPRi screen, to define the state of antibiotic persistence and identify key genes in persister formation and maintenance. To facilitate their endeavor, they work, alongside wild-type E. coli with two hyper-persistent mutants, metG* and hipA7. Comparing the single-cell transcriptomes of the three strains (metG* and hipA7 after subculture post overnight, and wild-type post 6-day stationary phase culture), they define a universal state of persistence. Using the CRISPRi screen, they identify genes important for the levels of persisters and extent of lag phase in wild-type and metG* mutant.
This work follows up on previous work by this group, Balaban, and others showing that in E. coli MG1655 in the experimental setup used here, the length of the lag phase correlates with antibiotic persistence levels and metG* and hipA7 are hyper-persistent mutants. Authors should be lauded for their efforts in using scRNAseq and the high-density CRISPRi screen more extensively than before to characterize how E.coli heterogeneously exits the stationary phase. These are valuable approaches and information. While the CRISPRi screen is powerful, the scRNAseq is inherently limited by the number of mRNAs (30-50 per cell) that the technique gives access to but is nonetheless interesting. Thus, the descriptions provided here of how E. coli exits stationary phase, although largely descriptive, are important.
That said, the authors claim of having described a universal, or archetypal, persister state is not substantiated by their work, the altered translation is mostly documented by RNAseq on 30 transcripts and production of one fluorescent protein, the use of two mutants affected in the translation process as a way to access persisters is biased towards one path. The amalgam of mutant persisters to hypothetical wild-type as being an “archetypal” does not then stand when CRISPRi reveal that different genes influence persisters in wild-type vs metG*.
Finally, while this work advances the understanding of how E. coli heterogeneously exits stationary phase, this is by no mean leading to progress in understanding what underpins antibiotic persistence responsible for recurring infections and this should not be claimed as such. There is no evidence, anywhere, in this paper or others, that how fast or not E. coli MG1655 returns to exponential phase at the single cell level in artificial laboratory medium, has any relevance to how bacteria persist antibiotics during infection. The work shoud not be framed in what it is not.
Major concerns:
1- Authors chose to facilitate their approach by studying two hyper-persistent mutants, metG* and hipA7, which are mutated in aspects of translation. Therefore, it is unsurprising that the signature they document (on the basis of 30-40 mRNAs) is about altered translation, and they admit it

themselves. They perpetuate this with tetracycline treated cells. There is no evidence that these two mutants, or the tet treated populations, form more of the same persisters as wild-type. They may just form more of one type of cells that can persist (because they do not exit stationary phase as quickly because of their altered translation). In absence of scRNAseq of wild-type persisters, treated in the same conditions, not after 6 days of stationary phase, the biggest claim of the manuscript does not stand.
L76 and Fig 1d, they display 1,000 representative cells for wt which forms 0.01% persisters, which is extreme undersampling and makes it impossible to determine whether metG* and WT form the same state of persisters or not.
What is happening to the wt sample middle panel for the 2h subculture transcripts?
L82- “were likely persister cells” this needs to be qualified throughout. First, authors need to show these are persisters, by sampling ampicillin cells and show where they map, or proving that cells with this signature are the ones surviving antibiotics. Second, these are likely to be persister cells of metG* not wild-type. Authors equate these mutant persisters and wild-type persisters, which is an overreach. This shortcut is perpetuated throughout the paper.
2- 40 transcripts, although might be sufficient to obtain a gross signature (i.e. altered translation) does not define or describe extensively a transcriptional state. Many differences may be entirely missed here between strains and conditions. Such statement as that in l139 (and many other places in the ms) is therefore difficult to support.
3- L102-103 three markers of persister state are “validated” by transcriptional fusions. Rather, the expression of three genes detected by RNAseq are validated by transcriptional fusions. Three genes are not enough. Again, this is an overreach.
4- Fig 2e, why is the hipA7 mutant sampled at a different time than metG*?
5- L180-181 “given that tetracycline pharmacologically recapitulates the persister state” is a bold statement that I am uncertain of.
6- after describing a “universal” persister state, authors move on to show that different genes underpin the state of persistence in wt vs the metG* mutant. This contradicts the statement that the same state is adopted by these bacteria. Authors claim that by CRISPRi they catch the “paths” to the universal state, but these claims are not substantiated. Another likely explanation is that metG* persisters are of one kind, wt persisters of another, and therefore there is no “archetypal” persister state.
7- CRISPRi provides a list of genes important for wt and others for metG* persistence. It does not lead to any mechanistic understanding of such state. This is just descriptive. Many of these genes had already been found and reported by others (Lon, TCA related genes) as the authors openly describe.
8- YqgE is potentially interesting but has no impact on wt persisters. Its overexpression increases persistence in wt, which overexpression of so many genes does. Additionally, no mechanistic understanding of the role of YqgE in metG* persistence is provided.
Minor:
L15 (and elsewhere) “dormant” is ill defined, does it refer to production of proteins or lack of growth.
L29-30 “these approaches may not yield pure persister populations and prone to technical artifacts” is inaccurate. Authors do not need to invoke this to justify the use of scRNAseq.
L38-39 “identified a stereotyped persister transcriptional state” is inaccurate.

Author Rebuttals to Initial Comments:

We appreciate the reviewers' thoughtful assessment of our manuscript. We have thoroughly considered their critical insights and have carried out extensive additional experiments and analyses to address all remaining points of concern. The revised manuscript is significantly strengthened both in terms of novel observations and deeper mechanistic insights. In particular, we significantly advance understanding of persister cells by showing that, across five models (now including a clinical isolate), persisters are transcriptionally active, responsive to fresh nutrients, and converge to similar transcriptional states dominantly defined by translational deficiency rather than dramatic upregulation of specific known pathways. We have extended CRISPRi screening to encompass three distinct models, now including *hipA7.* The breadth of these studies has revealed a surprising principle: despite their convergent transcriptional states, different persister types form by distinct, yet also overlapping, upstream mechanisms. Even for *metG** and *hipA7*, two mutants with defects in tRNA-aminoacylation, the most critical driver genes are not the same. Specifically, for *metG** cells, persistence is not simply a result of biosynthetic defects from hypoactive methionine-tRNA ligase but rather due to amplification of downstream regulatory events mediated by Lon and YqgE. Through extensive additional experiments on the functions of Lon and YqgE, including an additional (fourth) CRISPRi screen, now of the *yqgE* overexpression phenotype, and proteomics of Δ*lon* cells, we propose a model for how the *metG** mutation triggers hyper-persistence. With the addition of these new studies, our work is unique in scope and scale, including, in multiple models, the first characterization of single-cell persister transcriptional states and comprehensive genetic interrogation of mechanisms of persister formation.

The key new experiments and analyses are outlined below and our detailed responses follow:

1. PETRI-seq of a uropathogenic *E. coli* (UPEC) clinical isolate before and after ampicillin, which reveals transcriptomes of pre-existing states and antibiotic-enriched persisters. These persisters converge to a similar state as all MG1655 persister models and tetracycline-treated cells.
2. PETRI-seq of wildtype MG1655 cells after ampicillin, which converge to a similar persister state as hyper-persistent models and tetracycline-treated cells.
3. PETRI-seq of *metG** cells after ampicillin treatment to confirm the identity of persister transcriptomes.
4. PETRI-seq of wildtype MG1655 during transcriptional inhibition by rifampicin. This important control demonstrates that assayed persister transcriptomes do not reflect transcriptional deficiency.
5. Bulk proteomics of wildtype and *metG** cells during stationary phase and after dilution into fresh media, which confirms that the molecular state of *metG** persisters is primarily defined by translational deficiency.
6. Bulk proteomics of *metG**-Δ*lon*-Δ*sulA* cells to discover targets of Lon in the context of *metG** hyper-persistence. Top hits include iron-sulfur cluster assembly proteins and iron-sulfur cluster binding proteins.
7. CRISPRi screen of *hipA7* cells, which reveals that major driver genes (*lon*, *yqgE, hipA*) are distinct across models. Other driver genes, particularly components of the TCA cycle, are common across models.
8. CRISPRi screen of *yqgE* overexpression in wildtype MG1655 cells and validation experiments demonstrating that the function of Lon is epistatic to that of YqgE.
9. Dual marker FACS to confirm co-expression of persister markers in *metG** persisters.
10. *In silico* prediction of YqgE as a protein disulfide isomerase.
11. Antibiotic survival assays with Lon inhibitor bortezomib to demonstrate small molecule potentiation of ampicillin and ciprofloxacin activity.

Referee #1 (Remarks to the Author):

Despite the tremendous interest in the phenomenon of persistence, understanding the mechanisms for generating and maintaining this state have been elusive; thus, the authors are to be commended for this major effort to apply the very latest genomic technologies to elucidate persisters. Nevertheless, the originality and significance are debatable as translation and even specifically, the lon protease, have been previously described to be implicated in persistence. Overall, the CRISPRi work is the more functional, and thus revealing approach, though without much follow up on individual genes to reveal new biology.

Response: We thank the reviewer for supportive comments and the important issue of novelty. We have now expanded the scope of our studies and discussion to bolster our case for the importance and novelty of what we find. We have added new data to show in a more unbiased way that multiple models of persistence, now including wildtype MG1655 (Fig. 2f,g) and a clinical UPEC isolate (Fig. S10i,j), generate persisters that converge to the same persister cluster as *metG** persisters and tetracycline-treated cells (reference for translational deficiency). In addition, experiments using rifampicin to mimic transcriptional deficiency support the specificity of low translation as the core feature of assayed persisters (Fig. S12f,g). It is true that aspects of translation have been previously implicated in persistence. However, we now show that a global state of translational deficiency is the dominant feature of lag-dependent persistence. We show this conclusively for *metG** persisters through transcriptomic, proteomic (new; Fig. S11), and genetic experiments. For four additional persister types, which are rare and thus challenging to study, convergence of transcriptional states demonstrates translational deficiency as the key common element across models (Fig. 2f, S10j).

