Cell Systems

Transparent Peer Review Record

Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer

Lea Schuh, Michael Saint-Antoine, Eric M. Sanford, Benjamin L. Emert, Abhyudai Singh, Carsten Marr, Arjun Raj, and Yogesh Goyal

Summary

Initial Submission: Received July 16, 2019
Preprint: https://doi.org/10.1101/704247
Deposited on bioRxiv, July 18, 2019
Scientific editor: QuinceyJustman, Ph.D.

Editorial Decision after first round of review: Revision invited September 9, 2019
Major changes anticipated
Revision received February 3, 2020

Editorial Decision after second round of review: Accepted March 24, 2020

Editor's View

Scientific context and initial assessment. Many of us are drawn to biology because it seems to do special and surprising things and have characteristics that beggar the imagination. To me, the question has always been "How?" How does biology work? This question can be posed sharply, in terms of mechanistic specifics, but when it's abstracted, it demonstrates where a scientist thinks causation originates inside a biological system. I'll push this idea to a playful extreme for illustration's sake: Choose a team.

Team Master Regulator. Master Regulators drive biological function. They are in the right place at the right time. They exert disproportionate power over phenotype by sensing, integrating information, coordinating responses, and ensuring specificity. According Team Master Regulator, large cellular processes are a series of precisely orchestrated hand-offs between Master Regulators of sub-processes.

Master Regulators are molecular-level puppeteers of supreme intelligence if the words commonly used to describe their activities are given the weight of fact.

Team Dumb Luck. Dumb Luck drives biological function. This team argues that while biology might look special, special-looking things aren't, in fact, surprising. Instead, there is something about the way biology is wired, something about how probabilities play out in biological contexts, that allows special-looking things to happen at some characteristic frequency without any apparent biological intention. Specific regulation is needed comparatively rarely; most things work because in the cell, chances are that they will.

My own personal hunch is that despite the fact that almost all textbooks, review articles, and experiments lean on Master Regulators in their conception and description of biological phenomena, Dumb Luck is more prevalent and important than it might otherwise seem. Demonstrating that, though, is very, very hard. In a biological experiment, how are we to ensure that a Master Regulator is not lurking outside our field of view, sensing and responding in manner that links cause with effect?

This is one place math and modeling can really shine. We know what is and isn't in a model because we've made deliberate choices (there are no lurkers). The behavior of a conservatively constructed, minimal model can be explored in depth; parameters and architectures can be swept systematically to gain and develop intuition. This sort of work can set baseline expectation for biological behavior, e.g. in principle, given this limited set of reasonable assumptions, Interesting Behavior X ***can*** happen in the cell.

Cell Systems is deeply interested in work that develops principled baseline expectations for biological behavior. That's precisely what Schuh et al. did in this manuscript. A recent Nature paper from Arjun Raj and colleagues (https://doi.org/10.1038/nature22794) found a special and surprising biological phenomenon: about 1 in 3000 human melanoma cells enter a physiological state that is "pre-resistant" to chemotherapy. These cells express a handful of phenotypically important genes simultaneously, at a high level. The authors call this gene expression pattern a "rare coordinated state;" it occurs with defined statistics but apparently at random within the cellular population. In present manuscript, using a very conservative set of assumptions, Schuh et al. demonstrate that bursty transcription is sufficient to generate these rare coordinated states; additional regulation is not necessary. That doesn't mean there isn't additional regulation in the cell, it just means that in principle, there doesn't have to be. This conclusion is fundamentally important. Score one for Team Dumb Luck!

Review strategy and Editorial Decisions. Cell Systems has an exceptional and continually growing community of reviewers who understand and appreciate the goals of manuscripts like these. They approach reviewing manuscripts with open minds, incisive creativity, and deep expertise. Choosing reviewers from this community was straightforward, but that doesn't mean that this manuscript's life was straightforward. On the contrary. The first round of reviews was pretty tough. But I had a hunch that I was seeing something familiar and fixable that sat at the base of what looked like pervasive and quite damning problems.

Researchers live and breathe their work. That's one of the things that allows science to be extraordinary, but it can also cause miscommunication during peer review. The way authors conceive of their work is so

intimately familiar that they often forget to explain their conceptions to reviewers. Those conceptions need to be explained: they're very highly developed and well-honed. As a result, both the authors' and the reviewers' intentions can get muddled.

This happened in several places in this manuscript, but I'll highlight two. First, the authors were actually making an argument about ***sufficiency*** (i.e. "in principle, this minimal set of relationships is sufficient to generate this phenomenon inside the cell") ***not exclusivity*** (i.e. "this minimal set of relationships is exclusively responsible for this phenomenon inside the cell"). This wasn't clear enough in the first version of the paper, and it got them into trouble. Second, from the author's perspective, their focus on symmetrical networks was deliberate, important, and correct. This wasn't clear to the reviewers though, who objected to it as biologically unreasonable and over-simplified. Resolving this difference was a fundamentally important goal of the revision.

After the authors took a bit of time to digest the reviews, we had really wonderful meeting-of-the-minds over Zoom, during which Arjun memorably described epigenetics as "spooky," as in, "you don't need spooky epigenetics to achieve rare coordinated states." Happily, as Reviewer 2 said of the revision, "The authors have taken the reviews seriously and used them to improve the manuscript enormously. Now it is an important contribution to the field." We agreed and were very glad to accept this thought-provoking paper.

Editorial decision letter with reviewers' comments, first round of review

Dear Arjun,

I hope this email finds you well. I'm enclosing the comments that reviewers made on your paper, which I hope you will find useful and constructive. As you'll see, they express interest in the study, but they also have a number of criticisms and suggestions. Based on these comments, it seems premature to proceed with the paper in its current form; however, if it's possible to address the concerns raised with additional experiments and/or analysis, we'd be interested in considering a revised version of the manuscript.

As a matter of principle, I usually only invite a revision when I'm reasonably certain that the authors' work will align with the reviewers' concerns and produce a publishable manuscript. I'll be honest, though, that in the case of this manuscript, I'm unclear whether we're aligned. It's not a matter of whether this

manuscript ***can*** be revised to meet our bar, but rather whether you ***want*** to revise it along the lines we'd need. I'll do my best to articulate our vision for where we think this manuscript should go and you can make an informed choice about whether you'd like to keep it at *Cell Systems*. As always, we can always talk over the phone if it's helpful, especially with respect to making that choice.

Essentially, to move forward at *Cell Systems*, we'd need the revision to ring two bells:

1. We would need to understand the mathematical basis for the phenomenon in question (e.g. its critical dependencies, the relationships that drive it, etc. essentially mechanistically). One difficult thing about this manuscript is that you are really focussing on two phenomena rather than one:

> **A**) the existence of rare, coordinated states (along with clear, mathematical definitions of "rare" and "coordinated"), and
>
> **B**) the idea that these rare, coordinated states require a network larger than a given size and with a sufficient degree of connectivity.

My instinct says that understanding how rare, coordinated states arise (that is, **A**) is a very difficult task on its own and that trying to do both **A** and **B** in a single paper is probably too much. *Cell Systems* would be perfectly happy with a paper that focuses on **A** and interrogates it to the standards I've laid out below with the reviewers' help. I'll note, though, that there's some distance between where this manuscript is right now and where the revision would need to be. To give you a sense of where the gaps in understanding **A** are right now, here's my current read: I understand that you're able to recapitulate the existence of rare, coordinated states using the formalism you've chosen and I can follow your arguments, but I don't understand enough about how those states arise to make sense of the observation that their dependencies are so sharp. Why should they be so sharp and specific? For example, the observed sensitivity to 3 model parameters -- where does the sensitivity come from, why those three, why is the behavior insensitive to the others? Or the observed connectivity requirements >/= specifically six -- why a discontinuity and why does it happen between 5 and 7?

(As an aside, I'll also mention that I think making too much of **B** is risky. We have a series of papers coming out this month that look at network behavior across computationally assembled "atlases" of networks with a fixed number of nodes (i.e. connectivities are systematically enumerated and parameters are swept; this approach pays homage to https://www.ncbi.nlm.nih.gov/pubmed/19703401). That's hard even when the network behavior is comparatively simple and the networks are comparatively small. I understand that your work aims to do something different and that you've made choices that simply and limit the "space" of network topologies you're looking at, but if you choose to make arguments along the lines of **B** in the revision, they'd need to meet the same standards as these forthcoming published papers. Accordingly, if you want to expand on **B** in the revision, we'd likely need to ask at least one additional reviewer to take a look at that section of the manuscript. At present, there are too many moving parts to make reviewing **B** productive, but I hope this is resolved during revision. I'm sorry this is a bit unusual.)

2. We would need the mathematical basis of the rare, coordinated states needs to be generalizable in some sense. That is, the critical dependencies/relationships underlying this phenomenon shouldn't be the consequences of the specific, largely arbitrary choices made during modeling. Here, the reviewers have a lot to say, as I've highlighted in blue below (you'll also note particularly salient comments are highlighted in yellow). The reviewers' comments are part of the reason that I think focusing on **A** exclusively is a paper-unto-itself -- addressing them reasonably comprehensively in a reasonably narrow set of circumstances is already quite challenging.

I hope you find this feedback helpful. If you choose to move forward with *Cell Systems*, it's important that you and I stay on the same page. I'm always happy to talk, either over email or by phone, if you'd like feedback about whether your efforts are moving the manuscript in a productive direction. If you have any questions or concerns, please let me know. More technical information and advice about resubmission can be found below my signature. Please read it carefully, as it can save substantial time and effort later.

I look forward to seeing your revised manuscript.

All the best,

Quincey

Quincey Justman, Ph.D.
Editor-in-Chief, Cell Systems

Reviewers' comments:

Reviewer #1: Arjun's lab recently made an important observation that many jackpot genes are expressed in cells that are predisposed to being resistance to drug therapy. One of the questions that arises from this surprising observation is that how are these jackpot genes turned on in a subpopulation of cells. The paradox is that these genes are not that correlated in their expression level in the jackpot cells, yet tend to be enriched in the resistant cells. This manuscript tackles this fundamental question with simulations of networks of coupled and stochastic genes and showed that similar behavior as observed in experiments are seen in symmetric networks. They further showed that the coordinated stochastic behavior arises only in networks with a certain level of connectivity.

One of interesting findings is that entrance of the jackpot state is initiated by large burst size of one gene, whereas exit is not correlated with burst sizes. The authors further examined the parameter space and found that Kon, Koff and Kadd are the 3/7 parameters that contribute to the generation of the jackpot states. It is also illuminating to read the discussion about how variations of the models including non-symmetric and negative interactions will impact the results.

Understanding the phase space of gene networks and how they contribute to phenotypes is important to understanding cell type differentiation. This is an important contribution to the field and I strongly support the publication of this paper.

A few minor questions and comments. First, because the model can be applied to any gene network, would we expect to see these coordinated high states for many genes? And are the jackpot genes just happen to be important for drug resistance? Or are there something special about the architecture of these genes?

Second, it might be helpful to illustrate some of the networks used in the simulation in Figure 2. It is not entirely clear to me all the network topologies used in the simulation.

Reviewer #2: In this manuscript Schuch et al. present a theoretical framework to provide a mechanistic understanding of an increasingly well known phenomenon, namely the association of transient states of coordinated gene expression with persistent states of drug resistance in bacteria and mammalian cells. There is much that is interesting here but there is also much that is either missing or not properly reasoned which results in a missed opportunity. If the authors were to do more work, there could be something of general interest here.

The work is introduced in an odd grand manner that sets up a very high level frame of reference for what ends up being a very special case of an interesting but rare situation. Going from Taleb to cancer and drug resistance in a few paragraphs is a good thing in a review but not in a specialized research paper which is what this is. Importantly, the manuscript is motivated by experimental work (Shaffer et al, 2018 and Torre et al. 2019) but there is no proper explicit reference to the details of this work. The data appears to be scattered with ad hoc references to it as needed by the models. This is not right, the authors should state up front what are the features that are being modeled and then discuss the approach with reference to it. Rare states of coordinated gene expression in cancer cells may well do for the title but it is not enough for the development of the manuscript. The impression that one gets is that this, or a better organized version of this, could have been the modeling part of either of the two experimental works but, even for this, it would need more work.

I now raise a few issues that I would expect the authors to address if they want to capitalize on what is a good start to an interesting topic. The list is not exhaustive but hopefully highlights some of the problems that the manuscript has.

Introduction (which needs to be structured and elaborated in a very different manner more in accord with what the paper is about)

As already pointed out, the introduction is very wayward with sweeping statements on many topics and not suitable for a research report such as this one. It also contains some surprising statements such as the one referring to noise based models of the same phenomenon "Specifically, the classical models have largely described the variability that results in relatively normally distributed counts of mRNAs of a

given gene per cell". This is not true as the literature shows that these models give rise to log-normal and gamma distributions that are far from normal and often match the experiments. It is very worrying that this is stated as part of the introduction as it raises issues about how much the authors understand the problem.

In the introduction the two opposing models are introduced in different places (Fig1A and FigS1A) which is unhelpful and distracting.

They should also justify the restriction of the model to transcription and why they don't explore translational coupling since it has been shown that this can alter the distribution of cell states by creating biases in the cell states (as defined by the presence of protein products). I appreciate that they are modeling transcriptional data which, as also already stated, is not properly introduced but, when dealing with networks, proteins and translation matter. Their model appear as only having mRNA as the product and the regulator.

Results
An important issue is that they need to be very precise about the definition of a 'highly correlated' state. This will help understand what the work is about and what is being discussed and addressed. At the moment it is not. We are told that experimentally there is no correlation between the genes that characterize the state and this seems to be supported by the ad hoc reference to data (FigS1F) but then their networks assume some correlation. Even the basic model in which one gene activates another has this correlation implicit. Need clarification and explanation.

Page 7: they should justify the ergodic assumption since, intuitively, it will be broken by the entry into the rare stable states. They might argue that in the very large sampling they are doing this will apply but, does a cell have access to a similar phase space?

Some of the results presented in Figure 2 raise questions. In Figure 2B we are shown a comparison between data and simulation in terms of the genes coexpressed in these rare states. A superficial viewing suggests a similarity but it is obvious that the experimental data has a tail which is lacking in the simulation. One can see that there is a non-zero number of cells with 8 highly expressed genes. Need an explanation.

Figure 2C suffers from a similar problem as, in detail, the distributions are different. It would be helpful if to show similarity or differences, they would do a semi-log plot.

Another important issues arises in 2D where one wonders if this the result is not a trivial consequence of the intuitive fact that as the number of nodes in the network increases, the number of states increases. Maybe if they normalized to network size they might get a different result.

In this regard the limitation of the network size to <8 is not properly justified or explained. Looking at how ron is calculated (page 19) the limitation could be a trivial consequence of the addition that is used in the calculation. Above six nodes, the system migh blow up. They claim but not show that ≥8 networks will also display coordinated states but will they? And how does the number correlate with the experimental

data? Could the authors discuss?

Even admitting that it was clear what exactly they are trying to model, or better find an explanation to, the reliance on simulations obscures some of the potential explanations. One example is on Figure 3A. Surprisingly there appears to be no correlation between the maintenance of expression and stopping the burst e.g take the 'black gene', it is surprising that the expression is maintained for so long after the first long burst. How can this happen? The same applies to the others. If there is a memory in the system, where does it come from?

Along the same lines, it would be helpful if they could dig into the causes for the exit from the state and present some experimental data. And, of course, what determines the duration of the state?

They discussion of the parameters is confusing: there are 8 parameter sets but they only discuss 3. It would be good to know what the others are. Also, what is the relationship of these 8 parameters from the simulation to the 7 free parameters of the model? One can do some work, interpret the authors and know what they are talking about but they should tell us. Also, they should be more explicit about the point of the screen.

And, indeed, it would be good to have a phase space of the screens, parameters.

What is the importance of the frequency of the bursts in the entry into and maintenance of the state?

Discussion
The discussion seems to be focused in reinforcing some prejudices derived from their analysis and providing some excuses for work that should have been done. Here is an important paragraph:
"One limitation of our model is that we have performed quantitative analysis only on symmetric networks with positive interactions between nodes. It is likely that our findings hold more generally for asymmetric networks, as partially demonstrated for two cases of randomly selected asymmetric networks". If this is the way they feel, they should prove it. Also they state that "Inhibitory interactions between nodes is a separate and perhaps more interesting point. In principle, the model can be adapted to include inhibitory interactions" This is very important in light of the possibility that networks with <8 nodes will explode. They should definitely explore the influence of inhibitory interactions as this is a glaring omission in the analysis.

Overall, a very interesting topic but a very casual and, to a certain degree, superficial analysis. The paper needs to change in structure and also have more in depth analysis and a better judgement of what is being explained and what are the assumptions; at the moment they seem to be very biased by their assumptions.

Reviewer #3: The manuscript by Schuh et al investigates stochastic genetic network models that are capable of producing rate transient coordinated expression states as observed experimentally in melanoma. The authors focus on a specific class of genetic networks with transcriptional bursting and

computationally explore a wide range of network size, topology and parameters. They find these networks below a threshold of connectivity are capable of producing behaviour that is similar to experimental data. They also identify this model behaviour is sensitive to values of certain parameters in the model but not others. Interestingly, they validate using network inference on experimental data a transition from low connectivity to high connectivity that is associated with transcriptionally stable states. Overall, this is an interesting paper with novel results. However, the effect of some model assumptions and also the mechanistic origin of the observed behaviour is not fully clear. I have the following specific comments:

- The model used has several assumptions, it is not clear what is the effect of these assumptions on the model behaviour. The model, ignores protein and translation, but transcriptional regulation works through proteins. Including proteins explicitly is computational expensive, but some exploration of this would be useful. Also, the models only include transcriptional activation and only consider additive interaction. What would happen if you relax these assumptions?

- While the effect of model parameters is explored, the topology of the networks are not explored fully (only at the level of connectivity). Are there specific model topologies that produce the rate transient expression more robustly that others? Do these tend to contain positive feedbacks?

- Do you have any mechanistic insight on the origin of this behaviour? Why some parameters are critical and some are not relevant?

- The model uses non mass action kinetics. Have you used a variant of the Gillespie algorithm to handle this and what kind of approximation are you making? Could this affect your results.

- How much do we know about the specific genetic network, their players and the wiring in melanoma. Could you argue the network and its parameters is in the operating regime you have obtained.

- Please rephrase this sentence in the introduction, it is not clear to me what you are saying: "However, in this classical context, most of the cells … "

- Several references are missing full journal information, e.g Shaffer et al 2018, Saint Anthoine And Singh 2019, Corrigan et al 2016 Symons et al and Torre et al. Please check all references carefully.

Authors' response to the reviewers' first round comments

Attached.

Revised manuscript

Attached.

Editorial decision letter with reviewers' comments, second round of review

Hello, everyone,

Thanks very much for your patience while I was sick last week. I'm very pleased to let you know that the reviews of your revised manuscript are back and we're looking forward to accepting your manuscript "in principle" (that is, pending our receipt of your final, properly formatted files) as soon as a few minor changes are made. Congratulations!

In addition to the final comments from the reviewers, I've made some suggestions about your manuscript within the "Editorial Notes" section, below. Please consider my editorial suggestions carefully, ask any questions of me that you need, make all warranted changes, and then upload your final files into Editorial Manager. ***We hope to receive your files within 5 business days.*** Please email me directly if this will be a problem.

I'm looking forward to going through these last steps with you. More technical information can be found below my signature, and please let me know if you have any questions.

All the best,
Quincey

Quincey Justman, Ph.D.
Editor-in-Chief, Cell Systems

Editorial Notes

Abstract: Your abstract is too long; please give it another try. Note that most effective abstracts have the following structure:

> *[One sentence of background.] However, X. [X essentially presents the problem that you will solve.] Here, we [solve X]. [A compound sentence that begins with a methodological phrase and ends with a phrase that describes the results that the methodology produces (e.g. "Using a combination of method Y and method Z, we show something new about X.").] [Sentences that*

*describe key results and include carefully selected details that allow the reader to place the key results within a broader context.] [Optional: Include a sentence that "zooms out" to suggest future experiments or demonstrate impact.] (**150 words, maximum, *not including* the Transparent Peer Review statement**)*

For a complementary view on how to write abstracts, read: http://crosstalk.cell.com/blog/how-to-hook-an-audience-with-a-great-abstract.

Manuscript Text: Parts of your text were a bit confusing, so I took the liberty of suggesting quite a large number of specific edits. The notes below are "called out" on a matching PDF that I'll send you separately in a separate email. There are also additional edits on that PDF that don't require explanatory notes. While all of my edits are formally suggestions, please note that if I'm making a suggestion, the relevant section of text could use improvement, even if you disagree with the specific improvements I'm suggesting. Also, I apologize in advance for any distortions in meaning that my edits may inadvertently introduce. ***Finally, note that if you make the edits I suggest, you'll meet our "length limits." Please don't cut your text.***

Note 1a: I don't think you need to limit this discussion to models; these ideas likely apply in vivo as well. Hence the suggested deletion.

Note 1b: Replace highlighted text with, "Yet another possibility is that stochastic gene expression alone fails to produce rare coordinated high states in the absence of additional regulation."

Note 2: I found this paragraph confusing as written and suggest the following edits (a few additions for readability, but mostly re-organization:

> We limited our study to networks that are symmetric, i.e., networks without a hierarchical structure (**Box 1**; **STAR Methods**, section Networks, **Figure S1C**). We also excluded networks that are compositions of independent subnetworks (non weakly-connected networks) and networks that can be formed by structure-preserving bijections of other networks (isomorphic networks) (**STAR Methods,** section Networks, **Box1**). These choices reduce the testable space of unique networks by several orders of magnitude (**Figure S1C**) and allow for comparisons of parameters between networks of different sizes. They also are a conservative starting point for our analysis given experimental observations. In the frequency matrix for experimental RNA FISH data describing the rare high state in drug naïve melanoma, in which each entry corresponds to the fraction of cells with each gene-pair being highly expressed (**Figure S1D**) (Shaffer et al., 2017, 2018), we do not observe a clear directionality of regulation or hierarchical structure within the highly expressed genes. While simulated symmetric networks can recapitulate this experimental observation, asymmetric networks can result in frequency matrices being highly asymmetric (**Figure S1E-F**). For these reasons, we restricted our initial analysis to symmetrical networks.

Note 3: Ditto above for the section of text within the green highlights. I suggest:

> For example, in a particular 8-node network, we found that the distribution qualitatively captures the experimental observations where most cells do not exhibit high expression states, while some cells are in a high state for one or more genes **(Figure 2B)**. Similarly, when we selected a gene and plotted its product count for the randomly selected time points, we observed a heavy-tailed distribution (**Figure 2C, right panel**), similar to the experimental observations **(Figure 2C, left panel)**. These observations, while shown for a particular 8-node network, also hold true for simulations of other 8-node networks as well as networks of other sizes (**Figure S2B**).
>
> Note that the simulated distributions of gene product counts for each gene are qualitatively similar because each gene is equivalent within our symmetrical networks (**Figure S2C**). This is not biologically realistic; the experimental data in drug naive melanoma cells for mRNA counts display different degrees of skewness of the distribution for different genes (e.g. EGFR vs. Jun, **Figure S3A**). These experimental observations can be recapitulated in the simulated networks by introducing asymmetries. For example, two asymmetric networks we tested were able to produce rare coordinated high states (**Figure S2G-S4M**) and distributions of gene product counts with different degrees of skewness ***<correct? Descriptor here should match above (also highlighted)>***(**Figure S2M**). When experimentally observed expression distributions (**Figure S3A**) are compared to simulated expression distributions using Gini coefficients, we observe that while the Gini coefficient is low for most of the simulations (99.2%, gray), it is much higher for the simulations that produce rare coordinated high states (red) and overlaps with experimental Gini coefficients observed for individual genes (**Figure ~~S3D~~ *<move to main text and incorporate into Figure 2>***). In total, these observations suggest that a simple transcriptional bursting model is able to produce states which recapitulate key aspects of rare coordinated high states observed in drug naive melanoma.

Notes 4a and 4b: These are super interesting points, but they should be moved to the Discussion (they're hypothetical and potentially distracting).

Note 5: I would consider replacing "auto-regulation" with "auto-activation" throughout this section. It seems more specific. Also, consider replacing the highlighted text with "network connectivity and simple positive feedback in the form of gene auto-activation." (See Note 6 as well.)

Note 6: The two statements highlighted in green are apparently contradictory. Also, it's unclear to me why you need to rank connectivity and auto-activation/regulation, especially since the data in Fig. S5E are (IMO) among the weakest in the paper. Saying that both can matter is enough, and it's likely more appropriate, too. My suggested strikethroughs reflect this preference, although my edits should be checked carefully for accuracy and please do not follow my suggestions if they're unintentionally misleading (and apologies, if so, for that). Finally, please move S5F to the main text (incorporate into Figure 2) and insert a sentence summarizing the conclusions you draw from it at the end of this paragraph.

Note 7: This doesn't seem like a complete thought? Add a "suggesting that…" to the end of the sentence to help your reader out. The writing in this section is comparatively rough and obscures the message. (Edits suggested)

Note 8: Please only use "significant" in the statistical sense. Either add the test and the metric or delete the word.

Also please note that ouse style disallows editorializing within the text (e.g. strikingly, surprisingly, importantly, etc.), especially the Results section. These terms are a distraction and they aren't needed—your excellent observations are certainly impactful enough to stand on their own. Please remove these words and others like them. "Notably" is suitably neutral to use once or twice if absolutely necessary.

Figures: As mentioned above, Figs. S3D and S5F should be moved to the main text and incorporated into Figure 2.

Legends: While the Figure's title can be the "take home" message, **figure legends need to provide the information that a reader needs to interpret your figures independently, for themselves**. You should not simply state what an author should conclude from the figure, or what you think the figure demonstrates. All of the main text figure legends highlighted in green must be fixed. I haven't reviewed the Supplemental Figure Legends but please adhere to the same standard with them as well. Please revise your Figure Legends to make them fulsome and complete; if you are worried about text length limits, please email me.

Thank you!

Reviewer comments:

Reviewer #1: The authors have addressed all of my concerns and I recommend publication.

Reviewer #2: The authors have taken the reviews seriously and used them to improve the manuscript enormously. Now it is an important contribution to the field.

Reviewer #3: The reviewers have fully addressed my comments and the paper is greatly improved.

Cell Systems

Transparent Peer Review Record

Attachments: Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer

Lea Schuh, Michael Saint-Antoine, Eric M. Sanford, Benjamin L. Emert, Abhyudai Singh, Carsten Marr, Arjun Raj, and Yogesh Goyal

Author-provided documents included with the Transparent Peer Review Record:

1. Authors' response to the reviewers' first round comments
2. Revised manuscript

Additional documents included with the Transparent Peer Review Record:

1. "Marked-up PDF" containing editing suggestions

Reviewer #1:

Arjun's lab recently made an important observation that many jackpot genes are expressed in cells that are predisposed to being resistance to drug therapy. One of the questions that arises from this surprising observation is that how are these jackpot genes turned on in a subpopulation of cells. The paradox is that these genes are not that correlated in their expression level in the jackpot cells, yet tend to be enriched in the resistant cells. This manuscript tackles this fundamental question with simulations of networks of coupled and stochastic genes and showed that similar behavior as observed in experiments are seen in symmetric networks. They further showed that the coordinated stochastic behavior arises only in networks with a certain level of connectivity.

One of interesting findings is that entrance of the jackpot state is initiated by large burst size of one gene, whereas exit is not correlated with burst sizes. The authors further examined the parameter space and found that Kon, Koff and Kadd are the 3/7 parameters that contribute to the generation of the jackpot states. It is also illuminating to read the discussion about how variations of the models including non-symmetric and negative interactions will impact the results.

Understanding the phase space of gene networks and how they contribute to phenotypes is important to understanding cell type differentiation. This is an important contribution to the field and I strongly support the publication of this paper.

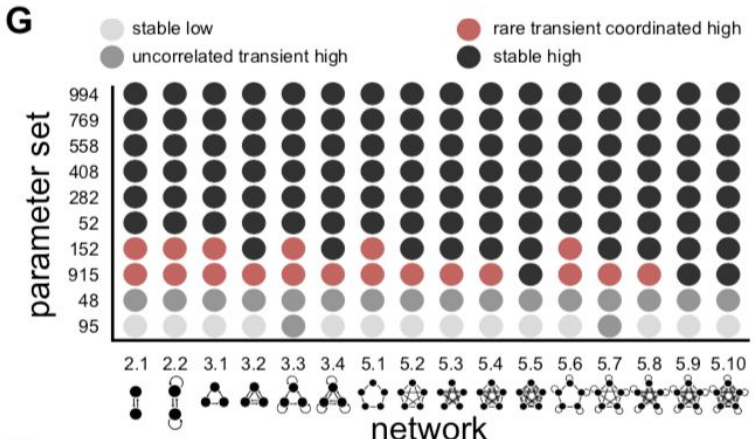
> We thank the reviewer for the positive comments and interest in our work. Based on their and other reviewers' suggestions, we have now further expanded on our analysis of the model, specifics of which are described below.

A few minor questions and comments. First, because the model can be applied to any gene network, would we expect to see these coordinated high states for many genes? And are the jackpot genes just happen to be important for drug resistance? Or are there something special about the architecture of these genes?

> We thank the reviewer for raising these conceptually deep and important discussion points. We indeed mention in the discussion section that "*the transcriptional bursting model predicts that every cell type is capable of entering the rare coordinated high gene state*". More specifically, in agreement with the reviewer, we posit that such coordinated high states can exist in general contexts and can have various (or no) functional consequences depending on the gene sets involved. We have expanded the discussion section in the revised manuscript (new text in red):
>
> *"Given the relative generality of the scenarios that produce rare coordinated high states, the transcriptional bursting model predicts that every cell type is capable of entering the rare coordinated high state. Furthermore, we show that canonical modes of transcription alone, namely the binding of the transcription factor at gene locus to produce mRNA via recruitment of RNA Polymerase II, can lead to these states without requiring other complex mechanisms such as DNA methylation, histone modifications, or phase separation. While such other mechanisms may still be operational in these cells to regulate their entry to or exit from these states, we posit that in principle, any set of genes interacting via traditional gene regulatory mechanisms are capable of exhibiting these rare coordinated high states, as long as they are interacting in a certain manner (e.g. sparsely connected) with appropriate kinetic parameters. In the case of melanoma cells, the transient state is characterized by an increased ability to survive drug therapy leading to uncontrolled proliferation of the resulting resistant cells. It is possible that these rare transient behaviors may exist across many sets of interacting genes which may or may not manifest into phenotypic consequences. Another possibility the transcriptional bursting model predicts is that even within the same cell, distinct modules of interacting genes can lead to distinct sets of rare coordinated high states that each can affect the cellular function and outcomes differently. These possibilities can be tested for by using increasingly accessible single cell RNA sequencing techniques on clonal population of cells."*

Second, it might be helpful to illustrate some of the networks used in the simulation in Figure 2. It is not entirely clear to me all the network topologies used in the simulation.

We thank the reviewer for the suggestion. We have now added the network topologies associated with each simulation and associated analyses in Figure 2 and all other revised main and supplementary figures. In particular, we have added the missing networks to the revised Figure 2A, 2E, and 2G. As an illustration, see the revised panel Figure 2G below:



Furthermore, all the networks used in the simulations are collectively provided in Figure S9.

Reviewer #2:

In this manuscript Schuch et al. present a theoretical framework to provide a mechanistic understanding of an increasingly well known phenomenon, namely the association of transient states of coordinated gene expression with persistent states of drug resistance in bacteria and mammalian cells. There is much that is interesting here but there is also much that is either missing or not properly reasoned which results in a missed opportunity. If the authors were to do more work, there could be something of general interest here.

The work is introduced in an odd grand manner that sets up a very high level frame of reference for what ends up being a very special case of an interesting but rare situation. Going from Taleb to cancer and drug resistance in a few paragraphs is a good thing in a review but not in a specialized research paper which is what this is. Importantly, the manuscript is motivated by experimental work (Shaffer et al, 2018 and Torre et al. 2019) but there is no proper explicit reference to the details of this work. The data appears to be scattered with ad hoc references to it as needed by the models. This is not right, the authors should state up front what are the features that are being modeled and then discuss the approach with reference to it. Rare states of coordinated gene expression in cancer cells may well do for the title but it is not enough for the development of the manuscript. The impression that one gets is that this, or a better organized version of this, could have been the modeling part of either of the two experimental works but, even for this, it would need more work.

We thank the reviewer for pointing out the need for a more focused and precise introduction. We have overhauled our introduction with these issues in mind, in particular highlighting the need to expand on the definitions and features of the rare coordinated high gene expression states being modeled. In the revised manuscript, we have added Box 1 which provides a detailed description of the model, associated assumptions, parameters, and relevant definitions of the features being modeled. We have also added Box 2, that describes the definitions of metrics used to quantify the features described in Box 1. Finally, we have rewritten large parts of introduction based on the reviewers' suggestions. In particular, we made the introductory paragraph concise and added relevant details on the experimental observations motivating our work from Shaffer et al. 2017:

''Rare and large heterogeneity in single cells have been reported to arise from non-genetic transcriptional variability, even in clonal, genetically homogeneous cells grown in identical conditions (Fallahi-Sichani et al., 2017; Gupta et al., 2011; Pisco and Huang, 2015; Shaffer et al., 2017; Sharma et al., 2018, 2010; Spencer et al., 2009; Su et al., 2017). Importantly, cells exhibiting these non-genetic deviations are resistant to anti-cancer drugs (e.g., Ras pathway inhibitors) and may lead to relapse in patients. For example, in a drug naive melanoma population, a small fraction (~1 in 3000) of cells are pre-resistant, meaning they are able to survive targeted drug therapy, resulting in their uncontrolled cellular proliferation (Shaffer et al. 2017). These rare pre-resistant cells are

marked by transient and coordinated high expression of dozens of marker genes. In other words, several genes are highly expressed simultaneously in a rare subset of cells, while the rest of the population have low or zero counts of mRNAs for these genes, resulting in a distribution of steady state mRNA counts per cell that peaks at or close to zero and has heavy tails. The rare cells in the tails, which transiently arise and disappear in the population by switching their gene expression state (Figure 1A), are much more likely to develop resistance to targeted therapies. Importantly, the rare and coordinated large fluctuations in the expression of multiple genes persist for several generations.''

I now raise a few issues that I would expect the authors to address if they want to capitalize on what is a good start to an interesting topic. The list is not exhaustive but hopefully highlights some of the problems that the manuscript has.

We thank the reviewer for the critical reading of our manuscript and providing insightful comments. We believe that their comments and suggestions have greatly strengthened our revised manuscript.

Introduction (which needs to be structured and elaborated in a very different manner more in accord with what the paper is about)

As already pointed out, the introduction is very wayward with sweeping statements on many topics and not suitable for a research report such as this one. It also contains some surprising statements such as the one referring to noise based models of the same phenomenon "Specifically, the classical models have largely described the variability that results in relatively normally distributed counts of mRNAs of a given gene per cell". This is not true as the literature shows that these models give rise to log-normal and gamma distributions that are far from normal and often match the experiments. It is very worrying that this is stated as part of the introduction as it raises issues about how much the authors understand the problem.

We completely agree with the reviewer's suggestion on restructuring the introduction. As also mentioned in the previous comment, we have now streamlined the introduction and added two boxes (attached at the end of the letter) to have a more focused and clear description of the experimental features and model formulation and assumptions.

The reviewer has also raised an important point with respect to the distribution of mRNA counts that can result from stochastic gene regulation models. We agree with the reviewer that long established models of stochastic gene expression can indeed give rise to distributions that are far from normal, as has been well documented throughout the literature. We have now eliminated any mention of the "particularity" of any kind of distribution. We have now rewritten this paragraph in the main text as follows:

"...In other words, several genes are highly expressed simultaneously in a rare subset of cells, while the rest of the population have low or zero counts of mRNAs for these genes, resulting in a distribution of steady state mRNA counts per cell that peaks at or close to zero and has heavy tails. The rare cells in the tails, which transiently arise and disappear in the population by switching their gene expression state (Figure 1A), are much more likely to develop resistance to targeted therapies. Importantly, the rare and coordinated large fluctuations in the expression of multiple genes persist for several generations. Classical probabilistic models of gene expression have predicted the possibility of various types of mRNA expression distributions across a population, including normal, log-normal, gamma, or heavy-tail distributions (Antolović et al., 2017; Chen and Larson, 2016; Corrigan et al., 2016; Golding et al., 2005; Ham et al., 2019, 2020; Iyer-Biswas et al., 2009; Raj and van Oudenaarden, 2008; Raj et al., 2006; So et al., 2011; Symmons and Raj, 2016; Thattai and van Oudenaarden, 2001). It is unclear if such models can recapitulate the non-genetic variability characterized by rare and transient high expression states for several genes simultaneously (from now on referred to as "rare coordinated high states"), and if so, under what conditions.

Might a stochastic system of interacting genes inside the cell facilitate transition in and out of the rare coordinated high state?"

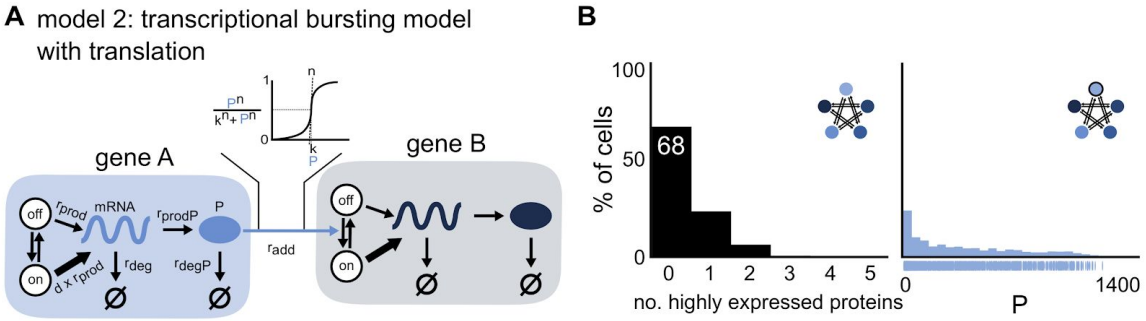
In the introduction the two opposing models are introduced in different places (Fig1A and FigS1A) which is unhelpful and distracting.

We agree with the reviewer that the two types of models discussed in the text should be presented side by side. We have now reworded our introduction to make it less distracting and more streamlined. We have also moved the constitutive-model to revised Figure 1.

They should also justify the restriction of the model to transcription and why they don't explore translational coupling since it has been shown that this can alter the distribution of cell states by creating biases in the cell states (as defined by the presence of protein products). I appreciate that they are modeling transcriptional data which, as also already stated, is not properly introduced but, when dealing with networks, proteins and translation matter. Their model appear as only having mRNA as the product and the regulator.

We completely agree with the reviewer that translation and protein production matter when studying reaction networks and should be explored, as also noted by Reviewer #3. We thus analyzed a model with translation to show that it can also produce rare coordinated high states, demonstrating the applicability of our approach in a more realistic scenario. We added the results of the analysis to the manuscript (see STAR Methods and Box 2):

"*We added one state (P) and two rate parameters, a protein synthesis rate r_{prodP} and a protein degradation rate r_{degP}, to the original transcriptional bursting model. [...] We tested three different translation scenarios: protein synthesis and degradation being (1) faster than (2) same as and (3) slower than mRNA synthesis and degradation. For network 5.3 and parameter set 968, giving rise to rare coordinated high states in the transcriptional bursting model without translation, we took a = b = 10 (faster), a = b = 1 (same) and a = b = 0.1 (slower) as additional parameters. We find that protein synthesis and degradation with faster (**Figure S4B**) and same rates as mRNA degradation and synthesis, also allows for the formation of rare coordinated high states in the case of translation. Only slower protein synthesis and degradation rates did not show rare coordinated high states, likely because for faster protein rates, the system dynamics is determined largely by the transcriptional dynamics. In sum, we demonstrate that the rare coordinated high states can arise in the revised model that includes translation.*"



Finally, we stress that performing the entire analysis on the revised model with translation is computationally extremely expensive, because adding translation increases the model dimensions and long stochastic simulations are required to study the rare coordinated high states. In the future, it will be important to systematically explore how the addition of translation steps to our model affects the networks and parameters that give rise to these states.