Beyond translational deficiency, we find that persisters transcriptionally respond to fresh media (Fig. 1f,g), thereby diverging from a stationary-like transcriptome, but do not express lag-specific pathways (Fig. 1i,j). Furthermore, we find a number of genes and pathways differentially expressed between persisters and tetracycline-treated cells. Our revised manuscript more clearly presents these marker genes and the differences across persister models (*metG**, *hipA7*, wildtype) (Fig. S6, S10k, S13a; Table S1, S3). Among these, Lon and YqgE are uniquely expressed in and are causal to *metG** persistence (Fig. 4d,e), demonstrating how scRNA-seq can enable prediction of key drivers even in rare persisters.

We are also grateful to the reviewer for recognizing the important functional insights from the CRISPRi screen. In our revision, we have expanded this part to show a more comprehensive view across persister types. Specifically, we have added a CRISPRi screen of *hipA7* cells, a hyper-persistent mutant affecting tRNA-aminoacylation (similar to *metG**) (Fig. 4c). Surprisingly, even in these two mutants related to tRNA-aminoacylation, the most important causal genes are different. Lon protease and YqgE, the major drivers we found for *metG**, do not contribute to *hipA7* hyper-persistence (Fig. 4d). Instead, *hipA* itself is the only highly significant *hipA7*-specific driver gene (Fig. 4c). Together, our simultaneous investigation of 3 models of persistence (*metG**, *hipA7*, wildtype) shows several unique and overlapping driver genes (Fig. 4, S14d). Thus, an important emerging principle is that despite convergence to similar transcriptional states, upstream mechanisms of persister formation are diverse.

Finally, in our revision, we have added more follow-up experiments, detailed below, to interrogate the mechanistic contributions of Lon and YqgE. In particular, we go to great lengths to resolve the outstanding controversial role of Lon in persistence, which previously was attributed to accumulation of cell division inhibitor SulA. Building on our finding that, in the absence of SulA, Lon significantly affects wildtype persistence and is necessary for *metG** hyper-persistence, we have used proteomics to implicate Lon-mediated dysregulation of iron-sulfur cluster assembly and the TCA cycle as potential causal contributions of Lon to persistence (Fig. S19). Furthermore, we have discovered that Lon activity is epistatic to YqgE function (Fig. 5g), expanding our understanding of this novel regulatory pathway, which may act as a critical checkpoint to reduce translation upon nutrient deprivation (Fig. 5h, Discussion).

Major points:

1. There is something quite circular about the study in that the persistence mutants examined (metG and hipA7) involve tRNA-synthetases involved in translation, with results pointing to translational deficiency. The study of wild-type cells after 6 days in stationary phase may well involve starvation of cells at that point, which would also point to translational deficiency. This is not to say that translation is not part of persistence, but this work can certainly not claim to be unbiased or comprehensive, and may partly explain the inability to uncover something truly novel, with translation previously well implicated.

Response: The reviewer raises an important point here, and we have added new experiments to make our study more unbiased. First, we have added proteomics of *metG** cells after dilution into fresh media, which clearly shows that *metG** proteomes are frozen from stationary phase and therefore reflect translational

deficiency (Fig. S11a). Thus, our finding that translational deficiency is the dominant phenomenon in *metG** persisters is corroborated by independent evidence.

On the bigger question of circularity, we now include WT MG1655 and a uropathogenic isolate (UPEC) after ampicillin treatment, which remarkably enriches cells in the same persister cluster as *metG** (Fig. 2g, S10i). Despite limited resolution for these rare persisters, we find highly significant overlap between their marker genes and those of *metG** persisters or tetracycline-treated cells (Fig. S6f, S10k).

Our new data clearly highlight that convergence to a translationally deficient state is not just a property of hyper-persisting strains that happen to have mutations in genes related to tRNA charging but a general feature of 'wild-type' persisters as well. However, even for the *metG** allele, our extensive CRISPRi and validation studies demonstrate that hyper-persistence is not due to a trivial biosynthetic defect in translation, as shown by complete loss of hyper-persistence by *lon* deletion (Fig. 5a) and by reversal of low stationary phase translation by the *yqgE* deletion (Fig. 5h). Rather, translational deficiency is due to downstream regulatory events, likely reflecting a checkpoint (see Discussion). Thus, our findings go substantially beyond an anecdotal role for low translation in persisters to encompass the convergence of multiple persister models to a global attractor state dominantly defined by stalled translation.

Given the novelty of scRNA-seq for bacteria, we view this study as foundational to understanding the transcriptional state of persisters. Our finding that tetracycline treatment yields a transcriptome that clusters with persisters is an important step towards making sense of transcriptional patterns in persisters. The fact that all of our models of persistence (*metG**, *hipA7*, and the expanded WT and UPEC strains) converge to a similar translationally deficient transcriptional state is a highly significant finding. We have changed the language of the manuscript to acknowledge that while all of our persister models cluster together, the mechanisms of convergence to this state are diverse. Acknowledging their similarity and the role of translational deficiency is critical to understanding each persister type individually. In Fig. S6 and Fig. S13a, we have more clearly presented these similar and different expression patterns. We then focus most on understanding specifically the expression patterns of *metG**, while we anticipate future studies will dig deeper into the other mechanisms of persister generation.

Finally, we note that our revised manuscript only uses the word *comprehensive* to describe the CRISPRi screen, which is extremely high resolution. We use *unbiased* only in the beginning of the Discussion ("In this study, we used unbiased systems approaches to characterize *E. coli* persister states and genetic factors promoting persister formation and maintenance.") We are claiming here that scRNA-seq and CRISPR-interference are unbiased approaches not that the entirety of decisions made in this study are unbiased. We appreciate the reviewer's attention to the need for precise language.

2. Another possible reason that new no unique, specific (much less new) pathways were found to be upregulated in persisters looking at the original scRNA seq data could be related to the quality of the scRNAseq data. As the median number of unique genes in many of the samples is below 50 (Fig S3), and the numbers of cells queried in each sample ranges from ~800-13K, the likely sparse coverage of the genome would impede the ability to confidently identify transcriptional states and call out significant genes/pathways. It would be helpful to actually know what is the coverage of the genome in the samples. (Pooling of samples for analysis in order to get better coverage is biased as it already presumes that they can be grouped together.)

Response: We thank the reviewer for this suggestion, and we have now included coverage of samples and cell clusters used for differential expression analysis in Table S6. For *metG** persisters, we have ~200x coverage of all genes in the genome, which should be sufficient to find upregulated genes and pathways, particularly given that <100x coverage of the early lag cluster is sufficient to discover strongly upregulated pathways (Fig. 1i-j, S5f). For early lag cells, we also capture only ~26 mRNAs per cell (Fig. S3e).

We have also loosened our criteria for reporting persister-specific pathways, which led to the identification of PspF upregulated genes (Fig. 1h). Though this pathway is also expressed in early lag, it is more highly expressed in the persister cluster and gives insight into the cellular state as expression of these genes occurs during cell envelope stress. We include the following explanation in the text: *"We also looked for*

known pathways enriched in persisters relative to all surrounding clusters (stationary, early/late lag, early exponential). No known pathways met these stringent criteria, but a single gene set, genes upregulated by PspF, was significantly enriched in all comparisons except versus early lag (Fig. 1h). Upregulation by PspF typically occurs during cell envelope stress[1] and has been seen in E. coli persisters[2] and biofilms[3]."

Regarding pooling of samples, the differential expression analysis in Fig. 1h uses all *metG** and wildtype cells in the persister cluster, which we agree could obscure the signal of upregulated pathways. Thus, we carried out the same analysis with only *metG** samples after dilution into fresh media. This analysis gives the same single upregulated pathway as analysis including all of the *metG** and wildtype cells. We have added this information to the legend of Fig. 1h: "The cells used here include *metG** and wildtype cells in the persister cluster (Table S5, S6), but restricting to *metG** cells after dilution into fresh media recapitulates the finding of only this upregulated pathway." For all other differential expression analysis in the paper, we only use a single cell type (Fig. S6, S9k-l, S10k) for comparison. Though there is still a risk of biasing our signal by combining samples of the same cell type, this should also give the most robust signal. Differential expression analysis of each *metG** sample in Fig. S8e (8-48x coverage each) to neighboring clusters confirms PspF upregulated genes as a consistent (3/3) marker of the persister state, while other pathways are significant for just a single sample.

3. Although the functional approach more clearly identifies specific genes and pathways in persister formation, no clear advance in biology is appreciated. While confirmation is provided for the role of Lon, which previously has been described albeit with a disputed role in persistence, and YqgE, a new gene, is identified, their roles are most implicated in the metG mutant and not wild-type. What is its more general significance or is this only in the context of specific mutants? Knockdown or overexpression of lon or yqgE alter the persistence phenotype, which could, if followed by the authors, reveal new biology related to persister cells.