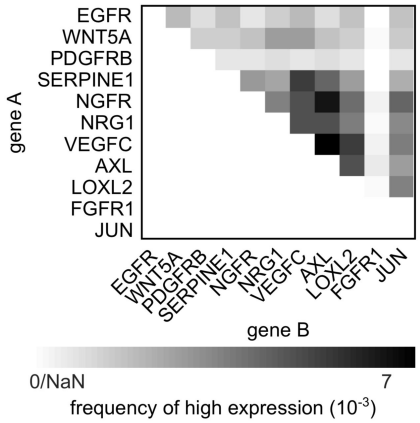
Results

An important issue is that they need to be very precise about the definition of a 'highly correlated' state. This will help understand what the work is about and what is being discussed and addressed. At the moment it is not. We are told that

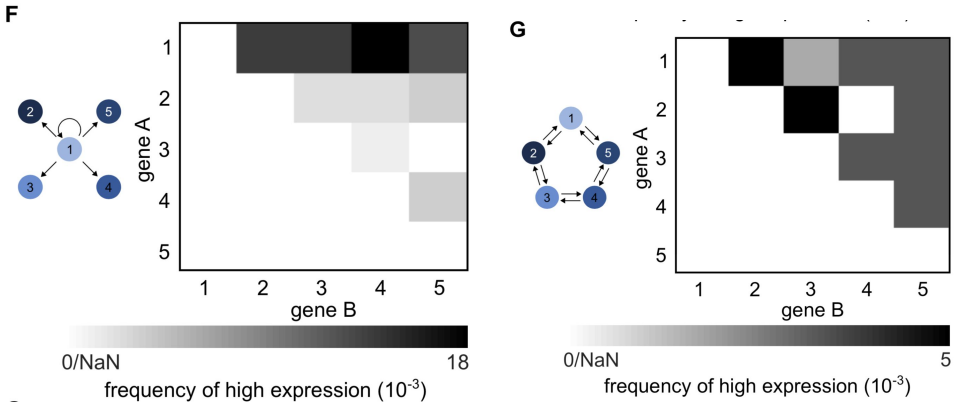
experimentally there is no correlation between the genes that characterize the state and this seems to be supported by the ad hoc reference to data (FigS1F) but then their networks assume some correlation. Even the basic model in which one gene activates another has this correlation implicit. Need clarification and explanation.

The reviewer has raised an important point about having precise definitions of the features being modeled. We completely agree with the reviewer about the confusion surrounding the 'highly correlated' state as it is currently used in the manuscript. In the revised manuscript, we have refrained from using the word "correlation" when describing the experimental observations. Specifically, we added the following in the revised results section:

*''We limited our study to networks that are symmetric, i.e., networks without a hierarchical structure (**Box 1**; **STAR Methods**, section Networks, **Figure S1C**), a simplification partially supported by the experimental observation that there doesn't seem to be a clear directionality of regulation or hierarchical structure within the highly expressed genes in the rare high state in drug-naive melanoma (**Figure S1D**) (Shaffer et al. 2018, 2017). The lack of hierarchy is inferred from the frequency matrix for the experimental RNA FISH data, in which each entry corresponds to the fraction of cells with each gene-pair being highly expressed (**Figure S1D**). Asymmetric networks can result in frequency matrices being highly asymmetric, as demonstrated by an example simulation of a star-shaped reaction network (**Figure S1E-F**).''*



Furthermore, we have provided an example simulation to demonstrate the above point. We simulated a star-shaped network with existing asymmetry where one central node (node 1) positively regulates itself and four peripheral nodes (nodes 2, 3, 4, and 5). The resultant frequency matrix is highly asymmetric (left) in that there is a high frequency of pairwise overlap between node 1 and all other peripheral nodes. However, the pairwise overlap between the peripheral nodes 2, 3, 4, and 5 is much smaller, which is inconsistent with the frequency matrix of experimental data. We also calculated the frequency matrix for a symmetric network (right) and found it to be qualitatively more similar to the experimental frequency matrix, where coordinated high expression is gene-independent (Figure S1D-F).



The corresponding text in the STAR Methods is as follows:

*"Experimental data from Shaffer et al. (Shaffer et al. 2017) implies the absence of any obvious hierarchical structure within the genes, and that the driver genes may interact in a relatively non-hierarchical manner (**Figure S1D**). The structural embedding of a node in its network can increase or decrease its ability of being involved in coordinated overexpression. For example, a centered node within a star-shaped network is involved more frequently in coordinated overexpression than the other nodes within the same network (**Figure S1E**), which is inconsistent with the experimental observations. To ensure for non-hierarchical behavior we define a set of symmetric networks (**Figure S1F**), where the number of in- and outgoing edges within a node and across nodes is identical and either all nodes in a network have a self-loop or not."*

We have also included Box 1 that explains the model and associated definitions, which we hope further clarifies the terms "coordinated" used in the text.

Page 7: they should justify the ergodic assumption since, intuitively, it will be broken by the entry into the rare stable states. They might argue that in the very large sampling they are doing this will apply but, does a cell have access to a similar phase space?

We thank the reviewer for pointing out the potential issue with our ergodic assumption, especially given the unique nature of the rare coordinated high state. We run each simulation for 1 million time units, and subsequently split the simulation into non-overlapping time interval of 1000 time units. For the ergodic assumption to hold, the behavior in one interval should not influence the behavior in another interval, essentially showing that the simulated events are Markovian (Van Kampen 1992). To check whether the independence was indeed the case, we compared the autocorrelation values between successive time-intervals, updated STAR Methods section, and included a summary of the analysis in Box 2:

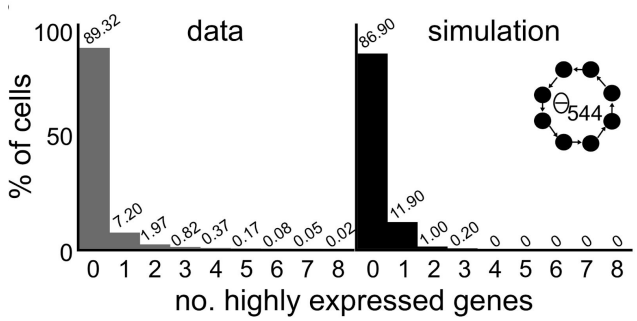
*"To show that sub-simulations of 1,000 time units are uncorrelated, we determine the autocorrelations for all 1,000 parameter sets of network 3.2 (**Figure S9**) for up to 1,000 lags (using the MATLAB autocorrelation function acf (Price, C. (2011). Autocorrelation function(ACF) (https://www.mathworks.com/matlabcentral/fileexchange/30540-autocorrelation-function-acf), MATLAB Central File Exchange. Retrieved June 13, 2019.). For each of these, we determine the first lag at which the autocorrelation is below the upper 95% confidence bound. For 88.2% of all simulations, the first lag below the upper 95% confidence bound occurs before 1,000 lags. For the 26 simulations with rare coordinated high states, 23 show a first lag below the upper 95% confidence bound before 1,000 lags. For the remaining three simulations the autocorrelation after 1,000 lags is at 0.0615, 0.0206 and 0.4363. Removing the simulation with high autocorrelation (0.4363) does not change the conclusions of our analysis."*

Some of the results presented in Figure 2 raise questions. In Figure 2B we are shown a comparison between data and simulation in terms of the genes coexpressed in these rare states. A superficial viewing suggests a similarity but it is obvious that the experimental data has a tail which is lacking in the simulation. One can see that there is a non-zero number of cells with 8 highly expressed genes. Need an explanation.

We thank the reviewer for their detailed reading of our manuscript and making insightful observations. The tail in the experimental data in original Figure 2B (at number of highly expressed genes = 8) was a mistake on our part when preparing the figure. We apologize for the oversight. The actual experimental data does not have that long of a tail, which we have corrected in the revised figure and added the percentages for clarity. Furthermore, we note that the distribution from model simulation shown here does not have non-zero values for higher number of genes, unlike the experimental data. The absence of of non-zero values may be because 1) The number of cells obtained from experimental RNA FISH imaging data is ~5-times more than the computationally simulated cells used to plot the histogram (as measured by splitting the entire simulation (10^6) into a simulated population of 1,000 cells); and 2) The network underlying the experimental data likely contains a much larger set of interacting

genes, which may increase the likelihood of obtaining the non-zero values for higher number of expressed genes. We have now added a discussion point in the main text:

"The absence of of non-zero values may be because the network underlying the experimental data contains a much larger set of interacting genes, thereby increasing the likelihood of non-zero values for higher number of expressed genes."



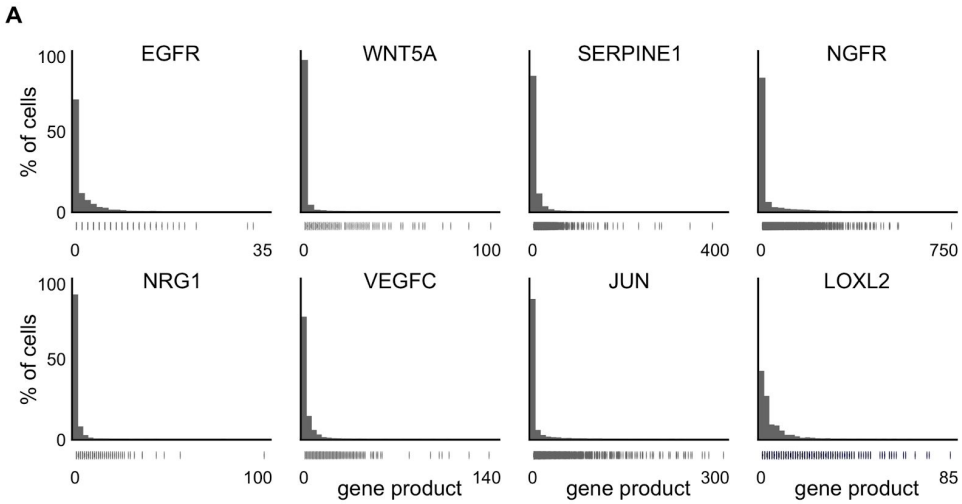
Finally, we emphasize that we are not aiming to fit the model parameters and the simulated data to the experimental data. Instead, our goal is to capture the features that describe the rare coordinated high states observed in drug naive melanoma cells

Figure 2C suffers from a similar problem as, in detail, the distributions are different. It would be helpful if to show similarity or differences, they would do a semi-log plot.

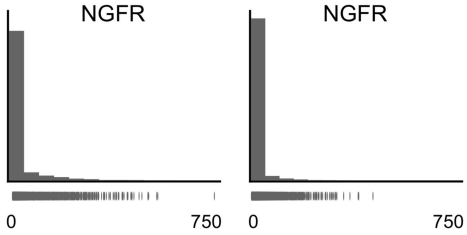
We thank the reviewer for their insightful observations and suggestions. In the revised manuscript, we sought to systematically compare the experimental and simulated data. First, we highlight that there is variability between the experimental distributions of genes that characterize rare coordinated high states (see the newly Figure S3A):

*"Similarly, when we selected a gene and plotted its product count for the randomly selected time points, we observed a heavy-tailed distribution (**Figure 2C, right panel**), similar to the experimental observations **(Figure 2C left panel** and **Figure S3A)**."*

*"Since there is both inter- and intra-gene variability between the experimental expression distributions characterizing these states (**Figure S3A**) ..."*



Furthermore, while the distributions of the same gene (e.g. NGFR, used in Figure 2C) of biological replicates have heavy tails, they don't appear exactly the same (plot not included in the revised manuscript).

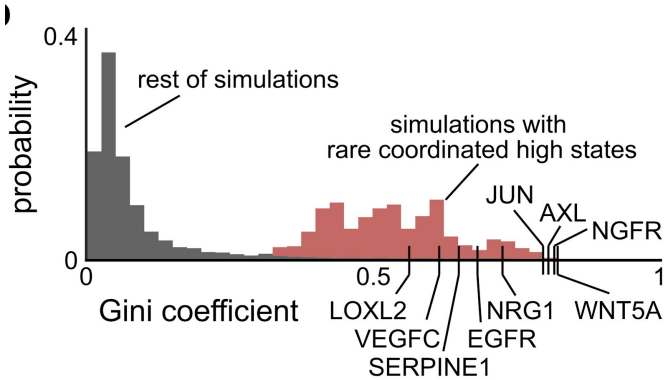


Therefore, we focused on performing the comparisons that capture general properties of distributions across all genes. We first note that comparisons on a semi-log plot are not appropriate as the majority of the points belong to x = 0. Removing these points fundamentally changes the distributions and adding a small offset to them is not ideal either. Therefore, we resorted to using other metrics to compare the simulated and experimental data.

Our two metrics for comparative analysis take into account these considerations. First, we computed Gini indices for both the simulated and experimental data, which was also used to characterize experimental distributions in the original study (Figure 3D, Shaffer et al, 2017). A Gini coefficient of 0 implies perfect equality such that for a given gene, all cells within a population have the same number of mRNA molecules, whereas 1 implies perfect inequality such that one cell expresses all the mRNA molecules while others express none. We added the new analysis to the results section in the revised manuscript and the STAR Methods:

*"Since there is both inter- and intra-gene variability between the experimental expression distributions characterizing these states (**Figure S3A**), we compared these expression distributions to simulated expression distributions using Gini coefficients, used to characterize experimental expression distributions in the original study (Shaffer et al, 2017). While the Gini coefficient is low for most of the simulations (99.2%, gray), it is much higher for the simulations that produce rare coordinated high states (red) and overlaps with experimental Gini coefficients (**Figure S3D**)."*

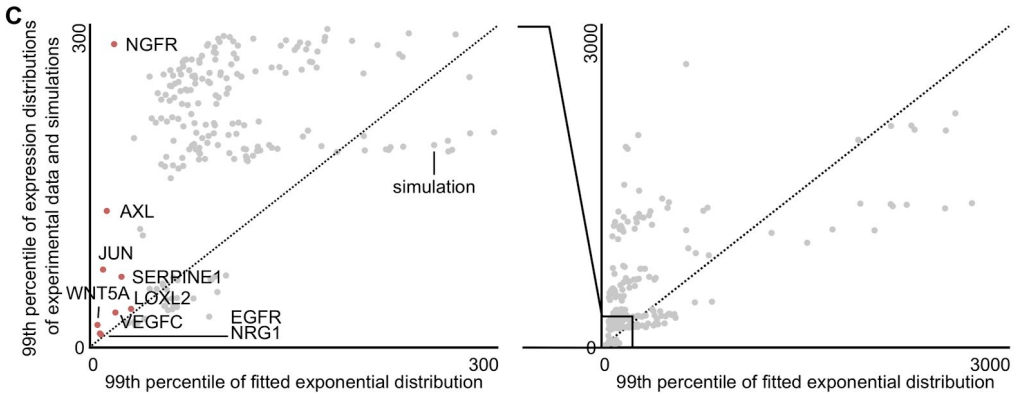
*"Additionally, we computed the Gini indices for the gene expression distributions of both the simulations showing rare coordinated high states and the experimental data (**Figure S3A and FigureS3D**) (Shaffer et al. 2017; Jiang et al. 2016). A Gini coefficient of 0 implies perfect equality such that for a given gene, all cells within a population have the same number of mRNA molecules, whereas 1 implies perfect inequality such that one cell expresses all the mRNA molecules while others express none. We used the MATLAB function gini (Yvan Lengwiler (2019). Gini coefficient and the Lorentz curve (https://www.mathworks.com/matlabcentral/fileexchange/28080-gini-coefficient-and-the-lorentz-curve), MATLAB Central File Exchange. Retrieved October 24, 2019.) for the computations."*



Second, we fitted exponential distributions to the simulated and now also experimental data and compared the 99th percentiles. Points above the x=y line correspond to distributions with fatter tails than of the fitted exponential distributions. Removing the simulations below the x=y line does not qualitatively change our downstream analysis. We added the following comparison to the STAR Methods section:

*"Most (82%) of the 99th percentile of the simulated expression distributions are above the diagonal, hence larger than the 99th percentile of the fitted exponential distributions (**Figure S3C, right panel**). The 99th percentile of all the nine marker genes in Shaffer et al. also lie above the diagonal in the general vicinity of the points corresponding to simulations with rare coordinated high states (**Figure S3C, left panel**)."*

The procedure and considerations taken while fitting exponentials is described in detail in STAR Methods, Simulation section.

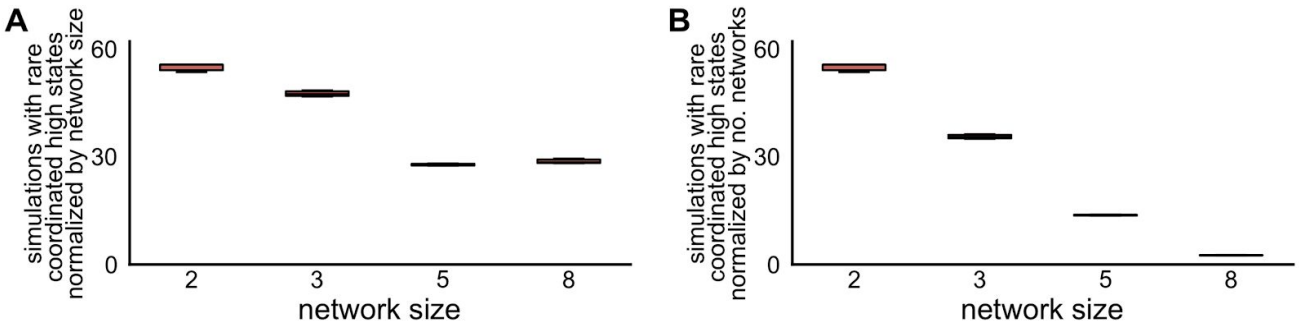


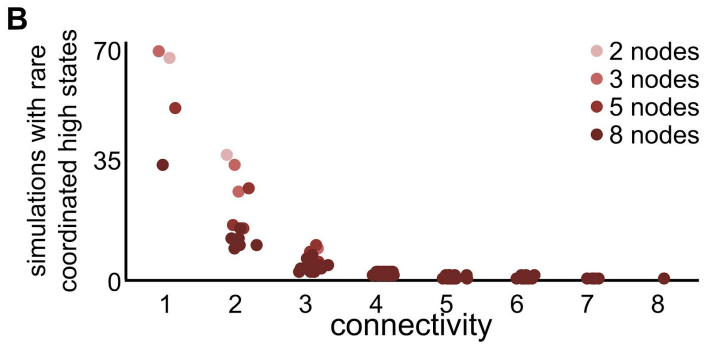
Together, our additional analysis on the experimental data demonstrates that the simulated data captures its key features.

Another important issues arises in 2D where one wonders if this the result is not a trivial consequence of the intuitive fact that as the number of nodes in the network increases, the number of states increases. Maybe if they normalized to network size they might get a different result.

We agree with the reviewer that the increase in the number of simulations with the rare coordinated high states with network size is likely a result of the increased number of possible networks for larger network sizes. We used the original figure panel to demonstrate that these states can be exhibited by networks of different sizes, but we see how it could be misleading. In the revised manuscript, we have provided two types of normalizations (Figures S5A,B) to take into account the network size: 1) Normalization by the network size as suggested by the reviewer; 2) Normalization by number of networks corresponding to each network size (Figure S9). We added the following to the main manuscript:

*"Indeed, we found that, within a particular network size, the ability to produce rare coordinated high states decreases dramatically (and monotonically) with increasing network connectivity (**Figure 2E and Figure S5C-D**). Consistently, the fraction of networks per network size (normalized by either network size or total networks per network size) exhibiting rare coordinated high states decreases with increasing size (**Figure S5A-B**) as a larger fraction of high connectivity networks exist in bigger networks (**Figure S5D**)."*





In this regard the limitation of the network size to <8 is not properly justified or explained. Looking at how ron is calculated (page 19) the limitation could be a trivial consequence of the addition that is used in the calculation. Above six nodes, the system might blow up. They claim but not show that ≥8 networks will also display coordinated states but will they? And how does the number correlate with the experimental data? Could the authors discuss?

The reviewer has brought up an important discussion point about our investigation being limited to network size ≤8. As the reviewer also points out, the space of networks increase with network sizes, in turn making it computationally prohibitive to perform extensive analysis across all possible topologies and parameters. Furthermore, it is possible that the rare coordinated high states may be a local feature of a subnetwork within an otherwise large network. In the original manuscript, we have already shown that a network with 8 nodes exhibits the rare coordinated high states. While our additional analysis in the revised manuscript shows that even a network with 10 nodes exhibits these states (Figure S2D-F), we completely agree with the reviewer that these additional observations don't ascertain that even larger networks will also display these states. Therefore, we have modified the text in the revised manuscript at various places to moderate our claims and provide clarification wherever necessary. An example statement is below:

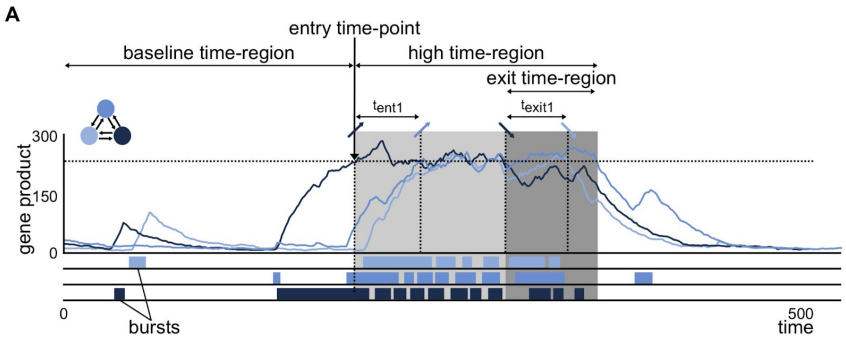
*"We found that the rare coordinated high states occur ubiquitously in networks with different numbers of nodes analysed (up to 10 nodes) (**Figure 2D and Figure S2B-F, Figure S5A-B**). We therefore hypothesize that even larger networks may also display rare coordinated high states, and can be explored in future studies."*

The reviewer has also asked us to discuss how the number (network size) correlates with the experimental data (also stated in the response to reviewer #3). While our lab used network inference algorithms to infer the underlying networks from RNA FISH (this manuscript) and RNASeq datasets (Figure 5, Shaffer et al, 2018), we lack a quantitative wiring diagram of the actual gene regulatory network. Therefore, making quantitative comparisons on the network size with the experimental data is not possible.

Even admitting that it was clear what exactly they are trying to model, or better find an explanation to, the reliance on simulations obscures some of the potential explanations. One example is on Figure 3A. Surprisingly there appears to be no correlation between the maintenance of expression and stopping the burst e.g take the 'black gene', it is surprising that the expression is maintained for so long after the first long burst. How can this happen? The same applies to the others. If there is a memory in the system, where does it come from?

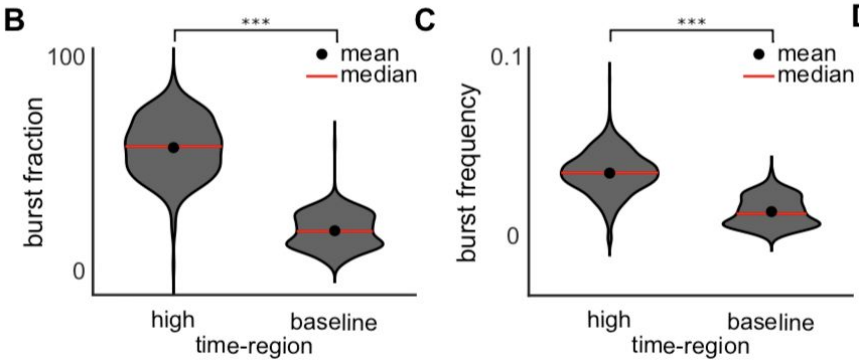
We thank the reviewer for noticing the potential inconsistencies in Figure 3A and encouraging us to think mechanistically about the rare behavior. We apologize for not only originally clarifying that the bursts shown in Figure 3A are schematics, but also for not accurately representing the actual scenario. We have now edited Figure 3A and associated figure caption to represent more transcriptional bursts:

"**Figure 3. Rare coordinated high state [...].**
(A) [...] The bursts below the exemplary simulation are representative schematics."



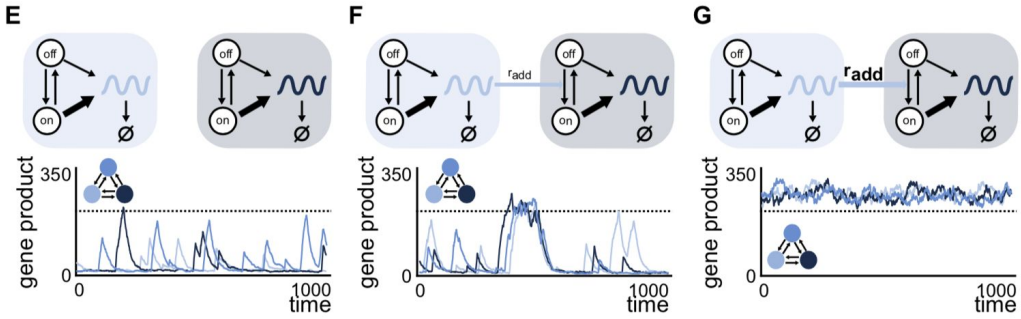
Specifically, inspired by the reviewers' comments, we wondered whether enhanced transcriptional activity facilitates the maintenance of rare coordinated high states. We now state in the revised manuscript:

*"We found a substantial increase in the transcriptional activity, as measured by the burst fraction, during the high expression time-region as compared to the baseline time-region (**Figure 3B**). Increased burst fraction could be a result of (1) longer transcriptional bursts or (2) a higher burst frequency. The former is not possible as the duration of each burst is distributed exponentially according to exp(r_{off}), which does not change between the baseline and high time-region. Indeed, we found an increase in the burst frequency in high time-region, thus establishing its role in the maintenance of the rare coordinated high state (**Figure 3C**). The increased transcriptional bursting seen in the transcriptional bursting model is consistent with the experimental data using labeled nascent transcripts which showed that the transcriptional activity occurred in frequent bursts in cells high for a marker gene (Shaffer et al. 2018)."*



Along the same lines, it would be helpful if they could dig into the causes for the exit from the state and present some experimental data. And, of course, what determines the duration of the state?

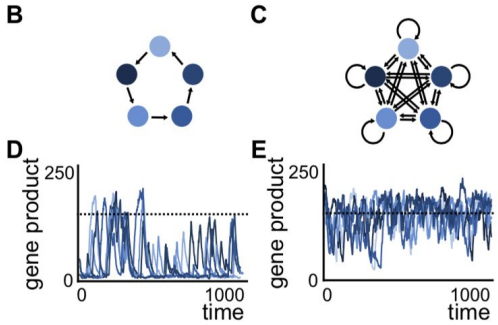
The reviewer has raised excellent points on finding the reasons leading to the exit from the high state and to present experimental data. While we do not have any experimental data on identifying the causes of exit, our work in the original manuscript established two things about the exit from the state: 1) exit for successive genes from the high expression state occurs independent of each other; 2) the exit from high expression state does not require changes in burst durations. We believe that the weak strength of coupling between nodes in the network is key to exiting from the high state, which, in turn, may also determine the duration of the state. We demonstrate that weak coupling is indeed critical by analyzing the parameter r_{add} that characterizes the strength of the positive interaction between the nodes. Specifically, we show that a fixed network exhibits the transient state only for low, non-zero values (<0.31) of r_{add}, which suggests existing but weak coupling. Keeping other parameters fixed, we show that too high values of r_{add} results in the disappearance of rare coordinated high states, giving way to stable high states, i.e., the network can transition into the high expression state but loses the ability to come out of it (Figure S6 E-G).



We have now added the corresponding text in the revised manuscript:

"*The exit from the high state could be a result of weak strength of coupling (as reflected by the moderate values of parameter radd) between nodes for the simulations that produce these states. Consistently, we found that too high values of radd results in the disappearance of rare coordinated high states, giving way to stable high states. In other words, the network can transition into the high expression state but loses the ability to come out of it (Figure S6 E-G).*"

Consistently, we see a similar effect upon increasing the network connectivity (and consequently the effective strength of interaction). Keeping all other things the same, we see a shift from a transient coordinated high expression state to stably high expression state upon increasing connectivity (Figure 4B-E).



They discussion of the parameters is confusing: there are 8 parameter sets but they only discuss 3. It would be good to know what the others are. Also, what is the relationship of these 8 parameters from the simulation to the 7 free parameters of the model? One can do some work, interpret the authors and know what they are talking about but they should tell us. Also, they should be more explicit about the point of the screen.

We thank the reviewer for pointing out the lack of clarity in the discussion of model parameters and apologize for this oversight. In the revised manuscript, we have added a new document named Box 1 which provides a detailed description of the model, associated assumptions, parameters, and relevant definitions. Specifically, we provide a table containing the definitions of independent (free) and dependent model parameters. Additionally, we provide a rationale for the definitions of dependent parameters by relaxing the definition of these parameters, as outlined in the footnote of Box 1. We have also edited the results section of main text to include the model parameters and direct readers to Box 1 as necessary. As an example:

"The dissociation constant k of the Hill function is dependent on the parameters r_{prod}, r_{deg}, and d, such that $k(r_{prod}, r_{deg}, d) = 0.95 * d * r_{prod}/r_{deg}$. In total, the model has seven free and one dependent model parameters, as outlined in Box 1."

The reviewer also raised that we explicitly explain the rationale behind the screen. Based on the reviewers suggestion, we have added text in the revised manuscript to include the inspiration behind analyzing the topologies and parameters:

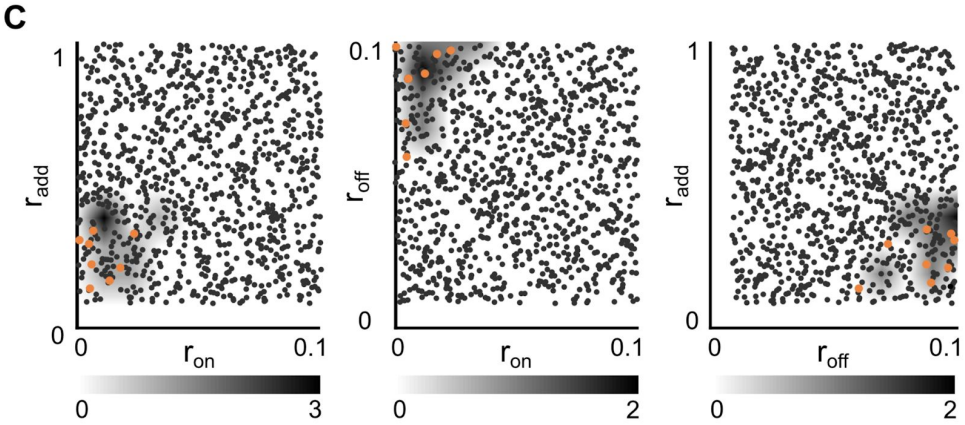
*"Since the rare coordinated high states occur in <1% of all simulations (**Figure S2A**), we wondered whether their occurrence depends on the network topologies and/or model parameters. Specifically, what are the features of the topologies and parameters that facilitate the occurrence of rare coordinated high states? For the simulations ..."*

*"Since the transcriptional bursting model has seven free parameters (r_{on}, r_{off}, r_{prod}, r_{add}, r_{add}, d, and n; see **Box 1** for details), we asked whether specific parameter combinations preferentially give rise to the rare coordinated high states, and if so, what features of such combinations facilitate it. The subsequent analysis is motivated by the initial observation that occurence of different classes of temporal gene product profiles across different network sizes and connectivities appear to also depend on the parameter sets (Figure 2G). Specifically, if …"*

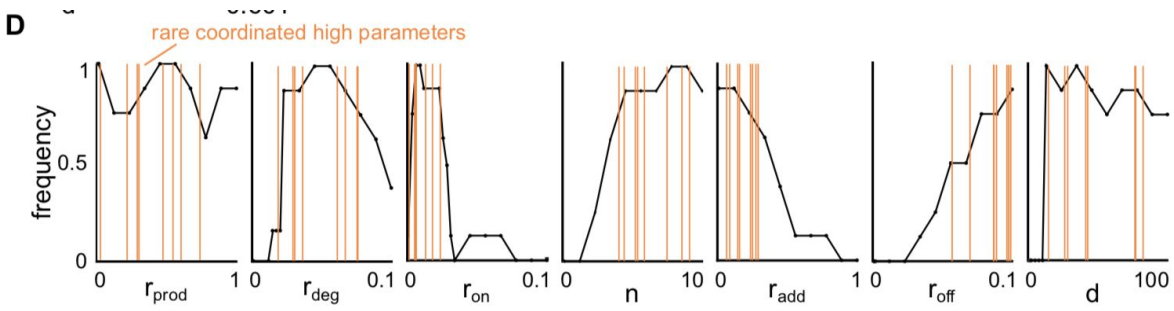
And, indeed, it would be good to have a phase space of the screens, parameters.

The reviewer has asked to provide a phase space of screens and parameters. We first highlight that our parameter space screen did not find any parameter combination that always resulted in rare coordinated high states. We did however find eight parameter sets that give rise to the rare coordinated high states much more frequently across different network architectures than others, but not necessarily always. Resampling using the bounds obtained from these eight parameter sets result in >14-fold enrichment of simulations giving rise to these states (Figure 2K).

In response to the reviewer, we have now added the phase space for the three parameters in the revised manuscript (Figure S6C).



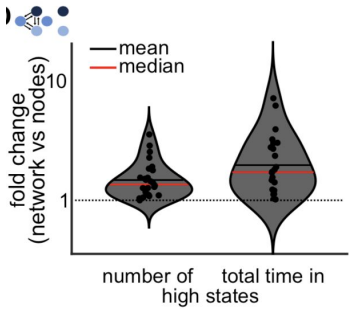
Furthermore, inspired by the reviewers' comments, we have now additionally performed parameter sensitivity analysis which confirmed that the three parameters (r_{on}, r_{off}, and r_{add}) identified by the decision tree algorithm and generalized linear model are indeed critical for producing the rare coordinated high states (Figure S6D). We also found a moderate dependence on the Hill coefficient n, also confirmed by the low p-value for n from generalized linear model analysis (Figure S6C).



What is the importance of the frequency of the bursts in the entry into and maintenance of the state?

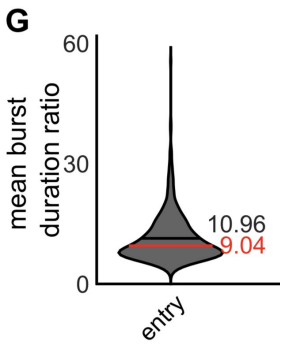
We thank the reviewer for raising important mechanistic questions on the importance of transcriptional frequency in the entry into and maintenance of the state. Reiterating the response to the reviewers' previous comment, we have now shown that the burst frequency is substantially increased during the high time-region as compared to the baseline time-region (**Figure 3B,C**). Therefore, burst frequency indeed plays an important role in maintaining the state, an observation consistent with the experimental data in Supplementary Figure 12 of Shaffer et al. 2018. We also performed additional analysis to answer whether the increase in frequency is promoted by positive regulatory interactions between the network nodes:

*"Next we wondered whether the increase in burst frequency is promoted by the interactions of genes organized within the network. We compared two networks of the same size (3 nodes), where one is comprised out of single unconnected (orphan) nodes and the other of an interdependent structure (network 3.2). We found that for any parameter set (screened for all 26 parameter sets giving simulations with rare coordinated high states in the previous analysis for network 3.2), the system with a connected network has (1) more high expression states and (2) prolonged time in high expression states, as compared to unconnected nodes (**Figure 3D**). Together, we find that the maintenance in the high state is because of increased burst frequency, which may be a result of the positive regulatory feedback intrinsic to the networks."*



The reviewer has also asked whether burst frequency plays a role in the entry into the high state. We posit that investigating the role of burst frequency first requires identifying a time-region before the entry into coordinated high state, which is not possible to define formally. Furthermore, we have shown in the original manuscript that it is the long duration of the transcriptional burst right before and during the entry into the state (entry time-point) that drives entry into the high state. In the revised text and Figure 3, we have now added that the conclusion on long duration of bursts at the entry time-point holds true when measured for all simulations that exhibit rare coordinated high states (Figure 3G):

*"Importantly, both of these conclusions hold true when measured for all simulations with rare coordinated high states (**Figure 3G**)."*



Discussion

The discussion seems to be focused in reinforcing some prejudices derived from their analysis and providing some excuses for work that should have been done. Here is an important paragraph:
"One limitation of our model is that we have performed quantitative analysis only on symmetric networks with positive interactions between nodes. It is likely that our findings hold more generally for asymmetric networks, as partially demonstrated for two cases of randomly selected asymmetric networks". If this is the way they feel, they should prove it. Also they state that "Inhibitory interactions between nodes is a separate and perhaps more interesting point. In principle, the model can be adapted to include inhibitory interactions" This is very important in light of the possibility that networks with <8 nodes will explode. They should definitely explore the influence of inhibitory interactions as this is a glaring omission in the analysis.

The reviewer is absolutely correct in pointing out that we have not proven our claims about the generality of our findings for asymmetric networks. We have now modified the text in discussion to moderate our claim.

*"One limitation of the transcriptional bursting model is that we have performed quantitative analysis only on symmetric networks with positive interactions between nodes. While the preliminary analysis on two cases of randomly selected asymmetric networks shows that they do exhibit the rare coordinated high states (**Figure S2G-S4M**), it remains to be seen whether these findings hold more generally for asymmetric networks."*

Similar to reviewer #3, the reviewer has also raised that we have limited our analysis to transcriptional activation, not taking into account transcriptional inactivation i.e. cases where gene interactions are inhibitory. We agree with the reviewer that including transcriptional inactivation will reveal important aspects including whether a revised model can still produce rare coordinated high states, and if so, under what model conditions. At the same time, including the inhibitory interactions necessitates performing the entire analysis on a large set of new parameter space and networks, which is computationally prohibitive, especially given the large number of relatively long stochastic simulations needed to search for rare events. We emphasize that the primary aim of our present study was to identify whether a minimal model based on established principles of transcriptional regulation *can* recapitulate the rare coordinated high states. It is possible that these behaviors can exist under more general conditions as well, and would be important for further studies to include additional layers of complexity. We believe that our framework will provide strong quantitative foundations for such studies.

We also agree with the reviewer that even smaller networks may explode when including the inhibitory interactions and have now added text in the revised manuscript, as follows:

''Inhibitory interactions between nodes is a separate and perhaps more interesting point. In principle, the model can be adapted to include inhibitory interactions. These inhibitory interactions may lead to non-monotonic effects of network connectivity on the occurrence of rare states, as positive and negative interactions can compete in non-linear ways. Similarly, a network with both negative and positive interactions may be more prone to instability, even for relatively smaller networks. Furthermore, inclusion of these interactions might also make the exit of genes from the high expression state dependent on one another, which occurs independently in the transcriptional bursting current model.''

Overall, a very interesting topic but a very casual and, to a certain degree, superficial analysis. The paper needs to change in structure and also have more in depth analysis and a better judgement of what is being explained and what are the assumptions; at the moment they seem to be very biased by their assumptions.

We thank the reviewer for their critical reading of the manuscript and providing a number of suggestions aimed at improving the manuscript. We have made substantial efforts to address most of the reviewers' comments and we believe that the revised manuscript is substantially improved as a result. Briefly, these include: 1) Restructuring parts of the introduction and other sections; 2) Inclusion of two boxes (attached at the end of the letter) to clearly lay down the model definitions and assumptions; 3) Additional analysis on network size and parameters, including

protein translation steps, and comparing experimental and simulated data with new metrics; and 4) Extensive analysis of factors driving the entry into, maintenance of, and exit from, the rare coordinated high state.

Reviewer #3:

The manuscript by Schuh et al investigates stochastic genetic network models that are capable of producing rate transient coordinated expression states as observed experimentally in melanoma. The authors focus on a specific class of genetic networks with transcriptional bursting and computationally explore a wide range of network size, topology and parameters. They find these networks below a threshold of connectivity are capable of producing behaviour that is similar to experimental data. They also identify this model behaviour is sensitive to values of certain parameters in the model but not others. Interestingly, they validate using network inference on experimental data a transition from low connectivity to high connectivity that is associated with transcriptionally stable states. Overall, this is an interesting paper with novel results. However, the effect of some model assumptions and also the mechanistic origin of the observed behaviour is not fully clear. I have the following specific comments:

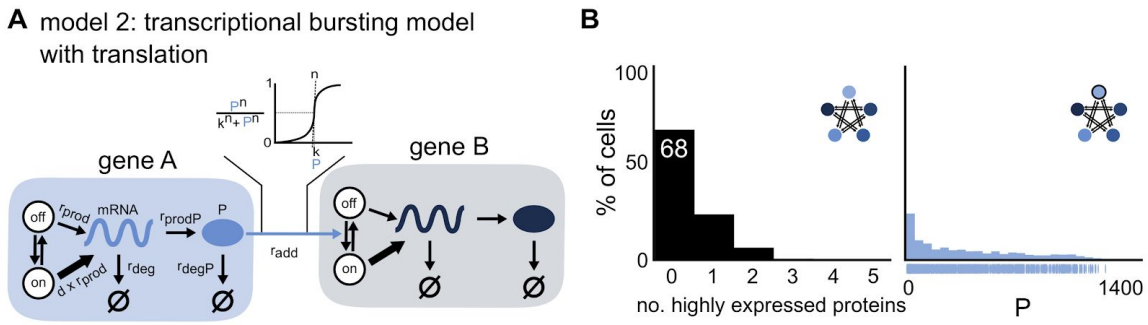
We thank the reviewer for their encouraging comments, and also for their suggestions that we believe have significantly improved our revised manuscript. We have now significantly expanded on the model assumptions and tested a subset of them for their effect on the model outcomes. We have also performed additional analysis on the mechanistic origins of the rare coordinated high states and expanded our investigation on network topologies. The specifics of our efforts are detailed in the point-by-point response below.

- The model used has several assumptions, it is not clear what is the effect of these assumptions on the model behaviour. The model, ignores protein and translation, but transcriptional regulation works through proteins. Including proteins explicitly is computational expensive, but some exploration of this would be useful. Also, the models only include transcriptional activation and only consider additive interaction. What would happen if you relax these assumptions?

The reviewer has brought up good points about the model and associated assumptions. In the revised manuscript, we have included a detailed description of the model and assumptions used to formulate the model as a part of Box 1. We also provide rationale for the definitions of dependent model and classification parameters by relaxing the definition of these parameters, as outlined in the footnote of Box 1. Based on the reviewer's suggestion, we have relaxed four out of the five major model assumptions, and provide a summary in Box 2. Details on the testing of relaxed assumptions are provided below:

1. We agree with the reviewer that while computationally expensive, an exploration of including translation and protein production in the model will be useful. We added the following analysis to the revised manuscript (see STAR Methods and Box 2):

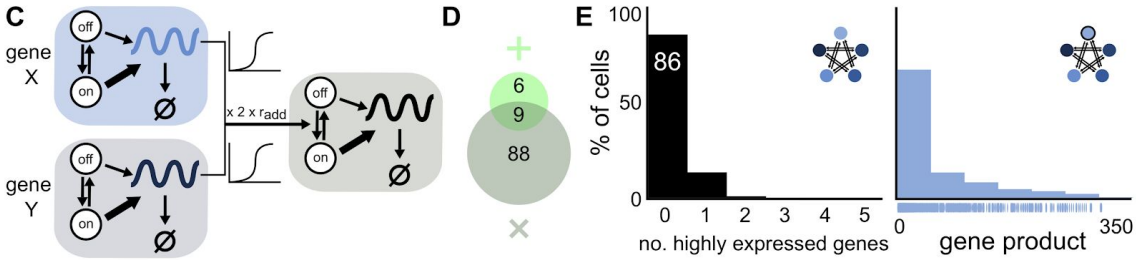
   *"We added one state (P) and two rate parameters, a protein synthesis rate r_{prodP} and a protein degradation rate r_{degP}, to the original transcriptional bursting model. [...] We tested three different translation scenarios: protein synthesis and degradation being (1) faster than (2) same as and (3) slower than mRNA synthesis and degradation. For network 5.3 and parameter set 968, giving rise to rare coordinated high states in the transcriptional bursting model without translation, we took a = b = 10 (faster), a = b = 1 (same) and a = b = 0.1 (slower) as additional parameters. We find that protein synthesis and degradation with faster (**Figure S4B**) and same rates as mRNA degradation and synthesis, also allows for the formation of rare coordinated high states in the case of translation. Only slower protein synthesis and degradation rates did not show rare coordinated high states, likely because for faster protein rates, the system dynamics is determined largely by the transcriptional dynamics. In sum, we demonstrate that the rare coordinated high states can arise in the revised model that includes translation."*



2. We extended our analysis on multi-gene regulation using additive interactions to also include multiplicative interactions, i.e. regulation where we multiply the Hill functions of the influencing genes. We found that network models with either additive or multiplicative mode of inter-gene regulation can, in principle, give rise to rare coordinated high states. We added details on the additional analysis to STAR Methods and Box 2:

*"We also tested for multiplicative regulation, i.e. regulation where we multiply the reaction rates (and consequently the reaction propensities) of the influencing genes (**Figure S4C**). [...] We show that for network 5.3, 97 out of 1000 simulations show rare coordinated high states in case of multiplicative regulation (**Figure S2D-E**). In comparison, 15 simulations show rare coordinated high states in case of additive regulation. 9 simulations show rare coordinated high states in both cases."*

"Multi-gene regulatory effects: The joint regulatory effects experienced by a gene which is regulated by several other genes can be modeled using different approaches. While the majority of analysis in our study uses an additive model of joint-regulation, we performed a subset of simulations (STAR Methods) for cases where the regulation by multiple gene nodes is multiplicative (Figure S4C and E). We find that for network architecture 5.3, 15 and 97 out of 1000 parameter sets give rise to simulations with rare coordinated high states in the additive and multiplicative joint-regulation, respectively (Figure S4D). Nine simulations are found to show rare coordinated high states in both definitions of multi-gene regulation."

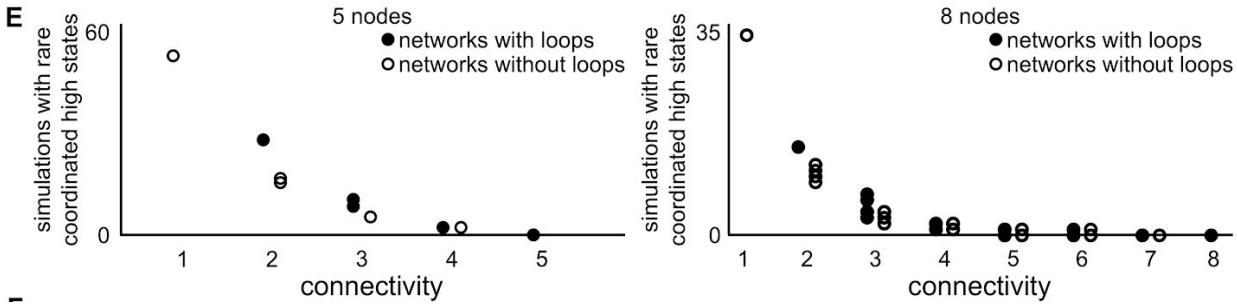


We also note that the reviewer raised that we have limited our analysis to transcriptional activation, not taking into account transcriptional inactivation i.e. cases where gene interactions are inhibitory. We completely agree with the reviewer that including transcriptional inactivation will reveal important aspects including whether the revised model can still produce rare behaviors, and if so, under what model conditions. However, as also stated above, incorporating inhibitory interactions necessitates performing the entire analysis on a large set of new network architecture and accompanying parameter sets, which is computationally prohibitive, especially given the large number of relatively long stochastic simulations needed to search for rare events. We emphasize that the primary aim of our present study was to identify whether a minimal model based on established principles of transcriptional regulation *can* recapitulate the rare coordinated high states. It is possible that these behaviors can exist under more general conditions as well, and would be important for further studies to include the inhibitory interactions. The framework developed in our study can be extended for future studies on systematic investigation of additional layers of complexity.

- While the effect of model parameters is explored, the topology of the networks are not explored fully (only at the level of connectivity). Are there specific model topologies that produce the rate transient expression more robustly that others? Do these tend to contain positive feedbacks?

The reviewer has brought up an important point about exploring the dependence of rare coordinated high states on parameters and network topologies. The reviewers' particular comment on lack of exploration of network topologies other than connectivity-based analysis has inspired our systematic analysis in the revised manuscript. We show that for a fixed size and connectivity, networks with gene auto-regulation (or self-loops) result in higher number of rare coordinated high states than networks without self-loops (Figure S5E). The enhanced ability of networks with self-loops to produce these states is likely because of the presence of *direct* positive feedback(s) for each gene, as also insightfully pointed out by the reviewer. We have now included the additional analysis in the revised manuscript:

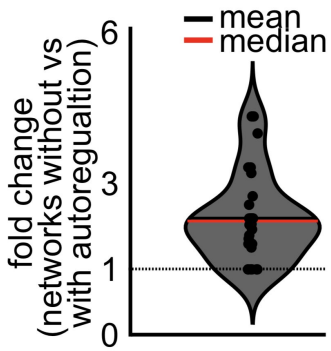
*''We next wondered whether gene auto-regulation (networks with self-loops) have any effect on a networks ability to produce the rare coordinated high states. Indeed we found that for a fixed size and connectivity, networks with auto-regulation result in higher number of simulations with rare coordinated high states than networks without auto-regulation (**Figure S5E**).''*



The analysis above inspired us to go even a step further and consider the dominating effect between connectivity and auto-regulation (self-loops). Indeed, we found that the effects of connectivity strongly overpower the effect of having self-loops on a network's ability to exhibit rare coordinated high states. We have now added the new analysis to the revised manuscript and STAR Methods:

*'"At the same time, we wondered whether connectivity or auto-regulation has a more dominating effect on a networks' ability to produce these states. We found that adding self-loops on otherwise identical networks reduced the occurrence number of simulations with rare coordinated high states (**Figure 2F**), demonstrating the stronger effect of connectivity than auto-regulation."*

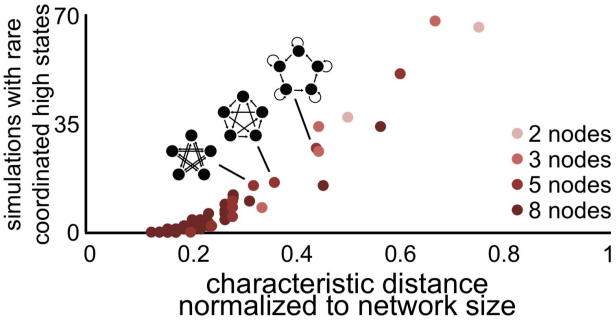
"Self-loops
*A network with a direct auto-regulation is called a network with a self-loop. Due to the restriction of symmetric networks, all networks can be classified as having self-loops for all nodes or not having self-loop for any node. Due to non-isomorphism, the set of networks contains for each network without self-loops an identical network with self-loops. We evaluate the ability of these different edge classes on the formation of rare coordinated high states (**Figure 2F and Figure S5E**). "*



We additionally considered the topological feature *characteristic distance* on the formation of simulations with rare coordinated high states:

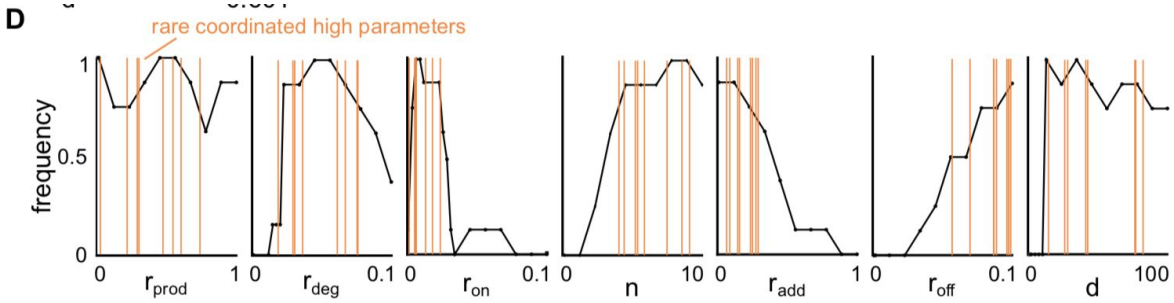
*"Finally, we analyzed network topologies based on characteristic length, defined as the average shortest path length between pairs of nodes of the network (see **STAR Methods, Box1**). Characteristic length recapitulates the effects of not only network connectivity (inversely correlated with characteristic distance), but also differentiates topologies with the same connectivity (**Figure S5F**), for example networks with or without auto-regulation."*

"Characteristic distance
*The characteristic distance of a network is defined as the average shortest path length for all pairs of nodes within a given network. To calculate this distance, we used the MATLAB function shortestpath on all pairs of nodes. We evaluated the ability of the characteristic distance normalized to the network size on the formation of rare coordinated high states (**Figure 5F**)."*



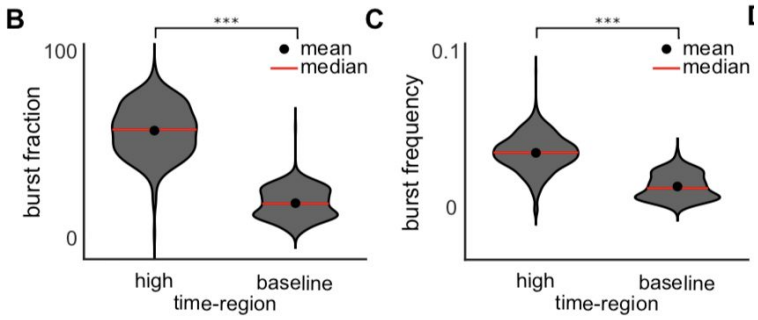
- Do you have any mechanistic insight on the origin of this behaviour? Why some parameters are critical and some are not relevant?

We thank the reviewer for encouraging us to think mechanistically about the origins of rare coordinated high states. In the revised manuscript, we first systematically performed parameter sensitivity analysis and confirmed that the three parameters (r_{on}, r_{off} , and r_{add}) identified by the decision tree algorithm and generalized linear model are indeed critical for producing the rare coordinated high states (**Figure S6D**).



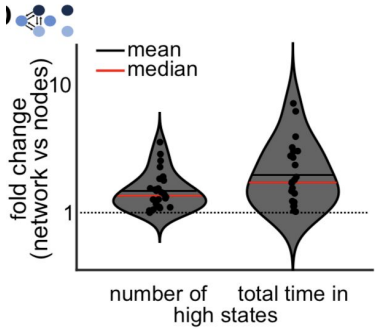
All the three parameters affect transcriptional bursting activity—two of these parameters (r_{on} and r_{off}) characterize the activity at each node and the third parameter (r_{add}) controls how the activity is affected by inter-gene(node) interaction. Given the strong dependence on three parameters regulating transcriptional bursting, we wondered whether enhanced transcriptional activity facilitates the maintenance of rare coordinated high states. We now state our results in the revised manuscript:

*''We found a substantial increase in the transcriptional activity, as measured by the burst fraction, during the high expression time-region as compared to the baseline time-region (**Figure 3B**). Increased burst fraction could be a result of (1) longer transcriptional bursts or (2) a higher burst frequency. The former is not possible as the duration of each burst is distributed exponentially according to exp(r_{off}), which does not change between the baseline and high time-region. Indeed, we found an increase in the burst frequency in high time-region, thus establishing its role in the maintenance of the rare coordinated high state (**Figure 3C**). The increased transcriptional bursting seen in the transcriptional bursting model is consistent with the experimental data using labeled nascent transcripts which showed that the transcriptional activity occurred in frequent bursts in cells high for a marker gene (Shaffer et al. 2018).''*



We next hypothesized, partly inspired by the reviewers previous comment, that the change in frequency is promoted by positive regulatory interactions between the network nodes. The revised manuscript now includes:

*"Next we wondered whether the increase in burst frequency is promoted by the interactions of genes organized within the network. We compared two networks of the same size (3 nodes), where one is comprised out of single unconnected (orphan) nodes and the other of an interdependent structure (network 3.2). We found that for any parameter set (screened for all 26 parameter sets giving simulations with rare coordinated high states in the previous analysis for network 3.2), the system with a connected network has (1) more high expression states and (2) prolonged time in high expression states, as compared to unconnected nodes (**Figure 3D**). Together, we find that the maintenance in the high state is because of increased burst frequency, which may be a result of the positive regulatory feedback intrinsic to the networks.''*



Together, since both the entry into and the maintenance of the rare coordinated high state is regulated by transcriptional bursting, the parameters directly controlling it are relevant, while others are not.

- The model uses non mass action kinetics. Have you used a variant of the Gillespie algorithm to handle this and what kind of approximation are you making? Could this affect your results.

We thank the reviewer for bringing up an important point about the use of non-mass action kinetics, namely using Hill functions, to model the regulation of one gene by another. It is true that this is not a "pure" Gillespie simulation of the full chemical master equation. However, given the commonly held assumption that the binding and unbinding of transcription factors to and from DNA is much faster than the rate of transcript production and degradation, a common way to model this binding is by using the quasiequilibrium assumption (Phillips et al. 2019), in which an average binding probability is used to modulate the transcription rate. The quasiequilibrium assumption allows us to accurately model transcriptional regulation while dramatically reducing computational costs, making our study feasible. We did not mention this clearly in the original manuscript but have now modified the text to discuss the approximation:

"In particular, we lump steps leading to transcription by implementing the commonly used quasiequilibrium assumption (Phillips et al. 2019), where binding and unbinding occurs much faster as compared to mRNA transcription and degradation."

Furthermore, we analyzed the impact of the dissociation constant k, which takes into account the steady state concentration of mRNA for each gene (see Box 2 and STAR Methods). As expected, for low values of k the threshold resulting in an effective gene activation is exceeded too often and the regulated DNA states are activated more frequently leading to the high gene expression states, and loss of rareness of the coordinated high gene expression event, eventually leading to bimodal distributions.

Explicit coupling between binding and transcription in a kinetic model will be interesting to model using Gillespie Algorithm or an appropriate approximation, but is beyond the scope of our current study.

- How much do we know about the specific genetic network, their players and the wiring in melanoma. Could you argue the network and its parameters is in the operating regime you have obtained.

The reviewer has raised an important point, namely whether the networks and parameters studied here can be compared to the melanoma-specific gene regulatory networks. Since the transient, rare, and coordinated nature of the variability has come to light only recently, we lack a quantitative wiring diagram of the actual gene regulatory networks and associated network parameters. Our lab has performed network inference algorithms on the RNA FISH imaging (this manuscript) and RNAseq data (Figure 5, Shaffer et al, 2018). We have found that these inferred networks contain genes that belong to a variety of pathways and span multiple chromosomes. While these efforts provide a first view of the putative gene networks, they need to be experimentally validated and also revised to identify directions and potential self-loops. Furthermore, the rare nature of the behavior makes it difficult to quantitatively characterize the genes and measure the parameters of their interactions.

We emphasize that our goal here is to capture the key features that describe this class of rare cell variability. We agree with the reviewer that it will be important to compare the experimentally identified gene regulatory networks with those identified by our model, and should be pursued in the future studies.

- Please rephrase this sentence in the introduction, it is not clear to me what you are saying: "However, in this classical context, most of the cells … "

Thanks to the reviewer for drawing attention to this sentence. We have now re-written this sentence and restructured the entire paragraph in the revised manuscript:

''These rare pre-resistant cells are marked by transient and coordinated high expression of dozens of marker genes. In other words, several genes are highly expressed simultaneously in a rare subset of cells, while the rest of the population have low or zero counts of mRNAs for these genes, resulting in a distribution of steady state mRNA counts per cell that peaks at or close to zero and has heavy tails. The rare cells in the tails, which transiently arise and disappear in the population by switching their gene expression state (Figure 1A), are much more likely to develop resistance to targeted therapies. Importantly, the rare and coordinated large fluctuations in the expression of multiple genes persist for several generations. Classical probabilistic models of gene expression have predicted the possibility of various types of mRNA expression distributions across a population, including normal, log-normal, gamma, or heavy-tail distributions (Antolović et al., 2017; Chen and Larson, 2016; Corrigan et al., 2016; Golding et al., 2005; Ham et al., 2019, 2020; Iyer-Biswas et al., 2009; Raj and van Oudenaarden, 2008; Raj et al., 2006; So et al., 2011; Symmons and Raj, 2016; Thattai and van Oudenaarden, 2001). It is unclear if such models can recapitulate the non-genetic variability characterized by rare and transient high expression states for several genes simultaneously (from now on referred to as "rare coordinated high states"), and if so, under what conditions.''

- Several references are missing full journal information, e.g Shaffer et al 2018, Saint Anthoine And Singh 2019, Corrigan et al 2016 Symons et al and Torre et al. Please check all references carefully.

We thank the reviewer for catching these errors on our part in the references section. We have now corrected these and all other references in the revised manuscript.

References:

Jiang, Lan, Huidong Chen, Luca Pinello, and Guo-Cheng Yuan. 2016. “GiniClust: Detecting Rare Cell Types from Single-Cell Gene Expression Data with Gini Index.” *Genome Biology* 17 (1): 144.

Phillips, Rob, Nathan M. Belliveau, Griffin Chure, Hernan G. Garcia, Manuel Razo-Mejia, and Clarissa Scholes. 2019. “Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression.” *Annual Review of Biophysics* 48 (May): 121–63.

Shaffer, Sydney M., Margaret C. Dunagin, Stefan R. Torborg, Eduardo A. Torre, Benjamin Emert, Clemens Krepler, Marilda Beqiri, et al. 2017. “Rare Cell Variability and Drug-Induced Reprogramming as a Mode of Cancer Drug Resistance.” *Nature* 546 (7658): 431–35.

Shaffer, Sydney M., Benjamin L. Emert, Ann E. Sizemore, Rohit Gupte, Eduardo Torre, Danielle S. Bassett, and Arjun Raj. 2018. “Memory Sequencing Reveals Heritable Single Cell Gene Expression Programs Associated with Distinct Cellular Behaviors.” *bioRxiv*, July. https://doi.org/10.1101/379016.

Van Kampen, N. G. 1992. *Stochastic Processes in Physics and Chemistry*. Elsevier.

Box 1. Model description, assumptions, parameters, and definitions

Model description: The transcriptional bursting model is comprised of single-gene expression modules described by the telegraph model: the DNA can take on an active and inactive state and transcribe mRNA at high and low rates (transcriptional bursting), respectively. These expression modules are coupled by an underlying network architecture, where regulation is modeled by a Hill function: the regulating gene influences the activation rate r_{on} of the respective regulated gene. The chemical reactions and propensities are described below:

Chemical reaction	Reaction propensity
$I \rightarrow A$	$(r_{on}+r_{add}*mRNA_X^n/(k^n+mRNA_X^n))*I$
$A \rightarrow I$	$r_{off} * A$
$I \rightarrow I + mRNA$	$r_{prod} * I$
$A \rightarrow A + mRNA$	$d*r_{prod} * A$
$mRNA \rightarrow \varnothing$	$r_{deg} * mRNA$

where $I,A \in \{0,1\}$, and $I+A = 1$, where $I = 0$ ($A = 1$) denotes that the DNA is in an active state and $I = 1$ ($A = 0$) denotes that the DNA is in an inactive state. $mRNA_X$ is the mRNA count of gene X at the given time. The model aims to recapitulate rare coordinated high states, where *rare* means that at the population level the expression distributions are unimodal and exhibit heavy tails; *coordinated* means that at least once throughout a simulation more than half the genes (nodes) show mRNA expressions above a specified threshold simultaneously; and *high* means that the mRNA expression of a gene exceeds a specified threshold (thres).

Model assumptions: (1) mRNA is able to influence the gene expression of its regulated gene directly, hence we refer to it as *gene product* throughout this work; (2) all genes are relationally identical (weakly-connected, non-isomorphic and symmetric gene regulatory networks); (3) all genes share the same model parameters; (4) gene regulation is only considered to be activating; and (5) if regulation occurs from several genes, their effects are additive. We discuss and check the generality of our model by testing many of these assumptions on a subset of cases, as described in **Box 2**.

Parameters: The model is described by 8 model parameters, as defined in the table below along with the corresponding ranges.

parameters		**sampling range**
independent model parameters		
r_{on}	The rate at which DNA is activated.	0.001 - 0.1
r_{off}	The rate at which DNA is inactivated.	0.01 - 0.1
r_{prod}	Synthesis rate of gene product.	0.01 - 1
r_{deg}	Degradation rate of gene product.	0.001 - 0.1
r_{add}	Parameter determining the contribution of the additional DNA activation rate upon gene regulation.	0.1 - 1
d	Factor by which the mRNA synthesis rate is increased when in an active DNA state. d >1.	2 - 100
n	Hill coefficient.	0.1 - 10
dependent model parameters		
k*	Dissociation constant of the Hill function, where $k(r_{prod}, r_{deg}, d) = 0.95 * d * r_{prod}/r_{deg}$	-
dependent classification parameters		
thres**	Threshold above which a gene is thought of being highly expressed, where $thres = 0.8 * d * r_{prod}/r_{deg}$	-

Here, r_{prod}/r_{deg} is the steady state in the baseline expression state (when there is no transcriptional burst) and $d * r_{prod}/r_{deg}$ is the steady state in the high expression state (if the DNA would continuously be in the active state).

Model Definitions:

- *weakly-connected network* - a directed network that when replacing the directed edges by undirected ones produces a connected graph in which every pair of nodes is connected by a path.
- *non-isomorphic* - two graphs are called non-isomorphic if there exists no structure-preserving bijection between them.
- *symmetric* - within a graph the number of in- and outgoing edges of a node and across nodes is identical and either all nodes in a network have a self-loop or not.
- *rare coordinated high state* - (1) at least once within a simulation more than half the genes are highly expressed simultaneously, (2) the histogram of simultaneously highly expressed genes at the population level decreases and (3) the gene expression distributions at the population are heavy-tailed.
- *connectivity* - number of ingoing edges for any node of the network.
- *characteristic distance*- the average shortest path length between pairs of nodes of the network.

*The parameter k is dependent on the parameters r_{prod}, r_{deg}, and d, such that: $k = x * d * r_{prod}/r_{deg}$, where $x \in \{0.75, 0.8, 0.85, 0.9, 0.95, 1\}$, which ensures a consistent definition of k throughout the network architectures and parameter sets. Here x represents the fraction of the value corresponding to the steady state value in the high expression state. We showed that for x = 0.75, none of the 100 simulations show rare coordinated gene expression because the threshold resulting in an effective gene regulation is exceeded too often—the regulated DNA states are activated more frequently leading to the high gene expression states and loss of rareness of the coordinated high gene expression event (leading to bimodal distributions). For x > 0.75, there is an increase in the number of simulations showing rare behavior, peaking at x = 0.95. Furthermore, throughout different values of x, the same parameter sets give rise to rare coordinated high states. We take x = 0.95 to maximize the number of simulations positive for the rare coordinated high states.

**We test several values for the threshold above which a gene is highly expressed: $thres = y * d * r_{prod}/r_{deg}$, where $y \in \{0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1\}$. For all $y \geq 0.6$, the set of simulations showing rare coordinated high states largely remains the same. Even for y = 0.3, half of the simulations identified previously to show rare behavior are still classified as such. We chose x = 0.8. Though arbitrarily chosen, the choice of x = 0.8 will not change the conclusions of our analysis.

Box 2.

Relaxing model assumptions

Protein translation: The original transcriptional bursting model does not include a step for translation and is assumed to be captured by the hill function term which not only greatly reduces the computational costs of long stochastic simulations but also allows for analyzing smaller set of parameters. To check if our model can produce rare coordinate high states even when the model includes the translation step, we focused on a particular network (5.3) and associated parameter values that give rise to these states in the original model. We show that for specific rates of translation and protein degradation (STAR Methods), the model including translation exhibits the rare coordinated high states.

Network architectures: By reducing the network architectures to weakly-connected, non-isomorphic and symmetric networks, we systematically reduce the number of possible network architectures. The reduced space of networks is partly supported by experimental observations (Shaffer et al. 2017, 2018), reporting that (1) there is no obvious hierarchical relationship between the expressed genes; and (2) no particular signaling pathway appears to be solely responsible for the observed behavior (see also Figure S1D). Furthermore, these network architectures allows for direct comparisons between network sizes, connectivities and parameter sets (not a given for other topologies). Although the analysis here primarily focuses on the constrained set of network architectures, we show for a subset of cases (STAR Methods) that asymmetric network architectures can also exhibit rare coordinated high gene expression states (Figure S2 G-I), paving the way for a more systematic analysis in the future studies.

Model parameters: While we primarily focus on keeping the same parameter set for each node, we analyzed a subset of networks with asymmetric parameters (STAR Methods) such that each node had distinct underlying parameter sets. We show that a model with asymmetric parameter sets is also capable of producing rare coordinated high gene expression states (Figure S2 J-M).

Multi-gene regulatory effects: The joint regulatory effects experienced by a gene which is regulated by several other genes can be modeled using different approaches. While the majority of analysis here uses an additive model of joint-regulation, we performed a subset of simulations (STAR Methods) for cases where the regulation by multiple gene nodes is multiplicative (Figure S4C and E). We find that for network architecture 5.3, 15 and 97 out of 1000 parameter sets give rise to simulations with rare coordinated high states in the additive and multiplicative joint-regulation, respectively (Figure S4D). Nine simulations are found to show rare coordinated high states in both definitions of multi-gene regulation.

Defining model-output metrics

Population level—sub-simulation size to determine a single cell: To qualitatively compare our results to experimental data, we convert the 1,000,000 time units long single-cell simulation to 1,000 single-cell sub-simulations of length 1,000 time units. We show that the simulations are largely (88.2%) uncorrelated after 1,000 time units, justifying our analysis (STAR Methods).

Heavy-tails: We test different levels of stringency in our definition of heavy-tailed/sub-exponential distributions. The analysis in Figure 2 is performed using the criteria described in STAR Methods, section Simulation classes. We perform further analysis similar to Figure 2 by using more stringent definitions, i.e. fit exponentials and compare the 99th percentiles (Figure S3C). We demonstrate that these results and conclusions are similar to the ones obtained using less stringent criteria (Box 1) shown in Figure 2 (see Figure S4F-M). For example, 6 and 7 out of 8 rare coordinated high parameter sets also appear in the two more stringent analyses (Figure S4H and L). We further validate that our model recapitulates the experimentally observed heavy-tails by comparing the Gini coefficients (Jiang et al. 2016) of experimental and model distributions (Figure S3D).

Number of nodes highly expressed to be called a 'coordinated' state: We define a simulation to show coordinated high gene expression if at least once throughout the simulation more than half of the gene product counts exceed the threshold. Furthermore, we show that for different node counts (2, 3, 4, 5) the number of simulations showing rare coordinated high states does not vary significantly. As an example, for a count of 2, we get 6 out of 100 simulations showing rare behavior; for a count of 3, we get 7. Note that the sets of simulations were overlapping between different scenarios.

Definition of rare coordinated high parameter sets: We define rare coordinated high parameter sets as parameter sets showing rare coordinated high expression in ≥ 20% of all 96 networks. The threshold was defined by inspecting the histogram (Figure 2H), where we see a separation at 20%. Notably, the same rare coordinated high parameter sets also appear in other analysis — they show increased frequencies of simulations with rare coordinated high states when considering the network sizes separately (Figure S6A). Additionally, stricter definitions for heavy-tailed expression distributions result in similar rare coordinated high parameter sets (Figure S4H and L).

Bootstrapping controls in Phixer algorithm: As the number of connections predicted by the Phixer algorithm can depend on the sample size, we bootstrapped the original data set into 4000-sample datasets. The number 4000 was chosen arbitrarily; bootstrapped sample sizes of 1000, 2000, and 6000 also produced qualitatively similar results.

Edge weight in Phixer algorithm: We created a randomized control consisting of permutations of each gene column from the original dataset. We then performed the Phixer analysis on these randomized controls. The resulting edge weight distributions give us a baseline or control edge weight for Phixer that, in principle, reflects potential false positives. We found that in the controls, nearly all of the predicted edge weights were below 0.45 (Figure S8B). Therefore, we decided to choose 0.45 as a threshold for our non-control analysis, thus eliminating edges that could have been predicted by chance alone.

Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer

Lea Schuh[1,2,3], Michael Saint-Antoine[4], Eric Sanford[1], Benjamin L. Emert[1], Abhyudai Singh[4], Carsten Marr[2], Yogesh Goyal[1,*], Arjun Raj[1,*,**]

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA
[2]Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, 85764, Germany
[3]Department of Mathematics, Technical University of Munich, Garching, 85748, Germany
[4]Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716, USA
*Authors for correspondence: yogesh.goyal0308@gmail.com, arjunrajlab@gmail.com
**Lead contact: arjunrajlab@gmail.com

SUMMARY

Non-genetic transcriptional variability at the single-cell level is a potential mechanism for therapy resistance in melanoma. Specifically, rare subpopulations of drug naive melanoma cells occupy a transient pre-resistant state characterized by coordinated high expression of several genes. Importantly, these rare cells are able to survive drug treatment and develop resistance. How might these extremely rare states arise and disappear within the population? It is unclear whether the canonical stochastic models of probabilistic transcriptional pulsing can explain this behavior, or if it requires special, hitherto unidentified molecular mechanisms. Here we use mathematical modeling to show that a minimal model comprising of transcriptional bursting and interactions between genes can give rise to rare coordinated high expression states. We next show that although these states occur across networks of different sizes, they occur more frequently in networks with low connectivity and depend strongly on three (of seven) independent model parameters. Interestingly, we find that while entry into the rare coordinated high state is initiated by a long transcriptional burst that also triggers entry of other genes, the exit from it occurs through the independent inactivation of individual genes. Finally, the transcriptional bursting model predicts that increased network connectivity can lead to transcriptionally stable states, which we verify using experimental data. In sum, we demonstrate that established principles of gene regulation are sufficient to describe the observed cell expression variability and argue for its general existence in other biological contexts.

Keywords: stochasticity, network, gene expression, melanoma, drug resistance, non-genetic

INTRODUCTION

Rare and large heterogeneity in single cells have been reported to arise from non-genetic transcriptional variability, even in clonal, genetically homogeneous cells grown in identical conditions (Fallahi-Sichani et al., 2017; Gupta et al., 2011; Pisco and Huang, 2015; Shaffer et al., 2017; Sharma et al., 2018, 2010; Spencer et al., 2009; Su et al., 2017). Importantly, cells exhibiting these non-genetic deviations are resistant to anti-cancer drugs (e.g., Ras pathway inhibitors) and may lead to relapse in patients. For example, in a drug naive melanoma population, a small fraction (~1 in 3000) of cells are pre-resistant, meaning they

are able to survive targeted drug therapy, resulting in their uncontrolled cellular proliferation (Shaffer et al., 2017). These rare pre-resistant cells are marked by transient and coordinated high expression of dozens of marker genes. In other words, several genes are highly expressed simultaneously in a rare subset of cells, while the rest of the population have low or zero counts of mRNAs for these genes, resulting in a distribution of steady state mRNA counts per cell that peaks at or close to zero and has heavy tails. The rare cells in the tails, which transiently arise and disappear in the population by switching their gene expression state (**Figure 1A**), are much more likely to develop resistance to targeted therapies. Importantly, the rare and coordinated large fluctuations in the expression of multiple genes persist for several generations. Classical probabilistic models of gene expression have predicted the possibility of various types of mRNA expression distributions across a population, including normal, log-normal, gamma, or heavy-tail distributions (Antolović et al., 2017; Chen and Larson, 2016; Corrigan et al., 2016; Golding et al., 2005; Ham et al., 2019, 2020; Iyer-Biswas et al., 2009; Raj and van Oudenaarden, 2008; Raj et al., 2006; So et al., 2011; Symmons and Raj, 2016; Thattai and van Oudenaarden, 2001). It is unclear if such models can recapitulate the non-genetic variability characterized by rare and transient high expression states for several genes simultaneously (from now on referred to as "rare coordinated high states"), and if so, under what conditions.

Might a stochastic system of interacting genes inside the cell facilitate transition in and out of the rare coordinated high state? One hypothesis is that within the canonical modeling framework, only a rare set of unique (and perhaps complex) networks can facilitate reversible transitions into the rare coordinated high states. Alternatively, relatively generic gene regulatory networks may be capable of producing such behaviors, suggesting that a large ensemble of such networks may admit rare-cell formation. Both of these scenarios have different implications—for instance, the latter hypothesis suggests that this behavior could be more common in biological systems than hitherto appreciated. The alternatives described above can also be posed in terms of the nature of model parameters—whether the set of values that give rise to rare coordinated high states are constrained to lie within a narrow window of parameter space or whether such behavior may occur across broad swaths of parameter space. Yet another possibility is that standard stochastic gene expression models fail to produce rare coordinated high states entirely, no matter what combinations of networks and parameters are used. In that case, one may argue that the reversible transition into the rare coordinated high state is driven by highly specialized processes (e.g. initiated by a master regulator) or other unknown mechanisms.

Here we describe a mathematical framework to test the hypotheses proposed above for the appearance and disappearance of rare coordinated high states (**Box 1**). Recent studies from our lab suggest that no particular molecular pathway is solely responsible for the formation of these rare cells (Shaffer et al., 2018; Torre et al., 2019). Specifically, in these rare cells, a sequencing and imaging based scheme identified a collection of marker genes, which are targets of multiple signaling pathways ranging from type 1 interferon to PI3K-Akt signaling. The implication is that instead of a single signaling pathway leading to the observed behavior, a network of interacting genes appears to be responsible. Accordingly, we used network modeling to see whether genes interacting within a network were capable of producing transitions to coordinated high expression states. We systematically formulated

and simulated networks of increasing size and complexity defined by a broad range for all independent parameters (Box 1 and 2; and **STAR Methods**, section Networks & section Parameters).

Computational screens on more than 96 million simulated cells reveal that many networks with interactions between genes are capable of producing rare coordinated high states. Critically, transcriptional bursting, a ubiquitous phenomenon in which genes flip between transcriptionally active and inactive state, is necessary for the transcriptional bursting model to produce these rare coordinated high states. Subsequent quantitative analysis shows that rare coordinated high states occur across networks of all sizes investigated (up to 10 nodes), but that (i) they depend on three (out of seven) independent model parameters and (ii) their frequency of occurrence decreases monotonically with increasing network connectivity. The transition into the rare coordinated high state is initiated by a long transcriptional burst, which, in turn, triggers the entry of subsequent genes into the rare coordinated high state. In contrast, the transition out of rare coordinated high state is independent of the duration of transcriptional bursts, rather it happens through the independent inactivation of individual genes. We also confirm model predictions using experimental gene expression data (RNA FISH data) taken from melanoma cell lines. Together, we demonstrate that the standard model of stochastic gene regulation with transcriptional bursting is capable of producing rare coordinated high states.