Response: We thank the reviewer for again challenging us to emphasize the novelty and significance of our findings. One of the key insights from our work is that despite convergence of transcriptional states across distinct models of persistence, it seems that there are diverse molecular mechanisms by which this is achieved. Substantiating this conclusion, we have added a CRISPRi screen of *hipA7*, which shows that HipA drives hyper-persistence through a lon- and yqgE-independent mechanism (Fig. 4c-d). One major contribution here is the functional interrogation of three models of persistence (wildtype, *metG**, *hipA7*) by the same methods to be able to compare causal genes (Fig. 4b-d, S14d). We find that upstream mechanisms are not universal, even for tRNA aminoacylation-related mutants, which suggests that there are diverse upstream mechanisms by which cells can achieve lag-dependent persistence. This, in itself, is an important finding which we emphasize in our revised Discussion.

Also, though Lon has been implicated before in persistence, it has been explained by accumulation of cell division inhibitor SulA. Here we show a SulA-independent contribution of Lon that is significant both in wildtype and *metG**, though of course the impact is greater in *metG** (Fig. 5a). In our revision, we have now added proteomics to identify the relevant targets of Lon (Fig. S19). Though we do not follow-up on these hits in depth, we provide a resource of candidate genes whose degradation may contribute to persister formation. Top candidates include iron-sulfur cluster assembly proteins (IscU, IscS) and iron-sulfur cluster binding proteins overall (Fig. S19a-b), depletion of which could be causal to extended lag times and persistence. In *metG** and wildtype, our proteomics data show that Lon deletion causes reduction in TCA cycle enzymes during stationary phase (Fig. S19c,d). As noted in the CRISPRi screen (Fig. 4g), reduction in TCA cycle decreases persistence across models. If Lon is critical for TCA cycle to be active, then loss of TCA cycle proteins upon Lon deletion could be one explanation for reduction in persistence.

Finally, we have added follow-up experiments exploring the role of YqgE. The reviewer correctly notes that YqgE is most important in the *metG** model, but we also find that YqgE overexpression extends lag times and increases persistence in wildtype (Fig. 5e, S17j). Furthermore, we find that YqgE directly or indirectly regulates translation during stationary phase (Fig. 5h), suggesting an important role for this protein outside the *metG** context. To explore the function of YqgE further, we have now added a CRISPRi screen in the YqgE overexpression strain (wildtype background). This screen and validation experiments reveal that Lon

is essential for the YqgE overexpression phenotype (Fig. 5g, S17k-l). We have also used the tool COFACTOR to predict *in silico* that YqgE is a disulfide isomerase (Fig. S20).

Overall, an important point about our functional study of *metG** is that *metG** hyper-persistence is not simply a result of a biosynthetic defect but instead is caused by downstream regulatory events via Lon and YqgE. In our Discussion, we propose a model in which YqgE acts as a checkpoint activating Lon to degrade critical proteins and shut down translation. In the absence of YqgE, baseline Lon activity could be sufficient to over time deplete proteins critical for ramping up translation upon addition of fresh nutrients.

Minor points:

1. The persistence phenotype studied is in lab growing– all the references used by the authors are studies that used lab growing/in vitro techniques and as such, its relevance to human infection is overstated.

Response: Thank you. We have changed the wording in the abstract and acknowledged this limitation of our work in the Discussion. Specifically, we have changed the final sentence of the Abstract from "*Our work defines the precise physiologic and genetic basis for persister formation at single-cell resolution, providing a rational foundation for targeted strategies against persisters in recalcitrant bacterial infections*" to "*Our work reveals key physiologic and genetic factors underlying starvation-triggered persistence, a critical step towards targeting persisters in recalcitrant bacterial infections*." We have also added the following to the Discussion: "*This convergent signature in persister transcriptomes implicates translational deficiency as a core feature of persistence even while active transcription occurs across persister models. Our work is done under laboratory conditions and primarily using non-pathogenic MG1655, but convergence of UPEC strain CFT073 persisters to the same persister cluster (Fig. S10i-j) suggests relevance of these findings to clinical strains*."

2. Phrases like "comprehensive single-cell atlas" and "precise physiological and genetic bases of persistence" seem like overstatements.

Response: Thank you. We have changed these statements. Instead of "comprehensive single-cell atlas", we now say "we generated a single-cell atlas of *E. coli* growth transitions". In the abstract, instead of saying "Our work defines the precise physiological and genetic bases of persistence," we now say "Our work reveals key physiologic and genetic factors underlying starvation-triggered persistence".

3. It is unclear whether the persisters are formed in stationary phase or in lag phase. Generally speaking, type-I persisters are thought to form during starvation in stationary phase. Surprisingly, the "persister population" identified in this study is from the early lag-phase (Fig. 1d). Can the authors explain why the "persister cluster" was rarely identified in the stationary phase (green cells in Fig. 1d)?

Response: We agree with the reviewer that this is an interesting result. Although we have found that transition into early stationary phase is critical for generating a long-tailed distribution of lag-times, the recurring persister transcriptional state we discover requires transition into fresh media with nutrients. Therefore, there is clear contingency on passing through stationary phase which eventually manifests as a bimodal population during lag-phase.

Our interpretation is that the persister transcriptional state reflects a response to fresh media (Fig. 1g) in addition to translational deficiency. In contrast, during stationary phase, *metG** and wildtype transcriptomes reflect the same state of starvation. Importantly, the difference between them is instead now apparent at the protein level, which we have assayed by proteomics (Fig. S11c, S19b,f). Our hypothesis, which is included in the Discussion, is that Lon-mediated degradation of critical proteins causes this divergence and results in extended dormancy upon addition of fresh nutrients.

4. The survival assay was measured using cells directly diluted into fresh medium with antibiotics (Fig. 1b), which corresponds to the green cells in Fig. 1d. The survival assay should also be measured using cells diluted into fresh medium and cultured for 20 mins, which corresponds to the orange cells in Fig. 1d. If the authors are correct and the persister cluster is true, a much higher survival rate should be observed. Thus,

this additional experiment will be crucial for the validation of the “persister population.”

Response: We thank the reviewer for this suggestion. We have added analysis related to this question to Fig. S4. In Fig. S4d, we show that after culturing for 20 minutes, the persistence rate is the same as immediately after dilution. Our model is that after dilution into fresh medium, persisters form regardless of the presence of antibiotics (Fig. S4e). In 20 minutes, very little growth of non-persisters occurs, and few persisters transition to a non-persister state, so the persistence rate does not change. Based on the PETRI-seq data, we see that *metG** cells can transition to the persister transcriptional state even in the presence of ampicillin (Fig. S4c), so for the survival assay in Fig. 1b, we are also looking at the persister cluster (Fig. 1f).

5. No biological replicates were performed for the most important sampling point, which was 5.1 hours cells and diluted then culture for 20 min (Fig. 1d, upper right panel).

Response: We agree that the 5.1 hour + 20 minute timepoint is critical. The chemostat experiment reveals the dynamics of persister emergence, and we show replication of these dynamics with antibiotic survival (Fig. 1b). However, for *metG** persister transcriptomes, we view overnight culture + 1-4 hours growth after dilution to be equally useful as replicates, particularly because we find the populations to be very consistent. Our revised manuscript individually shows 6 replicate experiments sampling *metG** persisters after overnight growth + dilution (Figs. 2a, S3a, S3b, S8e). The 3 populations in Fig. S8e, which were sampled ~1 hour after dilution, contain >74% cells in the persister cluster. These are thus similar to the homogenous population seen in the 5.1 hour + 20 minute condition in Fig. 1d.

6. From the pooled analysis (all samples included), it appeared that “persister population” is a homogeneous population. Are there enough cells with enough genes detected to allow unsupervised analysis for this population to show heterogeneity in this population?

Response: This is a great point, and actually we do find heterogeneity when we pool the persister population, as shown in Fig. 3c (top) and Fig. S6. We have changed the language in the manuscript to acknowledge that while multiple models converge to a similar persister state, it is not exactly the same state. Specifically, we have changed the title from “Discovery and genetic dissection of an archetypal persister cell state” to “Identification and genetic dissection of convergent persister cell states,” a conceptual change that is now reflected throughout the manuscript. We explicitly mention this in the Results section as well: “*Co-clustering of different persister types is an important signal of convergence of persister states but does not mean that gene expression is identical across models. While we find that marker genes and pathways significantly overlap, many genes and pathways also differ between persister models (Fig. S6, S9k-l, S10k). Like metG* persisters, hipA7 persisters do not upregulate lag-specific pathways. 6-day wildtype persisters upregulate one of three lag-specific pathways, oligopeptide transport genes (Fig. S9l).*” By demonstrating convergence of persister states but also presenting their differences, we believe the manuscript is substantially strengthened.

7. Fig. 2d, e, g-i: please provide the UMAP plots for biological replicates, and the percentages of the “persister population” in each replicate. Without this information, it is difficult to assess how reproducible the “persister population” is in hipA and wilt-type strains.

Response: We agree that this is important information for readers and have added every replicate with percentages in the persister cluster as Fig. S8.

8. The FACS experiments (Fig. S4) were important experimental validation for the “persister population.” However, the clusters were defined by multiple genes, rather than one. A double-marker RNA FISH experiment is needed to validate these marker genes at single-cell and RNA level.

Response: We appreciate this suggestion to improve the validation. We have added double-marker FACS, which confirms that the persister markers are upregulated together in the same cell populations (Fig. S7d-e). As further validation of the persister cluster, we have also treated *metG** cells with ampicillin and sequenced the remaining cells (Fig. 2a *right* and Fig. S4c). The cells after ampicillin treatment are highly enriched in the persister cluster, while the proportion of cells in the exponential cluster decreases (Fig. 2a).

This additional experiment supports that the multi-gene signature we find for the persister cluster does correspond to the persister phenotype of antibiotic survival.