RESULTS

Framework selection

Identifying the minimal network model generating rare coordinated high states

We focused on a network-based mathematical framework that models cell-intrinsic biochemical interactions and wondered what would be the minimal set of biochemical reactions that constitutes it. Since network models without gene activation (i.e. constitutive mode of gene expression) were not able to produce rare coordinated high states (see **Supplementary information**; **Figure 1B and Figure S1A-B**; **STAR Methods**, section Model 1), we use a leaky telegraph model as the building block of our framework. In terms of chemical reactions, a gene can reversibly switch between an active (r_{on}) and inactive state (r_{off}), where binding of the transcription factor at a gene locus controls the effective rate of gene production (**Box 1**; **Figure 1C, STAR Methods**). Specifically, when inactive (or unbound), the gene is transcribed as a Poisson process at a low basal rate (r_{prod}); when active, the rate becomes higher ($d \times r_{prod}$, where $d > 1$). We modeled degradation of the gene product as a Poisson process with degradation rate r_{deg}. The inter-node interaction parameter, r_{add}, has a Hill-function-based dependency on the gene product amount (Hill coefficient n) of the respective regulating node to account for the multistep nature of the interaction (**Figure 1C**). In particular, we lump steps leading to transcription by implementing the commonly used quasiequilibrium assumption (Phillips et al., 2019), where binding and unbinding occurs much faster as compared to mRNA transcription and degradation. The dissociation constant k of the Hill function is dependent on the parameters r_{prod}, r_{deg}, and d, such that $k(r_{prod}, r_{deg}, d) = 0.95 * d * r_{prod}/r_{deg}$. In total, the model has seven independent and

one dependent model parameters, as outlined in Box 1. All chemical reactions, propensities, and model parameters are presented in **STAR Methods**. We used Gillespie's Stochastic Simulation Algorithm (Gillespie, 1977) to systematically simulate networks of various sizes and architectures across a broad range of parameters (**Box 1**; **STAR Methods**, section Networks & section Parameters).

We limited our study to networks that are symmetric, i.e., networks without a hierarchical structure (**Box 1**; **STAR Methods**, section Networks, **Figure S1C**), a simplification partially supported by the experimental observation that there doesn't seem to be a clear directionality of regulation or hierarchical structure within the highly expressed genes in the rare high state in drug naive melanoma (**Figure S1D**) (Shaffer et al., 2017, 2018). The lack of hierarchy is inferred from the frequency matrix for the experimental RNA FISH data, in which each entry corresponds to the fraction of cells with each gene-pair being highly expressed (**Figure S1D**). Asymmetric networks can result in frequency matrices being highly asymmetric, as demonstrated by an example simulation of a star-shaped reaction network (**Figure S1E-F**). Symmetric models also allow for comparisons of parameters between networks of different sizes. Additionally, we excluded networks that are compositions of independent subnetworks (non weakly-connected networks) and networks that can be formed by structure-preserving bijections of other networks (isomorphic networks) (**STAR Methods,** section Networks, **Box1**). With these operations, we also reduce the testable space of unique networks by several orders of magnitude (**Figure S1C**).

Characterization of the transcriptional bursting model

When genes are organized in the system described above and simulated over long intervals, the transcriptional bursting model produced a range of temporal profiles for gene products (**Figure 1D-G** and **Figure S2A**). Importantly, the model was able to faithfully capture the qualitative features of experimental data, i.e., rare, transient, and coordinated high expression states (**Figure 1G**). We defined a set of rules to screen for the occurrence of different classes of states (**Figure 1D-G** and **Figure S2A**); these include stably low expression (class I), stably high expression (class II), uncoordinated transient high expression (class III), and rare transient coordinated high expression (class IV) (see **STAR Methods**, section Simulation classes), and used a heuristic approach to distinguish between these different classes (**Boxes 1** and **2**). For a detailed description of the rules and quantitative metrics used to define class IV, see **Boxes 1** and **2**; **Figure S3** and **Figure S4**; and **STAR Methods**, section Simulation classes.

To better compare the computational results with the experimental data from static RNA FISH images, we split the entire simulation into non-overlapping time interval of 1000 time units, as justified by the ergodic theory (**Box 2** and **STAR Methods**). We took snapshots of gene products at randomly selected time points in these time-intervals and noted the number of simultaneously highly expressed genes as well as their gene product counts, allowing us to represent the static states of a population of simulated cells (**Figure 2A**). For example, in a particular 8-node network, we found that the distribution qualitatively captures the experimental observations where most cells do not exhibit high expression states, while some cells are in a high state for one or more genes **(Figure 2B)**. Note that unlike the

experimental data, the model simulation does not have non-zero values for higher number of genes. The absence of of non-zero values may be because the network underlying the experimental data contains a much larger set of interacting genes, thereby increasing the likelihood of non-zero values for higher number of expressed genes. Similarly, when we selected a gene and plotted its product count for the randomly selected time points, we observed a heavy-tailed distribution (**Figure 2C, right panel**), similar to the experimental observations **(Figure 2C left panel** and **Figure S3A)**. Furthermore, these observations, while shown for a particular 8-node network, hold true for simulations of other 8-node networks as well as networks of other sizes (**Figure S2B**). Note that the distributions of gene product counts for each gene are qualitatively similar because of the symmetric nature of the networks (**Figure S2C**). The experimental data in drug naive melanoma cells for mRNA counts display different degrees of skewness of the distribution for different genes (**Figure S3A**) which may be recapitulated by introducing asymmetries in the networks. Two asymmetric networks we tested were both able to produce rare coordinated high states (**Figure S2G-S4M**). Importantly, the distributions of gene product counts for various genes displayed different levels of heavy-tails, as also observed in the experimental data (**Figure S2M**). Since there is both inter- and intra-gene variability between the experimental expression distributions characterizing these states (**Figure S3A**), we compared these expression distributions to simulated expression distributions using Gini coefficients, used to characterize experimental expression distributions in the original study (Shaffer et al, 2017). While the Gini coefficient is low for most of the simulations (99.2%, gray), it is much higher for the simulations that produce rare coordinated high states (red) and overlaps with experimental Gini coefficients (**Figure S3D**). Together, the transcriptional bursting model is able to produce states which recapitulate key aspects of rare coordinated high states observed in drug naive melanoma.

Rare coordinated high states depend on network topologies and model parameters

Since the rare coordinated high states occur in <1% of all simulations (**Figure S2A**), we wondered whether their occurrence depends on the network topologies and/or model parameters. Specifically, what are the features of the topologies and parameters that facilitate the occurrence of rare coordinated high states? For the simulations that produced rare coordinated high states, we extracted and quantitatively analyzed the corresponding networks. We found that the rare coordinated high states occur ubiquitously in networks with different numbers of nodes analyzed (up to 10 nodes) (**Figure 2D and Figure S2B-F, Figure S5A-B**). We therefore hypothesize that even larger networks may also display rare coordinated high states, and can be explored in future studies. Next, we wondered if the occurrence of rare coordinated high states depends on the network connectivity (**Box 1**). Indeed, within a particular network size, the ability to produce rare coordinated high states decreases monotonically with increasing network connectivity (**Figure 2E and Figure S5C-D**). Consistently, the fraction of networks per network size (normalized by either network size or total networks per network size) exhibiting rare coordinated high states decreases with increasing size (**Figure S5A-B**) as a larger fraction of high connectivity networks exist in bigger networks (**Figure S5D**).

We next wondered whether gene auto-regulation (networks with self-loops) have any effect on a networks ability to produce the rare coordinated high states. Indeed we found that for a fixed size and connectivity, networks with auto-regulation result in higher numbers of simulations with rare coordinated high states than networks without auto-regulation (**Figure S5E**). At the same time, we wondered whether connectivity or auto-regulation has a more dominating effect on a networks' ability to produce these states. We found that adding self-loops on otherwise identical networks reduced the occurrence number of simulations with rare coordinated high states (**Figure 2F**), demonstrating the stronger effect of connectivity than auto-regulation. Finally, we analyzed network topologies based on characteristic length, defined as the average shortest path length between pairs of nodes of the network (see **STAR Methods, Box1**). Characteristic length recapitulates the effects of not only network connectivity (inversely correlated with characteristic distance), but also differentiates topologies with the same connectivity (**Figure S5F)**, for example networks with or without auto-regulation. Together, we demonstrate that the occurrence of rare coordinated high states depends on network topologies.

Since the transcriptional bursting model has seven independent parameters (r_{on}, r_{off}, r_{prod}, r_{add}, r_{add}, d, and n; see **Box 1** for details), we asked whether specific parameter combinations preferentially give rise to the rare coordinated high states, and if so, what features of such combinations facilitate it. The subsequent analysis is motivated by the initial observation that occurence of different classes of temporal gene product profiles across different network sizes and connectivities appear to also depend on the parameter sets (**Figure 2G**). Specifically, if a parameter set gave a specific expression profile (e.g. rare coordinated high or stably high) for one network, it displayed a higher propensity to display the same profile for other networks as well (**Figure 2G** and **Figure S3E**), implying that parameters indeed play a major role in the occurrence of rare coordinated high states. To avoid biases in the parameter sets investigated, all 1,000 parameter sets were sampled from a broad range for each parameter using a Latin Hypercube Sampling algorithm (**Supplementary Information** ParSetsAnalysis.xlsx**; STAR Methods**, section Parameters).

We first measured the percentage of simulations per parameter set that gave rise to the rare coordinated high states. Out of the 1,000 parameter sets, eight parameter sets, from now on called rare coordinated high parameter sets (**Box 2**), clustered together at the tail-end of the distribution (orange, **Figure 2H**), meaning they generated simulations with frequent occurrence of rare coordinated high states in at least 20% of all networks tested (**Figure 2H**). Furthermore, these eight parameter sets robustly generated rare coordinated high states across all network sizes and architectures (**Figure S6A**). Therefore, we wondered if these eight parameter sets have any special or distinguishing features compared to the remaining 992 parameter sets.

We used a decision tree algorithm (Breiman et al., 1984) (see **STAR Methods**, section Decision tree optimization and generalized linear models) to identify the differentiating features of the rare coordinated high parameter sets from the rest. The decision tree analysis revealed that only three (r_{on}, r_{off}, and r_{add}) of the seven independent parameters showed a strong correlation with the rare coordinated high parameter sets (**Figure 2I**). We validated these findings with complementary analysis using generalized linear models

(**STAR Methods**, section Decision tree optimization and generalized linear models) where we found precisely these three specific parameters (r_{on}, r_{off}, and r_{add}) to be critical to produce the rare coordinated high states with high statistical significance (p values: r_{on} = 0.003; r_{off} = 0.005; r_{add} = 0.014) (**Figure S6B**). These observations became readily evident when we plotted all the 1,000 parameter sets for r_{on}, r_{off}, and r_{add} together and found the rare coordinated high parameters sets to occupy a narrow region of the parameter phase space (**Figure 2J** and **Figure S6C**). Furthermore, parameter sensitivity analysis across the parameter space also confirmed that these three parameters are indeed critical for producing the rare coordinated high states (**Figure S6D**). These three parameters are related to transcriptional bursting and inter-gene(node) regulation. Two of these parameters, r_{on} and r_{off}, define the transitioning between the active and inactive state of the DNA respectively. The third parameter is the gene activation rate, r_{add}, which corresponds to the positive regulation of transcriptional bursting rate of a gene by the gene product of another interacting gene. Interestingly, too high values (> 0.31) of r_{add} result in the disappearance of rare coordinated high states, as does a complete absence (r_{add} = 0) of this term (**Figure S6E-S6G**). To confirm that these three parameters (r_{on}, r_{off}, and r_{add}) and their corresponding range of values are indeed critical to producing simulations with rare coordinated high states, we sampled new 1,000 parameter sets from a constrained region containing all eight rare coordinated high parameter sets (**Figure 2J, orange box,** and **STAR Methods**) and ran simulations for two test networks, a 3-node and a 5-node network. We found that the frequency of simulations with rare coordinated high states for the constrained region is ~14-fold and ~21-fold higher than that for the original parameter space, respectively (**Figure 2K**). We note that while parameter sets with parameters r_{on}, r_{off}, and r_{add} within the identified critical parameter ranges give rise to simulations with rare coordinated high states much more frequently than other parameter sets, it is not 100% of the time.

Distinct mechanisms regulate the transition into and out of rare coordinated high states

We have identified the networks and parameter sets for which the transcriptional bursting model exhibits rare coordinated high states more frequently. Next, we wondered if we could dissect the features of the model that facilitate the occurrence of rare coordinated high states. Specifically, we wanted to know the factors that 1) trigger the entry into the rare coordinated high states, 2) facilitate its maintenance, and 3) trigger the escape from it. We began by analyzing various features of transcriptional activity, since including transcriptional bursting was found to be critical for the model to display the rare coordinated high states. These include the burst fraction, length of transcriptional bursts (burst duration) and burst frequency. To measure these features, we defined four regions for each simulation: low expression state (baseline time-region), entry into the high expression state (entry time-point), the high expression state (high time-region), and exit from the high expression state (exit time-region) (**Figure 3A, STAR Methods**, section Entry and Exit mechanisms).

We found an increase in the transcriptional activity, as measured by the burst fraction, during the high expression time-region as compared to the baseline time-region (**Figure 3B**). Increased burst fraction could be a result of (1) longer transcriptional bursts or (2) a higher burst frequency. The former is not possible as the duration of each burst is distributed

exponentially according to exp(r_{off}), which does not change between the baseline and high time-region. Indeed, we found an increase in the burst frequency in high time-region, thus establishing its role in the maintenance of the rare coordinated high state (**Figure 3C**). The increased transcriptional bursting seen in the transcriptional bursting model is consistent with the experimental data using labeled nascent transcripts which showed that the transcriptional activity occurred in frequent bursts in cells high for a marker gene (Shaffer et al., 2018). Next we wondered whether the increase in burst frequency is promoted by the interactions of genes organized within the network. We compared two networks of the same size (3 nodes), where one is comprised out of single unconnected (orphan) nodes and the other of an interdependent structure (network 3.2). We found that for any parameter set (screened for all 26 parameter sets giving simulations with rare coordinated high states in the previous analysis for network 3.2, **Supplementary Information**), the system with a connected network has (1) more high expression states and (2) prolonged time in high expression states, as compared to unconnected nodes (**Figure 3D**). Together, we find that the maintenance in the high state is because of increased burst frequency, which may be a result of the positive regulatory feedback intrinsic to the networks.

Next, we wanted to identify the factors triggering the entry into the rare coordinated high states. We found that for any gene in the network, the transcriptional burst duration right before/during the entry into a rare coordinated high state was significantly higher than that in the baseline time-region (i.e., regular bursting kinetics) (**Figure 3E**). In the example shown in **Figure 3E**, the average time of transcriptional burst at the entry time-point is 84.82 (time units) as compared to only 15.08 (time units) in the baseline time-region. Therefore, prolonged transcriptional bursts play a role in driving the cell to a coordinated high expression state. Conversely, we asked if the opposite is true at the exit time-region, such that transcriptional bursts for the exit time-region are shorter than for the high time-region. We found no difference in the distributions of burst durations between the high and the exit time-regions, as demonstrated by the example in **Figure 3F**, suggesting that the exit from high expression state occurs independently of the burst durations. Importantly, both of these conclusions hold true when measured for all simulations with rare coordinated high states (**Figure 3G**). Together, unlike the entry into the high time-region, the exit from it is not dependent on the transcriptional burst duration.

We also wondered if the entry into the high expression state of one gene influences the entry of other genes, or that the genes enter the high expression state independently of each other. We reasoned that if the time duration between two successive genes (t_{ent}, **Figure 3A**) entering the high expression state is exponentially distributed, it would imply that the genes enter the high expression state independent of each other. Instead, we found that the distributions of entry time intervals rejected the null-hypothesis of the Lilliefors' test for most of the simulations (84%), meaning they are not exponentially distributed (**Figure 3H**). The remaining 16% of cases were found to be largely falsely identified as exponentially distributed due to limited data (see a representative example in **Figure S7A)**. Similarly, we tested if the exit for successive genes from the high expression state occurs independent of each other. Interestingly, contrary to the situation during the entry into the high expression state, many distributions of exit time intervals satisfied the null-hypothesis of the Lilliefors' test, implying they are indistinguishable from exponential distributions (**Figure 3I**). The

simulations that did not satisfy the stringent Lilliefors' test mainly appear to be exponentially distributed nevertheless; a representative example is shown in **Figure S7B**. Together, the entry into and exit from the rare coordinated high state occur through fundamentally different mechanisms—the entry of one gene into the high expression state affects the entry of next gene, while they exit from it largely independently of each other. The exit from the high state could be a result of weak strength of coupling (as reflected by the moderate values of parameter r_{add}) between nodes for the simulations that produce these states. Consistently, we found that too high values of r_{add} results in the disappearance of rare coordinated high states, giving way to stable high states. In other words, the network can transition into the high expression state but loses the ability to come out of it (**Figure S6 E-G**).

Increasing network connectivity leads to transcriptionally stable states

So far, we have used the transcriptional bursting model to understand the potential origins of rare pre-resistant states in drug naive melanoma cells. Upon treatment with anti-cancer drugs, the transient pre-resistant cells reprogram and acquire resistance resulting in their uncontrolled proliferation. The resistant cells are characterized by the stabilization of the high expression of the marker genes which were transiently high in the drug naive pre-resistant cells (**Figure 4A**) (Shaffer et al., 2017). Studies using network inference of gene expression data have suggested that the genetic networks undergo significant rearrangements upon cellular transitions or reprogramming (Moignard et al., 2015; Schlauch et al., 2017). We wondered if the transcriptional bursting model can explain how the transient high expression in drug naive cells might become permanent upon treatment with anti-cancer drugs. The modeling framework produces a range of gene expression profiles, depending on the network properties and model parameters (**Figure 1D-G**). Increasing the network connectivity (for fixed parameter sets) is one way to shift from a rare transient coordinated high expression state to stably high expression state (**Figure 4B-E**). As an example, for a fixed network size (five) and associated parameters, increasing the network connectivity from one to five resulted in a shift from transient coordinated to stably high expression states (**Figure 4D** and **Figure 4E**). The shift from transient coordinated to stably high expression states is also reflected by the bimodal distribution of genes product counts for in the highly connected network (**Figure 4F** and **Figure 4G**), where genes stay permanently in the high state once they leave the low expression state. These results mimic the experimentally measured mRNA expression states of the drug-induced reprogrammed melanoma cells.

To test if the computational prediction holds true in melanoma, we performed network inference using φ-mixing coefficient-based (Ibragimov, 1962) Phixer algorithm (Singh et al., 2018) on the experimental data (**Box 2**; **STAR Methods**, section Comparative Network Inference). Specifically, we used the Phixer algorithm on the mRNA counts obtained from fluorescent in situ hybridization (FISH) imaging data of marker genes in drug naive cells and the resistant colonies that emerge post-drug treatment to infer the underlying network. Consistent with the model prediction, we found that the number of edge connections (for a range of edge weight thresholds) between marker genes increased substantially for 6/7 resistant colonies compared to the drug-naive cells (**Figure 4H**). To control for biases from subsampling of the experimental data and the nature of Phixer algorithm itself (see **STAR Methods**, section Comparative Network Inference), we ran the entire network inference

analysis 1,000 times. Again, in all 1,000 runs, we saw a higher number of total edges for 6/7 resistant colonies compared to the drug-naive cells (**Figure 4I, Figure S8A** and **Figure S8C**).

Besides the dependence on networks, our framework predicts that for a given network, stronger interactions between nodes (defined by the interaction parameter r_{add}) can also result in stable gene expression profiles (**Figure S6E-S6G**). It is possible that reprogramming results from a combination of increased edge connectivity as well as the enhanced interactions (given by parameter r_{add}) between existing edges. Biologically, it would translate into stronger and increased number of interactions between genes and associated transcription factors during reprogramming. Together, network inference of the experimental data is consistent with model findings about the cellular progression from a transient coordinated high expression state to a stably high expression state.

DISCUSSION

We developed a computational framework to model rare cell behaviors in the context of a drug naive melanoma population where a rare subpopulation of cells displays transient and coordinated high gene expression states. We found that a relatively parsimonious stochastic model consisting of transcriptional bursting and stochastic interactions between genes in a network is capable of producing rare coordinated high states that mimic the experimental observations. To systematically investigate their origins, we screened networks of increasing sizes and connectivities for a broad range of parameter values. Our study revealed that they occur more frequently for networks with low connectivity and depend on 3/7 independent model parameters. Furthermore, we showed that the mechanisms that lead to the transition into- and out of- the rare coordinated high state are fundamentally different from each other. Collectively, our framework provides an excellent basis for further mechanistic and quantitative studies of the origins of rare, transient, and coordinated high expression states.

Given the relative generality of the networks that produce rare coordinated high states, the transcriptional bursting model predicts that every cell type is capable of entering the rare coordinated high state. Furthermore, we show that canonical modes of transcription alone, namely the binding of the transcription factor at gene locus to produce mRNA via recruitment of RNA Polymerase II, can lead to these states without requiring other complex mechanisms such as DNA methylation, histone modifications, or phase separation. While such other mechanisms may still be operational in these cells to regulate their entry to or exit from these states, we posit that in principle, any set of genes interacting *via* traditional gene regulatory mechanisms are capable of exhibiting these rare coordinated high states, as long as they are interacting in a certain manner (e.g. sparsely connected) with appropriate kinetic parameters. In the case of drug naive melanoma cells, the transient state is characterized by an increased ability to survive drug therapy leading to uncontrolled proliferation of the resulting resistant cells. It is possible that these rare transient behaviors may exist across many sets of interacting genes which may or may not manifest into phenotypic consequences. Another possibility the transcriptional bursting model predicts is that even within the same cell, distinct modules of interacting genes can lead to distinct sets of rare coordinated high states that each can affect the cellular function and outcomes differently.

These possibilities can be tested for by using increasingly accessible single cell RNA sequencing techniques on clonal population of cells.

The transcriptional bursting model also makes two key predictions regarding transitions into and out of rare coordinated high states. The first is that prolonged transcriptional bursts drive entry into the high expression state while exit from it is independent of the burst duration. The second is that genes entering the high expression state promote the entry of subsequent genes, whereas genes exiting the high expression state do so independently of each other. Both these predictions can be readily tested experimentally by simultaneous visualization of transcriptional bursting and mRNA counts using live cell (e.g. by using RNA-binding fluorescent proteins) or fixed cell (intron and exon FISH) imaging approaches (Bartman et al., 2016; Rodriguez et al., 2019).

Additionally, we showed that increasing the network connectivity is one way to reach a drug-induced reprogrammed state, a prediction we verified by performing network inference analysis on the experimental data. The transcriptional bursting model proposes that there are many plausible ways to transition from networks that produce transient coordinated high expression states to stable high expression states. For example, the transition could be facilitated by different amounts of increases in connectivity between nodes (genes) and/or changes in parameter values of the gene expression model. Furthermore, it is possible that these changes may take place only for a subset of nodes and edges within the network. These computational scenarios suggest that there could be significant heterogeneity in the stable expression levels of network genes in the resistant colonies emerging even from clonal population of drug naive cells, a possibility that can be tested experimentally by isolating individual colonies and profiling them for molecular markers to identify the paths. Identification of dominant paths has relevance for rational targeted drug therapy design. Therefore, in addition to modeling rare-behaviors, our framework can be adapted for investigating the plasticity and reprogramming paradigm in cancer.

One limitation of the transcriptional bursting model is that we have performed quantitative analysis only on symmetric networks with positive interactions between nodes. While the preliminary analysis on two cases of randomly selected asymmetric networks shows that they do exhibit the rare coordinated high states (**Figure S2G-S4M**), it remains to be seen whether these findings hold more generally for asymmetric networks. Inhibitory interactions between nodes is a separate and perhaps more interesting point. In principle, the model can be adapted to include inhibitory interactions. These inhibitory interactions may lead to non-monotonic effects of network connectivity on the occurrence of rare states, as positive and negative interactions can compete in non-linear ways. Similarly, a network with both negative and positive interactions may be more prone to instability, even for relatively smaller networks. Furthermore, inclusion of these interactions might also make the exit of genes from the high expression state dependent on one another, which occurs independently in the transcriptional bursting current model.

While we have focused on rare, transient, and coordinated high expression states in melanoma, our study provides conceptual insights into other biological contexts such as stem cell reprogramming. Particularly, there is increasing evidence to suggest that stem cell

reprogramming to desired cellular states proceeds *via* non-genetic mechanisms in a very rare subset of cells (Hanna et al., 2009; Pour et al., 2015; Takahashi and Yamanaka, 2016). The transcriptional bursting model may explain the origins and transient nature of this type of rare cell variability. In sum, we have established the plausibility that a relatively parsimonious model comprising of transcriptional bursting and stochastic interactions of genes organized within a network can give rise to a new class of biological heterogeneities. Therefore, we believe that established principles of transcription and gene expression dynamics may be sufficient to explain the extreme heterogeneities that are being reported increasingly in a variety of biological contexts.

SUPPLEMENTAL INFORMATION

Supplemental Information includes 9 figures and 2 tables.

ACKNOWLEDGEMENTS

We thank the Raj lab members, especially Ian Mellis and Amy Azaria, for scientific discussions and comments on the manuscript. We also thank Ravi Radhakrishnan and Alok Ghosh for helpful discussion during the initial stages of this project. We thank Cesar A Vargas-Garcia for his help during the initial discussions on network inference. A.R. acknowledges NIH/NCI PSOC award number U54 CA193417, NSF CAREER 1350601, P30 CA016520, SPORE P50 CA174523, NIH U01 CA227550, NIH 4DN U01 HL129998, NIH Center for Photogenomics (RM1 HG007743), and the Tara Miller Foundation. C.M. acknowledges support from the Deutsche Forschungsgemeinschaft DFG through the SFB 1243. A.S. acknowledges support from the NIH grant 5R01GM124446-02. L.S. would like to acknowledge the support of the PROMOS fellowship of the DAAD, Germany. Y.G. would like to acknowledge the Schmidt Science Fellows in partnership with the Rhodes Trust. Y.G. is a fellow of The Jane Coffin Childs Memorial Fund for Medical Research and this investigation has been aided by a grant from The Jane Coffin Childs Memorial Fund for Medical Research.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.G. and A.R.; Methodology, L.S., A.R. and Y.G.; Software, L.S. and A.R.; Validation, L.S.; Formal Analysis, L.S. and M.S.A.; Resources, A.R. and A.S.; Investigation, B.E., E.M.S. and Y.G.; Data Curation, L.S. and Y.G.; Writing - Original Draft, Y.G.; Writing - Review & Editing, A.R., L.S., Y.G., C.M., E.M.S., B.E. and M.S.A.; Visualization, L.S. and Y.G.; Supervision, Y.G., A.R. and C.M.; Project Administration, Y.G. and A.R.; Funding Acquisition, A.R., A.S. and C.M.

DECLARATION OF INTERESTS

A.R. receives royalties related to Stellaris RNA FISH probes. All other authors declare no conflict of interests.

MAIN FIGURE CAPTIONS

Figure 1. A transcriptional bursting model is able to mimic the rare coordinated high states observed in drug naive melanoma.
(A) Drug naive melanoma cells exist in low (white cells) as well as rare coordinated high (blue cells) states. Cells in the rare coordinated high state characterize the pre-resistant state observed in drug naive melanoma. A schematic of the corresponding expression pattern is shown in the panel below. The cells in a high expression state are more likely to survive and acquire resistance upon drug administration.
(B) Schematic of the constitutive model for two nodes. Gene product is either produced at rate r_{prod} or degraded with rate r_{deg}. Gene regulation is modeled by a Hill function, where the gene product count of the regulating gene A increases the production rate of the gene product of the regulated gene B.
(C) Schematic of the transcriptional bursting model for two nodes. DNA is either in an inactive (off) or active (on) state. Transitions take place with rates r_{on} and r_{off}, where gene product is synthesized with rates r_{pod} and $d*r_{prod}$, respectively, $d>1$. Gene product degrades with rate r_{deg}. Gene regulation is modeled by a Hill function, where the gene expression of the regulating gene A increases the activation of the DNA of the regulated gene B.
(D-G) Depending on the network and the parameters of the transcriptional bursting model, we observe stably low expression (D), stably high expression (E), uncoordinated transient high expression (F) and rare transient coordinated high expression (G).
See also Figure S1.

Figure 2. Simulations of the transcriptional bursting model show similar behavior at the population level as the drug naive melanoma cells.
(A) Frame of simulation showing rare coordinated high state (shaded area). The 1,000,000 time unit simulation is split into frames of 1,000 time units to create a simulated cell population (shown for cell N). For a randomly determined time-point t_{rand}, the number of simultaneously highly expressed genes and the gene count per gene per cell are evaluated. The network of the corresponding simulation is given in the top left corner.
(B,C) The simulated number of simultaneously highly expressed genes and expression distribution at the population level are qualitatively similar to experimental data from a drug naive melanoma population (data from (Shaffer et al., 2017)). The percentages are indicated above the histogram (in B). The network and parameter set as well as the particular node (in C) used for comparison are shown in the right panel.
(D) Rare coordinated high states occur ubiquitously across networks of all analyzed network sizes (for three independent and randomly sampled t_{rand})(median, 25th and 75th percentiles).
(E) Increasing connectivity within all networks of size 5 leads to a decrease in the number of simulations with rare coordinated high states.
(F) Networks with auto-regulation exhibit rare coordinated high states less frequently than the same networks without auto-regulation.
(G) Simulations of a particular parameter set across different networks and sizes show largely the same class of gene expression profiles.
(H) The eight rare coordinated high parameter sets give rise to simulations with rare coordinated high states more frequently than others for all 96 networks and cluster at the tail of the histogram. The cut-off is defined at 20%.

(I) Decision tree optimization of resulting parameter set classes reveal that only three out of seven parameters, r_{on}, r_{off} , and r_{add}, show a strong correlation with the rare coordinated high parameter sets.
(J) Three dimensional representation of all tested 1000 parameter sets for r_{on}, r_{off} , and r_{add} show that the rare coordinated high parameter sets are narrowly constrained in the 3D space (orange dots). The orange box indicates the constrained parameter space enclosing all rare coordinated high parameter sets used for analysis in (K).
(K) The constrained subregion defined by the rare coordinated high parameter sets (see Figure 2J) heavily favors the formation of rare coordinated high states.
See also Figure S2, Figure S3, Figure S4, Figure S5 and Figure S6.

Figure 3. Rare coordinated high state is initiated by a long transcriptional burst, maintained by an increase in burst frequency and terminated according to a random process.
(A) An exemplary high region, with a baseline time-region, entry time-point, high time-region and an exit time-region. The time intervals for an additional gene to enter and exit the high region are marked by t_{ent} and t_{exit}, respectively. The bursts below the exemplary simulation are representative schematics.
(B) The system exhibits enhanced transcriptional activity, as measured by the burst fraction, when in the high time-region (two-sample Kolmogorov-Smirnov test, p-value < 0.001).
(C) The frequency of transcriptional bursts is increased in the high time-region (two-sample Kolmogorov-Smirnov test, p-value < 0.001).
(D) Positive regulatory interactions between the connected nodes (network) leads to an increased number of and total time in high states in comparison to independent nodes.
(E) The bursts during entry time-points are significantly longer than bursts not in a high time-region (two-sample Kolmogorov-Smirnov test).
(F) There is no statistical significant difference between the distributions underlying the duration of bursts in the high time-region and the exit time-region (two-sample Kolmogorov-Smirnov test).
(G) While the mean burst duration ratio between entry time-point and baseline time-region is significantly increased, the mean burst duration ratio between bursts in the exit time-region and in the rest of the high time-region are comparable for all simulations with rare coordinated high states. Ratio close to 1 suggests no difference between the two regions.
(H,I) The time intervals between genes entering and exiting the high time-region are distributed differently, as demonstrated by two representative simulations. While the time intervals for entering the high time-region are not exponentially distributed (H) (and hence not random), the time intervals for exiting the high time-region are exponentially distributed (I) (Lilliefors test, p-value < 0.001 and > 0.05, respectively).
See also Figure S7.

Figure 4. Increased connectivity of a network leads to stable high expression which is also observed in emerging resistant colonies post-drug treatment.
(A) Upon drug treatment, the surviving cells acquire stable resistance. A schematic gene expression pattern is shown below.
(B,C) Networks of size 5 with low (B) (1) and high (C) (5) connectivity and corresponding (D,E) simulations.

(F,G) The expression distribution underlying the simulation of the highly connected network (G) does not exhibit heavy-tails while the simulation of the lowly connected network (F) exhibits heavy-tails.
(H) Network inference analysis shows that 6 out of 7 resistant colonies (post drug treatment) have higher connectivity in comparison to two biological replicates of drug naive cells for many edge weight thresholds.
(I) Distribution of number of edges for the drug naive cells (red) is lower than an exemplary resistant colony (black) when the network inference analysis is run 1,000 times.
See also Figure S8.

STAR METHODS

Networks

In our framework, the nodes in the network represent genes, where the expression of a gene is regulated by the expression of other genes. Gene regulation is represented by directed edges in the network, e.g. if the expression of gene Y is regulated by the expression of gene X, then the network contains an edge from node X to node Y. These networks can be defined by adjacency matrices given by:

$$A_{ij} = \begin{cases} 1, \text{ if there is an edge from node i to j} \\ 0, \text{ else.} \end{cases}$$

Any node in a network of size N can be connected with up to N-1 other nodes and in the case of self-loops, to N other nodes. Hence, the adjacency matrix A is of size N*N. This means that there are 2^{NxN} possible adjacency matrices for a network of size N - each of the possible N*N matrix entries can take on one of the values of 0 (no edge) and 1 (edge). For example a network of size 3 has $2^{(3*3)} = 512$ possible networks.

Here, we focus on symmetric networks, where we assume a relational identity between all nodes in a network. Experimental data from Shaffer et al. (Shaffer et al., 2017) implies the absence of any obvious hierarchical structure within the genes, and that the driver genes may interact in a relatively non-hierarchical manner (**Figure S1D**). The structural embedding of a node in its network can increase or decrease its ability of being involved in coordinated overexpression. For example, a centered node within a star-shaped network is involved more frequently in coordinated overexpression than the other nodes within the same network (**Figure S1E**), which is inconsistent with the experimental observations. To ensure for non-hierarchical behavior we define a set of symmetric networks (**Figure S1F**), where the number of in- and outgoing edges within a node and across nodes is identical and either all nodes in a network have a self-loop or not, leading to adjacency matrices of which the rows are cyclic permutations (to the right) with offset one of each other. We first compute all possible vectors $\{0,1\}^N$, in total 2^N vectors. From each of these resulting vectors, we create an NxN matrix by using the given (row) vector as template, and creating the other N-1 rows by cycling the prior row vector to the right by one step, where the right-most entry in the row vector is added to the (so far empty) left-most entry. By applying this permutation N-1 times, all possible cyclic permutations are captured within a matrix, and each node in the given

network is completely relational identical. We make use of the *circshift* function in MATLAB to receive the possible cyclic permutations of the initial row vectors.

We further constrain the analysis to weakly-connected networks -- any node in a network has to be connected to at least one other node, without taking into account the directionality of the edges. In terms of the adjacency matrix:

$$\forall\ i \in \{1,...,N\}: \sum_{j \in \{1,...,N\},\ j \neq i} A_{ij} + A_{ji} \geq 1.$$

The above restriction allows us to exclude the consideration of compositions of smaller and unconnected networks, which could otherwise lead to double counting. These subnetworks of smaller sizes are analyzed in the sets of networks of respective node sizes. To perform this operation, we analyze all the previously constructed adjacency matrices using the MATLAB function *conncomp(X,'Type','weak')*, which assigns each node with a bin number according to the connected component of its underlying undirected graph. If all nodes of a network belong to the same bin number i.e. to the same connected component, the adjacency matrix encodes for a weakly-connected graph. Finally, we further restrict the analysis to non-isomorphic networks. Two networks are called isomorphic if there exists a bijection from the edge space of one network to the other, such that any edge of one network is projected to a particular edge in the other network. Here, the labeling of the nodes (gene 1, gene 2, ...) in the networks is arbitrary and hence relabeling of nodes in an adequate fashion leads to identical networks. To ensure that all the final networks analyzed are of a non-isomorphic set of networks, we test all networks with MATLAB's function *isisomorphic*. We initiate the final set of networks with one adjacency matrix, and then sequentially test all other networks for isomorphism. If the given network is non-isomorphic to the current final set, it is added to the final set. Conversely, if the network is isomorphic to one of the networks in the final set, it is discarded.

By reducing the possible set to weakly-connected, non-isomorphic and symmetric networks, we greatly reduce the possible number of networks. For example, in the previous example, we had 512 possible networks for 3 nodes. By applying all the mentioned constraints (weakly-connected, non-isomorphic and symmetric), 4 networks remain (**Figure S1C**). We perform the analysis on networks of sizes 2, 3, 5 and 8 each consisting of 2, 4,10 and 80 networks, respectively, adding up to a total of 96 networks (**Figure S9**). In principle, the transcriptional bursting model can easily be extended to larger network sizes without the loss of generality (**Figure S2D-F**).

Model 2 - Transcriptional bursting model

The transcriptional bursting model is an expansion of the telegraph model, where DNA can take on one of the two states, active and inactive, e.g. based on the presence or absence of transcription factors (**Figure 1C**). The active and inactive state directly translates into high and low rates of production of gene products, respectively. We add interaction terms to the model, where the expression of a gene influences the rate of DNA activation of another gene depending on how they are organized in a respective network. Here we use the number of mRNA as a faithful proxy for the number of proteins. In other words, we only model the number of mRNA counts and assume that any mRNA is immediately translated into one

single functional protein after its translation. Therefore, the mRNA count determines the strength of the regulation. Here, we model the regulation of one gene by another using the Hill function, given by:

$$f(mRNA_X) = mRNA_X^n/(k^n + mRNA_X^n),$$

where $mRNA_X$ is the mRNA count of gene X, n is the Hill coefficient and k is the dissociation constant, n,k > 0. The Hill coefficient determines the steepness of the Hill function, i.e., the extremeness of its switch-like effect. The dissociation constant determines the half-maximal value, $f(mRNA_X) = 0.5$.

The reversible transitions between the inactive and active states, as well as the mRNA synthesis and degradation, are modeled by chemical reactions. For each gene, we have three chemical species - the DNA inactive state, the DNA active state and mRNA. These three species interact with one another according to the following 5 chemical reactions:

$$\begin{aligned} I &\rightarrow A \\ A &\rightarrow I \\ I &\rightarrow I + mRNA \\ A &\rightarrow A + mRNA \\ mRNA &\rightarrow \varnothing, \end{aligned}$$

defining the corresponding stoichiometric matrix:

$$\begin{matrix} -1 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1. \end{matrix}$$

The stoichiometric matrix encodes the net change in each chemical species resulting from any of the chemical reactions where the chemical reactions are assumed to occur stochastically. Under the assumptions of the law of mass action, the probability of a specific molecular collision to occur in the infinitesimal time interval [t, t + dt) is proportional to the product of the molecule counts of the educt chemical species. The reaction propensity $a_j(x)$ for a given chemical reaction R_j and state x, determines the probability density function such that $a_j(x)dt$ gives the probability of the chemical reaction R_j taking place in dt, for small dt. Examples of reaction propensities for so called elementary reactions are given here:

Reaction	Reaction propensity
$\varnothing \rightarrow$ products	k
$X_i \rightarrow$ products	kx_i
$X_i + X_j \rightarrow$ products	kx_ix_j

where k is called the reaction rate.

The gene regulation influences the reaction rate of the DNA activating chemical reaction. To explain the above-mentioned chemical reactions, we introduce eight rates/parameters:

Parameter	Description
r_{on}	The rate at which DNA is activated.
r_{off}	The rate at which DNA is inactivated.
r_{prod}	Synthesis rate of mRNA.
r_{deg}	Degradation rate of mRNA.
r_{add}	Parameter determining the contribution of the additional DNA activation rate upon gene regulation.
d	Factor by which the mRNA synthesis rate is increased when in an active DNA state (in comparison to basal synthesis rate in DNA inactive state), >1.
k	Dissociation constant of the Hill function.
n	Hill coefficient.