9. Since the single-cell data is available, the expression of identified marker genes, e.g., lon, yqgE, and rpoH in Fig. 4c, could be shown with expression and percentage of cells expressed in each cluster (FeaturePlot and dotplot function using Seurat). This information is important to assess the distribution of these marker genes and how specific they are expressed in the "persister population."

Response: This is a helpful suggestion, and we have added plots with mean expression by cluster in Fig. S16. Because expression is roughly binary in single cells, percentage of cells expressing each gene is about the same as mean expression and would be redundant.

10. The authors concluded that persister cells are translational deficient, but it seems that these cells also have lower numbers of transcript per cell (Fig. S3c). How do you exclude the possibility that the translational deficiency was not caused by transcriptional deficiency?

Response: We thank the reviewer for this question. We have now explored this in depth in Fig. S12. First, we find that *metG** persisters have the same mRNAs per cell as tetracycline-treated wildtype cells (Fig. S12g). Thus, translational deficiency is sufficient to explain the reduction in transcripts seen for *metG** persisters. Additionally, we used rifampicin, an inhibitor of transcription, to demonstrate that persisters are transcriptionally active. Rifampicin-treated cells have enough rRNA to be detected by PETRI-seq (Fig. S12d), but they have effectively no mRNA (Fig. S12e). Using this as a reference, we defined cells as "transcriptionally deficient" if the mRNA ratio was below a certain threshold. We then find that while a small proportion of non-rifampicin-treated cells are below the transcriptional deficiency threshold, these cells are much less frequent than those in the orange persister cluster (Fig. S12f). *metG** and tetracycline-treated populations again exhibit similar levels of "transcriptional deficiency" by the analysis in Fig. S12f, supporting that reduction in mRNA transcripts can be explained by translational deficiency.

11. Testing existence of a similar persister sub-population in other in vitro models such as the hipA7 mutant strain and 6-day WT culture is valuable and makes the persister cluster definition much more robust. What was less clear was how the "occupancy" in the persister cluster was defined. Did all 3.6% of hipA7 persister cells express all the markers that are expressed in the metG persisters or was it a certain fraction of markers that were common to both types of persisters?

Response: We appreciate the opportunity to clarify this point. The occupancy in the persister cluster is defined by unsupervised clustering using all genes (Methods), so the marker genes are not expressed in all cells but instead are found by comparing population-wide expression levels. The reviewer asks an important question about how marker genes compare between persister types, and we have added more analysis to address this. In Fig. S6a-c, we have added differential expression analysis for each persister type relative to exponentially growing non-persisters. We have annotated the top markers for each persister type in the plots on the left. All genes with fold-changes and p-values are also listed in Table S1. These data show that *metG**, *hipA7*, and 6-day wildtype persisters have both common and distinct markers. Genes highly upregulated in all types include *rpoS*, *clpA*, and *nlpD*, all of which are also expressed in tetracycline. Overall, there is highly significant concordance between all three persister types ($p<10^{-14}$), though clearly the persister states are not identical. Fig. S6a-c also shows that many of the markers found by comparing persisters to exponential non-persisters can be explained by translational deficiency (green in right panels), a key point of the paper. Thus, we also include differential expression between tetracycline-treated cells and *metG**, *hipA7*, or 6-day wildtype persisters in Fig. S13a (and pathways in Fig. 3c and S13b). The Venn diagram in Fig. S13a shows again that some markers are shared between persister types, and some are unique. All two-way overlaps are highly significant (hyper-geometric test; $p<10^{-40}$). For comparisons vs. tetracycline, all fold-changes and p-values are listed in Table S3.

12. Using tetracycline-induced persistence model is a clever way to distinguish between transient persisters and self-maintained persisters. It is interesting that there are only 6 genes common to hipA7 and metG persisters and only 2 common to all 3 models of self-maintained persisters (Fig S7a).

Response: This is an important point that was not clear in our original manuscript. There are actually more shared hits than listed on the Venn diagram, which we have clarified in Fig. S13a (formerly S7) and its legend: "*Selected genes are annotated, and the number in each region gives the number of genes in that category (all listed in Table S3). All two-way overlaps are highly significant (hyper-geometric test; $p<10^{-40}$). The most significant markers shared by all persister types are ssrA, ompC, and clpA. metG* markers identified in Fig. S5d are also shown (mdtK, yhaM), as are the markers validated by FACS (Fig. S7a-e; cysK, rmf, mdtK).*"

12 (contd.) Are all or some of these genes essential? Did the authors observe any phenotypes for knock down of these genes in the CRISPRi dataset described in the next section? Including CRISPRi or gene deletion data on some or all of these genes before moving to the genome-wide approach can help improve the flow and continuity of this manuscript. The readers now need to dig into supplemental data to figure out if any of the dysregulated genes in persisters were picked up the CALM screen.

Response: We thank the reviewer for this suggestion. None of these genes are essential, and we were able to assay their functional importance. We have added these data to Fig. S15b, which now shows CRISPRi results for genes highly expressed in persisters (including those found in Fig. S13a [formerly S7a]). However, we see minimal effect from knockdown of these genes, which we note in the manuscript: "*Importantly, many top marker genes found by PETRI-seq did not prove functionally important by CRISPRi (Fig. S15b), demonstrating the importance of our dual approach and supporting a model in which causal genes are enriched but still a minority of those expressed*."

13. The handful of genes described appear to be cherry-picked genes; it is hard to know the overall performance of the screen. It would be helpful to have some performance metric of the screen in order to get a sense of whether identified genes contain a lot of false positives or are likely real.

Response: We agree that this information should be included. We selected 3 genes from the screen to validate, and 2 of them (*lon*, *yqgE*) were successful. We chose *lon* and *yqgE* as they are clearly the top hits from the aggregated screen data (Fig. 4b). 5 other genes (*rpoH*, *sucA*, *priA*, *dnaC*, *dnaT*) also shortened *metG** lag in all replicates (Fig. 4d, S15a). We considered all of these for validation and chose only *priA* for follow-up. Our reasoning is now in the manuscript:

> *"Of the genes not selected, sucA is part of the 2-oxoglutarate dehydrogenase complex, loss of which has already been shown to reduce lag-dependent persistence in wildtype cells[4]. dnaC, dnaT, and rpoH are essential genes[5], making validation challenging. They are also functionally related to selected genes (dnaC/T part of primosome with priA, lon induced by rpoH). As detailed below, deletion of lon or yqgE significantly reduced metG* lag-dependent persistence. However, deletion of priA did not affect metG* lag and was not pursued further (Fig. S17a)."*

Our false discovery rate is ~25-33%, depending on if we include *sucA*, which is validated by another study. However, we had expected *priA* could be a false positive, as the primosome is involved in plasmid maintenance, and our CRISPRi system is expressed from plasmids.

14. It is interesting that there were only 3 significant hits from the CRISPRi experiment and about 60 when stringency was reduced. Is it possible that most of the genes in these other pathways are essential and guides towards them were depleted before persisters testing steps (lag phase or lag phase-abx) in Fig 4a? Did the guide library include guides that caused a milder downregulation of essential genes and hence reduced dropout in the dcas9 induction phase in the CRISPRi protocol?

Response: We agree that the low number of highly reproducible hits was surprising and an interesting result. The uniquely strong effects of Lon or YqgE depletion point to a highly specific mechanism of persister formation. In our revision, we have updated the analysis pipeline and sequenced libraries for more coverage, so the number of hits has changed slightly. We now have 4 significant hits plus the 3 primosomal genes (which are likely confounded by plasmid effects). We then have 51 when stringency is reduced (Fig. 4f). The reviewer raises an important point about maintaining guides targeting essential genes. Our CRISPRi library targets every gene at an average of 56 positions (noted in Methods), including anti-sense and sense strands, which cause relatively strong and mild repression, respectively. High coverage and inclusion of sense gRNAs should reduce dropout. In fact, 3 of our 7 highly reproducible hits are essential

(*rpoH*, *dnaC*, *dnaT*). For *dnaT* and *dnaC*, we are able to detect their positive effect on lag time specifically by looking at milder downregulation from sense gRNAs (Fig. S15a).

In our revised manuscript, we have added a Venn diagram comparing the CRISPRi hits across genotypes (*metG**, *hipA7*, wildtype). Fig. S14d shows that there are 131 hits for wildtype, 51 for *metG**, and 56 for *hipA7*. It is interesting that there are more gene perturbations that can reduce lag time than in wildtype than in the hyper-persistent mutants.

15. How many of the 61 CALM screen hits with the metG strain also confirm in the 6-day WT model or in the hipA7 model of persisters? This data can give a better sense of which genes are most important to persister creation and survival.

Response: We are grateful for this question and have explored this more thoroughly in our revised manuscript. We have now added a CRISPRi screen to find persistence drivers in *hipA7* cells, which reveals distinct and overlapping causal genes between *metG** or *hipA7*. Based on the screen (Fig. 4c,d) and bortezomib treatment (Fig. S18e), *lon* and *yqgE* are not involved in *hipA7* hyper-persistence. Instead, *hipA* alone is the most important driver of *hipA7* persistence (Fig. 4c). Looking at all three cell types screened (*metG**, *hipA7*, wildtype), there is highly significant overlap between causal genes (Fig. S14d). We see that repression of TCA cycle consistently reduces lag times across genotypes (Fig. 4g, S14d) and thus could be an important universal target. However, our interpretation is not that common genes are necessarily most important. Instead, it seems that mechanisms of persister formation are diverse; the most important drivers for *metG** are Lon and YqgE, while for *hipA7* it is *hipA*. We have not looked into the major drivers of 6-day wildtype persisters, though we do find that *lon* and *yqgE* are not drivers (Fig. S17i).

Referee #2 (Remarks to the Author):

This study uses current technologies to investigate the physiology and underlying mechanism of E. coli persister cells. It is exciting to see that the authors could identify with single-cell scRNA a distinct subset of cells that pre-exists prior to antibiotic exposure. Additional CRISPRi screening revealed some mechanisms involved in forming persisters in a high-persister mutant.

Response: We thank the reviewer for supportive and constructive comments on our manuscript. Below, we have addressed all points through new experiments and/or additional discussion.

I have a few comments:
1) A previous study shows that E. coli persisters are cells that have undergone a higher number of divisions when entering stationary phase (DOI: 10.1038/s41564-021-00900-4). These cells have a smaller size and thus retain fewer resources for re-initiating replication. This could explain the low transcription and translational activity of the persisters as well as their unremarkable gene expression profile. It seems important to clarify the mechanistic link between this previous study and the current findings.

Response: We thank the reviewer for bringing this important study to our attention. We have added the following explanation to the Discussion:

> *"TCA cycle disruption significantly reduces lag times across models. For wildtype persistence, the effect of TCA cycle disruption is most profound (Fig. 4g) and is corroborated by similar effects from disrupting biosynthetic pathways that act on TCA cycle intermediates (Fig. 4i, S15c)[6]. Previous work has implicated the TCA cycle in persistence and attributed this to self-digestion to generate reducing power[4]. It has also been found that lag-dependent persisters undergo more divisions as a population enters stationary phase, which may lead to stochastic generation of cells with very low protein levels[7]. Taken together, the combination of self-digestion and reductive division could lead to rare cells with extremely low protein abundance and translation rates. Once fresh nutrients are available, positive feedback from translation can exacerbate small differences and establish bimodality in the population. Compounding this divergence, translationally deficient persisters do not appear to express important lag pathways, including redox defense and iron uptake (Fig. 1i-j, S5f, S9l)."*

2) Limited insights into the persister state. The authors observe an unremarkable transcriptome signature

with no clear activation of any specific pathways. This might be in part because the main lag-promoting mechanisms might be a translational impairment according to the authors' model, and this is not directly visible in RNA data. Thus, more interesting differences might exist at the protein levels and turnover. The authors might consider FACS sorting of GFP reporters as shown in Fig. S4 followed by proteome analysis of subsets to characterize such differences. This would provide deeper insights into the persister state.

Response:

We thank the reviewer for challenging us to emphasize the significance of our findings. Though the reviewer is correct that the transcriptional states of persisters demonstrate no significant activation of specific pathways and primarily reflect translational deficiency, we believe this result on its own is important. We note the following in the Discussion:

> "*Rather than dramatically upregulating specific known pathways, persisters seem to be in a dysregulated state in which transcriptional patterns primarily reflect collateral responses to low translation (Fig. 2j). This convergent signature in persister transcriptomes implicates translational deficiency as a core feature of persistence even while active transcription occurs across persister models.*"

In our revised manuscript, we have added PETRI-seq of wildtype MG1655 (Fig. 2g) and clinical UPEC (Fig. S10i) persisters, which support the convergence of many persister types to states of translational deficiency. Thus, we discover that a global state of translational deficiency is the dominant feature of lag-dependent persistence. Additionally, we note other important patterns in the persister transcriptomes. Persisters transcriptionally respond to fresh media (Fig. 1f,g), thereby diverging from a stationary-like transcriptome, but do not express lag-specific pathways (Fig. 1i,j). We also find a number of genes and pathways differentially expressed between persisters and tetracycline-treated cells. Our revised manuscript more clearly presents these marker genes and the differences across persister models (*metG**, *hipA7*, wildtype) (Fig. S6, S10k, S13a; Table S1, S3). Among these, Lon and YqgE are uniquely expressed in and causal to *metG** persisters (Fig. 4d,e), demonstrating how scRNA-seq can enable prediction of key drivers even in rare persisters.

Additionally, the reviewer makes a great suggestion to use proteomics for deeper insights into the persister state, and we have now added proteomics data to the revised manuscript. Specifically, we assayed proteomes of *metG** cells during stationary phase and 30 minutes after dilution into fresh media. It would be challenging to obtain enough cells by sorting, so we opted to take *metG** cells < 1 hour after dilution when they are almost homogenously in the persister state (Fig. S8e). For reference, we also assayed wildtype cells in stationary phase or 90 minutes after dilution (exponential). We projected the bulk proteomes with the single-cell RNA atlas (Fig. S11a-b), which reveals that *metG** persister proteomes are frozen in a stationary-like state. This finding supports our translational deficiency model. In addition, comparing *metG** persisters to stationary proteomes yields only 7 significantly upregulated proteins, 3 of which are cold-shock proteins known to be expressed when translation is low (noted in Fig. S11d legend). Together, the transcriptomics and proteomics implicate translational deficiency as the major explanation for the cellular state of *metG** persisters. For the other persister types (*hipA7*, wildtype), it would be much more challenging to isolate persisters for proteomics, so the scRNA-seq is a powerful tool to detect translationally deficient persister cells. In light of *metG** persister proteomes being minimally changed relative to stationary cells, it is significant that persister transcriptomes do change enough after stationary phase to become a distinct cluster.

In addition to using proteomics to look at persister cells, we reasoned that stationary phase proteomes could be informative of proteins functionally important during persister formation. We know that Lon protease is a necessary driver of *metG** hyper-persistence (Fig. 5a) and that its action is critical during stationary phase (Fig. 5c). Thus, we expect that Lon degradation of specific targets is key to persister generation, and those targets should be stabilized by Lon deletion. In Fig. S19, we compare proteomes of stationary *metG*-Δlon-ΔsulA* to stationary *metG** cells to discover multiple enriched proteins. We also compared wildtype to *metG** stationary cells; if degradation of a protein is causing *metG** hyper-persistence, then that protein should not also be degraded in wildtype cells (Fig. S19b). Top candidates in both comparisons include iron-sulfur cluster assembly proteins (IscU, IscS) and iron-sulfur cluster binding

proteins overall (Fig. S19a-b), depletion of which could be causal to extended lag times and persistence. Though it is beyond the scope of our study to follow-up on these hits further, we anticipate these proteomics data will be an important resource for future studies. Interestingly, we also find that Lon deletion decreases expression of TCA cycle proteins, which is likely one component of how Lon deletion decreases persistence particularly in wildtype cells (Fig. S19c-d; Fig. 4g). All dysregulated proteins are listed in Table S4.

3) If the main difference of persisters is indeed impaired induction of defense against redox stress and iron uptake, low-oxygen iron-replete conditions would reduce the lag of these cells. This seems testable.

Response: This is an important point, and we have added more explanation to the manuscript. Though we have not tested it, previous work has shown that anaerobic conditions during stationary phase decrease lag-dependent persistence, which is now noted in the main text ("*Knockout of oxyR has previously been shown to be sufficient to extend lag times in E. coli*[8]*. Conversely, non-oxidizing anaerobic conditions decrease lag times*[4]*.*"). Additionally, we did not mean to say the main difference of persisters is failure to turn on lag-specific processes (redox defense, iron uptake). Instead, we are proposing that failure to turn on these processes exacerbates persister dormancy, which we later tie to translational deficiency. We have revised the Discussion, specifically in the excerpt included in response to the reviewer's first point, to propose this model in which translational deficiency puts cells in a challenging position because only more protein production can ramp up cellular processes, repair damage, and allow growth to resume.

4) Universal state. The authors propose a universal ("archetypal") persister state but it remains unclear how similar the various persister subsets really are. The striking differences in upstream mechanisms could actually suggest that the metG* persister population are quite different from persisters in wild-type E. coli. The UMAP positions are interesting but might be driven by just a few genes. The pathway expressions are seemingly quite different (Fig. 3c) and this would need to be compared to respective non-persisters. The main markers for the metG* and hipA7 persisters are not shared with WT 6d persisters (Fig. S7a). The partial overlap shown in this figure could also be driven by comparing the different persister cells against tetracyclin-treated cells which are clearly heavily disturbed. Comparing to this special condition might let the various other persisters appear more similar among each other. A more relevant comparison might be the overlap between the different strains for significantly different genes of persister cells vs. concurrent non-persisters.

Response: We are grateful for this suggestion to further explore and clarify the shared and distinct expression patterns between the persister types. As the reviewer suggested, we have now added a comparison between persisters and concurrent non-persisters in Fig. S6a-c with all source data in Table S1. The three persister types share many markers, including top hits *clpA*, *rpoS*, and *nlpD*. We have also highlighted gene sets up- or down-regulated in persisters relative to concurrent non-persisters (Fig. S6a-iv,b-iv,c-iv). Many of these differentially expressed genes and pathways can be explained by translational deficiency, as they are also upregulated during tetracycline treatment (Fig. S6a-iv,b-iv,c-iv). For this reason, we also highlight differential expression between persisters and tetracycline treatment. In Fig. S13a (formerly S7a), we show the Venn diagram of unique and shared hits, which the reviewer mentioned. To make these data more accessible, we include all of the source data in Table S3.

The reviewer also asks a deeper question: is a state universal if many markers between cell types are not shared? In the text, we have softened this claim and instead present both the similarities and differences between the persister types as useful information to make sense of this complex phenotype. Specifically, we have changed the title from "Discovery and genetic dissection of an archetypal persister cell state" to "Identification and genetic dissection of convergent persister cell states," a conceptual change that is now reflected throughout the manuscript. We explicitly mention this in the Results section as well: "*Co-clustering of different persister types is an important signal of convergence of persister states but does not mean that gene expression is identical across models. While we find that marker genes and pathways significantly overlap, many genes and pathways also differ between persister models (Fig. S6, S9k-l, S10k). Like metG* persisters, hipA7 persisters do not upregulate lag-specific pathways. 6-day wildtype persisters upregulate one of three lag-specific pathways, oligopeptide transport genes (Fig. S9l).*" By demonstrating convergence of persister states but also acknowledging nuance, we believe the manuscript is substantially strengthened.