The full model description for one gene regulated by a single gene X is given below:

Chemical reaction	Reaction rate	Reaction propensity
$I \rightarrow A$	$r_{on} + r_{add} * mRNA_X^n/(k^n + mRNA_X^n)$	$(r_{on} + r_{add} * mRNA_X^n/(k^n + mRNA_X^n)) * I$
$A \rightarrow I$	r_{off}	$r_{off} * A$
$I \rightarrow I + mRNA$	r_{prod}	$r_{prod} * I$
$A \rightarrow A + mRNA$	$d * r_{prod}$	$d * r_{prod} * A$
$mRNA \rightarrow \varnothing$	r_{deg}	$r_{deg} * mRNA$

where $I,A \in \{0,1\}$, and $I+A = 1$, where $I = 0$ ($A = 1$) denotes that the DNA is in an active state and $I = 1$ ($A = 0$) denotes that the DNA is in an inactive state. $mRNA_X$ is the mRNA count of gene X at the given time, r_{on} is the basal DNA activation rate, r_{add} is the additional activation rate due to gene regulation, r_{off} is the DNA inactivation rate, r_{prod} is the basal mRNA synthesis rate in the DNA inactive state, d denotes the increase in the mRNA synthesis rate when the DNA is in the active state, where $d > 1$, and r_{deg} is the mRNA degradation rate. The chemical reactions are identical for all N nodes in a given network of size N. The reaction rate of

activation (I → A), composed of terms with parameters r_{on} and r_{add}, is the only node-specific rate. It depends on the underlying network and has to be adapted accordingly for each node, where the in-going edges of a node determine which gene regulations are active. The addition of hill function-based activation terms corresponds to the adaptation of the standard telegraph model, highlighted in blue in the above rates. We model gene regulation additively: if there is more than one influencing gene, we add the Hill function terms of the respective genes. As an example, if the gene of interest is influenced not only by gene X, but by gene X and gene Y, the activation rate from above will expand to:

$$r_{on} + r_{add} * (mRNA_X^n/(k^n + mRNA_X^n) + mRNA_Y^n/(k^n + mRNA_Y^n)).$$

We also tested for multiplicative regulation, i.e. regulation where we multiply the reaction rates (and consequently the reaction propensities) of the influencing genes (**Figure S4C**). In the example above the activation rate then expands to

$$r_{on} + r_{add} * 2 *(mRNA_X^n/(k^n + mRNA_X^n) * mRNA_Y^n/(k^n + mRNA_Y^n))$$

instead. By definition the Hill function is restricted to values between 0 and 1. While a multiplication of two Hill functions results in a maximal value of 1, an addition results in a maximal value of 2. As the Hill function is an important factor in these simulations we hence add a scaling factor to the activation rate in case of multiplicative regulation. We show that for network 5.3, 97 out of 1000 simulations show rare coordinated high states in case of multiplicative regulation (**Figure S2D-E**). In comparison, 15 simulations show rare coordinated high states in case of additive regulation. 9 simulations show rare coordinated high states in both cases.

Additionally, we tested for translation events (**Figure S4A**). We added one state (P) and two rate parameters, a protein synthesis rate r_{prodP} and a protein degradation rate r_{degP}, to the original transcriptional bursting model. The extended model description accounting for translation for one gene regulated by gene X is given below:

Chemical reaction	Reaction rate	Reaction propensity
$I \rightarrow A$	$r_{on}+r_{add}*P_X^n/(k^n+P_X^n)$	$(r_{on}+r_{add}*P_X^n/(k^n+P_X^n))*I$
$A \rightarrow I$	r_{off}	$r_{off} * A$
$I \rightarrow I + mRNA$	r_{prod}	$r_{prod} * I$
$A \rightarrow A + mRNA$	$d*r_{prod}$	$d*r_{prod} * A$
$mRNA \rightarrow \varnothing$	r_{deg}	$r_{deg} * mRNA$
$mRNA \rightarrow mRNA + P$	r_{prodP}	$r_{prodP} * mRNA$
$P \rightarrow \varnothing$	r_{degP}	$r_{deg} * P$

where we define k again as 0.95 of the high steady state, this time for the protein count:

$$k(r_{prod},r_{deg},d) = 0.95 * r_{prodP}/r_{degP} * d * r_{prod}/r_{deg},$$

which itself is dependent on the high steady state of the mRNA ($d * r_{prod}/r_{deg}$). Redefining $r_{prodP} = a * r_{prod}$ and $r_{degP} = b * r_{deg}$ gives

$$k(r_{prod},r_{deg},d) = 0.95 * d * a/b * r_{prod}^2 * r_{deg}^2.$$

We tested three different translation scenarios: protein synthesis and degradation being (1) faster than (2) same as and (3) slower than mRNA synthesis and degradation. For network 5.3 and parameter set 968, giving rise to rare coordinated high states in the transcriptional bursting model without translation, we took a = b = 10 (faster), a = b = 1 (same) and a = b = 0.1 (slower) as additional parameters. We find that protein synthesis and degradation with faster (**Figure S4B**) and same rates as mRNA degradation and synthesis, also allows for the formation of rare coordinated high states in the case of translation. Only slower protein synthesis and degradation rates did not show rare coordinated high states, likely because for faster protein rates, the system dynamics is determined largely by the transcriptional dynamics. In sum, we demonstrate that the rare coordinated high states can arise in the revised model that includes translation.

Model 1 - Constitutive model

Model 1 is a simple gene regulatory expression model, where mRNA can either be transcribed or degraded and the mRNA of a regulatory gene influences the transcription rate of a regulated gene (**Figure 1B**). Here again, we assume the number of mRNA to be a faithful proxy for the protein number and hence, only model the mRNA expression of a gene. The gene regulation is modeled according to the Hill function (**STAR Methods,** Model 2 - Transcriptional bursting model).

The synthesis and degradation are modeled by chemical reactions. For each gene, we have one chemical species, its mRNA, described by the following two chemical reactions:

$$\varnothing \rightarrow \text{mRNA}$$
$$\text{mRNA} \rightarrow \varnothing,$$

defining the corresponding stoichiometric matrix:

$$(1\ -1).$$

The full model description for one gene regulated by a single gene X is given below:

Chemical reaction	Reaction rate	Reaction propensity
$\varnothing \rightarrow$ mRNA	$r_{prod}+r_{add}*mRNA_X^n/(k^n + mRNA_X^n)$	$r_{prod}+r_{add}*mRNA_X^n/(k^n+mRNA_X^n)$
mRNA $\rightarrow \varnothing$	r_{deg}	$r_{deg} * mRNA$

where r_{prod} the basal mRNA synthesis rate, r_{deg} the mRNA degradation rate, r_{add} the additional synthesis rate due to gene regulation and $mRNA_X$ the mRNA count of gene X at the given time.

The chemical reactions are identical for all N nodes in a given network of size N. The synthesis rate is a node-specific rate (**STAR Methods,** Model 2 - Transcriptional bursting model). We model gene regulation additively (**STAR Methods,** Model 2 - Transcriptional bursting model). For k we tested two different definitions: one closer and one further away from the low expression taking into account the intrinsic stochasticity. We therefore first run a test simulation with a random k for 1,000 time units and determine the standard deviation of the expression of the node denoted as 'node 1'. K is latin hypercube sampled with the rest of the parameters with lower and upper boundary 100 and 1000. We set k to be:

$$k = r_{prod}/r_{deg}+x*std,$$

where std is the standard deviation of the expression of the node denoted as 'node 1' and $x \in \{3,5\}$. We then re-initiate the simulation with the adapted k value.

Parameters

The goal of our study is to model the emergence of rare transient coordinated high expression of several genes. The theoretical idea behind the transcriptional bursting model is that each time the DNA is in an active state, corresponding to a transcriptional burst, the steady-state of the mRNA count is shifted from r_{prod}/r_{deg} to $d*r_{prod}/r_{deg}$. Accordingly, the mRNA attempts to reach its new steady-state which results in a rapid increase in their counts. Depending on the length of the transcriptional burst, which is exponentially distributed with rate parameter r_{off}, the mRNA count is able to reach the new steady-state. We use the dynamical system behavior when modeling the rare coordinated overexpression. In principle, for most transcriptional bursts, the sudden mRNA increase should not initiate a DNA activation of its regulated genes; only in some rare cases, the transcriptional burst in one gene is long enough such that its mRNA count exceeds a certain threshold that may be able to affect the state of another gene locus on DNA. Exceeding of the mRNA threshold can lead to an increased probability of the DNA states of its regulated genes to be activated and hence to an increased mRNA synthesis in the respective genes. The increased mRNA synthesis of regulated genes may lead to positive feedback loops network-wide resulting in the transient coordinated overexpression of genes.

The threshold to be overcome by the mRNA count of a gene to make its gene regulation effective is given by the dissociation constant of the Hill function, k. k determines the 'switching point' from (almost) no gene regulation to (almost) complete gene regulation. Therefore, we define k to be a function of r_{prod}, r_{deg} and d as follows:

$$k(r_{prod},r_{deg},d) = 0.95 * d*r_{prod}/r_{deg},$$

where $d*r_{prod}/r_{deg}$ gives the steady-state mRNA count of the respective regulating gene in the DNA active state. Here, we arbitrarily determine the threshold k to 0.95 of its high-expression steady-state to restrict the emergence of coordinated overexpression to being rare and for the system to demonstrate a significant difference between the low and high gene expression state. The simulations and the analysis are all performed according to the above definition of k. We tested the robustness of this definition for a particular network 5.3 (**Figure S9**) where we performed the same simulations (for 100 latin hypercube sampled parameter

sets (**Supplementary Information** ParSetsAnalysis.xlsx)) as for the final analysis as before using five different definitions of k:

$$k(r_{prod},r_{deg},d) = x * d*r_{prod}/r_{deg},$$

where $x \in \{0.75, 0.8, 0.85, 0.9, 1\}$ (**Supplementary Information** ParSetsAnalysis.xlsx). Our analysis shows that for x = 0.75, none of the 100 simulations show rare coordinated high states: the threshold leading to an effective gene regulation is exceeded too often: the regulated DNA states are activated, the high state emerges and we lose the rareness of the coordinated high gene expression event. The number of simulations showing rare coordinated high states increases with increasing x, reaching its maximum for x = 0.95 (standard, 7 out of the 100 simulations show rare behavior). For x = 1 (high expression steady-state), we also see rare behavior in 7 out of 100 simulations, showing overlapping results in 6 out of the 7 simulations.

Together, we are left with a set of seven parameters consisting of: r_{on}, r_{add}, n, r_{off}, r_{prod}, d, r_{deg}, which may be split into inter-gene (r_{on}, r_{off}, r_{prod}, d, r_{deg}) and intra-gene (r_{add}, n) parameters and the dependent parameter k. Potentially, these parameter sets are node-dependent resulting in a N * 7-dimensional parameter space for a network of size N.

To emphasize the equality between the nodes, we use the same 7-dimensional parameter set for all nodes in a network. Hence, the nodes are relationally and parametrically identical, thereby also allowing us to directly compare the simulations of different network sizes, otherwise not possible, and to determine the effects of network size and architecture on the ability of forming the rare coordinated high state. Therefore, we latin-hypercube sample 1000 parameter sets out of the parameter space with upper and lower boundaries (chosen arbitrarily, but typically spanning two orders of magnitude):

Parameter	Lower boundary	Upper boundary
r_{prod}	0.01	1
r_{deg}	0.001	0.1
r_{on}	0.001	0.1
r_{off}	0.01	0.1
d	2	100
r_{add}	0.1	1
n	0.1	10

by using the MATLAB function *lhsdesign_modified (*Khaled, N. Latin Hypercube (https://de.mathworks.com/matlabcentral/fileexchange/45793-latin-hypercube), MATLAB Central File Exchange. Retrieved May 5, 2018.). The 1000 parameter sets are shown in the Supplementary Information (ParSetsAnalysis.xlsx). For some plots, we used a y-axis break function in MATLAB (Mike, C.F. Break Y Axis (https://www.mathworks.com/matlabcentral/fileexchange/45760-

break-y-axis), MATLAB Central File Exchange. Retrieved December 21, 2018.)

Simulations

We simulated model 2 for a total of 96 networks (for all weakly-connected, non-isomorphic, symmetric networks of sizes 2, 3, 5 and 8 with 2, 4, 10 and 80 networks, respectively)(**Figure S9**), each for 1,000 sampled parameter sets, resulting in a total of 96,000 simulations across four different network sizes. The simulations were performed according to Gillespie's next reaction method and were computed for 1,000,000 time units, which is critical for capturing rare behaviors. For all simulations, the DNA state was initiated (t = 0) to be in its inactive state and the mRNA count was arbitrarily set to 20 for all nodes. The mRNA counts quickly reach their low-expression steady state, such that we are certain that our analysis is not impaired by the given initial conditions. The simulations were implemented in MATLAB R2017a and R2018a. One single simulation of 1,000,000 time units took between 20 minutes and 9 hours depending on the parameter set and the network. The complete simulations took over 1.5 months to run, where we parallelised all 96 networks and and let each of them run on four cores simultaneously.

Simulation classes

We analyzed all of the 96,000 simulations, and assign them to the following four classes, initially by visual inspection, and subsequently by defined criteria (see below):

I - stably low gene expression
II - stably high gene expression
III - uncoordinated transient high gene expression
IV- rare, transient coordinated high gene expression

Therefore we constructed three criteria, for which all the simulations were tested. We primarily focus on the rare, transient coordinated high gene expression states, as defined by the following criteria:

1) **Coordinated high gene expression state.** We call a simulation to show coordinated high expression, if at least once within the 1,000,000 time unit simulation more than half of the mRNA counts are above a specified threshold (e.g. for 5 nodes, at least once three or more mRNA counts have to be above a defined threshold; for 8 nodes, at least once 5 or more mRNA counts have to be above a defined threshold). Similar to the definition of the dissociation constant k, we set the threshold to

$$\text{thres} = 0.8 * d * r_{prod}/r_{deg},$$

where $d * r_{prod}/r_{deg}$ gives the high-expression steady state. Again, we want to detect the rare occurrence of a large mRNA count deviation from the low-steady state and hence, set the threshold arbitrarily to 0.8 (see below for details on the choice of this value).

To compare the simulated results with the experimental data from a drug naive melanoma cell population, we split the 1,000,000 time unit simulations into 1,000 time unit sub-simulations, each accounting for a cell. Hence, we receive simulations of 1,000 cells for

1,000 time units, a procedure justified by the ergodic theory. To show that sub-simulations of 1,000 time units are uncorrelated, we determine the autocorrelations for all 1,000 parameter sets of network 3.2 (**Figure S9**) for up to 1,000 lags (using the MATLAB autocorrelation function *acf* (Price, C. (2011). Autocorrelation function(ACF) (https://www.mathworks.com/matlabcentral/fileexchange/30540-autocorrelation-function-acf), MATLAB Central File Exchange. Retrieved June 13, 2019.). For each of these, we determine the first lag at which the autocorrelation is below the upper 95% confidence bound. For 88.2% of all simulations, the first lag below the upper 95% confidence bound occurs before 1,000 lags. For the 26 simulations with rare coordinated high states, 23 show a first lag below the upper 95% confidence bound before 1,000 lags. For the remaining three simulations the autocorrelation after 1,000 lags is at 0.0615, 0.0206 and 0.4363. Removing the simulation with high autocorrelation (0.4363) does not change the conclusions of our analysis.

2) **Rareness/transience.** To mimic the results given by RNA-FISH in a drug naive melanoma population, where we only see a snapshot of the mRNA counts within a melanoma cell, we randomly determine a time point t_{rand}, where $t_{rand} \in [0,999]$ (uniformly distributed), at which we count the number of mRNA counts above the threshold (for each simulation t varies). We summarize the result of all 1,000 cells in a histogram, for which we expect a decrease with increasing mRNA count above the threshold.

3) **Heavy-tailed gene expression distributions.** At the population level, the single mRNA distributions of marker genes show heavy-tails. We use the same time point t as sampled for criterion 2) and consider the mRNA counts of all genes. If we plot these in gene-dependent histograms, we expect to find right-skewed and unimodal distributions. Here, we use the MATLAB function *skewness(X)* for evaluating the right-skewness of the histogram, where skewness(X) > 0, denotes that the data is spread out more to the right of the mean. Skewness is defined as

$$\text{skewness}(X) = E[(X-\mu)^3/\sigma^3]$$

where μ is the mean of X, σ is the standard deviation of X and E(.) the expectation. For determining unimodality, we test whether the maximum of the last quarter of histogram bins with bin width of one is less than the minimum of the first quarter of histogram bins. Although the definition above only characterizes a heavy-tailed distribution, we find it to be sufficient for our analysis.

Classes I and III, are both defined by criterion 1 only, where criterion 1 is not met in both cases. For class I, none of the genes in a network ever express above the given threshold. For class III, genes express above the given threshold but not once are more than half of the genes above the given threshold at any given time of the simulation. Only if a simulation is able to fulfill all three criteria, will we call it a simulation of class IV - rare transient coordinated high gene expression. If a simulation fulfills criteria 1, but fails to meet both other criteria, we classify it into class II.

To receive numbers of simulations in class IV - rare transient coordinated high expression - per network size, we randomly determine three different t_{rand}, where each $t_{rand} \in [0,999]$ (uniformly distributed) and evaluate all 96000 simulations for being in class IV at the respective snapshot (**Figure 2A**). Note that all these requirements are tested automatically using a script without manual/human intervention.

To show that criterion 3) is sufficient for defining heavy-tailed simulations in class IV in our analysis, we constrain criterion 3) further aiming to identify sub-exponentially decaying, heavy-tailed distributions more directly. We therefore reevaluate all simulations so far identified as class IV and compare their 99th percentiles of their expression distributions with those of fitted exponential distributions (**Figure S3C, right panel**). We expect most of the 99th percentile of the expression distributions to be larger than the 99th percentile of the fitted exponentials. Due to the symmetry of the networks and the resulting similarity between the expression distributions (**Figure S2C**), we only consider node one here, without the loss of generality. To avoid that the fitted exponentials account for the heavy-tails, we constrain the fits to have a maximal bin number (bin size of one) within ∓ 1 of the maximal bin number (bin size one) of the expression distributions. We do so by sequentially increasing/decreasing the exponential parameter μ by steps of 10, sampling 1000 times from the resulting exponential distribution with the MATLAB function *exprnd*(μ,1,1000) and comparing the maximal bin number of the resulting histograms. We repeat the above until the maximal bin number of the exponential distribution is within the predefined range of ∓ 1. As expression distributions with a large maximum bin are more similar to lognormal distributions with small variances and less to exponentials, we restrict the analysis to expression distributions with a maximum bin of ≤ 15 (**Figure S3B**). The threshold of a maximum bin of 15 was determined by considering the simulations and their exponential fits. We additionally discard simulations for which the optimization takes more than 1000 iterations or is producing non-positive parameter values.
Most (82%) of the 99th percentile of the simulated expression distributions are above the diagonal, hence larger than the 99th percentile of the fitted exponential distributions (**Figure S3C, right panel**). The 99th percentile of all the nine marker genes in Shaffer et al. also lie above the diagonal in the general vicinity of the points corresponding to simulations with rare coordinated high states (**Figure S3C, left panel**). We therefore conclude that criterion 3) sufficiently selects for sub-exponentially decaying heavy-tailed distributions.
We additionally, perform parts of the analysis again on two different levels of stricter stringency for criterion of heavy-tailed distributions (**Figure S4F-M**):

A) All simulations fulfilling criteria 1) - 3) which additionally comply to the above mentioned analysis (maximum bin ≤ 15, 99th percentile of expression distribution > 99th percentile of fitted exponential, <1000 iterations to reach a ∓ 1 of the maximal bin number (bin size one) in the optimization for determining the exponential fit and producing non-positive parameter values) (**Figure S4J-M**).
B) All simulations fulfilling criteria 1) - 3) which additionally comply to the above mentioned analysis or have a maximum bin > 15 (**Figure S4F-I**).

The results are qualitatively very similar to the results we receive if we perform the analysis only on criteria 1) - 3) (**Figure 2 and Figure S4**). The 6 and 7 rare coordinated high parameter sets identified by the more stringent analyses A) and B), respectively, are subsets of the original eight rare coordinated high parameter sets (**Figure 2H, FigureS4H** and **S4L**).

Although the resulting optimized decision trees vary slightly, they still identify all three parameters, r_{on}, r_{add} and r_{off}, controlling rare transient coordinated states, as in the original analysis. Together, we conclude that the simple characterization of heavy-tailed distributions is sufficient for further analysis.

The analysis above is a prerequisite for further findings and statements. Due to its importance, we tested its robustness with respect to the definition of the threshold, marking the mRNA count above which a gene is called to be in the high-gene expression state, and with respect to the number of mRNA counts required above the threshold to call it a coordinated high state (both determining criterion 1).
For the test network 5.3, we hence repeated the analysis for thresholds:

$$thres = x * d * r_{prod}/r_{deg},$$

where x = 0.3 : 0.05 : 1 (here, for 100 latin hypercube sampled parameter sets (**Supplementary Information** ParSetsAnalysis.xlsx), and we only test for class IV). Decreasing the threshold down to 0.6 of the high-expression steady state does not change the set of simulations with rare behavior in comparison to the results for x = 0.8. Even a further decrease of the threshold (down to 0.3 of the high-expression steady state) manifests in a similar result: half of the simulations identified previously to show rare behavior are still classified as such. Hence, we keep x = 0.8 for the rest of the analysis (**Supplementary Information** ParSetsAnalysis.xlsx).
Next, for network 5.3 and the 100 parameter sets (**Supplementary Information** ParSetsAnalysis.xlsx), we repeated the analysis requiring at least 1, 2, 4, and 5 mRNA counts to be above the threshold at least once, in order for the simulation to fulfill criterion 1. The lower the required mRNA count, the more simulations fulfill criterion 1 (peaking at a required mRNA count of at least 1 with 11 out of the 100 simulations showing rare behavior according to this definition). The above set of simulations entails the set of simulations fulfilling criterion 1 at the standard required mRNA count of at least 3 (7 out of 100 simulations). Hence, we keep the definition of coordinated overexpression to more than half the nodes being above the threshold.
Additionally, we computed the Gini indices for the gene expression distributions of both the simulations showing rare coordinated high states and the experimental data (**Figure S3A and FigureS3D**) (Jiang et al., 2016; Shaffer et al., 2017). A Gini coefficient of 0 implies perfect equality such that for a given gene, all cells within a population have the same number of mRNA molecules, whereas 1 implies perfect inequality such that one cell expresses all the mRNA molecules while others express none. We used the MATLAB function *gini* (Yvan Lengwiler (2019). Gini coefficient and the Lorentz curve (https://www.mathworks.com/matlabcentral/fileexchange/28080-gini-coefficient-and-the-lorentz-curve), MATLAB Central File Exchange. Retrieved October 24, 2019.) for the computations.

Network topologies

Connectivity

We define a measure for the connectivity of the networks, where

connectivity = number of ingoing edges for any node of the network,

where a self-loop is also considered to be an ingoing edge. As we constrain our analysis to symmetric networks (same number of in-going edges for all nodes in a network per definition), we are able to define one single connectivity per network. The constraints enable us to directly evaluate the impact of the connectivity of the network on the ability to form rare behavior.

Self-loops

A network with a direct auto-regulation is called a network with a self-loop. Due to the restriction of symmetric networks, all networks can be classified as having self-loops for all nodes or not having self-loop for any node. Due to non-isomorphism, the set of networks contains for each network without self-loops an identical network with self-loops. We evaluate the ability of these different edge classes on the formation of rare coordinated high states (**Figure 2F and Figure S5E**).

Characteristic distance

The characteristic distance of a network is defined as the average shortest path length for all pairs of nodes within a given network. To calculate this distance, we used the MATLAB function *shortestpath* on all pairs of nodes. We evaluated the ability of the characteristic distance normalized to the network size on the formation of rare coordinated high states (**Figure 5F**).

Quantitative Analysis

For each of the 96,000 simulations showing rare coordinated high states we performed a quantitative analysis. First, we define a high expression region as a region which is initiated by the first mRNA count to exceed the threshold, terminated by the last mRNA count to drop below the threshold and requires to contain a coordinated high expression state (criterion 1: more than half the mRNA counts have to exceed the defined threshold) between the initiation and termination time points. Breaks of up to 50 time unit intervals are accepted due to the stochastic nature of the simulations. For example, in a 3 node network, where we require at least 2 mRNA counts to exceed the threshold for a coordinated high state: the first mRNA count exceeds the threshold (initiation), then the second mRNA count exceeds the threshold (initiation of high state) but then drops below the threshold for 50 time units before exceeding the threshold again, is still counted as one high-expression region. The length of 50 time units were defined arbitrarily. Due to the stochasticity of the system and the conservative definition of the threshold (located close to the high-expression steady state), we observe these temporary violations of criterion 1. In order to create sensible statistics on the quantitative behavior of the simulations, the temporary relaxation of criterion 1 is necessary.

In the quantitative analysis we extract the total time spent in a high state (out of 1,000,000 time units) from all simulations showing rare behavior (**Figure S3E**).

Decision tree optimization, generalized linear models and constrained simulations

We classify all parameter sets into two classes, rare coordinated high parameter sets and non-rare coordinated high parameter sets, according to the percentage of total simulations per parameter set (96 simulations) in which rare coordinated high states are observed. The threshold above which a parameter set is called a rare coordinated high parameter set is at

20%. More than 19 of the 96 simulations have to show rare behavior in order for a parameter set to be called a rare coordinated high parameter set. The threshold was set according to a summarizing histogram, in which we see a clear distinction between the two groups: the main body of the histogram being located below 20% and the few parameter sets deviating extremely from that main group (> 20%). According to this binary classification, we performed a decision tree optimization (MATLAB function fitctree).
To validate the results of the decision tree optimization, we used generalized linear models on all seven independent parameters r_{on}, r_{add}, n, r_{off}, r_{prod}, d and r_{deg} with the MATLAB function fitglm(X,Y,'Distribution','binomial').
To validate that the parameter region determined by the decision tree optimization favors the formation of simulations with rare coordinated high states, we generate a new set of parameters constrained to values close to the minimal and maximal values of r_{on}, r_{add} and r_{off} for the rare coordinated high parameter sets:

Parameter	Lower boundary	Upper boundary
r_{prod}	0.01	1
r_{deg}	0.001	0.1
r_{on}	0.001	0.025
r_{off}	0.06	0.1
d	2	100
r_{add}	0.15	0.36
n	0.1	10

where altered boundaries are indicated in blue. We latin hypercube sample 1000 parameter sets from that constrained parameter space. For all 1000 parameter sets we simulate 1000000 time units by Gillespie's next reaction method for networks 3.2 and 5.3 (**Figure S9**). Each of these simulations was evaluated for having rare coordinated high states according to the three criteria (**STAR Methods**, section Simulation classes).

Sensitivity Analysis

For each parameter, we tested its sensitivity across its corresponding parameter space (see **STAR METHODS,** section Parameters). Briefly, we take network 3.2 (**Figure 9**) for the detailed analysis as network 3.2 shows rare coordinated high states in all eight rare coordinated high parameter sets. For each of the seven independent parameters (r_{on}, r_{off}, r_{pod}, r_{deg}, n, d, r_{add}), we determine 10 equidistant points across its parameter space, and create new parameter sets by swapping these new parameters one-by-one with ones from the eight rare coordinated high parameter sets, resulting in 8*7*10 = 560 new parameter sets. We simulate 1,000,000 time units with Gillepsie's next reaction method for these newly created parameter sets and evaluate all new simulations for showing rare coordinated high states. For each of the 10 newly sampled parameter values per parameter we receive 8 binary decisions where '1' indicates that the simulation exhibits rare coordinated high states and '0'

that it does not. Our analysis confirmed that the three parameters (r_{on}, r_{off}, and r_{add}) identified by the decision tree algorithm and generalized linear model are indeed critical for producing the rare coordinated high states (**Figure S6D**). We also found a moderate dependence on the Hill coefficient n, also confirmed by the low p-value for n from generalized linear model analysis (**Figure S6C**).

Burst analysis: maintenance of rare coordinated high states

For all simulations showing rare coordinated high states, we determine the fraction and frequency of transcriptional bursts in both the high and baseline time-regions (**Figure 3B-C**). By fraction we mean the percentage of the total time the system is bursting. By frequency we mean the number of bursts per unit time. Additionally, we determine the number of high states and the total time spent in a high state for a network of size three (network 3.2, **Figure 9**) and three independent nodes for each of the parameter sets showing rare coordinated high states in the connected network (**Figure 3D).**

Entry and Exit mechanisms

Entering/Exiting of high expression region - Transcriptional bursts

For all of the simulations in class IV showing rare coordinated high states - we analyze whether the durations of transcriptional bursts are coordinated with the entering and exiting of high time-regions (**Figure 3A**, **STAR Methods**, section Quantitative Analysis).

Entering high expression regions

For each of the defined high expression regions, we determine the entering gene - the gene corresponding to the gene count exceeding the threshold at the initial time point of the high expression region. We then extract all transcriptional bursts which do not start within a high expression region, determine their durations and classify them as either an entering burst or a non-entering burst. An entering burst is the last burst of a particular entering gene before or during its gene count exceeds the threshold. All other bursts are called non-entering bursts. We then perform a two-sample Kolmogorov-Smirnov test on the duration of the entering and non-entering bursts not in high expression regions with the MATLAB function *kstest2* at the significance level 0.05.

Exiting high expression regions

Transcriptional bursts: For each of the determined high gene expression regions we define an exiting region - the region between the first gene in the last quarter of the high expression region permanently leaving the high state (permanently having its gene count below the threshold for the rest of the high expression region) to the last time point of the high expression region. We again determine all transcriptional bursts - within the high expression regions. To exclude potentially prolonged entering bursts, we only consider bursts which start within a high expression region. Also, for bursts exceeding the high expression region, we only account for their durations within the high expression region. If a burst overlaps with an exiting region for at least one time point we call the burst an exiting burst. All other bursts which are not overlapping with an exiting region are called non-exiting bursts. We apply the two-sample Kolmogorov-Smirnov test to the duration of the exiting and non-exiting bursts in high expression regions with the MATLAB function *kstest2* at the significance level 0.05.

Entering/Exiting of high expression region - Times

For all of the simulations showing rare transient coordinated high gene expression, we analyze the distributions of waiting times between genes entering and exiting the high expression region (see Quantitative Analysis).

Entering high expression regions: For all high expression regions, we determine the first time points at which the gene counts exceed the threshold (only for genes with a gene count exceeding the threshold during a particular high expression region at least once). We then consider the waiting times - the time interval between the ascending sorted time points of genes entering the high expression region. These distributions - at most N-1 distributions for a network of size N, one for each waiting time between the genes - are compared to exponential distributions by the Lilliefors test according to the MATLAB function *lillietest(X, 'Distr', 'exp')* at a significance level of 0.05.

Exiting high expression regions: For all high expression regions we determine the last time points at which the gene counts exceed the threshold (again, only for genes with a gene count exceeding the threshold during a particular high expression region at least once). We consider the waiting times and compare their distributions to exponential distributions by the Lilliefors test by applying the MATLAB function *lillietest(X, 'Distr', 'exp')* at a significance level of 0.05.

Comparative Network Inference

Here we describe the computational techniques we used to infer the gene interaction network structure of the pre-drug and post-drug cells. When studying regulatory interactions between genes in a network, it can be useful to abstract the problem into a graph theory framework. Let us assume a set of N genes, with the expression level of each gene represented by the random variable X_i, with $i \in \{1,...,N\}$. The network of interactions between genes can then be represented as a graph of N nodes. An edge $X_i \rightarrow X_j$ signifies a regulatory relationship in which X_i either upregulates or downregulates X_j (Singh et al., 2018).

The computational challenge of network inference is to uncover the true edges of the gene interaction network from statistical relationships between gene expression levels. Many different algorithms, often based on mutual information, conditional probability, or regression analysis, have been developed (Huynh-Thu and Sanguinetti, 2019; Saint-Antoine and Singh, 2019; Singh et al., 2018). The output of an inference algorithm is a matrix of edge weights, which we will call W with dimensions NxN. In this matrix, the element w_{ij} is a measure of how confident we can be that the edge $X_i \rightarrow X_j$ exists in the network. A final network prediction will typically set a threshold for edge weights, and exclude any edges that fall below the threshold. Edges $X_i \rightarrow X_i$, called “self-edges” are typically excluded for the final network prediction, except in cases when temporal data is being analyzed. Since we are using atemporal expression data here, self-edges will be excluded from the analysis.

It is common to judge a network inference algorithm’s reliability by testing it on a “gold standard” dataset, for which the true structure of the network is already known, to see how well it can recover the real edges from the expression data (Huynh-Thu and Sanguinetti, 2019). We have chosen to use the Phixer algorithm (Singh et al., 2018), based on its impressive performance when benchmarked on the DREAM5 Challenge gold standard datasets

(weblink: http://dreamchallenges.org/project/dream-5-network-inference-challenge/; last accessed: 05/06/2019).

Phixer

Phixer computes edge weights using the phi-mixing coefficient. For discrete random variables X and Y taking values in sets A and B, the phi-mixing coefficient φ(X|Y) is defined as:

$$\varphi(X_i|X_j) = \max_{S \subseteq A, T \subseteq B} | Pr\{X_i \in S | X_j \in T\} - Pr\{X_i \in S\} | \quad (1)$$

We then assign $\varphi(X_i|X_j)$ as the weight of the edge $X_j \rightarrow X_i$. The phi-mixing coefficient is an asymmetric measure, so the weight of the edge $X_i \rightarrow X_j$ may be different (Singh et al., 2018). The original Phixer algorithm includes a pruning step, which attempts to correct for false positives by minimizing redundancy in the network. For every possible triplet of nodes X_i, X_j, and X_k, the following inequality is checked:

$$\varphi(X_i|X_k) \leq \min\{\varphi(X_i|X_j), \varphi(X_j|X_k)\} \quad (2)$$

If Equation 2 holds, the edge $X_k \rightarrow X_i$ is eliminated. However, previous work has found that the pruning step, though theoretically sensible, typically reduces accuracy in practice (Saint-Antoine and Singh, 2019), possibly due to the prevalence of redundant connections, such as feed forward loops in gene regulatory networks. So, we removed this part of the algorithm in order to achieve the highest possible level of accuracy.

The Phixer software is available online at the creator's Github page: https://github.com/nitinksingh/phixer/ (last accessed: 05/06/2019). We used the original C code, and kept the default parameter values the same, except for changing "NROW" to 19 and "TSAMPLE" to 4000, to reflect the dimensions of the input data files. The original Phixer code includes, by default, 10 bootstrapping runs, as well as a built-in procedure for binning the raw data, which we did not alter. We removed the pruning step from the code, but otherwise left the edge weight calculation process unchanged.

Data description

The two pre-drug datasets are referred to as NoDrug1 and NoDrug2 in the supplementary data files (**Supplementary Information** PhixerData.xlsx). The datasets containing clusters of resistant cells after four weeks of drug exposure are referred to as Fourweeks1-cluster1, Fourweeks1-cluster2, etc. where we differentiate between Fourweeks1 with four clusters and Fourweeks2 with three clusters. Details of how these datasets were acquired are presented in (Shaffer et al., 2017).

Bootstrapping controls

We found that the Phixer algorithm tends to predict more connections for larger sample sizes, even when the samples are taken from the same dataset. To control for the differences in original sample sizes of various samples, we bootstrapped the original datasets into 4000-sample datasets before performing the Phixer analysis. The number

4000 was chosen arbitrarily; bootstrapped sample sizes of 1000, 2000, and 6000 also appeared to produce similar results.

Randomized controls

For each size-controlled dataset to be analyzed, we created a randomized control consisting of permutations of each gene column from the original dataset (**Supplementary Information** PhixerData.xlsx). We then performed the Phixer analysis on these randomized controls. The resulting edge weight distributions give us a baseline or control edge weight for Phixer that, in principle, reflects potential false positives. We found that in the controls, nearly all of the predicted edge weights were below 0.45 (**Figure S8B**). Therefore, we decided to choose 0.45 as a threshold for the non-control analysis, thus eliminating edges that could have been predicted by chance alone.

Finally, since the analysis contains two stochastic elements (the bootstrapping to correct for the sample size issue and the bootstrapping step in the Phixer algorithm itself) we had to be sure that the observed differences in connectivity were not due to chance. For each dataset, we ran the entire analysis (including both the bootstrapping size correction and the Phixer algorithm) 1000 times, and provide the distributions of the number of edges with weight greater than 0.45 (**Supplementary Information** PhixerData.xlsx).

Asymmetric networks or parameter sets

To test the generality of the results, we generate asymmetric simulations. We introduce asymmetry in both network architectures and the parameter sets.

Asymmetric network

We randomly determine a weakly-connected but asymmetric five-node network (**Figure S2G**). We simulate the network with 100 parameter sets which are latin hypercube sampled out of the same parameter space as the 1000 parameter sets of the main analysis. Out of these 100 simulations, two simulations are classified as showing rare, transient coordinated high gene expression (fulfills all three criteria in **STAR Methods**, section Simulation classes**, Figure S2H-I**).

Asymmetric parameter sets

For the main analysis, we use the same parameter set, consisting of seven independent parameters (**STAR Methods**, section Parameters), for all nodes in a network. We introduce asymmetry by assigning each node in a network a separate set of parameters. Hence, we latin-hypercube sample 100 parameter sets out of a 7 x N parameter space, where N is the number of nodes of the network, with the MATLAB function *lhsdesign_modified*. Due to the high dimensionality, we here confine the parameter space to:

Parameter	Lower boundary	Upper boundary
r_prod	0.01	1
r_deg	0.001	0.1
r_on	0.001	0.1
r_off	0.001	0.1

d	2	100
r_add	0.2	0.4
n	5	10

where the changes in the boundaries are highlighted in blue. We confine the parameter space according to the clustering of rare coordinated high parameter sets. In total, six parameter sets give rise to rare-states more frequently than others for all 96 networks. Only two out of the seven independent parameters, r_{add} and n, show a strong correlation with the rare coordinated high state producing parameter sets as determined by a decision tree optimization. The boundaries in the table above are formed according to these decision tree boundaries in which five out of the six rare coordinated high state producing parameters lie (**Supplementary Information** ParSetsAnalysis.xlsx).
For these 100 parameter sets, we generated simulations for five-node network 5.3 (**Figure S2J**). Out of the resulting 100 simulations, we find two showing rare, transient coordinated high gene expression (fulfills all three criteria in **STAR Methods**, section Simulation classes**, Figure S2K-M**).

Implementation and code

The MATLAB code used for the analysis of this manuscript is available *via* Dropbox *(https://www.dropbox.com/sh/t3793528chgb66x/AAAo7cqaJYkZTQQtAuTwqxGGa?dl=0*).
The analysis was performed with MATLAB R2017a and R2018a.

Supplementary Information

ParSetsAnalysis.xlsx
PhixerData.xlsx

REFERENCES

Antolović, V., Miermont, A., Corrigan, A.M., and Chubb, J.R. (2017). Generation of Single-Cell Transcript Variability by Repression. Curr. Biol. *27*, 1811–1817.e3.

Bartman, C.R., Hsu, S.C., Hsiung, C.C.-S., Raj, A., and Blobel, G.A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol. Cell *62*, 237–247.

Breiman, L., Friedman, J., Stone, C.J., and Olshen, R.A. (1984). Classification and Regression Trees (Wadsworth Statistics/Probability) (Chapman and Hall/CRC).

Chen, H., and Larson, D.R. (2016). What have single-molecule studies taught us about gene expression? Genes Dev. *30*, 1796–1810.

Corrigan, A.M., Tunnacliffe, E., Cannon, D., and Chubb, J.R. (2016). A continuum model of

transcriptional bursting. Elife *5*, e13051.

Fallahi-Sichani, M., Becker, V., Izar, B., Baker, G.J., Lin, J.-R., Boswell, S.A., Shah, P., Rotem, A., Garraway, L.A., and Sorger, P.K. (2017). Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing de-differentiated state. Mol. Syst. Biol. *13*, 905.

Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. *81*, 2340–2361.

Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. Cell *123*, 1025–1036.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633–644.

Ham, L., Brackston, R.D., and Stumpf, M.P.H. (2019). Extrinsic noise and heavy-tailed laws in gene expression.

Ham, L., Schnoerr, D., Brackston, R.D., and Stumpf, M.P.H. (2020). Exactly solvable models of stochastic gene expression.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601.

Huynh-Thu, V.A., and Sanguinetti, G. (2019). Gene Regulatory Network Inference: An Introductory Survey. In Gene Regulatory Networks: Methods and Protocols, G. Sanguinetti, and V.A. Huynh-Thu, eds. (New York, NY: Springer New York), pp. 1–23.

Ibragimov, I. (1962). Some Limit Theorems for Stationary Processes. Theory Probab. Appl. *7*, 349–382.

Iyer-Biswas, S., Hayot, F., and Jayaprakash, C. (2009). Stochasticity of gene products from transcriptional pulsing. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. *79*, 031911.

Jiang, L., Chen, H., Pinello, L., and Yuan, G.-C. (2016). GiniClust: detecting rare cell types from single-cell gene expression data with Gini index. Genome Biol. *17*, 144.

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat. Biotechnol. *33*, 269–276.

Phillips, R., Belliveau, N.M., Chure, G., Garcia, H.G., Razo-Mejia, M., and Scholes, C. (2019). Figure 1 Theory Meets Figure 2 Experiments in the Study of Gene Expression. Annu. Rev. Biophys. *48*, 121–163.

Pisco, A.O., and Huang, S. (2015). Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: “What does not kill me strengthens me.” Br. J. Cancer *112*, 1725–1732.

Pour, M., Pilzer, I., Rosner, R., Smith, Z.D., Meissner, A., and Nachman, I. (2015).

Epigenetic predisposition to reprogramming fates in somatic cells. EMBO Rep. *16*, 370–378.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Rodriguez, J., Ren, G., Day, C.R., Zhao, K., Chow, C.C., and Larson, D.R. (2019). Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity. Cell *176*, 213–226.e18.

Saint-Antoine, M.M., and Singh, A. (2019). Evaluating Pruning Methods in Gene Network Inference. arXiv [q-bio.MN].

Schlauch, D., Glass, K., Hersh, C.P., Silverman, E.K., and Quackenbush, J. (2017). Estimating drivers of cell state transitions using gene regulatory network models. BMC Syst. Biol. *11*, 139.

Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature *546*, 431–435.

Shaffer, S.M., Emert, B.L., Sizemore, A.E., Gupte, R., Torre, E., Bassett, D.S., and Raj, A. (2018). Memory sequencing reveals heritable single cell gene expression programs associated with distinct cellular behaviors. bioRxiv.

Sharma, A., Cao, E.Y., Kumar, V., Zhang, X., Leong, H.S., Wong, A.M.L., Ramakrishnan, N., Hakimullah, M., Teo, H.M.V., Chong, F.T., et al. (2018). Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. Nat. Commun. *9*, 4931.

Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell *141*, 69–80.

Singh, N., Ahsen, M.E., Challapalli, N., Kim, H., White, M.A., and Vidyasagar, M. (2018). Inferring Genome-Wide Interaction Networks Using the Phi-Mixing Coefficient, and Applications to Lung and Breast Cancer. IEEE Transactions on Molecular, Biological and Multi-Scale Communications *4*, 123–139.

So, L.-H., Ghosh, A., Zong, C., Sepúlveda, L.A., Segev, R., and Golding, I. (2011). General properties of transcriptional time series in Escherichia coli. Nat. Genet. *43*, 554–560.

Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature *459*, 428–432.

Su, Y., Wei, W., Robert, L., Xue, M., Tsoi, J., Garcia-Diaz, A., Homet Moreno, B., Kim, J., Ng, R.H., Lee, J.W., et al. (2017). Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. Proc. Natl. Acad. Sci. U. S. A. *114*, 13679–13684.

Symmons, O., and Raj, A. (2016). What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. Mol. Cell *62*, 788–802.

Takahashi, K., and Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. Nat. Rev. Mol. Cell Biol. *17*, 183–193.

Thattai, M., and van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. U. S. A. *98*, 8614–8619.

Torre, E.A., Arai, E., Bayatpour, S., Beck, L.E., Emert, B.L., Shaffer, S.M., Mellis, I.A., Budinich, K.A., Weeraratna, A., Shi, J., et al. (2019). Genetic screening for single-cell variability modulators driving therapy resistance. bioRxiv.

Box 1. Model description, assumptions, parameters, and definitions

Model description: The transcriptional bursting model is comprised of single-gene expression modules described by the telegraph model: the DNA can take on an active and inactive state and transcribe mRNA at high and low rates (transcriptional bursting), respectively. These expression modules are coupled by an underlying network architecture, where regulation is modeled by a Hill function: the regulating gene influences the activation rate r_{on} of the respective regulated gene. The chemical reactions and propensities are described below:

Chemical reaction	Reaction propensity
$I \rightarrow A$	$(r_{on}+r_{add}*mRNA_X^n/(k^n+mRNA_X^n))*I$
$A \rightarrow I$	$r_{off} * A$
$I \rightarrow I + mRNA$	$r_{prod} * I$
$A \rightarrow A + mRNA$	$d*r_{prod} * A$
$mRNA \rightarrow \varnothing$	$r_{deg} * mRNA$

where $I,A \in \{0,1\}$, and $I+A = 1$, where $I = 0$ ($A = 1$) denotes that the DNA is in an active state and $I = 1$ ($A = 0$) denotes that the DNA is in an inactive state. $mRNA_X$ is the mRNA count of gene X at the given time. The model aims to recapitulate rare coordinated high states, where *rare* means that at the population level the expression distributions are unimodal and exhibit heavy tails; *coordinated* means that at least once throughout a simulation more than half the genes (nodes) show mRNA expressions above a specified threshold simultaneously; and *high* means that the mRNA expression of a gene exceeds a specified threshold (thres).

Model assumptions: (1) mRNA is able to influence the gene expression of its regulated gene directly, hence we refer to it as *gene product* throughout this work; (2) all genes are relationally identical (weakly-connected, non-isomorphic and symmetric gene regulatory networks); (3) all genes share the same model parameters; (4) gene regulation is only considered to be activating; and (5) if regulation occurs from several genes, their effects are additive. We discuss and check the generality of our model by testing many of these assumptions on a subset of cases, as described in **Box 2**.

Parameters: The model is described by 8 model parameters, as defined in the table below along with the corresponding ranges.

parameters		**sampling range**
independent model parameters		
r_{on}	The rate at which DNA is activated.	0.001 - 0.1
r_{off}	The rate at which DNA is inactivated.	0.01 - 0.1
r_{prod}	Synthesis rate of gene product.	0.01 - 1
r_{deg}	Degradation rate of gene product.	0.001 - 0.1
r_{add}	Parameter determining the contribution of the additional DNA activation rate upon gene regulation.	0.1 - 1
d	Factor by which the mRNA synthesis rate is increased when in an active DNA state. d >1.	2 - 100
n	Hill coefficient.	0.1 - 10
dependent model parameters		
k*	Dissociation constant of the Hill function, where $k(r_{prod}, r_{deg}, d) = 0.95 * d * r_{prod}/r_{deg}$	-
dependent classification parameters		
thres**	Threshold above which a gene is thought of being highly expressed, where $thres = 0.8 * d * r_{prod}/r_{deg}$	-

Here, r_{prod}/r_{deg} is the steady state in the baseline expression state (when there is no transcriptional burst) and $d * r_{prod}/r_{deg}$ is the steady state in the high expression state (if the DNA would continuously be in the active state).

Model Definitions:

- *weakly-connected network* - a directed network that when replacing the directed edges by undirected ones produces a connected graph in which every pair of nodes is connected by a path.
- *non-isomorphic* - two graphs are called non-isomorphic if there exists no structure-preserving bijection between them.
- *symmetric* - within a graph the number of in- and outgoing edges of a node and across nodes is identical and either all nodes in a network have a self-loop or not.
- *rare coordinated high state* - (1) at least once within a simulation more than half the genes are highly expressed simultaneously, (2) the histogram of simultaneously highly expressed genes at the population level decreases and (3) the gene expression distributions at the population are heavy-tailed.
- *connectivity* - number of ingoing edges for any node of the network.
- *characteristic distance*- the average shortest path length between pairs of nodes of the network.

*The parameter k is dependent on the parameters r_{prod}, r_{deg}, and d, such that: $k = x * d * r_{prod}/r_{deg}$, where $x \in \{0.75, 0.8, 0.85, 0.9, 0.95, 1\}$, which ensures a consistent definition of k throughout the network architectures and parameter sets. Here x represents the fraction of the value corresponding to the steady state value in the high expression state. We showed that for x = 0.75, none of the 100 simulations show rare coordinated gene expression because the threshold resulting in an effective gene regulation is exceeded too often—the regulated DNA states are activated more frequently leading to the high gene expression states and loss of rareness of the coordinated high gene expression event (leading to bimodal distributions). For x > 0.75, there is an increase in the number of simulations showing rare behavior, peaking at x = 0.95. Furthermore, throughout different values of x, the same parameter sets give rise to rare coordinated high states. We take x = 0.95 to maximize the number of simulations positive for the rare coordinated high states.

**We test several values for the threshold above which a gene is highly expressed: $thres = y * d * r_{prod}/r_{deg}$, where $y \in \{0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1\}$. For all $y \geq 0.6$, the set of simulations showing rare coordinated high states largely remains the same. Even for y = 0.3, half of the simulations identified previously to show rare behavior are still classified as such. We chose x = 0.8. Though arbitrarily chosen, the choice of x = 0.8 will not change the conclusions of our analysis.

Box 2.

Relaxing model assumptions

Protein translation: The original transcriptional bursting model does not include a step for translation and is assumed to be captured by the hill function term which not only greatly reduces the computational costs of long stochastic simulations but also allows for analyzing smaller set of parameters. To check if our model can produce rare coordinate high states even when the model includes the translation step, we focused on a particular network (5.3) and associated parameter values that give rise to these states in the original model. We show that for specific rates of translation and protein degradation (STAR Methods), the model including translation exhibits the rare coordinated high states.

Network architectures: By reducing the network architectures to weakly-connected, non-isomorphic and symmetric networks, we systematically reduce the number of possible network architectures. The reduced space of networks is partly supported by experimental observations (Shaffer et al. 2017, 2018), reporting that (1) there is no obvious hierarchical relationship between the expressed genes; and (2) no particular signaling pathway appears to be solely responsible for the observed behavior (see also Figure S1D). Furthermore, these network architectures allows for direct comparisons between network sizes, connectivities and parameter sets (not a given for other topologies). Although the analysis here primarily focuses on the constrained set of network architectures, we show for a subset of cases (STAR Methods) that asymmetric network architectures can also exhibit rare coordinated high gene expression states (Figure S2 G-I), paving the way for a more systematic analysis in the future studies.

Model parameters: While we primarily focus on keeping the same parameter set for each node, we analyzed a subset of networks with asymmetric parameters (STAR Methods) such that each node had distinct underlying parameter sets. We show that a model with asymmetric parameter sets is also capable of producing rare coordinated high gene expression states (Figure S2 J-M).

Multi-gene regulatory effects: The joint regulatory effects experienced by a gene which is regulated by several other genes can be modeled using different approaches. While the majority of analysis here uses an additive model of joint-regulation, we performed a subset of simulations (STAR Methods) for cases where the regulation by multiple gene nodes is multiplicative (Figure S4C and E). We find that for network architecture 5.3, 15 and 97 out of 1000 parameter sets give rise to simulations with rare coordinated high states in the additive and multiplicative joint-regulation, respectively (Figure S4D). Nine simulations are found to show rare coordinated high states in both definitions of multi-gene regulation.

Defining model-output metrics

Population level—sub-simulation size to determine a single cell: To qualitatively compare our results to experimental data, we convert the 1,000,000 time units long single-cell simulation to 1,000 single-cell sub-simulations of length 1,000 time units. We show that the simulations are largely (88.2%) uncorrelated after 1,000 time units, justifying our analysis (STAR Methods).

Heavy-tails: We test different levels of stringency in our definition of heavy-tailed/sub-exponential distributions. The analysis in Figure 2 is performed using the criteria described in STAR Methods, section Simulation classes. We perform further analysis similar to Figure 2 by using more stringent definitions, i.e. fit exponentials and compare the 99th percentiles (Figure S3C). We demonstrate that these results and conclusions are similar to the ones obtained using less stringent criteria (Box 1) shown in Figure 2 (see Figure S4F-M). For example, 6 and 7 out of 8 rare coordinated high parameter sets also appear in the two more stringent analyses (Figure S4H and L). We further validate that our model recapitulates the experimentally observed heavy-tails by comparing the Gini coefficients (Jiang et al. 2016) of experimental and model distributions (Figure S3D).

Number of nodes highly expressed to be called a 'coordinated' state: We define a simulation to show coordinated high gene expression if at least once throughout the simulation more than half of the gene product counts exceed the threshold. Furthermore, we show that for different node counts (2, 3, 4, 5) the number of simulations showing rare coordinated high states does not vary significantly. As an example, for a count of 2, we get 6 out of 100 simulations showing rare behavior; for a count of 3, we get 7. Note that the sets of simulations were overlapping between different scenarios.

Definition of rare coordinated high parameter sets: We define rare coordinated high parameter sets as parameter sets showing rare coordinated high expression in ≥ 20% of all 96 networks. The threshold was defined by inspecting the histogram (Figure 2H), where we see a separation at 20%. Notably, the same rare coordinated high parameter sets also appear in other analysis — they show increased frequencies of simulations with rare coordinated high states when considering the network sizes separately (Figure S6A). Additionally, stricter definitions for heavy-tailed expression distributions result in similar rare coordinated high parameter sets (Figure S4H and L).

Bootstrapping controls in Phixer algorithm: As the number of connections predicted by the Phixer algorithm can depend on the sample size, we bootstrapped the original data set into 4000-sample datasets. The number 4000 was chosen arbitrarily; bootstrapped sample sizes of 1000, 2000, and 6000 also produced qualitatively similar results.

Edge weight in Phixer algorithm: We created a randomized control consisting of permutations of each gene column from the original dataset. We then performed the Phixer analysis on these randomized controls. The resulting edge weight distributions give us a baseline or control edge weight for Phixer that, in principle, reflects potential false positives. We found that in the controls, nearly all of the predicted edge weights were below 0.45 (Figure S8B). Therefore, we decided to choose 0.45 as a threshold for our non-control analysis, thus eliminating edges that could have been predicted by chance alone.

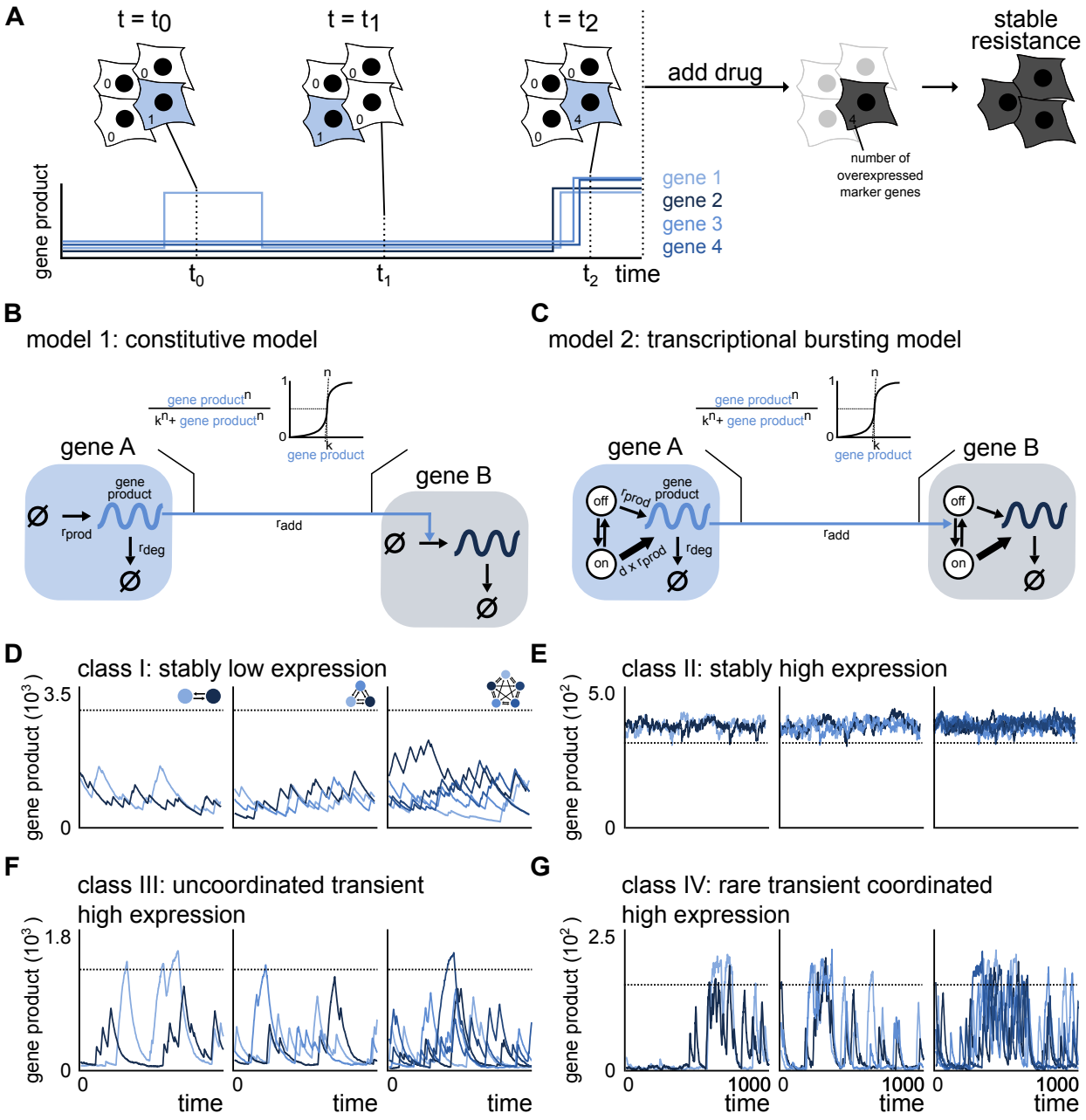

A
t = t0
t = t1
t = t2
add drug
stable resistance
number of overexpressed marker genes
gene product
gene 1
gene 2
gene 3
gene 4
time
B
model 1: constitutive model
gene A
gene B
gene product
C
model 2: transcriptional bursting model
off
on
D
class I: stably low expression
E
class II: stably high expression
F
class III: uncoordinated transient high expression
G
class IV: rare transient coordinated high expression

Figure 1.

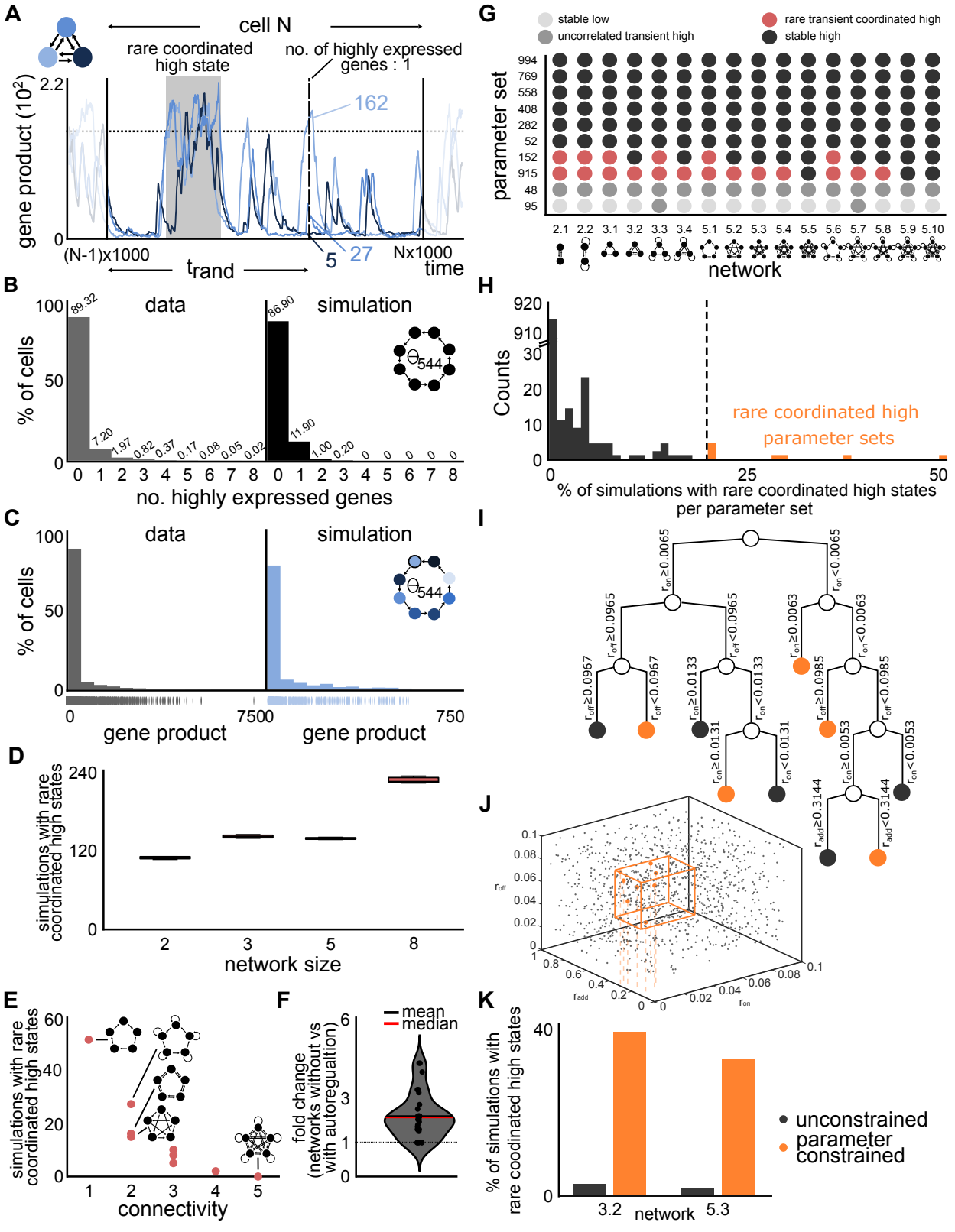
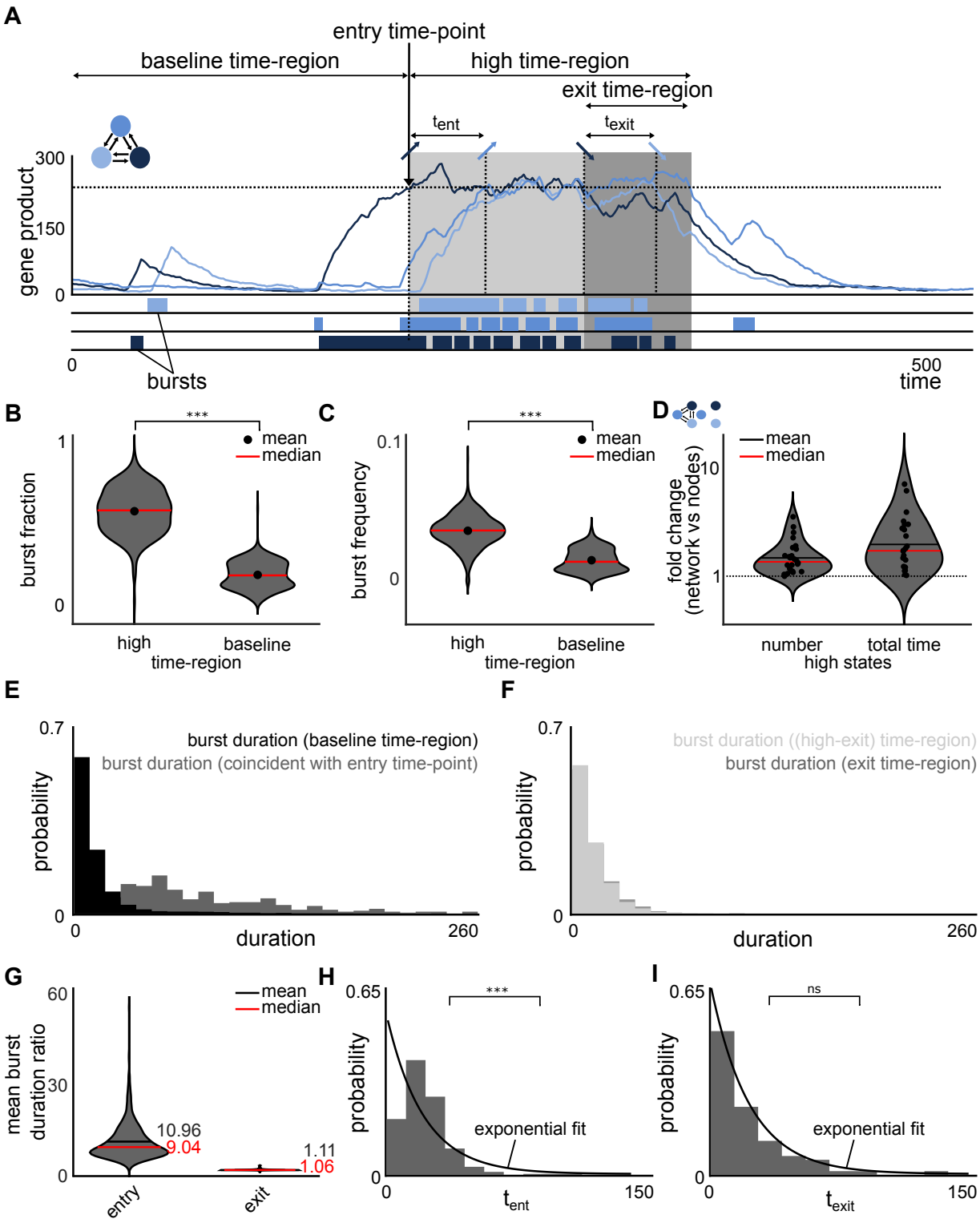

A
cell N
rare coordinated high state
no. of highly expressed genes : 1
gene product (10^2)
2.2
0
162
5
27
(N-1)x1000
Nx1000
time
trand
B
data
simulation
% of cells
100
50
0
89.32
7.20
1.97
0.82
0.37
0.17
0.08
0.05
0.02
86.90
11.90
1.00
0.20
0
Θ544
0 1 2 3 4 5 6 7 8
no. highly expressed genes
C
data
simulation
% of cells
100
50
0
Θ544
0
7500
750
gene product
gene product
D
simulations with rare coordinated high states
240
120
0
2
3
5
8
network size
E
simulations with rare coordinated high states
60
40
20
0
1 2 3 4 5
connectivity
F
fold change (networks without vs with autoregualtion)
6
3
1
0
mean
median
G
stable low
uncorrelated transient high
rare transient coordinated high
stable high
parameter set
994
769
558
408
282
52
152
915
48
95
2.1 2.2 3.1 3.2 3.3 3.4 5.1 5.2 5.3 5.4 5.5 5.6 5.7 5.8 5.9 5.10
network
H
Counts
920
910
30
20
10
0
rare coordinated high parameter sets
0
25
50
% of simulations with rare coordinated high states per parameter set
I
r_on≥0.0065
r_on<0.0065
r_off≥0.0965
r_off<0.0965
r_on≥0.0063
r_on<0.0063
r_off≥0.0967
r_off<0.0967
r_on≥0.0133
r_on<0.0133
r_off≥0.0985
r_off<0.0985
r_on≥0.0131
r_on<0.0131
r_on≥0.0053
r_on<0.0053
r_add≥0.3144
r_add<0.3144
J
r_off
0.1
0.08
0.06
0.04
0.02
0
r_add
r_on
K
% of simulations with rare coodinated high states
40
0
3.2
network
5.3
unconstrained
parameter constrained

Figure 2.


A
entry time-point
baseline time-region
high time-region
exit time-region
t_{ent}
t_{exit}
gene product
300
150
0
bursts
0
500
time
B
burst fraction
1
0
***
mean
median
high
baseline
time-region
C
burst frequency
0.1
0
***
mean
median
high
baseline
time-region
D
fold change (network vs nodes)
10
1
mean
median
number high states
total time high states
E
probability
0.7
0
burst duration (baseline time-region)
burst duration (coincident with entry time-point)
0
duration
260
F
probability
0.7
0
burst duration ((high-exit) time-region)
burst duration (exit time-region)
0
duration
260
G
mean burst duration ratio
60
30
0
mean
median
10.96
9.04
1.11
1.06
entry
exit
H
probability
0.65
0
***
exponential fit
0
t_{ent}
150
I
probability
0.65
0
ns
exponential fit
0
t_{exit}
150

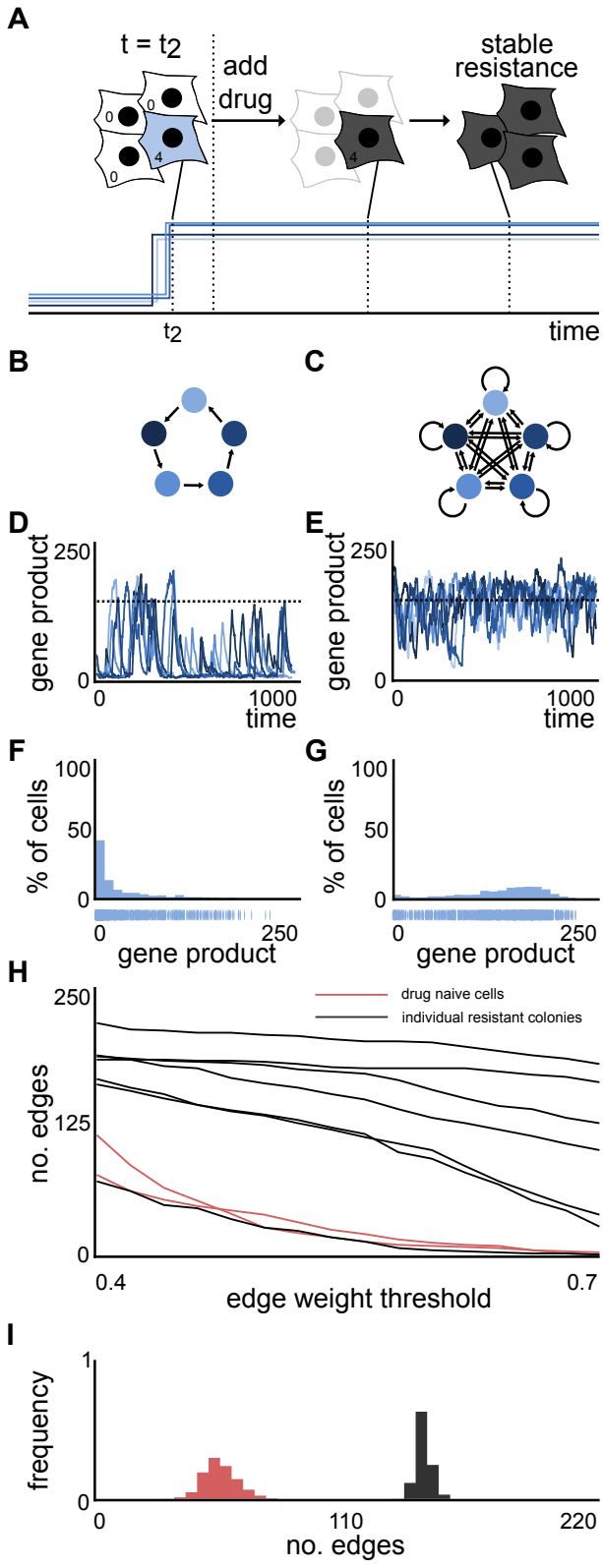

A
t = t2
add
drug
stable
resistance
0
0
4
4
t2
time
B
C
D
gene product
250
0
0
1000
time
E
gene product
250
0
0
1000
time
F
% of cells
100
50
0
0
250
gene product
G
% of cells
100
50
0
0
250
gene product
H
250
125
0
no. edges
drug naive cells
individual resistant colonies
0.4
0.7
edge weight threshold
I
frequency
1
0
0
110
220
no. edges

Figure 4.

Supplementary Information

We decided to develop a network-based framework that models the cell-intrinsic biochemical interactions. One of the first goals we had was to identify the minimal set of biochemical reactions that constitutes this network model. We asked whether a simple network model lacking gene activation step (Model1), i.e. with constitutive mode of gene expression, is sufficient to capture rare coordinated high states (**Figure 1B**; **STAR Methods**, section Model 1)? Or that we need to incorporate gene activation step *via* transcriptional bursting (Model 2) at each node, a phenomenon in which genes flip reversibly between transcriptionally active and inactive state regulated by the binding of a transcription factor(s) (**Figure 1C**; **STAR Methods**, section Model 2)?

In terms of chemical reactions, the critical difference between the two models is that, while in Model 1 the gene is transcribed as a Poisson process with a single rate, r_{prod} (**Figure 1B**), in Model 2, a gene can reversibly switch between active (r_{on}) and inactive state (r_{off}), where binding of the transcription factor at a gene locus defines the effective rate of gene production (**Figure 1C**). Specifically, when inactive, the gene is transcribed as a Poisson process at a basal rate (r_{prod}); when active, this rate becomes higher ($d \times r_{prod}$, where $d > 1$). For both the models, we modeled degradation of the gene product as a Poisson process with degradation rate r_{deg}. For both the models, the inter-node interaction parameter, r_{add}, has a Hill-function-based dependency on the gene product amount (Hill coefficient n) of the respective regulating node to account for the multistep nature of the interaction (**Figure 1B,C**). All chemical reactions, propensities, and model parameters are presented in **STAR Methods**. To test these two models, we used Gillespie's next reaction method (Gillespie 1977) and simulated test cases of small networks (of two or three nodes) for a range of parameters.

For a vast majority of the networks and parameter combinations, Model 1 either produced always low or always high expression states (**Figure S1A**). In some cases, while Model 1 could indeed produce a transition from low to high expression states, the transition happens for all gene products at the same time (**Figure S1A**). However, this model is not consistent with the experimental observations; in particular, if a cell is positive for one marker gene, then it is more likely to be positive for another marker gene, but not necessarily so (**Figure S1B**) (Shaffer et al. 2017). Furthermore, this mode of transition resulted in bimodal distributions of cellular state as determined by the amount of gene product (**Figure S1B**), which is different from the rare nature of the transitions, as reflected by the heavy-tailed distributions of gene products observed in melanoma. Model 2, which incorporates transcriptional bursting-dependent activation of a node (gene), also produced a range of gene expression states (**Figure 1C-1F**). Importantly, this model was able to faithfully capture the qualitative features of the experimental data i.e. rare, transient, and coordinated high expression states (**Figure 1F**). In contrast to Model 1, Model 2 captures another property of the experimental data, i.e. if one gene is in the high expression state, the other genes in the network are likely to be in high expression state, but not always (**Figure 2B** and **S2B**). Based on these initial observations, we decided to pursue Model 2 systematically and simulated networks of different sizes and architectures across a broad range of model parameters.

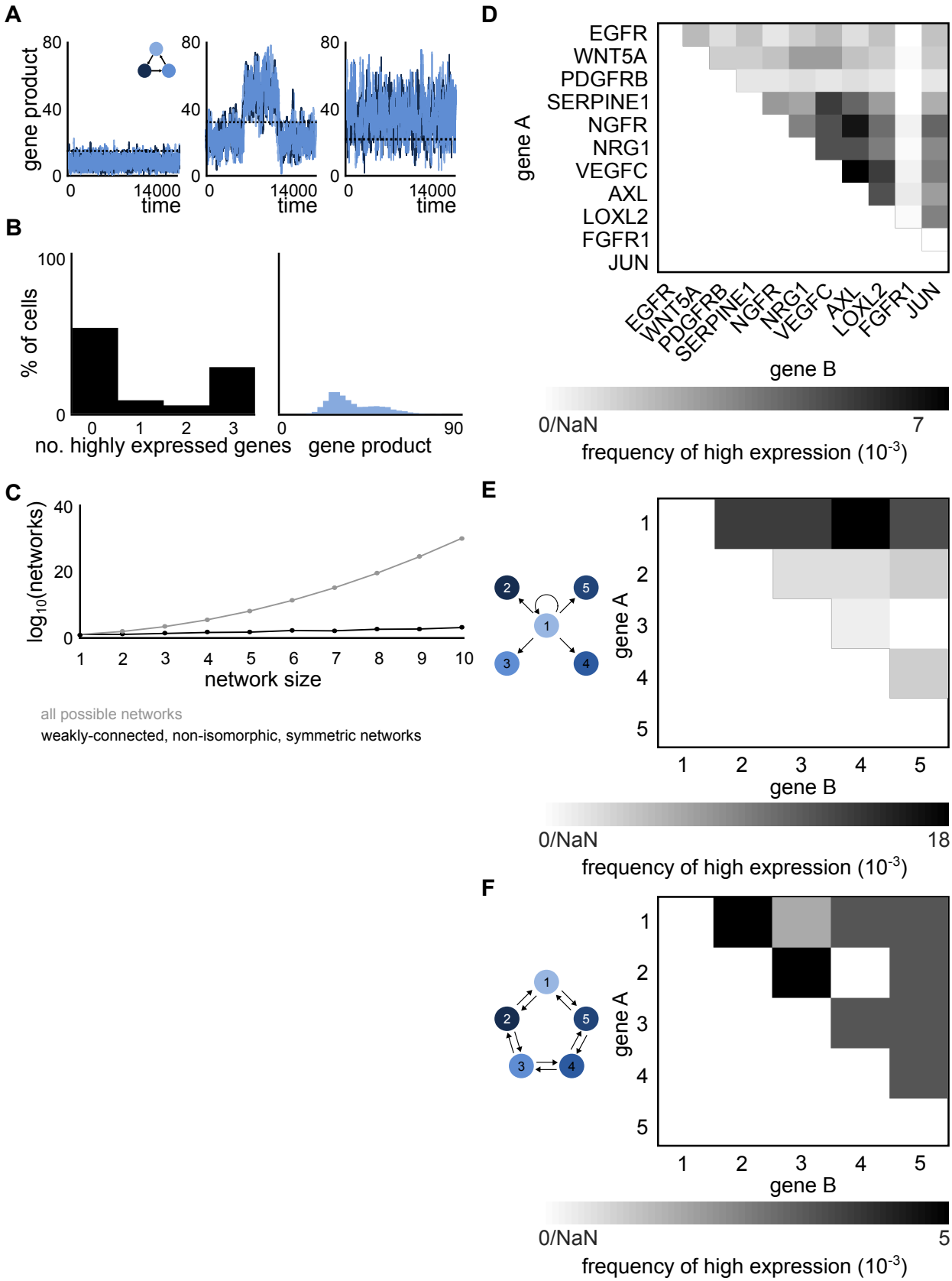

A
gene product
80
40
0
0
14000
time
B
% of cells
100
0
0 1 2 3
no. highly expressed genes
0
90
gene product
C
log10(networks)
40
20
0
1 2 3 4 5 6 7 8 9 10
network size
all possible networks
weakly-connected, non-isomorphic, symmetric networks
D
EGFR
WNT5A
PDGFRB
SERPINE1
NGFR
NRG1
VEGFC
AXL
LOXL2
FGFR1
JUN
gene A
gene B
0/NaN
7
frequency of high expression (10-3)
E
1 2 3 4 5
gene A
gene B
0/NaN
18
frequency of high expression (10-3)
F
1 2 3 4 5
gene A
gene B
0/NaN
5
frequency of high expression (10-3)

Figure S1.