5) E. coli K-12. The strain MG1655 used throughout the study has been passaged under laboratory conditions since its isolation in 1922. Lag-phases of this laboratory strain are likely not representative for more recent wild-type isolates as shown recently (doi: 10.1038/s41586-021-04114-w). This raises questions about the general relevance that the authors might want to discuss.

Response: We thank the reviewer for this suggestion, which led us to expand our study to include the uropathogenic clinical isolate (UPEC) strain CFT073. In Fig. S10a, we show that CFT073 exhibits higher antibiotic survival and more cells with long lag times than MG1655. We then used PETRI-seq to capture CFT073 cells 1 hour after dilution from stationary phase or after treatment with ampicillin (Fig. S10d-j). In the 1-hour population, we saw cell clusters corresponding to fimbriae or flagellin expression (Fig. S10d,e), but after integrating CFT073 cells into the MG1655 atlas, we also found elevated occupancy in the same persister cluster as *metG** for UPEC CFT073 [relative to wildtype MG1655] (Fig. S10h,j). After ampicillin, we then see a dramatic increase in cells in this persister cluster (Fig. S10i,j). Despite limited resolution (few cells, few counts per cell), differential expression analysis finds patterns consistent with other persisters (translation low, IHF upregulated genes high; Fig. S10k, S6a-c). Additionally, genes upregulated in UPEC CFT073 persisters significantly overlap with those expressed in MG1655 cells in tetracycline, again pointing to an overarching phenomenon of translational deficiency (Fig. S10k).

We have also added Discussion about this, particularly acknowledging that our work is done under laboratory conditions and primarily using MG1655. In the Discussion, we have written: "*This convergent signature in persister transcriptomes implicates translational deficiency as a core feature of persistence even while active transcription occurs across persister models. Our work is done under laboratory conditions and primarily using non-pathogenic MG1655, but convergence of UPEC strain CFT073 persisters to the same persister cluster (Fig. S10i-j) suggests relevance of these findings to clinical strains.*"

Finally, as we note in the introduction, in a previous study, "the *hipA7* mutation was found in 5% of a small sample of clinical *E. coli* isolates[9]", which also suggests relevance of MG1655 hyper-persistent mutants to infection.

Referee #3 (Remarks to the Author):

In their manuscript, Blattman et al. study the phenomenon of antibiotic persistence in the lab strain of E. coli MG1655. They carry out two unbiased approaches, namely scRNAseq and a CRISRPRi screen, to define the state of antibiotic persistence and identify key genes in persister formation and maintenance. To facilitate their endeavor, they work, alongside wild-type E. coli with two hyper-persistent mutants, metG* and hipA7. Comparing the single-cell transcriptomes of the three strains (metG* and hipA7 after subculture post overnight, and wild-type post 6-day stationary phase culture), they define a universal state of persistence. Using the CRISPRi screen, they identify genes important for the levels of persisters and extent of lag phase in wild-type and metG* mutant.
This work follows up on previous work by this group, Balaban, and others showing that in E. coli MG1655 in the experimental setup used here, the length of the lag phase correlates with antibiotic persistence levels and metG* and hipA7 are hyper-persistent mutants. Authors should be lauded for their efforts in using scRNAseq and the high-density CRISPRi screen more extensively than before to characterize how E.coli heterogeneously exits the stationary phase. These are valuable approaches and information. While the CRISPRi screen is powerful, the scRNAseq is inherently limited by the number of mRNAs (30-50 per cell) that the technique gives access to but is nonetheless interesting. Thus, the descriptions provided here of how E. coli exits stationary phase, although largely descriptive, are important.
That said, the authors claim of having described a universal, or archetypal, persister state is not substantiated by their work, the altered translation is mostly documented by RNAseq on 30 transcripts and production of one fluorescent protein, the use of two mutants affected in the translation process as a way to access persisters is biased towards one path. The amalgam of mutant persisters to hypothetical wild-type as being an "archetypal" does not then stand when CRISPRi reveal that different genes influence persisters in wild-type vs metG*.
Finally, while this work advances the understanding of how E. coli heterogeneously exits stationary phase, this is by no mean leading to progress in understanding what underpins antibiotic persistence responsible for recurring infections and this should not be claimed as such. There is no evidence, anywhere, in this

paper or others, that how fast or not E. coli MG1655 returns to exponential phase at the single cell level in artificial laboratory medium, has any relevance to how bacteria persist antibiotics during infection. The work shoud not be framed in what it is not.

Response: We thank the reviewer for supportive comments and for important feedback on the boundary of what we can claim given our observations. In the following responses, we address all the points and their remedies, including additional experiments/analyses or further discussion.

In our revised manuscript, we have now added scRNA-seq of non-mutant strains, which supports translational deficiency as a common mode of persistence. Additionally, we have used bulk proteomics to demonstrate extremely low translation in *metG** cells. Thus, though the reviewer is correct that scRNA-seq of persisters is inherently low-resolution for each cell, our hypothesis that the transcriptome primarily reflects translational deficiency is supported by other modalities, increasing confidence that scRNA-seq provides accurate information about heterogeneous populations. That said, we agree with the reviewer that "archetypal" is a bold claim that cannot be sufficiently supported, and, as such, we have changed the title and wording throughout the manuscript. Specifically, we have changed the title from "Discovery and genetic dissection of an archetypal persister cell state" to "Identification and genetic dissection of convergent persister cell states," a conceptual change that is now reflected throughout the manuscript. We explicitly mention this in the Results section as well: "*Co-clustering of different persister types is an important signal of convergence of persister states but does not mean that gene expression is identical across models….*" By demonstrating convergence of persister states but also acknowledging nuance, we believe the manuscript is substantially strengthened.

Similarly, we appreciate the reviewer's comments about the limits of studying laboratory strains in conditions that do not reflect infection. We have added PETRI-seq of a uropathogenic *E. coli* (UPEC) isolate to our study to begin to bridge the gap, but we agree that while our experimental setup sheds light on how *E. coli* exits stationary phase, the relevance to clinical infection is not explored here. We have changed the language of the manuscript to acknowledge these limitations and nuance. We have changed the final sentence of the abstract from "*Our work defines the precise physiologic and genetic basis for persister formation at single-cell resolution, providing a rational foundation for targeted strategies against persisters in recalcitrant bacterial infections*" to "*Our work reveals key physiologic and genetic factors underlying starvation-triggered persistence, a critical step towards targeting persisters in recalcitrant bacterial infections*." We have also added the following to the Discussion: "*This convergent signature in persister transcriptomes implicates translational deficiency as a core feature of persistence even while active transcription occurs across persister models. Our work is done under laboratory conditions and primarily using non-pathogenic MG1655, but convergence of UPEC strain CFT073 persisters to the same persister cluster (Fig. S10i-j) suggests relevance of these findings to clinical strains*."

Finally, although it is unclear whether the timing of stationary phase exit for MG1655 is relevant to recalcitrant infections, we note that *hipA7* mutations, which cause lag-dependent hyper-persistence, have been isolated from infections. We mention this in the introduction: "*As evidence that similar evolution occurs during treatment of infection, the hipA7 mutation was found in 5% of a small sample of clinical E. coli isolates*[9]*."*

Major concerns:
1- Authors chose to facilitate their approach by studying two hyper-persistent mutants, metG* and hipA7, which are mutated in aspects of translation. Therefore, it is unsurprising that the signature they document (on the basis of 30-40 mRNAs) is about altered translation, and they admit it themselves. They perpetuate this with tetracycline treated cells. There is no evidence that these two mutants, or the tet treated populations, form more of the same persisters as wild-type. They may just form more of one type of cells that can persist (because they do not exit stationary phase as quickly because of their altered translation). In absence of scRNAseq of wild-type persisters, treated in the same conditions, not after 6 days of stationary phase, the biggest claim of the manuscript does not stand.

Response: We appreciate the reviewer's attention to this important point and have now added significant additional experimental data to bolster our case for convergence to a translationally deficient state for

wildtype persisters. Indeed, capturing wildtype persisters is challenging due to their extremely low frequency. However, we have now conducted additional experiments including scRNA-seq profiling of wildtype MG1655 (Fig. 2f-g) and a uropathogenic clinical isolate (UPEC CFT073) (Fig. S10i-j) following ampicillin treatment. Although ampicillin treatment complicates the analysis due to the confounding effects of dead cells and possible transcriptional responses to ampicillin itself, dramatic enrichment in the persister cluster indicates that surviving cells are captured and cluster with *metG** persisters. Because these persisters are so rare, we could only capture ~40 cells. Nonetheless, differential expression analysis finds common patterns (translation downregulated, genes induced by IHF upregulated) between wildtype (MG1655, CFT073) cells after ampicillin (Fig. S6f, S10k) and other persister models (Fig. S6a-c). Also, genes upregulated in wildtype (MG1655, CFT073) cells after ampicillin significantly overlap with those expressed in tetracycline-treated cells (Fig. S6f, S10k; noted in legends). Overall, these additional data support that transcriptional states of wildtype and even clinical persisters are similar to hyper-persisters *and* exhibit the transcriptional signature of translational deficiency.