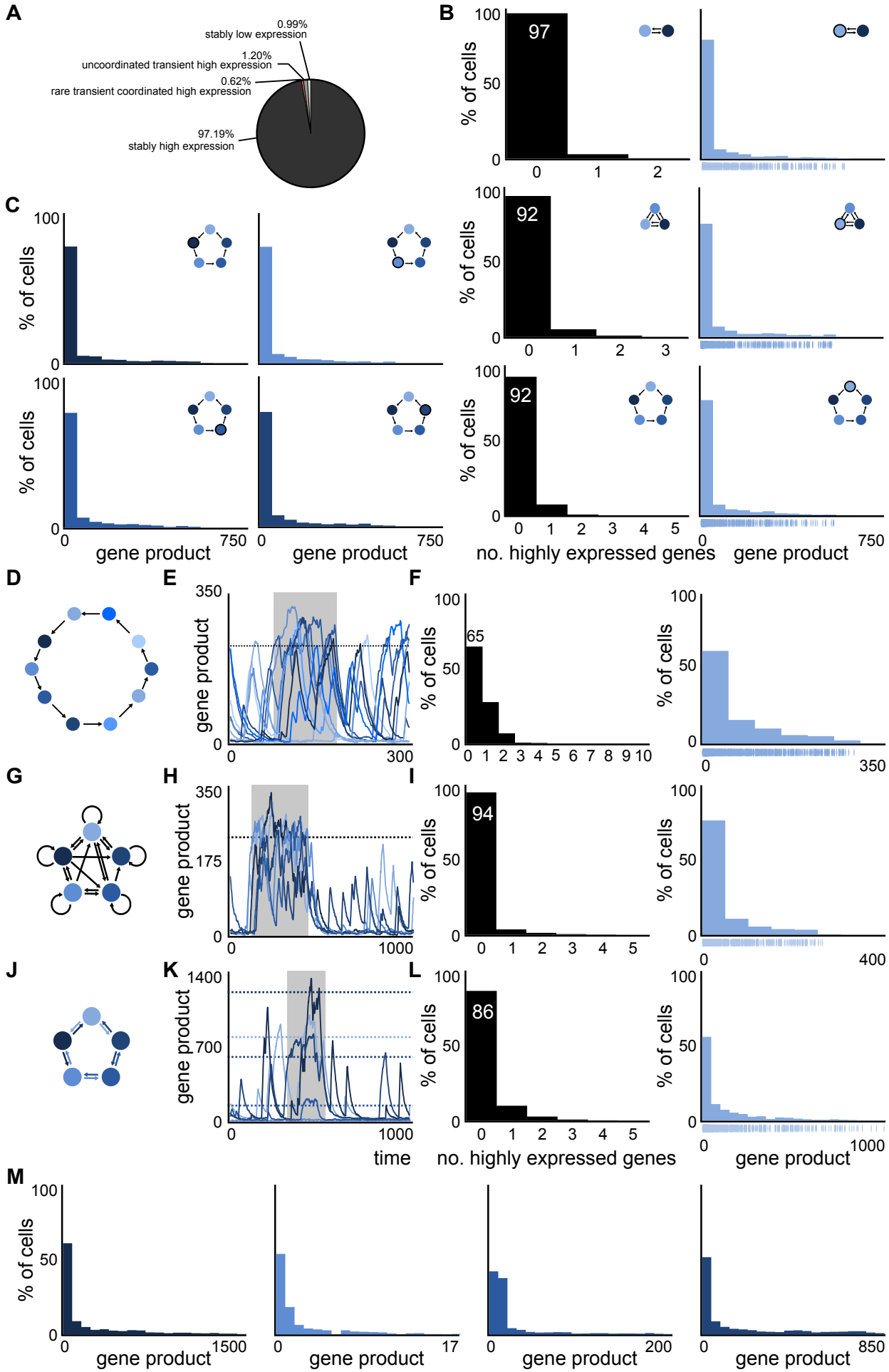

A
0.99%
stably low expression
1.20%
uncoordinated transient high expression
0.62%
rare transient coordinated high expression
97.19%
stably high expression
B
97
92
92
% of cells
no. highly expressed genes
gene product
C
% of cells
gene product
D
E
gene product
F
65
% of cells
G
H
gene product
I
94
% of cells
J
K
gene product
time
L
86
% of cells
no. highly expressed genes
gene product
M
% of cells
gene product

Figure S2.

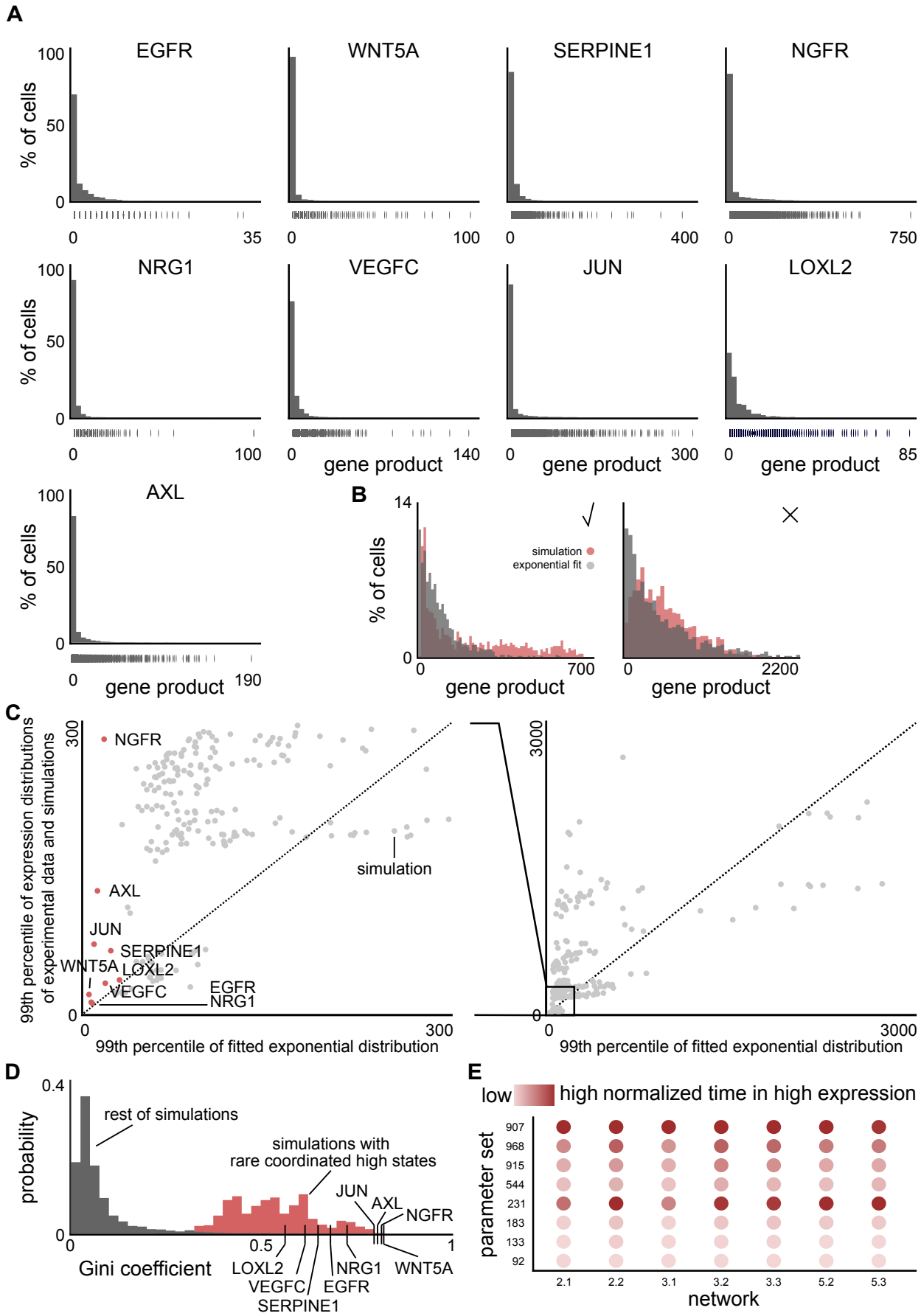

A
EGFR
WNT5A
SERPINE1
NGFR
NRG1
VEGFC
JUN
LOXL2
AXL
% of cells
gene product
B
simulation
exponential fit
C
99th percentile of expression distributions of experimental data and simulations
99th percentile of fitted exponential distribution
simulation
NGFR
AXL
JUN
SERPINE1
WNT5A
LOXL2
VEGFC
EGFR
NRG1
D
probability
rest of simulations
simulations with rare coordinated high states
Gini coefficient
E
low
high normalized time in high expression
parameter set
network

Figure S3.

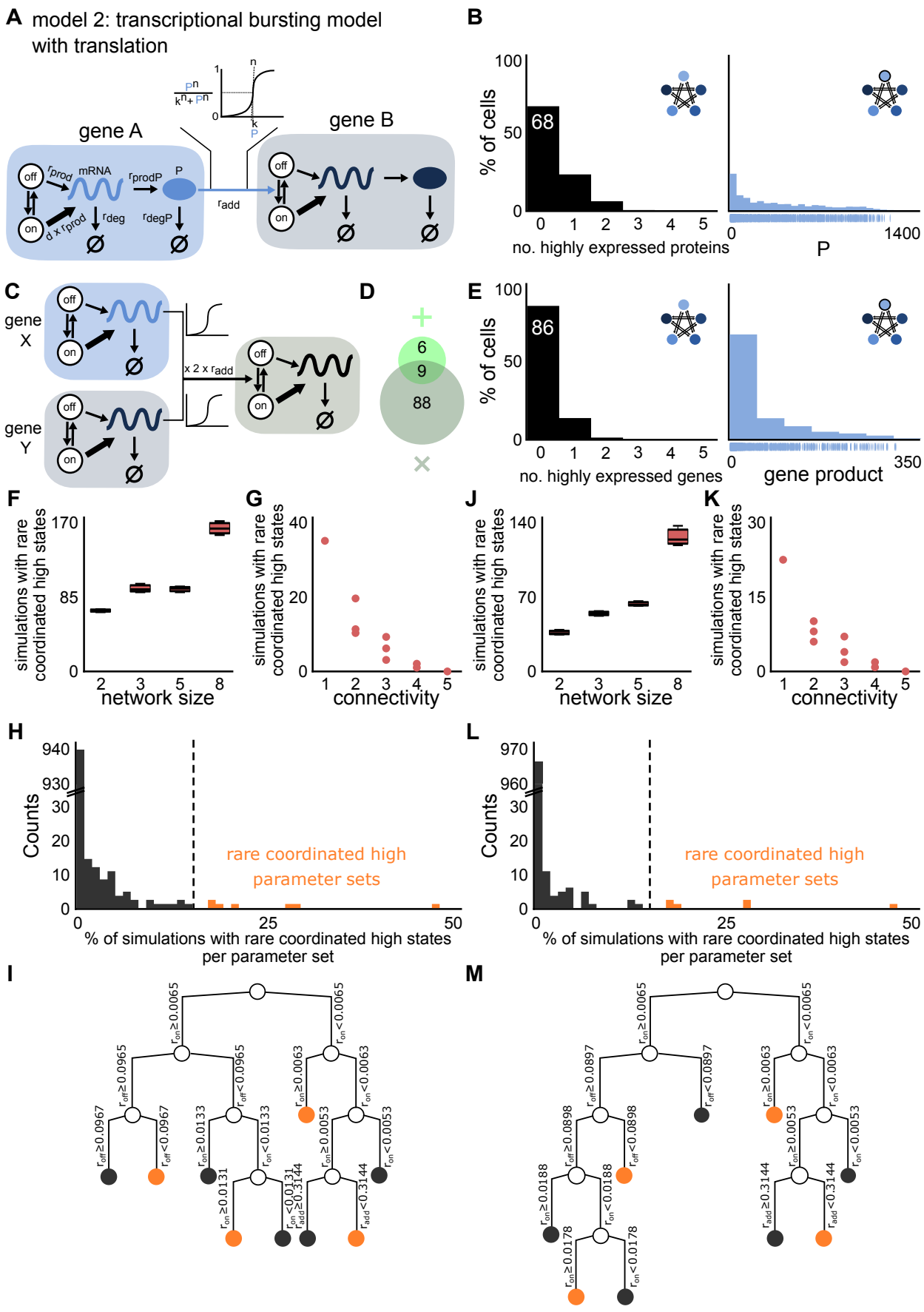

A
model 2: transcriptional bursting model with translation
gene A
gene B
off
on
mRNA
P
no. highly expressed proteins
% of cells
68
B
no. highly expressed genes
gene product
86
E
C
gene X
gene Y
x 2 x r_add
D
6
9
88
F
G
J
K
simulations with rare coordinated high states
network size
connectivity
H
L
Counts
rare coordinated high parameter sets
% of simulations with rare coordinated high states per parameter set
I
M

Figure S4.

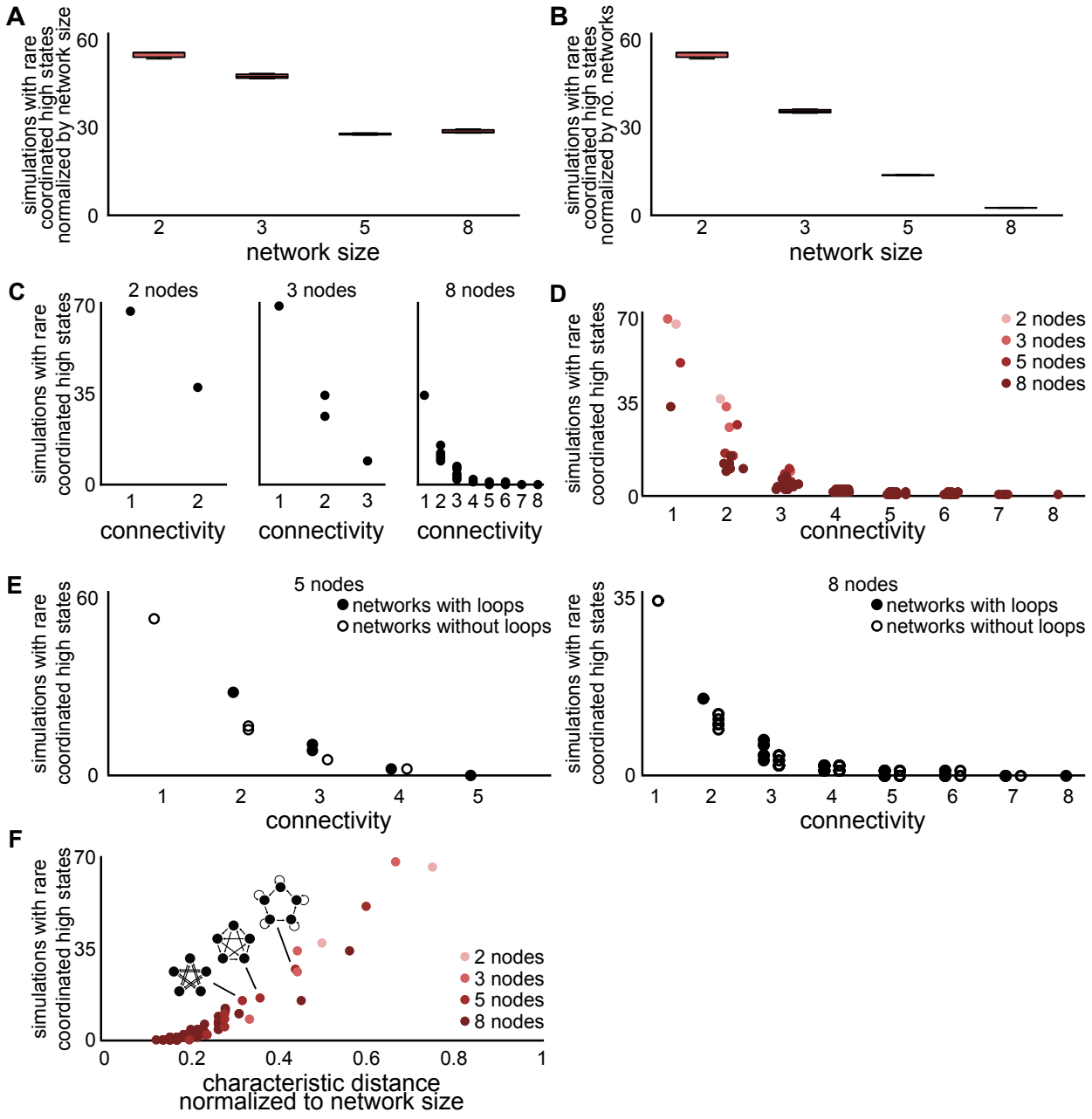

A
simulations with rare coordinated high states normalized by network size
60
30
0
2
3
5
8
network size
B
simulations with rare coordinated high states normalized by no. networks
60
30
0
2
3
5
8
network size
C
2 nodes
3 nodes
8 nodes
simulations with rare coordinated high states
70
35
0
1 2
1 2 3
1 2 3 4 5 6 7 8
connectivity
D
simulations with rare coordinated high states
70
35
0
2 nodes
3 nodes
5 nodes
8 nodes
1 2 3 4 5 6 7 8
connectivity
E
5 nodes
networks with loops
networks without loops
simulations with rare coordinated high states
60
0
1 2 3 4 5
connectivity
8 nodes
networks with loops
networks without loops
simulations with rare coordinated high states
35
0
1 2 3 4 5 6 7 8
connectivity
F
simulations with rare coordinated high states
70
35
0
2 nodes
3 nodes
5 nodes
8 nodes
0 0.2 0.4 0.6 0.8 1
characteristic distance
normalized to network size

Figure S5.

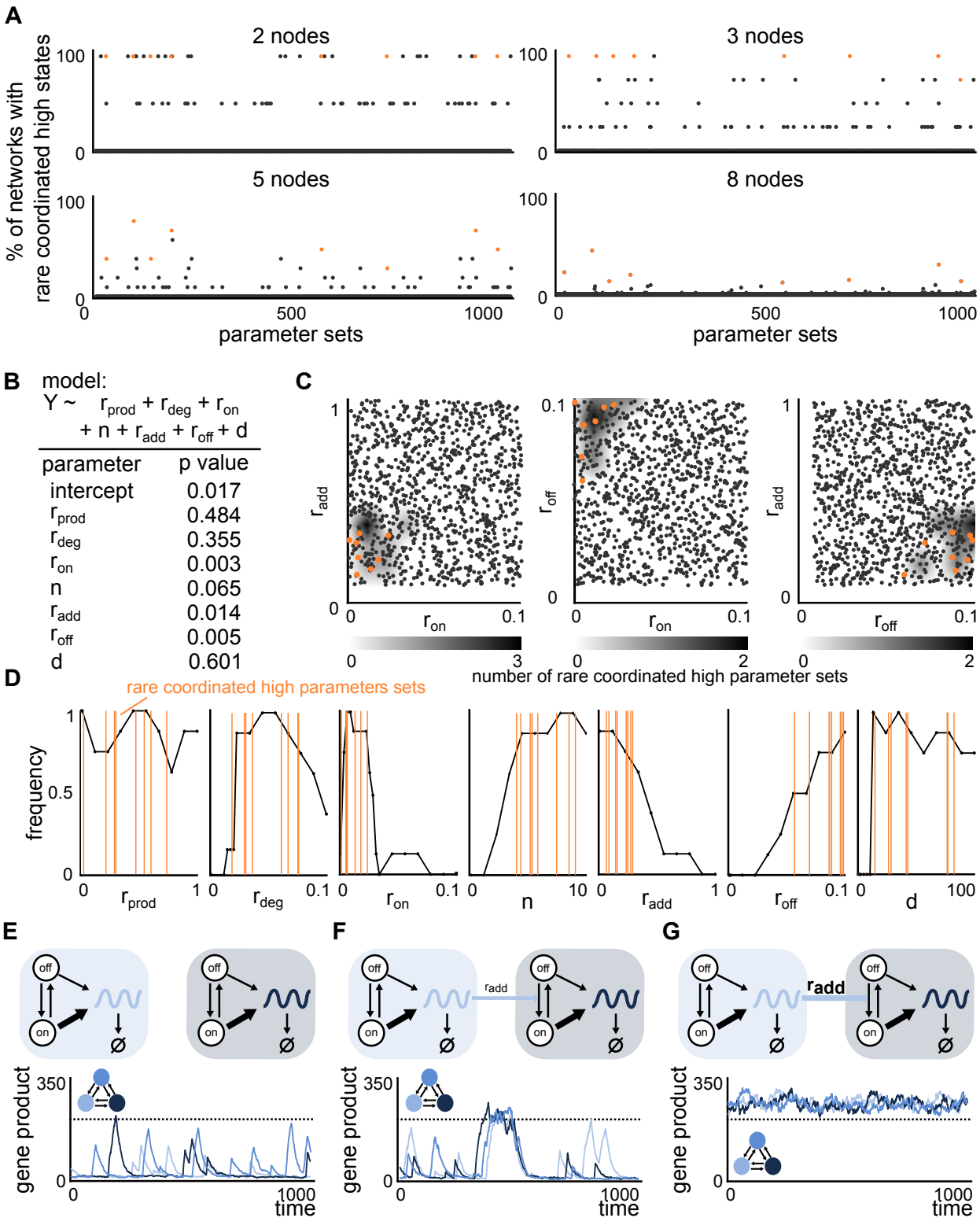

A
2 nodes
3 nodes
5 nodes
8 nodes
% of networks with rare coordinated high states
parameter sets
B
model:
Y ~ r_prod + r_deg + r_on + n + r_add + r_off + d
parameter | p value
intercept | 0.017
r_prod | 0.484
r_deg | 0.355
r_on | 0.003
n | 0.065
r_add | 0.014
r_off | 0.005
d | 0.601
C
number of rare coordinated high parameter sets
D
rare coordinated high parameters sets
frequency
E
F
G
gene product
time

Figure S6.

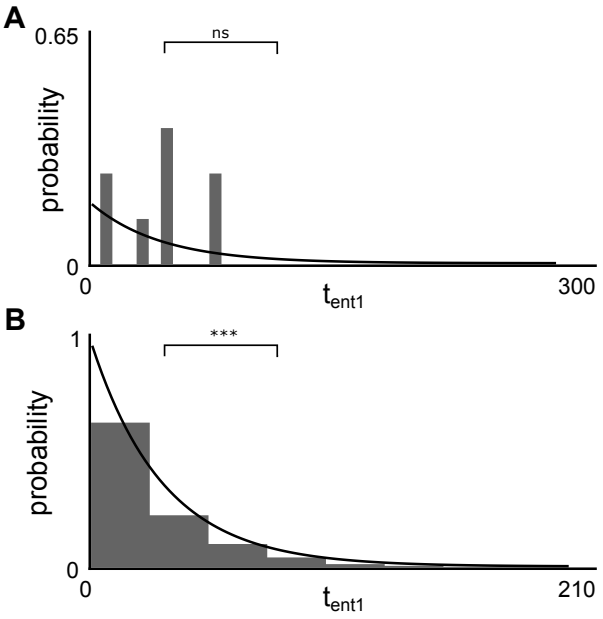

A
0.65
ns
probability
0
0
300
t_{ent1}
B
1
***
probability
0
0
210
t_{ent1}

Figure S7.

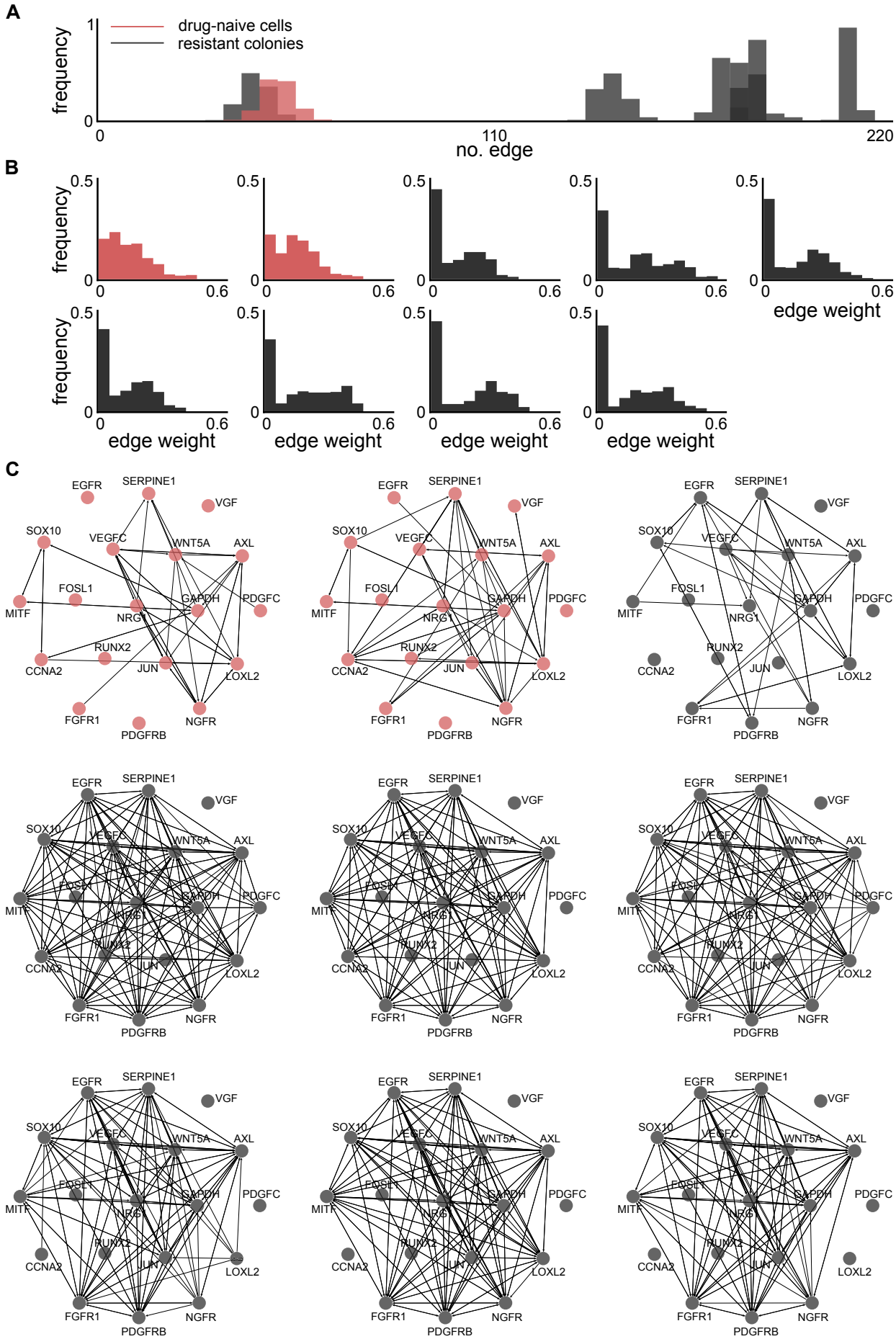

A
drug-naive cells
resistant colonies
frequency
no. edge
B
frequency
edge weight
C
EGFR
SERPINE1
VGF
SOX10
VEGFC
WNT5A
AXL
FOSL1
GAPDH
PDGFC
MITF
NRG1
RUNX2
CCNA2
JUN
LOXL2
FGFR1
NGFR
PDGFRB

Figure S8.

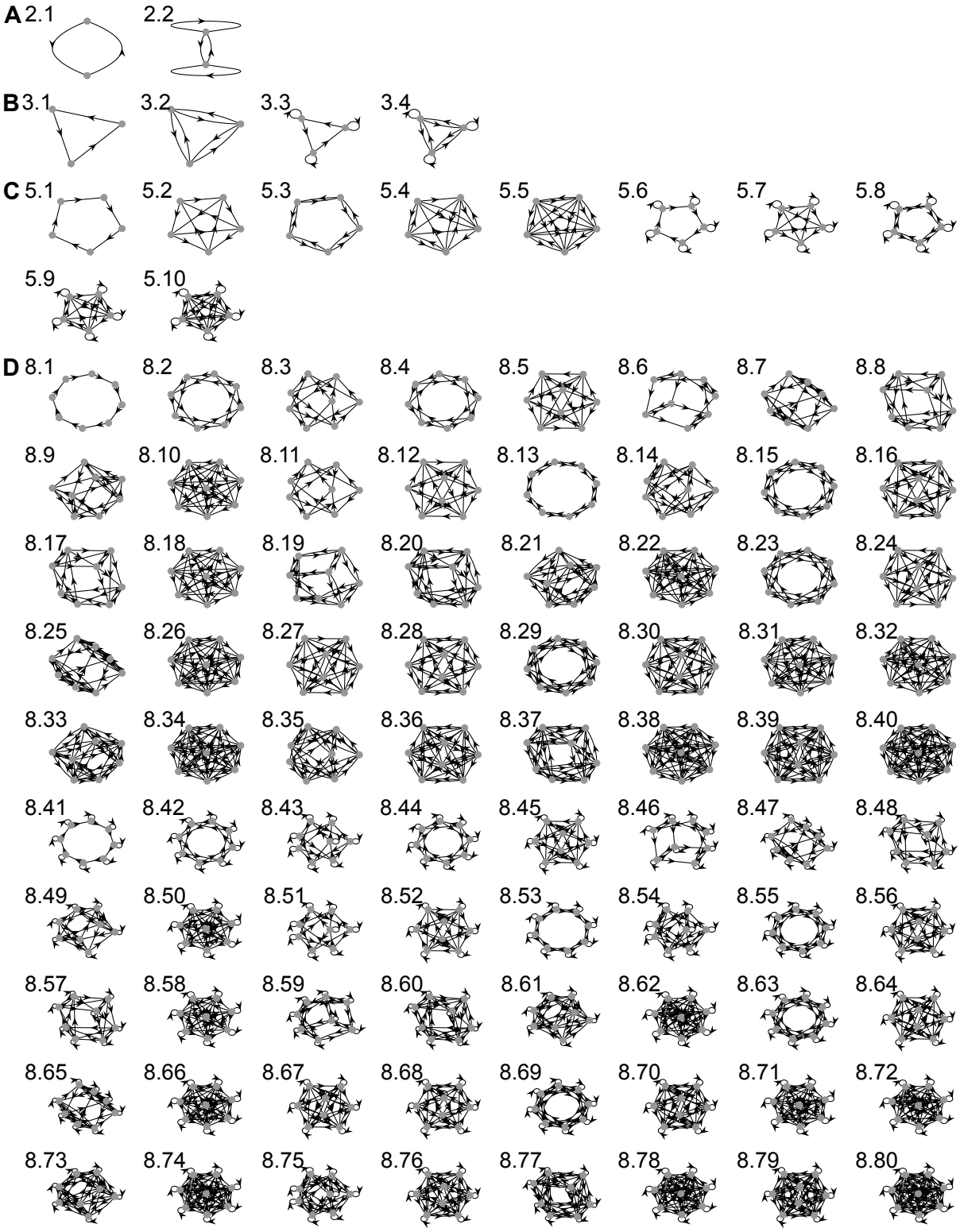

A
2.1
2.2
B
3.1
3.2
3.3
3.4
C
5.1
5.2
5.3
5.4
5.5
5.6
5.7
5.8
5.9
5.10
D
8.1
8.2
8.3
8.4
8.5
8.6
8.7
8.8
8.9
8.10
8.11
8.12
8.13
8.14
8.15
8.16
8.17
8.18
8.19
8.20
8.21
8.22
8.23
8.24
8.25
8.26
8.27
8.28
8.29
8.30
8.31
8.32
8.33
8.34
8.35
8.36
8.37
8.38
8.39
8.40
8.41
8.42
8.43
8.44
8.45
8.46
8.47
8.48
8.49
8.50
8.51
8.52
8.53
8.54
8.55
8.56
8.57
8.58
8.59
8.60
8.61
8.62
8.63
8.64
8.65
8.66
8.67
8.68
8.69
8.70
8.71
8.72
8.73
8.74
8.75
8.76
8.77
8.78
8.79
8.80

Figure S9.

Figure S1. Related to Figure 1 and STAR Methods.
(A) Depending on the network architecture and the parameters of the gene expression model, we observe either stably low expression (left), stably high expression (right) or transient coordinated high expression (middle).
(B) The distributions of simultaneously overexpressed genes and the gene products at the population level (B middle) show bimodal distributions and are inconsistent with the observations in drug naive melanoma cells.
(C) The subset of weakly-connected, non-isomorphic, symmetric networks decreases the testable architecture space by many orders of magnitude.
(D) No clear driver gene or hierarchy is apparent from the two-dimensional RNA FISH experimental data.
(E) Star-shaped networks (left) may lead to hierarchies within the joint frequencies of genes exhibiting the high expression state (right).
(F) Symmetric networks (left) do not form hierarchical structures in the joint frequencies of high expression (right).

Figure S2. Related to Figure 2 and STAR Methods. Simulations of varying network sizes and asymmetries are able to recapitulate the number of simultaneously highly expressed genes and expression distribution as seen in drug naive melanoma.
(A) 0.62% of simulations show rare transient coordinated high expression.
(B) The simulated distributions of simultaneously highly expressed genes and expression are qualitatively similar to data from a pre-resistant melanoma population ubiquitously in networks with different numbers of nodes. Shown for a two node (top), three node (middle) and eight node network (bottom).
(C) The gene expression distributions of all five nodes (the gene expression distribution of node one is shown in (B)) are qualitatively similar.
(D-F) Network of size 10 (A) with corresponding simulation (B) and distributions of simultaneously overexpressed genes and gene expression (C). The distributions show qualitatively the same behavior as drug naive melanoma cells.
(G-I) Asymmetric network architecture (D) with corresponding simulation (E) and distributions of simultaneously overexpressed genes and gene expression (F). The distributions show qualitatively the same behavior as drug naive melanoma cells.
(J-L) Symmetric network architecture and an asymmetric parameter set (G) with corresponding simulation (H) and distributions of simultaneously overexpressed genes and gene expression (I) . The distributions show qualitatively the same behavior as drug naive melanoma cells.
(M) The gene expression distributions of all five nodes (the gene expression distribution of node one is shown in (I)) generated with an asymmetric parameter set display different levels of heavy-tails.

Figure S3. Related to Figure 2 and STAR Methods. Most simulations with rare coordinated high states show heavy-tails in their gene expression distributions.
(A) Expression distributions determined by single cell RNA-FISH of nine identified marker genes (data from Schaffer et al., 2017) show heavy-tails.

(B) Simulated gene expression distributions deviating too much from exponential distributions (left panel) are discarded in the analysis shown in (C).
(C) The tails of the simulated distributions for gene expressions are fatter than of fitted exponential distributions (right panel) (see STAR Methods). The same is true for the experimentally observed expression distributions (left panel).
(D) The Gini indices of simulations of rare coordinated high states are substantially higher than of simulations not showing rare coordinated high states. The experimentally measured expression distributions have similar Gini indices than simulations with rare coordinated high states.
(E) Simulations of particular parameter sets across different network architectures and sizes show similar (normalized) time in high expression relative to other parameter sets.

Figure S4. Related to Figure 2 and STAR Methods. Transcriptional bursting model with fast protein synthesis and degradation or multiplicative gene regulation shows simulations with rare coordinated high states.
(A) Schematic of the transcriptional bursting model with translation for two nodes. DNA is either in an inactive (off) or active (on) state. Transitions take place with rates r_{on} and r_{off}, where mRNA is synthesized with rates r_{pod} and $d*r_{prod}$, respectively, d>1. mRNA degrades with rate r_{deg}. Protein is synthesized with rate r_{prodP} and degraded with r_{degP}. Gene regulation is modeled by a Hill function, where the protein count of the regulating gene A increases the activation of the regulated gene B.
(B) Fast translation events where protein synthesis and degradation is ten times faster than mRNA synthesis and degradation leads to simulations with rare coordinated high states. The simulated distributions of simultaneously highly expressed proteins and protein expression qualitatively capture features of experimental data from a pre-resistant melanoma population. The networks for simulation are indicated in the top right corner.
(C) Schematic of multiplicative gene regulation. Gene regulation on the gene activation of the regulated gene is the product of the Hill functions of regulating genes X and Y, rate r_{add} and a factor (the number of regulating genes, see STAR Methods).
product of Hill functions for gene X and Y
(D) Multiplicative gene regulation leads to more simulations showing rare coordinated high states than additive gene regulation.
(E) The simulated distributions of simultaneously highly expressed genes and expression are qualitatively similar to data from a pre-resistant melanoma population. The networks for simulation are indicated in the top right corner.
(F-M) Two levels of stringencies for the definition of heavy-tailed distributions show qualitatively similar results (F-I and J-M) to each other and to the stringency defined in main text (Figure 2).

Figure S5. Related to Figure 2 and STAR Methods. Networks with higher connectivity or without autoregulation show less simulations with rare coordinated high states.
(A-B) Number of simulations with rare coordinated high states normalized by network size (A) and number of networks within each network size (B).

(C) Increasing connectivity within all networks of sizes two (left), three (middle) and eight (right) leads to a decrease in the number of simulations with rare coordinated high states.
(D) All network sizes show the same trend of inverse relation between connectivity and number of simulations with rare coordinated high states.
(E) For the same connectivity, networks with autoregulation show a higher number of simulations with rare coordinated high states.
(F) With increasing characteristic distance (normalized to network size) more simulations show rare coordinated high states.

Figure S6. Related to Figure 2 and STAR Methods. Three out of seven parameters regulate the formation of rare coordinated high states.
(A) The rare coordinated high parameter sets (orange) give rise to rare coordinated high states more frequently than others in any given network of sizes two, three, five, and eight (from top left to bottom right).
(B) Analysis of the parameter sets by the generalized linear model where the model specification, parameters, and the respective p values are shown. Parameters with p-value less than 0.05 are considered significant.
(C) Phase space overlaid with all tested 1000 parameter sets for r_{on} - r_{add} , r_{on} - r_{off} and r_{off} - r_{add} show that the rare coordinated parameters are narrowly constrained in the respective 2D spaces (orange).
(D) The sensitivity analysis reflects the findings of the decision tree optimization and generalized linear model. Three parameters, r_{on}, r_{off} and r_{add} are more sensitive to changes.
(E-G) Increasing parameter radd leads to more stable high expression shown for radd = 0 (E), radd = 0.29 (F) and radd = 100,000 (G).

Figure S7. Related to Figure 3 and STAR Methods. Counterexamples of
(A) Representative plot of distribution that satisfies the Lilliefors test corresponding to Figure 3H.
(B) Representative plot of distribution that rejects Lilliefors test corresponding to Figure 3I.

Figure S8. Related to Figure 4 and STAR Methods. With increasing connectivity, simulations are more likely to enter the coordinated high state but are no longer able to leave it.
(A) The number of edges in the inferred gene regulatory networks are higher in 6/7 resistant colonies than in the two biological replicates of drug naive cells.
(B) For randomized controls, the edge weight is below 0.45, shown for all biological replicates.
(C) The resistant colonies (gray) have more edges in their respective inferred gene regulatory networks than drug naive melanoma cells (red).

Figure S9. Related to Star Methods. All architectures of sizes two (A), three (B), five (C) and eight (D).