Furthermore, we strongly agree with the reviewer that we should soften our claim of "universality" which we have done in the text and title. However, we hope that these additional observations bolster our case for convergent transcriptional states of lag-dependent persisters across hyper-persister mutants and wild-type strains. Our study is the most comprehensive single-cell phenomenological characterization and genetic interrogation of lag-dependent persistence and, as such, serves to establish an experimental framework for studying other modes of persistence.

L76 and Fig 1d, they display 1,000 representative cells for wt which forms 0.01% persisters, which is extreme undersampling and makes it impossible to determine whether metG* and WT form the same state of persisters or not.

Response: We agree that it is impossible to see wildtype persisters in Fig. 1d. Our main goal in this figure is to show the emergence of *metG** persisters. Wildtype populations are included for reference, which is why we fixed every population to the same number of cells.

However, the reviewer raises an important point, which we have now addressed in Fig. S8. Specifically, we show 7 replicates of wildtype MG1655 45-60 minutes after dilution from overnight. For comparison, we again show the same number of cells on every plot, here 1700 each (11,900 total; 40,245 in full dataset on GEO). The fraction in the persister cluster is ~1%, which is close to the survival rate seen if MG1655 is diluted into media + ampicillin and incubated for ~45 minutes (Fig. S8b). Elsewhere in the paper, we show the 0.01% persistence rate (measured after 4 hours treatment; Fig. 1b, 2b, S1a-b) because killing reaches a plateau by that point. However, persisters can be short-lived and quickly switch into a non-persister state.

As discussed above, we have also now added PETRI-seq of wildtype cells (MG1655 and UPEC CFT073) after ampicillin treatment, which are dramatically enriched in the persister cluster (Fig. 2f-g, S10i-j).

What is happening to the wt sample middle panel for the 2h subculture transcripts?

Response: We agree that this was confusing. The sample was not taken because of constraints during the experiment. We have added a note to the legend of Fig. 1d: "In the bottom middle panel, the wildtype 2-hour population was not sampled and thus is not shown."

L82- "were likely persister cells" this needs to be qualified throughout. First, authors need to show these are persisters, by sampling ampicillin cells and show where they map, or proving that cells with this signature are the ones surviving antibiotics. Second, these are likely to be persister cells of metG* not wild-type. Authors equate these mutant persisters and wild-type persisters, which is an overreach. This shortcut is perpetuated throughout the paper.

Response: We appreciate the reviewer's suggestion to sample cells from ampicillin. For *metG**, we have now added PETRI-seq of cells before and after ampicillin (Fig. 2a). Comparison of the UMAPs on the left and right shows that ampicillin clears the exponential cells (blue and purple) and enriches the persister cluster (orange).

We have also removed the line mentioned by the reviewer. Instead, we write, "*We hypothesized that this distinct population corresponds to metG* persisters, as it emerges during lag phase precisely at the transition to hyper-persistence… Sequencing of metG* cells after ampicillin treatment confirmed our annotation of the unique persister cluster (Fig. 2a).*"

Finally, we appreciate the reviewer's comment about being precise in our language about mutant and wildtype persisters. As mentioned, we have added PETRI-seq of wildtype (MG1655 and UPEC) cells after ampicillin and found that these persisters cluster with *metG** persisters (Fig. 2f-g, S10i-j). However, resolution is limited with only ~40 cells captured. We are now clearer in the text about what we can claim about each persister type and the differences between them. For all of the persister types, we show that they cluster with tetracycline-treated cells and thereby find that their transcriptomes reflect translational deficiency (Fig. 2, S10j). The other claim we make about all persister types is that they are in a transitional state between stationary and exponential phase; this comes from the positions of all persister types in the middle of the UMAP (Fig. 1f), which correspond well with the PCA in Fig. S5c.

For specific gene and pathway expression patterns, we are also explicit about differences between persister types. In the Results section, we include the following:

> *"Co-clustering of different persister types is an important signal of convergence of persister states but does not mean that gene expression is identical across models. While we find that marker genes and pathways significantly overlap, many genes and pathways also differ between persister models (Fig. S6, S9k-l, S10k). Like metG* persisters, hipA7 persisters do not upregulate lag-specific pathways. 6-day wildtype persisters upregulate one of three lag-specific pathways, oligopeptide transport genes (Fig. S9l)."*

2- 40 transcripts, although might be sufficient to obtain a gross signature (i.e. altered translation) does not define or describe extensively a transcriptional state. Many differences may be entirely missed here between strains and conditions. Such statement as that in l139 (and many other places in the ms) is therefore difficult to support.

Response: We thank the reviewer for pointing this out and have now modified the title and text to acknowledge this limitation and nuance. We have changed the specific line in question to "*We conclude that metG* and hipA7 persisters converge to similar cell states*."

However, we also have evidence that 40 transcripts per cell gives a lot of information about a cell state. Looking at early lag cells as a reference, we capture a median of 26 mRNAs per cell but still find multiple strongly upregulated pathways (Fig. 1i-j, S5f). Our dataset includes 13,000 early lag cells and >30,000 *metG** persister cells (Table S6), so the resolution for *metG** persisters should be even higher.

Furthermore, we have added experiments to bolster our finding that translational deficiency is the dominant explanation for the observed persister transcriptome. Specifically, bulk proteomics of *metG** persisters reveal a stationary-like proteome that has not responded to addition of fresh nutrients (Fig. S11a-d). Given that persisters transcriptionally respond to fresh media (Fig. 1f-g, Fig. S11f), the proteome should change in response, but in this case, extremely low translation prevents adaptation of the proteome. The reviewer is correct that our per-cell resolution may be limited, but the proteomics support our conclusion that translational deficiency is the defining feature of the *metG** persister state. For the four additional persister types, which are rare and thus challenging to study, convergence of transcriptional states demonstrates translational deficiency as the key common element across models (Fig. 2f, S10j).

Additionally, we challenged our model further by testing transcriptional deficiency as an alternative explanation for the persister state (Fig. S12). Treating cells with rifampicin, which inhibits RNA polymerase, demonstrates that translational, not transcriptional, perturbation can explain the persister transcriptomes (across models; Fig. S12f). Finally, even with low per-cell resolution, we find many common and distinct genes upregulated across different persister types, which are shown in Figs. S6, S10k, and S13a.

3- L102-103 three markers of persister state are “validated” by transcriptional fusions. Rather, the expression of three genes detected by RNAseq are validated by transcriptional fusions. Three genes are not enough. Again, this is an overreach.

Response: The reviewer is correct, and we have updated the text accordingly: “Top genes upregulated in *metG** persisters include *rmf*, *cysK*, and *mdtK* (Fig. S6a-*i*), expression of which we validated using transcriptional fusions[10] (Fig. S7a-e).”

4- Fig 2e, why is the hipA7 mutant sampled at a different time than metG*?

Response: We have changed Figure 2 to show the *metG** cells before and after ampicillin (Fig. 2a). It is now critical to show cells 3 hours after dilution (Fig. 2a, *left*), since at that point the population is split between persisters and exponential cells, making enrichment of persisters clear after ampicillin. However, the reviewer raises an important point about visibility of all replicates with clear labeling of the timepoints. In Fig. S8, we have added all replicate experiments used for Fig. 2 with the time of outgrowth noted for all of them. Most samples were taken after ~1 hour, and there is a *metG** 52-minute sample for best comparison (Fig. S8e, 3rd panel). For *metG**, 52 minutes vs. 2 hours after dilution does not make a large difference since most cells are in the persister cluster and not growing. For *hipA7*, most cells are non-persister growing cells, so sequencing after 1 hour makes more sense than 2 hours, after which point exponential cells will be ~4x more abundant (2 doublings).

5- L180-181 “given that tetracycline pharmacologically recapitulates the persister state” is a bold statement that I am uncertain of.

Response: We agree and have now softened the wording here – “Given that tetracycline pharmacologically recapitulates a persister-like state…”.

6- after describing a “universal” persister state, authors move on to show that different genes underpin the state of persistence in wt vs the metG* mutant. This contradicts the statement that the same state is adopted by these bacteria. Authors claim that by CRISPRi they catch the “paths” to the universal state, but these claims are not substantiated. Another likely explanation is that metG* persisters are of one kind, wt persisters of another, and therefore there is no “archetypal” persister state.

Response: We agree with the reviewer. In light of this feedback, we have modified our interpretation to describe convergence of the transcriptional states across persister models. Our nuanced claim is now that despite diverse upstream mechanisms of persister formation, multiple models of lag-dependent persistence converge to transcriptional states that cluster together and reflect translational deficiency.

7- CRISPRi provides a list of genes important for wt and others for metG* persistence. It does not lead to any mechanistic understanding of such state. This is just descriptive. Many of these genes had already been found and reported by others (Lon, TCA related genes) as the authors openly describe.

Response: We thank the reviewer for the opportunity to emphasize the significance of our findings. The CRISPRi screens provide global and specific insight into persister mechanisms. Globally, a key insight from our work is that despite the common convergence of distinct models of persistence to transcriptional states reflecting translational deficiency, it seems that there are diverse molecular mechanisms by which this is achieved. Substantiating this conclusion, we have now added a CRISPRi screen of *hipA7*, which shows that *hipA* drives hyper-persistence through a *lon*- and *yqgE*-independent mechanism (Fig. 4c-d). One major contribution here is the functional interrogation of three models of persistence (wildtype, *metG**, *hipA7*) by the same methods to be able to compare causal genes (Fig. 4d, S14d). We find that upstream mechanisms are not universal, even for tRNA aminoacylation-related mutants, which suggests that there are diverse upstream mechanisms by which cells can achieve lag-dependent persistence. This, in itself, is an important finding which we emphasize in our revised Discussion.

Regarding the specific genes and pathways we discover, the reviewer is correct that some have been implicated before. For TCA-related genes, our novel finding is that this gene-set is unique in affecting all three models of persistence (Fig. 4g, S14d), suggesting it may be a particularly impactful target.

Also, though Lon has been implicated before in persistence, it has been explained by accumulation of cell division inhibitor SulA. Here we show a SulA-independent contribution of Lon that is significant both in wildtype and *metG**, though of course the impact is greater in *metG** (Fig. 5a). In our revision, we have added proteomics to identify the relevant targets of Lon (Fig. S19). Though we do not follow-up on these hits in depth, we provide a resource of candidate genes whose degradation may contribute to persister formation. Top candidates include iron-sulfur cluster assembly proteins (IscU, IscS) and iron-sulfur cluster binding proteins overall (Fig. S19a-b), depletion of which could be causal to extended lag times and persistence. In *metG** and wildtype, our proteomics data show that Lon deletion causes reduction in TCA cycle enzymes during stationary phase (Fig. S19c-d). As noted in the CRISPRi screen (Fig. 4g), reduction in TCA cycle decreases persistence across models. If Lon is critical for TCA cycle to be active, then loss of TCA cycle proteins upon Lon deletion could be one explanation for reduction in persistence.

Finally, we have now added follow-up experiments exploring the role of YqgE, which is a novel persister gene. YqgE is most important in the *metG** model, but we find that YqgE overexpression extends lag times and increases persistence in wildtype (Fig. 5e, S17j). Also, YqgE directly or indirectly regulates translation during stationary phase (Fig. 5h), suggesting another important role for this protein outside the *metG** context. To explore the function of YqgE further, we have added a CRISPRi screen in the YqgE overexpression strain (wildtype background). This screen and validation experiments reveal that Lon is essential for the YqgE overexpression phenotype (Fig. 5g, S17k-l). We have also used the tool COFACTOR to predict *in silico* that YqgE is a disulfide isomerase (Fig. S20).

Overall, an important point about our functional study of *metG** is that *metG** hyper-persistence is not simply a result of a biosynthetic defect but instead is caused by downstream regulatory events via Lon and YqgE. In our Discussion, we propose a model in which YqgE acts as a checkpoint activating Lon to degrade critical proteins and shut down translation. In the absence of YqgE, baseline Lon activity could be sufficient to over time deplete proteins critical for ramping up translation upon addition of fresh nutrients.

8- YqgE is potentially interesting but has no impact on wt persisters. Its overexpression increases persistence in wt, which overexpression of so many genes does. Additionally, no mechanistic understanding of the role of YqgE in metG* persistence is provided.

Response: The reviewer raises important points here, and we have revised the manuscript with new insight into YqgE, which is detailed in the previous response. Though overexpression of many genes can increase persistence, our newly added CRISPRi screen in YqgE+ cells adds nuance to this finding and reveals an important functional relationship between YqgE and Lon even in the absence of *metG** mutation.

Minor:
L15 (and elsewhere) “dormant” is ill defined, does it refer to production of proteins or lack of growth.

Response: We agree and have defined dormancy now as lack of growth: “*Antibiotic tolerance is the ability of bacteria to survive high concentrations of antibiotics, often by assuming a dormant, or non-growing, state.*”

L29-30 “these approaches may not yield pure persister populations and prone to technical artifacts” is inaccurate. Authors do not need to invoke this to justify the use of scRNAseq.

Response: We thank the reviewer for this important point. We have removed this statement from the manuscript.

L38-39 “identified a stereotyped persister transcriptional state” is inaccurate.

Response: Thank you. This has been changed to “Within this atlas, we identified a unique persister cluster that is common across five genetic and physiological models of persistence.”

References

1 Darwin, A. J. Stress Relief during Host Infection: The Phage Shock Protein Response Supports Bacterial Virulence in Various Ways. *PLoS Pathogens* **9**, e1003388 (2013).

2 Vega, N. M., Allison, K. R., Khalil, A. S. & Collins, J. J. Signaling-mediated bacterial persister formation. *Nature Chemical Biology* **8**, 431 (2012).

3 Beloin, C. *et al.* Global impact of mature biofilm lifestyle on Escherichia coli K-12 gene expression. *Molecular Microbiology* **51**, 659-674 (2004).

4 Orman, M. A. & Brynildsen, M. P. Inhibition of stationary phase respiration impairs persister formation in E. coli. *Nature Communications* **6**, 7983 (2015).

5 Goodall, E. C. A. *et al.* The Essential Genome of *Escherichia coli* K-12. *mBio* **9**, e02096-02017 (2018).

6 Berg, J. M., Tymoczko, J. L. & Stryer, L. (New York: WH Freeman, 2012).

7 Bakshi, S. *et al.* Tracking bacterial lineages in complex and dynamic environments with applications for growth control and persistence. *Nature Microbiology* **6**, 783-791 (2021).

8 Johnson, J. R., Clabots, C. & Rosen, H. Effect of Inactivation of the Global Oxidative Stress Regulator oxyR on the Colonization Ability of Escherichia coli O1:K1:H7 in a Mouse Model of Ascending Urinary Tract Infection. *Infect. Immun.* **74**, 461-468 (2006).

9 Schumacher, M. A. *et al.* HipBA–promoter structures reveal the basis of heritable multidrug tolerance. *Nature* **524**, 59 (2015).

10 Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat Methods* **3**, 623-628 (2006).

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

All concerns have been addressed

Referee #2 (Remarks to the Author):

The authors should be commended for their careful attention to my comments and questions. Critical issues have been clarified, in part through extensive new experimental data sets. I thus have no crucial reservations.

Previously, I expressed concerns about the representativeness of the lab strain MG1655 for clinical strains. The authors now include data for CFT073, a clinical strain from a patient with pyelonephritis. While this is indeed a clinical isolate, it has undergone extensive laboratory passages since at least 1989 (DOI: 10.1128/iai.58.5.1281-1289.1990). This long in vitro passaging has clearly altered its biology in complex ways (DOI: 10.1128/spectrum.02085-23). Therefore, it remains questionable how representative these results are for fresh isolates that are clinically most relevant.

However, both MG1655 and CFT073 have been used as models for numerous studies, and a better understanding of their persister phenotypes is of broad interest. It would be sufficient to mention that this study focused on heavily passaged strains, and further research is needed to determine if similar mechanisms operate in fresh isolates.

Referee #3 (Remarks to the Author):

I appreciate the authors' diligence in addressing my comments and those of other reviewers and thank them for amending their conclusions and text in response to the reviews.
While I am impressed by the vast amount of work they added to their manuscript and their rigor in the revisions, this only addresses part of my comments. I remain convinced that the WT data are underpowered (they added more but still only report the transcriptome of very few cells) and that scRNAseq only picks up a very narrow number of genes. I understand the authors' arguments that it is already informative, but I stand by my comment that at the moment it is only the tip of the iceberg. In addition, although the authors added screen to the high-throughput information, there is still a lack of mechanistic understanding of the persister states. The techniques are implemented to the highest standards but are not new, and the biological or even more mechanistic inputs are limited.

Author Rebuttals to First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

All concerns have been addressed

Response: We thank the referee for their careful review of our paper.

Referee #2 (Remarks to the Author):

The authors should be commended for their careful attention to my comments and questions. Critical issues have been clarified, in part through extensive new experimental data sets. I thus have no crucial reservations.

Previously, I expressed concerns about the representativeness of the lab strain MG1655 for clinical strains. The authors now include data for CFT073, a clinical strain from a patient with pyelonephritis. While this is indeed a clinical isolate, it has undergone extensive laboratory passages since at least 1989 (DOI: 10.1128/iai.58.5.1281-1289.1990). This long in vitro passaging has clearly altered its biology in complex ways (DOI: 10.1128/spectrum.02085-23). Therefore, it remains questionable how representative these results are for fresh isolates that are clinically most relevant.

However, both MG1655 and CFT073 have been used as models for numerous studies, and a better understanding of their persister phenotypes is of broad interest. It would be sufficient to mention that this study focused on heavily passaged strains, and further research is needed to determine if similar mechanisms operate in fresh isolates.

Response: We thank the reviewer for their kind comments and careful attention to our work. We appreciate this comment about extensive laboratory passaging of CFT073. We have added the following to the discussion: "As MG1655 and CFT073 strains have both been passaged for decades[51], it will be important for future research to extend scRNA-seq to fresh clinical isolates."

Referee #3 (Remarks to the Author):

I appreciate the authors' diligence in addressing my comments and those of other reviewers and thank them for amending their conclusions and text in response to the reviews.
While I am impressed by the vast amount of work they added to their manuscript and their rigor in the revisions, this only addresses part of my comments. I remain convinced that the WT data are underpowered (they added more but still only report the transcriptome of very few cells) and that scRNAseq only picks up a very narrow number of genes. I understand the authors' arguments that it is already informative, but I stand by my comment that at the

moment it is only the tip of the iceberg. In addition, although the authors added screen to the high-throughput information, there is still a lack of mechanistic understanding of the persister states. The techniques are implemented to the highest standards but are not new, and the biological or even more mechanistic inputs are limited.

Response: We are grateful to the reviewer for their thoughtful comments and careful reading of our manuscript. We believe that the number of cells and capture-rates are adequate to support our key inferences and conclusions. However, we do point out possible limitations in the regime of low persistence-rate:

"Together, these data support shared transcriptional programs in wildtype and metG* persisters, though resolution at this low persistence rate is limited."