ParSetsAnalysis.xlsx

PhixerData.xlsx

Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer

Lea Schuh[1,2,3], Michael Saint-Antoine[4], Eric Sanford[1], Benjamin L. Emert[1], Abhyudai Singh[4], Carsten Marr[2], Yogesh Goyal[1,*], Arjun Raj[1,*,**]

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA
[2]Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg, 85764, Germany
[3]Department of Mathematics, Technical University of Munich, Garching, 85748, Germany
[4]Electrical and Computer Engineering, University of Delaware, Newark, Delaware 19716, USA
*Authors for correspondence: yogesh.goyal0308@gmail.com, arjunrajlab@gmail.com
**Lead contact: arjunrajlab@gmail.com

SUMMARY

Non-genetic transcriptional variability at the single-cell level is a potential mechanism for therapy resistance in melanoma. Specifically, rare subpopulations of drug naive melanoma cells occupy a transient pre-resistant state characterized by coordinated high expression of several genes. Importantly, these rare cells are able to survive drug treatment and develop resistance. How might these extremely rare states arise and disappear within the population? It is unclear whether the canonical stochastic models of probabilistic transcriptional pulsing can explain this behavior, or if it requires special, hitherto unidentified molecular mechanisms. Here we use mathematical modeling to show that a minimal model comprising of transcriptional bursting and interactions between genes can give rise to rare coordinated high expression states. We next show that although these states occur across networks of different sizes, they occur more frequently in networks with low connectivity and depend strongly on three (of seven) independent model parameters. Interestingly, we find that while entry into the rare coordinated high state is initiated by a long transcriptional burst that also triggers entry of other genes, the exit from it occurs through the independent inactivation of individual genes. Finally, the transcriptional bursting model predicts that increased network connectivity can lead to transcriptionally stable states, which we verify using experimental data. In sum, we demonstrate that established principles of gene regulation are sufficient to describe the observed cell expression variability and argue for its general existence in other biological contexts.

Keywords: stochasticity, network, gene expression, melanoma, drug resistance, non-genetic

INTRODUCTION

Replace with "Cellular heterogeneity has"

~~Rare and large heterogeneity in single cells have~~ been reported to arise from non-genetic transcriptional variability, even in clonal, genetically homogeneous cells grown in identical conditions (Fallahi-Sichani et al., 2017; Gupta et al., 2011; Pisco and Huang, 2015; Shaffer et al., 2017; Sharma et al., 2018, 2010; Spencer et al., 2009; Su et al., 2017). Importantly, cells exhibiting these non-genetic deviations are resistant to anti-cancer drugs (e.g., Ras pathway inhibitors) and may lead to relapse in patients. For example, in a drug naive melanoma population, a small fraction (~1 in 3000) of cells are pre-resistant, meaning they

are able to survive targeted drug therapy, resulting in their uncontrolled cellular proliferation (Shaffer et al., 2017). These rare pre-resistant cells are marked by transient and coordinated high expression of dozens of marker genes. In other words, several genes are highly expressed simultaneously in a rare subset of cells, while the rest of the population have low or zero counts of mRNAs for these genes, resulting in a distribution of steady state mRNA counts per cell that peaks at or close to zero and has heavy tails. The rare cells in the tails, which transiently arise and disappear in the population by switching their gene expression state (**Figure 1A**), are much more likely to develop resistance to targeted therapies. Importantly, the rare and coordinated large fluctuations in the expression of multiple genes persist for several generations. Classical probabilistic models of gene expression have predicted the possibility of various types of mRNA expression distributions across a population, including normal, log-normal, gamma, or heavy-tail distributions (Antolović et al., 2017; Chen and Larson, 2016; Corrigan et al., 2016; Golding et al., 2005; Ham et al., 2019, 2020; Iyer-Biswas et al., 2009; Raj and van Oudenaarden, 2008; Raj et al., 2006; So et al., 2011; Symmons and Raj, 2016; Thattai and van Oudenaarden, 2001). It is unclear if such models can recapitulate the non-genetic variability characterized by rare and transient high expression states for several genes simultaneously (from now on referred to as "rare coordinated high states"), and if so, under what conditions.

Insert paragraph break

Might a stochastic system of interacting genes inside the cell facilitate transition in and out of the rare coordinated high state? One hypothesis is that ~~within the canonical modeling framework,~~ only a rare set of unique (and perhaps complex) networks can facilitate reversible transitions into the rare coordinated high states. Alternatively, relatively generic gene regulatory networks may be capable of producing such behaviors, suggesting that a large ensemble of such networks may admit rare-cell formation. Both of these scenarios have different implications—for instance, the latter hypothesis suggests that this behavior could be more common in biological systems than hitherto appreciated. The alternatives described above can also be posed in terms of the nature of model parameters—whether the set of values that give rise to rare coordinated high states are constrained to lie within a narrow window of parameter space or whether such behavior may occur across broad swaths of parameter space. Yet another possibility is that standard stochastic gene expression models fail to produce rare coordinated high states entirely, no matter what combinations of networks and parameters are used. In that case, one may argue that the reversible transition into the rare coordinated high state is driven by highly specialized processes (e.g. initiated by a master regulator) or other unknown mechanisms.

Note 1a

Note 1b

Here we describe a mathematical framework to test the hypotheses proposed above for the appearance and disappearance of rare coordinated high states (**Box 1**). Recent studies from our lab suggest that no particular molecular pathway is solely responsible for the formation of these rare cells (Shaffer et al., 2018; Torre et al., 2019). Specifically, in these rare cells, a sequencing and imaging based scheme identified a collection of marker genes, which are targets of multiple signaling pathways ranging from type 1 interferon to PI3K-Akt signaling. The implication is that instead of a single signaling pathway leading to the observed behavior, a network of interacting genes appears to be responsible. Accordingly, we used network modeling to see whether genes interacting within a network were capable of producing transitions to coordinated high expression states. We systematically formulated

and simulated networks of increasing size and complexity defined by a broad range for all independent parameters (Box 1 and 2; and **STAR Methods**, section Networks & section Parameters).

insert before the highlighted period "within the context of our models."

Computational screens on more than 96 million simulated cells reveal that many networks with interactions between genes are capable of producing rare coordinated high states. Critically, transcriptional bursting, a ubiquitous phenomenon in which genes flip between transcriptionally active and inactive state, is necessary ~~for the transcriptional bursting model~~ to produce these rare coordinated high states. Subsequent quantitative analysis shows that rare coordinated high states occur across networks of all sizes investigated (up to 10 nodes), but that (i) they depend on three (out of seven) independent model parameters and (ii) their frequency of occurrence decreases monotonically with increasing network connectivity. The transition into the rare coordinated high state is initiated by a long transcriptional burst, which, in turn, triggers the entry of subsequent genes into the rare coordinated high state. In contrast, the transition out of rare coordinated high state is independent of the duration of transcriptional bursts, rather it happens through the independent inactivation of individual genes. We also confirm model predictions using experimental gene expression data (RNA FISH data) taken from melanoma cell lines. Together, we demonstrate that the standard model of stochastic gene regulation with transcriptional bursting is capable of producing rare coordinated high states~~.~~ **in the absence of additional regulation.**

RESULTS

Framework selection

Identifying the minimal network model generating rare coordinated high states

insert "comprised of only constituitively expressed genes"

We focused on a network-based mathematical framework that models cell-intrinsic biochemical interactions and wondered what would be the minimal set of biochemical reactions that constitutes it. Since network models ~~without gene activation (i.e. constitutive mode of gene expression)~~ were not able to produce rare coordinated high states (see **Supplementary information**; **Figure 1B and Figure S1A-B**; **STAR Methods**, section Model 1), we use a leaky telegraph model as the building block of our framework. In terms of chemical reactions, a gene can reversibly switch between an active (r_{on}) and inactive state (r_{off}), where binding of the transcription factor at a gene locus controls the effective rate of gene production (**Box 1**; **Figure 1C, STAR Methods**). Specifically, when inactive (or unbound), the gene is transcribed as a Poisson process at a low basal rate (r_{prod}); when active, the rate becomes higher ($d \times r_{prod}$, where $d > 1$). We modeled degradation of the gene product as a Poisson process with degradation rate r_{deg}. The inter-node interaction parameter, r_{add}, has a Hill-function-based dependency on the gene product amount (Hill coefficient n) of the respective regulating node to account for the multistep nature of the interaction (**Figure 1C**). In particular, we lump steps leading to transcription by implementing the commonly used quasiequilibrium assumption (Phillips et al., 2019), where binding and unbinding occurs much faster as compared to mRNA transcription and degradation. The dissociation constant k of the Hill function is dependent on the parameters r_{prod}, r_{deg}, and d, such that $k(r_{prod}, r_{deg}, d) = 0.95 * d * r_{prod}/r_{deg}$. In total, the model has seven independent and

one dependent model parameters, as outlined in Box 1. All chemical reactions, propensities, and model parameters are presented in **STAR Methods**. We used Gillespie's Stochastic Simulation Algorithm (Gillespie, 1977) to systematically simulate networks of various sizes and architectures across a broad range of parameters (**Box 1**; **STAR Methods**, section Networks & section Parameters).

Note 2

We limited our study to networks that are symmetric, i.e., networks without a hierarchical structure (**Box 1**; **STAR Methods**, section Networks, **Figure S1C**), a simplification partially supported by the experimental observation that there doesn't seem to be a clear directionality of regulation or hierarchical structure within the highly expressed genes in the rare high state in drug naive melanoma (**Figure S1D**) (Shaffer et al., 2017, 2018). The lack of hierarchy is inferred from the frequency matrix for the experimental RNA FISH data, in which each entry corresponds to the fraction of cells with each gene-pair being highly expressed (**Figure S1D**). Asymmetric networks can result in frequency matrices being highly asymmetric, as demonstrated by an example simulation of a star-shaped reaction network (**Figure S1E-F**). Symmetric models also allow for comparisons of parameters between networks of different sizes. Additionally, we excluded networks that are compositions of independent subnetworks (non weakly-connected networks) and networks that can be formed by structure-preserving bijections of other networks (isomorphic networks) (**STAR Methods,** section Networks, **Box1**). With these operations, we also reduce the testable space of unique networks by several orders of magnitude (**Figure S1C**).

Characterization of the transcriptional bursting model

When genes are organized in the system described above and simulated over long intervals, the transcriptional bursting model produced a range of temporal profiles for gene products (**Figure 1D-G** and **Figure S2A**). ~~Importantly,~~ the model was able to faithfully capture the qualitative features of experimental data, i.e., rare, transient, and coordinated high expression states (**Figure 1G**). We defined a set of rules to screen for the occurrence of different classes of states (**Figure 1D-G** and **Figure S2A**); these include stably low expression (class I), stably high expression (class II), uncoordinated transient high expression (class III), and rare transient coordinated high expression (class IV) (see **STAR Methods**, section Simulation classes), and used a heuristic approach to distinguish between these different classes (**Boxes 1** and **2**). For a detailed description of the rules and quantitative metrics used to define class IV, see **Boxes 1** and **2**; **Figure S3** and **Figure S4**; and **STAR Methods**, section Simulation classes.

To better compare the computational results with the experimental data from static RNA FISH images, we split the entire simulation into non-overlapping time interval of 1000 time units, as justified by the ergodic theory (**Box 2** and **STAR Methods**). We took snapshots of gene products at randomly selected time points in these time-intervals and noted the number of simultaneously highly expressed genes as well as their gene product counts, allowing us to represent the static states of a population of simulated cells (**Figure 2A**). For example, in a particular 8-node network, we found that the distribution qualitatively captures the experimental observations where most cells do not exhibit high expression states, while some cells are in a high state for one or more genes **(Figure 2B)**. Note that unlike the

experimental data, the model simulation does not have non-zero values for higher number of genes. The absence of of non-zero values may be because the network underlying the experimental data contains a much larger set of interacting genes, thereby increasing the likelihood of non-zero values for higher number of expressed genes. Similarly, when we selected a gene and plotted its product count for the randomly selected time points, we observed a heavy-tailed distribution (**Figure 2C, right panel**), similar to the experimental observations **(Figure 2C left panel** and **Figure S3A)**. Furthermore, these observations, while shown for a particular 8-node network, hold true for simulations of other 8-node networks as well as networks of other sizes (**Figure S2B**). Note that the distributions of gene product counts for each gene are qualitatively similar because of the symmetric nature of the networks (**Figure S2C**). The experimental data in drug naive melanoma cells for mRNA counts display different degrees of skewness of the distribution for different genes (**Figure S3A**) which may be recapitulated by introducing asymmetries in the networks. Two asymmetric networks we tested were both able to produce rare coordinated high states (**Figure S2G-S4M**). Importantly, the distributions of gene product counts for various genes displayed different levels of heavy-tails, as also observed in the experimental data (**Figure S2M**). Since there is both inter- and intra-gene variability between the experimental expression distributions characterizing these states (**Figure S3A**), we compared these expression distributions to simulated expression distributions using Gini coefficients, used to characterize experimental expression distributions in the original study (Shaffer et al, 2017). While the Gini coefficient is low for most of the simulations (99.2%, gray), it is much higher for the simulations that produce rare coordinated high states (red) and overlaps with experimental Gini coefficients (**Figure S3D**). Together, the transcriptional bursting model is able to produce states which recapitulate key aspects of rare coordinated high states observed in drug naive melanoma.

Note 4a

Rare coordinated high states depend on network topologies and model parameters

Since the rare coordinated high states occur in <1% of all simulations (**Figure S2A**), we wondered whether their occurrence depends on the network topologies and/or model parameters. Specifically, what are the features of the topologies and parameters that facilitate the occurrence of rare coordinated high states? For the simulations that produced rare coordinated high states, we extracted and quantitatively analyzed the corresponding networks. We found that the rare coordinated high states occur ubiquitously in networks with different numbers of nodes analyzed (up to 10 nodes) (**Figure 2D and Figure S2B-F, Figure S5A-B**). ~~We therefore hypothesize that even larger networks may also display rare coordinated high states, and can be explored in future studies.~~ ~~Next, we wondered if the occurrence of rare coordinated high states depends on the network connectivity (**Box 1**). Indeed,~~ within a particular network size, the ability to produce rare coordinated high states decreases monotonically with increasing network connectivity (**Figure 2E and Figure S5C-D**). Consistently, the fraction of networks per network size (normalized by either network size or total networks per network size) exhibiting rare coordinated high states decreases with increasing size (**Figure S5A-B**) as a larger fraction of high connectivity networks exist in bigger networks (**Figure S5D**).

Note 4b

We next wondered whether gene auto-regulation (networks with self-loops) have any effect on a networks ability to produce the rare coordinated high states. ~~Indeed we found that for a fixed size and connectivity, networks with auto-regulation result in higher numbers of simulations with rare coordinated high states than networks without auto-regulation (**Figure S5E**). At the same time, we wondered whether connectivity or auto-regulation has a more dominating effect on a networks' ability to produce these states.~~ We found that adding self-loops on otherwise identical networks reduced the occurrence number of simulations with rare coordinated high states (**Figure 2F**), ~~demonstrating the stronger effect of connectivity than auto-regulation. Finally,~~ also we analyzed network topologies based on characteristic length, defined as the average shortest path length between pairs of nodes of the network (see **STAR Methods, Box1**). Characteristic length recapitulates the effects of not only network connectivity (inversely correlated with characteristic distance), but also differentiates topologies with the same connectivity (**Figure S5F)**, for example networks with or without auto-regulation. Together, we demonstrate that the occurrence of rare coordinated high states depends on network topologies.

Note 5

Note 6

Insert a sentence summarizing the conclusions drawn from S5F at the highlighted period.

Since the transcriptional bursting model has seven independent parameters (r_{on}, r_{off}, r_{prod}, r_{add}, r_{add}, d, and n; see **Box 1** for details), we asked whether specific parameter combinations preferentially give rise to the rare coordinated high states, and if so, what features of such combinations facilitate it. The subsequent analysis is motivated by the initial observation that occurence of different classes of temporal gene product profiles across different network sizes and connectivities appear to also depend on the parameter sets (**Figure 2G**). Specifically, if a parameter set gave a specific expression profile (e.g. rare coordinated high or stably high) for one network, it displayed a higher propensity to display the same profile for other networks as well (**Figure 2G** and **Figure S3E**), implying that parameters indeed play a major role in the occurrence of rare coordinated high states. To avoid biases in the parameter sets investigated, all 1,000 parameter sets were sampled from a broad range for each parameter using a Latin Hypercube Sampling algorithm (**Supplementary Information** ParSetsAnalysis.xlsx**; STAR Methods**, section Parameters).

We first measured the percentage of simulations per parameter set that gave rise to the rare coordinated high states. Out of the 1,000 parameter sets, eight parameter sets, from now on called rare coordinated high parameter sets (**Box 2**), clustered together at the tail-end of the distribution (orange, **Figure 2H**), meaning they generated simulations with frequent occurrence of rare coordinated high states in at least 20% of all networks tested (**Figure 2H**). Furthermore, these eight parameter sets robustly generated rare coordinated high states across all network sizes and architectures (**Figure S6A**). Therefore, we wondered if these eight parameter sets have any special or distinguishing features compared to the remaining 992 parameter sets.

We used a decision tree algorithm (Breiman et al., 1984) (see **STAR Methods**, section Decision tree optimization and generalized linear models) to identify the differentiating features of the rare coordinated high parameter sets from the rest. The decision tree analysis revealed that only three (r_{on}, r_{off}, and r_{add}) of the seven independent parameters showed a strong correlation with the rare coordinated high parameter sets (**Figure 2I**). We validated these findings with complementary analysis using generalized linear models

(**STAR Methods**, section Decision tree optimization and generalized linear models) where we found precisely these three specific parameters (r_{on}, r_{off}, and r_{add}) to be critical to produce the rare coordinated high states with high statistical significance (p values: r_{on} = 0.003; r_{off} = 0.005; r_{add} = 0.014) (**Figure S6B**). These observations became readily evident when we plotted all the 1,000 parameter sets for r_{on}, r_{off}, and r_{add} together and found the rare coordinated high parameters sets to occupy a narrow region of the parameter phase space (**Figure 2J** and **Figure S6C**). ~~Furthermore, parameter~~ sensitivity analysis across the parameter space also confirmed that these three parameters are indeed critical for producing the rare coordinated high states (**Figure S6D**). These three parameters are related to transcriptional bursting and inter-gene(node) regulation. Two of these parameters, r_{on} and r_{off}, define the transitioning between the active and inactive state of the DNA respectively. The third parameter is the gene activation rate, r_{add}, which corresponds to the positive regulation of transcriptional bursting rate of a gene by the gene product of another interacting gene. ~~Interestingly,~~ *** too high values (> 0.31) of r_{add} result in the disappearance of rare coordinated high states, as does a complete absence (r_{add} = 0) of this term (**Figure S6E-S6G**). To confirm that these three parameters (r_{on}, r_{off}, and r_{add}) and their corresponding range of values are indeed critical to producing simulations with rare coordinated high states, we sampled new 1,000 parameter sets from a constrained region containing all eight rare coordinated high parameter sets (**Figure 2J, orange box,** and **STAR Methods**) and ran simulations for two test networks, a 3-node and a 5-node network. We found that the frequency of simulations with rare coordinated high states for the constrained region is ~14-fold and ~21-fold higher than that for the original parameter space, respectively (**Figure 2K**). We note that while parameter sets with parameters r_{on}, r_{off}, and r_{add} within the identified critical parameter ranges give rise to simulations with rare coordinated high states much more frequently than other parameter sets, it is not 100% of the time.

Move to ***

Distinct mechanisms regulate the transition into and out of rare coordinated high states

We have identified the networks and parameter sets for which the transcriptional bursting model exhibits rare coordinated high states more frequently. Next, we ~~wondered if we could dissect~~ dissected the features of the model that facilitate the occurrence of rare coordinated high states. Specifically, we ~~wanted to know~~ indentified the factors that 1) trigger the entry into the rare coordinated high states, 2) facilitate its maintenance, and 3) trigger the escape from it. We began by analyzing various features of transcriptional activity, since including transcriptional bursting was found to be critical for the model to display the rare coordinated high states. These include the burst fraction, length of transcriptional bursts (burst duration) and burst frequency. To measure these features, we defined four regions for each simulation: low expression state (baseline time-region), entry into the high expression state (entry time-point), the high expression state (high time-region), and exit from the high expression state (exit time-region) (**Figure 3A, STAR Methods**, section Entry and Exit mechanisms).

We found an increase in the transcriptional activity, as measured by the burst fraction, during the high expression time-region as compared to the baseline time-region (**Figure 3B**). ~~Increased burst fraction could be a result of (1) longer transcriptional bursts or (2) a higher burst frequency. The former is not possible as the duration of each burst is distributed~~

Note 7

exponentially according to exp(r_{off}), which does not change between the baseline and high time-region. Indeed, we found an increase in the burst frequency in high time-region, thus establishing its role in the maintenance of the rare coordinated high state (**Figure 3C**). The increased transcriptional bursting seen in the ~~transcriptional bursting model~~ replace with "models capable of generating rare coordinated high states" is consistent with the experimental ~~data using labeled nascent transcripts which showed~~ observations that the transcriptional activity occurred in frequent bursts in cells high for a marker gene (Shaffer et al., 2018). Next we wondered whether ~~the increase in~~ burst frequency ~~is promoted by~~ increases with the interactions of genes ~~organized~~ within the network. We compared two networks of the same size (3 nodes), where one is comprised out of single unconnected (orphan) nodes and the other of an interdependent structure (network 3.2). We found that for any parameter set (screened for all 26 parameter sets giving simulations with rare coordinated high states in the previous analysis for network 3.2, **Supplementary Information**), the system with a connected network has (1) more high expression states and (2) prolonged time in high expression states, as compared to unconnected nodes (**Figure 3D**). Together, we find that the maintenance in the high state is because of increased burst frequency, ~~which may be a result of the positive regulatory feedback intrinsic to the networks.~~ too speculative.

Next, we wanted to identify the factors triggering the entry into the rare coordinated high states. We found that for any gene in the network, the transcriptional burst duration right before/during the entry into a rare coordinated high state was significantly higher than that in the baseline time-region (i.e., regular bursting kinetics) (**Figure 3E**). In the example shown in **Figure 3E**, the average time of transcriptional burst at the entry time-point is 84.82 (time units) as compared to only 15.08 (time units) in the baseline time-region. Therefore, prolonged transcriptional bursts play a role in driving the cell to a coordinated high expression state. Conversely, we asked if the opposite is true at the exit time-region, such that transcriptional bursts for the exit time-region are shorter than for the high time-region. We found no difference in the distributions of burst durations between the high and the exit time-regions, as demonstrated by the example in **Figure 3F**, suggesting that the exit from high expression state occurs independently of the burst durations. ~~Importantly,~~ both of these conclusions hold true when measured for all simulations with rare coordinated high states (**Figure 3G**). Together, unlike the entry into the high time-region, the exit from it is not dependent on the transcriptional burst duration. Note 8

We also wondered if the entry into the high expression state of one gene influences the entry of other genes, or that the genes enter the high expression state independently of each other. We reasoned that if the time duration between two successive genes (t_{ent}, **Figure 3A**) entering the high expression state is exponentially distributed, it would imply that the genes enter the high expression state independent of each other. Instead, we found that the distributions of entry time intervals rejected the null-hypothesis of the Lilliefors' test for most of the simulations (84%), meaning they are not exponentially distributed (**Figure 3H**). The remaining 16% of cases were found to be largely falsely identified as exponentially distributed due to limited data (see a representative example in **Figure S7A)**. Similarly, we tested if the exit for successive genes from the high expression state occurs independent of each other. ~~Interestingly,~~ contrary to the situation during the entry into the high expression state, many distributions of exit time intervals satisfied the null-hypothesis of the Lilliefors' test, implying they are indistinguishable from exponential distributions (**Figure 3I**). The

simulations that did not satisfy the stringent Lilliefors' test mainly appear to be exponentially distributed nevertheless; a representative example is shown in **Figure S7B**. Together, the entry into and exit from the rare coordinated high state occur through fundamentally different mechanisms—the entry of one gene into the high expression state affects the entry of next gene, while they exit from it largely independently of each other. The exit from the high state could be a result of weak strength of coupling (as reflected by the moderate values of parameter r_{add}) between nodes for the simulations that produce these states. Consistently, we found that too high values of r_{add} results in the disappearance of rare coordinated high states, giving way to stable high states. In other words, the network can transition into the high expression state but loses the ability to come out of it (**Figure S6 E-G**).

Increasing network connectivity leads to transcriptionally stable states

So far, we have used the transcriptional bursting model to understand the potential origins of rare pre-resistant states in drug naive melanoma cells. Upon treatment with anti-cancer drugs, the transient pre-resistant cells reprogram and acquire resistance resulting in their uncontrolled proliferation. The resistant cells are characterized by the stabilization of the high expression of the marker genes which were transiently high in the drug naive pre-resistant cells (**Figure 4A**) (Shaffer et al., 2017). Studies using network inference of gene expression data have suggested that the genetic networks undergo significant rearrangements upon cellular transitions or reprogramming (Moignard et al., 2015; Schlauch et al., 2017). We wondered if the transcriptional bursting model can explain how the transient high expression in drug naive cells might become permanent upon treatment with anti-cancer drugs. The modeling framework produces a range of gene expression profiles, depending on the network properties and model parameters (**Figure 1D-G**). Increasing the network connectivity (for fixed parameter sets) is one way to shift from a rare transient coordinated high expression state to stably high expression state (**Figure 4B-E**). As an example, for a fixed network size (five) and associated parameters, increasing the network connectivity from one to five resulted in a shift from transient coordinated to stably high expression states (**Figure 4D** and **Figure 4E**). The shift from transient coordinated to stably high expression states is also reflected by the bimodal distribution of genes product counts for in the highly connected network (**Figure 4F** and **Figure 4G**), where genes stay permanently in the high state once they leave the low expression state. These results mimic the experimentally measured mRNA expression states of the drug-induced reprogrammed melanoma cells.

To test if the computational prediction holds true in melanoma, we performed network inference using φ-mixing coefficient-based (Ibragimov, 1962) Phixer algorithm (Singh et al., 2018) on the experimental data (**Box 2**; **STAR Methods**, section Comparative Network Inference). Specifically, we used the Phixer algorithm on the mRNA counts obtained from fluorescent in situ hybridization (FISH) imaging data of marker genes in drug naive cells and the resistant colonies that emerge post-drug treatment to infer the underlying network. Consistent with the model prediction, we found that the number of edge connections (for a range of edge weight thresholds) between marker genes increased substantially for 6/7 resistant colonies compared to the drug-naive cells (**Figure 4H**). To control for biases from subsampling of the experimental data and the nature of Phixer algorithm itself (see **STAR Methods**, section Comparative Network Inference), we ran the entire network inference

analysis 1,000 times. Again, in all 1,000 runs, we saw a higher number of total edges for 6/7 resistant colonies compared to the drug-naive cells (**Figure 4I, Figure S8A** and **Figure S8C**).

Besides the dependence on networks, our framework predicts that for a given network, stronger interactions between nodes (defined by the interaction parameter r_{add}) can also result in stable gene expression profiles (**Figure S6E-S6G**). It is possible that reprogramming results from a combination of increased edge connectivity as well as the enhanced interactions (given by parameter r_{add}) between existing edges. Biologically, it would translate into stronger and increased number of interactions between genes and associated transcription factors during reprogramming. Together, network inference of the experimental data is consistent with model findings about the cellular progression from a transient coordinated high expression state to a stably high expression state.

DISCUSSION

We developed a computational framework to model rare cell behaviors in the context of a drug naive melanoma population where a rare subpopulation of cells displays transient and coordinated high gene expression states. We found that a relatively parsimonious stochastic model consisting of transcriptional bursting and stochastic interactions between genes in a network is capable of producing rare coordinated high states that mimic the experimental observations. To systematically investigate their origins, we screened networks of increasing sizes and connectivities for a broad range of parameter values. Our study revealed that they occur more frequently for networks with low connectivity and depend on ~~5/7~~ 3 of the 7 independent model parameters. Furthermore, we showed that the mechanisms that lead to the transition into- and out of- the rare coordinated high state are fundamentally different from each other. Collectively, our framework provides an excellent basis for further mechanistic and quantitative studies of the origins of rare, transient, and coordinated high expression states.

Given the relative generality of the networks that produce rare coordinated high states, the transcriptional bursting model predicts that every cell type is capable of entering the rare coordinated high state. Furthermore, we show that canonical modes of transcription alone, namely the binding of the transcription factor at gene locus to produce mRNA via recruitment of RNA Polymerase II, can lead to these states without requiring other complex mechanisms such as DNA methylation, histone modifications, or phase separation. While such other mechanisms may still be operational in these cells to regulate their entry to or exit from these states, we posit that in principle, any set of genes interacting *via* traditional gene regulatory mechanisms are capable of exhibiting these rare coordinated high states, as long as they are interacting in a certain manner (e.g. sparsely connected) with appropriate kinetic parameters. In the case of drug naive melanoma cells, the transient state is characterized by an increased ability to survive drug therapy leading to uncontrolled proliferation of the resulting resistant cells. It is possible that these rare transient behaviors may exist across many sets of interacting genes which may or may not manifest into phenotypic consequences. Another possibility the transcriptional bursting model predicts is that even within the same cell, distinct modules of interacting genes can lead to distinct sets of rare coordinated high states that each can affect the cellular function and outcomes differently.

These possibilities can be tested for by using increasingly accessible single cell RNA sequencing techniques on clonal population of cells.

~~The transcriptional bursting model also makes two key predictions regarding transitions into and out of rare coordinated high states. The first is that prolonged transcriptional bursts drive entry into the high expression state while exit from it is independent of the burst duration. The second is that genes entering the high expression state promote the entry of subsequent genes, whereas genes exiting the high expression state do so independently of each other. Both these predictions can be readily tested experimentally by simultaneous visualization of transcriptional bursting and mRNA counts using live cell (e.g. by using RNA-binding fluorescent proteins) or fixed cell (intron and exon FISH) imaging approaches (Bartman et al., 2016; Rodriguez et al., 2019).~~

~~Additionally, we showed that increasing the network connectivity is one way to reach a drug-induced reprogrammed state, a prediction we verified by performing network inference analysis on the experimental data. The transcriptional bursting model proposes that there are many plausible ways to transition from networks that produce transient coordinated high expression states to stable high expression states. For example, the transition could be facilitated by different amounts of increases in connectivity between nodes (genes) and/or changes in parameter values of the gene expression model. Furthermore, it is possible that these changes may take place only for a subset of nodes and edges within the network. These computational scenarios suggest that there could be significant heterogeneity in the stable expression levels of network genes in the resistant colonies emerging even from clonal population of drug naive cells, a possibility that can be tested experimentally by isolating individual colonies and profiling them for molecular markers to identify the paths. Identification of dominant paths has relevance for rational targeted drug therapy design. Therefore, in addition to modeling rare-behaviors, our framework can be adapted for investigating the plasticity and reprogramming paradigm in cancer.~~

One limitation of the transcriptional bursting model is that we have performed quantitative analysis only on symmetric networks with positive interactions between nodes. While the preliminary analysis on two cases of randomly selected asymmetric networks shows that they do exhibit the rare coordinated high states (**Figure S2G-S4M**), it remains to be seen whether these findings hold more generally for asymmetric networks. Inhibitory interactions between nodes is a separate and perhaps more interesting point. In principle, the model can be adapted to include inhibitory interactions. These inhibitory interactions may lead to non-monotonic effects of network connectivity on the occurrence of rare states, as positive and negative interactions can compete in non-linear ways. Similarly, a network with both negative and positive interactions may be more prone to instability, even for relatively smaller networks. Furthermore, inclusion of these interactions might also make the exit of genes from the high expression state dependent on one another, which occurs independently in the transcriptional bursting current model.

While we have focused on rare, transient, and coordinated high expression states in melanoma, our study provides conceptual insights into other biological contexts such as stem cell reprogramming. Particularly, there is increasing evidence to suggest that stem cell

reprogramming to desired cellular states proceeds *via* non-genetic mechanisms in a very rare subset of cells (Hanna et al., 2009; Pour et al., 2015; Takahashi and Yamanaka, 2016). The transcriptional bursting model may explain the origins and transient nature of this type of rare cell variability. In sum, we have established the plausibility that a relatively parsimonious model comprising of transcriptional bursting and stochastic interactions of genes organized within a network can give rise to a new class of biological heterogeneities. Therefore, we believe that established principles of transcription and gene expression dynamics may be sufficient to explain the extreme heterogeneities that are being reported increasingly in a variety of biological contexts.

SUPPLEMENTAL INFORMATION

Supplemental Information includes 9 figures and 2 tables.

ACKNOWLEDGEMENTS

We thank the Raj lab members, especially Ian Mellis and Amy Azaria, for scientific discussions and comments on the manuscript. We also thank Ravi Radhakrishnan and Alok Ghosh for helpful discussion during the initial stages of this project. We thank Cesar A Vargas-Garcia for his help during the initial discussions on network inference. A.R. acknowledges NIH/NCI PSOC award number U54 CA193417, NSF CAREER 1350601, P30 CA016520, SPORE P50 CA174523, NIH U01 CA227550, NIH 4DN U01 HL129998, NIH Center for Photogenomics (RM1 HG007743), and the Tara Miller Foundation. C.M. acknowledges support from the Deutsche Forschungsgemeinschaft DFG through the SFB 1243. A.S. acknowledges support from the NIH grant 5R01GM124446-02. L.S. would like to acknowledge the support of the PROMOS fellowship of the DAAD, Germany. Y.G. would like to acknowledge the Schmidt Science Fellows in partnership with the Rhodes Trust. Y.G. is a fellow of The Jane Coffin Childs Memorial Fund for Medical Research and this investigation has been aided by a grant from The Jane Coffin Childs Memorial Fund for Medical Research.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.G. and A.R.; Methodology, L.S., A.R. and Y.G.; Software, L.S. and A.R.; Validation, L.S.; Formal Analysis, L.S. and M.S.A.; Resources, A.R. and A.S.; Investigation, B.E., E.M.S. and Y.G.; Data Curation, L.S. and Y.G.; Writing - Original Draft, Y.G.; Writing - Review & Editing, A.R., L.S., Y.G., C.M., E.M.S., B.E. and M.S.A.; Visualization, L.S. and Y.G.; Supervision, Y.G., A.R. and C.M.; Project Administration, Y.G. and A.R.; Funding Acquisition, A.R., A.S. and C.M.

DECLARATION OF INTERESTS

A.R. receives royalties related to Stellaris RNA FISH probes. All other authors declare no conflict of interests.

MAIN FIGURE CAPTIONS

Important: the green highlighted text is problematic as noted in my email. Please fix.

Figure 1. A transcriptional bursting model is able to mimic the rare coordinated high states observed in drug naive melanoma.
(A) Drug naive melanoma cells exist in low (white cells) as well as rare coordinated high (blue cells) states. Cells in the rare coordinated high state characterize the pre-resistant state observed in drug naive melanoma. A schematic of the corresponding expression pattern is shown in the panel below. The cells in a high expression state are more likely to survive and acquire resistance upon drug administration.
(B) Schematic of the constitutive model for two nodes. Gene product is either produced at rate r_{prod} or degraded with rate r_{deg}. Gene regulation is modeled by a Hill function, where the gene product count of the regulating gene A increases the production rate of the gene product of the regulated gene B.
(C) Schematic of the transcriptional bursting model for two nodes. DNA is either in an inactive (off) or active (on) state. Transitions take place with rates r_{on} and r_{off}, where gene product is synthesized with rates r_{pod} and $d*r_{prod}$, respectively, $d>1$. Gene product degrades with rate r_{deg}. Gene regulation is modeled by a Hill function, where the gene expression of the regulating gene A increases the activation of the DNA of the regulated gene B.
(D-G) Depending on the network and the parameters of the transcriptional bursting model, we observe stably low expression (D), stably high expression (E), uncoordinated transient high expression (F) and rare transient coordinated high expression (G).
See also Figure S1.

Figure 2. Simulations of the transcriptional bursting model show similar behavior at the population level as the drug naive melanoma cells.
(A) Frame of simulation showing rare coordinated high state (shaded area). The 1,000,000 time unit simulation is split into frames of 1,000 time units to create a simulated cell population (shown for cell N). For a randomly determined time-point t_{rand}, the number of simultaneously highly expressed genes and the gene count per gene per cell are evaluated. The network of the corresponding simulation is given in the top left corner.
(B,C) The simulated number of simultaneously highly expressed genes and expression distribution at the population level are qualitatively similar to experimental data from a drug naive melanoma population (data from (Shaffer et al., 2017)). The percentages are indicated above the histogram (in B). The network and parameter set as well as the particular node (in C) used for comparison are shown in the right panel.
(D) Rare coordinated high states occur ubiquitously across networks of all analyzed network sizes (for three independent and randomly sampled t_{rand})(median, 25th and 75th percentiles).
(E) Increasing connectivity within all networks of size 5 leads to a decrease in the number of simulations with rare coordinated high states.
(F) Networks with auto-regulation exhibit rare coordinated high states less frequently than the same networks without auto-regulation.
(G) Simulations of a particular parameter set across different networks and sizes show largely the same class of gene expression profiles.
(H) The eight rare coordinated high parameter sets give rise to simulations with rare coordinated high states more frequently than others for all 96 networks and cluster at the tail of the histogram. The cut-off is defined at 20%.

(I) Decision tree optimization of resulting parameter set classes reveal that only three out of seven parameters, r_{on}, r_{off} , and r_{add}, show a strong correlation with the rare coordinated high parameter sets.
(J) Three dimensional representation of all tested 1000 parameter sets for r_{on}, r_{off} , and r_{add} show that the rare coordinated high parameter sets are narrowly constrained in the 3D space (orange dots). The orange box indicates the constrained parameter space enclosing all rare coordinated high parameter sets used for analysis in (K).
(K) The constrained subregion defined by the rare coordinated high parameter sets (see Figure 2J) heavily favors the formation of rare coordinated high states.
See also Figure S2, Figure S3, Figure S4, Figure S5 and Figure S6.

Figure 3. Rare coordinated high state is initiated by a long transcriptional burst, maintained by an increase in burst frequency and terminated according to a random process.
(A) An exemplary high region, with a baseline time-region, entry time-point, high time-region and an exit time-region. The time intervals for an additional gene to enter and exit the high region are marked by t_{ent} and t_{exit}, respectively. The bursts below the exemplary simulation are representative schematics.
(B) The system exhibits enhanced transcriptional activity, as measured by the burst fraction, when in the high time-region (two-sample Kolmogorov-Smirnov test, p-value < 0.001).
(C) The frequency of transcriptional bursts is increased in the high time-region (two-sample Kolmogorov-Smirnov test, p-value < 0.001).
(D) Positive regulatory interactions between the connected nodes (network) leads to an increased number of and total time in high states in comparison to independent nodes.
(E) The bursts during entry time-points are significantly longer than bursts not in a high time-region (two-sample Kolmogorov-Smirnov test).
(F) There is no statistical significant difference between the distributions underlying the duration of bursts in the high time-region and the exit time-region (two-sample Kolmogorov-Smirnov test).
(G) While the mean burst duration ratio between entry time-point and baseline time-region is significantly increased, the mean burst duration ratio between bursts in the exit time-region and in the rest of the high time-region are comparable for all simulations with rare coordinated high states. Ratio close to 1 suggests no difference between the two regions.
(H,I) The time intervals between genes entering and exiting the high time-region are distributed differently, as demonstrated by two representative simulations. While the time intervals for entering the high time-region are not exponentially distributed (H) (and hence not random), the time intervals for exiting the high time-region are exponentially distributed (I) (Lilliefors test, p-value < 0.001 and > 0.05, respectively).
See also Figure S7.

Figure 4. Increased connectivity of a network leads to stable high expression which is also observed in emerging resistant colonies post-drug treatment.
(A) Upon drug treatment, the surviving cells acquire stable resistance. A schematic gene expression pattern is shown below.
(B,C) Networks of size 5 with low (B) (1) and high (C) (5) connectivity and corresponding (D,E) simulations.

(F,G) The expression distribution underlying the simulation of the highly connected network (G) does not exhibit heavy-tails while the simulation of the lowly connected network (F) exhibits heavy-tails.
(H) Network inference analysis shows that 6 out of 7 resistant colonies (post drug treatment) have higher connectivity in comparison to two biological replicates of drug naive cells for many edge weight thresholds.
(I) Distribution of number of edges for the drug naive cells (red) is lower than an exemplary resistant colony (black) when the network inference analysis is run 1,000 times.
See also Figure S8.